

# A Stochastic Theory of Adiabatic Invariance

R. Cogburn and J.A. Ellison\*

Department of Mathematics and Statistics, The University of New Mexico, Albuquerque, New Mexico 87131, USA

Received August 19, 1991; in revised form February 21, 1992

**Abstract.** Let  $I$  be a set of invariants for a system of differential equations with an order  $o(\varepsilon)$  vector field. When order  $\varepsilon$  perturbations of zero mean are added to the system we show that, under suitable regularity and ergodicity conditions,  $I$  becomes an adiabatic invariant with maximal variations of order one on time scales of order  $1/\varepsilon^2$ . In the stochastically perturbed case,  $I$  behaves asymptotically (for small  $\varepsilon$ ) like a diffusion process on  $1/\varepsilon^2$  time scales. The results also apply to an interesting class of deterministic perturbations. This study extends the results of Khas'minskii on stochastically averaged systems, as well as some of the deterministic methods of averaging, to such invariants.

## 0. Introduction

We consider the behavior of the stochastic differential equation in  $R^p$ ,

$$\dot{x} = \varepsilon F(x, t, \varepsilon, \omega), \quad x(0) = x_0, \quad (0.1)$$

where

$$F(x, t, \varepsilon, \omega) = f(x, t) + F^{(0)}(x, t, \omega) + \varepsilon F^{(1)}(x, t, \omega) + o(\varepsilon) \quad (0.2)$$

as  $\varepsilon \rightarrow 0$ . We require that  $EF^{(0)}(x, t) = 0$  and that

$$\bar{f}(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(x, s) ds \quad (0.3)$$

exists for all  $x$ . Khas'minskii [13] examined these systems at time scales of  $O(1/\varepsilon)$  and gave general conditions for the asymptotic approximation of solutions of (0.1) by solutions of the deterministic equations

$$\dot{x} = \varepsilon \bar{f}(x) \quad (0.4)$$

---

\* Supported by NSF grant DMR-8704348

as well as a Markov correction to this approximation. Then in [14] Khas'minskii considers the same systems at time scales of  $O(1/\varepsilon^2)$  when the averaged problem is degenerate (i.e.,  $\bar{f}(x) = 0$ ). A diffusion process is obtained in the scaled limit to (0.1) under suitable regularity conditions.

The unperturbed problem (0.4) is slowly varying. Knapp, Papanicolaou and White [15] discuss the case of a perturbed Hamiltonian system, where the unperturbed problem,  $\dot{I} = 0, \dot{\phi} = \omega(I)$ , is an  $O(1)$  vector field in action angle-variables. It should also be noted that the change of scale  $\sigma = \varepsilon t$  transforms (0.1) into

$$\frac{dx}{d\sigma} = F\left(x, \frac{\sigma}{\varepsilon}, \varepsilon, \omega\right), \quad x(0) = x_0 \tag{0.5}$$

in which the perturbation appears as an  $O(1)$  rapidly varying field.

In this paper, we will show that similar results can be obtained at  $O(1/\varepsilon^2)$ -times even when  $\bar{f} \neq 0$  provided we consider a suitable collection of invariants  $I = I(x)$  of the unperturbed problem (0.4). That is, we require

$$\frac{\partial I}{\partial x} \bar{f}(x) = 0, \tag{0.6}$$

which entails that  $I(x(t))$  is constant along solutions of (0.4). If  $\bar{f} = 0$ , then we can take  $I(x) = x$  and we recover Khas'minskii's result at a certain stage of our analysis. Our goal is to study the evolution of  $y(t, \varepsilon) = I(x(t, \varepsilon))$  under the flow defined by (0.1) on  $O(1/\varepsilon^2)$  time intervals.

Clearly, the analog of (0.1) for  $y$  is

$$\dot{y} = \frac{\partial I}{\partial x} \dot{x} = \varepsilon G(x, t, \varepsilon, \omega), \quad y(0) = I(x_0), \tag{0.7}$$

where

$$G = g + G^{(0)} + \varepsilon G^{(1)} + o(\varepsilon),$$

and

$$g(x, t) = \frac{\partial I}{\partial x} (f(x, t) - \bar{f}(x)),$$

$$G^{(i)}(x, t, \omega) = \frac{\partial I}{\partial x} F^{(i)}(x, t, \omega), \quad i = 1, 2. \tag{0.8}$$

Note that the time average  $\bar{g}$  of  $g$  is zero, so the analog of (0.4) for  $y$  is  $\dot{y} = 0$ , or  $y = \text{constant}$ .

Equation (0.7) is similar to the equations treated in [14], except that  $G$  depends on  $x$ , not  $y$ . Under conditions similar to those in [14] analogous estimates can be obtained for increments of the solutions to (0.7). The  $y$ -increments depend on  $x$ , however under the assumption that the deterministic solutions to (0.4) behave ergodically on the surfaces corresponding to  $I(x) = \text{constant}$  the increments are shown, asymptotically as  $\varepsilon \rightarrow 0$ , to depend on  $x$  only through  $y$ . In the final step of the analysis, convergence to a diffusion process in  $y$  is obtained.

In Sect. 1, we introduce our hypotheses, state the main theorem and discuss some related points. In Sect. 2, we discuss several examples which illustrate the

theorem. Preliminary estimates are presented in Sect. 3, and these are gathered together in Sect. 4 to prove the theorem.

### 1. Formulation of the Main Results

We will consider the solution to the  $x$ -equation (0.1) in an (open, connected) domain  $D_0$ , possibly all of  $R^p$ . Let  $I = I(x): R^p \rightarrow R^q$  be a function satisfying (0.6), and let  $D_1 = I(D_0)$ . We require  $D_0$  to be chosen so  $D_1$  is open in  $R^q$ . We assume that the vector field  $F$  is bounded, continuous and locally Lipschitz in  $x$  for almost every  $\omega$ , and this implies ([11]) that the solution,  $x(t, \varepsilon)$ , to (0.1) exists uniquely in  $D_0$  on a maximal right interval  $[0, B_\varepsilon(\omega))$  and that either  $B_\varepsilon = \infty$  (i.e., the solution stays in  $D_0$  for all  $t \geq 0$ ) or that  $x(t, \varepsilon) \rightarrow x_\varepsilon^B \in \partial D_0$  as  $t \uparrow B_\varepsilon < \infty$ . This allows us to define the process on  $[0, \infty)$  as

$$\hat{x}(t, \varepsilon) = \begin{cases} x(t, \varepsilon) & 0 \leq t < B_\varepsilon \\ x_\varepsilon^B & B_\varepsilon \leq t < \infty \end{cases}$$

and let  $\hat{y}(t, \varepsilon) = I(\hat{x}(t, \varepsilon))$ .

In what follows,  $(\Omega, \mathcal{F}, P)$  is a probability space and for each  $x \in D_0, t \geq 0$  and  $\varepsilon > 0$  the functions  $F, F^{(0)}$  and  $F^{(1)}$  are  $R^p$  valued random variables on  $\Omega$ . In the statement of the assumptions, we use the conventions that, for a function  $h: D_0 \times [0, \infty) \times \Omega \rightarrow R^r$  or  $h: D_0 \times [0, \infty) \times \Omega \rightarrow C^r$  ( $r$ -dimensional complex space), letting  $|h| = \sum_{i=1}^r |h^{(i)}|$ ,

$$\|h\| = P - \text{ess sup } \sup_{x \in D_0} \sup_{t \geq 0} |h(x, t, \omega)|,$$

$$\|h\|_1 = P - \text{ess sup } \sup_{\substack{x_1, x_2 \in D_0 \\ x_1 \neq x_2}} \sup_{t \geq 0} |h(x_1, t, \omega) - h(x_2, t, \omega)| / |x_1 - x_2| + \|h\|,$$

$$\|h\|_2 = \sup_{1 \leq i \leq p} \|\partial h / \partial x_i\|_1 + \|h\|,$$

and use the same conventions when  $h$  depends on a subset of these arguments. Note that

$$\|h\| \leq \|h\|_1 \leq \|h\|_2.$$

Now, consider the assumptions:

(A1) Relations (0.1) and (0.2) hold, where, in relation (0.2) and what follows, all order statements for the limit as  $\varepsilon \rightarrow 0$  are to hold uniformly in  $x \in D_0, t \geq 0$  and  $\omega \in \Omega_\varepsilon$  with  $P\Omega_\varepsilon = 1$ . Moreover, we assume that  $F, F^{(0)}$  and  $F^{(1)}$  are continuous functions in  $x, t$ , that  $F$  is bounded, locally  $x$ -Lipschitz and  $\partial F^{(0)} / \partial x$  exists and

$$\|F^{(0)}\|_2 < \infty \text{ and } \|F^{(1)}\|_1 < \infty.$$

(A2) We require  $f(x, t)$  to be almost periodic ([9]) with the Fourier representation  $f(x, t) = \sum a_k(x) e^{i\lambda_k t}$ . Since  $f$  is real, we must have  $a_{-k} = a_k^*$  (where  $a^*$  denotes the

complex conjugate of  $a$ ) and  $\lambda_{-k} = -\lambda_k$ . Note that  $a_0(x) = \bar{f}(x)$  as defined by (0.3). The derivatives  $\partial a_k / \partial x$  must exist and we require

$$\sum \|a_k\|_1 < \infty, \quad \sum_{k \neq 0} \|a_k\|_1 / |\lambda_k| < \infty \quad \text{and} \quad \sum_{k \neq 0} \|a_k\|_2 / |\lambda_k|^{1+\theta} < \infty$$

for some  $0 < \theta \leq 1$ .

(A3) Relations (0.6), (0.7) and (0.8) hold, the invariant  $I$  has first and second partial derivatives with respect to  $x$  and

$$\|\partial I / \partial x_j\| < \infty, \quad \|\partial^2 I / \partial x_j \partial x_k\|_1 < \infty \quad \text{for } 1 \leq j, k \leq p.$$

*Note.* It follows that  $g(x, t) = \sum_{k \neq 0} b_k(x) e^{i\lambda_k t}$  where  $b_k = (\partial I / \partial x) a_k$ . The assumptions in (A3) imply that the three series in (A2) converge when  $a_k$  is replaced by  $b_k$ .

(A4) For  $0 \leq s \leq t \leq \infty$  let  $\mathcal{F}_s^t$  be sub- $\sigma$ -fields of  $\mathcal{F}$  such that for  $s_1 \leq s_2 \leq t_2 \leq t_1$   $\mathcal{F}_{s_2}^{t_2} \subset \mathcal{F}_{s_1}^{t_1}$  and such that  $F^{(0)}(x, r)$  and  $F^{(1)}(x, r)$  are  $\mathcal{F}_r^r$  measurable for all  $x \in D_0$ . If the initial value  $x_0$  of (0.1) is random, then we also require  $x_0$  to be  $\mathcal{F}_0^0$  measurable. We require a  $\varphi$ -mixing condition: let

$$\varphi(t) = \sup_{s \geq 0} \sup_{A \in \mathcal{F}_s^0: PA > 0} \sup_{B \in \mathcal{F}_{s+t}^\infty} |P(B|A) - PB|.$$

We require  $t^r \varphi(t) \leq C < \infty$  for all  $t > 1$  and some  $r > 4$ .

(A5) Let

$$\mu_{1,k}(x) := \frac{i}{\lambda_k} \frac{\partial b_k}{\partial x}(x) a_k(x)^*$$

and let

$$\mu_2(x, s, r) := E \left[ \frac{\partial G^{(0)}}{\partial x}(x, s) F^{(0)}(x, r) \right],$$

$$\mu_3(x, t) := EG^{(1)}(x, t),$$

$$\mathcal{Z}(x, s, r) := E[G^{(0)}(x, s) G^{(0)}(x, r)^T].$$

The assumptions (A2) and (A3) imply

$$\tilde{\mu}_1(x) := \sum_{k \neq 0} \mu_{1,k}(x)$$

exists and the series converges uniformly in  $x \in D_0$ . Moreover, assume the following limits exist uniformly in  $x \in D_0$  and  $t$ :

$$\tilde{\mu}_2(x) := \lim_{l \rightarrow \infty} \frac{1}{l} \int_t^{t+l} ds \int_t^s dr \mu_2(x, s, r)$$

$$\tilde{\mu}_3(x) := \lim_{l \rightarrow \infty} \frac{1}{l} \int_t^{t+l} \mu_3(x, s) ds,$$

$$\tilde{\mathcal{Z}}(x) := \lim_{l \rightarrow \infty} \frac{1}{l} \int_t^{t+l} ds \int_t^{t+l} dr \mathcal{Z}(x, s, r).$$

Finally, let  $\tilde{\mu} = \tilde{\mu}_1 + \tilde{\mu}_2 + \tilde{\mu}_3$ .

It is easy to see that

$$\mu_{1,k}(x) + \mu_{1,-k}(x) = \frac{i}{\lambda_k} \frac{\partial I(x)}{\partial x} \left( \frac{\partial a_k(x)}{\partial x} a_k^*(x) - \frac{\partial a_k^*(x)}{\partial x} a_k(x) \right).$$

Also, letting

$$K(x, s, r) := E \left[ \frac{\partial F^{(0)}}{\partial x}(x, s) F^{(0)}(x, r) \right],$$

$$C(x, s, r) := E [ F^{(0)}(x, s) F^{(0)}(x, r)^T ] = C(x, r, s)^T,$$

and assuming that

$$\tilde{K}(x) := \lim_{l \rightarrow \infty} \frac{1}{l} \int_0^l ds \int_0^s dr K(x, s, r),$$

$$\tilde{C}(x) := \lim_{l \rightarrow \infty} \frac{1}{l} \int_0^l ds \int_0^l dr C(x, s, r) = \tilde{C}(x)^T$$

exist, it follows that

$$\tilde{\mu}_2(x) = \frac{1}{2} \sum_{j,k=1}^p \frac{\partial^2 I}{\partial x_j \partial x_k}(x) \tilde{C}_{j,k}(x) + \frac{\partial I}{\partial x}(x) \tilde{K}(x),$$

$$\tilde{\mathcal{F}}(x) = \frac{\partial I}{\partial x}(x) \tilde{C}(x) \frac{\partial I}{\partial x}(x)^T,$$

where  $\tilde{C}_{j,k}$  is the  $j, k$  entry of  $\tilde{C}$ .

(A6) Let  $\bar{x}(t, x_0)$  be the solution to the initial value problem

$$\frac{dx}{dt} = \bar{f}(x), \quad x(0) = x_0.$$

Assume there exist continuous functions  $\bar{\mu}(y), \bar{\mathcal{F}}(y)$  of  $y$  on  $D_1$  such that the following limits exist uniformly in  $x \in D_0$ :

$$\bar{\mu}(I(x)) = \lim_{l \rightarrow \infty} \frac{1}{l} \int_0^l \tilde{\mu}(\bar{x}(s, x)) ds,$$

$$\bar{\mathcal{F}}(I(x)) = \lim_{l \rightarrow \infty} \frac{1}{l} \int_0^l \tilde{\mathcal{F}}(\bar{x}(s, x)) ds. \tag{1.1}$$

(A6') Assume

1. The limits in (1.1) exist for almost every  $x$  (they need not be uniform, but they must be functions of  $y = I(x)$ ).
2. Let  $X_\varepsilon(\tau) = x(\tau/\varepsilon^2, \varepsilon)$ . For  $0 \leq \tau < \tau + h$ , we assume the conditional distribution of  $X_\varepsilon(\tau + h)$  given  $\mathcal{F}_0^{\tau/\varepsilon^2}$  has a conditional density  $p_\varepsilon(\omega, \tau, \tau + h, x)$  on  $D_0$  such that for some  $\varepsilon_0 > 0$  and each compact subset  $K$  of  $D_0$ ,

$$p_K(\tau, \tau + h) := P - \text{ess sup} \sup_{\omega} \sup_{0 < \varepsilon < \varepsilon_0} \sup_{x \in K} p_\varepsilon(\omega, \tau, \tau + h, x)$$

satisfies

$$\int_{\tau+s}^{\tau+t} p_K(\tau, \tau + h)dh < \infty$$

for each  $0 < s < t < \infty$ .

*Note.* The functions  $\bar{\mu}(I(x))$  and  $\bar{\mathcal{F}}(I(x))$  defined by (1.1) are necessarily continuous functions of  $x$ . We require them to be continuous functions of  $y = I(x)$  as well. Suppose the limits in (1.1) exist uniformly for all  $x$ , as in (A6). Then it is easy to see that  $\bar{\mu}$  and  $\bar{\mathcal{F}}$  are continuous in  $y \in D_1$  provided, when  $y_n \rightarrow y$  in  $D_1$ , the distance  $d(I^{-1}\{y_n\}, I^{-1}\{y\}) \rightarrow 0$ . In other words, provided there exist  $x'_n, x''_n$  in  $D_0$  such that  $y_n = I(x'_n)$ ,  $y = I(x''_n)$  and  $x'_n - x''_n \rightarrow 0$ . This condition is met in all the examples we have considered. Under (A6') an application of Egoroff's theorem shows that the limits coincide almost everywhere with continuous functions of  $y$  under the above condition on  $I^{-1}$ .

Results of the type considered here are best stated in terms of a scaled time. To this purpose, we introduce the "slow time"  $\tau = \varepsilon^2 t$ .

**Theorem 1.1.** *Let assumptions (A1) through (A5) and either (A6) or (A6') hold. Let  $\hat{y}(t, \varepsilon)$  be the solution to (0.7) stopped on reaching  $I(\partial D_0)$  with  $\hat{y}(0, \varepsilon) = I(x_0) \in D_1$ . Then the scaled processes*

$$Y_\varepsilon(\tau) := \hat{y}(\tau/\varepsilon^2, \varepsilon)$$

*converge weakly to a Markov diffusion process  $Y_0(\tau)$  on  $D_1$  with infinitesimal generator  $A$  defined on  $C^{(2)}(D_1)$  by*

$$Ah(y) = \sum_{i=1}^q \bar{\mu}^{(i)}(y) \frac{\partial h}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^q \bar{\mathcal{F}}^{(i,j)}(y) \frac{\partial^2 h}{\partial y_i \partial y_j} \tag{1.2}$$

*and with  $I(\partial D_0)$  an absorbing set.*

*Remarks.*

1. In the deterministic case the weak convergence yields the following: The functions  $Y_\varepsilon(\tau)$  are equicontinuous and for each  $\tau$ ,  $Y_\varepsilon(\tau) \rightarrow Y_0(\tau)$  as  $\varepsilon \rightarrow 0$ . It follows that  $Y_\varepsilon(\tau)$  converges to  $Y_0(\tau)$  uniformly in  $\tau \in [0, \tau_0)$  for any fixed  $\tau_0$ , where in this case  $Y_0(\tau)$  is the solution to the initial value problem

$$\frac{dz}{d\tau} = \bar{\mu}_1(z), \quad z(0) = I(x_0). \tag{1.3}$$

2. Note that if  $\bar{f} = 0$  then we can take  $y = x$ , conditions (A3) and (A6) are automatic and  $\bar{\mu} = \tilde{\mu}$ ,  $\bar{\mathcal{F}} = \tilde{\mathcal{F}}$ . In this sense our result extends that of Khas'minskii [14], though our regularity assumptions, and in particular (A2) and our treatment of  $f$ , differ from his.

3. In all examples we have considered  $|I(x)| \rightarrow \infty$  as  $|x| \rightarrow \infty$ . In such cases, since  $I$  is continuous,  $I(F)$  is closed for any closed set  $F$ . Thus for any choice of  $D_0$ ,  $I(\partial D_0)$  and  $I(\bar{D}_0)$  are closed, and the  $Y_0$  process can exit  $D_1$  only into the set  $I(\partial D_0)$ . In any case  $Y_\varepsilon$  is confined to  $I(\bar{D}_0)$  for  $\varepsilon > 0$  and  $Y_0$  is confined to  $\bar{D}_1 \supset I(\bar{D}_0)$ .

4. Let  $X_\varepsilon(\tau) := x(\tau/\varepsilon^2, \varepsilon)$  and let  $\beta_\varepsilon = \varepsilon^2 B_\varepsilon$  be the scaled hitting time of  $\partial D_0$ . Since  $I$  is continuous,  $\beta_\varepsilon$  is also the hitting time of  $I(\partial D_0)$  by  $Y_\varepsilon(\tau)$ , and the process in

$D_1$  can exit only into the set  $I(\partial D_0)$ , which in general will be a subset of the boundary of the set  $D_1$ . Since  $X_\varepsilon$  is stopped on  $\partial D_0$ ,  $Y_\varepsilon$  is stopped on reaching  $I(\partial D_0)$ , both at the time  $\beta_\varepsilon$ .

Now let  $d_\varepsilon(\tau)$  be the minimum distance of  $\{Y_\varepsilon(s) : 0 \leq s \leq \tau\}$  from  $\partial D_1$ . Note that this is a continuous functional of the  $Y_\varepsilon$  processes. Also let  $\beta_0$  be the hitting time of  $\partial D_1$  by  $Y_0$ . Then it is easy to see that  $[\beta_\varepsilon > \tau] = [d_\varepsilon(\tau) > 0]$  for  $\varepsilon \geq 0$ . It follows from the weak convergence of  $d_\varepsilon(\tau)$  to  $d_0(\tau)$  that

$$\liminf_{\varepsilon \rightarrow 0+} P[d_\varepsilon(\tau) > 0] \geq P[d_0(\tau) > 0],$$

hence

$$\liminf_{\varepsilon \rightarrow 0+} P[\beta_\varepsilon > \tau] \geq P[\beta_0 > \tau]. \tag{1.4}$$

Thus, asymptotically the  $Y_\varepsilon$  processes stay in  $D_1$  at least as long (in terms of the distribution of the hitting time of the boundary) as  $Y_0$  does.

While Theorem 1 is typical of results in the literature on stochastic averaging, the obvious problem with it is that often the domain  $D_0$  is unbounded and the regularity assumptions do not hold uniformly for  $x \in D_0$  as required, though they may hold uniformly on suitable compact subsets of  $D_0$ . The next result extends the theorem to this situation. Let  $D_0^{(n)}$  be an increasing sequence of bounded subsets of  $D_0$  such that  $D_0 = \bigcup D_0^{(n)}$  and such that  $D_1^{(n)} = I(D_0^{(n)})$  is open, and let  $\beta_\varepsilon^{(n)}$  be the hitting time of  $I(\partial D_0^{(n)}) = \partial D_1^{(n)}$  by  $Y_\varepsilon(\tau)$  for  $\varepsilon \geq 0$ .

**Corollary 1.1.** *Suppose that the assumptions (A1) through (A5) and (A6) or (A6') hold when  $D_0$  is replaced by  $D_0^{(n)}$  for each  $n$ . Then (1.2) defines a diffusion on  $D_1$  such that, for each  $n$ , the  $Y_\varepsilon$  process stopped upon reaching the set  $I(\partial D_0^{(n)})$  converges weakly to  $Y_0$  stopped on the same set.*

*Remarks.*

1. Under the conditions of the theorem, the coefficients of the diffusion operator  $A$  are bounded, so explosions will not occur. But in the more general setting of the Corollary these coefficients may be unbounded and explosions may be possible for the  $Y_0$  process. Nevertheless, the  $Y_0$  process will remain finite on the interval  $[0, \beta_0)$ , where  $\beta_0 = \lim_{n \rightarrow \infty} \beta_0^{(n)}$ . Of course the process on  $D_1$  will not explode if  $D_1$  is bounded or if the coefficients of  $A$  have at most a linear growth rate.
2. The relation (1.4) holds for the  $\beta_\varepsilon^{(n)}$  and each  $n$ . Letting  $\beta_\varepsilon = \lim_{n \rightarrow \infty} \beta_\varepsilon^{(n)}$ , we have  $\beta_\varepsilon^{(n)} \uparrow \beta_\varepsilon$ , and it follows that (1.4) holds for the  $\beta_\varepsilon$  as well.
3. In the case of a one degree of freedom oscillator in phase space  $R^2 = R^p$ , where  $I(x)$  is the energy, we could take  $D_0^{(n)} = \{x : 1/n < I(x) < n\}$  so  $D_1 = (1/n, n)$ . Or if we need to consider energy restricted to be at most some finite  $e_c$ , take  $D_1^{(n)} = (1/n, e_c - 1/n)$ . Then  $D_1 = (0, e_c)$ . In this case 0 and  $e_c$  will be reflecting or absorbing boundaries for the diffusion  $Y_0$ , depending on how the oscillator is perturbed near these boundary energies.

*Discussion of the Almost Periodic Assumption.* Readers familiar with Khas'minskii's approach will recall that he does not require the function  $f$  to be almost periodic.

Instead there is a condition that there exist times  $T_n \uparrow \infty$  increasing at most geometrically fast such that

$$\sup_{t > T_n} \sup_x \left| T_n^r \int_t^{t+T_n} (f(x, s) - \bar{f}(x)) ds \right| \rightarrow 0 \tag{1.5}$$

as  $n \rightarrow \infty$  for some  $r > 0$  (in fact  $r = 6$  in [14] and  $r = 3/2$  in [2]). Now, suppose this condition holds uniformly in  $t \in R^1$ . Instead of the convergence to 0, as in (1.5), we will require only that

$$\sup_{t, x} \left| T_n^r \int_t^{t+T_n} (f(x, s) - \bar{f}(x)) ds \right| \leq C < \infty .$$

This condition, for any  $r > 0$ , actually implies that the function  $h(x, t) := \int^t (f(x, s) - \bar{f}(x)) ds$  is uniformly almost periodic. To see this note that, choosing a subsequence of the  $T_n$  if necessary, we can assume  $\rho T_n \leq T_{n+1} < \rho^2 T_n$  for each  $n$  and some  $\rho > 1$ . Given integer  $m$  and  $U > 0$  we will choose  $t \in (U, U + T_m]$  so  $\sup_{x, s} \|h(x, s + t) - h(x, s)\|$  is small. To do this, let  $n_1 = \max\{n: T_n \leq U\}$  and for  $k > 1$  define  $n_k$  inductively by  $n_k = \max\{n: T_n \leq U - \sum_{j=1}^{k-1} T_{n_j}\}$ . Let  $v = \max\{k: n_k \geq m\}$ . Note there are at most  $\rho^2 - 1$  values of  $n_k$  that are equal to any given  $n \geq 1$ . Let

$$t = \sum_{k=1}^v T_{n_k} + T_m .$$

Then  $U < t \leq U + T_m$  and, since  $\|f\| < \infty$ ,

$$\begin{aligned} |h(x, s + t) - h(x, s)| &\leq \sum_{k=2}^v \left| h\left(x, s + \sum_{j=1}^k T_{n_j}\right) - h\left(x, s + \sum_{j=1}^{k-1} T_{n_j}\right) \right| \\ &\quad + |h(x, s + T_{n_1}) - h(x, s)| \\ &\quad + |h(x, s + t) - h(x, s + t - T_m)| \\ &\leq C\rho^2 T_m^{-r} (1 + \rho^{-r} + \rho^{-2r} + \dots) = C \frac{\rho^2}{1 - \rho^{-r}} T_m^{-r} . \end{aligned}$$

For  $m$  sufficiently large, this bound is less than any given  $\delta > 0$ , and the almost periodic claim follows.

By a standard result, the almost periodicity of the integral  $h(x, t)$  implies that  $f(x, t)$  is almost periodic provided  $f(x, t)$  is uniformly continuous in  $t$ . Thus, we see that Khas'minskii's condition (1.5) is close to the assumption that  $f$  is almost periodic.

On the other hand, even the simplest quasi-periodic function, of the form

$$f(t) = A \cos 2\pi t + B \cos 2\pi\gamma t ,$$

where  $\gamma$  is irrational may not satisfy (1.5) with  $r = 1$ , though it certainly satisfies (A2). To see why it may fail to satisfy (1.5), take  $A = 2\pi a$ ,  $B = 2\pi\gamma$ , where  $\gamma$  is a quadratic irrational and set

$$\begin{aligned} h_n(t) &:= \int_t^{t+T_n} (2\pi a \cos 2\pi s + 2\pi\gamma \cos 2\pi\gamma s) ds \\ &= 2a \sin \pi t_n \cos \pi(2t + t_n) + 2 \sin \pi\gamma t_n \cos \pi\gamma(2t + t_n) , \end{aligned}$$

where we will choose the  $t_n \uparrow \infty$  to make  $\|t_n h_n\|$  as small as possible.<sup>1</sup>

Now, it is known that a quadratic irrational is “approximable to order 2 and to no higher order” (see e.g., Theorem 188 of ref. [12]), which leads us to the following lemma.

**Lemma 1.1.** *If  $\gamma$  is a quadratic irrational, then there is no sequence  $t_n \uparrow \infty$  such that*

$$\sup_{t \geq t_n} |t_n h_n(t)| \rightarrow 0 .$$

*Proof.* Since  $\gamma$  is an irrational,

$$\sup_{t \geq t_n} |t_n h_n(t)| = 2t_n |a| |\sin \pi t_n| + 2t_n |\sin \pi \gamma t_n| . \tag{1.6}$$

This follows from the observation that the orbit of

$$(\pi(2t + t_n) \bmod 2\pi, \pi\gamma(2t + t_n) \bmod 2\pi), t_n \leq t < \infty$$

fills the 2-torus densely. If the right-hand side of (1.6) approaches zero, then both  $t_n |\sin \pi t_n| \rightarrow 0$  and  $t_n |\sin \pi \gamma t_n| \rightarrow 0$ , which implies

$$t_n = q_n + o\left(\frac{1}{t_n}\right) \quad \text{and} \quad \gamma t_n = p_n + o\left(\frac{1}{t_n}\right),$$

where  $q_n$  and  $p_n$  are integers. But this entails

$$p_n = \gamma q_n + o\left(\frac{1}{t_n}\right) = \gamma q_n + o\left(\frac{1}{q_n}\right),$$

hence

$$\gamma = \frac{p_n}{q_n} + o\left(\frac{1}{q_n^2}\right).$$

However, this contradicts the fact that a quadratic irrational cannot be approximated to an order higher than 2. □

In our first approach to this problem we tried to carry through arguments similar to those in [14] and [2] using the value  $r = 1$  in (1.5). We found that some of the estimates obtained in those papers still apply when  $r = 1$ , but that no smaller value of  $r$  is possible. It is for these reasons that we chose an alternative approach based on almost periodicity.

Finally, we note that it is possible to weaken the almost periodic assumptions to a form of asymptotic almost periodicity. If some function  $\tilde{f}$  satisfies condition (A2) and

$$\int_0^\infty \sup_{x \in D_0} |(\tilde{f}(x, t) - f(x, t))| dt < \infty \tag{1.7}$$

then the results presented in this paper still apply.

## 2. Examples and Discussion

In this section we discuss several examples which illustrate the theory. Our examples will be of the form (0.1) with

$$F(x, t, \omega) = \varepsilon \tilde{f}(x) + \varepsilon p(x, t) + \varepsilon F^{(0)}(x, t, \omega) ,$$

where  $p(x, t) := f(x, t) - \bar{f}(x)$ , and in most cases the vector field  $\bar{f}(x)$  will be derivable from a Hamiltonian,  $H(x)$ . Furthermore, in the case where there is only one invariant, we take phase space volume to be the invariant. More precisely, we take

$$I(x) = V(H(x)), \tag{2.1a}$$

where

$$V(h) = \int_{H(x) \leq h} dx. \tag{2.1b}$$

This is convenient because it simplifies the infinitesimal generator (1.2). Under the assumption of ergodicity on energy surfaces  $H = h$ , for any given continuous function  $\psi(x)$  define

$$\bar{\psi}(y) = \lim_{l \rightarrow \infty} \frac{1}{l} \int_0^l \psi(\bar{x}(s, x)) ds = \frac{1}{DV(h)} \frac{d}{dh} \int_{H(x) \leq h} \psi(x) dx, \tag{2.2}$$

where  $y = I(x)$  and  $h = h(y)$  is defined by  $y = V(h)$ . It is convenient to use the notation  $D_m^n$  for the  $n^{\text{th}}$  derivative with respect to its  $m^{\text{th}}$  argument. If  $n = 1$  we may suppress the superscript and if there is only one argument, we may suppress the subscript. If there are no parentheses, then  $D_m^n$  operates only on the next indicated function. The averages in (1.1) are obtained by setting  $\psi = \tilde{\mu}$  and  $\tilde{\mathcal{F}}$  and  $\bar{\psi} = \bar{\mu}$  and  $\bar{\mathcal{F}}$ . It then follows that

$$\begin{aligned} \bar{\mu}_2(y) = & \frac{1}{2} D^2 V(h(y)) \overline{DH(x) \tilde{C}(x) DH(x)^T} + DV(h(y)) \left\{ \frac{1}{2} \overline{\text{Trace}[\mathcal{H}(x) \tilde{C}(x)]} \right. \\ & \left. + \overline{DH(x) \tilde{K}(x)} \right\}, \end{aligned} \tag{2.3}$$

where  $\mathcal{H}(x)$  is the Hessian matrix of  $H(x)$ , the bar denotes the average of (2.2) and  $\tilde{C}$  and  $\tilde{K}$  are defined in (A5). In addition,

$$\begin{aligned} \bar{\mathcal{F}}(y) &= (DV(h))^2 \overline{DH(x) \tilde{C}(x) DH(x)^T} \\ &= DV(h) \frac{d}{dh} \int_{H(x) \leq h} (DH(x) \tilde{C}(x) DH(x)^T) dx. \end{aligned} \tag{2.4}$$

The infinitesimal generator simplifies when  $\bar{\mu}_2(y) = \frac{1}{2} D \bar{\mathcal{F}}(y)$ . Sufficient conditions for this are

$$K = 0 \tag{2.5a}$$

and

$$\frac{d}{dh} \int_{H \leq h} DH(x) \tilde{C}(x) DH(x)^T dx = \int_{H \leq h} \text{Trace}[\mathcal{H}(x) \tilde{C}(x)] dx. \tag{2.5b}$$

Under sufficient smoothness assumptions, the infinitesimal generator, Fokker-Planck equation and backward Kolmogorov equation can be written

$$A(y) = \frac{1}{2} \frac{\partial}{\partial y} \bar{\mathcal{F}}(y) \frac{\partial}{\partial y}, \tag{2.6a}$$

$$\frac{\partial \rho(y, \tau; z, \tau')}{\partial \tau} = A(y) \rho, \tag{2.6b}$$

and

$$\frac{\partial \rho}{\partial \tau'}(y, \tau; z, \tau') = -A(z)\rho, \tag{2.6c}$$

where the transition probability density  $\rho(y, \tau; z, \tau')$  satisfies  $\rho(y, \tau; z, \tau') = \delta(y - z)$  and the appropriate absorbing boundary condition.

If  $D_0 = \{x: H(x) < h_b\}$  and  $D_1 = [0, y_b)$ , then the quantity  $\beta_0$  discussed in Remark 5 to Theorem 1.1 is given by

$$P[\beta_0 > \tau] = \int_0^{y_b} \rho(y, \tau; z, 0) dy := G(z, \tau) \tag{2.7}$$

and  $G$  satisfies the IBVP

$$\begin{aligned} \frac{\partial G}{\partial \tau} &= A(z)G, \\ G(z, 0) &= \begin{cases} 1 & 0 \leq z < y_b \\ 0 & z = y_b \end{cases}, \\ G(y_b, \tau) &= 0. \end{aligned} \tag{2.8}$$

The mean first passage time  $\tau_m = -\int_0^\infty \tau G_\tau(z, \tau) d\tau$ , thus from (2.8),

$$A(z)\tau_m = \begin{cases} -1 & 0 \leq z < y_b \\ 0 & z = y_b \end{cases} \tag{2.9}$$

from which the mean time from  $z$  to  $y_b$  is

$$\tau_m(z, y_b) = 2 \int_z^{y_b} \frac{s}{\mathcal{F}(s)} ds. \tag{2.10}$$

*Example A.* Here we consider the deterministic IVP

$$\begin{aligned} \dot{x}_1 &= \varepsilon x_2, \\ \dot{x}_2 &= -\varepsilon U'(x_1) + \varepsilon(x_2^2 \cos \lambda t + \lambda \alpha \sin \lambda t), \quad x(0) = x_0, \end{aligned} \tag{2.11}$$

where  $\lambda$  and  $\alpha$  are parameters. In addition,  $U(-x_1) = U(x_1)$ ,  $U(0) = 0$ ,  $U(x_1) \rightarrow \infty$  as  $|x_1| \rightarrow \infty$  so that all solutions of the unperturbed problem are periodic. From (A5) we have

$$\tilde{\mu}_1(x) = -\alpha x_2^2 DV(H(x)),$$

where  $H(x) = \frac{1}{2}x_2^2 + U(x_1)$  and (2.2) gives

$$\bar{\mu}_1(y) = -\alpha y.$$

The choice of the action as the invariant makes  $\bar{\mu}_1$  particularly simple.

Now  $y(t) = V(H(x(t)))$ , and Theorem 1.1 (see Remark 1) gives

$$y(t, \varepsilon) - z(\varepsilon^2 t) = o(1) \quad \text{for } 0 < t \leq \tau_0/\varepsilon^2,$$

where  $z(\tau) = y(0) \exp(-\alpha\tau)$  since  $z$  is defined by (1.3). In this case, we can actually show that  $y(t, \varepsilon) = z(\varepsilon^2 t) + O(\varepsilon)$  by refining the estimates in Sect. 3. This result, in

which  $\alpha$  behaves like a damping constant, seems remarkable to us. That is, simply by fixing  $\alpha$  in (2.11) to be positive, zero or negative, the action increases  $O(1)$ , remains constant to  $O(\varepsilon)$  or decreases  $O(1)$  on  $O(1/\varepsilon^2)t$ -intervals. One might think, a priori, that a KAM type theorem might apply to (2.11), since without the perturbation  $p(x, t)$  all solutions lie on closed invariant curves. Our result shows this cannot be the case. However, it should be noted that the vector field in (2.11) is not divergence free.

To obtain more insight into this result, we apply the averaging theorem [7, 8] at second order to (2.11). For this purpose, we introduce the averaged IVP at second order,

$$\begin{aligned} \dot{v}_1 &= \varepsilon v_2, \quad \dot{v}_2 = -\varepsilon U'(v_1) - \varepsilon^2 \alpha v_2, \\ v(0) &= x_0 - \varepsilon \mathcal{P}(x_0, 0), \end{aligned} \tag{2.12}$$

and

$$w(t, \varepsilon) := v(t, \varepsilon) + \varepsilon \mathcal{P}(v(t, \varepsilon), t), \tag{2.13}$$

where

$$\mathcal{P}(v, t) = \begin{pmatrix} 0 \\ \frac{1}{\lambda} v_2^2 \sin \lambda t - \alpha \cos \lambda t \end{pmatrix}. \tag{2.14}$$

The averaging theorem now states that for fixed  $x_0 \in R^2$  and positive  $T < \infty$ , there exist positive  $\varepsilon^* = \varepsilon^*(x_0, T)$  and  $C = C(x_0, T)$  such that  $0 < \varepsilon < \varepsilon^*$  implies that for  $0 \leq t \leq T/\varepsilon$  the IVPs (2.11) and (2.12) have unique solutions and  $\|x(t, \varepsilon) - w(t, \varepsilon)\| \leq C\varepsilon^2$ . The equation for  $v_1$  is  $\ddot{v}_1 + \varepsilon^2 \alpha \dot{v}_1 + \varepsilon^2 U'(v_1) = 0$ , so that  $\alpha$  does indeed behave like a damping constant on  $O(1/\varepsilon)$   $t$ -intervals consistent with our result. Our result is deeper however, since it is valid on  $O(1/\varepsilon^2)$   $t$ -intervals.

*Example B.* Here we consider the deterministic IVP (0.1), where

$$\dot{x} = \varepsilon f(x, t), \tag{2.15}$$

when  $\tilde{\mu}_1(x) = 0$ . In this case (A6) is trivially satisfied and we do not need any form of ergodicity for the unperturbed problem,  $\dot{x} = \varepsilon \bar{f}(x)$ . It then follows from the proof of Theorem 1.1 and Remark 1 that the change in  $y$  on  $O(1/\varepsilon^2)$   $t$ -intervals is  $o(1)$ .

The theory of adiabatic invariance for deterministic systems has a long history and much work has been done, however we are not aware of any result in the literature which gives this result. Although results on adiabatic invariance of the action do overlap with the present result.

In order to gain further insight we proceed *formally* with a second order averaging transformation, that is, we look for a transformation

$$x = u + \varepsilon \mathcal{P}_1(u, t) + \varepsilon^2 \mathcal{P}_2(u, t) \tag{2.16}$$

which transforms (2.15) into

$$\dot{u} = \varepsilon V_1(u) + \varepsilon^2 V_2(u) + \varepsilon^3 R(u, t, \varepsilon). \tag{2.17}$$

We require  $\mathcal{P}_1$  and  $\mathcal{P}_2$  to be of the same type as  $p(x, t)$ , that is periodic, quasiperiodic or almost periodic. If we differentiate (2.16) along the solutions of (2.15) and use (2.17), then at  $O(\varepsilon)$  we obtain  $V_1(u) = \bar{f}(u)$  and

$$\mathcal{P}_1(u, t) = \sum_k \frac{1}{i\lambda_k} a_k(u) e^{i\lambda_k t},$$

and at  $O(\varepsilon^2)$  we obtain

$$V_2(u) + D_2 \mathcal{P}_2(u, t) + D_1 \mathcal{P}_1(u, t) V_1(u) = D\bar{f}(u) \mathcal{P}_1(u, t) + D_1 p(u, t) \mathcal{P}_1(u, t).$$

It follows that  $V_2(u)$  must be the  $t$ -average of  $D_1 p(u, t) \mathcal{P}_1(u, t)$  since these are the only terms of nonzero mean; this gives

$$V_2(u) = \sum_k \frac{1}{i\lambda_k} Da_k(u) a_k^*(u).$$

Thus  $\tilde{\mu}_1(x) = DI(x) V_2(x)$  and is zero when  $V_2 = 0$ . Therefore, it is a result of our theorem that when  $V_2$  is zero in averaging at second order, then invariants of the averaged problem become adiabatic invariants of the full problem on  $O(1/\varepsilon^2)$   $t$ -intervals. This extends the standard averaging result which would give this result on  $O(1/\varepsilon)$   $t$ -intervals.

This result can be used to show that in planar and axial channeling in perfect crystals [7], transverse energy is approximately conserved for very long times.

*Example C.* Here we consider particle motion in a rapidly varying field [3]

$$\begin{aligned} \dot{x}_1 &= \varepsilon x_2, \\ \dot{x}_2 &= -\varepsilon U'(x_1) + \varepsilon P(x_1, t) + \varepsilon Q(x_1, t, \omega), \end{aligned} \tag{2.18}$$

where  $U$  is a symmetric bowl potential as in Example A. Clearly  $H(x) = \frac{1}{2} x_2^2 + U(x_1)$  and gives rise to an ergodic flow. It is easy to check that  $Da_k(x) a_k^*(x) = 0$ , hence  $\tilde{\mu}_1 = 0$ . Furthermore, the calculation in (2.4) gives

$$\bar{\mathcal{F}}(y) = T(h(y)) 4\sqrt{2} \int_0^{a(y)} \sqrt{h(y) - U(x_1)} \tilde{C}_{22}(x_1) dx_1, \tag{2.19}$$

where the action is given by

$$V(h) = 4\sqrt{2} \int_0^{a(y)} \sqrt{h - U(x_1)} dx_1, \tag{2.20}$$

$T(h) = DV(h)$  is the oscillation period as a function of energy,  $a(y)$  is the unique positive solution of  $h(y) = U(a(y))$ ,  $C_{22}(x_1, s, r) = E(Q(x_1, s)Q(x_1, r))$  and we have assumed  $\tilde{C}_{22}(x_1)$  is an even function. It is easy to check that (2.5) is satisfied, hence

$$\bar{\mu}_2(y) = \frac{1}{2} D\bar{\mathcal{F}}(y), \tag{2.21}$$

and Eqs. (2.6) to (2.10) follow.

To apply Theorem 1.1, we take  $D_0 = \{x: H(x) < h_b\}$  for some  $h_b > 0$ , which gives  $D_1 = [0, y_b]$ , where  $y_b = V(h_b)$ . We take  $P(x_1, t)$  to be periodic, quasiperiodic or almost periodic with sufficient smoothness in  $x_1$  and  $t$  so that the series in (A2)

converge. If  $Q$  is sufficiently smooth then (A1) and (A2) are satisfied. Since  $U$  is smooth,  $V(h)$  and  $H(x)$  are smooth, so  $I(x) = V(H(x))$  satisfies (A3). We assume  $Q$  satisfies the mixing assumption (A4), where the  $\sigma$ -fields  $\mathcal{F}_s^t$  could be those generated by  $Q(x_1, r)$  for  $r \in [s, t]$  and  $x \in D_0$ . The uniformity of the limits in (A5) is another assumption on  $C_{22}(x_1, s, r)$  and (A6) is satisfied because of the trivial ergodicity in these one degree of freedom problems. Thus, by Theorem 1.1, the process  $Y_\varepsilon(\tau) = V(H(\hat{x}(\tau/\varepsilon^2, \varepsilon)))$  converges weakly to the Markov process on  $D_1$  defined by (2.19) and (2.21).

If we are interested in the process on  $D_0 = R^2$  and  $D_1 = [0, \infty)$  then, clearly, Corollary 1.1 applies with  $D_0^{(n)} = \{x: H(x) < n\}$  and  $D_1^{(n)} = [0, y_n)$ , and we obtain the limit diffusion on  $[0, \infty)$ . Remark 5 helps refine this result in that it gives us some qualitative information on the effect of the absorbing boundary condition. For example, does the distribution of the process  $Y_\varepsilon(\varepsilon^2 t)$  on  $D_1^{(n)}$  with absorbing boundary at  $y_n$  concentrate at  $y_n$  before  $t$  becomes  $O(1/\varepsilon^2)$ ? The answer is no, as can be seen from (1.4). Furthermore, for  $z < y_n$ , the mean time to hit  $y_n$  is  $\tau_m(z, y_n)/\varepsilon^2$ , where  $\tau_m$  is given by (2.10), thus for times  $\tau \ll \tau_m(z)$  we would expect the process  $Y_0$  without the absorbing boundary condition to be a good approximation to  $Y_\varepsilon$  without the absorbing boundary condition.

To gain a little more insight, we take  $\tilde{C}_{22} = K$  a constant which gives  $\bar{\mathcal{F}}(y) = KyT(h(y))$ . If  $U(x_1) = \frac{1}{2}\alpha^2 x_1^2$  then  $T = 2\pi/\alpha$  and  $\tau_m = \frac{\alpha}{K\pi}(y_b - z)$ , and if  $U(x_1) = b \tan^2 ax_1$  then

$$T(h(y)) = 2\pi \left( \frac{2a^2}{4\pi} y + a\sqrt{2b} \right)^{-1} \text{ and}$$

$$\tau_m = \frac{1}{K} \left[ \frac{a^2}{4\pi^2} (y_b^2 - z^2) + a\sqrt{2b}(y_b - z) \right].$$

This gives a clear picture of possible dependencies of first passage times on  $K$  and  $y_b$ .

*Example D.* Here we consider the randomly forced pendulum equation,

$$\begin{aligned} \dot{x}_1 &= \varepsilon x_2 + \varepsilon Q(t, \omega), \\ \dot{x}_2 &= -\varepsilon U'(x_1), \end{aligned} \tag{2.22}$$

where  $U(x_1) = \alpha(1 - \cos x_1)$ . Clearly  $p(x, t) = 0$ ,  $F^{(0)}(x, t) = (Q, 0)^T$  and  $H(x) = \frac{1}{2}x_2^2 + U(x_1)$  gives rise to an ergodic flow. The only nonzero element of  $\tilde{C}$  is  $\tilde{C}_{11} = K$  a constant. It is now easy to show

$$\bar{\mathcal{F}}(y) = KT(h(y))4\sqrt{2} \int_0^{a(y)} U''(x_1)\sqrt{h - U(x_1)} dx_1 \tag{2.23}$$

and

$$\bar{\mu}(y) = \frac{1}{2} D\bar{\mathcal{F}}(y). \tag{2.24}$$

Here  $T(h) = DV(h)$ , where  $V$  is given by (2.20) and  $a(y)$  is defined as in the remarks following (2.20). The separatrix energy and action are given by  $h_s = U(\pi) = 2\alpha$  and  $y_s = V(2\alpha)$ .

This problem arises in the study of the influence of RF noise on the longitudinal motion of particles in modern accelerators such as the SPS at CERN or the SSC.

We will discuss this problem in more detail in a separate paper [4]. Here, we will focus on the dynamics inside the separatrix. A heuristic analysis of the RF noise problem was presented in the paper by Dôme [5].

To apply Theorem 1.1, we take  $D_1 = [0, y_b)$  and  $D_0 = \{x: H(x) < h(y_b)\}$  for  $y_b \leq y_s$ . It is easy to check that (A1) to (A6) are satisfied as long as  $Q$  satisfies the mixing condition and the uniformity of the limit in (A5).

In the application to the RF noise problem, the absorbing boundary condition needs to be eliminated. We are presently studying the coefficients  $\bar{\chi}$  and  $\bar{\mu}$  across and beyond the separatrix and the applicability of Theorem 1.1. Here we apply Remark 5 after the theorem. The mean first passage time is given by

$$\tau_m(z, y_b) = 2 \int_z^{y_b} \frac{s}{\bar{\chi}(s)} ds$$

and if times  $\tau \ll \tau_m(z)$ , where  $z = V(H(\zeta)) < y_b$ , then the processes with and without the absorbing boundary conditions should be fairly close.

*Example E.* Here we extend Example C to the  $n$  degree of freedom case. The equations have the form of (2.18) where  $x_1, x_2, P$  and  $Q$  are now vectors in  $R^n$ . The Hamiltonian  $H$  becomes  $H(x) = \frac{1}{2} x_2^T x_2 + U(x_1)$ , where  $U$  is a smooth potential with zero minimum. We assume that the energy surfaces  $H(x) = h$  in  $R^{2n}$  are compact and that the motion on these surfaces is ergodic. It should be noted, however, that no Hamiltonians of the form kinetic energy plus smooth potential have been proven to be ergodic for  $n \geq 2$ . As in Example C, it is easy to see that  $\mu_1 = 0$  and that  $\bar{\chi}(y)$  is defined by (2.4). If we define the accessible area by

$$A(h) := \int_{U(x_1) \leq h} dx_1,$$

then it can be shown that

$$V(h) = \frac{(2\pi)^{n/2}}{\Gamma(n/2)} \int_0^h (h - s)^{n/2-1} A(s) ds.$$

Also  $C_{22} = E[Q(x, s)Q(x, r)^T]$ , and if  $\tilde{C}_{22}$  is  $K$  times the  $n$ -dimensional identity matrix, then a lengthy calculation gives  $\bar{\chi}(y) = nKyV'(h(y))$  and  $\bar{\mu}_2(y) = \frac{1}{2}D\bar{\chi}(y)$ .

Assumptions (A1) to (A5) go through as before, but even with the ergodicity, we would not expect (A6) to hold and would have to assume the stochastic process  $Q$  suffices to give (A6').

*Example F.* In the final example, we illustrate the case of two invariants. We take

$$\begin{aligned} \bar{f}(x) &= \begin{pmatrix} x_2 \\ -\alpha_1^2 x_1 \\ x_3 \\ -\alpha_2^2 x_4 \end{pmatrix}, \quad p(x, t) = \begin{pmatrix} 0 \\ p_2(x_1, x_3, t) \\ 0 \\ p_4(x_1, x_3, t) \end{pmatrix}, \\ F^{(0)}(x, t, \omega) &= \begin{pmatrix} 0 \\ q_2(x_1, x_3, t, \omega) \\ 0 \\ q_4(x_1, x_3, t, \omega) \end{pmatrix}, \end{aligned} \tag{2.25}$$

where  $\alpha_1$  and  $\alpha_2$  have irrational ratio. Thus the unperturbed problem is a pair of uncoupled simple harmonic oscillators and its motion is ergodic on two-tori. Here we take the energies

$$\begin{aligned} I_1(x) &= H_1(x) = \frac{1}{2}x_2^2 + \frac{1}{2}\alpha_1^2x_1^2, \\ I_2(x) &= H_2(x) = \frac{1}{2}x_4^2 + \frac{1}{2}\alpha_2^2x_3^2, \end{aligned} \quad (2.26)$$

as the invariants. The limits in (1.1) then become

$$\bar{\sigma}(y_1, y_2) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \tilde{\sigma} \left( \frac{r_1}{\alpha_1} \cos \theta_1, r_1 \sin \theta_1, \frac{r_2}{\alpha_2} \cos \theta_2, r_2 \sin \theta_2 \right) d\theta_1 d\theta_2, \quad (2.27)$$

where  $r_i = \sqrt{2y_i}$ . Furthermore, since the ergodization rate is the same for each torus the limits in (A6) are uniform and (A6) is satisfied. It should be pointed out that it is necessary to take two independent invariants because otherwise the limits in (1.1) would not depend on  $x$  only through  $y$ . The structure of  $p$  and  $q$  give  $\tilde{\mu}_1 = 0 = \tilde{K}$ ,

$$\mu_2(x, s, r) = 2\pi \begin{pmatrix} C_{22}(x, s, r)/\alpha_1 \\ C_{44}(x, s, r)/\alpha_2 \end{pmatrix}, \quad (2.28)$$

and

$$\bar{\chi}(x, s, r) = \begin{pmatrix} x_2^2 C_{22} & x_2 x_4 C_{24} \\ x_2 x_4 C_{42} & x_4^2 C_{44} \end{pmatrix}. \quad (2.29)$$

The infinitesimal generator is then easily computed from (1.2). Theorem 1.1 now applies by choosing  $p$  and  $F^{(0)}$  so that (A1)–(A5) are satisfied.

### 3. Preliminary Estimates

Using the slow time  $\tau = \varepsilon^2 t$ , Eq. (0.1) becomes

$$\begin{aligned} \frac{dX_\varepsilon}{d\tau} &= \frac{1}{\varepsilon} \left( f \left( X_\varepsilon, \frac{\tau}{\varepsilon^2} \right) + F^{(0)} \left( X_\varepsilon, \frac{\tau}{\varepsilon^2}, \omega \right) \right) \\ &\quad + F^{(1)} \left( X_\varepsilon, \frac{\tau}{\varepsilon^2}, \omega \right) + o(1), \quad X_\varepsilon(0) = x_0. \end{aligned} \quad (3.1)$$

The corresponding  $y$ -equation is

$$\begin{aligned} \frac{dY_\varepsilon}{d\tau} &= \frac{1}{\varepsilon} \left( g \left( X_\varepsilon, \frac{\tau}{\varepsilon^2} \right) + G^{(0)} \left( X_\varepsilon, \frac{\tau}{\varepsilon^2}, \omega \right) \right) \\ &\quad + G^{(1)} \left( X_\varepsilon, \frac{\tau}{\varepsilon^2}, \omega \right) + o(1), \quad Y_\varepsilon(0) = I(x_0). \end{aligned} \quad (3.2)$$

We will analyze the behavior of  $Y_\varepsilon(\tau)$  as  $\varepsilon \rightarrow 0$ .

To simplify notation, we will denote the conditional expectations  $E(Z|\mathcal{F}_0^t)$  by  $E^{(t)}Z$ . The following standard mixing result is noted for later reference (e.g. see [14]):

**Lemma 3.1.** *Let  $H(x, \omega)$  be an  $\mathcal{F}_{s+t}^\infty$  measurable random variable with values in  $R^k$  for each  $x \in R^p$  and be Borel measurable in  $x$  for each  $\omega$ . Let  $\|H\| < \infty$  and let  $h(x) = EH(x)$ . Let  $Z$  be an  $\mathcal{F}_0^s$  measurable  $R^p$  valued random variable. Then*

$$\|E^{(s)}H(Z) - h(Z)\| \leq 2\|H\|\varphi(t). \tag{3.3}$$

Let  $\mathcal{H} = \{h: R^p \times [0, \infty) \times \Omega \rightarrow C^q \text{ and } \|h\|_2 < \infty\}$ .

**Lemma 3.2.** *There is a positive constant  $C < \infty$ , independent of  $h$ , such that for all  $h \in \mathcal{H}$ ,  $\varepsilon > 0$ ,  $l > 0$ ,  $0 \leq r \leq l$  and  $0 \leq \tau_1 < \tau_2$ ,*

$$\begin{aligned} & \left\| \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} h\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds - \sum_{k=\lceil \tau_1/\varepsilon^2 \rceil}^{\lceil \tau_2/\varepsilon^2 \rceil - 1} (\varepsilon \Xi_k(h, \varepsilon, l, r) + \varepsilon^2 \Gamma_k(h, \varepsilon, l, r)) \right\| \\ & \leq C((\tau_2 - \tau_1)\|h\|_2 \varepsilon l^2 + \|h\|_1 \varepsilon^2 l^2 + \|h\| \varepsilon l), \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} \Xi_k(h, \varepsilon, l, r) &= \int_{kl+r}^{(k+1)l+r} h(X_\varepsilon(\varepsilon^2 kl), t) dt, \\ \Gamma_k(h, \varepsilon, l, r) &= \int_{kl+r}^{(k+1)l+r} \frac{\partial h}{\partial X}(X_\varepsilon(\varepsilon^2 kl), t) \int_{kl}^t (f + F^{(0)})(X_\varepsilon(\varepsilon^2 kl), s) ds dt. \end{aligned}$$

*Proof.* First, note that

$$\begin{aligned} \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} h\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds &= \varepsilon \int_{\tau_1/\varepsilon^2}^{\tau_2/\varepsilon^2} h(X_\varepsilon(\varepsilon^2 t), t) dt \\ &= \sum_{k=\lceil \tau_1/\varepsilon^2 \rceil}^{\lceil \tau_2/\varepsilon^2 \rceil - 1} \varepsilon \int_{kl+r}^{(k+1)l+r} h(X_\varepsilon(\varepsilon^2 t), t) dt + \|h\|_0(\varepsilon l). \end{aligned} \tag{3.5}$$

To simplify notation let  $X_\varepsilon^{(k)} = X_\varepsilon(\varepsilon^2 kl)$ . Condition (A1) implies that, for  $kl \leq t < (k+1)l+r$ ,

$$|X_\varepsilon(\varepsilon^2 t) - X_\varepsilon^{(k)}| \leq C_1 \varepsilon l \tag{3.6}$$

and that, provided  $\beta_\varepsilon \notin [kl, (k+1)l+r)$ ,

$$X_\varepsilon(\varepsilon^2 t) - X_\varepsilon^{(k)} = \varepsilon \int_{kl}^t (f + F^{(0)})(X_\varepsilon^{(k)}, s) ds + 0(\varepsilon^2 l^2),$$

and, applying Taylor's expansion to  $h$ ,

$$h(X_\varepsilon(\varepsilon^2 t), t) = h(X_\varepsilon^{(k)}, t) + \varepsilon \frac{\partial h}{\partial X}(X_\varepsilon^{(k)}, t) \int_{kl}^t (f + F^{(0)})(X_\varepsilon^{(k)}, s) ds + \|h\|_2 0(\varepsilon^2 l^2).$$

It follows that

$$\int_{kl+r}^{(k+1)l+r} h(X_\varepsilon(\varepsilon^2 t), t) dt = \Xi_k(h, \varepsilon, l, r) + \varepsilon \Gamma_k(h, \varepsilon, l, r) + \|h\|_2 0(\varepsilon^2 l^3). \tag{3.7}$$

There are at most two values of  $k$  with  $\beta_\varepsilon \in [kl, (k + 1)l + r)$ , and for any value of  $k$  the left-hand side of (3.7) is  $\|h\|O(l)$  while the right side is  $\|h\|O(l) + \varepsilon \|h\|_1 O(l^2)$ . The lemma follows on applying these estimates to the terms of the sum in (3.5), since there are at most  $(\tau_2 - \tau_1)/\varepsilon^2 l$  such terms (using the convention  $\sum_{k=n}^{n-1} = 0$ ).  $\square$

**Proposition 3.1.** *For each  $\varepsilon_0 > 0$  there is a  $C < \infty$  such that for all  $0 \leq \tau_1 < \tau_2$  and  $0 < \varepsilon \leq \varepsilon_0$ ,*

$$\left\| \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds \right\| \leq C(\tau_2 - \tau_1 + \varepsilon). \tag{3.8}$$

Moreover, with probability one, as  $\varepsilon \rightarrow 0$ ,

$$\begin{aligned} E^{\tau_1/\varepsilon^2} \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds &= E^{\tau_1/\varepsilon^2} \int_{\tau_1}^{\tau_2} \tilde{\mu}_1(X_\varepsilon(s)) ds \\ &+ (\tau_2 - \tau_1 + \varepsilon^{1/2}) o(1). \end{aligned} \tag{3.9}$$

*Proof.* 1. Apply Lemma 3.2 to  $g_j(x, t) := b_j(x)e^{i\lambda_j t}$ , with  $l = l_j = 2\pi n/|\lambda_j|$ , where  $n$  is a given positive integer. Then  $\Xi_k(g_j, \varepsilon, l_j, r) = 0$  and

$$\begin{aligned} &\left| \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} b_j(X_\varepsilon(s)) e^{i\lambda_j s/\varepsilon^2} ds - \varepsilon^2 \sum_{k=\lceil \tau_1/\varepsilon^2 l_j \rceil}^{\lceil \tau_2/\varepsilon^2 l_j \rceil - 1} \Gamma_k(g_j, \varepsilon, l, r) \right| \\ &\leq C_1 \left( (\tau_2 - \tau_1) \|b_j\|_2 \frac{\varepsilon n^2}{\lambda_j^2} + \|b_j\|_1 \frac{\varepsilon^2 n^2}{\lambda_j^2} + \|b_j\| \frac{\varepsilon n}{|\lambda_j|} \right) \\ &=: e_j(\varepsilon, n). \end{aligned} \tag{3.10}$$

By (A2), (A3) and Fubini's theorem,

$$\frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds = \sum_j \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g_j\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds$$

and so, letting  $\alpha = \theta/(2 - \theta)$ , where  $\theta$  is the quantity in (A2),

$$\begin{aligned} &\left\| \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds - \varepsilon^2 \sum_{j: |\lambda_j| \geq \varepsilon^{1+\alpha}} \sum_{k=\lceil \tau_1/\varepsilon^2 l_j \rceil}^{\lceil \tau_2/\varepsilon^2 l_j \rceil - 1} \Gamma_k(g_j, \varepsilon, l_j, r) \right\| \\ &\leq \varepsilon^\alpha \sum_{|\lambda_j| \geq \varepsilon^{1+\alpha}} (\tau_2 - \tau_1) \frac{\|b_j\|}{|\lambda_j|} + C_1 \sum_{|\lambda_j| \geq \varepsilon^{1+\alpha}} \left\{ (\tau_2 - \tau_1) \frac{\|b_j\|_2}{|\lambda_j|^{1+\theta}} \varepsilon^\alpha n^2 \right. \\ &\quad \left. + \frac{\|b_j\|_1}{|\lambda_j|^{1+\theta}} \varepsilon^{1+\alpha} n^2 + \frac{\|b_j\|}{|\lambda_j|} \varepsilon n \right\} \\ &\leq C_2((\tau_2 - \tau_1)\varepsilon^\alpha n^2 + \varepsilon n), \end{aligned} \tag{3.11}$$

provided  $\varepsilon^\alpha n \leq C_3$  for any fixed  $C_3 < \infty$ .

The first assertion follows from the estimate in (3.11) with  $n = 1$  and the following estimates of the  $\Gamma_k$ 's, where we let  $X = X_\varepsilon(\varepsilon^2 kl_j)$ ,

$$\begin{aligned} \|\Gamma_k(g_j, \varepsilon, l_j, 0)\| &= \left\| \int_{kl_j}^{(k+1)l_j} ds \int_s^{(k+1)l_j} dt \frac{\partial b_j}{\partial X}(X) e^{i\lambda_j t} (f + F^{(0)})(X, s) \right\| \\ &\leq 2l_j \|f + F^{(0)}\| \cdot \|b_j\|_1 / |\lambda_j|, \end{aligned} \tag{3.12}$$

so, for  $n = 1$ ,

$$\varepsilon^2 \sum_j \sum_k \|\Gamma_k(g_j, \varepsilon, l_j, 0)\| \leq C_4(\tau_2 - \tau_1).$$

2. Let

$$\bar{\Gamma}_k(g_j, \varepsilon, l_j) := \frac{|\lambda_j|}{2\pi} \int_0^{2\pi/|\lambda_j|} \Gamma_k(g_j, \varepsilon, l_j, r) dr.$$

Since (3.10) and (3.11) hold for  $\Gamma_k(g_j, \varepsilon, l_j, r)$  and each  $r \leq 2\pi/|\lambda_j|$ , they also hold when  $\Gamma_k$  is replaced by  $\bar{\Gamma}_k$ .

Since  $l_j$  is a period of  $g_j$  we have

$$\int_{kl_j+r}^{(k+1)l_j+r} \frac{\partial g_j}{\partial X}(x, t) dt = 0,$$

and it follows that, again letting  $X = X_\varepsilon(\varepsilon^2 kl_j)$ ,

$$\begin{aligned} E^{(kl_j)} \bar{\Gamma}_k(g_j, \varepsilon, l_j) &= \frac{|\lambda_j|}{2\pi} \int_0^{2\pi/|\lambda_j|} dr \int_{kl_j+r}^{(k+1)l_j+r} ds \int_s^{(k+1)l_j+r} dt \frac{\partial g_j}{\partial X}(X, t) f(X, s) \\ &\quad + \frac{|\lambda_j|}{2\pi} \int_0^{2\pi/|\lambda_j|} dr \int_{kl_j+r}^{(k+1)l_j+r} ds \int_s^{(k+1)l_j+r} dt \frac{\partial g_j}{\partial X}(X, t) E^{(kl_j)} F^{(0)}(X, s) \\ &=: I_j + \Pi_j. \end{aligned}$$

The third integral in  $\Pi_j$  is at most  $2\|b_j\|_1/|\lambda_j|$  and  $\|E^{(kl_j)} F^{(0)}(X, s)\| \leq C_5 \varphi(s - kl_j)$  by Lemma 3.1, hence  $\|\Pi_j\| \leq C_6 \|b_j\|_1 / |\lambda_j|$ .

Now, let  $f_\nu(x, s) = a_\nu(x) e^{i\lambda_\nu s}$ , and note that we can write  $I_j = \sum_\nu I_{j,\nu}$ , where  $I_{j,\nu}$  is obtained by replacing  $f$  by  $f_\nu$  in  $I_j$ . With a change of variables, we have

$$I_{j,0} = \frac{|\lambda_j|}{2\pi} \int_{kl_j}^{(k+1)l_j} du \int_u^{(k+1)l_j} dv \int_0^{2\pi/|\lambda_j|} dr e^{i\lambda_j(v+r)} \frac{\partial b_j}{\partial X}(X) f_0(X) = 0.$$

For  $\nu \neq 0$  and  $\nu \neq -j$  direct calculation shows that

$$\|I_{j,\nu}\| \leq 2\|b_j\|_1 \|a_\nu\| |\lambda_\nu(\lambda_\nu + \lambda_j)|^{-1},$$

while  $I_{j,-j} = l_j \mu_{1,j}(X)$ , defined in (A5).

Let  $f^{(m)} = \sum_{|\nu|>m} f_\nu$ . Then

$$\begin{aligned} \left\| \sum_{|\nu|>m} I_{j,\nu} \right\| &= \left\| \frac{|\lambda_j|}{2\pi} \int_0^{2\pi/|\lambda_j|} dr \int_{kl_j+r}^{(k+1)l_j+r} ds \int_s^{(k+1)l_j+r} dt \frac{\partial g_j}{\partial X}(X, t) f^{(m)}(X, s) \right\| \\ &\leq 2l_j \|f^{(m)}\| \|b_j\|_1 / |\lambda_j|. \end{aligned}$$

Letting  $C_{j,m} := C_6 + \sum_{1 \leq |v| \leq m, v \neq -j} \|a_v\| |\lambda_j| |\lambda_v(\lambda_v + \lambda_j)|^{-1}$ ,

$$\|E^{(kl_j)} \bar{\Gamma}_k(g_j, \varepsilon, l_j) - I_{j,-j}\| \leq (C_{j,m} + 2l_j \|f^{(m)}\|) \|b_j\|_1 / |\lambda_j|.$$

Summing over  $k$  and using (3.10) and the rectangular Riemann sums approximation to the integral of  $\mu_{1,j}$  yields

$$\begin{aligned} & \left\| E^{(\tau_1/\varepsilon^2)} \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g_j \left( X_\varepsilon(s), \frac{s}{\varepsilon^2} \right) ds - E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \mu_{1,j}(X_\varepsilon(s)) ds \right\| \\ & \leq C_7(\tau_2 - \tau_1) \left\{ \frac{\|b_j\|_1}{n} C_{j,m} + \frac{\|b_j\|_1}{|\lambda_j|} \|f^{(m)}\| + \varepsilon n \frac{\|b_j\|_2}{\lambda_j^2} \right\} + e_j(\varepsilon, n), \end{aligned} \quad (3.13)$$

where in the Riemann sum approximation we use the estimate

$$\left\| \frac{\partial \mu_{1,j}}{\partial x} \right\| \leq \frac{\|b_j\|_2 \|a_j\| + \|b_j\|_1 \|a_j\|_1}{|\lambda_j|} \leq C_8 \frac{\|b_j\|_2}{|\lambda_j|}$$

since  $\sup \|a_j\|_1 \leq \sum \|a_j\|_1 < \infty$ .

Let  $g^{(m)} = \sum_{|j| > m} g_j$ . If we replace  $g$  by  $g^{(m)}$  in the first part of the proof, it is clear that the bound in (3.11) applies. Combining this with the estimates in (3.12) for  $|j| > m$  yields (use  $n = 1$  for this estimate)

$$\left\| \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g^{(m)} \left( X_\varepsilon(s), \frac{s}{\varepsilon^2} \right) ds \right\| \leq C_9(\tau_2 - \tau_1) \left( \sum_{|j| > m} \frac{\|b_j\|_1}{|\lambda_j|} + \varepsilon^\alpha \right) + C_2 \varepsilon. \quad (3.14)$$

Combining (3.13) and (3.14),

$$\begin{aligned} & \left\| E^{(\tau_1/\varepsilon^2)} \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g \left( X_\varepsilon(s), \frac{s}{\varepsilon^2} \right) ds - E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}_1(X_\varepsilon(s)) ds \right\| \\ & \leq C_{10}(\tau_2 - \tau_1) \left\{ \|f^{(m)}\| + \sum_{1 \leq |j| \leq m} \left( \frac{\|b_j\|_1}{n} C_{j,m} + \varepsilon n \frac{\|b_j\|_2}{\lambda_j^2} \right) \right. \\ & \quad \left. + \sum_{|j| > m} \frac{\|b_j\|_1}{|\lambda_j|} + \varepsilon^\alpha \right\} + C_2 \varepsilon + \sum_{1 \leq |j| \leq m} e_j(\varepsilon, n). \end{aligned} \quad (3.15)$$

If we let  $\varepsilon \rightarrow 0$  and  $n \rightarrow \infty$  so  $\varepsilon n^2 \rightarrow 0$  then each  $e_j(\varepsilon, n) = (\tau_2 - \tau_1 + \varepsilon^{1/2}) o(1)$  by (3.10), and so this estimate applies to the final sum in (3.15) for each  $m$ . Thus, letting  $\varepsilon \rightarrow 0$  then  $n \rightarrow \infty$  then  $m \rightarrow \infty$  in the iterated limit, all terms in the brackets also go to 0, and the upper bound is  $(\tau_2 - \tau_1 + \varepsilon^{1/2}) o(1)$ .  $\square$

*Remark.* If condition (A2) is replaced by the asymptotic almost periodic condition (1.7), Proposition 3.1 remains true. To see this let  $\tilde{g} = \frac{\partial I}{\partial x} \tilde{f}$ . Then

$$\varepsilon \int_{\tau_1/\varepsilon^2}^{\tau_2/\varepsilon^2} (\tilde{g}(X_\varepsilon(\varepsilon^2 t), t) - g(X_\varepsilon(\varepsilon^2 t), t)) dt = O(\varepsilon)$$

and the estimates in the proof of the proposition apply to  $\tilde{g}$ .

The remaining analysis centers on the random perturbing terms. For this part, we break the process into blocks of length  $L = L(\varepsilon)$ , where  $L(\varepsilon)\varepsilon^{1/2} \rightarrow 0$

as  $\varepsilon \rightarrow 0$  slower than any power of  $\varepsilon$ , and let  $t_k = kL$  and  $X_\varepsilon^{(k)} = X_\varepsilon(\varepsilon^2 t_k)$ . We regard the slow times  $0 \leq \tau_1 < \tau_2$  as fixed (but arbitrary) and let  $n_1 = \lceil \tau_1/\varepsilon^2 L \rceil$ ,  $n_2 = \lceil \tau_2/\varepsilon^2 L \rceil - 1$ .

**Lemma 3.3.** *With probability one, a solution  $Y_\varepsilon(\tau)$  to (3.2) satisfies (as  $\varepsilon \rightarrow 0$ )*

$$Y_\varepsilon(\tau_2) - Y_\varepsilon(\tau_1) = \frac{1}{\varepsilon} \int_{\tau_1}^{\tau_2} g\left(X_\varepsilon(s), \frac{s}{\varepsilon^2}\right) ds + \sum_{k=n_1}^{n_2} \{\varepsilon \xi_k + \varepsilon^2(\gamma_k + \zeta_k)\}, \\ + (\tau_2 - \tau_1 + \varepsilon^{1/2})o(1), \tag{3.16}$$

where  $\xi_k = \Xi_k(G^{(0)}, \varepsilon, L, L)$ ,  $\gamma_k = \Gamma_k(G^{(0)}, \varepsilon, L, L)$  and  $\zeta_k = \Xi_k(G^{(1)}, \varepsilon, L, L)$ .

*Proof.* By (A1) and (A3) we have  $\|G^{(0)}\|_2 < \infty$  and  $\|G^{(1)}\|_1 < \infty$ . Thus, for  $t_k \leq t \leq t_k + 2L$ ,

$$G^{(1)}(X_\varepsilon(\varepsilon^2 t), t) = G^{(1)}(X_\varepsilon^{(k)}, t) + 0(\varepsilon L)$$

and

$$\varepsilon^2 \int_{\tau_1/\varepsilon^2}^{\tau_2/\varepsilon^2} G^{(1)}(X_\varepsilon(\varepsilon^2 t), t) dt = \varepsilon^2 \sum_{k=n_1}^{n_2} \zeta_k + (\tau_2 - \tau_1 + \varepsilon^{1/2})o(1).$$

The lemma follows on applying (3.4) to the  $G^{(0)}$  term. □

It is convenient to write  $\gamma_k = \bar{\gamma}_k + \hat{\gamma}_k$ , where

$$\bar{\gamma}_k = \int_{t_{k+1}}^{t_{k+2}} dt \int_{t_k}^t ds \frac{\partial G^{(0)}}{\partial X}(X_\varepsilon^{(k)}, t) f(X_\varepsilon^{(k)}, s).$$

**Lemma 3.4.** *With probability one, as  $\varepsilon \rightarrow 0$ ,*

$$E^{(t_k)} \xi_k = 0(L\varphi(L)), \\ E^{(t_k)} \bar{\gamma}_k = 0(L^2 \varphi(L)), \\ E^{(t_k)} \hat{\gamma}_k = L\tilde{\mu}_2(X_\varepsilon^{(k)}) + o(L), \\ E^{(t_k)} \zeta_k = L\tilde{\mu}_3(X_\varepsilon^{(k)}) + o(L). \tag{3.17}$$

*Proof.* For the first relation, note that

$$|E^{(t_k)} \xi_k| = \left| \frac{\partial I}{\partial X}(X_\varepsilon^{(k)}) \int_{t_{k+1}}^{t_{k+2}} E^{(t_k)} F^{(0)}(X_\varepsilon^{(k)}, t) dt \right| \\ \leq C \int_{t_{k+1}}^{t_{k+2}} \varphi(t - t_k) dt \leq CL\varphi(L).$$

The second relation follows by a similar estimate since

$$E^{(t_k)} \frac{\partial}{\partial X} G^{(0)}(X_\varepsilon^{(k)}, t) = 0(\varphi(t - t_k)),$$

and the fourth relation follows from

$$E^{(t_k)} G^{(1)}(X_\varepsilon^{(k)}, t) = \mu_3(X_\varepsilon^{(k)}, t) + 0(\varphi(t - t_k))$$

and (A5).

To establish the third relation, first note that, for  $t_k \leq s \leq t$ ,

$$E^{(t_k)} \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, t) F^{(0)}(X_\varepsilon^{(k)}, s) = E^{(t_k)} \left[ \left( E^{(s)} \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, t) \right) F^{(0)}(X_\varepsilon^{(k)}, s) \right] \\ = 0(\varphi(t - s)).$$

Hence

$$\left| E^{(t_k)} \int_{t_{k+1}}^{t_{k+2}} dt \int_{t_k}^{t_{k+1}} ds \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, t) F^{(0)}(X_\varepsilon^{(k)}, s) \right| \\ \leq C \int_{t_{k+1}}^{t_{k+2}} dt \int_{t_k}^{t_{k+1}} ds \varphi(t - s) \leq C \int_0^\infty r\varphi(r) dr < \infty$$

by (A4). Also, for  $t_{k-1} \leq s \leq t$ ,

$$\left| E^{(t_k)} \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, t) F^{(0)}(X_\varepsilon^{(k-1)}, s) - \mu_2(X_\varepsilon^{(k)}, t, s) \right| = 0(\varphi(s - t_k)),$$

so by (A4) and (A5)

$$E^{(t_k)} \int_{t_{k+1}}^{t_{k+2}} dt \int_{t_{k+1}}^t ds \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, t) F^{(0)}(X_\varepsilon^{(k)}, s) = L\tilde{\mu}_2(X_\varepsilon^{(k)}) + o(L). \quad \square$$

Combining the estimates in (3.17) and (3.9) and using the Riemann sums approximation to the integral  $\int_{\tau_1}^{\tau_2} \tilde{\mu}(X_\varepsilon(s)) ds$  we obtain:

**Proposition 3.2.** *With probability one, as  $\varepsilon \rightarrow 0$ ,*

$$E^{(\tau_1/\varepsilon^2)}(Y_\varepsilon(\tau_2) - Y_\varepsilon(\tau_1)) = E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(X_\varepsilon(s)) ds + (\tau_2 - \tau_1 + \varepsilon^{1/2})o(1). \quad (3.18)$$

In the following results involving higher moments, we will take  $Y \in R^1$  for convenience. The more general case follows by the same arguments with suitable changes in notation.

**Lemma 3.5.** *With probability one, as  $\varepsilon \rightarrow 0$ ,*

$$E^{(t_k)} \xi_k^2 = L\tilde{\mathcal{F}}(X_\varepsilon^{(k)}) + o(L), \quad E^{(t_k)} \xi_k \xi_{k+1} = 0(1) \quad \text{and} \\ E^{(t_k)} \xi_k \xi_l = 0(L^{3/2} \varphi(L)) \quad (3.19)$$

for  $l > k + 1$ .

*Proof.* The first assertion follows from (A5) since

$$|E^{(t_k)} G^{(0)}(x, s) G^{(0)}(x, t) - \tilde{\mathcal{F}}(x, s, t)| = 0(\varphi(L))$$

for  $s, t \geq t_{k+1}$ , and since  $L^2 \varphi(L) = o(1)$  by (A4).

Next note that, for  $t_{k+1} \leq s \leq t_{k+2} \leq t$ ,

$$E^{(t_k)} [G^{(0)}(X_\varepsilon^{(k)}, s) G^{(0)}(X_\varepsilon^{(k+1)}, t)] = E^{(t_k)} [G^{(0)}(X_\varepsilon^{(k)}, s) E^{(s)} G^{(0)}(X_\varepsilon^{(k+1)}, t)] \\ = 0(\varphi(t - s)).$$

Integrating this expression over  $t_{k+1} \leq s \leq t_{k+2}$  and  $t_{k+2} \leq t \leq t_{k+3}$  yields the second assertion. For the third, use

$$E^{(t_k)} \xi_k \xi_l = E^{(t_k)} (\xi_k E^{(t)} \xi_l) = E^{(t_k)} |\xi_k| 0(L\varphi(L)) \leq (E^{(t_k)} \xi_k^2)^{1/2} 0(L\varphi(L)) \\ = 0(L^{3/2} \varphi(L))$$

by (3.17). □

**Lemma 3.6.** For  $0 \leq \tau_1 < \tau_2 \leq \tau_0$  and any fixed  $\tau_0 < \infty$ , with probability one, as  $\varepsilon \rightarrow 0$ ,

$$E^{(\tau_1/\varepsilon^2)} \left( \varepsilon \sum_{k=n_1}^{n_2} \zeta_k \right)^2 = E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mathcal{F}}(X_\varepsilon(\tau)) d\tau + (\tau_2 - \tau_1 + \varepsilon) o(1). \quad (3.20)$$

*Proof.* We have

$$\begin{aligned} E^{(\tau_1/\varepsilon^2)} \left( \varepsilon \sum_k \zeta_k \right)^2 &= E^{(\tau_1/\varepsilon^2)} \varepsilon^2 \sum_k \zeta_k^2 + 2E^{(\tau_1/\varepsilon^2)} \varepsilon^2 \sum_k \zeta_k \zeta_{k+1} + 2E^{(\tau_1/\varepsilon^2)} \varepsilon^2 \sum_{k+1 < l} \zeta_k \zeta_l \\ &=: I_1 + I_2 + I_3, \end{aligned}$$

where  $k$  and  $l$  lie between  $n_1$  and  $n_2$ . By (3.19)

$$I_1 = E^{(\tau_1/\varepsilon^2)} \varepsilon^2 \sum_k E^{(t_k)} \zeta_k^2 = E^{(\tau_1/\varepsilon^2)} \varepsilon^2 L \sum_k \tilde{\mathcal{F}}(X_\varepsilon^{(k)}) + (\tau_2 - \tau_1) o(1)$$

and, since  $\tilde{\mathcal{F}}(X_\varepsilon(\varepsilon^2 t)) - \tilde{\mathcal{F}}(X_\varepsilon^{(k)}) = 0(\varepsilon L)$  for  $t_k \leq t \leq t_{k+2}$ ,

$$\varepsilon^2 L \sum_k \tilde{\mathcal{F}}(X_\varepsilon^{(k)}) = \int_{\tau_1}^{\tau_2} \tilde{\mathcal{F}}(X_\varepsilon(\tau)) d\tau + (\tau_2 - \tau_1 + \varepsilon) o(1).$$

Also, by (3.19),  $I_2 = (\tau_2 - \tau_1) 0(L^{-1})$  and by (A4),  $\varepsilon^{-2} \varphi(L) = o(1)$ , so by (3.19)

$$I_3 = 0((\tau_2 - \tau_1)^2 \varepsilon^{-2} L^{-1/2} \varphi(L)) = (\tau_2 - \tau_1) o(1). \quad \square$$

**Lemma 3.7.** For some finite  $C$

$$\|\tilde{\gamma}_k\| \leq CL^2 \quad \text{and} \quad \|\zeta_k\| \leq CL, \quad (3.21)$$

and with probability one, as  $\varepsilon \rightarrow 0$ ,

$$E^{(t_k)} \tilde{\gamma}_k^2 = 0(L^3) \quad \text{and} \quad E^{(t_k)} \tilde{\gamma}_k \tilde{\gamma}_l = 0(L^4 \varphi(L)) \quad (3.22)$$

for  $l - k > 1$ .

*Proof.* The bounds in (3.21) are obvious. For (3.22) note that

$$\begin{aligned} E^{(t_k)} \tilde{\gamma}_k^2 &= 2E^{(t_k)} \int_{t_{k+1}}^{t_{k+2}} dr_1 \int_{t_k}^{t_{k+2}} ds_1 \int_{r_1}^{r_2} dr_2 \int_{t_k}^{r_2} ds_2 \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, r_1) f(X_\varepsilon^{(k)}, s_1) \\ &\quad \cdot \left( E^{(r_1)} \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, r_2) \right) f(X_\varepsilon^{(k)}, s_2). \end{aligned}$$

The first relation follows since

$$E^{(r_1)} \frac{\partial G^{(0)}}{\partial x}(X_\varepsilon^{(k)}, r_2) = 0(\varphi(r_2 - r_1)).$$

For the second relation, use

$$E^{(t_k)} \tilde{\gamma}_k \tilde{\gamma}_l = E^{(t_k)} (\tilde{\gamma}_k E^{(t_l)} \tilde{\gamma}_l) = 0(L^4 \varphi(L))$$

by (3.17) and (3.21). □

**Lemma 3.8.** With probability one, as  $\varepsilon \rightarrow 0$ ,

$$E^{(t_k)} \zeta_k^4 = 0(L^2), \quad E^{(t_k)} \hat{\gamma}_k^2 = 0(L^2) \quad \text{and} \quad E^{(t_k)} \hat{\gamma}_k^4 = 0(L^4). \quad (3.23)$$

These results are established by the same arguments given in [14], since  $G^{(0)}$ ,  $(\partial/\partial x)G^{(0)}$  and  $F^{(0)}$  are each bounded and have zero mean. It should be noted that Khas'minskii's proofs of these relations in [14] refer back to a mixing result he establishes in Lemma 2.1 of [13]. The final assertion in Lemma 3.8 requires the strongest hypothesis in order to apply his Lemma 2.1, namely that there is an  $\alpha$ -mixing coefficient  $\alpha(t)$  satisfying  $\int_0^\infty t^3 \alpha(t) dt < \infty$ . Of course, our assumption (A4) on the stronger  $\varphi$ -mixing coefficient implies this.

Now, let

$$\hat{Y}_\varepsilon(\tau) = \frac{1}{\varepsilon} \int_0^\tau \left\{ G^0 \left( X_\varepsilon(s), \frac{s}{\varepsilon^2} \right) + \varepsilon G^{(1)} \left( X_\varepsilon(s), \frac{s}{\varepsilon^2} \right) \right\} ds .$$

**Proposition 3.3.** *For  $0 \leq \tau_1 < \tau_2 \leq \tau_0$  and any fixing  $\tau_0 < \infty$  there is a finite  $C$  such that*

$$|E^{(\tau_1/\varepsilon^2)}(\hat{Y}_\varepsilon(\tau_2) - \hat{Y}_\varepsilon(\tau_1))^2 - E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mathcal{K}}(X_\varepsilon(t)) dt| \leq C(\tau_2 - \tau_1)^{3/2} + o(1) \quad (3.24)$$

with probability one, as  $\varepsilon \rightarrow 0$ .

*Proof.* Let  $\hat{Y}_\varepsilon(\tau_2) - \hat{Y}_\varepsilon(\tau_1) = U + V$ , where  $U = \varepsilon \sum_{k=n_1}^{n_2} \xi_k$ . We will show that  $E^{(\tau_1/\varepsilon^2)} V^2 \leq C_1(\tau_2 - \tau_1)^2 + o(1)$  a.s., and (3.24) will follow from this estimate and (3.20) by the Schwartz inequality, since (3.20) also implies  $E^{(\tau_1/\varepsilon^2)} U^2 \leq C_2(\tau_2 - \tau_1) + o(1)$ .

Letting  $\sum = \sum_{k=n_1}^{n_2}$ , (3.16) and the  $c_r$ -inequality yield

$$E^{(\tau_1/\varepsilon^2)} V^2 \leq 4 \{ \varepsilon^4 E^{(\tau_1/\varepsilon^2)} [(\sum \bar{\gamma}_k)^2 + (\sum \hat{\gamma}_k)^2 + (\sum \zeta_k)^2] + o(1) \} .$$

Now, by (3.21)

$$\varepsilon^4 E^{(\tau_1/\varepsilon^2)} (\sum \zeta_k)^2 \leq C_3(\tau_2 - \tau_1)^2 .$$

Also, by (3.22)

$$\begin{aligned} \varepsilon^4 E^{(\tau_1/\varepsilon^2)} (\sum \bar{\gamma}_k)^2 &= \varepsilon^4 E^{(\tau_1/\varepsilon^2)} \left[ \sum \bar{\gamma}_k^2 + 2 \sum \bar{\gamma}_k \bar{\gamma}_{k+1} + 2 \sum_{k+1 < l} \bar{\gamma}_k \bar{\gamma}_l \right] \\ &\leq C_4((\tau_2 - \tau_1)\varepsilon^2 L^2 + (\tau_2 - \tau_1)^2 L^2 \varphi(L)) = o(1) , \end{aligned}$$

and by (3.23)

$$\varepsilon^4 E^{(\tau_1/\varepsilon^2)} (\sum \hat{\gamma}_k)^2 \leq C_5(\tau_2 - \tau_1)^2 . \quad \square$$

**Lemma 3.9.** *For  $0 \leq \tau_1 < \tau_2 \leq \tau_0$  and any fixed  $\tau_0 < \infty$  there is a constant  $C$  depending only on  $\tau_0$  such that*

$$E(\hat{Y}_\varepsilon(\tau_2) - \hat{Y}_\varepsilon(\tau_1))^4 \leq C(\tau_2 - \tau_1)^2 + o(1) . \quad (3.25)$$

*Proof.* Letting  $\sum = \sum_{k=n_1}^{n_2}$ , we have by (3.16) and the  $c_r$ -inequality that the left-hand side of (3.25) is bounded by

$$C_0 E[\varepsilon^4 (\sum \xi_k)^4 + \varepsilon^8 \{ (\sum \bar{\gamma}_k)^4 + (\sum \hat{\gamma}_k)^4 + (\sum \zeta_k)^4 \}] + o(1) .$$

By (3.21)

$$\varepsilon^8 \|(\sum \zeta_k)^4\| \leq C_1 \varepsilon^8 L^4 \left( \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \right)^4 \leq C'_1 (\tau_2 - \tau_1)^2 .$$

Since  $E\hat{\gamma}_k^{4l} = 0(L^4)$ , it follows that

$$\varepsilon^8 E(\sum \hat{\gamma}_k)^4 \leq C_2 \varepsilon^8 L^4 \left( \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \right)^4 \leq C_2' (\tau_2 - \tau_1)^2.$$

By (3.17) and (3.21), if  $i \leq j \leq k < l - 1$ ,

$$|E(\bar{\gamma}_i \bar{\gamma}_j \bar{\gamma}_k \bar{\gamma}_l)| \leq E|\bar{\gamma}_i \bar{\gamma}_j \bar{\gamma}_k E^{(tl)} \bar{\gamma}_l| = 0(L^8 \varphi(L)),$$

and if  $i \leq j < k - 1$  and  $k = l - 1$  or  $k = l$  then by (3.22),

$$|E(\bar{\gamma}_i \bar{\gamma}_j \bar{\gamma}_k \bar{\gamma}_l)| \leq E|\bar{\gamma}_i \bar{\gamma}_j E^{(tk)} \bar{\gamma}_k \bar{\gamma}_l| = 0(L^7).$$

Now write

$$E(\sum \bar{\gamma}_k^4) = E(\sum \bar{\gamma}_i \sum \bar{\gamma}_j \sum \bar{\gamma}_k \sum \bar{\gamma}_l) = (\Sigma^{(1)} + \Sigma^{(2)} + \Sigma^{(3)})E(\bar{\gamma}_i \bar{\gamma}_j \bar{\gamma}_k \bar{\gamma}_l),$$

where, letting  $i' \leq j' \leq k' \leq l'$  be  $i, j, k, l$  ordered increasingly,  $\Sigma^{(1)}$  is the sum over indices  $i, j, k, l$  with  $k' < l' - 1$ ,  $\Sigma^{(2)}$  is the sum over  $j' < k' - 1$  and  $k' = l' - 1$  or  $k' = l'$ , and  $\Sigma^{(3)}$  is the sum over  $j' = k' - 1$  or  $j' = k'$  and  $k' = l' - 1$  or  $k' = l'$ . Then

$$\varepsilon^8 |\Sigma^{(1)}| \leq C_3 \varepsilon^8 L^8 \varphi(L) \left( \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \right)^4 = C_3 L^4 \varphi(L) (\tau_2 - \tau_1)^4,$$

$$\varepsilon^8 |\Sigma^{(2)}| \leq C_4 \varepsilon^8 L^7 \left( \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \right)^3 = C_4 \varepsilon^2 L^4 (\tau_2 - \tau_1)^3,$$

$$\varepsilon^8 |\Sigma^{(3)}| \leq C_5 \varepsilon^8 L^8 \left( \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \right)^2 = C_5 \varepsilon^4 L^6 (\tau_2 - \tau_1)^2.$$

By (A4),  $\varepsilon^8 E(\sum \bar{\gamma}_k^4) \leq C_6 (\tau_2 - \tau_1)^2$ . Finally, let  $\Sigma_i$  be  $\Sigma^{(i)}$  with  $\bar{\gamma}_k$  replaced by  $\xi_k$ . Arguments in Lemma 3.2 of [14] show  $\Sigma_2$  and  $\Sigma_3$  are  $0((\tau_2 - \tau_1)^2)$ . To estimate  $\Sigma_1$ , note that by (3.20), for  $n_1 \leq k \leq n_2$ ,

$$E \left| \sum_{i=n_1}^k \xi_i \right|^3 \leq E \left( \sum_{i=n_1}^k \xi_i \right)^2 \left\| \sum_{i=n_1}^k \xi_i \right\| \leq C_7 \frac{(\tau_2 - \tau_1)^2}{\varepsilon^4},$$

so by (3.17) and (A4),

$$\begin{aligned} \varepsilon^4 |\Sigma_1| &= 4\varepsilon^4 \left| \sum_{l=n_1+2}^{n_2} E \left[ \left( \sum_{i=n_1}^{l-2} \xi_i \right)^3 E^{(tl)} \xi_l \right] \right| \leq C_8 (\tau_2 - \tau_1)^2 L \varphi(L) \frac{\tau_2 - \tau_1}{\varepsilon^2 L} \\ &= (\tau_2 - \tau_1)^2 o(1). \end{aligned} \quad \square$$

**Lemma 3.10.** For some  $C < \infty$ , for all  $0 \leq s \leq t$ ,

$$E^{(s)} \left| \frac{1}{T} \int_t^{t+T} (f(x, s) - \bar{f}(x) + F^{(0)}(x, s)) ds \right|^2 \leq \frac{C}{T} \quad (3.26)$$

with probability one.

*Proof.* Condition (A2) implies

$$\left| \frac{1}{T} \int_t^{t+T} (f(x, s) - \bar{f}(x)) ds \right| \leq \frac{C_1}{T}.$$

The assertion follows from this and the estimate

$$E^{(s)} \left| \frac{1}{T} \int_t^{t+T} F^{(0)}(x, s) ds \right|^2 = 2E^{(s)} \frac{1}{T^2} \int_t^{t+T} du \int_u^{t+T} dv F^{(0)}(x, u)^T E^{(u)} F^{(0)}(x, v) \leq \frac{C_2}{T}$$

since  $E^{(u)} F^{(0)}(x, v) = 0(\varphi(v - u))$ . □

*Note.* By (3.24), for  $1 \leq l \leq \varepsilon^{-1}$ ,

$$E^{(\tau/\varepsilon^2)} |Y_\varepsilon(\tau + l\varepsilon) - Y_\varepsilon(\tau)|^2 \leq C\varepsilon l + o(1)$$

for some fixed  $C < \infty$ , so

$$P(|Y_\varepsilon(\tau + l\varepsilon) - Y_\varepsilon(\tau)| \geq d | \mathcal{F}_0^{\tau/\varepsilon^2}) \leq \frac{C\varepsilon l + o(1)}{d^2}.$$

Now, if  $X_\varepsilon(\tau) \in D_0$  then  $Y_\varepsilon(\tau) \in D_1$  and there is a positive distance  $d = d(Y_\varepsilon(\tau))$  from  $Y_\varepsilon(\tau)$  to  $\partial D_1$ . Recall that  $\beta_\varepsilon$  is the time  $X_\varepsilon(\tau)$  hits  $\partial D_0$ . Since  $X_\varepsilon(\tau)$  is stopped when it reaches  $\partial D_0$ ,  $Y_\varepsilon$  is stopped on reaching  $I(\partial D_0)$  and

$$\begin{aligned} P(\beta_\varepsilon > \tau + l\varepsilon | \mathcal{F}_0^{\tau/\varepsilon^2}, Y_\varepsilon(\tau) \in D_1) &= P(Y_\varepsilon(\tau + l\varepsilon) \in D_1 | \mathcal{F}_0^{\tau/\varepsilon^2}, Y_\varepsilon(\tau) \in D_1) \\ &\geq 1 - \frac{C\varepsilon l + o(1)}{d^2} \rightarrow 1 \end{aligned}$$

as  $\varepsilon \rightarrow 0$  and  $\varepsilon l \rightarrow 0$ .

The next result is a minor variation on the standard first order averaging result (see Theorem 1.1 of [13] and the related Theorem 1.2 of Chap. 7 of [10]).

**Proposition 3.4.** *For each fixed  $l$ ,*

$$P - \text{ess sup}_\tau \sup_{0 \leq t \leq \min(l, (\beta_\varepsilon - \tau)/\varepsilon)} E^{(\tau/\varepsilon^2)} |X_\varepsilon(\tau + \varepsilon t) - \bar{x}(t, X_\varepsilon(\tau))| \rightarrow 0 \quad (3.27)$$

as  $\varepsilon \rightarrow 0$ .

On these time scales, we can and do take  $F^{(1)} = 0$  and  $F = f + F^{(0)}$ . Our result replaces expectation by the conditional expectation  $E^{(\tau/\varepsilon^2)}$  in the results mentioned above. Since  $|X_\varepsilon(\tau + \varepsilon t)| \leq |X_\varepsilon(\tau)| + Ct$  with probability one, it suffices to prove that, for each  $\alpha > 0$ ,

$$P - \text{ess sup} P \left( \sup_{0 \leq t \leq \min(l, (\beta_\varepsilon - \tau)/\varepsilon)} |X_\varepsilon(\tau + \varepsilon t) - \bar{x}(t, X_\varepsilon(\tau))| > \alpha \mid \mathcal{F}_0^{\tau/\varepsilon^2} \right) \rightarrow 0$$

uniformly in  $\tau$ . This follows from the estimates in the proof of Theorem 1.3 of Chapter 2 of [10], replacing probability and expectation by the conditional probability and expectation given  $\mathcal{F}_0^{\tau/\varepsilon^2}$ . The regularity conditions (A1), (A2) and Lemma 3.10, which they imply, are more than is needed, and the uniformity in  $\tau$  and almost all  $\omega$  for the conditional probability follows from the uniformity of these conditions.

**Lemma 3.11.** *With probability one, as  $\varepsilon \rightarrow 0$ ,*

$$\begin{aligned} E^{(\tau_1/\varepsilon^2)}(Y_\varepsilon(\tau_2) - Y_\varepsilon(\tau_1)) &= E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(Y_\varepsilon(s)) ds + (\tau_2 - \tau_1)o(1), \\ E^{(\tau_1/\varepsilon^2)}(Y_\varepsilon(\tau_2) - Y_\varepsilon(\tau_1))^2 &= E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mathcal{X}}(Y_\varepsilon(s)) ds \\ &\quad + 0((\tau_2 - \tau_1)^{3/2}) + (\tau_2 - \tau_1)o(1). \end{aligned} \quad (3.28)$$

*Proof.* We prove the first assertion, the proof of the second being similar.

1. First note that

$$\begin{aligned} \int_{\tau}^{\tau+\varepsilon l} \tilde{\mu}(X_\varepsilon(s)) ds &= \varepsilon \int_0^l \tilde{\mu}(X_\varepsilon(\tau + \varepsilon r)) dr \\ &= \varepsilon \int_0^l \tilde{\mu}(\bar{x}(r, X_\varepsilon(\tau))) dr + \varepsilon \int_0^l (\tilde{\mu}(X_\varepsilon(\tau + \varepsilon r)) \\ &\quad - \tilde{\mu}(\bar{x}(r, X_\varepsilon(\tau)))) dr \\ &=: I_1 + I_2. \end{aligned}$$

Assumptions (A1) and (A3) imply that  $\tilde{\mu}$  has bounded derivative, and applying (3.27),

$$E^{(\tau/\varepsilon^2)}|I_2| \leq C\varepsilon \int_0^l E^{(\tau/\varepsilon^2)}|X_\varepsilon(\tau + \varepsilon r) - \bar{x}(r, X_\varepsilon(\tau))| dr = \varepsilon l o(1). \quad (3.29)$$

Now, let

$$e(x, l) := \tilde{\mu}(y(x)) - \frac{1}{l} \int_0^l \tilde{\mu}(\bar{x}(s, x)) ds.$$

Then

$$I_1 = \varepsilon l (\tilde{\mu}(Y_\varepsilon(\tau)) - e(X_\varepsilon(\tau), l)). \quad (3.30)$$

2. Let (A6) hold. Then, by (3.29) and (3.30)

$$\begin{aligned} E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(X_\varepsilon(s)) ds &= E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} ds \frac{1}{\varepsilon l} \int_s^{s+\varepsilon l} dr \tilde{\mu}(X_\varepsilon(r)) + 0(\varepsilon l) \\ &= E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(Y_\varepsilon(s)) ds + \int_{\tau_1}^{\tau_2} (e(X_\varepsilon(s), l) + o(1)) ds + 0(\varepsilon l), \end{aligned}$$

and the last two terms are at most

$$(\tau_2 - \tau_1) \left( \sup_x e(x, l) + o(1) \right) + 0(\varepsilon l) \rightarrow 0$$

as  $\varepsilon \rightarrow 0$  then  $l \rightarrow \infty$  by (A6).

3. Now, suppose (A6') holds. We will establish the result assuming  $D_0$  bounded. The lemma will then follow for this case and can be extended to unbounded  $D_0$  by choosing a suitable sequence of bounded  $D_0^{(n)} \uparrow D_0$ .

By the first assumption of (A6') and Egoroff's theorem,  $e(x, l) \rightarrow 0$  almost uniformly on  $D_0$ . Thus, given  $\delta > 0$ , there is an  $l_\delta$  such that  $m\{x: e(x, l) > \delta\} < \delta$  for  $l \geq l_\delta$  (where  $m$  denotes Lebesgue measure in  $R^p$ ). By the third assumption of (A6'), for  $l \geq l_\delta$ ,

$$\begin{aligned} E^{(\tau_1/\varepsilon^2)} e(X_\varepsilon(\tau_1 + h), l) &= \int_{D_0} e(x, l) p_\varepsilon(\tau_1, \tau_1 + h, x) dx \\ &\leq \delta + B p_{\bar{D}_0}(\tau_1, \tau_1 + h) \delta \\ &= \delta(1 + B p_{\bar{D}_0}(\tau_1, \tau_1 + h)), \end{aligned}$$

where  $B = \|e\|$ . Thus

$$\begin{aligned} E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(X_\varepsilon(s)) ds &= E^{(\tau_1/\varepsilon^2)} \int_{\tau_1}^{\tau_2} \tilde{\mu}(Y_\varepsilon(s)) ds \\ &\quad + \int_{\tau_1}^{\tau_2} E^{(\tau_1/\varepsilon^2)} e(X_\varepsilon(\tau_1 + s), l) ds + 0(\varepsilon l). \end{aligned}$$

The second two terms are at most

$$tB + \delta \int_{\tau_1+t}^{\tau_2} (1 + B p_{\bar{D}_0}(\tau_1, \tau_1 + s)) ds + 0(\varepsilon l).$$

The first assertion follows by letting  $\varepsilon \rightarrow 0$  then  $l \rightarrow \infty$  then  $\delta \rightarrow 0$  then  $t \rightarrow 0$ . □

Finally, we note the following standard result (see [14]):

**Lemma 3.12.** *Let  $Z_n \rightarrow Z$  in probability and  $P(AA_n^c \cup A^c A_n) \rightarrow 0$ . Suppose  $E|Z_n|^{1+\alpha} \leq C < \infty$  for all  $n$  and some  $\alpha > 0$ . Then*

$$\int_{A_n} Z_n dP \rightarrow \int_A Z dP.$$

#### 4. Proof of the Theorem

The estimates in Proposition 3.1 and Lemma 3.9 imply that the processes  $\{Y_\varepsilon(\tau), 0 \leq \tau \leq \tau_0\}$  are tight as  $\varepsilon \rightarrow 0$  and that  $\sup_{\tau \leq \tau_0} E|Y_\varepsilon(\tau)|^4 < \infty$  (see [1]). Now, consider any weakly convergent sequence  $Y_{\varepsilon_n} \rightarrow Y_0$ . It follows from Propositions 3.1 and 3.3 and Chebychev's inequality that, for any  $\delta > 0$ ,

$$\lim_{h \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{|\tau_2 - \tau_1| < h} P[|Y_{\varepsilon_n}(\tau_2) - Y_{\varepsilon_n}(\tau_1)| > \delta] = 0.$$

A theorem of Skorokhod (Sect. 6, Chap. 1 of [16]) then states that there exists a sequence of stochastic processes  $\{Y'_{\varepsilon_n}\}$  and a process  $Y'_0$  on some probability space  $(\Omega', \mathcal{F}', P')$  such that the finite dimensional distributions of the  $Y'_{\varepsilon_n}$  are the same as those of the  $Y_{\varepsilon_n}$  and  $Y'_{\varepsilon_n}(\tau) \rightarrow Y'_0(\tau)$  in probability as  $n \rightarrow \infty$  for each  $\tau$ .

Using this Skorokhod imbedding, we will treat the  $Y_{\varepsilon_n}(\tau)$  as though  $Y_{\varepsilon_n}(\tau) \rightarrow Y_0(\tau)$  in probability in the following argument. First, note that this and

boundedness and continuity of  $\bar{\mu}$  imply

$$\lim_{n \rightarrow \infty} E \left| \int_{\tau}^{\tau+h} \bar{\mu}(Y_{\varepsilon_n}(s)) ds - \int_{\tau}^{\tau+h} \bar{\mu}(Y_0(s)) ds \right| = 0 .$$

Now, let  $\tau_1 \leq \tau_2 \leq \dots \leq \tau_k \leq \tau < \tau + h$ ,  $B$  be a Borel set in  $(R^q)^k$  and let

$$A_n = [(Y_{\varepsilon_n}(\tau_1), \dots, Y_{\varepsilon_n}(\tau_k)) \in B] ,$$

$$A = [(Y_0(\tau_1), \dots, Y_0(\tau_k)) \in B] .$$

Choose  $B$  so  $P[(Y_0(\tau_1), \dots, Y_0(\tau_k)) \in \partial B] = 0$ . Then  $P(A_n A^c \cup A_n^c A) \rightarrow 0$  and, by Lemmas 3.11 and 3.12,

$$\begin{aligned} \int_A (Y_0(\tau+h) - Y_0(\tau)) dP &= \lim_{n \rightarrow \infty} \int_{A_n} (Y_{\varepsilon_n}(\tau+h) - Y_{\varepsilon_n}(\tau)) dP \\ &= \lim_{n \rightarrow \infty} \int_{A_n} E^{(\tau/\varepsilon^2)}(Y_{\varepsilon_n}(\tau+h) - Y_{\varepsilon_n}(\tau)) dP \\ &= \lim_{n \rightarrow \infty} \int_{A_n} \int_{\tau}^{\tau+h} \bar{\mu}(Y_{\varepsilon_n}(s)) ds dP \\ &= \int_A \int_{\tau}^{\tau+h} \bar{\mu}(Y_0(s)) ds dP . \end{aligned}$$

Thus, letting  $\mathcal{A}_{\tau}$  be the  $\sigma$ -field generated by  $Y_0(t)$ ,  $0 \leq t \leq \tau$ , we have

$$E(Y_0(\tau+h) - Y_0(\tau) | \mathcal{A}_{\tau}) = E \left( \int_{\tau}^{\tau+h} \bar{\mu}(Y_0(s)) ds | \mathcal{A}_{\tau} \right) ,$$

and by a similar argument

$$E((Y_0(\tau+h) - Y_0(\tau))^2 | \mathcal{A}_{\tau}) = E \left( \int_{\tau}^{\tau+h} \bar{\mathcal{K}}(Y_0(s)) ds | \mathcal{A}_{\tau} \right) + 0(h^{3/2}) .$$

The remainder of our argument is similar to that of Borodin [2]. Let

$$\bar{Y}_0(\tau) = Y_0(\tau) - \int_0^{\tau} \bar{\mu}(Y_0(s)) ds .$$

Then  $\bar{Y}_0$  is a martingale. Since  $\bar{\mu}$  is bounded,

$$\left| \int_A (\bar{Y}_0(\tau+h) - \bar{Y}_0(\tau))^2 dP - \int_A \int_{\tau}^{\tau+h} \bar{\mathcal{K}}(Y_0(s)) ds dP \right| \leq Ch^{3/2}$$

for  $A \in \mathcal{A}_{\tau}$ . It follows easily that

$$E((\bar{Y}_0(\tau+h) - \bar{Y}_0(\tau))^2 | \mathcal{A}_{\tau}) = E \left( \int_{\tau}^{\tau+h} \bar{\mathcal{K}}(Y_0(s)) ds | \mathcal{A}_{\tau} \right) .$$

Then a theorem of Doob [6] yields that

$$\bar{Y}_0(\tau) = \bar{Y}_0(0) + \int_0^{\tau} \sqrt{\bar{\mathcal{K}}(Y_0(s))} dW(s) ,$$

where  $W(s)$  is the  $q$ -dimensional Wiener process. Hence

$$Y_0(\tau) = I(x_0) + \int_0^\tau \bar{\mu}(Y_0(s)) ds + \int_0^\tau \sqrt{\bar{\Sigma}(Y_0(s))} dW(s). \quad (4.1)$$

Since the distribution of  $Y_0(\tau)$  is the same for all convergent sequences  $Y_{\varepsilon_n}$  of  $Y_\varepsilon$ , we have that  $Y_\varepsilon(\tau)$  converges weakly to  $Y_0(\tau)$  on  $[0, \infty)$ , and the form of the infinitesimal generator follows from (4.1) (e.g., see Sect. 1.5 of [10]).

The final assertion follows from the weak convergence when  $I(\bar{D}_0)$  is closed since  $P[Y_\varepsilon(\tau) \in I(\bar{D}_0) \text{ for } 0 \leq \tau \leq \tau_0] = 1$  for all  $\tau_0$  and  $\varepsilon > 0$ .  $\square$

## References

1. Billingsley, P.: Convergence of probability measures. New York: Wiley 1968
2. Borodin, A.N.: A limit theorem for solutions of differential equations with random right-hand side. *Theory Prob. Appl.* **22**, 482–497 (1977)
3. Cogburn, R., Ellison, J.A.: Particle motion in a rapidly varying field. 1990 (submitted)
4. Cogburn, R., Ellison, J.A., Newberger, B.S., Shih, H.J.: private communication
5. Dôme, G.: Diffusion due to RF noise. CERN Advanced Accelerator School, Advanced Accelerator Physics, CERN Report No. 87-03 (1987), pp. 370–401
6. Doob, J.L.: Stochastic processes. New York: Wiley 1956
7. Dumas, H.S., Ellison, J.A., Sáenz, A.W.: Axial channeling in perfect crystals, the continuum model and the method of averaging. *Ann. Phys.* **209**, 97–123 (1991)
8. Ellison, J.A., Sáenz, A.W., Dumas, H.S.: Improved  $N^{\text{th}}$  order averaging theory for periodic systems. *J. Diff. Eq.* **84**, 383 (1990)
9. Fink, A.M.: Almost periodic differential equations. *Lecture Notes in Mathematics* vol. **377**, Berlin, Heidelberg, New York: Springer 1974
10. Freidlin, M.I., Wentzell, A.D.: Random perturbations of dynamical systems. Berlin, Heidelberg, New York: Springer 1984
11. Hale, J.K.: Ordinary differential equations. New York: Wiley 1969
12. Hardy, G.H., Wright, E.M.: An introduction to the theory of numbers. Oxford: Oxford University Press, fifth edition, 1979
13. Khas'minskii, R.Z.: On processes defined by differential equations with a small parameter. *Theory Prob. Appl.* **11**, 211–228 (1966)
14. Khas'minskii, R.Z.: A limit theorem for solutions of differential equations with random right-hand side. *Theory Prob. Appl.* **11**, 390–406 (1966)
15. Knapp, R., Papanicolaou, G., White, B.: Nonlinearity and localization in one dimensional random media. In: *Disorder and Nonlinearity*. Bishop, A.R., Campbell, D.K., Pnevmticas, S. (eds.). *Proceedings in Physics* vol. **39**, pp. 2–26. Berlin, Heidelberg, New York: Springer 1989 See also Knapp, R.J.: Nonlinearity and localization in one dimensional random media. Ph.D. thesis, New York University, 1988
16. Skorokhod, A.V.: Studies in the theory of random processes. Reading, MA: Addison-Wesley 1965

Communicated by J.L. Lebowitz