Averaging principle for quasi-linear parabolic PDE's and related diffusion processes

M. Freidlin[*], L. Koralov[†]

Abstract

Quasi-linear perturbations of a two-dimensional flow with a first integral and the corresponding parabolic PDE's with a small parameter at the second order derivatives are considered in this paper.

2000 Mathematics Subject Classification Numbers: 35K55

Keywords: averaging, quasi-linear parabolic equations, perturbations of dynamical systems.

1 Introduction

Let v be a C^2-smooth incompressible vector field on $\mathbb{R}^2$ with the stream function H. Let a domain $D \subset \mathbb{R}^2$ be topologically equivalent to an annulus bounded by the curves γ_1 and γ_2, such that $H(x) = h_1$ for $x \in \gamma_1$, and $H(x) = h_2$ for $x \in \gamma_2$ with $h_1 < h_2$. We also assume at first that H does not have critical points in $\overline{D}$.

Let g be a C^2 function defined in a neighborhood of $[h_1, h_2]$ and $c_1 = g(h_1)$, $c_2 = g(h_2)$. Let $a_{ij} = a_{ji} \in C^2(\mathbb{R}^3)$, $1 \le i, j \le 2$, and $b_i \in C^2(\mathbb{R}^3)$, $1 \le i \le 2$. We assume that there is a positive constant k such that $k|\xi|^2 \le \sum_{i,j=1}^2 a_{ij}(x,u)\xi_i\xi_j$, $x \in D, u \in \mathbb{R}$, $\xi \in \mathbb{R}^2$.

In this paper we study the asymptotic behavior, as $\varepsilon \downarrow 0$, of solutions to the quasi-linear parabolic equation

$$\frac{\partial u^\varepsilon(t,x)}{\partial t} = \frac{1}{2}\sum_{i,j=1}^2 a_{ij}(x,u^\varepsilon)\frac{\partial^2 u^\varepsilon(t,x)}{\partial x_i \partial x_j} + \left(\frac{1}{\varepsilon}v(x) + b(x,u^\varepsilon)\right)\cdot \nabla_x u^\varepsilon(t,x), \quad x \in D, \quad t > 0, \tag{1}$$

$$u^\varepsilon(0,x) = g(H(x)),\ x \in D;\ \ u^\varepsilon(t,x) = c_1,\ t \ge 0,\ x \in \gamma_1,\ \ u^\varepsilon(t,x) = c_2,\ t \ge 0,\ x \in \gamma_2 \tag{2}$$

and the diffusion processes related to (1)-(2). Under the above assumptions, the solution u^ε exists and is unique in the class of functions that are continuous on $[0,\infty) \times \overline{D}$, and

[*]Dept of Mathematics, University of Maryland, College Park, MD 20742, mif@math.umd.edu

[†]Dept of Mathematics, University of Maryland, College Park, MD 20742, koralov@math.umd.edu

have partial derivative in t and partial derivatives up to the second order in x which are bounded and continuous in $(0, \infty) \times D$ (see Theorem 5, Chapter 6.2 of [5]). Later, we also briefly discuss the case of general boundary and initial conditions.

We will show that the u^ε converges to the solution of the averaged quasi-linear equation. Namely, let

$$T(h) = \int_{\gamma(h)} \frac{1}{|\nabla H|} dl, \tag{3}$$

$$\overline{a}(h, u) = \frac{1}{T(h)} \int_{\gamma(h)} \frac{\langle a(x, u)\nabla H, \nabla H\rangle}{|\nabla H|} dl, \quad \text{and} \tag{4}$$

$$\overline{v}(h, u) = \frac{1}{2T(h)} \int_{\gamma(h)} \frac{a(x, u) \cdot H'' + 2\langle b(x, u), \nabla H\rangle}{|\nabla H|} dl, \tag{5}$$

where $a \cdot H''(x) = \sum_{1 \le i,j \le 2} a_{ij}(x) \partial^2 H(x)/\partial_{x_i}\partial_{x_j}$, $\gamma(h) = \{x \in \overline{D} : H(x) = h\}$, and u is a real parameter. Let $\overline{u}$ solve

$$\frac{\partial \overline{u}(t, h)}{\partial t} = \frac{1}{2}\overline{a}(h, \overline{u})\overline{u}''_{hh}(t, h) + \overline{v}(h, \overline{u})\overline{u}'_h(t, h), \quad h \in [h_1, h_2], \quad t > 0, \tag{6}$$

$$\overline{u}(0, h) = g(h),\ h \in [h_1, h_2]; \quad \overline{u}(t, h_1) = c_1,\ t \ge 0, \quad \overline{u}(t, h_2) = c_2,\ t \ge 0.$$

The solution exists and is unique in the same class of functions as above with D replaced by (h_1, h_2). The main result of this paper is the following.

Theorem 1.1. *For each $x \in \overline{D}$ and $t > 0$ we have*

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(t, x) = \overline{u}(t, H(x)). \tag{7}$$

The convergence is uniform on $[0, t_0] \times \overline{D}$ for each $t_0 > 0$.

Before we proceed with the proof of the theorem, let us briefly discuss the situation in the linear case ($b(x, u) = b(x)$ and $a(x, u) = a(x)$), where the result is well-known. Namely, consider the process $X_t^{x,\varepsilon}$ that starts at x and solves

$$dX_t^{x,\varepsilon} = \frac{1}{\varepsilon} v(X_t^{x,\varepsilon})dt + b(X_t^{x,\varepsilon})dt + \sigma(X_t^{x,\varepsilon})dW_t, \quad t \le \tau^\varepsilon, \tag{8}$$

$$\tau^\varepsilon = \min\{t : X_t^{x,\varepsilon} \in \partial D\}, \qquad X_t^{x,\varepsilon} = X_{\tau^\varepsilon}^{x,\varepsilon}, \quad \tau^\varepsilon \le t.$$

Here W_t is a Wiener process in $\mathbb{R}^2$, and $\sigma(x) = a^{1/2}(x)$. The motion in (8) consists of the fast (due to the presence of $1/\varepsilon$ factor in front of v) advection along the flow lines of v perturbed by the slower advection and diffusion. The limiting behavior (as $\varepsilon \downarrow 0$) of the slow component of $X_t^{x,\varepsilon}$ can be described in terms of the diffusion process on the graph corresponding to the structure of the level sets of H (the general result can be found in [3], Chapter 8). Here, we restrict ourselves to the case when $H(x)$ has no critical points in $\overline{D}$, and so the corresponding graph is just the segment $[h_1, h_2]$. The general case is briefly considered in Section 4.

We introduce the process Y_t^h on $[h_1, h_2]$ that will serve as the limit of $H(X_t^{x,\varepsilon})$. Let $\overline{a}(h)$ and $\overline{v}(h)$ be defined by (4) and (5), respectively, where the right hand sides now don't depend on u. Let $\overline{\sigma}(h) = \sqrt{\overline{a}(h)}$. Let Y_t^h be the process that starts at h and solves

$$dY_t^h = \overline{v}(Y_t^h)dt + \overline{\sigma}(Y_t^h)d\overline{W}_t, \quad t \leq \overline{\tau},$$

$$\overline{\tau} = \min\{t : Y_t^h \in \{h_1, h_2\}\}, \qquad Y_t^h = Y_{\overline{\tau}}^h, \quad \overline{\tau} \leq t,$$

where $\overline{W}_t$ is a one-dimensional Wiener process. The following theorem can be found in [3], Chapter 8.

Theorem 1.2. *Let $x \in D$. Under the above assumptions, the measures on the space $C([0,\infty), [h_1, h_2])$ induced by the processes $H(X_t^{x,\varepsilon})$ converge weakly, as $\varepsilon \downarrow 0$, to the measure induced by the process $Y_t^{H(x)}$.*

Remark 1.3. *The coefficients b and σ, and consequently $\overline{v}$ and $\overline{\sigma}$, may be allowed to depend on t. Moreover, the weak convergence claimed in Theorem 1.2 can be proved to be uniform in certain parameters. For example, for each $t_0, C > 0$, and continuous function F on $[h_1, h_2]$, the following limit*

$$\lim_{\varepsilon \downarrow 0} \mathrm{E}\left(F(H(X_t^{x,\varepsilon})) - F(Y_t^{H(x)}) \right) = 0$$

is uniform in $t \in [0, t_0]$, $x \in D$ and b, $\sigma \in \mathcal{S}_C$, where $\mathcal{S}_C$ is the space of all functions of t and x that are bounded by C, in absolute value, together with the first order partial derivatives.

This theorem together with the representation of the solution of the PDE as the expectation of the functional of the process

$$u^\varepsilon(t,x) = \mathrm{E}g(H(X_t^{x,\varepsilon})), \tag{9}$$

easily imply Theorem 1.1 in the case when a does not depend on u. (Note that in (9) we used the fact that $c_1 = g(h_1)$ and $c_2 = g(h_2)$.)

The difficulty in the non-linear case is due to the fact that now the family of diffusion processes corresponding to the operator in the right hand side of (1) depends on the unknown function u^ε because of the dependence of the coefficients a and b on u^ε. Namely, for $t > 0$ and $x \in \overline{D}$, we can define $X_s^{t,x,\varepsilon}$, $s \in [0,t]$, as the process which starts at x and solves

$$dX_s^{t,x,\varepsilon} = \frac{1}{\varepsilon} v(X_s^{t,x,\varepsilon})ds + b(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))ds + \sigma(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))dW_s, \quad s \leq \tau^\varepsilon \wedge t, \tag{10}$$

$$\tau^\varepsilon = \min\{s : X_s^{t,x,\varepsilon} \in \partial D\}, \qquad X_s^{t,x,\varepsilon} = X_{\tau^\varepsilon}^{t,x,\varepsilon}, \quad \tau^\varepsilon \leq s \leq t, \tag{11}$$

where $\sigma = a^{1/2}$. As in the linear case, we have the following relation between u^ε and the process $X_s^{t,x,\varepsilon}$:

$$u^\varepsilon(t,x) = \mathrm{E}g(H(X_t^{t,x,\varepsilon})). \tag{12}$$

In fact, (10)-(12) can be viewed as a system of equations for the unknown function u^ε and the family of processes $X_s^{t,x,\varepsilon}$ that is equivalent to (1)-(2) (see [2], Chapter 5).

In the next section we use the relationship between the processes $X_s^{t,x,\varepsilon}$ and the solution u^ε in order to prove Theorem 1.1. A lemma that is used in the proof is, in turn, proved in Section 3. In Section 4 we discuss several generalizations. One concerns the systems with several degrees of freedom. The other concerns more general initial and boundary value functions. We also discuss the case when the domain D contains critical points of H.

2 Proof of the main theorem

First, let us prove the uniform convergence for $(t,x) \in [0,t_0] \times \overline{D}$ if t_0 is sufficiently small. The value of t_0 will be specified later. If the convergence does not hold, then there are $\varkappa > 0$ and sequences $\varepsilon^k \downarrow 0$, $s_0^k \in [0,t_0]$, $x^k \in D$ such that

$$|u^{\varepsilon^k}(s_0^k, x^k) - \overline{u}(s_0^k, H(x^k))| \geq \varkappa. \tag{13}$$

Due to the continuity of u^ε and $\overline{u}$, we can pick a smaller (if necessary) value of s_0^k such that (13) still holds and

$$\sup_{(s,x)\in[0,s_0^k]\times D} |u^{\varepsilon^k}(s,x) - \overline{u}(s,H(x))| \leq \varkappa, \tag{14}$$

and consequently the inequality in (13) becomes an equality. We will drop the superscript k from the notation, simply keeping in mind that an arbitrarily small ε can be taken.

For $t > 0$, $h \in [h_1,h_2]$, let $Y_s^{t,h}$, $s \in [0,t]$, be the process which starts at h and solves

$$dY_s^{t,h} = \overline{v}(Y_s^{t,h})ds + \overline{\sigma}(Y_s^{t,h}, \overline{u}(t-s, Y_s^{t,h}))d\overline{W}_s, \quad s \leq \overline{\tau} \wedge t,$$

$$\overline{\tau} = \min\{s : Y_s^{t,h} \in \{h_1,h_2\}\}, \qquad Y_s^{t,h} = Y_{\overline{\tau}}^{t,h}, \quad \overline{\tau} \leq s \leq t,$$

where $\overline{\sigma} = \sqrt{\overline{a}}$. For $t > 0$ and $x \in \overline{D}$, we also introduce $Z_s^{t,x,\varepsilon}$, $s \in [0,t]$, as the process which starts at x and solves

$$dZ_s^{t,x,\varepsilon} = \frac{1}{\varepsilon}v(Z_s^{t,x,\varepsilon})ds + b(Z_s^{t,x,\varepsilon}, \overline{u}(t-s, H(Z_s^{t,x,\varepsilon})))ds +$$
$$\sigma(Z_s^{t,x,\varepsilon}, \overline{u}(t-s, H(Z_s^{t,x,\varepsilon})))dW_s, \quad s \leq \widetilde{\tau}^\varepsilon \wedge t,$$

$$\widetilde{\tau}^\varepsilon = \min\{s : Z_s^{t,x,\varepsilon} \in \partial D\}, \qquad Z_s^{t,x,\varepsilon} = Z_{\widetilde{\tau}^\varepsilon}^{t,x,\varepsilon}, \quad \widetilde{\tau}^\varepsilon \leq s \leq t.$$

Then

$$|u^\varepsilon(s_0,x) - \overline{u}(s_0,H(x))| = |\mathrm{E}g(H(X_{s_0}^{s_0,x,\varepsilon})) - \mathrm{E}g(Y_{s_0}^{s_0,H(x)})| \leq$$
$$|\mathrm{E}g(H(X_{s_0}^{s_0,x,\varepsilon})) - \mathrm{E}g(H(Z_{s_0}^{s_0,x,\varepsilon}))| + |\mathrm{E}g(H(Z_{s_0}^{s_0,x,\varepsilon})) - \mathrm{E}g(Y_{s_0}^{s_0,H(x)})|.$$

By taking $\widetilde{b}(s,x) = b(x, \overline{u}(s_0 - s, H(x)))$, $\widetilde{\sigma}(s,x) = \sigma(x, \overline{u}(s_0 - s, H(x)))$ and $F = g$, we see from Remark 1.3 that the second term on the right hand side here can be made smaller than $\varkappa/2$ for all $(s_0, x) \in [0, t_0] \times \overline{D}$ and all sufficiently small ε. In order to deal with the first term we need the following lemma that will be proved in the next section.

Lemma 2.1. *Consider two processes $Z_t^{1,\varepsilon}$ and $Z_t^{2,\varepsilon}$ which solve*

$$dZ_t^{i,\varepsilon} = \frac{1}{\varepsilon} v(Z_t^{i,\varepsilon})dt + b_i(t, Z_t^{i,\varepsilon})dt + \sigma_i(t, Z_t^{i,\varepsilon})dW_t, \quad t \le \tau^{i,\varepsilon}, \qquad Z_0^{i,\varepsilon} = x \in D,$$

$$\tau^{i,\varepsilon} = \min\{t : Z_t^{i,\varepsilon} \in \partial D\}, \qquad Z_t^{i,\varepsilon} = Z_{\tau^{i,\varepsilon}}^{i,\varepsilon}, \quad \tau^{i,\varepsilon} \le t.$$

We assume that $|b_1|, |b_2|, |\sigma_1|, |\sigma_2| \le M$ for some positive constant M, σ_1 and σ_2 are uniformly elliptic with a constant $m > 0$, b_1 and σ_1 are Lipschitz continuous on $[0, \infty) \times \overline{D}$ with the constant M, b_2 and σ_2 are Lipschitz continuous and

$$\sup_{(t,x) \in (0,\infty) \times D} |b_1(t,x) - b_2(t,x)| \le \eta. \tag{15}$$

$$\sup_{(t,x) \in (0,\infty) \times D} |\sigma_1(t,x) - \sigma_2(t,x)| \le \eta. \tag{16}$$

Let $f \in C^2((h_1, h_2), \mathbb{R}) \bigcap C([h_1, h_2], \mathbb{R})$ be a function whose C^2-norm is bounded by M. For each choice of m, M and $S > 0$ there exists a positive constant K such that

$$|\mathrm{E} f(H(Z_t^{1,\varepsilon})) - \mathrm{E} f(H(Z_t^{2,\varepsilon}))| \le K\eta\sqrt{t} \tag{17}$$

whenever $x \in D$, $\eta > 0$, $\varepsilon \le 1$, $t \le S$, and b_2, σ_2 satisfy (15)-(16).

By (14), the drift and diffusion coefficients of the process $X_s^{s_0,x,\varepsilon}$ differ from the respective coefficients of the process $Z_s^{s_0,x,\varepsilon}$ by no more than $\eta = L\varkappa$, where L is the Lipschitz constant of the functions $b(x,u)$ and $\sigma(x,u)$. We can thus apply Lemma 2.1 (with $f = g$ and M determined by b, σ, $\overline{u}$ and g) to the processes $Z_s^{1,\varepsilon} = Z_s^{s_0,x,\varepsilon}$ and $Z_s^{2,\varepsilon} = X_s^{s_0,x,\varepsilon}$ to obtain that

$$|\mathrm{E} g(H(X_{s_0}^{s_0,x,\varepsilon})) - \mathrm{E} g(H(Z_{s_0}^{s_0,x,\varepsilon}))| \le KL\varkappa\sqrt{s_0}.$$

If we take t_0 such that $KL\sqrt{t_0}/2 < 1$, then the right hand side is bounded from above by $\varkappa/2$, which leads to a contradiction with (13), thus proving uniform convergence for $(t,x) \in [0, t_0] \times \overline{D}$.

Now let us show the uniform convergence for $(t,x) \in [t_0, T_0] \times \overline{D}$, where $T_0 > t_0$ is arbitrary. Again, if the convergence does not hold, then there are $\varkappa > 0$ and sequences $\varepsilon^k \downarrow 0$, $s_0^k \in [t_0, T_0]$, $x^k \in D$ such that (13) holds. Let us choose positive constants α, β and $r < t_0$ such that

$$\alpha \exp(\beta T_0) < \varkappa, \tag{18}$$

$$\beta r < 1, \tag{19}$$

$$\beta \sqrt{r} \ge \widetilde{K}, \tag{20}$$

where the value of $\widetilde{K}$ will be specified later. Due to (18) and the continuity of u^ε and $\overline{u}$, by choosing if necessary a smaller value of s_0^k we can assume, without loss of generality, that instead of (13) we have for some x

$$|u^{\varepsilon^k}(s_0^k, x) - \overline{u}(s_0^k, H(x))| = \alpha \exp(\beta s_0^k), \tag{21}$$

while

$$\sup_{x\in D} |u^{\varepsilon^k}(s, x) - \overline{u}(s, H(x))| \le \alpha \exp(\beta s) \tag{22}$$

for each $s \le s_0^k$. Note that $s_0^k > t_0$ for all sufficiently small ε^k, since otherwise (21) would contradict the uniform convergence on $[0, t_0] \times \overline{D}$ established above. Let us drop the the superscript k again. By (21) and (22), we have

$$\begin{aligned}
\alpha \exp(\beta s_0) = |u^\varepsilon(s_0, x) - \overline{u}(s_0, H(x))| = \\
|\mathrm{E}\left(u^\varepsilon(s_0 - r, X_r^{s_0,x,\varepsilon}) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)| \le \\
|\mathrm{E}\left(u^\varepsilon(s_0 - r, X_r^{s_0,x,\varepsilon}) - \overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon}))\right)| + \\
|\mathrm{E}\left(\overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)| \le \\
\alpha \exp(\beta(s_0 - r)) + |\mathrm{E}\left(\overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)|.
\end{aligned}$$

Therefore, due to (19),

$$|\mathrm{E}\left(\overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)| \ge \alpha\beta r \exp(\beta s_0)/2.$$

As before, we have

$$\begin{aligned}
|\mathrm{E}\left(\overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)| \le \\
|\mathrm{E}\left(\overline{u}(s_0 - r, H(X_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, H(Z_r^{s_0,x,\varepsilon}))\right)| + \\
|\mathrm{E}\left(\overline{u}(s_0 - r, H(Z_r^{s_0,x,\varepsilon})) - \overline{u}(s_0 - r, Y_r^{s_0,H(x)})\right)|.
\end{aligned}$$

The second term on the right hand side is smaller than $\alpha\beta r \exp(\beta s_0)/4$ for all sufficiently small ε by Remark 1.3. The first term is estimated from above, using Lemma 2.1, by $KL\alpha \exp(\beta s_0)\sqrt{r}$. We obtain a contradiction with (20) if we take $\widetilde{K} = 4KL$, thus proving the theorem.

3 Lemma on closeness of the processes

In this section we prove Lemma 2.1. Let us consider the processes $(A_t^{i,\varepsilon}, T_t^{i,\varepsilon})$, $i = 1, 2$, which are defined by

$$dA_t^{i,\varepsilon} = T(H(A_t^{i,\varepsilon}))\left(\frac{1}{\varepsilon} v(A_t^{i,\varepsilon}) + b_i(T_t^{i,\varepsilon}, A_t^{i,\varepsilon})\right) dt + T^{\frac{1}{2}}(H(A_t^{i,\varepsilon}))\sigma_i(T_t^{i,\varepsilon}, A_t^{i,\varepsilon}) dW_t, \quad t \le \beta^{i,\varepsilon},$$

$$A_0^{i,\varepsilon} = x \in D,$$
$$\beta^{i,\varepsilon} = \min\{t : A_t^{i,\varepsilon} \in \partial D\}, \qquad A_t^{i,\varepsilon} = A_{\beta^{i,\varepsilon}}^{i,\varepsilon}, \quad \beta^{i,\varepsilon} \le t.$$
$$dT_t^{i,\varepsilon} = T(H(A_t^{i,\varepsilon}))dt, \quad T_0^{i,\varepsilon} = 0,$$

where the function T was defined in (3). Thus $A_t^{i,\varepsilon}$ can be obtained from $Z_t^{i,\varepsilon}$, $i = 1, 2$, via a random change of time. Let Γ be a smooth curve connecting γ_1 and γ_2 and transversal to the vector field v. For $h \in [h_1, h_2]$, let $p(h)$ be the point on Γ such that $H(p(h)) = h$. For $h \in [h_1, h_2]$, let $y_h(\varphi)$ solve

$$\frac{dy_h(\varphi)}{d\varphi} = T(H(y_h(\varphi)))v(y_h(\varphi)), \quad y_h(0) = p(h).$$

Thus $y_h(\varphi)$ represents deterministic periodic (with period one) motion along the level sets of H. Let us make the change of variables (action-angle coordinates after time change) $\Psi : \overline{D} \to \overline{U} := [h_1, h_2] \times S$, where S is a circle of length one. Namely,

$$\Psi(x) = (h(x), \varphi(x)) := (H(x), y^{-1}_{H(x)}(x)).$$

We denote the process $A_t^{i,\varepsilon}$ in the new variables by $B_t^{i,\varepsilon} = (H_t^{i,\varepsilon}, \varphi_t^{i,\varepsilon})$. Then

$$dH_t^{i,\varepsilon} = \widetilde{b}_i(T_t^{i,\varepsilon}, B_t^{i,\varepsilon})dt + \widetilde{\sigma}_i(T_t^{i,\varepsilon}, B_t^{i,\varepsilon})dW_t, \quad t \le \beta^{i,\varepsilon}, \qquad H_0^{i,\varepsilon} = H(x),$$
$$d\varphi_t^{i,\varepsilon} = \frac{1}{\varepsilon}dt + \widehat{b}_i(T_t^{i,\varepsilon}, B_t^{i,\varepsilon})dt + \widehat{\sigma}_i(T_t^{i,\varepsilon}, B_t^{i,\varepsilon})dW_t, \quad t \le \beta^{i,\varepsilon}, \qquad \varphi_0^{i,\varepsilon} = \varphi(x),$$
$$\beta^{i,\varepsilon} = \min\{t : H_t^{i,\varepsilon} \in \{h_1, h_2\}\}, \qquad H_t^{i,\varepsilon} = H_{\beta^{i,\varepsilon}}^{i,\varepsilon}, \quad \varphi_t^{i,\varepsilon} = \varphi_{\beta^{i,\varepsilon}}^{i,\varepsilon}, \quad \beta^{i,\varepsilon} \le t.$$
$$dT_t^{i,\varepsilon} = T(H_t^{i,\varepsilon})dt, \quad T_0^{i,\varepsilon} = 0,$$

where the scalar functions $\widetilde{b}_i$, $\widehat{b}_i$, $i = 1, 2$, and the vector valued functions $\widetilde{\sigma}_i$, $\widehat{\sigma}_i$, $i = 1, 2$, have the following properties: $\widetilde{b}_1$, $\widehat{b}_1$, $\widetilde{\sigma}_1$ and $\widehat{\sigma}_1$ are bounded by M' and are Lipschitz continuous with the same constant M'; $\widetilde{b}_2$, $\widehat{b}_2$, $\widetilde{\sigma}_2$ and $\widehat{\sigma}_2$ are Lipschitz continuous, $|\widetilde{\sigma}_1|$ and $|\widetilde{\sigma}_2|$ are bounded from below by a positive constant, and there is a constant $\overline{m}$ such that

$$\sup_{(t,y)\in[0,\infty)\times\overline{U}} \Big(|\widetilde{b}_1(t, y) - \widetilde{b}_2(t, y)| + |\widehat{b}_1(t, y) - \widehat{b}_2(t, y)|\Big) \le \overline{m}\eta, \tag{23}$$
$$\sup_{(t,y)\in[0,\infty)\times\overline{U}} (|\widetilde{\sigma}_1(t, y) - \widetilde{\sigma}_2(t, y)| + |\widehat{\sigma}_1(t, y) - \widehat{\sigma}_2(t, y)|) \le \overline{m}\eta. \tag{24}$$

These properties immediately follow from the corresponding properties for the vectors b_1 and b_2 and the matrices σ_1 and σ_2 if one takes into account that the change of time and the change of variables were smooth.

In what follows, C will stand for a generic constant that does not depend on η, but may vary from line to line. Let $\tau \le \beta^{1,\varepsilon} \wedge \beta^{2,\varepsilon}$ be a stopping time for the process W_t. Using the Ito formula, (23) and (24), we obtain

$$\mathrm{E}|H_{t\wedge\tau}^{1,\varepsilon} - H_{t\wedge\tau}^{2,\varepsilon}|^2 = 2\mathrm{E}\int_0^{t\wedge\tau} (H_{s\wedge\tau}^{1,\varepsilon} - H_{s\wedge\tau}^{2,\varepsilon})(\widetilde{b}_1(T_s^{1,\varepsilon}, B_s^{1,\varepsilon}) - \widetilde{b}_2(T_s^{2,\varepsilon}, B_s^{2,\varepsilon}))ds +$$

$$\mathrm{E}\int_0^{t\wedge\tau}|\widetilde{\sigma}_1(T_s^{1,\varepsilon},B_s^{1,\varepsilon})-\widetilde{\sigma}_2(T_s^{2,\varepsilon},B_s^{2,\varepsilon})|^2ds\le$$

$$\mathrm{E}\int_0^{t\wedge\tau}|H_{s\wedge\tau}^{1,\varepsilon}-H_{s\wedge\tau}^{2,\varepsilon}|^2ds+\mathrm{E}\int_0^{t\wedge\tau}|\widetilde{b}_1(T_s^{1,\varepsilon},B_s^{1,\varepsilon})-\widetilde{b}_2(T_s^{2,\varepsilon},B_s^{2,\varepsilon})|^2ds+$$

$$\mathrm{E}\int_0^{t\wedge\tau}|\widetilde{\sigma}_1(T_s^{1,\varepsilon},B_s^{1,\varepsilon})-\widetilde{\sigma}_2(T_s^{2,\varepsilon},B_s^{2,\varepsilon})|^2ds\le$$

$$\mathrm{E}\int_0^{t\wedge\tau}|H_{s\wedge\tau}^{1,\varepsilon}-H_{s\wedge\tau}^{2,\varepsilon}|^2ds+$$

$$\mathrm{E}\int_0^{t\wedge\tau}|(\widetilde{b}_1(T_{s\wedge\tau}^{1,\varepsilon},B_{s\wedge\tau}^{1,\varepsilon})-\widetilde{b}_1(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon}))+(\widetilde{b}_1(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon})-\widetilde{b}_2(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon}))|^2ds+$$

$$\mathrm{E}\int_0^{t\wedge\tau}|(\widetilde{\sigma}_1(T_{s\wedge\tau}^{1,\varepsilon},B_{s\wedge\tau}^{1,\varepsilon})-\widetilde{\sigma}_1(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon}))+(\widetilde{\sigma}_1(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon})-\widetilde{\sigma}_2(T_{s\wedge\tau}^{2,\varepsilon},B_{s\wedge\tau}^{2,\varepsilon}))|^2ds\le$$

$$C\left(t\eta^2+\int_0^t\mathrm{E}(|T_{s\wedge\tau}^{1,\varepsilon}-T_{s\wedge\tau}^{2,\varepsilon}|^2+|B_{s\wedge\tau}^{1,\varepsilon}-B_{s\wedge\tau}^{2,\varepsilon}|^2)ds\right).$$

In the same way we obtain

$$\mathrm{E}|\varphi_{t\wedge\tau}^{1,\varepsilon}-\varphi_{t\wedge\tau}^{2,\varepsilon}|^2\le C\left(t\eta^2+\int_0^t\mathrm{E}(|T_{s\wedge\tau}^{1,\varepsilon}-T_{s\wedge\tau}^{2,\varepsilon}|^2+|B_{s\wedge\tau}^{1,\varepsilon}-B_{s\wedge\tau}^{2,\varepsilon}|^2)ds\right)$$

and

$$\mathrm{E}|T_{t\wedge\tau}^{1,\varepsilon}-T_{t\wedge\tau}^{2,\varepsilon}|^2\le C\int_0^t\mathrm{E}|B_{s\wedge\tau}^{1,\varepsilon}-B_{s\wedge\tau}^{2,\varepsilon}|^2ds.$$

Combining these three inequalities we see that

$$\mathrm{E}\left(|T_{t\wedge\tau}^{1,\varepsilon}-T_{t\wedge\tau}^{2,\varepsilon}|^2+|B_{t\wedge\tau}^{1,\varepsilon}-B_{t\wedge\tau}^{2,\varepsilon}|^2\right)\le$$

$$C\left(t\eta^2+\int_0^t\mathrm{E}(|T_{s\wedge\tau}^{1,\varepsilon}-T_{s\wedge\tau}^{2,\varepsilon}|^2+|B_{s\wedge\tau}^{1,\varepsilon}-B_{s\wedge\tau}^{2,\varepsilon}|^2)ds\right).$$

The Gronwall inequality then implies that

$$\mathrm{E}\left(|T_{t\wedge\tau}^{1,\varepsilon}-T_{t\wedge\tau}^{2,\varepsilon}|^2+|B_{t\wedge\tau}^{1,\varepsilon}-B_{t\wedge\tau}^{2,\varepsilon}|^2\right)\le\eta^2(\exp(Ct)-1).$$

In particular, if $t\le S$ and τ is bounded by t, then

$$\mathrm{E}\left(|T_\tau^{1,\varepsilon}-T_\tau^{2,\varepsilon}|^2+|B_\tau^{1,\varepsilon}-B_\tau^{2,\varepsilon}|^2\right)\le C\eta^2t, \tag{25}$$

where C now depends on S.

Let $\alpha^{i,\varepsilon}(t)$, $i=1,2$, be the stopping times defined by

$$\int_0^{\alpha^{i,\varepsilon}(t)}T(H_s^{i,\varepsilon})ds=t. \tag{26}$$

Note that $f(H(Z_t^{i,\varepsilon}))$ has the same distribution as $f(H^{i,\varepsilon}_{\alpha^{i,\varepsilon}(t)\wedge\beta^{i,\varepsilon}})$. We need therefore to estimate the difference $\mathrm{E}f(H^{1,\varepsilon}_{\alpha^{1,\varepsilon}(t)\wedge\beta^{1,\varepsilon}}) - \mathrm{E}f(H^{2,\varepsilon}_{\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}})$. Since we are interested in $t \le S$ and T is bounded from below, the stopping times $\alpha^{i,\varepsilon}(t)$, $i = 1, 2$, are bounded by $t/\inf_{h\in[h_1,h_2]} T(h) \le S/\inf_{h\in[h_1,h_2]} T(h) =: t_0$.

Let $\alpha^\varepsilon(t) = \alpha^{1,\varepsilon}(t)\wedge\alpha^{2,\varepsilon}(t)$ and $\mu^\varepsilon(t) = \alpha^{1,\varepsilon}(t)\wedge\beta^{1,\varepsilon}\wedge\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}$. Since T is a Lipschitz continuous function and $\mu^\varepsilon(t)$ is bounded by a constant times t, by (25) we have for all s:

$$\mathrm{E}|T(H^{1,\varepsilon}_{s\wedge\mu^\varepsilon}) - T(H^{2,\varepsilon}_{s\wedge\mu^\varepsilon})| \le C\eta\sqrt{t}.$$

Therefore,

$$\mathrm{E}|\int_0^{\mu^\varepsilon} T(H^{1,\varepsilon}_s)ds - \int_0^{\mu^\varepsilon} T(H^{2,\varepsilon}_s)ds| \le \int_0^{t_0} \mathrm{E}|T(H^{1,\varepsilon}_{s\wedge\mu^\varepsilon}) - T(H^{2,\varepsilon}_{s\wedge\mu^\varepsilon})| \le C\eta\sqrt{t}.$$

Since T is bounded from below, this together with (26) implies that

$$\mathrm{E}(\chi_{\{\mu^\varepsilon=\alpha^\varepsilon(t)\}}|\alpha^{1,\varepsilon}(t) - \alpha^{2,\varepsilon}(t)|) \le C\eta\sqrt{t}. \tag{27}$$

If $\tau \le \beta^{i,\varepsilon}$ is a stopping time, then by the Ito formula

$$|\mathrm{E}f(H^{i,\varepsilon}_\tau) - f(H^{i,\varepsilon}_0)| \le \mathrm{E}\int_0^\tau |f'(H^{i,\varepsilon}_s)||\widetilde{b}_i|(T^{i,\varepsilon}_s, B^{i,\varepsilon}_s)ds + \frac{1}{2}\mathrm{E}\int_0^\tau |f''(H^{i,\varepsilon}_s)||\widetilde{\sigma}_i|^2(T^{i,\varepsilon}_s, B^{i,\varepsilon}_s)ds \le C\mathrm{E}\tau, \tag{28}$$

which holds for any initial point of the process $(T^{i,\varepsilon}_t, B^{i,\varepsilon}_t)$.

Now,

$$\begin{aligned}|\mathrm{E}f(H^{1,\varepsilon}_{\alpha^{1,\varepsilon}(t)\wedge\beta^{1,\varepsilon}}) - \mathrm{E}f(H^{2,\varepsilon}_{\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}})| \le \\ |\mathrm{E}f(H^{1,\varepsilon}_{\mu^\varepsilon}) - \mathrm{E}f(H^{2,\varepsilon}_{\mu^\varepsilon})| + \mathrm{E}|\chi_{\{\mu^\varepsilon=\alpha^{1,\varepsilon}(t)\}}\left(f(H^{2,\varepsilon}_{\mu^\varepsilon}) - f(H^{2,\varepsilon}_{\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}})\right)| + \\ \mathrm{E}|\chi_{\{\mu^\varepsilon=\alpha^{2,\varepsilon}(t)\}}\left(f(H^{1,\varepsilon}_{\alpha^{1,\varepsilon}(t)\wedge\beta^{1,\varepsilon}}) - f(H^{1,\varepsilon}_{\mu^\varepsilon})\right)| + \\ \mathrm{E}|\chi_{\{\mu^\varepsilon=\beta^{1,\varepsilon}\}}\left(f(H^{2,\varepsilon}_{\mu^\varepsilon}) - f(H^{2,\varepsilon}_{\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}})\right)| + \mathrm{E}|\chi_{\{\mu^\varepsilon=\beta^{2,\varepsilon}\}}\left(f(H^{1,\varepsilon}_{\alpha^{1,\varepsilon}(t)\wedge\beta^{1,\varepsilon}}) - f(H^{1,\varepsilon}_{\mu^\varepsilon})\right)|.\end{aligned} \tag{29}$$

The first term on the right hand side is estimated by $C\eta\sqrt{t}$ due to (25) since the derivative of f is bounded. The second and third terms are estimated by $C\eta\sqrt{t}$ using the Markov property, (27) and (28).

The last two terms are estimated using the proximity of $H^{1,\varepsilon}_{\mu^\varepsilon}$ and $H^{2,\varepsilon}_{\mu^\varepsilon}$. Indeed, as shown above, $\mathrm{E}|H^{1,\varepsilon}_{\mu^\varepsilon} - H^{2,\varepsilon}_{\mu^\varepsilon}| \le C\eta\sqrt{t}$. Note that $H^{1,\varepsilon}_{\mu^\varepsilon}$ takes values h_1 or h_2 when $\mu^\varepsilon = \beta^{1,\varepsilon}$, and therefore

$$\mathrm{E}\left(\chi_{\{\mu^\varepsilon=\beta^{1,\varepsilon}\}}\min(H^{2,\varepsilon}_{\mu^\varepsilon} - h_1, h_2 - H^{2,\varepsilon}_{\mu^\varepsilon})\right) \le C\eta\sqrt{t}.$$

Also note that there is a constant $C > 0$ with the property that if then initial data for the process $(H^{2,\varepsilon}_t, \varphi^{2,\varepsilon}_t, T^{2,\varepsilon}_t)$ is such that $\min(H^{2,\varepsilon}_0 - h_1, h_2 - H^{2,\varepsilon}_0) \le \rho$, then $\mathrm{E}|H^{2,\varepsilon}_{\widetilde{\mu}} - H^{2,\varepsilon}_0| \le C\rho$

for each ρ and stopping time $\widetilde{\mu}$. This easily follows from the boundedness from below of the diffusion coefficient $\widetilde{\sigma}_i$ for the process $H_t^{2,\varepsilon}$ and the fact that $H_t^{2,\varepsilon}$ is stopped when it reaches the boundary. By applying the strong Markov property with respect to the stopping time μ^ε and using the smoothness of f, we obtain that

$$\mathrm{E}|\chi_{\{\mu^\varepsilon=\beta^{1,\varepsilon}\}}\left(f(H^{2,\varepsilon}_{\mu^\varepsilon}) - f(H^{2,\varepsilon}_{\alpha^{2,\varepsilon}(t)\wedge\beta^{2,\varepsilon}})\right)| \leq C\eta\sqrt{t}.$$

The last term on the right hand side of (29) is estimated similarly. This completes the proof of the lemma.

4 Remarks and generalizations

1. Consider nonlinear perturbations of systems with more than one degree of freedom. Let the unperturbed system be completely integrable and in action-angle coordinates have the form

$$\dot{I}_t = 0, \dot{\Phi}_t = \Omega(I_t), \quad I \in \mathbb{R}^m, \Phi \in \mathbb{T}^m, \tag{30}$$

where $\mathbb{T}^m$ is the m-torus, and $\Omega(I) = (\omega_1(I), ..., \omega_m(I))$ is assumed to be sufficiently smooth and bounded.

Consider the following perturbation of the original system (in the fast time):

$$dX_s^{t,x,\varepsilon} = \frac{1}{\varepsilon}\widehat{\Omega}(I_s^{t,x,\varepsilon})ds + b(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))ds + \sigma(X_s^{t,x,\varepsilon})dW_s, \quad s \leq \tau^\varepsilon \wedge t,$$

$$\tau^\varepsilon = \min\{s : X_s^{t,x,\varepsilon} \in \partial D\}, \qquad X_s^{t,x,\varepsilon} = X_{\tau^\varepsilon}^{t,x,\varepsilon}, \quad \tau^\varepsilon \leq s \leq t, \qquad X_0^{t,x,\varepsilon} = x,$$

$$u^\varepsilon(t,x) = \mathrm{E}g(I_t^{t,x,\varepsilon}).$$

Here $X_s^{t,x,\varepsilon} = (I_s^{t,x,\varepsilon}, \Phi_s^{t,x,\varepsilon})$, W_s is a Wiener process in $\mathbb{R}^{2m}$, $\widehat{\Omega}(I) = (0, \Omega(I)) \in \mathbb{R}^{2m}$, and the entries of $\sigma(x)$ and $b(x,u)$ are bounded and Lipschitz continuous. We assume as before that the first integrals are constant on connected components of the boundary ∂D of the domain $D \subset \mathbb{R}^{2m}$ and the boundary-initial function g depends only on I. We restrict ourselves to the case when the diffusion coefficients do not depend on u.

Together with the process $X_s^{t,x,\varepsilon}$, consider the process $\widetilde{X}_s^{x,\varepsilon} = (\widetilde{I}_s^{x,\varepsilon}, \widetilde{\Phi}_s^{x,\varepsilon})$ defined by the equation

$$d\widetilde{X}_s^{x,\varepsilon} = \frac{1}{\varepsilon}\widehat{\Omega}(\widetilde{I}_s^{x,\varepsilon})ds + \sigma(\widetilde{X}_s^{x,\varepsilon})dW_s, \quad s \leq \tau^\varepsilon \wedge t,$$

$$\tau^\varepsilon = \min\{s : \widetilde{X}_s^{x,\varepsilon} \in \partial D\}, \qquad \widetilde{X}_s^{x,\varepsilon} = \widetilde{X}_{\tau^\varepsilon}^{x,\varepsilon}, \quad \tau^\varepsilon \leq s \leq t, \qquad \widetilde{X}_0^{x,\varepsilon} = x.$$

We assume that the matrix $a(x) = \sigma(x)\sigma^*(x)$ is non-degenerate. Then, using the Girsanov formula, we can write down the following equation for $u^\varepsilon(t,x)$:

$$u^\varepsilon(t,x) = \mathrm{E}g(\widetilde{I}_t^{t,x,\varepsilon})\exp(\int_0^t \sigma^{-1}(\widetilde{X}_s^{x,\varepsilon})b(\widetilde{X}_s^{x,\varepsilon}, u^\varepsilon(t-s, \widetilde{X}_s^{x,\varepsilon}))dW_s - \frac{1}{2}\int_0^t |\sigma^{-1}(\widetilde{X}_s^{x,\varepsilon})b(\widetilde{X}_s^{x,\varepsilon}, u^\varepsilon(t-s, \widetilde{X}_s^{x,\varepsilon}))|^2 ds). \tag{31}$$

In the case of many degrees of freedom we should add one more assumption: because of resonances, the invariant measure of the fast motion, in general, is not unique for some values of I. Assume that the set

$$\{I \in \mathbb{R}^m : \ \text{frequencies} \ \ \omega_1(I), ..., \omega_m(I) \ \ \text{are} \ \ \text{rationally} \ \ \text{dependent} \ \}$$

has zero Lebesgue measure.

Under this assumption, one can apply the averaging principle to the process $(\widetilde{I}_s^{x,\varepsilon}, \widetilde{\Phi}_s^{x,\varepsilon})$ and describe the limiting slow motion (see [4]). Using this result and equation (31), one can prove that $\lim_{\varepsilon \downarrow 0} u^\varepsilon(t, I, \Phi) = \overline{u}(t, I)$ exists and solves the problem

$$\frac{\partial \overline{u}}{\partial t} = \frac{1}{2} \sum_{i,j=1}^{m} \overline{a}_{ij}(I) \frac{\partial^2 \overline{u}}{\partial I_i \partial I_j} + \sum_{i=1}^{m} \overline{b}_i(I, \overline{u}) \frac{\partial \overline{u}}{\partial I_i},$$

$$\overline{u}(0, I) = g(I), \quad \overline{u}(t, I)|_{\partial D} = g(I)|_{\partial D},$$

where

$$\overline{a}_{ij}(I) = \frac{1}{(2\pi)^m} \int_{\mathbb{T}^m} a_{ij}(I, \Phi) d\Phi, \quad \overline{b}_i(I, u) = \frac{1}{(2\pi)^m} \int_{\mathbb{T}^m} b_i(I, \Phi, u) d\Phi, \quad i, j \in \{1, ..., m\}.$$

2. We have assumed that the initial and the boundary functions were independent of the fast variable Φ. It is easy to show that this assumption, in general, can not be omitted: if, for instance, in the case of one degree of freedom $\omega(I) = \omega_0 = \text{const}$, and the initial function depends on the angle coordinate, then $\lim_{\varepsilon \downarrow 0} u^\varepsilon(t, I, \varphi)$ may not exist. On the other hand, one can prove, at least in the case of one degree of freedom, that if $d\omega(I)/dI$ has just a finite number of zeros, then $\lim_{\varepsilon \downarrow 0} u^\varepsilon(t, I, \varphi) = \overline{u}(t, I)$ exists and is the solution of the same equation as in the case of initial-boundary functions independent of φ, but the initial and boundary conditions should be replaced by the functions appropriately averaged in φ.

3. Let us again consider the system with one degree of freedom (10)-(12), but now allow H to have a finite number of non-degenerate critical points inside D. If $\sigma = \sigma(x)$ and $b = b(x)$ are independent of u, the limiting behavior of the slow component of the process can be described in terms of a diffusion process on a graph Γ homeomorphic to the set of connected components of the level sets of H inside D (see [3], Chapter 8).

Let $\mathcal{H} : D \to \Gamma$ be the identification mapping. The corresponding result for the parabolic equation is that $u^\varepsilon(t, x)$ converges to $\overline{u}(t, \mathcal{H}(x))$ as $\varepsilon \downarrow 0$, where the limiting function $\overline{u}$ is defined on $[0, \infty) \times \Gamma$, satisfies the differential equations on the edges of the graph (as in (6), but with coefficients independent of $\overline{u}$) and certain gluing conditions at the interior vertices of Γ ([3], Chapter 8). Now we claim that a similar result holds even if b is allowed to depend on u. Namely, using the Girsanov formula, one can write an equation for $u^\varepsilon(t, x)$ in the form (31). Using this equation and the bounds obtained in Sections 2 and 3, one can check that the gluing conditions at the vertices will be the same as in the case $b(x, u) \equiv 0$ (compare with [1]). The differential operators inside the edges of the graph are obtained in the same way as in Theorem 1.1.

Acknowledgements: While working on this article, M. Freidlin was supported by NSF grants DMS-0803287 and DMS-0854982 and L. Koralov was supported by NSF grants DMS-0706974 and DMS-0854982.

References

[1] Brin M, Freidlin M.I., *On stochastic behavior of perturbed Hamiltonian system*, Ergodic Th. and Dynamical Sys., 2000, 20, pp 55-76.

[2] Freidlin M.I., *Functional Integration and Partial Differential Equations*, Princeton University Press, 1985.

[3] Freidlin M. I., Wentzell A. D., *Random Perturbations of Dynamical Systems*, Springer 1998.

[4] Freidlin M. I., Wentzell A. D., *Averaging Principle for stochastic perturbations of multifrequency systems*, Stochastics and Dynamics, 3, pp 393-408, 2003.

[5] Krylov N.V., *Nonlinear Elliptic and Parabolic Equations of the Second Order (Mathematics and its Applications)*, Springer 1987.