STOCHASTIC AVERAGING NEAR HOMOCLINIC ORBITS VIA SINGULAR PERTURBATIONS

Richard B. Sowers
Department of Mathematics
University of Illinois at Urbana–Champaign
Urbana, IL 61801
r-sowers@math.uiuc.edu
http://www.math.uiuc.edu/~r-sowers

Abstract We outline a singular-perturbations approach to the graph-valued stochastic averaging results of Freidlin-Wentzell and Freidlin-Weber. We specifically consider the Freidlin-Weber problem (a Newtonian particle in a double-well potential). To show the Freidlin-Weber convergence result, we develop a perturbed test function via a boundary-layer PDE near the homoclinic orbit. Solvability of this PDE is equivalent to the glueing conditions of Freidlin-Wentzell. Details of our calculations will appear elsewhere.

Keywords: Markov processes, stochastic averaging, stratified space

1. Introduction

One of the principal raisons d'être of "applied mathematics" is in *model reduction*; i.e., the development of rigorous methods to replace, often in some limiting regime, a complicated system by a simpler, or lower-dimensional one. We here study a model-reduction problem in *Markov processes*, namely a problem in *stochastic averaging*.

The underpinning of classical averaging is a separation of time scales; there is a coordinate which slowly varies (often thought of as the energy) and a coordinate which quickly varies (typically thought of as an angle). As the ratio of the fast to slow speed increases, it becomes possible to in a sense fix the slowly varying coordinate and carry out a long-term average in the quickly varying component, in the process identifying effective coefficients for the dynamics of the slow motion. A simple example of such a system is a Hamiltonian system with small perturbations. The slowly varying coordinate is the value of the Hamiltonian and the quickly varying coordinate is the position (or angle) in the appropriate level set of the Hamiltonian. One can seek a reduced model for the slow variable by first rescaling time so that the variation of the slow variable is of order one. As the fast motion becomes faster, the behavior of the slow variable can (often) be described via a closed set of equations (without reference to the fast variable). Stratonovich [21] and Gikhman [9] developed some of the first averaging results. Completely rigorous arguments involving diffusion processes were given by Khas'minskii [11, 12] and several others [2, 3, 17, 18].

The nature and complexity of the reduced or averaged process depends on the complexity of the Hamiltonian H; roughly, the reduced process takes values in the space of orbits of the fast motion. When H is fairly simple, viz., something like a paraboloid with a single isolated elliptic critical point, the reduced process is simply a Markov process on a line—a classical result [12]. This machinery has only recently been extended to handle more complicated Hamiltonians; e.g., Hamiltonians with a finite collection of distinct minima and saddle points; in this case the reduced Markov process takes its values in a graph which encodes the topology of the level sets of H [6, 7, 8]; see also the work of Neishtadt [16] and Wolansky [23]). Each leg of the graph corresponds to a region in which the the dynamics are diffeomorphic to a fast deterministic rotation around a cylinder and a slow diffusion and drift in the axial direction of the cylinder. At the end of the legs, however, *glueing conditions* need to be imposed. Heuristically, the glueing conditions generate a family of independent tosses of a die. When the limiting process hits the vertex of the graph, a new leg is chosen via the toss of the die. The next excursion out of the vertex occurs on this chosen leg. When the process returns to the vertex, the process repeats with a new coin. See [5] for a more rigorous discussion of this. See also [20] for another example of a process defined on a *stratified space*, which is more general than a graph.

Our goal here is to motivate the glueing conditions of Freidlin, Wentzell, and Weber, via a singular perturbations argument. The basic idea stems from a coordinate change introduced by Khasminskii in [8]. We shall not give the entirety of the proof; see [19] for the details. The original work of Freidlin and Wentzell [8] relied upon a specific type of noise—divergence-free noise, which allowed the invariant measure of the original process to be explicitly identified (Lebesgue measure). Roughly, this allowed one to develop an analysis near the homoclinic orbit by comparing probabilities of events where the trajectory starts at different points near the homoclinic orbit to integration of such probabilities where initial distribution is distributed according to the known invariant measure. More general noise was considered by Freidlin and Weber in [6]. There, fairly complicated hypoelliptic estimates were used to effect a reduction to the calculations for divergence-free noise. Our arguments, however, are entirely local and nowhere require the knowledge of the globally-defined invariant measure. Our main insight is to identify the proper boundary coordinates (these are motivated by [10]). We then pose a problem involving a coupled collection of partial differential equations. It turns out that the solvability of these coupled PDE's exactly corresponds to the glueing conditions. These coordinate have appeared in related problems (see [20]), and we hope that our efforts here will allow insight into other related problems of interest to the engineering and mathematics communities.

2. Problem Formulation

Fix a double-well potential function U on $\mathbb{R}$. We assume that U is nondegenerate; i.e., U'' is nonzero at the points where $U'=0$ and that $\lim_{|x|\nearrow\infty} U(x)=\infty$. We also assume that the local maximum of U occurs at 0, and that $U(0)=0$. Define the energy function

$$H(x,p) \overset{\text{def}}{=} \frac{p^2}{2} + U(x)$$

for all $(x,p)\in\mathbb{R}^2$. Consider the stochastic differential equation

$$\begin{aligned} dx^\varepsilon_t &= p^\varepsilon dt \\ dp^\varepsilon_t &= -U'(x^\varepsilon_t)dt + \varepsilon^2\beta(x^\varepsilon_t, p^\varepsilon_t)dt + \varepsilon dW_t \end{aligned}$$

where W is a standard Brownian motion, and where β represents dissipation. Also fix an $(x_\circ, p_\circ)\in\mathbb{R}^2$ as our initial conditions; i.e., $(x^\varepsilon_0, p^\varepsilon_0) = (x_\circ, p_\circ)$. Our aim is to understand the fluctuations of the energy $H(x^\varepsilon_t, p^\varepsilon_t)$. This is the problem studied by Freidlin and Weber in [6] and [7].

A proper understanding of $H(x^\varepsilon_t, p^\varepsilon_t)$ hinges upon first looking on the proper time scale. Define $Z^\varepsilon_t \overset{\text{def}}{=} (x^\varepsilon_{t/\varepsilon^2}, p^\varepsilon_{t/\varepsilon^2})$; then in law

$$dZ^\varepsilon_t = \frac{1}{\varepsilon^2}\bar{\nabla}H(Z^\varepsilon_t)dt + b(Z^\varepsilon_t)dt + \sigma(Z^\varepsilon_t)dW_t,$$

where $\bar{\nabla}H(x,p) \overset{\text{def}}{=} (\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial x})(x,p)$, $B(x,p) \overset{\text{def}}{=} (0,\beta(x,p))$, and $\sigma(x,p) \overset{\text{def}}{=} (0,1)$. The initial condition is that $Z^\varepsilon = (x_\circ, p_\circ)$. Define the second order operator $\mathscr{L}$, symbol $\langle\cdot,\cdot\rangle_{\mathscr{L}}$, and pseudometric $\|\cdot\|_{\mathscr{L}}$ (on $T^*\mathbb{R}^2$) by

$$\begin{aligned}
(\mathscr{L}f)(x,p) &\overset{\text{def}}{=} \frac{1}{2}\frac{\partial^2 f}{\partial p^2}(x,p) + \beta(x,p)\frac{\partial f}{\partial p}(x,p) \\
\langle df, dg\rangle_{\mathscr{L}}(x,p) &\overset{\text{def}}{=} \frac{\partial f}{\partial p}(x,p)\frac{\partial g}{\partial p}(x,p) \qquad\qquad (x,p)\in\mathbb{R}^2 \\
\|df\|^2_{\mathscr{L}}(x,p) &\overset{\text{def}}{=} \left(\frac{\partial f}{\partial p}(x,p)\right)^2
\end{aligned}$$

for all f and g in $C^2(\mathbb{R}^2)$. Then the generator of Z^ε is

$$\mathscr{L}^\varepsilon \overset{\text{def}}{=} \frac{1}{\varepsilon^2}\bar{\nabla} H + \mathscr{L}$$

and

$$H(Z^\varepsilon_t) = H(Z^\varepsilon_0) + \int_0^t (\mathscr{L}H)(Z^\varepsilon_s)ds + M^\varepsilon_t$$

where M is a martingale with respect to the filtration of W and with quadratic variation

$$\langle M\rangle_t = \int_0^t \|dH\|^2_{\mathscr{L}}(Z^\varepsilon_s)ds$$

for all $t > 0$.

The dominant part of the dynamics of Z^ε corresponds to the flow

$$\begin{aligned}
\dot{\mathfrak{z}}_t(x) &= \bar{\nabla} H(\mathfrak{z}_t(x)) \qquad t\in\mathbb{R},\ x\in\mathbb{R}^2 \\
\mathfrak{z}_0(x) &= x.
\end{aligned} \tag{1}$$

The requirements on U imply that there is a homoclinic orbit at $H^{-1}(0)$ which divides the phase space of $\mathfrak{z}$ (viz. $\mathbb{R}^2$) into three regions

$$\begin{aligned}
\mathbf{S}_O &\overset{\text{def}}{=} \{(x,p)\in\mathbb{R}^2 :\ H(x,p) > 0\} \\
\mathbf{S}_L &\overset{\text{def}}{=} \{(x,p)\in\mathbb{R}_-\times\mathbb{R} :\ H(x,p) < 0\} \\
\mathbf{S}_R &\overset{\text{def}}{=} \{(x,p)\in\mathbb{R}_+\times\mathbb{R} :\ H(x,p) < 0\}.
\end{aligned}$$

Define for convenience the index set

$$\Lambda \overset{\text{def}}{=} \{O, L, R\}.$$

Inside each of the $\mathbf{S}_\ell$'s, the orbits of $\mathfrak{z}$ are periodic. Thus as long as Z^ε stays inside any given $\mathbf{S}_\ell$, we should be able to asymptotically find closed dynamics for $H(Z^\varepsilon)$. We should (c.f. the remarks in [8]) then be able

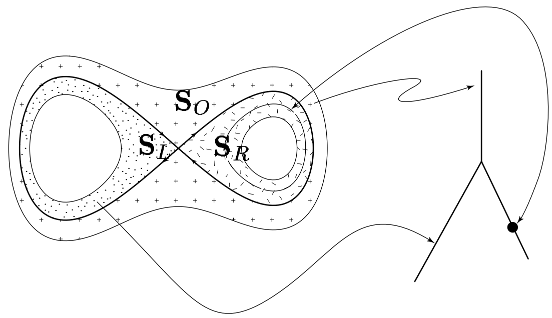

Figure 1. The "Y" graph.

to in a sense extend these averaging results to all of $\mathbb{R}^2$ if, in addition to keeping track of the slow variable $H(Z^\varepsilon)$, we also keep track of which of the $\mathbf{S}_\ell$'s Z^ε is in. Define $\mathcal{R}_O \stackrel{\text{def}}{=} \mathbb{R}_+$ and $\mathcal{R}_L = \mathcal{R}_R = \mathbb{R}_-$; i.e., the space for the limiting process should roughly be the disjoint union

$$E \stackrel{\text{def}}{=} \cup_{\ell \in \Lambda} (\mathcal{R}_\ell \times \{\ell\})$$

We say roughly since in fact, we want to identify all three points in $\{0\} \times \Lambda$ (they all correspond to the homoclinic orbit $H^{-1}(0)$. We will let $\mathfrak{M} \stackrel{\text{def}}{=} E/\sim$, where $\sim$ is the equivalence relation such that $(h, \ell) \in E$ is equivalent only to itself if $h \neq 0$, and such that $(0, O) \sim (0, L) \sim (0, R)$; we let $\star \stackrel{\text{def}}{=} \{0\} \times \Lambda$, the "vertex", be this equivalence class. We equip $\mathfrak{M}$ with the standard quotient topology (see [1, p. 60]). Topologically, $\mathfrak{M}$ is equivalent to a "wye"-shaped subset of $\mathbb{R}^2$; see Figure 1. We define the map $\mathscr{Y} : \mathbb{R}^2 \to \mathfrak{M}$ by setting

$$\mathscr{Y}(z) \stackrel{\text{def}}{=} (H(z), \ell)$$

if $z \in \mathbf{S}_\ell$, and we define $\mathscr{Y}(z) \stackrel{\text{def}}{=} \star$ if $z \in H^{-1}(0)$. The main result of Freidlin and Wentzell [8] was to show that $\mathscr{Y}(Z^\varepsilon)$ converges in law to an explicitly-defined Markov process on $\mathfrak{M}$ (see also [22]; a diffusion process on such a space is sometimes called a *spider*). Since the limiting process is not on a Euclidean space, we need a formalism more general than that of stochastic differential equations, namely the theory of generators and domains. Define the function spaces

$$C(\mathfrak{M}) \stackrel{\text{def}}{=} \{f \in C(E) : f(0, O) = f(0, L) = f(0, R)\}$$

$$C^2(\mathfrak{M} \setminus \star) \stackrel{\text{def}}{=} \left\{f \in C(\mathfrak{M}) : f(\cdot, \ell) \in C^2(\mathcal{R}_\ell^\circ) \text{ for all } \ell \in \Lambda\right\};$$

then $C(\mathfrak{M})$ is (isomorphic to) the collection of continuous functions on $\mathfrak{M}$ (where $\mathfrak{M}$ is endowed with its appropriate topology) and $C^2(\mathfrak{M})$ is the collection of continuous functions on $\mathfrak{M}$ which are twice differentiable on the open manifolds corresponding to the legs of the "wye". Note that while $\mathfrak{M}$ is a topological space, it does *not* have a differentiable structure at the vertex $\star$. Define the coefficients

$$\begin{aligned} m_2(h,\ell) &\overset{\text{def}}{=} \lim_{T\nearrow\infty} \frac{1}{T}\int_0^T \|dH\|^2_{\mathscr{L}}(\mathfrak{z}_s(x))ds \\ &= \frac{\int_{x'\in\mathbb{R}^2:\mathscr{Y}(x')=(h,\ell)} \frac{\|dH\|^2_{\mathscr{L}}(x')}{\|\bar\nabla H(x')\|}\mathscr{H}^1(dx')}{\int_{x'\in\mathbb{R}^2:\mathscr{Y}(x')=(h,\ell)} \|\bar\nabla H(x')\|^{-1}\mathscr{H}^1(dx')} \\ m_1(h,\ell) &\overset{\text{def}}{=} \lim_{T\nearrow\infty} \frac{1}{T}\int_0^T (\mathscr{L}H)(\mathfrak{z}_s(x))ds \\ &= \frac{\int_{x'\in\mathbb{R}^2:\mathscr{Y}(x')=(h,\ell)} \frac{(\mathscr{L}H)(x')}{\|\bar\nabla H(x')\|}\mathscr{H}^1(dx')}{\int_{x'\in\mathbb{R}^2:\mathscr{Y}(x')=(h,\ell)} \|\bar\nabla H(x')\|^{-1}\mathscr{H}^1(dx')} \end{aligned}$$

for all $(h,\ell)\in E\setminus\star$, where x is any element of $\mathbb{R}^2\setminus H^{-1}(0)$ such that $\mathscr{Y}(x)=(h,\ell)$ and where $\mathscr{H}^1$ is 1-dimensional Hausdorff measure.

If $f\in C^2(\mathfrak{M})$, we define

$$\mathscr{L}^\dagger f(h,\ell) \overset{\text{def}}{=} \frac{1}{2}m_2(h,\ell)\frac{\partial^2 f}{\partial h^2}(h,\ell) + m_1(h,\ell)\frac{\partial f}{\partial h}(h,\ell)$$

for all $(h,\ell)\in E\setminus\star$. It is fairly easy to show that (for each $\ell\in\Lambda$) if $\sup_{(h,\ell)\in E\setminus\star}|\mathscr{L}^\dagger f(h,\ell)|$ is finite, then $\frac{\partial f}{\partial h}(0,\ell)$ exists. We then define the glueing coefficients

$$\mathsf{G}_\ell \overset{\text{def}}{=} \int_{x'\in\partial\mathbf{S}_\ell} \|\bar\nabla H(x')\|^{-1}\mathscr{H}^1(dx')$$

for all $\ell\in\Lambda$ and we define the generator

$$\begin{aligned} \mathscr{A}^\dagger \overset{\text{def}}{=} \Big\{(f,g)\in C(\mathfrak{M})\times C(\mathfrak{M}) : f\in C^2(\mathfrak{M}\setminus\star),\ (\mathscr{L}^\dagger f)=g \text{ on } E\setminus\star \\ \text{and } \mathsf{G}_O\frac{\partial f}{\partial h}(0,O) = \mathsf{G}_L\frac{\partial f}{\partial h}(0,L) + \mathsf{G}_R\frac{\partial f}{\partial h}(0,R)\Big\}. \end{aligned}$$

To state the main result of Freidlin and Wentzell devoid of technicalities, let's first restrict the region of interest. Let $\hbar>0$ be small enough that

$$\mathbf{S}_\hbar \overset{\text{def}}{=} \{(x,p)\in\mathbb{R}^2 : |H(x,p)|\le\hbar\}$$

has no critical points of H other than the origin. Let Y be the Markov process which starts at $\mathscr{Y}(x_\circ,p_\circ)$ and has generator $\mathscr{A}^\dagger$ up to the time it leaves $\mathscr{Y}(\mathbf{S}_\hbar)$. In other words, let Y be a stochastic process on some probability triple such that for each $(f,g) \in \mathscr{A}^\dagger$ such that $g \equiv 0$ on $\mathscr{Y}(\mathbb{R}^2 \setminus \mathbf{S}_\hbar)$ (this restriction corresponds to killing the process when it leaves $\mathscr{Y}(\mathbf{S}_\hbar)$), there is a martingale M (with respect to the filtration generated by Y) such that

$$f(Y_t) = f(\mathscr{Y}(x_\circ, p_\circ)) + \int_{s=0}^t g(Y_s)ds + M_t$$

for all $t \geq 0$. Define the stopping time

$$\tau^\varepsilon \stackrel{\text{def}}{=} \inf\{t \geq 0 : Z_t^\varepsilon \notin \mathbf{S}_\hbar\}$$

Theorem 1 *The processes* $\{\mathscr{Y}(Z^\varepsilon_{t\wedge\tau^\varepsilon});\, t \geq 0\}$ *converge (as* $\varepsilon \searrow 0$*) in law to* Y.

The glueing conditions roughly mean that when the process hits $\star$, it flips a coin to decide where to make the next excursion. With probability $\mathrm{G}_\ell / \sum_{\ell' \in \Lambda} \mathrm{G}_{\ell'}$, this excursion will be made into leg ℓ. Since transitions from one leg to another correspond to transitions across $H^{-1}(0)$, it is of course natural that the glueing coefficients are given by some statistic of the original generator at the $\partial\mathbf{S}_\ell$'s.

The compactness of the laws of $\mathscr{Y}(Z^\varepsilon)$'s and the uniqueness of the law defined by the generator $\mathscr{A}^\dagger$ are fairly standard. We can carefully use stochastic averaging to locally show that within each leg, $H(Z^\varepsilon)$ evolves according to the genereator $\mathscr{L}$. We also need to show that Z^ε does not "stick" at the homoclinic orbit. This can be done too.

Glueing Conditions

The main interest, of course, is to understand the glueing conditions.

There are several perspectives. Let's look first at the answer. The glueing coefficients are essentially the numerators of the averaged diffusion coefficient m_2. This is not too surprising. The *diffusion* is the dominant reason why the process moves between the $\mathbf{S}_\ell$'s; a Brownian motion crosses zero many more times than a smooth curve. This brings up a technical issue which gets us started on our analysis. A core of the domain of $\mathscr{L}^\varepsilon$, the generator of Z^ε, is $C_0^2(\mathbb{R}^2)$. In other words, we initially understand the law of Z^ε by looking at test functions which are twice-differentiable near $H^{-1}(0)$. In the limit, however, the dynamics of Y are singular due to the structure of $\mathfrak{M}$. In some way, we need to connect the smoothness requirements of the prelimit process to the singular nature of the limiting process. Another framework in which to study this is continuity conditions and conservation of flux. At each point x of

$H^{-1}(0)$, the density of Z^ε (if a density exists) must be continuous and the flux going across $H^{-1}(0)$ at x must sum to zero. The continuity requirement means that if we approach x from within $H^{-1}(0)$, the density has the same limit as if we approach x from outside $H^{-1}(0)$. We need to compress these continuity and conservation laws from all of $H^{-1}(0)$ into a corresponding requirement at $\star$.

Let's focus on how to collapse the generator of Z^ε to $\mathscr{A}^\dagger$. Fix an $(f,g) \in \mathscr{A}^\dagger$. We can then define $\hat{f}: \mathbb{R}^2 \to \mathbb{R}$ by setting

$$\hat{f}(x) \stackrel{\text{def}}{=} \sum_{\ell \in \Lambda} f(H(x),\ell)\chi_{\mathbf{S}_\ell}(x) + f(\star)\chi_{\{0\}}(H(x))$$

We first note that $\hat{f}$ is continuous. However, unless all of the $\frac{\partial f}{\partial h}(0,\ell)$'s are equal (this is clearly allowed since $\mathrm{G}_O = \mathrm{G}_L + \mathrm{G}_R$, but such functions do not form a large enough set to ensure uniqueness of the limiting process), $\hat{f}$ will *not* be differentiable at $H^{-1}(0)$, and thus it will *not* be in the domain of $\mathscr{L}^\varepsilon$. Thus we need to *correct* $\hat{f}$ near $H^{-1}(0)$. This suggests a link with singular perturbations, or more accurately, the notion of *perturbed test functions* (see [4], [13], [14], and [15]). Our main claim is the following.

Proposition 2 *Since the glueing conditions hold, there is a sequence* $\{\Phi^\varepsilon : \varepsilon \in (0,1)\}$ *of bounded functions on* $\mathbb{R}^2$ *such that*

i) $\hat{f} + \Phi^\varepsilon \in C^1(\mathbb{R}^2)$

ii) Φ^ε *and its derivatives tend to zero exponentially as* $|H(x,p)/\varepsilon| \nearrow \infty$.

iii) $\mathscr{L}^\varepsilon \Phi^\varepsilon$ *is* $o(1)$ *as* $\varepsilon \searrow 0$.

iv) Φ^ε *is* $o(1)$ *as* $\varepsilon \searrow 0$.

Thus, the Φ^ε's are essentially harmonic functions (for $\mathscr{L}^\varepsilon$) which smooth $\hat{f}$. It is natural to seek harmonic functions since such harmonic functions naturally describe exit probabilities.

In general, an inner expansion is an approximate solution of a given PDE in a boundary layer near the edge of a boundary. In canonical problems, the role of the inner expansion is to describe the transition between some specified boundary data and the outer expansion (which usually fails to satisfy the specified boundary conditions). In our case, the boundary in question is $H^{-1}(0)$, and the boundary data is implicit; it is given by the "hidden" constraint that $\hat{f} + \Phi^\varepsilon$ is required to be C^1 at $H^{-1}(0)$. This means that at $H^{-1}(0)$, we want the values of $\bar{\Phi}^\varepsilon$ and their appropriate transversal derivatives (i.e., transversal to $H^{-1}(0)$) to match. We also note that effectively we have *three* boundary layers, one in each $\mathbf{S}_\ell$, near $\partial\mathbf{S}_\ell$.

Often the approximate solution of a PDE for an inner expansion is given by using an expanded coordinate in the direction *transversal* to the boundary, and a coordinate in the direction *along* the boundary.

Since $H^{-1}(0)$ is naturally defined as a level set of H, the coordinate in the direction transversal to $H^{-1}(0)$ will naturally be H. Consider for a moment a boundary layer of order ε^α for some $\alpha > 0$. If we apply $\mathscr{L}$ to a nonlinear function of H/ε^α, then we get a term of order $\varepsilon^{-2\alpha}$ (since $\mathscr{L}$ is a second-order operator). This is due to the diffusion in the direction transversal to $H^{-1}(0)$. We want this to be of the same order as the fast drift, which is of order ε^{-2}, so we should take $\alpha = 1$ and consider a boundary layer of order ε.

Let's next discuss the boundary coordinate. If Θ is a twice-differentiable real-valued function on some open subset $\mathscr{O}$ of $\mathbb{R}^2$ and if we set $\psi^\varepsilon(x) \stackrel{\text{def}}{=} \psi\left(\Theta(x), \frac{H(x)}{\varepsilon}\right)$ for $x \in \mathscr{O}$ where ψ is twice-differentiable, then

$$
\begin{aligned}
(\mathscr{L}^\varepsilon \psi^\varepsilon)(x) = \frac{1}{\varepsilon^2}\left\{\frac{\partial\psi}{\partial\theta}(\bar{\nabla}H, \nabla\Theta)(x) + \frac{1}{2}\frac{\partial^2\psi}{\partial^2 h}\|dH\|^2_{\mathscr{L}}(x)\right\} \\
+ \frac{1}{\varepsilon}\left\{\frac{\partial\psi}{\partial h}(\mathscr{L}H)(x) + \frac{\partial^2\psi}{\partial\theta\partial h}\left\langle dH, d\Theta\right\rangle_{\mathscr{L}}(x)\right\} \\
+ \left\{\frac{\partial\psi}{\partial\theta}(\mathscr{L}\Theta)(x) + \frac{1}{2}\frac{\partial^2\psi}{\partial\theta^2}\|d\Theta\|^2_{\mathscr{L}}(x)\right\}
\end{aligned} \quad (2)
$$

for all $x \in \mathscr{O}$, where the various derivatives of ψ are all evaluated at $\left(\Theta(x), \frac{H(x)}{\varepsilon}\right)$. The *Khasminskii* coordinates, which we will denote by Θ_K, equate the coefficients of the two dominant terms; i.e., they solve the transport equation

$$(\bar{\nabla}H, \nabla\Theta_K) = \|dH\|^2_{\mathscr{L}} \quad (3)$$

on some appropriate subset of $\mathbf{S}$.

Remark 3 *Heuristically, the Khasminskii coordinates normalize the diffusion across $H^{-1}(0)$ (this transversal diffusion is responsible for causing the graph-valued limit to choose one leg over the other when it starts at the vertex). Roughly, equal increments of Khasminskii coordinates correspond to equal amounts of transversal diffusion across $H^{-1}(0)$. The glueing conditions are directly related to the relative lengths of the $\partial\mathbf{S}_\ell$'s in the Khasminskii coordinates.*

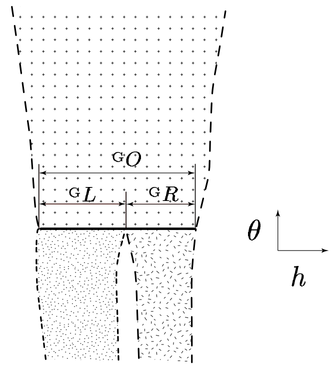

Figure 2. Khasminskii Coordinates (this picture was suggested to the author by N. Sri Namachchivaya, based on some work with Peter Imkeller.)

We are thus led to the coordinate transformation $x \mapsto (\Theta_K(x), H(x))$. One needs to select initial conditions for the PDE (3); these initial conditions will in general be specified along some manifold which is transveral to the level sets of H. In particular, one needs to specify initial data in each of the $\mathbf{S}_\ell$'s. Note that the range of Θ_K in each $\mathbf{S}_\ell$ near $\partial\mathbf{S}_\ell$ is an interval of width g_ℓ; see Figure 2. To make things convenient, we can impose initial data on (3) so that Θ_K starts at 0 just below $(0,0)$, and then increases to g_L as we travel around the left loop (in the direction of $\mathfrak{z}$) of $H^{-1}(0)$ and comes back to $(0,0)$, and then arrange things so that Θ_K further increases around the right loop of $H^{-1}(0)$ (again travelling in the direction of $\mathfrak{z}$) to finally reach $\mathsf{g}_L + \mathsf{g}_R = \mathsf{g}_O$ as we come back to $(0,0)$. Define

$$\mathsf{l}_O = \mathsf{l}_L \stackrel{\text{def}}{=} 0, \quad \mathsf{r}_L = \mathsf{l}_R = \mathsf{g}_L, \quad \text{and} \quad \mathsf{r}_R = \mathsf{r}_O = \mathsf{g}_O$$

and then define

$$\mathsf{I}_\ell \stackrel{\text{def}}{=} (\mathsf{l}_\ell, \mathsf{r}_\ell)$$

for all $\ell \in \Lambda$.

The following result, which is the heart of Proposition 2, now has some significance.

Proposition 4 *There is a triplet $(\Psi^K_O, \Psi^K_L, \Psi^K_R)$ of functions such that the following hold. For each $\ell \in \Lambda$, Ψ^K_ℓ is in $C^\infty(\overline{\mathsf{I}_\ell} \times \mathcal{R}_\ell)$ and*

$$\frac{\partial \Psi^K_\ell}{\partial \theta}(\theta, h) + \frac{1}{2}\frac{\partial^2 \Psi^K_\ell}{\partial^2 h}(\theta, h) = 0 \tag{4}$$

for all $(\theta, h) \in I_\ell \times \mathcal{R}_\ell^\circ$ *(this is a PDE). Secondly, for every multi-index* α*, there is a constant* $K > 0$ *such that*

$$\left| D^\alpha \Psi_\ell^K(\theta, h) \right| \le K e^{-|h|/K} \tag{5}$$

for all $\theta \in I_\ell$ *and* $h \in \mathcal{R}_\ell$ *such that* $|h| \ge 1$. *(these are growth conditions). Thirdly, we have that*

$$\Psi_\ell^K({\scriptstyle L}\ell, h) = \Psi_\ell^K({\scriptstyle R}\ell, h) \tag{6}$$

for all $h \in \mathcal{R}_\ell$ *(this is a periodicity condition). Fourthly and finally, we have that*

$$\begin{aligned} \lim_{h \searrow 0} \Psi_O^K(\cdot, h) &= \lim_{h \nearrow 0} \Psi_L^K(\cdot, h) \qquad \textit{on } I_L \\ \lim_{h \searrow 0} \Psi_O^K(\cdot, h) &= \lim_{h \nearrow 0} \Psi_R^K(\cdot, h) \qquad \textit{on } I_R \end{aligned} \tag{7}$$

and that

$$\begin{aligned} \lim_{h \searrow 0} \frac{\partial \Psi_O^K}{\partial h}(\cdot, h) &= \lim_{h \nearrow 0} \frac{\partial \Psi_L^K}{\partial h}(\cdot, h) \qquad \textit{on } I_L \\ \lim_{h \searrow 0} \frac{\partial \Psi_O^K}{\partial h}(\cdot, h) &= \lim_{h \nearrow 0} \frac{\partial \Psi_R^K}{\partial h}(\cdot, h) \qquad \textit{on } I_R \end{aligned} \tag{8}$$

(these are matching conditions at the boundaries).

See [19] for a proof; the sense of convergence in (7) and (8) is also specified there. We then set

$$\Phi^\varepsilon(x) \stackrel{\text{def}}{=} \varepsilon \Psi_\ell^K \left(\Theta_K(x), \frac{H(x)}{\varepsilon} \right) \tag{9}$$

if $x \in \mathbf{S}_\ell$. The PDE condition of (4) should imply that Φ^ε is approximately harmonic. The decay conditions of (5) mean that Φ^ε is a true boundary term. The periodicity conditions of (6) imply that if we approximate the range of Θ_K in $\mathbf{S}_\ell$ (and not just at $\partial \mathbf{S}_\ell$) by an interval of width ${\scriptstyle G}\ell$, then Φ^ε should be almost smooth (in particular where we have specified the initial data for Θ_K). The matching conditions of (7) and (8) are exactly what we need to make $\hat{f} + \Phi^\varepsilon$ continuous and differentiable at $H^{-1}(0)$.

We should note, however, that there are some wrinkles. In general, one has small discontinuities when the orbits of $\mathfrak{z}$ come back to the manifold along which initial data for Θ_K was specified (since the width of the range along the orbits of $\mathfrak{z}$ are exactly the ${\scriptstyle G}\ell$'s only on the homoclinic orbit $H^{-1}(0)$). Secondly, one must consider *singularities* at the origin. There is absolutely no assurance that the terminal data of the backward heat equations is such that (9) is smooth near the origin. These singularities need to be carefully treated.

References

[1] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press Inc., Orlando, Fla., second edition, 1986.

[2] A. N. Borodin. A limit theorem for the solutions of differential equations with a random right-hand side. *Teor. Verojatnost. i Primenen.*, 22(3):498–512, 1977.

[3] A. N. Borodin and M. I. Freidlin. Fast oscillating random perturbations of dynamical systems with conservation laws. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(3):485–525, 1995.

[4] L. C. Evans. The perturbed test function method for viscosity solutions of nonlinear PDE. *Proc. Roy. Soc. Edinburgh Sect. A*, 111(3-4):359–375, 1989.

[5] S. N. Evans and R. B. Sowers. Pinching and twisting Markov processes. *Annals of Probability.* to appear.

[6] M. Freidlin and M. Weber. Random perturbations of nonlinear oscillators. *Ann. Probab.*, 26(3):925–967, 1998.

[7] M. Freidlin and M. Weber. A remark on random perturbations of the nonlinear pendulum. *Ann. Appl. Probab.*, 9(3):611–628, 1999.

[8] M. I. Freidlin and A. D. Wentzell. Random perturbations of Hamiltonian systems. *Mem. Amer. Math. Soc.*, 109(523):viii+82, 1994.

[9] I. I. Gihman. Concerning a theorem of N. N. Bogolyubov. *Ukrain. Mat. Ž.*, 4:215–219, 1952.

[10] R. Z. Has'minskiĭ. Diffusion processes with a small parameter. *Izv. Akad. Nauk SSSR Ser. Mat.*, 27:1281–1300, 1963.

[11] R. Z. Has'minskiĭ. A limit theorem for solutions of differential equations with a random right hand part. *Teor. Verojatnost. i Primenen*, 11:444–462, 1966.

[12] R. Z. Has'minskiĭ. Stochastic processes defined by differential equations with a small parameter. *Teor. Verojatnost. i Primenen*, 11:240–259, 1966.

[13] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis*, 12:55–67, 1973.

[14] T. G. Kurtz. An abstract averaging theorem. *J. Functional Analysis*, 23(2):135–144, 1976.

[15] H. J. Kushner. *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory.* MIT Press, Cambridge, MA, 1984.

[16] A. I. Neĭshtadt. Probability phenomena due to separatrix crossing. *Chaos*, 1(1):42–48, 1991.

[17] G. C. Papanicolaou and W. Kohler. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.*, 27:641–668, 1974.

[18] G. C. Papanicolaou and W. Kohler. Asymptotic analysis of deterministic and stochastic equations with rapidly varying components. *Comm. Math. Phys.*, 45(3):217–232, 1975.

[19] R. B. Sowers. Stochastic averaging near a homoclinic orbit with multiplicative noise. submitted.

[20] R. B. Sowers. Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere). *Trans. Amer. Math. Soc.*, 354(3):853–900 (electronic), 2002.

[21] R. L. Stratonovich. *Topics in the Theory of Random Noise. Vol. I: General Theory of Random Processes. Nonlinear Transformations of Signals and Noise.* Gordon and Breach Science Publishers, New York, 1963. Revised English edition. Translated from the Russian by Richard A. Silverman.

[22] J. Walsh. A diffusion with discontinuous local time. In *Temps Locaux, Astérisque*, volume 52–53. Société Mathématique de France, Paris, 1978.

[23] G. Wolansky. Limit theorem for a dynamical system in the presence of resonances and homoclinic orbits. *J. Differential Equations*, 83(2):300–335, 1990.