

# INVARIANT MEASURE SELECTION BY NOISE AN EXAMPLE

ABSTRACT. We consider a deterministic system with two conserved quantities and infinity many invariant measures. However the systems possess a unique invariant measure when enough stochastic forcing and balancing dissipation are added. We then show that as the forcing and dissipation are removed a unique limit of the deterministic system is selected. The exact structure of the limiting measure depends on the specifics of the stochastic forcing.

JONATHAN C. MATTINGLY

Mathematics Department and Department of Statistical Science  
Duke University, Box 90320  
Durham, NC 27708-0320, USA

ETIENNE PARDOUX

Laboratoire d'Analyse, Topologie, Probabilités  
Université de Provence 39, rue F. Joliot-Curie  
F-13453 Marseille cedex 13, France

*Dedicated to the memory of José Real*

## 1. INTRODUCTION

There is much interest in the regularizing effects of noise on the longtime dynamics. One often speaks informally of adding a balancing noise and dissipation to a dynamical system with many invariant measures and then studying the zero noise/dissipation limit as a way of selecting the “physically relevant” invariant measure.

There are a number of settings where such a procedure is fairly well understood. In the case of a Hamiltonian or gradient system with sufficiently non-degenerate noise, Wentzell-Freidlin theory gives a rather complete description of the effective limiting dynamics [FW12] in terms of a limiting “slow” system derived through a quasi-potential and deterministic averaging. In the gradient case the stochastic invariant measures concentrate on the attracting structures of the dynamics. In the Hamiltonian setting, Wentzell-Freidlin theory considers the slow dynamics of the conserved quantity (the Hamiltonian) when the system is subject to noise. It is the zero noise limit of these dynamics which decides which mixture of the Hamiltonian invariant measures is selected in the zero noise limit.

In the case of system with an underlying hyperbolic structure, such as Axiom A, it is known that the zero noise limit of random perturbations selects a canonical SRB/”physical measure” [Sin68, Sin72, Rue82, Kif74]. This relies fundamentally on the expansion/contraction properties of the underlying deterministic dynamical system. See [You02] for a nice discussion of these issues. The Axiom A assumption

---

1991 *Mathematics Subject Classification*. Primary: 60H10, 37L40; Secondary: 37A60, 34C29.

*Key words and phrases*. inviscid limits, zero-noise limit, invariant measure selection, stochastic dynamically systems.

ensures that the deterministic dynamics has a rich attractor which attracts a set of positive Lebesgue measure.

One area where the idea of the relevant invariant measure being selected through a zero noise limit is prevalent is in the study of stochastically forced and damped PDEs. Two important examples are the stochastic Navier-Stokes equations and the stochastic KdV equation. Both of these equations have been studied in a sequence of works by Kuksin and his co-authors [Kuk04, Kuk07a, Kuk07b, KP08, Kuk10]. In all these works, tightness is established by balancing the noise and dissipation as the zero noise limit is taken. Any limiting invariant measure is shown to satisfy an appropriate limiting equation. Typically a number of properties are inherited from the pre-limiting invariant measure.

The hope is that the study of these limiting measures will give some insight into important questions for the original, unperturbed equations. In the case of the Navier-Stokes equations one would be interested in understanding questions such as the existence of energy cascades and turbulence. Setting aside the question of whether the regularity of the solutions in [Kuk04] is appropriate for turbulence, it is interesting to understand if the noise selects a unique limit and what are the obstructions to such uniqueness as they give information about the structure of the deterministic phase space. In all of the works [Kuk04, Kuk07a, Kuk07b, KP08, Kuk10] the question of uniqueness of the limit is not addressed and seems out of reach. (Though a recent work [Kuk13] makes progress in this direction. See the note at the end of the introduction.)

The equation for the evolution of a 2D incompressible fluid's vorticity  $q(x, t)$  (a scalar) on the 2-torus subject to stochastic excitation can be written as

$$\dot{q}(x, t) = \nu \Delta q(x, t) + B(q(x, t), q(x, t)) + \sqrt{\nu} \sum_{k \in \mathbf{Z}^2} \sigma_k e^{ik \cdot x} \dot{W}_t^{(k)}$$

where  $\nu > 0$  is the viscosity,  $\Delta$  is the Laplacian,  $\sigma_k$  are constants to be chosen,  $\{W_t^{(k)} : k \in \mathbf{Z}^2\}$  are a collection of standard one-dimensional Wiener processes and  $B(q, q)$  is a quadratic non-linearity such that  $\langle B(q, q), q \rangle_{L^2} = 0$ . The scaling of  $\nu$  is chosen to keep the spatial  $L^2$  norm of order one in the  $\nu \rightarrow 0$  limit and is the only scaling on a fixed torus which will result in a non-trivial sequence of tight processes. On a fixed interval, the formal  $\nu = 0$  limit of this equation is the Euler equation which conserves its Hamiltonian (the energy or  $L^2$  norm) but also has an infinite collection of other conserved quantities since the vorticity is simply transported about space. This means that *a priori* there will be many conserved quantities whose slow evolution must be analyzed.

Inspired by models in [Lor63, MTVE02] and the Euler equation itself, we construct a model problem in the form of an ODE in  $\mathbf{R}^3$  such that the non-linearity is quadratic and conserves the norm of the solution as in analogy with the Euler non-linearity. We will also see that our model system in fact possesses two conserved quantities (the most it could have without becoming trivial). In many ways our analysis follows the familiar pattern of [FW12] in that we change time to consider the evolution of the conserved quantities from the unforced system on a long time interval which grows as the noise is taken to zero. This produces a limiting system which captures the effect of the noise. However multiple conserved quantities are not usually treated in Wentzell-Freidlin theory and the complications of having more than one are non-trivial in our case. In particular, the limiting system does

not have a unique solution. Nonetheless, we are able to show that a particular solution is selected by the limiting procedure which in turn leads to a unique invariant measure for the limiting system being selected.

Since the limiting system does not have unique solutions there are many possible invariant measures depending on which of the solutions are chosen. It is interesting to note that identifying the limiting “averaged” solution is not sufficient to identify the likely limit of the invariant measure. Analysis of the limiting solution in isolation reveal domain walls which separate different regions of phase space and along which the diffusion degenerates giving rise to the possibility of solutions which could spend arbitrary amounts of time on the domain boundaries. Only through the analysis of pre-limiting systems do we discover that the systems selects the solutions which spend zero time on the domain walls.

These domain walls are the planes  $\{x = y\}$  in  $\mathbf{R}^3$  and correspond to heteroclinic cycles from the original deterministic system made up of homoclinic orbits connecting the fix points. Hence it is not surprising that the limiting system supports solutions which could spend arbitrarily long times on these orbits. However, it is interesting that the limiting procedure selects solutions which do not become trapped near the heteroclinic orbits.

After the completion and submission of this work, we became aware of a recent work by Kuksin which proves that the zero noise/damping limit of the stochastic Complex Ginzburg Landau (CGL) selects a unique invariant measure from the many possible measures which are invariant for the formal deterministic limit [Kuk13]. That problem is very much in the spirit of the one discussed here. Our example is finite dimensional. However, the associated limiting martingale problem for the fast variables is more complicated and does not have a unique solution. Like the CGL, our example has a simplified orbit structure which facilitates averaging. More complicated settings such as the stochastic Navier Stokes equations are still out of reach. We hope our paper helps clarifying some of the issues involved.

## 2. MODEL SYSTEM

As an exercise in studying the zero noise/dissipation limit of conservative systems, we have chosen to study the following three-dimensional system:

$$(1) \quad \dot{\xi}_t = \mathcal{B}(\xi_t, \xi_t)$$

with  $\xi_0 = (X_0, Y_0, Z_0) \in \mathbf{R}^3$  where if  $\xi = (x, y, z) \in \mathbf{R}^3$  and  $\hat{\xi} = (\hat{x}, \hat{y}, \hat{z})$  then  $\mathcal{B}$  is the symmetric bi-linear form defined by

$$(2) \quad \mathcal{B}(\xi, \hat{\xi}) \stackrel{\text{def}}{=} \frac{1}{2} \begin{pmatrix} y\hat{z} + \hat{y}z \\ x\hat{z} + \hat{x}z \\ -2x\hat{y} - 2\hat{x}y \end{pmatrix}.$$

We will write  $\varphi_t$  for the flow map induced by (1), i.e.  $\xi_t = \varphi_t(\xi_0)$ . We will constantly write  $(X_t, Y_t, Z_t)$  for  $\xi_t$  when we wish to speak of the components of  $\xi_t$ .

Since

$$(3) \quad B(\xi, \xi) \cdot \xi = 0$$

we see that  $|\xi_t|^2 = X_t^2 + Y_t^2 + Z_t^2$  is constant along trajectories of (1). Similarly one sees that  $X_t^2 - Y_t^2$  is also conserved by the dynamics of (1). Since any linear combination is also conserved, we are free to consider  $2X_t^2 + Z_t^2$  and  $2Y_t^2 + Z_t^2$ , which

are more symmetric. Since we will typically use the second pair, we introduce the map

$$(4) \quad \Phi: (x, y, z) \mapsto (u, v) = (2x^2 + z^2, 2y^2 + z^2).$$

A moments reflection shows that the existence of these two conserved quantities implies that all of the orbits of (1) are bounded and most are closed orbits, topologically equivalent to a circle. All orbits live on the surface of a sphere whose radius is dictated by the values of the conserved quantities. More precisely, given the initial condition  $\xi_0 \in \mathbf{R}^3$  the orbit  $\{\xi_t : t \geq 0\}$  is contained in the set

$$(5) \quad \Gamma = \left\{ \xi : \Phi(\xi) = \Phi(\xi_0) \right\}$$

To any initial point  $\xi_0 = (X_0, Y_0, Z_0)$  contained in a closed orbit, we can associate a measure defined by the following limit

$$(6) \quad \mu_{\xi_0}(dx, dy, dz) \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \delta_{\xi_s}(dx, dy, dz) ds.$$

We will show in Section 7.4 that this invariant measure depends only upon  $\Phi(\xi_0)$ , and a choice of a sign. Any such measure is an invariant measure for the dynamics given by (1). Hence we see that (1) has infinitely many invariant measures. It is reasonable to expect that the addition of sufficient driving noise and balancing dissipation, will result in a system with a unique invariant measure. Our goal is to study its limit as the noise/dissipation are scaled to zero. We are specifically interested in understanding whether this procedure selects a unique convex combination of the measures for the underlying deterministic system (1).

More concretely for  $\varepsilon > 0$ , we will explore the following stochastic differential system

$$(7) \quad \dot{\xi}_t^\varepsilon = \mathcal{B}(\xi_t^\varepsilon, \xi_t^\varepsilon) - \varepsilon \xi_t^\varepsilon + \sqrt{\varepsilon} \sigma \dot{W}_t,$$

with  $\xi_0^\varepsilon = (X_0, Y_0, Z_0) \in \mathbf{R}^3$ ,

$$\sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } W_t = (W_t^{(1)}, W_t^{(2)}),$$

where the two components  $W_t^{(1)}$  and  $W_t^{(2)}$  are mutually independent standard Brownian motions. In all this paper, we assume that  $\sigma_1 > 0$  and  $\sigma_2 > 0$ .

As above, we will write  $(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$  when we wish to discuss the coordinates of  $\xi_t^\varepsilon$ .

For each  $\varepsilon > 0$ , the three-dimensional hypoelliptic diffusion process is positive recurrent and ergodic, its unique invariant probability measure  $\mu_\varepsilon$  is absolutely continuous with respect to Lebesgue measure, with a density which charges all open sets.

Our aim is to study the limit of  $\mu_\varepsilon$ , as  $\varepsilon \rightarrow 0$ . We first note that as  $\varepsilon \rightarrow 0$ , the process  $(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$  converges to the solution of (1) on any finite time interval.

The main result of this article is that there exists a probability measure  $\mu$  which is absolutely continuous with respect to Lebesgue measure and so that  $\mu^\varepsilon$  converges weakly to  $\mu$  as  $\varepsilon \rightarrow 0$ . Of course,  $\mu$  is a mixture of the ergodic invariant measures appearing in (6), which we shall describe.

It is natural to ask what is the effect of adding noise also in the  $z$ -direction. Unfortunately this leads to unexpected complications which at present we are not able to handle.

## 3. MAIN RESULTS

For  $\varepsilon \geq 0$ , we define the Markov semigroup  $P_t^\varepsilon$  associated with (7) by

$$(P_t^\varepsilon \phi)(\xi) = \mathbf{E}_\xi \phi(\xi_t^\varepsilon)$$

for bounded  $\phi: \mathbf{R}^3 \rightarrow \mathbf{R}$ .

The dynamics obtained by formally setting  $\varepsilon = 0$  are deterministic. We will write  $P_t$  rather than  $P_t^0$  for the corresponding Markov semigroup which is defined by  $(P_t \phi)(\xi) = \phi(\xi_t)$ .

**Theorem 3.1.** *For each  $\varepsilon > 0$ ,  $P_t^\varepsilon$  has a unique invariant probability measure  $\mu^\varepsilon$  which has a  $C^\infty$  density which is everywhere positive.*

**Theorem 3.2.** *There exists a probability measure  $\mu$  such that  $\mu^\varepsilon \Rightarrow \mu$  as  $\varepsilon \rightarrow 0$  and furthermore such that  $\mu$  is invariant for the dynamics generated by (1) in that  $\mu P_t = \mu$  for all  $t > 0$ . In addition,  $\mu$  is absolutely continuous with respect to Lebesgue measure on  $\mathbf{R}^3$ , with a density which is positive on the complement of  $\{x = y\} \cup \{x = -y\}$ .*

The paper is organized as follows. Section 4 studies finite time convergence of  $\xi^\varepsilon$  as  $\varepsilon \rightarrow 0$ . Section 5 studies existence and uniqueness of the invariant measure for  $\xi^\varepsilon$ . Namely, this section proves Theorem 3.1. Section 6 studies the deterministic system on a faster time scale, more precisely it introduces the process  $(U_t^\varepsilon, V_t^\varepsilon) = \Phi(X_{t/\varepsilon}^\varepsilon, Y_{t/\varepsilon}^\varepsilon, Z_{t/\varepsilon}^\varepsilon)$ . Assuming some results from Section 8, we uniquely characterize the limit  $(U_t, V_t)$  of the process  $(U_t^\varepsilon, V_t^\varepsilon)$  as  $\varepsilon \rightarrow 0$ , and show that that limit has a unique invariant probability measure. Section 7 studies very precisely the deterministic dynamics behind the ODE (1) obtained by formally setting  $\varepsilon = 0$  in (7). Section 8 establishes crucial results which were assumed to hold in the discussion in Section 6, the main important and most delicate one being the convergence of the quadratic variation of  $(U^\varepsilon, V^\varepsilon)$ , which builds upon the analysis in Section 7. Finally Section 9 is devoted to the proof of Theorem 3.2.

## 4. FINITE TIME CONVERGENCE ON ORIGINAL TIMESCALE

In this section we show that the dynamics of stochastic dynamics given by (7) converge to the deterministic dynamics given by (1). Hence the limit as  $\varepsilon \rightarrow 0$  on this time scale does not help in understanding the selection of any limiting invariant measure as  $\varepsilon \rightarrow 0$ .

**Lemma 4.1.** *There exists a positive constant  $c$  so that if  $\xi_0 = \xi_0^\varepsilon \in \mathbf{R}^3$  then for all  $\varepsilon > 0$*

$$\mathbf{E}|\xi_t^\varepsilon - \xi_t|^2 \leq \varepsilon \frac{|\sigma|^2 + \varepsilon|\xi_0|}{c|\xi_0|} e^{c|\xi_0|t}.$$

**Corollary 1.** *For any  $t \geq 0$ ,  $P_t^\varepsilon$  converges weakly to  $P_t$  as  $\varepsilon \rightarrow 0$ . In other words, for any bounded and continuous  $\phi: \mathbf{R}^3 \rightarrow \mathbf{R}$ ,  $P_t^\varepsilon \phi(\xi) \rightarrow P_t \phi(\xi)$  for all  $\xi \in \mathbf{R}^3$ .*

*Proof of Lemma 4.1.* Defining  $\rho_t^\varepsilon = \xi_t^\varepsilon - \xi_t$  we have that

$$d\rho_t^\varepsilon = -\varepsilon \rho_t^\varepsilon - \varepsilon \xi_t + \mathcal{B}(\rho_t^\varepsilon, \xi_t) + \mathcal{B}(\xi_t, \rho_t^\varepsilon) + \mathcal{B}(\rho_t^\varepsilon, \rho_t^\varepsilon) + \sqrt{\varepsilon} \sigma dW_t.$$

We will make use of the following estimate which is straightforward to prove : there exists a  $c > 0$  so that

$$2|\mathcal{B}(\xi, \rho) \cdot \rho| + 2|\mathcal{B}(\rho, \xi) \cdot \rho| + 2\varepsilon|\xi \cdot \rho| \leq \varepsilon^2|\xi| + c|\xi||\rho|^2.$$

Applying Itô's formula to  $\rho \mapsto |\rho|^2$  and this estimate produces

$$\frac{d}{dt} \mathbf{E} |\rho_t^\varepsilon|^2 \leq c |\xi_t| \mathbf{E} |\rho_t^\varepsilon|^2 + \varepsilon (|\sigma|^2 + \varepsilon |\xi_t|)$$

Recalling that  $|\xi_t| = |\xi_0|$  and applying Gronwall's lemma produces the stated result.  $\square$

## 5. EXISTENCE AND UNIQUENESS OF INVARIANT MEASURES WITH NOISE

Similarly if we consider the evolution of the norm, we have the following result which is useful in establishing the existence of the invariant measure  $\mu^\varepsilon$  and the tightness of various objects.

**Proposition 1.** *For any integer  $p \geq 1$  there exists  $C(p) > 0$  so that for all  $t \geq 0$ ,  $\varepsilon > 0$ ,*

$$\mathbf{E} |\xi_t^\varepsilon|^{2p} \leq C(p) \left[ 1 + e^{-2\varepsilon t} \sum_{k=1}^p |\xi_0|^{2k} \right]$$

*Proof of Proposition 1.* Defining  $|\sigma|^2 = \sum_i \sigma_i^2$ , Itô's formula implies that

$$d|\xi_t^\varepsilon|^2 = -2\varepsilon |\xi_t^\varepsilon|^2 dt + |\sigma|^2 dt + dM_t^\varepsilon$$

for a martingale  $M_t^\varepsilon$  with quadratic variation satisfying

$$d\langle M^\varepsilon \rangle_t = (\sigma_1^2 (X_t^\varepsilon)^2 + \sigma_2^2 (Y_t^\varepsilon)^2) dt \leq \sigma_{max}^2 |\xi_t^\varepsilon|^2 dt$$

where  $\sigma_{max}^2 = \sigma_1^2 \vee \sigma_2^2$ . The proof then follows from Lemma 5.1 below.  $\square$

**Remark 1.** One can actually easily prove uniform in time bounds on  $\mathbf{E} \exp(\kappa X_t)$  for  $\kappa > 0$  but sufficiently small. See [HM08] for a proof using the exponential martingale estimate.

The following Lemma provides the key estimate to Proposition 1.

**Lemma 5.1.** *Let  $X_t$  be a semimartingale so that  $X_t \geq 0$ ,*

$$dX_t = (a - bX_t)dt + dM_t$$

*where  $a > 0$ ,  $b > 0$  and  $M_t$  is a continuous local martingale satisfying*

$$d\langle M \rangle_t \leq cX_t dt$$

*for some  $c > 0$ . Then for any integer  $p \geq 1$  there exist a constant  $C(p)$  (depending besides  $p$  only on  $a$ ,  $b$  and  $c$ ) so that for any  $X_0 \geq 0$  and  $t \geq 0$*

$$\mathbf{E}[X_t^p] \leq C(p) \left[ 1 + \sum_{k=1}^p e^{-bkt} X_0^k \right]$$

*Proof of Lemma 5.1.* Fixing an  $N > 0$  and defining the stopping time  $\tau = \inf\{t : X_t > N\}$  observe that

$$\mathbf{E}X_t \leq \mathbf{E}X_{t \wedge \tau} \leq at + X_0.$$

where the first inequality follows from Fatou's lemma applied to the limit  $N \rightarrow \infty$ . Using the assumption on the quadratic variation of  $M_t$  we now see that  $M_t$  is a  $L^2$ -martingale. Hence

$$(8) \quad \mathbf{E}X_t = e^{-bt} X_0 + \frac{a}{b} (1 - e^{-bt})$$

Now applying Itô's formula to  $X_t^p$  produces

$$dX_t^p = pX_t^{p-1}(a - bX_t)dt + \frac{p(p-1)}{2}X_t^{p-2}d\langle M \rangle_t + dM_t^{(p)}$$

where  $dM_t^{(p)} = pX_t^{p-1}dM_t$ . Using the same stopping time  $\tau$  and the same argument as before, we have

$$\mathbf{E}X_t^p \leq \mathbf{E}X_{t \wedge \tau}^p \leq X_0^p + (pa + cp(p-1)) \int_0^t \mathbf{E}X_s^{p-1} ds$$

Hence inductively we have a bound on  $\mathbf{E}X_t^p$  for all integer  $p \geq 1$  which implies that  $M_t^{(p)}$  is an  $L^2$ -martingale for all  $p \geq 1$ . Hence we have

$$\mathbf{E}X_t^p \leq e^{-bpt} X_0^p + (pa + \frac{c}{2}p(p-1)) \int_0^t e^{-bp(t-s)} \mathbf{E}X_s^{p-1} ds.$$

Proceeding inductively using this estimate and (8) as the base case produces the stated result.  $\square$

**Corollary 2.** *For each  $\varepsilon > 0$ , the Feller diffusion  $\{\xi_t^\varepsilon : t \geq 0\}$  possesses at least one invariant probability measure  $\mu^\varepsilon$ . Furthermore, any invariant probability measure  $\mu^\varepsilon$  satisfies*

$$\int_{\mathbf{R}^3} |\xi|^p \mu^\varepsilon(d\xi) \leq C(p)$$

for any integer  $p \geq 1$  where  $C(p)$  is the constant from Lemma 1 (which is independent of  $\varepsilon$ ). Hence the collection of probability measures which are invariant under the dynamics for any given  $\varepsilon > 0$  is tight.

*Proof of Corollary 2.* Since  $\xi_t^\varepsilon$  is a time-homogeneous Feller diffusion process and from Proposition 1 for fixed  $\varepsilon > 0$ , the collection of random vectors  $\{\xi_t^\varepsilon, t > 0\}$  is tight, the existence of an invariant probability measure  $\mu^\varepsilon$  follows by the Krylov-Bogolyubov theorem.

Defining  $\phi_{N,p}(\xi) = |\xi|^{2p} \phi(|\xi|/N)$  where  $\phi$  is a smooth function such that  $\phi(x) = 1$  for  $x \leq 1$ ,  $\phi(x) = 0$  for  $x \geq 2$ , and  $\phi$  decreases monotonically on  $(1, 2)$ , we see that

$$\begin{aligned} \int \phi_{N,p}(\xi) \mu^\varepsilon(d\xi) &= \int \mathbf{E}_{\xi_0} \phi_{N,p}(\xi_t^\varepsilon) \mu^\varepsilon(d\xi_0) \\ &\leq C(p) [1 + e^{-2\varepsilon t} \int \phi_{N,p}(\xi) \mu^\varepsilon(d\xi)] \\ &\leq C(p) [1 + e^{-2\varepsilon t} (N+1)^{2p}] \end{aligned}$$

Taking  $t \rightarrow \infty$ , followed by  $N \rightarrow \infty$ , the result follows from Fatou's Lemma. Since these bounds are uniform in  $\varepsilon$ , tightness follows immediately.  $\square$

The next result follows by hypoellipticity and the Stroock and Varadhan support theorem.

**Proposition 2.** *For any  $\varepsilon > 0$ , there exists a transition density  $p_t^\varepsilon(\xi, \eta)$  which is jointly smooth in  $(t, \xi, \eta)$  so that for all  $\xi \in \mathbf{R}^3$  and Borel  $A \subset \mathbf{R}^3$  one has*

$$P_t^\varepsilon(\xi, A) = \int_A p_t^\varepsilon(\xi, \eta) d\eta.$$

Additionally,  $\int_B p_t^\varepsilon(\xi, \eta) d\eta > 0$  for every  $\varepsilon > 0$ ,  $t > 0$ ,  $\xi \in \mathbf{R}^3$ , and any ball  $B \subset \mathbf{R}^3$ .

*Proof of Proposition 2.* Hypocoellipticity follows from the fact that taking Lie brackets of the drift with  $\partial_x$  and then  $\partial_y$  (the two noise directions) produces the third and missing direction  $\partial_z$ . This ensures the existence of a smooth density with respect to Lebesgue measure [Str08, Hör94a, Hör94b]. Positivity will then follow by showing that the support of the transition density is all of  $\mathbf{R}^3$ .

We will invoke the support theorem of Stroock and Varadhan [SV72]. Indeed, consider the controlled system associated to the SDE for  $(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$ , which reads

$$(9) \quad \begin{aligned} \frac{dx^\varepsilon}{dt}(t) &= y^\varepsilon(t)z^\varepsilon(t) - \varepsilon x^\varepsilon(t) + \sqrt{\varepsilon}\sigma_1 f_1(t) \\ \frac{dy^\varepsilon}{dt}(t) &= x^\varepsilon(t)z^\varepsilon(t) - \varepsilon y^\varepsilon(t) + \sqrt{\varepsilon}\sigma_2 f_2(t) \\ \frac{dz^\varepsilon}{dt}(t) &= -2x^\varepsilon(t)y^\varepsilon(t) - \varepsilon z^\varepsilon(t), \end{aligned}$$

where  $\{(f_1(t), f_2(t)), t \geq 0\}$  is the control at our disposal. Now by choosing appropriately the control, we can drive the two components  $(x^\varepsilon(t), y^\varepsilon(t))$  in time as short as we like to any desired position, which permits us to drive the last component  $z^\varepsilon(t)$  to any prescribed position in any prescribed time. The result follows.  $\square$

We are now in a position to give the proof of Theorem 3.1.

*Proof of Theorem 3.1.* Since by Proposition 2,  $P_t^\varepsilon$  has a smooth transition density, any invariant probability measure must have a smooth density which charges any ball  $B \subset \mathbf{R}^3$ . Recall the fact that in our setting any two distinct ergodic invariant probability measures must have disjoint supports which is impossible since the measures have densities and charge any open set. Uniqueness of the invariant probability measure follows immediately from the fact that any invariant measure can be decomposed into ergodic components [CFS82].  $\square$

## 6. THE FAST DYNAMICS

Since by the results in Section 4  $\xi_t^\varepsilon = (X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$  converges to  $\xi_t = (X_t, Y_t, Z_t)$ , in order to study the limiting invariant probability measure one needs to consider the system on ever increasing time intervals as  $\varepsilon \rightarrow 0$ . One must pick a time scale, depending on  $\varepsilon$ , so that the amount of randomness injected into the system is sufficient to keep the system from settling onto a deterministic trajectory as  $\varepsilon \rightarrow 0$ .

With this in mind consider the process  $\xi_t^\varepsilon$  on the fast scale  $t/\varepsilon$ . In other words, consider the process  $\tilde{\xi}_t^\varepsilon = \xi_{t/\varepsilon}^\varepsilon$  which solves the SDE

$$(10) \quad \dot{\tilde{\xi}}_t^\varepsilon = \frac{1}{\varepsilon} \mathcal{B}(\tilde{\xi}_t^\varepsilon, \tilde{\xi}_t^\varepsilon) - \tilde{\xi}_t^\varepsilon + \sigma \dot{W}_t.$$

Here we have used a slight abuse of notations, replacing the  $\varepsilon$ -dependent standard Brownian motion  $W_t^\varepsilon = \sqrt{\varepsilon}W_{t/\varepsilon}$  by  $W_t$ . In coordinates we will write  $(\tilde{X}_t^\varepsilon, \tilde{Y}_t^\varepsilon, \tilde{Z}_t^\varepsilon) = (X_{t/\varepsilon}^\varepsilon, Y_{t/\varepsilon}^\varepsilon, Z_{t/\varepsilon}^\varepsilon)$ .

Let  $\tilde{P}_t^\varepsilon$  be the Markov semigroup associated to (10) and defined for  $\psi : \mathbf{R}^3 \rightarrow \mathbf{R}$  by

$$(11) \quad (\tilde{P}_t^\varepsilon \psi)(\xi) = \mathbf{E}_\xi \psi(\tilde{\xi}_t^\varepsilon).$$

Associated with this right-action on functions we have a dual action on measures. We will denote this by left action rather than the often used  $(\tilde{P}_t^\varepsilon)^*$  notation. Hence



if  $\mu$  is a measure on  $\mathbf{R}^3$  and  $\psi$  a real-valued function on  $\mathbf{R}^3$  then

$$\mu \tilde{P}_t^\varepsilon \psi = \int_{\mathbf{R}^3} (\tilde{P}_t^\varepsilon \psi)(\xi) \mu(d\xi).$$

Of course, this time change does not change the set of invariant measures for the dynamics since the time change was not state dependent. Hence  $\tilde{P}_t^\varepsilon$  has the same unique invariant probability measure  $\mu^\varepsilon$  as  $P_t^\varepsilon$ .

**6.1. Fast evolution of conserved quantities.** One indication that this is the right time scale is that the conserved quantities  $(u, v) = \Phi(x, y, z)$  now continue to evolve randomly as  $\varepsilon \rightarrow 0$ . More precisely, defining the processes  $(U_t^\varepsilon, V_t^\varepsilon)$  by  $(U_t^\varepsilon, V_t^\varepsilon) = \Phi(\tilde{X}_t^\varepsilon, \tilde{Y}_t^\varepsilon, \tilde{Z}_t^\varepsilon)$  and applying Itô's formula shows that

$$(12) \quad \begin{aligned} dU_t^\varepsilon &= [2\sigma_1^2 - 2U_t^\varepsilon]dt + 4\sigma_1 \tilde{X}_t^\varepsilon dW_t^{(1)}, \\ dV_t^\varepsilon &= [2\sigma_2^2 - 2V_t^\varepsilon]dt + 4\sigma_2 \tilde{Y}_t^\varepsilon dW_t^{(2)}. \end{aligned}$$

We will show below in Section 8 that  $\{(U_t^\varepsilon, V_t^\varepsilon)\}_{\varepsilon>0}$  is tight, and that any accumulation point  $(U_t, V_t)$  solves the SDE

$$(13) \quad \begin{aligned} dU_t &= [2\sigma_1^2 - 2U_t]dt + \sigma_1 \sqrt{8(U_t - \Gamma(U_t, V_t))} dW_t^{(1)}, \\ dV_t &= [2\sigma_2^2 - 2V_t]dt + \sigma_2 \sqrt{8(V_t - \Gamma(U_t, V_t))} dW_t^{(2)}, \end{aligned}$$

where

$$(14) \quad \Gamma(u, v) = (u \wedge v) \Lambda \left( \frac{u \wedge v}{u \vee v} \right).$$

The function  $\Lambda$  will be defined in Section 7.3.2. However for our present discussion, it will be sufficient to state a few important facts whose proofs will be given later in Section 7.3.2.

**Proposition 3.**  $\Lambda(r)$  is a continuous and strictly increasing function on  $[0, 1]$  with  $\Lambda(0) = \frac{1}{2}$  and  $\Lambda(1) = 1$ . Furthermore as  $\varepsilon \rightarrow 0^+$ ,

$$\begin{aligned} \Lambda(\varepsilon) &= \frac{1}{2} + \frac{1}{16}\varepsilon + \frac{1}{32}\varepsilon^2 + o(\varepsilon^2), \\ \Lambda(1 - \varepsilon) &= 1 - \frac{2}{|\ln(\varepsilon)|} + o\left(\frac{1}{|\ln(\varepsilon)|}\right). \end{aligned}$$

In addition, on any closed interval in  $[0, 1]$ ,  $\Lambda$  is uniformly Lipschitz.

**6.2. Finite time behavior  $(U_t, V_t)$ .** Before stating and proving the main theorem of this section, let us establish three Lemmata.

**Lemma 6.1.** Let  $\{X_t, t \geq 0\}$  be a continuous  $\mathbf{R}_+$ -valued  $\mathcal{F}_t$ -adapted process which satisfies

$$\begin{aligned} dX_t &= (a - bX_t)dt + \sqrt{cX_t}dW_t, \\ X_0 &= x, \end{aligned}$$

where  $b, c > 0$ ,  $\{W_t, t \geq 0\}$  is a standard Brownian adapted to  $\mathcal{F}_t$  motion and  $x > 0$ . If  $a \geq c/2$ , then with probability one  $X_t > 0$  for all  $t \geq 0$ .

*Proof of Lemma 6.1.* We consider the SDE

$$\begin{aligned} dY_t &= c^{-1} (aY_t - bY_t^2) dt + Y_t dW_t, \\ Y_0 &= x, \end{aligned}$$

whose solution satisfies clearly

$$Y_t = x \exp \left( \left[ \frac{a}{c} - \frac{1}{2} \right] t - \frac{b}{c} \int_0^t Y_s ds + W_t \right),$$

and define  $A_t = \frac{1}{c} \int_0^t Y_s ds$ ,  $\eta_t = \inf\{s > 0, A_s > t\}$  and  $X_t = Y_{\eta_t}$ . It is not hard to see that there exists a standard Brownian motion, which by an abuse of notation we denote again by  $W$ , which is such that

$$\begin{aligned} dX_t &= (a - bX_t)dt + \sqrt{cX_t}dW_t, \\ X_0 &= x. \end{aligned}$$

Since  $X_{A_t} = Y_t$  and  $Y_t > 0$  for all  $t < \infty$ , for  $X_t$  to hit zero in finite time, it is necessary that  $A_\infty < \infty$  and  $Y_\infty = 0$ . But from the above formula for  $Y_t$ , we deduce that if  $a \geq c/2$ , on the event  $\{A_\infty < \infty\}$ ,  $\limsup_{t \rightarrow \infty} Y_t = +\infty$ , which implies that  $A_\infty = +\infty$ .  $\square$

We will need a slightly better result which generalizes the preceding Lemma.

**Lemma 6.2.** *Let  $\{X_t, t \geq 0\}$  and  $\{Y_t, t \geq 0\}$  be continuous  $\mathbf{R}_+$ -valued  $\mathcal{F}_t$ -adapted processes which satisfy  $0 \leq Y_t \leq X_t$  for all  $t \geq 0$ , with  $Y_0 > 0$ ,*

$$\begin{aligned} dX_t &= (a - bX_t)dt + \sqrt{cY_t}dW_t, \\ X_0 &= x, \end{aligned}$$

where  $b, c > 0$ ,  $\{W_t, t \geq 0\}$  is a standard  $\mathcal{F}_t$ -Brownian motion and  $x > 0$ . If  $a \geq c/2$ , then a. s.  $X_t > 0$  for all  $t \geq 0$ .

*Proof of Lemma 6.2.* We define

$$B_t = \int_0^t \frac{Y_s}{X_s} ds, \quad \sigma_t = \inf\{s > 0, B_s > t\}, \quad \text{and} \quad Z_t = X_{\sigma_t}.$$

There exists a standard Brownian motion, still denoted by  $W$ , such that

$$\begin{aligned} dZ_t &= (a - bZ_t) \frac{Z_t}{Y_{\sigma_t}} dt + \sqrt{cZ_t} dW_t, \\ X_0 &= x. \end{aligned}$$

Define two sequences of stopping times as follows.  $S_0 = 0$ , and for  $k \geq 1$ ,

$$T_k = \inf \left\{ t > S_{k-1}, Z_t < \frac{a}{2b} \right\} \quad \text{and} \quad S_k = \inf \left\{ t > T_k, Z_t > \frac{a}{b} \right\}.$$

On each interval  $[T_k, S_k]$ , since  $Z_t/Y_{\sigma_t} \geq 1$ , by a standard comparison theorem for SDEs we can bound from below  $Z_t$  by the solution of the equation of the previous Lemma, starting from  $a/2b$ . Hence  $Z_t$  never hits zero.  $\square$

Using Lemma 6.2, we are now in a position to prove the following result.

**Theorem 6.3.** *Assume that the initial condition  $(U_0, V_0)$  satisfies  $U_0 > 0$ ,  $V_0 > 0$ . Any solution of equation (13) lives in the set  $(0, +\infty) \times (0, +\infty)$  for all positive times.*

*Proof.* We first prove that any solution  $(U_t, V_t)$  never hits the two axis, except possibly at  $(0, 0)$ . The fact that  $(U_t, V_t)$  cannot hit  $(0, v)$  with  $v > 0$  follows clearly from the equation for  $U_t$  and Lemma 6.2, once we have noted that whenever  $U_t < V_t$ ,  $U_t - \Gamma(U_t, V_t) \leq U_t/2$ , as follows from Proposition 3.

The same proof shows that  $(U_t, V_t)$  cannot hit  $(u, 0)$  with  $u > 0$ . It remains to show that  $(U_t, V_t)$  cannot hit  $(0, 0)$ . Let  $a = \sigma_1^{-2}$ ,  $b = \sigma_2^{-2}$ ,  $K_t = aU_t + bV_t$ . There exists a standard Brownian motion  $W_t$  such that

$$dK_t = (4 - 2K_t)dt + \sqrt{8[K_t - (a+b)\Gamma(U_t, V_t)]}dW_t.$$

The result again follows from Lemma 6.2, since  $\Gamma(U_t, V_t) \geq 0$ .  $\square$

We next establish a crucial property shared by all the possible accumulation points of the collection  $\{(U^\varepsilon, V^\varepsilon)\}_{\varepsilon > 0}$ .

**Theorem 6.4.** *If  $(U, V) = \lim_{n \rightarrow \infty} (U^{\varepsilon_n}, V^{\varepsilon_n})$ , for some sequence  $\varepsilon_n \rightarrow 0$ , then*

$$\int_0^t \mathbf{1}_{\{U_s = V_s\}} ds = 0 \text{ for all } t > 0 \text{ almost surely.}$$

*Proof.* For any  $M, \delta > 0$ , let

$$\psi_{M,\delta}(x) = \left[ \log \left( \frac{1}{|x|} \right) \wedge M \right] \mathbf{1}_{[-\delta, \delta]}(x), \quad \psi_\delta(x) = \log \left( \frac{1}{|x|} \right) \mathbf{1}_{[-\delta, \delta]}(x).$$

We define  $\varphi_{M,\delta}$  and  $\varphi_\delta \in C^1(\mathbf{R}) \cap W^{2,\infty}(\mathbf{R})$  by

$$\begin{aligned} \varphi_{M,\delta}(0) &= 0, \quad \varphi'_{M,\delta}(0) = 0, \quad \varphi''_{M,\delta} = \psi_{M,\delta}, \\ \varphi_\delta(0) &= 0, \quad \varphi'_\delta(0) = 0, \quad \varphi''_\delta = \psi_\delta. \end{aligned}$$

Let  $J_t^\varepsilon = U_t^\varepsilon - V_t^\varepsilon$ . It follows from Itô's formula, which can be applied here although  $\varphi_{M,\delta} \notin C^2(\mathbf{R})$ , that for each  $M > 0$ ,  $n \geq 1$ ,

$$\begin{aligned} 8\mathbf{E} \int_0^t \left[ \sigma_1^2 (\tilde{X}_s^{\varepsilon_n})^2 + \sigma_2^2 (\tilde{Y}_s^{\varepsilon_n})^2 \right] \psi_{M,\delta}(J_s^{\varepsilon_n}) ds \\ \leq \mathbf{E} \left( \varphi_{M,\delta}(J_t^{\varepsilon_n}) - \varphi_{M,\delta}(J_0^{\varepsilon_n}) - 2 \int_0^t (\sigma_1^2 - \sigma_2^2 - J_s^{\varepsilon_n}) \varphi'_{M,\delta}(J_s^{\varepsilon_n}) ds \right). \end{aligned}$$

We now let  $M \rightarrow \infty$ . Since  $\sigma_1^2 (\tilde{X}_s^{\varepsilon_n})^2 + \sigma_2^2 (\tilde{Y}_s^{\varepsilon_n})^2 > 0$  ( $ds \times d\mathbf{P}$ )-almost everywhere, we deduce that

$$\begin{aligned} 8\mathbf{E} \int_0^t \left[ \sigma_1^2 (\tilde{X}_s^{\varepsilon_n})^2 + \sigma_2^2 (\tilde{Y}_s^{\varepsilon_n})^2 \right] \psi_\delta(J_s^{\varepsilon_n}) ds \\ \leq \mathbf{E} \left( \varphi_\delta(J_t^{\varepsilon_n}) - \varphi_\delta(J_0^{\varepsilon_n}) - 2 \int_0^t (\sigma_1^2 - \sigma_2^2 - J_s^{\varepsilon_n}) \varphi'_\delta(J_s^{\varepsilon_n}) ds \right). \end{aligned}$$

We now take the limit in the last inequality as  $n \rightarrow \infty$ , and deduce from Fatou's Lemma that

$$\begin{aligned} \mathbf{E} \int_0^t \left[ \sigma_1^2 (U_s - \Gamma(U_s, V_s)) + \sigma_2^2 (V_s - \Gamma(U_s, V_s)) \right] \psi_\delta(J_s) ds \\ \leq \mathbf{E} \left( \varphi_\delta(J_t) - \varphi_\delta(J_0) - 2 \int_0^t (\sigma_1^2 - \sigma_2^2 - J_s) \varphi'_\delta(J_s) ds \right). \end{aligned}$$

It follows from Proposition 3 that to any  $c > 0$ , we can associate  $\delta > 0$  and  $a > 0$  such that whenever  $u, v \geq c > 0$ , and  $-\delta \leq k = u - v \leq \delta$ ,

$$4 [\sigma_1^2(u - \Gamma(u, v)) + \sigma_2^2(v - \Gamma(u, v))] \log \left( \frac{1}{|k|} \right) \geq a > 0.$$

Consequently the above establishes that

$$\begin{aligned} a \mathbf{E} \int_0^t \mathbf{1}_{U_s \geq c, V_s \geq c} \mathbf{1}_{(-\delta, \delta)}(J_s) ds \\ \leq \mathbf{E} \left( \varphi_\delta(J_t) - \varphi_\delta(J_0) - 2 \int_0^t (\sigma_1^2 - \sigma_2^2 - J_s) \varphi'_\delta(J_s) ds \right), \end{aligned}$$

and letting finally  $\delta \rightarrow 0$ , we deduce that

$$\mathbf{E} \int_0^t \mathbf{1}_{U_s \geq c, V_s \geq c} \mathbf{1}_{\{0\}}(J_s) ds = 0.$$

Since we know that both  $U_t$  and  $V_t$  never reaches 0, and  $c > 0$  is arbitrary, this shows that  $J_t = U_t - V_t$  spends a.s. zero time at 0, i.e. that the process  $(U_t, V_t)$  spends a.s. zero time on the diagonal.  $\square$

Now that we know that  $(U, V)$  spends no time on the diagonal, we can introduce the following time change. Let for  $u, v > 0$ ,

$$(15) \quad F(u, v) = \begin{cases} 1 - \Lambda \left( \frac{u \wedge v}{u \vee v} \right), & \text{if } \frac{u \wedge v}{u \vee v} \geq \frac{1}{2}, \\ 1 - \Lambda \left( \frac{1}{2} \right), & \text{if } \frac{u \wedge v}{u \vee v} < \frac{1}{2}. \end{cases}$$

Let us define the time change

$$(16) \quad \begin{aligned} A_t &= \int_0^t F(U_s, V_s) ds, \quad \eta_t = \inf\{s > 0, A_s > t\}, \\ H_t &= U_{\eta_t}, \quad \text{and } K_t = V_{\eta_t}. \end{aligned}$$

There exists a two-dimensional Wiener process, which we still denote by  $(W_t^{(1)}, W_t^{(2)})$ , such that

$$(17) \quad \begin{aligned} dH_t &= 2 \frac{\sigma_1^2 - H_t}{F(H_t, K_t)} dt + 2\sqrt{2}\sigma_1 \sqrt{\frac{H_t - H_t \wedge K_t}{F(H_t, K_t)} + (H_t \wedge K_t)G(H_t, K_t)} dW_t^{(1)}, \\ dK_t &= 2 \frac{\sigma_2^2 - K_t}{F(H_t, K_t)} dt + 2\sqrt{2}\sigma_2 \sqrt{\frac{K_t - H_t \wedge K_t}{F(H_t, K_t)} + (H_t \wedge K_t)G(H_t, K_t)} dW_t^{(2)}, \end{aligned}$$

where

$$G(H_t, K_t) = \frac{1 - \Lambda \left( \frac{H_t \wedge K_t}{H_t \vee K_t} \right)}{F(H_t, K_t)}.$$

It is easily verified that the diffusion matrix of this system is locally uniformly elliptic in  $(0, +\infty) \times (0, +\infty)$  and continuous. However, the drift is unbounded near the diagonal. We will now prove uniqueness of the weak solution of (17), using methods and results from Portenko [Por90]. [Por90] constructs a weak solution to an equation like (17) from the solution without drift, using Girsanov's theorem, provided the drift is in  $L^p(\mathbf{R}^2)$ , with  $p > 4$ . His uniqueness theorem is proved under conditions which are difficult to verify. The condition is tailored to make sure that Girsanov's theorem can be used to show that the law of the equation with drift is

absolutely continuous with respect to that of the equation without drift. We will do that by verifying the condition of Lemma 1.1 from [Por90], which we now state

**Lemma 6.5.** *Assume that  $\{Z_t, t \geq 0\}$  is a non-negative progressively measurable process, adapted to the  $\sigma$ -algebra  $\{\mathcal{F}_t, t \geq 0\}$ . Suppose that there exists a mapping  $\rho$  from the set of subintervals of  $[0, T]$  into  $\mathbf{R}_+$ , such that*

- (i)  $\mathbf{E} \left[ \int_s^t Z_r dr \middle| \mathcal{F}_s \right] \leq \rho(s, t)$ , for all  $0 \leq s < t \leq T$ ;
- (ii)  $\rho(t_1, t_2) \leq \rho(t_3, t_4)$ , whenever  $(t_1, t_2) \subset (t_3, t_4)$ ;
- (iii)

$$\lim_{h \downarrow 0} \sup_{\substack{0 \leq s < t \leq T \\ t-s \leq h}} \rho(s, t) = \kappa$$

Then for any  $\lambda < \kappa^{-1}$ ,

$$\mathbf{E} \exp \left( \lambda \int_0^T Z_t dt \right) < \infty.$$

We intend to apply Proposition 6.5 for a  $Z_t$  which will give us sufficient control over the drift in (17) that we can use Novikov's criterion and Girsanov's Theorem to transform the SDE (17) into the same equation without drift, and hence prove uniqueness of the solution using an argument in the vein of Theorem 1.2 from [Por90].

To better understand how to use Lemma 6.5 to remove the drift from (17) en route to prove uniqueness of the solution, we take a close look at the drift term. Our uniqueness argument exploits the fact that, since uniqueness is a local property, we can modify the coefficients of the  $(H, K)$  equation outside the set  $\{(h, k), M^{-1} \leq h, k \leq M\}$  for some arbitrary  $M > 0$ , so that the resulting equation takes the form

$$(18) \quad \begin{aligned} dH_t &= \frac{f_1(H_t)}{F(H_t, K_t)} dt + \sigma_1 \sqrt{g(H_t, K_t)} dW_t^{(1)}, \\ dK_t &= \frac{f_2(K_t)}{F(H_t, K_t)} dt + \sigma_2 \sqrt{g(K_t, H_t)} dW_t^{(2)}. \end{aligned}$$

where for some  $C > 0$ ,  $C^{-1} \leq g(h, k) \leq C$  and  $|f_1(h)| + |f_2(k)| \leq C$ , for all  $h, k \in \mathbf{R}$ . Hence the only possible difficulty will arise if when  $F$  is small. From the definition of  $F$  in (15) and the asymptotics in Lemma 3 we note that  $F(h, k)$  is small when  $\frac{|h-k|}{h \vee k}$  is small.

Hence if we wish to use Girsanov's theorem to remove the drift from (18), the danger comes from  $|h-k|$  small if we restrict our attention to the set  $\{(h, k), M^{-1} \leq h, k \leq M\}$ . The following simple observation is useful to control the drift.

**Lemma 6.6.** *If  $a, b > 0$  then*

$$\log \left( \frac{b}{a} \right) \leq b + \log \left( \frac{1}{a} \right) \mathbf{1}_{\{a \leq 1\}} = b + |\log(a)| \mathbf{1}_{\{a \leq 1\}}.$$

*Proof.* The inequality follows from the fact that if  $b > 0$  then  $b > \log(b)$  and if  $a \in (0, 1]$  then  $\log(1/a) \geq 0$ . The equality follows from  $\log \left( \frac{1}{a} \right) \mathbf{1}_{\{a \leq 1\}} = |\log \left( \frac{1}{a} \right)| \mathbf{1}_{\{a \leq 1\}} = |\log(a)| \mathbf{1}_{\{a \leq 1\}}$ .  $\square$

Combining Lemma 6.6 and the asymptotics in Lemma 3 we note that

$$\frac{2}{F(h, k)} \lesssim |h \vee k| + |\log(|h-k|)| \mathbf{1}_{|h-k| \leq 1}.$$

Hence on the set  $M^{-1} \leq h \vee k \leq M$ , controlling  $F^{-1}(h, k)$  amounts to estimating  $|\log(|h - k|)|$  for  $|h - k| < a$ , where  $0 < a \leq 1$  is arbitrary.

Hence, defining  $N_t := H_t - K_t$ , if we desire to apply Lemma 6.5 to  $Z_t = 1/F^2(H_t, K_t)$  it will be sufficient to estimate

$$(19) \quad \mathbf{E} \left[ \int_s^t \mathbf{1}_{\{|N_r| \leq 1/2\}} \log^2 |N_r| dr \middle| \mathcal{F}_s \right] = \mathbf{E} \left[ \int_s^t \mathbf{1}_{\{|N_r| \leq 1/2\}} \log^2 |N_r| dr \middle| H_s, K_s \right],$$

for  $0 \leq H_s, K_s \leq M$  for some  $M > 0$ . Notice that as we will eventually prove that Lemma 6.5 holds with  $\kappa = 0$  the conclusion of Lemma 6.5 will be more than sufficient to invoke Novikov's criterion.

Recall that  $N_t$  satisfies

$$dN_t = \frac{f_1(H_t) - f_2(K_t)}{F(H_t, K_t)} dt + \sqrt{\bar{g}(H_t, K_t)} dW_t$$

with  $f_1, f_2, \bar{g}$  bounded and  $\bar{g}$  bounded away from zero. As a prologue to the needed estimate on (19), we prove the following result.

**Lemma 6.7.** *Let  $0 < a < b < 2a \leq 1$ . For any  $M > 0$ , there exists a constant  $C_M$  such that for any  $0 \leq s < t$  with  $t - s \leq 1$ ,  $|H_s|, |K_s| \leq M$ ,*

$$\mathbf{E} \left[ \int_s^t \mathbf{1}_{\{a \leq |N_r| \leq b\}} \middle| H_s, K_s \right] \leq [(t - s) \wedge C_M (b - a)^{4/5}].$$

PROOF: It clearly suffices to prove that

$$\mathbf{E} \left[ \int_s^t \mathbf{1}_{\{a \leq |N_r| \leq b\}} \middle| H_s, K_s \right] \leq C_M (b - a)^{4/5}.$$

We prove this result with  $\{a \leq |N_r| \leq b\}$  replaced by  $\{a \leq N_r \leq b\}$ . The same proof would estimate similarly  $\{-b \leq N_r \leq -a\}$ . Let  $\varphi_{a,b} \in C^1(\mathbf{R}) \cap W^{2,\infty}(\mathbf{R})$  be defined by  $\varphi_{a,b}(0) = \varphi'_{a,b}(0) = 0$ ,  $\varphi''_{a,b}(x) = \mathbf{1}_{[a,b]}(x)$ . Also  $\varphi_{a,b} \notin C^2$ , we can apply Itô's formula to obtain

$$\begin{aligned} \frac{1}{2} \int_s^t \bar{g}^2(H_r, K_r) \mathbf{1}_{\{a \leq N_r \leq b\}} dr &= \varphi_{a,b}(N_t) - \varphi_{a,b}(N_s) \\ &\quad - \int_s^t \frac{f_1(H_r) - f_2(K_r)}{F(H_r, K_r)} \varphi'_{a,b}(N_r) dr \\ &\leq (b - a)[|N_t| + |N_s|] \\ &\quad + C(b - a) \int_s^t [1 + \log(|N_r|^{-1}) \mathbf{1}_{\{a \leq N_r \leq 1\}}] dr \\ &\leq C[1 + |N_t| + |N_s| + \log(1/a)](b - a). \end{aligned}$$

Since there exists  $C_M$  such that  $\mathbf{E}[|N_t| + |N_s| \mid H_s, K_s] \leq C_M$  for  $|H_s| \leq M$ ,  $|K_s| \leq M$  and  $t - s \leq 1$ , and  $\sup_{0 < a \leq 1} \log(1/a)(b - a)^{1/5} < \infty$  (we recall that  $b - a \leq a$ ), the result is proved.  $\square$

With this result in hand, we now return to the needed estimate on (19) which is contained in the following Lemma.

**Lemma 6.8.** *For any  $M > 0$ , there exists a constant  $C_M$  such that, for any  $0 \leq s < t$  with  $t - s \leq 1$ , on the event  $|H_s|, |K_s| \leq M$ ,*

$$\mathbf{E} \left[ \int_s^t \mathbf{1}_{\{|N_r| \leq 1/2\}} \log^2 |N_r| dr \middle| H_s, K_s \right] \leq C_M (t - s)^{1/4}.$$

PROOF: In this proof,  $C$  stands for a constant which may vary on each line.

$$\begin{aligned} & \mathbf{E} \left[ \int_s^t \mathbf{1}_{\{|N_r| \leq 1/2\}} \log^2 |N_r| dr \middle| H_s, K_s \right] \\ &= \sum_{k=2}^{\infty} \mathbf{E} \left[ \int_s^t \mathbf{1}_{\{(k+1)^{-1} \leq |N_r| \leq k^{-1}\}} \log^2 |N_r| dr \middle| H_s, K_s \right] \\ &\leq \sum_{k=2}^{\infty} \log^2(k+1) \mathbf{E} \left[ \int_s^t \mathbf{1}_{\{(k+1)^{-1} \leq |N_r| \leq k^{-1}\}} dr \middle| H_s, K_s \right] \\ &\leq C \sum_{k=2}^{\infty} \log^2(k+1) \left( \frac{1}{k^{8/5}} \wedge (t-s) \right) \\ &\leq C(t-s) \sum_{k=2}^{N(s,t)} \log^2(k+1) + C \sum_{k=N(s,t)}^{\infty} \log^2(k+1) \frac{1}{k^{8/5}}, \end{aligned}$$

where  $N(s, t) = \lceil (t-s)^{-5/8} \rceil$ , and we have used for the second inequality Lemma 6.7 and  $k^{-1} - (k+1)^{-1} \leq k^{-2}$ . For some constant  $C$ , for all  $k \geq 2$ ,

$$\log^2(k+1) \leq Ck^{1/5},$$

hence the second term in the last right-hand side is bounded by

$$C \sum_{k=N(s,t)}^{\infty} k^{-7/5} \leq C(t-s)^{1/4}.$$

Now the first term is bounded by

$$C(t-s)^{3/8} \log^2((t-s)^{-5/8}) \leq C(t-s)^{1/4}.$$

The claimed result is proved.  $\square$

We are now in a position to establish the desired uniqueness result.

**Theorem 6.9.** *Equation (13) has a unique solution which spends zero time on the diagonal. Furthermore the whole sequence  $(U^\varepsilon, V^\varepsilon)$  converges weakly to this solution as  $\varepsilon \rightarrow 0$ .*

*Proof.* Consider any solution of (13) which spends zero time on the diagonal. The time-change defined by (16) transforms this process into a solution  $(H_t, K_t)$  of (17).

Lemma 6.8 combined with Lemma 6.5 shows that, locally in the open positive quadrant, the law of that solution to (17) is absolutely continuous with respect to that of the same SDE, but without drift. That last equation has a unique weak solution, according to Theorem 7.2.1 in [SV79]. Now any solution to (17) coincides with the one constructed via Girsanov's theorem in Theorem 1.1 of [Por90] whose assumptions clearly hold in our case (alternatively, Girsanov's theorem could also be used thanks to a simplified version of Lemma 6.8 together with Lemma 6.5 again).

Now that we have a unique weak solution  $(H_t, K_t)$  to (17), we can time-change it back to the original  $(H_t, K_t)$ , i.e. defining  $A'_t = \int_0^t \frac{ds}{F(H_s, K_s)}$ ,  $\eta'_t = \inf\{s > 0, A_s > t\}$ ,  $U_t = H_{\eta'_t}$ ,  $V_t = K_{\eta'_t}$ . Indeed a weaker version (without the power 2 of the

logarithm) of Lemma 6.9 shows that  $A'_t < \infty$ , for all  $t < \infty$ . Clearly  $A'_t \rightarrow \infty$ , as  $t \rightarrow \infty$ . So this defines  $(U_t, V_t)$  for all  $t \geq 0$ , and this process coincides with the arbitrary solution which spends zero time on the diagonal. But the law of that process is uniquely characterized as being the time-change of the unique weak solution of (17).

Finally, it follows from this conclusion and Theorem 6.4 that all accumulation points of the collection  $\{(U^\varepsilon, V^\varepsilon), \varepsilon > 0\}$  have the same law, hence the whole sequence converges. This proves the stated result.  $\square$

From now on  $(U_t, V_t)$  will always refer to the process whose law has just been uniquely characterized.

Since the martingale problem associated to (17) is well posed,  $(H_t, K_t)$  is a Markov process, and from Theorem 6.3 in [Var07] (this theorem is stated in dimension 1, but exactly the same argument works in our case), so is its time-change  $(U_t, V_t)$ . We call  $Q_t$  the Markov semigroup associated to that process, which is defined for  $\phi: \mathbf{R}_+^2 \rightarrow \mathbf{R}$  by

$$(20) \quad (Q_t \phi)(u, v) = \mathbf{E}_{(u,v)}[\phi(U_t, V_t)].$$

**Remark 2.** We note that in both cases  $\sigma_1 = \sigma_2$  and  $\sigma_1 \neq \sigma_2$ , the law of the uniquely characterized solution of (13) and that of a non-degenerate SDE in the quadrant  $\mathbf{R}_+ \times \mathbf{R}_+$  are equivalent.

This is in sharp contrast with the result one would get if the diffusion coefficient would degenerate in a more regular way on the diagonal (e.g. it would be Lipschitz). In the latter case, in the case  $\sigma_1 = \sigma_2$ , the solution would stay on the diagonal once it has hit it. In the case  $\sigma_1 \neq \sigma_2$ , the solution would stay in the set  $u \geq v$  or  $v \geq u$ , depending upon the sign of  $\sigma_1^2 - \sigma_2^2$ , after having hit once the diagonal.

Often the period around an orbit diverges like  $1/\log(\rho)$  while approaching a heteroclinic cycle or a homoclinic orbit. Here  $\rho$  is the distance from the limiting orbit. This often leads to coefficients which vanish very slowly. This is the situation in our setting. Hence while it may seem esoteric at first, in fact it is likely to be generic in many settings.

**6.3. Longtime behavior of  $(U, V)$ .** Unlike the pair  $(U_t^\varepsilon, V_t^\varepsilon)$ , the pair  $(U_t, V_t)$  constructed from  $(H_t, K_t)$  in the previous section form a Markov process. Hence, we can speak of an invariant probability measure for the Markov semigroup  $Q_t$ .

Observe that

$$d(U_t + V_t) = [a - 2(U_t + V_t)]dt + dM_t$$

where  $a > 0$ .  $M_t$  is a continuous local Martingale satisfying

$$d\langle M \rangle_t \leq c(U_t + V_t)dt$$

for some positive  $c$  (recall that  $\Gamma(u, v) \leq u \wedge v$ ). Hence the following result follows from Lemma 5.1.

**Proposition 4.** *For any  $p \geq 1$ , there exists a constant  $C(p)$  so that*

$$\sup_{t \geq 0} \mathbf{E}[U_t^p + V_t^p] \leq C(p)[1 + U_0^p + V_0^p].$$

We shall also need the an analogous result for the  $(H, K)$  process. We could prove stronger results but the following will be sufficient for our purposes.



**Lemma 6.10.**

$$\sup_{t \geq 0} \mathbf{E} (|H_t|^2 + |K_t|^2) < \infty.$$

*Proof.* It suffices to treat  $H_t$ . Defining  $G_t = [(H_t - \sigma_1^2)_+]^2$ , we note that

$$\frac{d}{dt} \mathbf{E}(G_t^2) \leq a - b \mathbf{E}[G_t^2],$$

for some positive constants  $a$  and  $b$ . The result follows readily.  $\square$

**Lemma 6.11.** *The process  $(H, K)$  possesses a unique invariant probability measure  $\nu$ . Additionally,  $\nu$  has a density which is everywhere positive in the interior of  $(0, \infty) \times (0, \infty)$  and*

$$\frac{1}{F} \in L^1(\nu).$$

where  $F$  was defined in (15) and was used in the time change between  $(U, V)$  and  $(H, K)$ .

*Proof of Lemma 6.11.* Lemma 6.10 gives the needed tightness of the averaged transition densities of  $(H, K)$  to employ the Krylov–Bogolyubov theorem to show the existence of an invariant measure (just as we did in the proof of Corollary 2). Since  $(H, K)$  was obtained by a Girsanov transformation from a non degenerate diffusion with zero drift, its transition probabilities have a density which is positive in the open positive quadrant. This immediately implies that any invariant measure must have a density which respect to Lebesgue which is positive in the open positive quadrant. This in turn implies that there can only be one invariant probability measure which we will henceforth denote by  $\nu$ .

The Birkoff ergodic theorem and the fact that  $\nu$  and any transition probability have densities which are positive in the open positive quadrant imply that for any initial distribution and any  $\phi \in L^1(\nu)$

$$(21) \quad \frac{1}{t} \int_0^t \phi(H_s, K_s) ds \rightarrow \int \phi d\nu \quad \text{a.s. as } t \rightarrow \infty.$$

It is not hard to see that whether  $\int (1/F) d\nu < \infty$  or  $= \infty$ , in both cases, as  $t \rightarrow \infty$ ,

$$\frac{1}{t} \int_0^t \frac{ds}{F(H_s, K_s)} \rightarrow \int \frac{1}{F} d\nu \quad \text{a.s.}$$

The case  $\int (1/F) d\nu = \infty$  follows as follows. For any  $M > 0$ ,

$$\int \left( \frac{1}{F} \wedge M \right) d\nu \leq \liminf_{t \rightarrow \infty} \frac{1}{t} \int_0^t \frac{ds}{F(H_s, K_s)},$$

which implies by monotone convergence that

$$+\infty \leq \liminf_{t \rightarrow \infty} \frac{1}{t} \int_0^t \frac{ds}{F(H_s, K_s)},$$

hence the result.

Consequently, in order to prove that  $1/F \in L^1(\nu)$ , as a consequence of Fatou's Lemma, all we have to show is that there exists  $C > 0$  and  $T > 0$  such that for all  $t \geq T$ ,

$$(22) \quad \mathbf{E} \int_0^t \frac{ds}{F(H_s, K_s)} \leq Ct.$$

It follows from Lemma 6.10 and Lemma 6.6 that all (22) will follow from the fact that for some constant  $C$  and all  $t \geq 1$ ,

$$\int_0^t \mathbf{E} [|\log |N_s|| \mathbf{1}_{\{|N_s| \leq 1\}}] ds \leq Ct.$$

Consider the function  $\varphi \in C^1(\mathbf{R}) \cap W^{2,\infty}(\mathbf{R})$  defined as

$$\varphi(x) = \begin{cases} \frac{7}{4} - x, & \text{if } x < -1; \\ \frac{x^2}{4}(3 - 2 \log(|x|)), & \text{if } -1 \leq x \leq 1; \\ x - \frac{1}{4}, & \text{if } x > 1. \end{cases}$$

We then have

$$\varphi'(x) = \begin{cases} -1, & \text{if } x < -1; \\ x(1 - \log(|x|)), & \text{if } -1 \leq x \leq 1; \\ 1, & \text{if } x > 1; \end{cases}$$

and

$$\varphi''(x) = (-\log(|x|))_+.$$

We now deduce from Itô's formula

$$\begin{aligned} \mathbf{E} \int_0^t |\log |N_s|| \mathbf{1}_{\{|N_s| \leq 1\}} \bar{g}(H_s, K_s) ds = \\ 2\mathbf{E}[\varphi(N_t) - \varphi(N_0)] - 2\mathbf{E} \int_0^t \frac{\varphi'(N_s)}{F(H_s, K_s)} [f_1(H_s) - f_2(K_s)] ds. \end{aligned}$$

The result now follows easily from Lemma 6.10 and the fact that the process  $F^{-1}(H_s, K_s)\varphi'(N_s)$  is bounded.  $\square$

**Theorem 6.12.** *The semigroup  $Q_t$  defined in the previous section possesses a unique invariant probability measure  $\lambda$ . Furthermore,  $\lambda(du, dv) = \rho(u, v) du dv$  where  $\rho$  is continuous away from the diagonal and positive in the positive open quadrant. Lastly  $\rho(u, v) \rightarrow \infty$ , as  $|u - v| \rightarrow 0$  in the positive quadrant.*

*Proof of Theorem 6.12.* Let  $\rho(u, v)$  be the density of the unique invariant measure  $\nu$  as guaranteed by Lemma 6.11. The same lemma also states that  $\rho$  is positive in the open positive quadrant.

Now, using the notations from the proof of Theorem 6.9, for any measurable and locally bounded  $f : (0, \infty) \times (0, \infty) \rightarrow \mathbf{R}$  one has

$$\begin{aligned} \int_0^t f(U_s, V_s) ds &= \int_0^t f(H_{\eta'_s}, K_{\eta'_s}) ds \\ &= \int_0^{\eta'_t} \frac{f(H_r, K_r)}{F(H_r, K_r)} dr \\ \frac{1}{t} \int_0^t f(U_s, V_s) ds &= \frac{\eta'_t}{t} \times \frac{1}{\eta'_t} \int_0^{\eta'_t} \frac{f(H_r, K_r)}{F(H_r, K_r)} dr \end{aligned}$$

If one assumes that  $f$  is bounded then Lemma 6.11 ensures that  $f/F \in L^1(\nu)$  and that for any initial distribution

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(U_s, V_s) ds = \left( \lim_{t \rightarrow \infty} \frac{\eta'_t}{t} \right) \times \int \frac{f}{F} d\nu \quad \text{a.s.}$$

where the convergence of  $\frac{1}{\eta_t} \int_0^{\eta_t} \frac{f(H_r, K_r)}{F(H_r, K_r)} dr$  to  $\int \frac{f}{F} d\nu$  is ensured by (21). The same computation with  $f \equiv 1$  permits one to conclude that

$$\lim_{t \rightarrow \infty} \frac{\eta_t'}{t} = \left( \int \frac{1}{F} d\nu \right)^{-1}.$$

In summary, we have that for any  $f$  measurable and bounded, for any initial probability measure,

$$(23) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(U_s, V_s) ds = \frac{\int \frac{f}{F} d\nu}{\int \frac{1}{F} d\nu}.$$

This is enough to show that our process  $(U_t, V_t)$  has the unique invariant probability measure

$$\lambda(du, dv) = \frac{F^{-1}(u, v) \rho(u, v) dudv}{\int_{\mathbf{R}_+^2} F^{-1}(u, v) \rho(u, v) dudv}.$$

Indeed, to see uniqueness, let  $\lambda'$  be any ergodic invariant probability measure for  $(U_t, V_t)$ . Then for any bounded  $f$ , the Birkoff ergodic theorem implies that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(U_s, V_s) ds = \int f d\lambda' \quad \text{a.s.}$$

for  $\lambda'$ -almost every initial  $(U_0, V_0)$ . Combining this with (23) implies that  $\int f d\lambda' = \int f d\lambda$  for all bounded  $f$  which in turn implies  $\lambda = \lambda'$ . Since any invariant measure can be decomposed into ergodic invariant measures the uniqueness of  $\lambda$  is proved.  $\square$

## 7. THE DETERMINISTIC DYNAMICS

We now investigate more fully the deterministic dynamics given in (1) and obtained by formally setting  $\varepsilon = 0$  in (7). As already mentioned, (1) has two conserved quantities  $(u, v) = \Phi(\xi_0)$  which are constant on any given orbit. If  $\xi_0 = (X_0, Y_0, Z_0)$  then  $u = 2x^2 + z^2$  and  $v = 2y^2 + z^2$  give two independent equations. Since we are working in three dimensions, the locus of the solutions, which contains the points in the orbits, is a one-dimensional curve. We undertake this study since the  $\frac{1}{\varepsilon} \mathcal{B}$  term in (10) implies that on the fast timescale the solution will make increasingly many turns very near a deterministic orbit of (1), before the stochastic or dissipative terms cause appreciable diffusion or drift from the current deterministic orbit.

**7.1. Structure of orbits.** If  $u \neq v$  then the orbit is a simple periodic orbit which is topologically equivalent to a circle. In this case, there are two disjoint orbits which are solutions. If  $u > v$ , one such orbit is given by

$$\Gamma_{u,v}^+ = \left\{ \left( \sqrt{\frac{u-z^2}{2}}, \pm \sqrt{\frac{v-z^2}{2}}, z \right) : z \in [-\sqrt{v}, \sqrt{v}] \right\},$$

and another by

$$\Gamma_{u,v}^- = \left\{ \left( -\sqrt{\frac{u-z^2}{2}}, \pm \sqrt{\frac{v-z^2}{2}}, z \right) : z \in [-\sqrt{v}, \sqrt{v}] \right\}.$$

Similarly if  $v > u$  then the corresponding orbits are given by

$$\begin{aligned} \Gamma_{u,v}^+ &= \left\{ \left( \pm \sqrt{\frac{u-z^2}{2}}, \sqrt{\frac{v-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\}, \\ \Gamma_{u,v}^- &= \left\{ \left( \pm \sqrt{\frac{u-z^2}{2}}, -\sqrt{\frac{v-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\}. \end{aligned}$$

Whether  $u > v$  or  $v > u$  is enough information to localize a given orbit to one of two orbits on sphere of radius  $\sqrt{(u+v)}/2$ . The remaining piece of information is contained in the sign of the function defined by

$$(24) \quad \mathfrak{sn}(x, y, z) = \text{sign}(\mathbf{1}_{|x|>|y|}x + \mathbf{1}_{|y|>|x|}y)$$

The value of  $\mathfrak{sn}$  corresponds to the sign decorating the  $\Gamma_{u,v}^\pm$ . Hence if one starts from the initial condition  $(x, y, z)$  such that the  $(u, v)$  computed from these orbits satisfies  $u \neq v$  then the deterministic dynamics will trace the set  $\Gamma_{u,v}^{\mathfrak{sn}}$ .

The exception to being topologically equivalent to a circle are the lines of fixed points given by  $\{(0, 0, z) : z \in \mathbf{R}\}$ ,  $\{(x, 0, 0) : x \in \mathbf{R}\}$ , and  $\{(0, y, 0) : y \in \mathbf{R}\}$  and the heteroclinic orbits which connect them which are contained in the locus of points where  $u = v$ . For a given such choice there are four heteroclinic orbits given by

$$\begin{aligned} \mathcal{H}_u^{(1)} &= \left\{ \left( \sqrt{\frac{u-z^2}{2}}, \sqrt{\frac{u-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\} \\ \mathcal{H}_u^{(2)} &= \left\{ \left( \sqrt{\frac{u-z^2}{2}}, -\sqrt{\frac{u-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\} \\ \mathcal{H}_u^{(3)} &= \left\{ \left( -\sqrt{\frac{u-z^2}{2}}, -\sqrt{\frac{u-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\} \\ \mathcal{H}_u^{(4)} &= \left\{ \left( -\sqrt{\frac{u-z^2}{2}}, \sqrt{\frac{u-z^2}{2}}, z \right) : z \in [-\sqrt{u}, \sqrt{u}] \right\} \end{aligned}$$

These heteroclinic orbits split each sphere into four regions which contain closed orbits of finite period. The following set limits hold

$$\begin{aligned} \lim_{v \rightarrow u^-} \Gamma_{u,v}^+ &= \lim_{u \rightarrow v^+} \Gamma_{u,v}^+ = \mathcal{H}_u^{(1)} \cup \mathcal{H}_u^{(2)} \\ \lim_{v \rightarrow u^+} \Gamma_{u,v}^+ &= \lim_{u \rightarrow v^-} \Gamma_{u,v}^+ = \mathcal{H}_u^{(1)} \cup \mathcal{H}_u^{(4)} \\ \lim_{v \rightarrow u^-} \Gamma_{u,v}^- &= \lim_{u \rightarrow v^+} \Gamma_{u,v}^- = \mathcal{H}_u^{(3)} \cup \mathcal{H}_u^{(4)} \\ \lim_{v \rightarrow u^+} \Gamma_{u,v}^- &= \lim_{u \rightarrow v^-} \Gamma_{u,v}^- = \mathcal{H}_u^{(2)} \cup \mathcal{H}_u^{(3)} \end{aligned}$$

In contrast to the case when  $u \neq v$ , the orbits starting from a given point  $(x, y, z)$  do not converge to one of these unions of heteroclinic trajectories since any given orbit is restricted to a single heteroclinic trajectory. This could be a point of concern, but we will see in the next section that it does not pose a problem, which is an interesting and important feature of this model.

**7.2. Symmetries and their implications.** Defining  $\mathfrak{s}_e : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  by  $\mathfrak{s}_e(x, y, z) = (y, x, z)$  and  $\mathfrak{s}_\pm : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  by  $\mathfrak{s}_\pm(x, y, z) = (-x, -y, z)$ , observe that if  $\xi_t$  is a solution to (1) then so are  $\mathfrak{s}_e(\xi_t)$  and  $\mathfrak{s}_\pm(\xi)$ . This implies that  $\Gamma_{u,v}^- = \mathfrak{s}_\pm(\Gamma_{u,v}^+)$  and  $\Gamma_{u,v}^+ = \mathfrak{s}_e(\Gamma_{v,u}^+)$ , and that if  $\mu$  is an invariant probability measure for  $P_t$  then necessarily  $\mu \mathfrak{s}_e^{-1}$  and  $\mu \mathfrak{s}_\pm^{-1}$  are also invariant probability measures for  $P_t$ .

The situation for the stochastic dynamics given in (7) is the same for  $\mathfrak{s}_\pm$  but depends on the choice of  $\sigma_1$  and  $\sigma_2$  for  $\mathfrak{s}_e$ . In all cases  $\mathfrak{s}_\pm(\xi_t^\varepsilon)$  is a solution (for a different Brownian motion) if  $\xi_t^\varepsilon$  is a solution. However  $\mathfrak{s}_e(\xi_t^\varepsilon)$  is again a solution if  $\xi_t^\varepsilon$  is one only when  $\sigma_1 = \sigma_2$ . In any case, we have the following observation which we formulate as a proposition for future reference.

**Proposition 5.** *Let  $\mathfrak{s} : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  be a map such that  $\mathfrak{s}(\xi_t^\varepsilon)$  is a solution (for possibly a different Brownian motion) whenever  $\xi_t^\varepsilon$  is a solution, then  $\mu^\varepsilon = \mu^\varepsilon \mathfrak{s}^{-1}$  where  $\mu^\varepsilon$  is the unique invariant probability measure of  $P_t^\varepsilon$  guaranteed by Theorem 3.1.*

*Proof of Proposition 5.* As before it is clear that  $\mu^\varepsilon \mathfrak{s}^{-1}$  is again an invariant probability measure, however we know that  $\mu^\varepsilon$  is the unique invariant probability measure given the assumptions on the  $\sigma$ 's. Hence we conclude that  $\mu^\varepsilon \mathfrak{s}^{-1} = \mu$ .  $\square$

**7.3. Averaging along the deterministic trajectories.** Since the separation of time scale between the fast and slow dynamics leads to the averaging of the coefficients of  $(U^\varepsilon, V^\varepsilon)$  equation around the deterministic orbits we now discuss averaging along the deterministic orbits in general. After this we will define the function  $\Lambda$  whose asymptotics was described in Proposition 3.

Given a function  $\psi : \mathbf{R}^3 \rightarrow \mathbf{R}$ , we define

$$(25) \quad (\mathcal{A}\psi)(\xi) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t (\psi \circ \varphi_s)(\xi) ds.$$

Notice that  $\mathcal{A}\psi$  is again a function from  $\mathbf{R}^3 \rightarrow \mathbf{R}$  and that it is constant on the connected components of the level sets of  $(u, v)$ .

7.3.1. *Averaging when  $u \neq v$ .* Let  $(u, v) = \Phi(\xi)$ . If  $u \neq v$  then  $\xi$  lies on a periodic orbit of finite period. Letting  $\tau$  denote the period, one has

$$(\mathcal{A}\psi)(\xi) = \frac{1}{\tau} \int_0^\tau (\psi \circ \varphi_s)(\xi) ds$$

To obtain a more explicit representation for the averaging operation we will switch to an angular variable  $\theta$ . Given any positive  $u$  and  $v$ , for  $\theta \in [0, 2\pi]$  we parametrize  $z$  by  $z(\theta) = \sqrt{u \wedge v} \sin(\theta)$ . To define the other coordinates we introduce the following auxiliary angles

$$\phi_1(\theta) = \begin{cases} \arcsin(\sqrt{\frac{u}{u}} \sin \theta) & u > v \\ \theta & u \leq v \end{cases}, \quad \phi_2(\theta) = \begin{cases} \theta & u \geq v \\ \arcsin(\sqrt{\frac{u}{v}} \sin \theta) & u < v \end{cases}.$$

and set  $x(\theta) = \sqrt{\frac{u}{2}} \cos(\phi_1(\theta))$ , and  $y(\theta) = \sqrt{\frac{v}{2}} \cos(\phi_2(\theta))$ . Putting everything together we have that the trace of the trajectory starting at  $(\sqrt{\frac{u}{2}}, \sqrt{\frac{v}{2}}, 0)$  is given by

$$\Gamma_{u,v}^+ = \{\gamma_{u,v}(\theta) : \theta \in [0, 2\pi]\}$$

where  $\gamma_{u,v}(\theta) \stackrel{\text{def}}{=} (x(\theta), y(\theta), z(\theta))$ . As already discussed depending on whether  $u > v$  or  $v > u$  this represents a closed orbit on the sphere of radius  $\sqrt{(u+v)/2}$  which rotates around respectively either the  $x$ -axis in the positive  $x$  half space or the  $y$ -axis in the positive  $y$  half space. The orbits in the negative half space are given by  $\Gamma_{u,v}^- = \mathfrak{s}_\pm(\Gamma_{u,v}^+)$ .

To define the occupation measure on these orbits we define a third auxiliary angle

$$\phi_{u,v}(\theta) = \arcsin\left(\sqrt{\frac{u \wedge v}{u \vee v}} \sin \theta\right) = \begin{cases} \phi_1(\theta) & u > v \\ \theta & u = v \\ \phi_2(\theta) & u < v \end{cases}$$

For  $u \neq v$ , we define a probability measure on  $\mathbf{R}^3$  by

$$(26) \quad \nu_{u,v}^+(dx dy dz) = \int_0^{2\pi} \frac{K_{u,v}}{|\cos(\phi_{u,v}(\theta))|} \delta_{\gamma_{u,v}(\theta)}(dx dy dz) d\theta$$

where  $K_{u,v}^{-1} = \int_0^{2\pi} \frac{1}{|\cos(\phi_{u,v}(\theta))|} d\theta$ . We let  $\nu_{u,v}^-(dx dy dz) = \nu_{u,v}^+ \mathfrak{s}_\pm^{-1}$ . For  $u = v$ , we define  $\nu_{u,u}^\pm(dx dy dz) = \delta_{(0,0,\pm\sqrt{u})}(dx dy dz)$ . Each of these probability measures is

supported on the corresponding set  $\Gamma_{u,v}^+$  or  $\Gamma_{u,v}^-$ . It is straightforward to see that for any  $\psi : \mathbf{R}^3 \rightarrow \mathbf{R}$  and  $(x, y, z) \in \mathbf{R}^3$  such that  $|x| \neq |y|$  one has

$$(27) \quad (\mathcal{A}\psi)(\xi) = \int_{\mathbf{R}^3} \psi(\eta) \nu_{u,v}^{\mathfrak{s}}(d\eta)$$

where  $\xi = (x, y, z)$ ,  $(u, v) = \Phi(\xi)$  and  $\mathfrak{s} = \mathfrak{sn}(\xi)$  where  $\mathfrak{sn}$  was defined in (24).

**7.3.2. Definitions of  $\Gamma$  and  $\Lambda$ .** The central quantities which need to be averaged in the  $(U_t^\varepsilon, V_t^\varepsilon)$  dynamics, given in equation (12), are the infinitesimal quadratic variations. They are given respectively by  $16\sigma_1^2 x^2$  and  $16\sigma_1^2 y^2$ . From (4), we have that  $x^2 = \frac{1}{2}(u - z^2)$  and  $y^2 = \frac{1}{2}(v - z^2)$ . Since  $u$  and  $v$  are constant along the deterministic trajectories, this in turn implies that

$$\mathcal{A}(x^2) = \frac{u - \mathcal{A}(z^2)}{2} \quad \text{and} \quad \mathcal{A}(y^2) = \frac{v - \mathcal{A}(z^2)}{2}.$$

Since  $z^2$  does not depend on the choice of sign in the definition  $\nu_{u,v}^\pm(dx dy dz)$  by defining the single function

$$(28) \quad \Gamma(u, v) = \mathcal{A}(z^2) = K_{u,v} \int_0^{2\pi} \frac{(u \vee v) \sin^2(\theta)}{|\cos(\phi_{u,v}(\theta))|} d\theta$$

we have access to all of the averaged quantities we will require.

Clearly the function  $\Gamma(u, v)$  is symmetric in  $(u, v)$  and can be written as a function of  $u \vee v$  and  $u \wedge v$  only. In fact one sees that if one defines

$$\Lambda(r) = K_r \int_0^{2\pi} \frac{\sin^2(\theta)}{|\cos(\arcsin(r \sin(\theta)))|} d\theta, \quad \text{where}$$

$$K_r^{-1} = \int_0^{2\pi} \frac{1}{|\cos(\arcsin(r \sin(\theta)))|} d\theta$$

for  $r \in [0, 1]$  then (14) holds. The properties of  $\Lambda$  given in Proposition 3 follow directly from this definition, Proposition 6 in the next subsection, and its proof.

### 7.3.3. Averaging near the diagonal.

**Proposition 6.** *Let  $\psi : \mathbf{R}^3 \rightarrow \mathbf{R}$  be a continuous function. If  $\delta = 1 - (u \wedge v)/(v \vee u)$  then as  $|u - v| \rightarrow 0$  (and hence  $\delta \rightarrow 0$ ) while  $(u, v)$  remains in a compact set, one has*

$$(29) \quad (\nu\psi)(u, v, \sigma) = \frac{1}{2}(\psi(0, 0, \sqrt{u \vee v}) + \psi(0, 0, -\sqrt{u \vee v})) + o(1)$$

If in addition for all  $u$

$$C_u(\psi) = \int_0^{2\pi} \frac{\psi(\sqrt{u} \cos(\theta), \sqrt{u} \cos(\theta), \sqrt{u} \sin(\theta))}{|\cos(\theta)|} d\theta < \infty,$$

then as  $u \rightarrow v$  one has

$$(\nu\psi)(u, v, \sigma) = \frac{C_u(\psi)}{2|\ln(1 - r)|} + o(|\ln(1 - r)|^{-1}),$$

where  $r = \frac{u \wedge v}{v \vee u}$ .

**Remark 3.** The asymptotic expansion given in Proposition 3 follows from the fact that  $C_u(1 - z^2) = 4u$ . The continuity properties follow from the formulas and the fact that the values at the ends of the intervals are finite.

*Proof.* We will begin by exploring the asymptotics of the constant  $K_r$  which equals  $K_{u,v}$  when  $r = \frac{u \wedge v}{v \vee u}$ . Making the change of variables  $\alpha = \sin \theta$  followed by  $\beta^2 = \sqrt{r} \alpha^2$ , one has

$$\begin{aligned} K_r^{-1} &= \int_0^1 \frac{4}{\sqrt{(1-r\alpha^2)(1-\alpha^2)}} d\alpha = \frac{4}{r^{\frac{1}{4}}} \int_0^{r^{\frac{1}{4}}} \frac{1}{\sqrt{(1-r^{\frac{1}{2}}\beta^2)(1-r^{-\frac{1}{2}}\beta^2)}} d\beta \\ &= \frac{4}{r^{\frac{1}{4}}} \int_0^{r^{\frac{1}{4}}} \frac{1}{1-\beta^2} \frac{1}{\sqrt{1-\gamma(r)\frac{\beta^2}{(1-\beta^2)^2}}} d\beta, \end{aligned}$$

where  $\gamma(r) = r^{\frac{1}{2}} + r^{-\frac{1}{2}} - 2$ . Now since for all  $\beta, r \in (0, 1]$  with  $\beta \leq r$

$$0 \leq \gamma(r) \frac{\beta^2}{(1-\beta^2)^2} \leq \gamma(r) \frac{r^2}{(1-r^2)^2} \leq \frac{1}{16}$$

we have

$$(30) \quad \frac{4}{r^{\frac{1}{4}}} \operatorname{arctanh}(r^{\frac{1}{4}}) \leq K_r^{-1} \leq \frac{16}{\sqrt{15}} \frac{1}{r^{\frac{1}{4}}} \operatorname{arctanh}(r^{\frac{1}{4}}),$$

again for all  $r \in (0, 1]$ . Furthermore it is clear that

$$(31) \quad \lim_{r \rightarrow 1} K_r |\ln(1-r)| = \frac{1}{2}.$$

Now

$$(32) \quad (\nu\psi)(u, v, \sigma) = K_{u,v} \int_0^{2\pi} \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi(\theta))|} d\theta.$$

As  $|u-v| \rightarrow 0$  and hence  $r \rightarrow 1$ , this integral concentrates around the two points  $\theta$  equal  $\pi/2$  and  $3\pi/2$  since around these points  $|\cos(\phi(\theta))| \rightarrow 0$  as  $r \rightarrow 1$ . At these points  $(x(\theta), y(\theta), z(\theta))$  converges to  $(0, 0, \sqrt{u \vee v})$  and  $(0, 0, -\sqrt{u \vee v})$  respectively. Around these points we have one behavior and away from the another. Consider the following representative portion of the integral which will converge to  $\frac{1}{2}\psi(0, 0, \sqrt{u \vee v})$ . Fixing any sufficiency small  $\varepsilon > 0$ , we define  $a = a(\varepsilon)$  so that  $\sin(\pi/2 - a) = \sin(\pi/2 + a) = 1 - \varepsilon$ . Then

$$\begin{aligned} K_{u,v} \int_0^\pi \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi_{u,v}(\theta))|} d\theta &= K_{u,v} \int_0^{\frac{\pi}{2}-a} \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi_{u,v}(\theta))|} d\theta \\ &\quad + K_{u,v} \int_{\frac{\pi}{2}-a}^{\frac{\pi}{2}+a} \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi_{u,v}(\theta))|} d\theta + K_{u,v} \int_{\frac{\pi}{2}+a}^\pi \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi_{u,v}(\theta))|} d\theta. \end{aligned}$$

The remaining half of the integral in (32) will converge to  $\frac{1}{2}\psi(0, 0, -\sqrt{u \vee v})$  in a completely analogous fashion. The first and third integral behave the same. We consider the first. If, as before, we have  $r = \frac{u \wedge v}{v \vee u}$  and then make the change of variables  $\alpha = \sin \theta$  followed by  $\beta^2 = \sqrt{r} \alpha^2$  to obtain

$$\begin{aligned} K_{u,v} \left| \int_0^{\frac{\pi}{2}-a} \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi_{u,v}(\theta))|} d\theta \right| &\leq \frac{K_{u,v} \|\psi\|_\infty}{r^{\frac{1}{4}}} \int_0^{(1-\varepsilon)r^{\frac{1}{4}}} \frac{1}{1-\beta^2} \frac{1}{\sqrt{1-\gamma(r)\frac{\beta^2}{(1-\beta^2)^2}}} d\beta \\ &\leq CK_r \|\psi\|_\infty \operatorname{arctanh}((1-\varepsilon)r^{\frac{1}{4}}). \end{aligned}$$

By the asymptotics on  $K_r$  given in (30), this goes to zero since  $\varepsilon > 0$  as  $|u-v| \rightarrow 0$  and hence  $r \rightarrow 1$ .

Now as  $r \rightarrow 1$  one has

$$\begin{aligned} K_r \int_{\frac{\pi}{2}-a}^{\frac{\pi}{2}+a} \frac{\psi(x(\theta), y(\theta), z(\theta))}{|\cos(\phi(\theta))|} d\theta &\approx \psi(0, 0, \sqrt{u \vee v}) \frac{2K_r}{r^{\frac{1}{4}}} \int_{(1-\varepsilon)r^{\frac{1}{4}}}^{r^{\frac{1}{4}}} \frac{1}{1-\beta^2} d\beta \\ &\approx \frac{1}{2} \psi(0, 0, \sqrt{u \vee v}). \end{aligned}$$

The last conclusion follows directly from the assumed finiteness of  $C_u(\psi)$  and the asymptotics of  $K_r$  as  $r \rightarrow 1$  as  $|u - v| \rightarrow 0$ . in (31).  $\square$

**7.4. The Ergodic Invariant Measures.** The set of ergodic invariant probability measures is the set of extremal invariant probability measures. The extremal measures are those which can not be decomposed. Clearly this corresponds to the collection of the occupancy measures of each periodic orbit along with the delta measures sitting on each of the fixed points  $(0, 0, \sqrt{u})$  and  $(0, 0, -\sqrt{u})$ . These are precisely the measures  $\nu_{u,v}^{\pm}$  defined in Section 7.3.1. Since the union of these orbits and fixed points covers all the space except for the heteroclinic connections which cannot support an invariant probability measure. Hence we have identified all the ergodic probability measures.

We summarize this discussion in the following result.

**Proposition 7.** *The set of ergodic invariant probability measure of (1) consists precisely of*

$$\{\nu_{u,v}^+, \nu_{u,v}^- : u, v > 0\}.$$

Given  $(u, v) \in \mathbf{R}_+^2$ , we define the probability measure  $\nu_{u,v}$  on  $\mathbf{R}^3$  by

$$(33) \quad \nu_{u,v}(d\xi) = \frac{1}{2} \nu_{u,v}^+(d\xi) + \frac{1}{2} \nu_{u,v}^-(d\xi)$$

where  $\nu_{u,v}^{\pm}(d\xi)$  we defined in (26) and the text below it.

The following corollary of Proposition 7 will be central to the proof of the convergence of  $\mu^\varepsilon$  to a unique limiting measure.

**Corollary 3.** *Any invariant probability measure  $m$  for (1) which satisfies  $m\mathfrak{s}_{\pm}^{-1} = m$  can be represented as*

$$m(dx dy dz) = \int_{[0,\infty)^2} \nu_{u,v}(dx dy dz) \gamma(du dv)$$

for some probability measure  $\gamma$  on  $(0, \infty)^2$ . Furthermore the measure  $\gamma$  is unique. Conversely, a probability measure which is invariant for (1) and satisfies  $m\mathfrak{s}_{\pm}^{-1} = m$  is uniquely specified by the measure  $\gamma = m\Phi^{-1}$ .

*Proof of Corollary 3.* The ergodic decomposition theorem [CFS82] implies that there exists a unique pair of measures  $(\gamma^+, \gamma^-)$  so that the total mass of  $\gamma^+ + \gamma^-$  is one and

$$\begin{aligned} m(dx dy dz) &= \int_{[0,\infty)^2} \nu_{u,v}^+(dx dy dz) \gamma^+(du dv) \\ &\quad + \int_{[0,\infty)^2} \nu_{u,v}^-(dx dy dz) \gamma^-(du dv). \end{aligned}$$



Now since  $m\mathfrak{s}_{\pm}^{-1} = m$ ,  $\nu_{u,v}^- = \nu_{u,v}^+\mathfrak{s}_{\pm}^{-1}$  and  $\nu_{u,v}^+ = \nu_{u,v}^-\mathfrak{s}_{\pm}^{-1}$ , we have that

$$m(dx dy dz) = \int_{[0,\infty)^2} \nu_{u,v}^-(dx dy dz) \gamma^+(du dv) + \int_{[0,\infty)^2} \nu_{u,v}^+(dx dy dz) \gamma^-(du dv).$$

Since  $\nu_{u,v}^-$  and  $\nu_{a,b}^+$  are mutually singular for all choices of positive  $u, v, a$ , and  $b$ , we see that  $\gamma^+ = \gamma^-$  and the total mass of both is  $\frac{1}{2}$ . Setting  $\gamma = 2\gamma^+ = 2\gamma^-$  we see that  $\gamma$  is a probability measure and that

$$\begin{aligned} m(dx dy dz) &= \int_{[0,\infty)^2} \left[ \frac{1}{2}\nu_{u,v}^+(dx dy dz) + \frac{1}{2}\nu_{u,v}^-(dx dy dz) \right] \gamma(du dv) \\ &= \int_{[0,\infty)^2} \nu_{u,v}(dx dy dz) \gamma(du dv) \end{aligned}$$

This proves that any invariant  $m$  satisfying the symmetry assumption can be represented as claimed. All that remains is to show is that  $\gamma$  is unique. Let  $\tilde{\gamma}$  be another probability measure so that

$$m(dx dy dz) = \int_{[0,\infty)^2} \nu_{u,v}(dx dy dz) \tilde{\gamma}(du dv)$$

which implies that

$$m(dx dy dz) = \int_{[0,\infty)^2} \left[ \frac{1}{2}\nu_{u,v}^+(dx dy dz) + \frac{1}{2}\nu_{u,v}^-(dx dy dz) \right] \tilde{\gamma}(du dv)$$

which in turn implies that  $\frac{1}{2}\tilde{\gamma} = \gamma^+$  since the ergodic decomposition is unique. However, this implies  $\tilde{\gamma} = \gamma$  as desired.  $\square$

**7.5. The Limiting Fast Semigroup.** We begin with a small detour to think about the limiting dynamics. Its action on a test function can be understood to instantly assign to each point on an orbit the average of the function around the orbit and to each point on the heteroclinic connection the value of the function at the limiting fixed point on the  $z$ -axis.

Recall the definition of  $\nu_{u,v}$  from (33), for  $\phi: \mathbf{R}^3 \rightarrow \mathbf{R}$  we define  $\nu\phi$  by

$$(\nu\phi)(u, v) = \int \phi(\xi)\nu_{u,v}(d\xi).$$

Recalling the definition of  $\Phi$  which maps  $\xi$  to  $(u, v)$  from (4), we note that for any  $\rho: \mathbf{R}_+^2 \rightarrow \mathbf{R}$ ,

$$(34) \quad \nu(\rho \circ \Phi)(u, v) = \rho(u, v).$$

Lastly recalling the definition of  $\tilde{P}_t^\varepsilon$  from (11),  $Q_t$  from (20) and let  $\lambda$  be the unique invariant probability measure of  $Q_t$  guaranteed by Theorem 6.12. For  $\phi: \mathbf{R}^3 \rightarrow \mathbf{R}$  we define

$$(35) \quad (\tilde{P}_t\phi)(\xi) = (Q_t\nu\phi) \circ \Phi(\xi)$$

**Remark 4.** If  $\phi$  is a test function such that  $\phi \circ \mathfrak{s}_{\pm} = \phi$  or  $m$  is an initial measure on  $\mathbf{R}^3$  such that  $m = m\mathfrak{s}_{\pm}^{-1}$  then is not hard to convince oneself that  $m\tilde{P}_t^\varepsilon\phi \rightarrow m\tilde{P}_t\phi$  as  $\varepsilon \rightarrow 0$ . If one neither starts with initial data which has this symmetry nor uses a symmetric test function, then things are more complicated. The orbit may average with respect to only one of the two measure:  $\nu_{u,v}^+$  or  $\nu_{u,v}^-$ . For definiteness

assume that we are on the  $\nu_{u,v}^+$  orbit. We believe that when the  $(U, V)$ -dynamics hits the line  $U = V$  then it is essentially spending all of its time at  $(0, 0, \sqrt{u})$  and  $(0, 0, -\sqrt{u})$ . (See Proposition 6.) With probability  $\frac{1}{2}$  it returns to a  $\nu_{u,v}^+$  orbit and with probability  $\frac{1}{2}$  it enters on to a  $\nu_{u,v}^-$  orbit. Hence to describe the  $\tilde{P}_t$  semigroup in the non-symmetric setting, it seems we need to add a sequence of independent Bernoulli random variables to make decision of whether one should average with respect to the  $+$  or the  $-$  orbit. Since we are primarily interested in the structure of the invariant probability measure we have not tried to make this picture rigorous.

Let  $\lambda$  be the unique invariant probability measure of  $Q_t$  and define  $\mu = \lambda\nu$ . Observe that  $\mu$  is invariant under  $\tilde{P}_t$  because for any bounded  $\phi: \mathbf{R}^3 \rightarrow \mathbf{R}$  one has

$$\mu\tilde{P}_t\phi = \lambda\nu(Q_t\nu\phi \circ \Phi) = \lambda Q_t\nu\phi = \lambda(\nu\phi) = \mu\phi.$$

Here the first equality is by definition, the second follows from (34), the third from the invariance of  $\lambda$  under  $Q_t$  and the last from the definition of  $\mu$ .

## 8. CONVERGENCE OF $(U_\varepsilon V_\varepsilon)$ TOWARDS $(U, V)$

We now prove the results which were taken for granted in Section 6, namely that  $\{(U_t^\varepsilon, V_t^\varepsilon)\}_{\varepsilon>0}$  is tight, and that any accumulation point  $(U_t, V_t)$  solves the SDE (13).

**8.1. Tightness.** Let us rewrite (12) in the form

$$(36) \quad \begin{cases} dU_t^\varepsilon = (C_u - 2U_t^\varepsilon)dt + dM_t^\varepsilon \\ dV_t^\varepsilon = (C_v - 2V_t^\varepsilon)dt + dN_t^\varepsilon, \end{cases}$$

where  $\{M_t^\varepsilon, t \geq 0\}$  and  $\{N_t^\varepsilon, t \geq 0\}$  are continuous local martingales such that

$$(37) \quad \frac{d}{dt}\langle M^\varepsilon \rangle_t \leq C U_t^\varepsilon, \quad \frac{d}{dt}\langle N^\varepsilon \rangle_t \leq C V_t^\varepsilon,$$

where  $C_u, C_v$  and  $C$  are three positive constants.

We want to show

**Proposition 8.** *Suppose that*

$$\sup_{\varepsilon>0} \mathbf{E} [(U_0^\varepsilon)^2 + (V_0^\varepsilon)^2] < \infty.$$

*Then the collection of processes  $\{(U_t^\varepsilon, V_t^\varepsilon), t \geq 0\}_{\varepsilon>0}$  is tight in  $C([0, +\infty); \mathbf{R}^2)$ .*

In light of (36) and (37), the following needed first step follows from Lemma 5.1.

**Lemma 8.1.** *Under the condition of Proposition 8,*

$$\sup_{\varepsilon>0} \sup_{t \geq 0} \mathbf{E} [(U_t^\varepsilon)^2 + (V_t^\varepsilon)^2] < \infty.$$

We can now proceed with the proof of tightness.

*Proof of Proposition 8.* We prove tightness of  $U^\varepsilon$  only,  $V^\varepsilon$  being treated completely similarly. We have

$$U_t^\varepsilon = \frac{C_u}{2} + \left( U_0^\varepsilon - \frac{C_u}{2} \right) e^{-2t} + e^{-2t} \int_0^t e^{2s} dM_s^\varepsilon.$$

Clearly the first two terms on the right are tight in  $C([0, \infty))$ , since the collection of  $\mathbf{R}$ -valued r.v.'s  $U_0^\varepsilon$  is tight. We only need check tightness in  $C([0, \infty))$  of the process  $W_t^\varepsilon := \int_0^t e^{2s} dM_s^\varepsilon$ . Since  $W_0^\varepsilon = 0$ , we need only verify condition (ii) from

Theorem 7.3 in Billingsley [Bil99], which follows from the condition of the Corollary of Theorem 7.4 again in [Bil99]. In other words it suffices to check that for any  $T$ ,  $\eta$  and  $\eta' > 0$ , there exists  $\delta \in (0, 1)$  such that for all  $\varepsilon > 0$ ,  $0 \leq t \leq T - \delta$ ,

$$(38) \quad \frac{1}{\delta} \mathbf{P} \left( \sup_{t \leq s \leq t+\delta} |W_s^\varepsilon - W_t^\varepsilon| \geq \eta \right) \leq \eta'.$$

Combining Chebycheff and Burkholder–Davis–Gundy inequalities, we deduce that (we use below the result from Lemma 8.1)

$$\begin{aligned} \mathbf{P} \left( \sup_{t \leq s \leq t+\delta} |W_s^\varepsilon - W_t^\varepsilon| \geq \eta \right) &\leq \eta^{-4} \mathbf{E} \left( |\langle W^\varepsilon \rangle_{t+\delta} - \langle W^\varepsilon \rangle_t|^2 \right) \\ &\leq \eta^{-4} e^{8T} C^2 \delta \int_t^{t+\delta} \mathbf{E}[(U_s^\varepsilon + V_s^\varepsilon)^2] ds \\ &\leq \eta^{-4} \bar{C} e^{8T} \delta^2, \end{aligned}$$

from which (38) follows if we choose  $\delta = e^{-4T} \eta^2 \sqrt{\eta'/\bar{C}}$ .  $\square$

**8.2. Tightness of  $\lambda^\varepsilon$ .** Since  $(U_t^\varepsilon, V_t^\varepsilon)$  is not a Markov process it does not have an invariant probability measure. However the projection  $\lambda^\varepsilon = \mu^\varepsilon \Phi^{-1}$  of  $\mu^\varepsilon$ , the unique invariant probability measure of the Markov process  $\xi_t^\varepsilon$ , is well defined. We now establish the following tightness result:

**Lemma 8.2.** *The sequence of measure  $\{\lambda^\varepsilon : \varepsilon > 0\}$  is tight on the space  $(0, \infty) \times (0, \infty)$ .*

**Remark 5.** We emphasize that Lemma 8.2 is tightness in the open set  $(0, \infty) \times (0, \infty)$  which implies the measure does not accumulate neither at the boundary at “infinity” nor at the boundary at zero. In other words, for any  $\delta > 0$  there exists a  $r > 0$  so that

$$\inf_{\varepsilon > 0} \lambda^\varepsilon \left( \left[ \frac{1}{r}, r \right] \times \left[ \frac{1}{r}, r \right] \right) > 1 - \delta.$$

The following result which implies the tightness at infinity follows immediately from the definition of  $\lambda^\varepsilon$ , the definition of  $\Phi$  and Corollary 2.

**Lemma 8.3.** *For any  $p \geq 1$ , there exists a  $C(p) > 0$  so that*

$$\sup_{\varepsilon > 0} \int (u^p + v^p) \lambda^\varepsilon(du, dv) < C(p)$$

We now handle the boundary at zero.

**Lemma 8.4.** *Let  $\zeta_t$  be a Markov process and  $f$  and  $g$  two real-valued functions on the state space of  $\zeta_t$  satisfying  $0 \leq g(\zeta_t) \leq f(\zeta_t)$  for all  $t \geq 0$  almost surely and such that  $f(\zeta_t)$  is a continuous semimartingale satisfying*

$$df(\zeta_t) = (a - f(\zeta_t))dt + c\sqrt{g(\zeta_t)}dW_t$$

where  $a$  and  $c$  are positive constants and  $W_t$  a standard Wiener process. If  $\mu$  is any invariant probability measure of  $\zeta_t$  with  $\mu[f^2] = \int f^2(\zeta)d\zeta < \infty$ , then for any  $\delta \in (0, 1)$ .

$$(39) \quad \mu\{f \leq \delta\} \leq \frac{\mu[f^2] + b}{a|\log \delta|},$$

with  $b = 1 + c^2/2$ .

*Proof of Lemma 8.4.* Defining

$$\phi(x) = \begin{cases} -\frac{1}{x}, & x \leq 1 \\ 1, & x \geq 1 \end{cases}, \quad I(x) = \begin{cases} 1 - \log x, & x \leq 1 \\ x, & x \geq 1 \end{cases},$$

$$H(x) = \begin{cases} 2x - x \log x, & x \leq 1 \\ \frac{1}{2}x^2 + \frac{3}{2}, & x \geq 1 \end{cases}.$$

Observe that  $H'(x) = I(x)$  and  $I'(x) = \phi(x)$  and that  $\phi$ ,  $I$  and  $H$  are well defined on the intervals  $(0, \infty)$ ,  $(0, \infty)$  and  $[0, \infty)$  respectively.  $H$  and  $I$  are everywhere positive, while  $\phi$  is positive on  $[1, +\infty)$  and negative on  $(0, 1)$ . It is plain that the discontinuity of  $H''$  at  $x = 1$  will not prevent us from using Itô's formula. Taking  $\zeta_0$  distributed according to  $\mu$ , noticing that since  $H(x) < 2 + x^2$  for  $x \geq 0$ , and setting  $X_t = f(\zeta_t)$  for notational convenience, we have that

$$(40) \quad \mu[\mathbf{E}_{\zeta_0} H(X_t)] = \mu[\mathbf{E}_{\zeta_0} (H \circ f)(\zeta_t)] = \mu[H \circ f] < \infty.$$

Now from Itô's formula

$$dH(X_t) = (a - X_t)I(X_t)dt + \frac{1}{2}c^2g(\zeta_t)\phi(X_t)dt + dM_t$$

where  $M_t$  is the Martingale defined by  $dM_t = c\sqrt{g(\zeta_t)}I(X_t)dW_t$ . We conclude that

$$a \int_0^t \mathbf{E}_{\zeta_0}[I(X_s)] ds \leq \mathbf{E}_{\zeta_0} H(X_t) - H(X_0) + \int_0^t \mathbf{E}_{\zeta_0} X_s I(X_s) ds \\ + \frac{c^2}{2} \int_0^t \mathbf{E}_{\zeta_0} g(\zeta_s) \phi^-(X_s) ds.$$

Now integrating over the initial conditions  $\zeta_0$  (which were distributed according to  $\mu$ ), we see that  $H$  terms are equal by the stationarity embodied in (40) (and hence they cancel) and that

$$a \mu[I \circ f] \leq \mu[f(I \circ f)] + \frac{c^2}{2} \mu[g(\phi^- \circ f)]$$

and since  $g(\phi^- \circ f) \leq 1$ ,

$$a \mu[|\log f| \mathbf{1}\{f \leq 1\}] \leq a \mu[I \circ f] \leq \mu[f(I \circ f)] + \frac{c^2}{2}$$

Finally, for any  $\delta \in (0, 1)$

$$a \mu\{f \leq \delta\} = a \mu[\{|\log f| \geq |\log \delta|\} \cap \{f \leq 1\}] \\ \leq \frac{a \mu[|\log f| \mathbf{1}\{f \leq 1\}]}{|\log \delta|} \leq \frac{\mu[f(I \circ f)] + \frac{c^2}{2}}{|\log \delta|}$$

The result follows, since  $xI(x) \leq 1 + x^2$ . □

The following Corollary is a direct consequence of the two last Lemmata

**Corollary 4.** *There exists a constant  $C > 0$  so that for any  $\delta \in (0, 1)$*

$$\sup_{\varepsilon > 0} \lambda^\varepsilon\{(u, v) : u + v < \delta\} \leq \frac{C}{|\log \delta|}$$

*Proof of Lemma 8.2.* The result follows immediately by combining Lemma 8.3 and Corollary 4. □

**8.3. Convergence of Quadratic variation.** Now that we know that the collection  $\{(U_t^\varepsilon, V_t^\varepsilon), t \geq 0\}_{\varepsilon > 0}$  is tight, in view of Theorem 6.3, the weak uniqueness result for (13), and comparing (12) and (13), the weak convergence  $(U^\varepsilon, V^\varepsilon) \Rightarrow (U, V)$  will follow from the convergence of the quadratic variations of  $U^\varepsilon$  and  $V^\varepsilon$  to those of  $U$  and  $V$ , which will be proved in the next Lemma.

For each  $M > 0$ , let

$$\kappa_M^\varepsilon := \inf\{t > 0, U_t^\varepsilon \vee V_t^\varepsilon > M\}.$$

Considering the three different cases of the behavior of  $(U, V)$ , it is not hard to see that in all cases  $\kappa_M$ , defined exactly as  $\kappa_M^\varepsilon$ , but with  $(U^\varepsilon, V^\varepsilon)$  replaced by  $(U, V)$ , is a.s. a continuous function of the  $(U, V)$  trajectory, hence

$$\kappa_M^\varepsilon \Longrightarrow \kappa_M \quad \text{as } \varepsilon \rightarrow 0$$

will follow from  $(U^\varepsilon, V^\varepsilon) \Rightarrow (U, V)$ .

In particular

$$\liminf_{\varepsilon \rightarrow 0} \mathbf{P}(\kappa_M^\varepsilon > t) \geq \mathbf{P}(\kappa_M > t).$$

Clearly for all  $t > 0$ ,

$$\mathbf{P}(\kappa_M > t) \rightarrow 1, \quad \text{as } M \rightarrow \infty.$$

It will then follow that for any  $t > 0$ , the  $\liminf$  as  $\varepsilon \rightarrow 0$  of  $\mathbf{P}(\kappa_M^\varepsilon > t)$  can be made arbitrarily close to 1, by choosing  $M$  large enough.

**Lemma 8.5.** *Let  $\nu_\varepsilon$  be any sequence of tight probability measures on  $\mathbf{R}^3$  and let  $(\tilde{X}_t^\varepsilon, \tilde{Y}_t^\varepsilon, \tilde{Z}_t^\varepsilon)$  be the solution to (10) with  $(\tilde{X}_0^\varepsilon, \tilde{Y}_0^\varepsilon, \tilde{Z}_0^\varepsilon)$  distributed as  $\nu_\varepsilon$ . Then for any  $t > 0$ , as  $\varepsilon \rightarrow 0$ ,*

$$\int_0^t (\tilde{X}_s^\varepsilon)^2 ds \Rightarrow \int_0^t \mathcal{A}(x^2)(U_s, V_s) ds \quad \text{and} \quad \int_0^t (\tilde{Y}_s^\varepsilon)^2 ds \Rightarrow \int_0^t \mathcal{A}(y^2)(U_s, V_s) ds.$$

*Proof.* Since  $2(\tilde{X}_s^\varepsilon)^2 = U_s^\varepsilon - (\tilde{Z}_s^\varepsilon)^2$  and  $2(\tilde{Y}_s^\varepsilon)^2 = V_s^\varepsilon - (\tilde{Z}_s^\varepsilon)^2$ , we only need to show that  $\int_0^t (\tilde{Z}_s^\varepsilon)^2 ds \Rightarrow \int_0^t \mathcal{A}(z^2)(U_s, V_s) ds$ . It suffices in fact to show that

$$\int_0^{t \wedge \kappa_M^\varepsilon} (\tilde{Z}_s^\varepsilon)^2 ds \Longrightarrow \int_0^{t \wedge \kappa_M} \mathcal{A}(z^2)(U_s, V_s) ds,$$

for all  $M > 0$ .

The values of  $t > 0$  and  $M$  will be fixed throughout this proof. For any  $\delta > 0$ , we define  $N_\delta = \lceil t/\delta \rceil$ ,  $t_n = n\delta \wedge \kappa_M^\varepsilon$  for  $0 \leq n < N_\delta$  and  $t_{N_\delta} = t \wedge \kappa_M^\varepsilon$ . Let now  $Z_s^{(n)}$  be the  $z$  component of the solution to the deterministic dynamics (1) at time  $s$  which started at time  $t_n$  from the point  $(X_{t_n}^\varepsilon, Y_{t_n}^\varepsilon, Z_{t_n}^\varepsilon)$ . Then clearly

$$(41) \quad \int_0^{t \wedge \kappa_M^\varepsilon} (\tilde{Z}_s^\varepsilon)^2 ds = \varepsilon \sum_{n=0}^{N_\delta-1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^\varepsilon)^2 ds = \Phi_{\varepsilon, \delta} + \Xi_{\varepsilon, \delta}$$

where

$$\begin{aligned} \Phi_{\varepsilon, \delta} &= \delta \sum_{n=0}^{N_\delta-1} \frac{\varepsilon}{\delta} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^{(n)})^2 ds \\ \Xi_{\varepsilon, \delta} &= \varepsilon \sum_{n=0}^{N_\delta-1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} \left[ (Z_s^\varepsilon)^2 - (Z_s^{(n)})^2 \right] ds. \end{aligned}$$

To control the error term observe that

$$|\Xi_{\varepsilon, \delta}| \leq \sqrt{\varepsilon \sum_{n=0}^{N_\delta-1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} [Z_s^\varepsilon + Z_s^{(n)}]^2 ds} \sqrt{\varepsilon \sum_{n=0}^{N_\delta-1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} [Z_s^\varepsilon - Z_s^{(n)}]^2 ds}.$$

The first term in the product on the righthand side is bounded due to the stopping time  $\kappa_M^\varepsilon$ . Using Lemma 4.1, we see that  $\mathbf{E}|\Xi_{\varepsilon, \delta}|$  is bounded by a constant times the square root of

$$\varepsilon \sum_{n=0}^{N_\delta-1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} \mathbf{E} [Z_s^\varepsilon - Z_s^{(n)}]^2 ds \leq C_M \varepsilon^2 \lceil \frac{t}{\delta} \rceil \exp \left[ C_M \frac{\delta}{\varepsilon} \right].$$

Hence if we choose

$$(42) \quad \delta = C_M^{-1} \varepsilon \log(1/\varepsilon),$$

then  $\Xi_{\varepsilon, \delta} \rightarrow 0$  in  $L^1(\Omega)$  as  $\varepsilon \rightarrow 0$ . Having made this choice of  $\delta$ , we now suppress it from notation designating dependence on parameters.

We now further divide  $\Phi_\varepsilon$  ( $\delta$  having been suppressed) depending on whether in phase space the starting point  $(X_{t_n/\varepsilon}^{(n)}, Y_{t_n/\varepsilon}^{(n)}, Z_{t_n/\varepsilon}^{(n)})$  lies the region where  $|U_{t_n}^\varepsilon - V_{t_n}^\varepsilon|$  is small or not. To accomplish this, for any  $\rho > 0$ , let  $\chi_\rho \in C(\mathbf{R}; [0, 1])$  be such that

$$\chi_\rho(x) = \begin{cases} 0 & , \text{ if } |x| \geq \rho, \\ 1 & , \text{ if } |x| \leq \rho/2 \end{cases}$$

and define  $\bar{\chi}_\rho = 1 - \chi_\rho$ . Consider the decomposition  $\Phi_\varepsilon = A_{\varepsilon, \rho} + B_{\varepsilon, \rho}$  where

$$A_{\varepsilon, \rho} = \delta \sum_{n=0}^{N_\delta-1} \frac{\varepsilon}{\delta} \chi_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^{(n)})^2 ds,$$

$$B_{\varepsilon, \rho} = \delta \sum_{n=0}^{N_\delta-1} \frac{\varepsilon}{\delta} \bar{\chi}_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^{(n)})^2 ds.$$

The reason for this decomposition is that why the time average of  $(Z^n)^2$  over the time interval  $[t_{n+1}, t_n]$  is close to the function  $\mathcal{A}(z^2)(U_{t_n}, V_{t_n})$  is different in the two regions. The terms which have  $|u - v| > \rho$  have periods uniformly bounded from above and hence as  $(t_{n+1} - t_n)/\varepsilon = \delta/\varepsilon \rightarrow \infty$  the number of periods contained in the interval over which we are averaging also goes to infinity. On the other hand, as the points approach the diagonal  $u = v$  the period grows to infinity. So for  $|u - v|$  small enough the period might be much greater than the length of the time interval  $(t_{n+1} - t_n)/\varepsilon = \delta/\varepsilon$  over which we are averaging. Hence the reason for convergence for the  $A_{\varepsilon, \rho}$  to the appropriate average values occurs by a different mechanism. Proposition 6 shows that  $\mathcal{A}(z^2)(u, v) \rightarrow u = v$  as  $|u - v| \rightarrow 0$ . To understand why  $\frac{\varepsilon}{\delta} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^{(n)})^2 ds \rightarrow U_{t_n} \wedge V_{t_n}$  one needs to recall the discussion from Section 7. The deterministic orbits when  $u = v$  consist of heteroclinic orbits connecting the fixed points at  $(0, 0, \sqrt{u})$  and  $(0, 0, -\sqrt{u})$ . Since the time to reach the fixed points on these orbits is infinite, it is not surprising that for  $|u - v|$  small the periodic orbit spends most of its time near  $(0, 0, \pm\sqrt{u}) \sim (0, 0, \pm\sqrt{v})$ . This can also be seen in the fact that the occupation measures given in (26) concentrates around  $\theta \sim \pi/2, 3\pi/2$ , which corresponds to the fixed points, if  $|u - v| \sim 0$ . Importantly, even when the time is not long enough to traverse the orbit completely, any average will be concentrated near the fixed points since the time to reach the neighborhood

of the fixed point is small relative to the time it will take to leave that neighborhood once it has arrived there. This idea will be made quantitative below.

Hence we define

$$\begin{aligned}\hat{A}_{\varepsilon,\rho} &= \delta \sum_{n=0}^{N_\delta-1} \chi_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) (U_{t_n}^\varepsilon \wedge V_{t_n}^\varepsilon) \quad \text{and} \quad \Theta_{\varepsilon,\rho} = A_{\varepsilon,\rho} - \hat{A}_{\varepsilon,\rho} \\ \hat{B}_{\varepsilon,\rho} &= \delta \sum_{n=0}^{N_\delta-1} \bar{\chi}_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) \gamma_\varepsilon(U_{t_n}^\varepsilon, V_{t_n}^\varepsilon) \mathcal{A}(z^2)(U_{t_n}^\varepsilon, V_{t_n}^\varepsilon) \quad \text{and} \quad \Upsilon_{\varepsilon,\rho} = B_{\varepsilon,\rho} - \hat{B}_{\varepsilon,\rho}\end{aligned}$$

where

$$\gamma_\varepsilon(u, v) = \left\lfloor \frac{\delta}{\varepsilon \tau(u, v)} \right\rfloor \frac{\varepsilon \tau(u, v)}{\delta},$$

and  $\tau(u, v)$  was the period of the deterministic orbit.

For all  $\beta > 0$  and  $\alpha > \rho$ , we define

$$\begin{aligned}\hat{\tau}_{\beta,\rho} &= \min \left\{ \tau(u, v) : |u - v| \leq \rho, \beta \leq u \wedge v \leq u \vee v \leq M \right\}, \\ \Psi_{\alpha,\rho} &= 4\sqrt{M} \int_0^{1-\alpha} \frac{1}{|1-a^2|} da.\end{aligned}$$

The utility of  $\Psi$  is the following which can be deduced from Section 7.3

$$\sup_{u, v \leq M} \text{Leb}(\{s > 0, |z_s| \notin [(1-\alpha)\sqrt{u \wedge v}, \sqrt{u \wedge v}]\}) \leq \Psi_{\alpha,\rho}.$$

Now

$$|\Theta_{\varepsilon,\rho}| \leq t \left[ M \left( \frac{\Psi_{\alpha,\rho}}{\delta/\varepsilon} + \frac{\Psi_{\alpha,\rho}}{\hat{\tau}_{\beta,\rho}} \right) + 2\beta \right] \stackrel{\text{def}}{=} K_{\varepsilon,\rho,\alpha,\beta}.$$

On the other hand, since

$$\Upsilon_{\varepsilon,\rho} = \varepsilon \sum_{n=0}^{N_\delta-1} \bar{\chi}_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) \int_{\frac{t_n}{\varepsilon} + \lfloor \frac{\delta}{\varepsilon \tau_n^\varepsilon} \rfloor \tau_n^\varepsilon}^{\frac{t_{n+1}}{\varepsilon}} (Z_r^{(n)})^2 dr,$$

we have the inequality

$$\begin{aligned}|\Upsilon_{\varepsilon,\rho}| &\leq M \frac{\varepsilon}{\delta} \delta \sum_{n=0}^{N_\delta-1} \bar{\chi}_\rho(U_{t_n}^\varepsilon - V_{t_n}^\varepsilon) \tau(U_{t_n}^\varepsilon, V_{t_n}^\varepsilon) \\ &\leq \frac{\varepsilon}{\delta} \bar{M}(\rho) \stackrel{\text{def}}{=} L_{\varepsilon,\rho},\end{aligned}$$

where  $\bar{M}(\rho) := \sup_{u, v \leq M} \bar{\chi}_\rho(u, v) \tau(u, v) < \infty$  if  $\rho > 0$ . Hence  $L_{\varepsilon,\rho} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , for any  $\rho > 0$ .

Now applying Lemma 8.6 below, we see that for all  $\rho > 0$ , as  $\varepsilon \rightarrow 0$  one has

$$\begin{aligned}\hat{A}_{\varepsilon,\rho} + \hat{B}_{\varepsilon,\rho} &\implies \hat{A}_\rho + \hat{B}_\rho \stackrel{\text{def}}{=} \int_0^{t \wedge \kappa_M} \chi_\rho(U_s - V_s) (U_s \wedge V_s) ds \\ &\quad + \int_0^{t \wedge \kappa_M} \bar{\chi}_\rho(U_s - V_s) \mathcal{A}(z^2)(U_s, V_s) ds\end{aligned}$$

Notice that as  $\rho \rightarrow 0$ ,  $\hat{A}_\rho + \hat{B}_\rho$  converges to

$$\int_0^{t \wedge \kappa_M} \mathbf{1}_{U_s=V_s} (U_s \wedge V_s) ds + \int_0^{t \wedge \kappa_M} \mathbf{1}_{U_s \neq V_s} \mathcal{A}(z^2)(U_s, V_s) ds.$$

By Proposition 6, we see that  $\mathbf{1}_{U_s=V_s}(U_s \wedge V_s) = \mathbf{1}_{U_s=V_s} \mathcal{A}(z^2)(U_s, V_s)$  since  $(x, y, z) \mapsto z^2$  evaluated at  $(0, 0, \sqrt{u \wedge v})$  is  $u \wedge v$ . (Of course  $u \wedge v = u = v$  since we are considering the case  $u = v$ .) In light of this, we conclude that  $\hat{A}_\rho + \hat{B}_\rho$  converges to

$$\int_0^{t \wedge \kappa_M} \mathcal{A}(z^2)(U_s, V_s) ds$$

as  $\rho \rightarrow 0$ .

Now let  $F \in C(\mathbf{R}^+, \mathbf{R}^+)$  be any increasing, bounded function. Then

$$\begin{aligned} F(\hat{A}_{\varepsilon, \rho} + \hat{B}_{\varepsilon, \rho} - K_{\varepsilon, \rho, \alpha, \beta} - L_{\varepsilon, \delta}) &\leq F(A_{\varepsilon, \rho} + B_{\varepsilon, \rho}) \leq \\ &F(\hat{A}_{\varepsilon, \rho} + \hat{B}_{\varepsilon, \rho} + K_{\varepsilon, \rho, \alpha, \beta} + L_{\varepsilon, \delta}). \end{aligned}$$

Observe that  $A_{\varepsilon, \rho} + B_{\varepsilon, \rho} = \Phi_\varepsilon$  and hence is independent of the choice of  $\rho$ . Since  $F$  is bounded and as already noted  $L_{\varepsilon, \rho} \rightarrow 0$  as  $\varepsilon \rightarrow 0$  for any  $\rho > 0$ , we have that

$$\begin{aligned} \mathbf{E}F(\hat{A}_\rho + \hat{B}_\rho - K_{\rho, \alpha, \beta}) &\leq \varliminf_{\varepsilon \rightarrow 0} \mathbf{E}F(\Phi_\varepsilon) \\ &\leq \varlimsup_{\varepsilon \rightarrow 0} \mathbf{E}F(\Phi_\varepsilon) \leq \mathbf{E}F(\hat{A}_\rho + \hat{B}_\rho + K_{\rho, \alpha, \beta}), \end{aligned}$$

where

$$K_{\rho, \alpha, \beta} = \lim_{\varepsilon \rightarrow 0} K_{\varepsilon, \rho, \alpha, \beta} = t \left[ M \left( \frac{\Psi_{\alpha, \rho}}{\hat{\tau}_{\beta, \rho}} \right) + 2\beta \right].$$

Now since  $K_{\rho, \alpha, \beta} \rightarrow 0$  if  $\rho \rightarrow 0$ , followed by  $\alpha \rightarrow 0$ , followed by  $\beta \rightarrow 0$  we obtain that  $\lim_{\varepsilon \rightarrow 0} \mathbf{E}F(\Phi_\varepsilon)$  exists and equals  $\lim_{\rho \rightarrow 0} \mathbf{E}F(\hat{A}_\rho + \hat{B}_\rho)$ .

It remains to exploit Lemma 8.7 below to deduce that

$$\Phi_\varepsilon \Rightarrow \int_0^{t \wedge \kappa_M} \mathcal{A}(z^2)(U_s, V_s) ds$$

as  $\varepsilon \rightarrow 0$ . □

**Lemma 8.6.** *Let  $X_n$  be a sequence of  $\mathcal{X}$ -valued r. v.'s, and  $X$  be such that  $X_n \Rightarrow X$ , where  $\mathcal{X}$  is a separable Banach space. Let  $\{F_n, n \geq 1\}$  be a sequence in  $C(\mathcal{X})$ , which is such that as  $n \rightarrow \infty$ ,  $F_n \rightarrow F$  uniformly on each compact subset of  $\mathcal{X}$ . Then  $F_n(X_n) \Rightarrow F(X)$ , as  $n \rightarrow \infty$ .*

*Proof of Lemma 8.6.* Choose  $\varepsilon > 0$  arbitrary, and let  $K$  be a compact subset of  $\mathcal{X}$  such that  $\mathbf{P}(X_n \notin K) \leq \varepsilon$ , for all  $n \geq 1$ . Now choose  $n$  large enough such that  $|F_n(x) - F(x)| \leq \varepsilon$ , for all  $x \in K$ . Choose an arbitrary  $G \in C_b(\mathbf{R})$ , such that  $\sup_x |G(x)| \leq 1$ . We have

$$\begin{aligned} |\mathbf{E}[G \circ F_n(X_n)] - \mathbf{E}[G \circ F(X)]| &\leq |\mathbf{E}[G \circ F_n(X_n) - G \circ F(X_n); X_n \in K]| \\ &\quad + 2\varepsilon + |\mathbf{E}[G \circ F(X_n) - G \circ F(X)]| \end{aligned}$$

The first term of the righthand side can be made arbitrarily small by choosing  $\varepsilon$  small, uniformly in  $n$ , since  $G$  is uniformly continuous on the union of the images of  $K$  by the  $F_n$ 's. The last term clearly goes to zero as  $n \rightarrow \infty$ . □

**Lemma 8.7.** *Let  $\{X_n, n \geq 1\}$  and  $X$  denote real-valued random variables, defined on a given probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ . A sufficient condition for  $X_n \Rightarrow X$  is that*

$$\mathbf{E}[F(X_n)] \rightarrow \mathbf{E}[F(X)],$$

for any continuous, bounded and increasing function  $F$ .



*Proof of Lemma 8.7.* It is plain that the condition of the Lemma implies that  $\mathbf{E}[F(X_n)] \rightarrow \mathbf{E}[F(X)]$  for any  $F$  continuous, bounded with bounded variations. Associating to each  $M > 0$  a continuous function  $F_M$  from  $\mathbf{R}$  into  $[0, 1]$ , which is decreasing on  $\mathbf{R}_-$  and increasing on  $\mathbf{R}_+$ , equal to zero on the interval  $[-M+1, M-1]$ , and to one outside the interval  $[-M, M]$ , we note that the condition of the Lemma implies that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{P}(|X_n| > M) &\leq \lim_{n \rightarrow \infty} \mathbf{E}[F_M(X_n)] \\ &= \mathbf{E}[F_M(X)]. \end{aligned}$$

Since the last right-hand side can be made arbitrarily small by choosing  $M$  large enough, the last statement implies tightness of the sequence  $\{X_n, n \geq 1\}$ . Consequently  $X_n \Rightarrow X$  will follow if  $\mathbf{E}[F(X_n)] \rightarrow \mathbf{E}[F(X)]$  for any  $F$  in a class of continuous and bounded functions which separates probability measures, which clearly is the case under the condition of the theorem.  $\square$

## 9. PROOF OF THEOREM 3.2

Recall that for each  $\varepsilon > 0$   $\mu^\varepsilon$  denotes the unique invariant probability measure of  $\xi^\varepsilon$ , and that  $\lambda$  denotes the unique invariant probability measure of the diffusion process  $(U_t, V_t)$  or equivalently of its semigroup  $Q_t$ .

The fact that any accumulation point of the collection  $\{\mu^\varepsilon, \varepsilon > 0\}$  satisfies  $\mu = \mu\mathfrak{s}_\pm^{-1}$  follows from Proposition 5. Corollary 3 states that at most one invariant probability measure of (1) satisfies both  $\mu = \mu\mathfrak{s}_\pm^{-1}$  and  $\mu\Phi^{-1} = \lambda$ .

From Corollary 2 we know the collection  $\{\mu^\varepsilon, \varepsilon > 0\}$  is tight. Consequently, there exists a sequence  $\varepsilon_n \rightarrow 0$  and a measure  $\tilde{\mu}$ , such that  $\mu^{\varepsilon_n} \Rightarrow \tilde{\mu}$ .

We now show that this  $\tilde{\mu}$  is invariant for the  $\xi_r$  dynamics. Fix an arbitrary  $t > 0$ . If we initialize  $\xi^{\varepsilon_n}$  with its invariant probability measure  $\mu^{\varepsilon_n}$ , then both marginal laws of the pair  $(\xi_0^{\varepsilon_n}, \xi_t^{\varepsilon_n})$  equal  $\mu^{\varepsilon_n}$ . Since  $(\xi_0^{\varepsilon_n}, \xi_t^{\varepsilon_n}) \Rightarrow (\xi_0, \xi_t)$ , we deduce that if  $\xi_0 \simeq \tilde{\mu}$ , then  $\xi_t \simeq \tilde{\mu}$ , and this is true for all  $t > 0$ , hence  $\tilde{\mu}$  is invariant for  $\xi$ .

Next recall that for each  $\varepsilon > 0$ , we defined  $\lambda^\varepsilon := \mu^\varepsilon\Phi^{-1}$ . Since both marginal laws of the pair  $(\xi_0^{\varepsilon_n}, \xi_{t/\varepsilon}^{\varepsilon_n})$  equal  $\mu^{\varepsilon_n}$ , we conclude that both marginal laws of the pair  $((U_0^{\varepsilon_n}, V_0^{\varepsilon_n}), (U_t^{\varepsilon_n}, V_t^{\varepsilon_n}))$  equal  $\lambda^{\varepsilon_n}$ . We now show that  $\tilde{\lambda} = \tilde{\mu}\Phi^{-1}$  is an invariant measure for  $Q_t$ . Combining the fact that

$$((U_0^\varepsilon, V_0^\varepsilon), (U_t^\varepsilon, V_t^\varepsilon)) \Rightarrow ((U_0, V_0), (U_t, V_t)),$$

$\Phi$  is continuous and that from Lemma 8.2 we know that  $\tilde{\lambda}$  is supported on  $(0, +\infty) \times (0, +\infty)$ , we conclude that  $\lambda^{\varepsilon_n} \Rightarrow \tilde{\lambda}$ . Since the marginals are equal for all  $t > 0$ , we conclude that  $\tilde{\lambda}$  is an invariant probability measure for  $Q_t$ . Since from Theorem 6.12,  $Q_t$  has the unique invariant probability measure  $\lambda$ , we conclude that  $\tilde{\lambda} = \lambda$ .

Hence  $\tilde{\mu} = \mu$ , and  $\mu^\varepsilon \Rightarrow \mu$ , as  $\varepsilon \rightarrow 0$ .

Because  $\lambda = \mu\Phi^{-1}$  does not charge the diagonal,  $\mu$  does not charge the set  $\{x = y\} \cup \{x = -y\}$ . Pick a point  $(x, y, z) \in \mathbf{R}^3$ , with  $|x| \neq |y|$  and let  $(u, v) = \Phi(x, y, z)$ . Assume that  $(x, y, z) \in \Gamma_{u,v}^+$  (the other case is treated exactly in the same way). To  $(x, y, z)$  corresponds a value  $\theta$  of the parameter on  $\Gamma_{u,v}^+$  defined in Section 7.3.1. There exists a smooth bijection  $\Psi$  with a smooth inverse from an open neighborhood  $\mathcal{O}$  of  $(x, y, z)$  onto an open neighborhood  $\mathcal{U}$  of  $(u, v, \theta)$ . The restriction of  $\mu$  to  $\mathcal{O}$  is the image by  $\Psi^{-1}$  of the restriction to  $\mathcal{U}$  of the measure

$$\frac{K_{u,v}}{|\cos(\phi_{u,v}(\theta))|} \rho(u, v) d\theta dudv,$$

where  $\rho$  denotes the density of  $\lambda$  with respect to Lebesgue measure and we have used formula (26). Hence the restriction of  $\mu$  to  $\mathcal{O}$  is absolutely continuous with respect to Lebesgue measure on  $\mathbf{R}^3$ , with a density which is positive at  $(x, y, z)$  since  $\rho(u, v) > 0$ .

#### ACKNOWLEDGMENTS

The authors are indebted to two anonymous referees who pointed some errors in the original version of this paper, to Alexander Veretennikov for bringing [Por90] to our attention, and to Hans–Juergen Engelbert, for discussions on uniqueness/non–uniqueness for one–dimensional SDEs. JCM thanks the National Science Foundation for its support through the grant NSF-DMS-08-54879 (FRG).

#### REFERENCES

- [Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [CFS82] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.
- [FW12] Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szűcs.
- [HM08] Martin Hairer and Jonathan C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.
- [Hör94a] Lars Hörmander. *The analysis of linear partial differential operators. III*, volume 274 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1994. Pseudo-differential operators, Corrected reprint of the 1985 original.
- [Hör94b] Lars Hörmander. *The analysis of linear partial differential operators. IV*, volume 275 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1994. Fourier integral operators, Corrected reprint of the 1985 original.
- [Kif74] Ju. I. Kifer. Some theorems on small random perturbations of dynamical systems. *Uspehi Mat. Nauk*, 29(3(177)):205–206, 1974.
- [KP08] Sergei B. Kuksin and Andrey L. Piatnitski. Khasminskii-Whitham averaging for randomly perturbed KdV equation. *J. Math. Pures Appl. (9)*, 89(4):400–428, 2008.
- [Kuk04] Sergei B. Kuksin. The Eulerian limit for 2D statistical hydrodynamics. *J. Statist. Phys.*, 115(1-2):469–492, 2004.
- [Kuk07a] S. B. Kuksin. Eulerian limit for 2D Navier-Stokes equation and damped/driven KdV equation as its model. *Tr. Mat. Inst. Steklova*, 259(Anal. i Osob. Ch. 2):134–142, 2007.
- [Kuk07b] S. B. Kuksin. Eulerian limit for 2D Navier-Stokes equation and damped/driven KdV equation as its model. *Tr. Mat. Inst. Steklova*, 259(Anal. i Osob. Ch. 2):134–142, 2007.
- [Kuk10] Sergei B. Kuksin. Damped-driven KdV and effective equations for long-time behaviour of its solutions. *Geom. Funct. Anal.*, 20(6):1431–1463, 2010.
- [Kuk13] Sergei B. Kuksin. Weakly nonlinear stochastic CGL equations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):1033–1056, 2013.
- [Lor63] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130141, Mar 1963.
- [MTVE02] Paul A. Milewski, Esteban G. Tabak, and Eric Vanden-Eijnden. Resonant wave interaction with random forcing and dissipation. *Studies in Applied Mathematics*, 108(1):123144, 2002.
- [Por90] Nikolai I. Portenko. *Generalized Diffusion Processes*, volume 83 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1990. Translated from the 1982 Russian original by H. H. McFaden.

- [Rue82] David Ruelle. Small random perturbations of dynamical systems and the definition of attractors. *Comm. Math. Phys.*, 82(1):137–151, 1981/82.
- [Sin68] Ja. G. Sinaĭ. Markov partitions and U-diffeomorphisms. *Funkcional. Anal. i Priložen*, 2(1):64–89, 1968.
- [Sin72] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.
- [Str08] Daniel W. Stroock. *Partial differential equations for probabilists*, volume 112 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008.
- [SV72] Daniel W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, pages 333–359, Berkeley, Calif., 1972. Univ. California Press.
- [SV79] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.
- [Var07] S. R. S. Varadhan. *Stochastic processes*, volume 16 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York, 2007.
- [You02] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.*, 108(5-6):733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

*E-mail address:* jonm@math.duke.edu

*E-mail address:* etienne.pardoux@univ-amu.fr