INVARIANT MEASURE SELECTION BY NOISE AN EXAMPLE

JONATHAN C. MATTINGLY AND ETIENNE PARDOUX

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 known that the zero noise limit of random perturbations selects a conical 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 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 limits 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 studies in a sequence of works by Kuskin and his co-authors[Kuk04, Kuk07a, Kuk07b, KP08, Kuk10b].

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 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 prove insight into important equations 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, Kuk10b] the question of uniqueness of the limit is not addressed and seems out of reach.

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

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

where $\nu > 0$ is the viscosity, Δ is the Laplacian, σ_k are constants chosen to enforce the reality of q, $\{W_t^{(k)} : k \in \mathbf{Z}^2\}$ 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 ν is chosen to keep the spatial L^2 norm of order one in the $\nu \to 0$ 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 is 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 by analysed.

Inspired by models in [Lor63, MTVE02] and the Euler equation itself, we construct a model problem in the form of a 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 with out becoming trivial). In many ways our analysis follows a 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. Using this dynamics were are able to show that a unique limiting measure is selected. However multiple conserved quantities are not usually treated in Wentzell-Freidlin theory and the complications of having more then one are non-trivial in our case. They lead to a surprising dependence of the nonlinearity on the structure of the noise.

In many examples, it is tempting to cite the fact that pre-limiting invariant measures are non degenerate in that they charge all open sets to suggest that any limiting measure also does so. One interesting feature of our study is that the qualitative structure of the limiting measure varies drastically depending on the structure of the noise despite the fact that in all cases the pre-limit measure will always charge all open balls. If isotropic noise is used we will see that the limiting measure is supported only on the z-axis and has a density with respect to its one dimensional Lebesgue measure. For other choices it will be concentrated on only two points, while with still another choice of forcing structure it will have a density with respect to Lebesgue measure on $\mathbf{R}^3$ but have support contained in $\{(x, y, z) : |x| \leq |y|\}$ and for still a difference choice in $\{(x, y, z) : |x| \geq |y|\}$.

Thus by choosing noise with different structures, the support of the limiting invariant measure can be either one-dimensional, two dimensional, or three-dimensional. We will see that the this is not completely surprising once we understand the symmetries of the problems and which noises break which symmetries. However, even in light of this it is still quite surprising that in the case when the measure charges an open subset of $\mathbf{R}^3$ that it does not charge all of $\mathbf{R}^3$ as does all of the pre-limiting measures. The fact that "walls" develop in phase space which segment the phase into chambers, some which trap the dynamics and others which eventually expel it. In this example, the walls of the chambers are made up of the heteroclinic connections. Since the period of the orbits diverge logarithmically as they approach the heteroclinic connections, this might not be completely surprising that they play a special role.

2. Model System

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

$$\dot{\xi}_t = \mathcal{B}(\xi_t, \xi_t) \tag{2.1}$$

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 symmetric bi-linear form defined by

$$\mathcal{B}(\xi, \hat{\xi}) \stackrel{\text{def}}{=} \frac{1}{2} \begin{pmatrix} y\hat{z} + \hat{y}z \\ x\hat{z} + \hat{x}z \\ -2x\hat{y} - 2\hat{x}y \end{pmatrix} . \tag{2.2}$$

We will write φ_t for the flow map induced by (2.1), i.e. $\xi_t = \varphi_t(\xi_0)$. We will constantly write (X_t, Y_t, Z_t) for ξ_t when we wish to speak of the components of ξ_t.

Since

$$B(\xi, \xi) \cdot \xi = 0 \tag{2.3}$$

we see that $|\xi_t|^2 = X_t^2 + Y_t^2 + Z_t^2$ is constant along trajectories of (2.1). Similarly one sees that $X_t^2 - Y_t^2$ is also conserved by the dynamics of (2.1). 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

$$\Phi\colon (x, y, z) \mapsto (u, v) = (2x^2 + z^2, 2y^2 + z^2) . \tag{2.4}$$

A moments reflection shows that the existence of these two conserved quantities implies that all of the orbits of (2.1) 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

$$\Gamma = \Big\{\xi : \Phi(\xi) = \Phi(\xi_0)\Big\} \tag{2.5}$$

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

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

Any such defined measure is an invariant measure for the dynamics given by (2.1). Hence we see that (2.1) has infinitely many invariant measures. It is reasonable to expect the addition of sufficient driving noise and balancing dissipation, wil result in a system with a unique invariant measure. Our goal is to study the limit as the noise/dissipation are scaled to zero. We are specifically interested in if this procedure selects a unique convex combination of the measures for the underlying deterministic system (2.1).

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

$$\dot{\xi}_t^\varepsilon = \mathcal{B}(\xi_t^\varepsilon, \xi_t^\varepsilon) - \varepsilon \xi_t^\varepsilon + \sqrt{\varepsilon}\sigma \dot{W}_t, \tag{2.6}$$

with $\xi_0^\varepsilon = (X_0, Y_0, Z_0) \in \mathbf{R}^3$ and where $W_t = (W_t^{(1)}, W_t^{(2)}, W_t^{(3)})$, $\{W_t^{(i)} i = 1, 2, 3\}$ is a collection of i.i.d. standard Brownian motions, and $\sigma \in \mathbf{R}^3 \times \mathbf{R}^3$ with $(\sigma)_{ij} = \delta_{i=j}\sigma_i$. As above, we will write $(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$ when we wish to discuss the coordinates of ξ_t^ε.

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

Our aim is to study the limit of μ_ε, as $\varepsilon \to 0$. We first note that as $\varepsilon \to 0$, the process $(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$ converges to the solution of (2.1) on any finite time interval.

From (2.3) we see that $|\xi|^2 = X_t^2 + Y_t^2 + Z_t^2$ is constant along trajectories of (2.1). Similarly one sees that $X_t^2 - Y_t^2$ is also conserved by the dynamics of (2.1). 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 $\Phi\colon (x, y, z) \mapsto (u, v)$ defined by $u = 2x^2 + z^2$ and $v = 2y^2 + z^2$. A moments reflection shows that the existence of these two conserved quantities implies that all of the orbits of (2.1) 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. To any initial point (X_0, Y_0, Z_0) on one of the closed orbits, we can associate a measure defined by the following limit

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \delta_{(X_s, Y_s, Z_s)} ds \,.$$

Any such defined measure is an invariant measure for the dynamics given by (2.1). Hence we see that (2.1) has infinitely many invariant measures.

The principle result of this article is that there exists a probability measure μ which is absolutely continuous with respect to Lebesgue measure and so that μ^ε converges weakly as μ as $\varepsilon \to 0$.

3. Main Results

For $\varepsilon \geq 0$, we define the Markov semigroup P_t^ε associated with (2.6) by

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

for bounded $\phi\colon \mathbf{R}^3 \to \mathbf{R}$.

Remark 3.1. *When $\varepsilon = 0$, then $(P_t\phi)(\xi) = \phi(\xi_t)$ since the dynamics is deterministic.*

Theorem 3.2. *Provided that at least two of the $\sigma_i > 0$, for $i = 1, 2, 3$, are strictly positive then for each $\varepsilon > 0$, P_t^ε has a unique invariant measure μ^ε which has a C^∞ density which is everywhere positive.*

Theorem 3.3. *If either $\sigma_1 = \sigma_2 > 0$, or $\sigma_1 > \sigma_2 > \sigma_3 = 0$, or $\sigma_2 > \sigma_1 > \sigma_3 = 0$, then there exists a measure μ which is invariant for the dynamics generated by (2.1), namely $\mu P_t = \mu$ for all $t > 0$, such that $\mu^\varepsilon \Rightarrow \mu$ as $\varepsilon \to 0$.*

Theorem 3.4. *Furthermore the following descriptions of μ hold:*

(1) *If $\sigma_1 = \sigma_2 = \sigma > 0$ and $\sigma_3 = 0$ then then*
$$\mu = \tfrac{1}{2}\delta_{(0,0,\sigma)} + \tfrac{1}{2}\delta_{(0,0,-\sigma)}$$
(2) *If $\sigma_1 = \sigma_2 = \sigma > 0$ and $\sigma_3 > 0$ then*
$$\mu = \int_0^\infty \left[\tfrac{1}{2}\delta_{(0,0,s)} + \tfrac{1}{2}\delta_{(0,0,-s)}\right]\rho(s)ds$$
and
$$\rho(s) = K|s|^{\sigma_1^2/\sigma_3^2}\exp\left(-\tfrac{s^2}{2\sigma_3^2}\right)$$
where K is the normalisation constant.

(3) *If $\sigma_1 > \sigma_2 > \sigma_3 = 0$ then μ has a density with respect to Lesbegue measure on $\mathbf{R}^3$ which possesses a C^∞ density in the interior of the set*
$$\{(x, y, z) : |x| \geq |y|\}$$
only isolated zeros inside this set and strictly zero outside this set.

(4) *If $\sigma_2 > \sigma_1 > \sigma_3 = 0$ then the situation is identical to the preceding case but the statements hold with resect to the set*
$$\{(x, y, z) : |x| \leq |y|\}$$

4. Finite Time Convergence on original timescale

Lemma 4.1. *There exist 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 4.2. *For any $t \geq 0$, P_t^ε converges weakly to P_t as $\varepsilon \to 0$. In other words, for any bounded and continuous $\phi : \mathbf{R}^3 \to \mathbf{R}$, $P_t^\varepsilon \phi(\xi) \to P_t\phi(\xi)$ 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 $K > 0$

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

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 quoted result. □

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 μ^ε and the tightness of various objects.

Proposition 5.1. *For any integer $p \geq 1$ there exists $C(p) > 0$ so that for all $t \geq 0$, $\varepsilon > 0$,*

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

Proof of Proposition 5.1. 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^ε with quadratic variation satisfying

$$d\langle M^\varepsilon\rangle_t = (\sigma_1^2(X_t^\varepsilon)^2 + \sigma_2^2(Y_t^\varepsilon)^2 + \sigma_3^2(Z_t^\varepsilon)^2)dt \leq \sigma_{max}^2|\xi_t^\varepsilon|^2 dt$$

where $\sigma_{max}^2 = \max(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. The proof then follows from Lemma 5.3 which is given a proven below. □

Remark 5.2. *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 martingle estimate.*

Lemma 5.3. *Let X_t be a semimartingle 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 $t \geq 0$

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

Proof of Lemma 5.3. 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 $N \to \infty$. Using the assumption on the quadratic variation of M_t we see that M_t is a L^2-Martingale. Hence

$$\mathbf{E}X_t = e^{-bt}X_0 + \frac{a}{b}(1 - e^{-bt}) \tag{5.1}$$

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 stoping time τ 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-Martingle 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 (5.1) as the base case produces the quoted result. □

Corollary 5.4. *For each $\varepsilon > 0$, the Feller diffusion $\{\xi_t^\varepsilon : t \geq 0\}$ possesses at least one invariant probability measure μ^ε. Furthermore, any invariant probability measure μ^ε satisfies*

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

for any integer $p \geq 1$ where $C(p)$ is the constant form Lemma 5.1 (which is independent of ε). Hence the collection of probability measures which are invariant under the dynamics for some $\varepsilon > 0$ is tight.

Proof of Corollary 5.4. Since ξ_t^ε is a time–homogeneous Feller diffusion process and from Proposition 5.1 for fixed $\varepsilon > 0$, the collection of random vectors $\{\xi_t^\varepsilon,\, t > 0\}$ is tight, the existence of an invariant measure μ^ε follows by the KrylovBogolyubov theorem.

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

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

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

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

Proposition 5.5. *If at least two of the σ_1, σ_2 and σ_3 are not zero then for any $\varepsilon > 0$, there exists a transition density $p_t^\varepsilon(\xi, \eta)$ which is jointly-smooth in (t, ξ, η) 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 5.5 . When all of the σ are positive then the system is uniformly elliptic and the results follow from the classical theory of uniformly elliptic diffusions.

As long as only one $\sigma_i = 0$, hypoellipticity follows from the fact that taking Lie brackets of the drift with any two coordinate directions in succession produces the third. This ensures the existence of a smooth density with respect to Lesbegue mesure [Str08, Hör94a, Hör94b]. Positivity will then follow by showing that the support of the transition density is all of $\mathbf{R}^3$.

We will consider the case $\sigma_3 = 0$, $\sigma_1 > 0$ and $\sigma_2 > 0$. The other cases follow identical reasoning. 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

$$\begin{aligned}\frac{dx^\varepsilon}{dt}(t) &= y^\varepsilon(t)z^\varepsilon(t) - \varepsilon x^\varepsilon(t) + \sqrt{\varepsilon}\sigma_1 f_1(t)\\ \frac{dy^\varepsilon}{dt}(t) &= x^\varepsilon(t)z^\varepsilon(t) - \varepsilon y^\varepsilon(t) + \sqrt{\varepsilon}\sigma_2 f_2(t)\\ \frac{dz^\varepsilon}{dt}(t) &= -2x^\varepsilon(t)y^\varepsilon(t) - \varepsilon z^\varepsilon(t),\end{aligned} \tag{5.2}$$

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. □

We are now in a position to give the proof of Theorem 3.2.

Proof of Theorem 3.2. Since by Proposition 5.5 , P_t^ε has a smooth, transition density any invariant 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 measure must have disjoint support which is impossible since the measures have densities which are strictly positive except possibly at isolated points. . Uniqueness of invariant measure follows immediately from the fact that any invariant measure can be decomposed into ergodic components [FKS87]. □

6. The fast dynamics

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

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

$$\dot{\widetilde{\xi}}_t^\varepsilon = \frac{1}{\varepsilon}\mathcal{B}(\widetilde{\xi}_t^\varepsilon, \widetilde{\xi}_t^\varepsilon) - \widetilde{\xi}^\varepsilon + \sigma \dot{W}_t, \tag{6.1}$$

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

Let $\widetilde{P}_t^\varepsilon$ be the Markov semigroup associated to (6.1) and defined for $\psi : \mathbf{R}^3 \to \mathbf{R}$ by

$$(\widetilde{P}_t^\varepsilon \psi)(\xi) = \mathbf{E}_\xi \psi(\widetilde{\xi}_t^\varepsilon)\,. \tag{6.2}$$

Associated with this right-action on functions we associated dual action on measures. We will denote this by left action rather then the often used $(\widetilde{P}_t^\varepsilon)^*$ notation. Hence is μ is a measure on $\mathbf{R}^3$ and ψ a real-valued function on $\mathbf{R}^3$ then

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

Of course, this time change does not change the set of invariant measures for the dynamics. Hence if we let $\mathcal{M}$ denote set of all probability measures and $\mathcal{I}^\varepsilon$ the set of all invariant probability measure for (2.6) then

$$\mathcal{I}^\varepsilon = \big\{ \mu \in \mathcal{M} : \mu \widetilde{P}_t^\varepsilon = \mu \quad \text{for all} \quad t \geq 0 \big\}$$

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 $\varepsilon \to 0$. More precisely, defining the processes $(U_t^\varepsilon\ V_t^\varepsilon)$ by $(U_t^\varepsilon, V_t^\varepsilon) = \Phi(\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon, \widetilde{Z}_t^\varepsilon)$ and application of Itô's formula shows that

$$\begin{aligned} dU_t^\varepsilon &= [2\sigma_1^2 + \sigma_3^2 - 2U_t^\varepsilon]dt + 4\sigma_1 \widetilde{X}_t^\varepsilon dW_t^{(1)} + 2\sigma_3 \widetilde{Z}_t^\varepsilon dW_t^{(3)}, \\ dV_t^\varepsilon &= [2\sigma_2^2 + \sigma_3^2 - 2V_t^\varepsilon]dt + 4\sigma_2 \widetilde{Y}_t^\varepsilon dW_t^{(2)} + 2\sigma_3 \widetilde{Z}_t^\varepsilon dW_t^{(3)}. \end{aligned} \tag{6.3}$$

We will show below that $(U_t^\varepsilon, V_t^\varepsilon)$ converges weakly as a process to (U_t, V_t) which solves

$$\begin{aligned} dU_t &= [2\sigma_1^2 + \sigma_3^2 - 2U_t]dt + \sigma_1 \sqrt{8\big(U_t - \Gamma(U_t, V_t)\big)}dW_t^{(1)} \\ &\qquad + 2\sigma_3 \sqrt{\Gamma(U_t, V_t)} dW_t^{(3)}, \\ dV_t &= [2\sigma_2^2 + \sigma_3^2 - 2V_t]dt + \sigma_2 \sqrt{8\big(V_t - \Gamma(U_t, V_t)\big)}dW_t^{(2)} \\ &\qquad + 2\sigma_3 \sqrt{\Gamma(U_t, V_t)} dW_t^{(3)}. \end{aligned} \tag{6.4}$$

where

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

The function Λ will be defined in a later section. However for our present discussion, it will be sufficient to state a few important facts

Proposition 6.1. *$\Lambda(r)$ is a continuous and strictly increasing function on $[0,1]$ with $\Lambda(0) = \frac{1}{2}$ and $\Lambda(1) = 1$. Furthermore as $\varepsilon \to 0^+$,*

$$\Lambda(\varepsilon) = \frac{1}{2} + \frac{1}{16}\varepsilon + \frac{1}{32}\varepsilon^2 + o(\varepsilon^2)$$
$$\Lambda(1-\varepsilon) = 1 - \frac{2}{|\ln(\varepsilon)|} + o\Big(\frac{1}{|\ln(\varepsilon)|}\Big)$$

In addition, on any closed interval in $[0,1)$, Λ is uniformly Lipschitz.

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

Lemma 6.2. *Let $\{X_t,\, t \geq 0\}$ be a continuous $\mathbf{R}_+$–valued $\mathcal{F}_t$–adapted process which satisfies*

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

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 consider the SDE

$$dY_t = c^{-1}\left(aY_t - bY_t^2\right)dt + Y_t dW_t,$$
$$Y_0 = x,$$

whose solution satisfies clearly

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

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

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

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 on the event $\{A_\infty < \infty\}$, $\limsup_{t\to\infty} Y_t = +\infty$, which implies that $A_\infty = +\infty$. □

We will need a slightly better result

Lemma 6.3. *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 6.3. 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. ☐

A similar argument allows us to prove

Lemma 6.4. *Let* X_t *be a recurrent* $\mathbf{R}_+$*–valued process, solution of the SDE*

$$\begin{aligned} dX_t &= b(X_t)dt + \sqrt{c(X_t)}dW_t, \\ X_0 &= x, \end{aligned}$$

where $x > 0$*,* $b : \mathbf{R}_+ \to \mathbf{R}$ *is measurable and upper bounded,* $c : \mathbf{R}_+ \to \mathbf{R}_+$ *is locally Lipschitz on* $(0, +\infty)$*,* $c(x) > 0$ *if* $x > 0$*, and* $c(x)/x \to +\infty$ *as* $x \to 0$*. Then* X_t *hits zero in finite time a. s.*

We now prove the existence an uniqueness of solutions for the (U, V) equation.

Theorem 6.5. *Assume that the initial condition* (U_0, V_0) *satisfies* $U_0 > 0$*,* $V_0 > 0$*. In both cases*

(1) $\sigma_1 = \sigma_2 \neq 0$ *and* $\sigma_3 \geq 0$*,*

(2) $\sigma_1 > \sigma_2 > \sigma_3 = 0$, *or* $\sigma_2 > \sigma_1 > \sigma_3 = 0$,

equation (6.4) *has a unique weak solution which lives in the set* $(0,+\infty)\times(0,+\infty)\cup\{(0,0)\}$ *(in case (1)) and in* $(0,+\infty)\times(0,+\infty)$ *(in case (2)) for all positive times.*

Note that we have strong existence and uniqueness in case (1).

Proof of Theorem 6.5. Existence of a weak solution follows from the fact that the coefficients are continuous.

We first prove that in the two above cases, any solution (U_t, V_t) never hits the two axis, except possibly at $(0,0)$ (the origin is possibly hit only in the first case). The fact that (U_t, V_t) cannot hit $(0, v)$ with $v > 0$ follows clearly from the equation for U_t and Lemma 6.3, once we have noted that whenever $U_t < V_t$, $U_t - \Gamma(U_t, V_t) \le U_t/2$, as follows from Lemma 6.1.

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 $(0,0)$ in the second case. 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 6.3, since $\Gamma(U_t, V_t) \ge 0$.

We now prove strong uniqueness in the first case. If $U_0 \ne V_0$, since the coefficients are locally Lipschitz away from the diagonal, the solution is unique until the diagonal is hit, which happens soon or later as a consequence of Lemma 6.4, since the process is recurrent and non degenerate in each sector $0 < u < v$ and $0 < v < u$, and moreover the diffusion coefficient of the process $U_t - V_t$ vanishes like $\sqrt{[\log(|U_t - V_t|/U_t \vee V_t)]^{-1}}$ near the diagonal. Once on the diagonal, it is easily seen that the process stays there. If $\sigma_3 = 0$, then the process on the diagonal satisfies a linear ODE, while if $\sigma_3 > 0$, the process on the diagonal satisfies a one–dimensional SDE to which the well–known strong uniqueness result of Yamada–Watanabe applies, see for example Theorem IX.3.5 in [RY99].

We now consider the second case. Here the process again hits the diagonal, but it will spend zero time there. Define

$$F(u,v) = \begin{cases} \log\left(\frac{u\vee v}{|u-v|}\right), & \text{if } u \vee v > 0 \text{ and } u \ne v, \\ 0, & \text{otherwise,} \end{cases},$$

$J_t = F(U_t, V_t)$, $A_t = \int_0^t \frac{1}{J_s} ds$, $\eta_t = \inf\{s > 0, A_s > t\}$, $H_t = U_{\eta_t}$ and $K_t = V_{\eta_t}$. There exists a two dimensional Wiener process, which we still denote by $(W_t^{(1)}, W_t^{(2)})$, such that

$$dH_t = (2\sigma_1^2 - 2H_t)F(H_t, K_t)dt + \sigma_1\sqrt{8(H_t - \Gamma(H_t, K_t))F(H_t, K_t)}dW_t^{(1)},$$
$$dK_t = (2\sigma_2^2 - 2K_t)F(H_t, K_t)dt + \sigma_2\sqrt{8(K_t - \Gamma(H_t, K_t))F(H_t, K_t)}dW_t^{(2)}.$$

It is easily verified that the diffusion matrix of this system is uniformly elliptic and locally Lipschitz, while the drift term belongs to $L^6_{\text{loc}}(\mathbf{R}_+ \times \mathbf{R}_+)$. Hence the strong uniqueness statement in Theorem 2.1 from [GM01] applies. This translates into a weak uniqueness result for the (U, V) equation. □

Remark 6.6. *We have not been able to prove that in the first case the process does not hit the diagonal first at the origin, also we suspect that it might be impossible. If the diagonal is hit first at the origin, the process enters instantaneously the open set $\{(u, v),\ u = v > 0\}$, and never hits the origin again.*

Our method for proving uniqueness in the second case does not allow us to consider the case $\sigma_1 \neq \sigma_2$, $\sigma_3 > 0$. Uniqueness of the solution in this last case is an open problem.

6.3. **Longtime behavior of (U, V).** We will let Q_t denote the Markov semigroup associated to (6.4) which can be defined by

$$(Q_t\phi)(u, v) = \mathbf{E}_{(u,v)}\big[\phi(U_t, V_t)\big] \tag{6.5}$$

for $\phi\colon \mathbf{R}_+^2 \to \mathbf{R}$.

Unlike the pair $(U_t^\varepsilon, V_t^\varepsilon)$, the pair (U_t, V_t) is a Markov process and hence we can speak of invariant measure for the Markov semigroup Q_t.

Observe that

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

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

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

for some positive c. Hence the following result follows from Lemma 5.3.

Proposition 6.7. *For any $p \geq 1$, there exists a constant $C(p)$ so that*

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

Theorem 6.8. *The semigroup Q_t generated by the dynamics of (U_t, V_t) possess an invariant measure λ. In each of the following the cases λ is the unique invariant measure and it has the stated characterization.*

(1) *If $\sigma_1 = \sigma_2 = \sigma \neq 0$ and $\sigma_3 = 0$, then $\lambda = \delta_{(\sigma^2,\sigma^2)}$*

(2) *$\sigma_1 = \sigma_2 = \sigma \neq 0$ and $\sigma_3 > 0$, $\lambda(du \times dv) = G(u)\delta_u(dv)du$ where G is a probability density function. See Remark 6.9 below for the form of G.*
(3) *If $\sigma_1 > \sigma_2 > \sigma_3 = 0$, then $\lambda(du \times dv) = \rho(u,v)\,du\,dv$ where ρ is a density which is smooth function in the interior of $\mathcal{D}_\geq = \{(u,v) : u, v \in [0,\infty), u \geq v\}$ with at most possibly isolated zeros in the interior of $\mathcal{D}_\geq$, and is zero outside of $\mathcal{D}_\geq$.*
(4) *If $\sigma_2 > \sigma_1 > \sigma_3 = 0$, then $\lambda(du \times dv) = \rho(v,u)\,du\,dv$ where ρ is the function in the preceding case. Notice $(u,v) \to \rho(v,u)$ is supported on $\mathcal{D}_\leq = \{(u,v) : (v,u) \in \mathcal{D}_\geq\}$.*

Proof of Theorem 6.8. Most of the needed observations are contained in the proof of Theorem 6.5. As noted there, in the first two cases the trajectory hits $u = v$ almost surely in finite time and then never leaves the diagonal again. Hence in the first two cases any invariant measure is supported on the diagonal $\{u = v, u > 0, v > 0\}$. In the first case when $\sigma_3 = 0$, the dynamics on the diagonal are deterministic since $\Lambda(u,u) = 1$. The dynamics has a unique globally attracting fix point at $(u,v) = (\sigma^2, \sigma^2)$. Which proves the result in the first case.

In the second case, we are again trapped on diagonal in finite time. Looking at (6.4) we see that the dynamics solve

$$dU_t = [2\sigma_1^2 + \sigma_3^2 - 2U_t]dt + 2\sigma_3\sqrt{U}dW_t^{(3)},$$

Direct calculation with the generator shows that at this system has an invariant measure with density given by G. Since $G(u) > 0$ for $u > 0$ we know it is the unique invariant measure since any invariant measure must have a density with respect to Lesbegue measure on $\{u = v, u > 0\}$.

The last two cases are the same only with the role of u and v switched. We will consider the case when $\sigma_1 > \sigma_2$. In this case as discussed in the proof of Theorem 6.5, If $U_0 > V_0$ then the systems never crosses $u = v$ almost surely. If $U_0 \leq V_0$ then the dynamics his $u = v$ in finite time and is instantaneously pushed in the $u > v$ where is stays for the rest of time almost surely. In the interior of $\{u > v > 0\}$ the coefficients are locally smooth, hence there is a positive probability in any finite time of hitting any open subset of $\{(u,v) : u > v > 0\}$. The existence of a transition density implies that any invariant measure has a density with respect to Lesbegue. The positivity of hitting open sets ensure that the invariant measure is unique. □

Remark 6.9. *Standard computations show that in case (2) the invariant density $G(u)$ equals, up to a normalizing constant,*

$$u^\rho \exp(-u/c), \qquad \text{with } \rho = \frac{\sigma_1^2}{\sigma_3^2} - \frac{1}{2}, \text{ and } c = 2\sigma_3^2.$$

7. The Deterministic Dynamics

We now in investigate more fully the deterministic dynamics given in (2.1) and obtained by formally setting $\varepsilon = 0$ in (2.6). As already mentioned, (2.1) 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 $v = 2y^2 + z^2$ gives 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 (6.1) implies that on the fast times scale the solution will make increasingly many orbits about the deterministic orbits of (2.1) 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 it 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

$$\Gamma^+_{u,v} = \Big\{ (\sqrt{\tfrac{u-z^2}{2}}, \pm\sqrt{\tfrac{v-z^2}{2}}, z) \colon z \in [-\sqrt{v}, \sqrt{v}] \Big\}$$

and another by

$$\Gamma^-_{u,v} = \Big\{ (-\sqrt{\tfrac{u-z^2}{2}}, \pm\sqrt{\tfrac{v-z^2}{2}}, z) \colon z \in [-\sqrt{v}, \sqrt{v}] \Big\}$$

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

$$\Gamma^+_{u,v} = \Big\{ (\pm\sqrt{\tfrac{u-z^2}{2}}, \sqrt{\tfrac{v-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}] \Big\}$$
$$\Gamma^-_{u,v} = \Big\{ (\pm\sqrt{\tfrac{u-z^2}{2}}, -\sqrt{\tfrac{v-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}] \Big\}$$

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

$$\mathfrak{sn}(x, y, z) = \text{sign}\left(\mathbf{1}_{|x|>|y|}x + \mathbf{1}_{|y|>|x|}y\right) \tag{7.1}$$

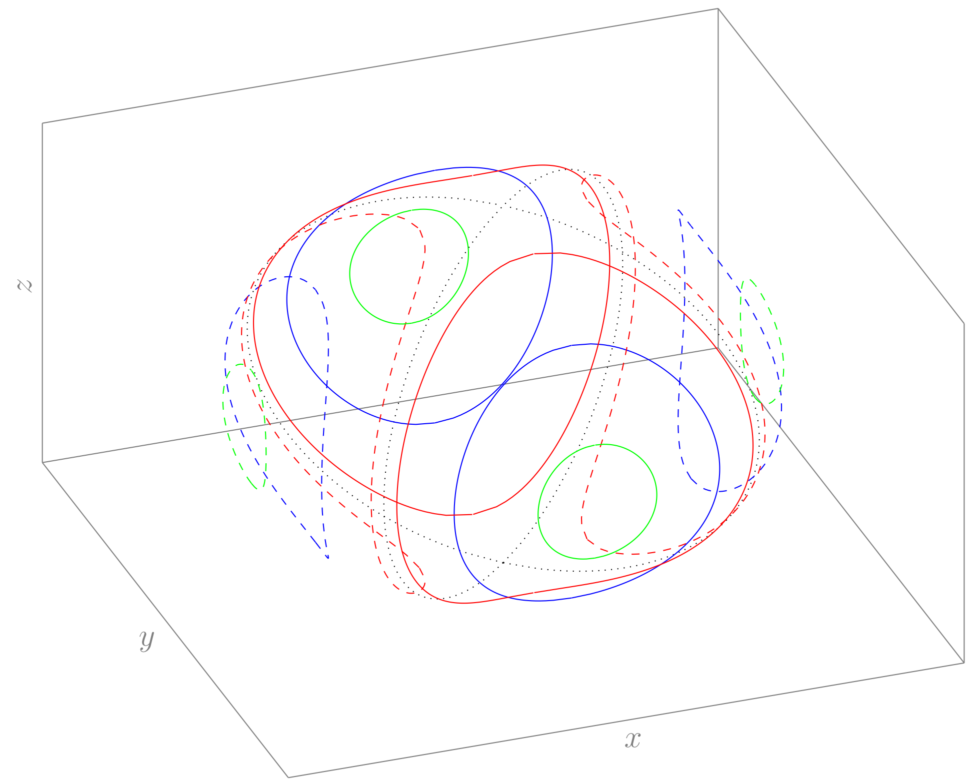

FIGURE 1. Trajectories of the system on the sphere of radius 1.The solid orbits correspond to $v > u$ (same as $|y| > |x|$). The dashed orbitscorrespond to $v < u$ (same as $|y| < |x|$). The dotted lines are hetroclinic connections which connect the fix points at the north and south poles, $(0, 0, 1)$ and $(0, 0, -1)$ respectively. The difference between the two collections of orbits of each type is the choice of the $\mathfrak{sn}$ defined in (7.1).

The value of sigma corresponds to the sign decorating the $\Gamma^{\pm}_{u,v}$. 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 $\Gamma^{\sigma}_{u,v}$.

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

$$\mathcal{H}_u^{(1)} = \Big\{(\sqrt{\tfrac{u-z^2}{2}}, \sqrt{\tfrac{u-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}]\Big\}$$
$$\mathcal{H}_u^{(2)} = \Big\{(\sqrt{\tfrac{u-z^2}{2}}, -\sqrt{\tfrac{u-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}]\Big\}$$
$$\mathcal{H}_u^{(3)} = \Big\{(-\sqrt{\tfrac{u-z^2}{2}}, -\sqrt{\tfrac{u-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}]\Big\}$$
$$\mathcal{H}_u^{(4)} = \Big\{(-\sqrt{\tfrac{u-z^2}{2}}, \sqrt{\tfrac{u-z^2}{2}}, z) \colon z \in [-\sqrt{u}, \sqrt{u}]\Big\}$$

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

$$\lim_{v\to u^-} \Gamma^+_{u,v} = \lim_{u\to v^+} \Gamma^+_{u,v} = \mathcal{H}_u^{(1)} \cup \mathcal{H}_u^{(2)}$$
$$\lim_{v\to u^+} \Gamma^+_{u,v} = \lim_{u\to v^-} \Gamma^+_{u,v} = \mathcal{H}_u^{(1)} \cup \mathcal{H}_u^{(4)}$$
$$\lim_{v\to u^-} \Gamma^-_{u,v} = \lim_{u\to v^+} \Gamma^-_{u,v} = \mathcal{H}_u^{(3)} \cup \mathcal{H}_u^{(4)}$$
$$\lim_{v\to u^+} \Gamma^-_{u,v} = \lim_{u\to v^-} \Gamma^-_{u,v} = \mathcal{H}_u^{(2)} \cup \mathcal{H}_u^{(3)}$$

In contrast to the case when $u \neq v$, the orbits starting from a given (x, y, z) point 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 it does not pose a problem which is an interesting and important feature of this model.

7.2. **Symmetries and their implications.** Defining $\mathfrak{s}_e \colon \mathbf{R}^3 \to \mathbf{R}^3$ by $\mathfrak{s}_e(x, y, z) = (y, x, z)$ and $\mathfrak{s}_\pm \colon \mathbf{R}^3 \to \mathbf{R}^3$ by $\mathfrak{s}_\pm(x, y, z) = (-x, -y, z)$, observe that if ξ_t is a solution to (2.1) then so is $\mathfrak{s}_e(\xi_t)$ and $\mathfrak{s}_\pm(\xi)$. This implies that $\Gamma^-_{u,v} = \mathfrak{s}_\pm(\Gamma^+_{u,v})$ and $\Gamma^+_{u,v} = \mathfrak{s}_e(\Gamma^+_{v,u})$. This implies that if μ is an invariant measure for P_t then necessarily $\mu\mathfrak{s}_e^{-1}$ and $\mu\mathfrak{s}_\pm^{-1}$ are also invariant measure for P_t.

The situation for the stochastic dynamics given in (2.6) is the same for $\mathfrak{s}_\pm$ but depends on the choice of σ_1 and σ_2 for $\mathfrak{s}_e$. In all cases $\mathfrak{s}_\pm(\xi_t^\varepsilon)$ is a solution (for a different Brownian motion) if ξ_t^ε is a solution. However $\mathfrak{s}_e(\xi_t^\varepsilon)$ is again a solution if ξ_t^ε is only when $\sigma_1 = \sigma_2$. In any case, we have the following observation which we desperate as a proposition for future reference.

Proposition 7.1. *Assume that at least two of $\{\sigma_1, \sigma_2, \sigma_3\}$ are strictly positive. Let $\mathfrak{s} \colon \mathbf{R}^3 \to \mathbf{R}^3$ be a map such that $\mathfrak{s}(\xi_t^\varepsilon)$ is a solution (for possibly a different Brownian motion) whenever ξ_t^ε is a solution, then*

$\mu^\varepsilon = \mu^\varepsilon \mathfrak{s}^{-1}$ *where* μ^ε *is the unique invariant measure of* P_t^ε *guaranteed by Theorem 3.2.*

Proof of Proposition 7.1. As before it is clear that $\mu^\varepsilon \mathfrak{s}^{-1}$ is again an invariant measure, however we know that μ^ε is the unique invariant measure given the assumptions on the σ's. Hence we conclude that $\mu^\varepsilon \mathfrak{s}^{-1} = \mu$. □

7.3. **Averaging along trajectories.** Given a function $\psi : \mathbf{R}^3 \to \mathbf{R}$, we will define

$$(\mathcal{A}\psi)(\xi) = \lim_{t\to\infty} \frac{1}{t}\int_0^t (\psi \circ \varphi_s)(\xi) ds \tag{7.2}$$

where $\xi_0 = \xi$. Notice that $\mathcal{A}\psi$ is again a function from $\mathbf{R}^3 \to \mathbf{R}$ and that it is constant the connected components of the on level sets of (u, v).

7.3.1. *Averaging when* $u \neq v$. Let $(u, v) = \Phi(\xi)$. If $u \neq v$ then ξ lies on a periodic orbit of finite period. Letting τ 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 θ. 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

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

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

$$\Gamma^+_{u,v} = \left\{\gamma_{u,v}(\theta) \colon \theta \in [0, 2\pi]\right\}$$

where $\gamma_{u,v}(\theta) \stackrel{\text{def}}{=} \big(x(\theta), y(\theta), z(\theta)\big)$. As already discussed depending on weather $u > v$ 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

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

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

$$\nu^+_{u,v}(dx\,dy\,dz) = \int_0^{2\pi} \frac{K_{u,v}}{|\cos(\phi_{u,v}(\theta))|}\delta_{\gamma_{u,v}(\theta)}(dx\,dy\,dz)d\theta \tag{7.3}$$

where $K^{-1}_{u,v} = \int_0^{2\pi}\frac{1}{|\cos(\phi_{u,v}(\theta))|}d\theta$. We define $\nu^-_{u,v}(dx\,dy\,dz) = \nu^+_{u,v}\mathfrak{s}^{-1}_{\pm}$. For $u = v$, we define $\nu^{\pm}_{u,u}(dx\,dy\,dz) = \delta_{(0,0,\pm\sqrt{u})}(dx\,dy\,dz)$. 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 $\psi : \mathbf{R}^3 \to \mathbf{R}$ and $(x, y, z) \in \mathbf{R}^3$ such that $|x| \neq |y|$ one has

$$(\mathcal{A}\psi)(\xi) = \int_{\mathbf{R}^3}\psi(\eta)\nu^s_{u,v}(d\eta) \tag{7.4}$$

where $\xi = (x, y, z)$, $(u, v) = \Phi(\xi)$ and $\mathfrak{s} = \mathfrak{sn}(\xi)$ where $\mathfrak{sn}$ was defined in (7.1).

7.3.2. *Averaging near the diagonal.*

Proposition 7.2. *Let $\psi : \mathbf{R}^3 \to \mathbf{R}$ be a localy bounded function. If $\delta = 1 - (u \wedge v)/(v \vee u)$ then as $|u - v| \to 0$ (and hence $\delta \to 0$) one has*

$$(\nu\psi)(u, v, \sigma) = \tfrac{1}{2}(\psi(0, 0, \sqrt{u \vee v}) + \psi(0, 0, -\sqrt{u \vee v})) + o(1) \tag{7.5}$$

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 $v \to v$ one has

$$(\nu\psi)(u, v, \sigma) = \frac{\sqrt{3}C_u(\psi)}{8}\frac{1}{|\ln(1-r)|} + o\big(|\ln(1-r)|^{-1}\big)$$

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

Proof. We will begin by exploring the asymptotics of the constant $K_{u,v}$. Making the change of variables $\beta^2 = \sqrt{r}\alpha^2$, one has

$$K_{u,v} = \int_0^1 \frac{4}{\sqrt{(1-r\alpha^2)(1-\alpha^2)}} d\alpha = \frac{1}{r^{\frac{1}{4}}} \int_0^{\sqrt{r}} \frac{1}{\sqrt{(1-r^{\frac{1}{2}}\beta^2)(1-r^{-\frac{1}{2}}\beta^2)}} d\beta$$
$$= \frac{4}{r^{\frac{1}{4}}} \int_0^{\sqrt{r}} \frac{1}{1-\beta^2} \frac{1}{\sqrt{1-\gamma(r)\frac{\beta^2}{(1-\beta^2)^2}}} d\beta$$

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

$$0 \le \gamma(r)\frac{r^2}{(1-r^2)^2} \le \frac{1}{4}$$

and hence we have

$$\frac{4}{r^{\frac{1}{4}}}\text{arctanh}\sqrt{r} \le K_{u,v} \le \frac{8}{\sqrt{3}}\frac{1}{r^{\frac{1}{4}}}\text{arctanh}\sqrt{r} \tag{7.6}$$

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

$$\lim_{|u-v|\to 0} \frac{K_{u,v}}{|\ln(1-r)|} = \frac{4}{\sqrt{3}} \tag{7.7}$$

Now

$$(\nu\psi)(u,v,\sigma) = K_{u,v} \int_0^{2\pi} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|} d\theta \tag{7.8}$$

As $|u-v| \to 0$ this integral concentrates around two points θ equal $\pi/2$ and $3\pi/2$ since around these points $|\cos(\phi(\theta))| \to 0$ as $|u-v| \to 0$. As 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 $\frac{1}{2}\psi(0,0,\sqrt{u\vee v})$:

$$K_{u,v}\int_0^{\pi} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|} d\theta = K_{u,v}\int_0^{\pi/2-\delta} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|} d\theta$$
$$+ K_{u,v}\int_{\pi/2-\delta}^{\pi/2+\delta} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|} d\theta + K_{u,v}\int_{\pi/2+\delta}^{\pi} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|} d\theta$$

The remaining half of the integral in 7.8 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 set $r = \frac{u\wedge v}{v \vee u}$ then

$$K_{u,v}\Big|\int_0^{\pi/2-\delta} \tfrac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|}d\theta\Big| \le \tfrac{K_{u,v}\|\psi\|_\infty}{r^{\frac14}}\int_0^{r^{\frac14}} \tfrac{1}{1-\beta^2}\frac{1}{\sqrt{1-\gamma(r)\frac{\beta^2}{(1-\beta^2)^2}}}d\beta$$
$$\le CK_{u,v}\|\psi\|_\infty \mathrm{arctanh}(r^{\frac14})$$

By the asymptotics on $K_{u,v}$ given in in (7.6), this goes to zero as $r \to 1$.

Now as $r \to 1$ one has

$$K_{u,v}\int_{\pi/2-\delta}^{\pi/2+\delta} \frac{\psi(x(\theta),y(\theta),z(\theta))}{|\cos(\phi(\theta))|}d\theta \approx \psi(0,0,\sqrt{u\vee v})\frac{4K_{u,v}}{\sqrt{3}}\int_{r^{\frac14}}^{r^{\frac14}} \tfrac{1}{1-\beta^2}d\beta$$
$$\approx \tfrac12 \psi(0,0,\sqrt{u \vee v})$$

The last conclusion follows directly from the assumed finiteness of C_u and the asymptotics of $K_{u,v}$ as $|u-v| \to 0$ given in (7.7). ☐

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

We summarize this discussion in the following result.

Proposition 7.3. *The set of ergodic invariant measure of* (2.1) *consists precisely of*

$$\left\{\nu^+_{u,v}, \nu^-_{u,v} : u,v > 0\right\}.$$

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

$$\nu_{u,v}(d\xi) = \frac12 \nu^+_{u,v}(d\xi) + \frac12 \nu^-_{u,v}(d\xi) \tag{7.9}$$

where $\nu^{\pm}_{u,v}(d\xi)$ we defined in 7.3 and the text below it.

The following corollary of Proposition 7.3 will be central to the proof of the convergence of μ^ε to a unique limiting measure.

Corollary 7.4. *Any invariant measure m for* (2.1) *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 γ on $[0,\infty)^2$. Furthermore the measure γ is unique. Conversely, a measure which is invariant for (2.1) *and satisfies $m\mathfrak{s}_\pm^{-1} = m$ is uniquely specified by the measure $m\Phi^{-1}$.*

Proof of Corollary 7.4. There ergodic decomposition theorem [FKS87] implies that there exists a unique pair of measure (γ^+, γ^-) so that the total mass of $\gamma^+ + \gamma^-$ is one and

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

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 for all choices of non-negative 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 γ is a probability measure and that

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

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

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

which implies that

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

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

8. The Limiting Fast Semigroup

We begin with a small detour to think about the limiting dynamics. it 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 fix point on the z-axis.

Recalling the definition of $\nu_{u,v}$ from (7.9), for $\phi\colon \mathbf{R}^3 \to \mathbf{R}$ we define $\nu\phi$ by

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

Recalling the definition of Φ which maps ξ to (u,v) from (2.4), note that for any $\rho : \mathbf{R}^2_+ \to \mathbf{R}$,

$$\nu(\rho\circ\Phi)(u,v) = \rho(u,v)\,. \tag{8.1}$$

Recalling the definition of $\widetilde{P}^\varepsilon_t$ from (6.2), Q_t from (6.5) and let λ be the unique invariant measure of Q_t guarantied by Theorem 6.8. For $\phi\colon \mathbf{R}^3 \to \mathbf{R}$ we define

$$(\widetilde{P}_t\phi)(\xi) = (Q_t\nu\phi)\circ\Phi(\xi) \tag{8.2}$$

Remark 8.1. *If ϕ 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 $m\widetilde{P}^\varepsilon_t\phi \to m\widetilde{P}^\phi_t$ as $\varepsilon \to 0$. If one neither starts with initial data which has this symmetry nor a symmetric test function, then things are more complicated. The orbit will average with respect to only one of the two measure: $\nu^+_{u,v}$ or $\nu^-_{u,v}$. For definiteness assume that $\sigma_1 > \sigma_2 > \sigma_3 = 0$, $u > v$ and that we are on a 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 7.2.) 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 $\widetilde{P}_t$ semigroup in the non-symmetric setting, it seems we need to add a sequence of independent Bernoulli random variables to make decision of weather to be averaging with respect to the $+$ or the $-$ orbit. Since we are primarily interested in the structure of the invariant measure we have not tried to make the picture rigorous.*

Let λ be the unique invariant of Q_t and define $\mu = \lambda\nu$. Observe that μ is invariant under $\widetilde{P}_t$ because for any bounded $\phi\colon \mathbf{R}^3 \to \mathbf{R}$ one has

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

Here the first equality is by definition, the second follows from (8.1), the third from the invariance of λ under Q_t and the last from the definition of μ.

9. Convergence of $(U^\varepsilon, V^\varepsilon)$ towards (U,V)

9.1. **Tightness.** Let us rewrite (6.3) in the form

$$\begin{cases} dU^\varepsilon_t = (C_u - 2U^\varepsilon_t)dt + dM^\varepsilon_t \\ dV^\varepsilon_t = (C_v - 2V^\varepsilon_t)dt + dN^\varepsilon_t, \end{cases} \tag{9.1}$$

where $\{M^\varepsilon_t,\, t\geq 0\}$ and $\{N^\varepsilon_t,\, t\geq 0\}$ are continuous local martingales such that

$$\frac{d}{dt}\langle M^\varepsilon\rangle_t \leq CU^\varepsilon_t, \quad \frac{d}{dt}\langle N^\varepsilon\rangle_t \leq CV^\varepsilon_t, \tag{9.2}$$

where C_u, C_v and C are three positive constants.

We want to show

Proposition 9.1. *Suppose that*

$$\sup_{\varepsilon>0}\mathbf{E}\left[(U^\varepsilon_0)^2 + (V^\varepsilon_0)^2\right] < \infty\,.$$

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

In light of (9.1) and (9.2), the following needed first step follows from Lemma 5.3.

Lemma 9.2. *Under the condition of Proposition 9.1,*

$$\sup_{\varepsilon>0}\sup_{t\geq 0}\mathbf{E}\left[(U^\varepsilon_t)^2 + (V^\varepsilon_t)^2\right] < \infty.$$

We can now proceed with the proof of tightness.

Proof of Proposition 9.1. We prove tightness of U^ε only, V^ε being treated completely similarly. We have

$$U^\varepsilon_t = U^\varepsilon_0 e^{-2t} + e^{-2t}\int_0^t e^{2s}dM^\varepsilon_s.$$

Clearly the first term on the right is tight in $C([0,\infty))$, since the collection of $\mathbf{R}$–valued r.v.'s U^ε_0 is tight. We only need check tightness in $C([0,\infty))$ of the process $W^\varepsilon_t := \int_0^t e^{2s}dM^\varepsilon_s$. Since $W^\varepsilon_0 = 0$, 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, η and $\eta' > 0$, there exists $\delta \in (0,1)$ such that for all $\varepsilon > 0$, $0 \le t \le T - \delta$,

$$\frac{1}{\delta}\mathbf{P}\left(\sup_{t\le s\le t+\delta} |W_s^\varepsilon - W_t^\varepsilon| \ge \eta\right) \le \eta'. \tag{9.3}$$

Combining Chebycheff and Burkholder–Davis–Gundy inequalities, we deduce that (we use below the result from Lemma 9.2)

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

from which (9.3) follows if we choose $\delta = e^{-4T}\eta^2\sqrt{\eta'/\bar{C}}$. □

9.2. **Tighness of λ^ε.** Since $(U_t^\varepsilon, V_t^\varepsilon)$ is not a Markov process it does not have an invariant measure, However the projection $\lambda^\varepsilon = \mu^\varepsilon\Phi^{-1}$ stationary measure have been defined using μ^ε which is the unique invariant measure of the Markov process ξ_t^ε. We now establish the following tightness result:

Lemma 9.3. *The sequence of measure $\{\lambda^\varepsilon : \varepsilon > 0\}$ is tight on the space $(0,\infty)\times(0,\infty)$.*

Remark 9.4. *We emphasis that Lemma 9.3 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 $\varepsilon > 0$ there exists a $r > 0$ so that*

$$\inf_{\varepsilon>0} \lambda^\varepsilon([\tfrac{1}{r}, r]\times[\tfrac{1}{r}, r]) > 1 - \varepsilon$$

The following result which implies the tightness at infinity follows immediately from the definition of λ^ε, the definition of Φ and Corollary 5.4.

Lemma 9.5. *For any $p \ge 1$, there exists a $C(p) > 0$ so that*

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

We now handle the boundary at zero.

Lemma 9.6. *Let ζ_t be a Markov process and f and g two real-values functions on the state space of ζ_t satisfying $0 \le g(\zeta_t) \le f(\zeta_t)$ 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 μ is any invariant measure of ζ_t with $\mu[\, f^2\,] = \int f^2(\zeta)d\zeta < \infty$, then for any $\delta \in (0,1)$.

$$\mu\{f \leq \delta\} \leq \frac{\mu[\, f^2\,]}{a\,|\log\delta|}\,. \tag{9.4}$$

Proof of Lemma 9.6. Defining

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

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

observe that $I(x) = \int_1^x \phi(z)\,dz$ and $H(x) = \int_0^x I(z)dz$ and that ϕ, I and H are well defined and positive on the intervals $(0,\infty)$,$(0,\infty)$ and $[0,\infty)$ respectively. Taking ζ_0 distributed according to μ, noticing that since $H(x) < 2x^2$ for $x \geq 0$, and setting $X_t = f(\zeta_t)$ for notational convenience, we have that

$$\mu\big[\,\mathbf{E}_{\zeta_0} H(X_t)\,\big] = \mu\big[\,\mathbf{E}_{\zeta_0}(H \circ f)(\zeta_t)\,\big] = \mu\big[\,H \circ f\,\big] < \infty \tag{9.5}$$

Now observe that

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

where M_t is the Martingale defined by $dM_t = c\sqrt{g(\zeta_t)}I(X_t)dW_t$. Since $g(\zeta_t)\phi(X_t) \geq 0$, we conclude that

$$a\int_0^t \mathbf{E}_{\zeta_0}[I(X_s)]\,ds \leq \mathbf{E}_{\zeta_0} H(X_t) - H(X_0) + \int_0^t \mathbf{E}_{\zeta_0}\, X_s I(X_s)\,ds\,.$$

Now integrating over the initial conditions ζ_0 (which were distributed according to μ), we see that H terms are equal by the stationarity embodied in (9.5) (and hence they cancel) and that

$$a\,\mu[I \circ f] \leq \mu[f(I \circ f)]$$

and

$$a\,\mu\big[\,|\log f|\,\mathbf{1}\{f \leq 1\}\,\big] \leq a\,\mu[\,I \circ f\,] \leq \mu[\,f(I \circ f)\,]$$

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

$$a\,\mu\{f \le \delta\} = a\,\mu\big[\,\{|\log f| \ge |\log \delta|\} \cap \{f \le 1\}\,\big] \le \frac{a\,\mu\big[\,|\log f|\,\mathbf{1}\{f \le 1\}\,\big]}{|\log \delta|} \le \frac{\mu\big[\,f(I \circ f)\,\big]}{|\log \delta|}$$

□

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

Corollary 9.7. *There exists a constant $C > 0$ so that for any $\delta \in (0,1)$*

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

Proof of Lemma 9.3. The result follows immediately by combining Lemma 9.5 and Corollary 9.7. □

9.3. **Convergence of Quadratic variation.** Now that we know that the collection $\{(U^{\varepsilon}_t, V^{\varepsilon}_t),\, t \ge 0\}_{\varepsilon>0}$ is tight, in view of Theorem 6.5, the weak uniqueness result for (6.4), and comparing (6.3) and (6.4), the weak convergence $(U^{\varepsilon}, V^{\varepsilon}) \Rightarrow (U, V)$ will follow from the convergence of the quadratic variations of U^{ε} and V^{ε} to those of U and V, which will be proved in the next Lemma.

For each $M > 0$, let

$$\kappa^{\varepsilon}_M := \inf\{t > 0,\ U^{\varepsilon}_t \vee V^{\varepsilon}_t > M\}.$$

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

$$\kappa^{\varepsilon}_M \Longrightarrow \kappa_M \quad \text{as } \varepsilon \to 0$$

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

In particular

$$\liminf_{\varepsilon \to 0} \mathbf{P}(\kappa^{\varepsilon}_M > t) \ge \mathbf{P}(\kappa_M > t).$$

Clearly for all $t > 0$,

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

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

Lemma 9.8. *Let ν_ε be any sequence of tight probability measures on $\mathbf{R}^3$ and let $(\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon, \widetilde{Z}_t^\varepsilon)$ be the solution to (6.1) with $(\widetilde{X}_0^\varepsilon, \widetilde{Y}_0^\varepsilon, \widetilde{Z}_0^\varepsilon)$ distributed as ν_ε. Then for any $t > 0$, as $\varepsilon \to 0$,*

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

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

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

for all $M > 0$.

$t > 0$ 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 $0 \le n < N_\delta$ 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 (2.1) 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 clear

$$\int_0^{t\wedge\kappa_M^\varepsilon} (\widetilde{Z}_s^\varepsilon)^2 ds = \varepsilon \sum_{n=0}^{N_\delta - 1} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z_s^\varepsilon)^2 ds = \Phi_{\varepsilon,\delta} + \Xi_{\varepsilon,\delta} \tag{9.6}$$

where

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

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

To control the error term observe that

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

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

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

Hence if we choose

$$\delta = C_M^{-1} \varepsilon \log(1/\varepsilon), \tag{9.7}$$

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

We now further divide Φ_ε (δ having been suppressed) depending on where in phase space the starting point $(X^{(n)}_{t_n/\varepsilon}, Y^{(n)}_{t_n/\varepsilon}, Z^{(n)}_{t_n/\varepsilon})$ lies the region where $|U^\varepsilon_{t_n} - V^\varepsilon_{t_n}|$ 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

$$\begin{aligned} A_{\varepsilon,\rho} =& \delta \sum_{n=0}^{N_\delta - 1} \frac{\varepsilon}{\delta} \chi_\rho(U^\varepsilon_{t_n} - V^\varepsilon_{t_n}) \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z^{(n)}_s)^2 ds, \\ B_{\varepsilon,\rho} =& \delta \sum_{n=0}^{N_\delta - 1} \frac{\varepsilon}{\delta} \bar{\chi}_\rho(U^\varepsilon_{t_n} - V^\varepsilon_{t_n}) \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z^{(n)}_s)^2 ds\,. \end{aligned}$$

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 $(t_{n+1} - t_n)/\varepsilon = \delta/\varepsilon \to \infty$ 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 7.2 shows that the $\mathcal{A}(z^2)(u,v) \to u = v$ as $|u - v| \to 0$. To understand why $\frac{\varepsilon}{\delta} \int_{t_n/\varepsilon}^{t_{n+1}/\varepsilon} (Z^{(n)}_s)^2 ds \to U_{t_n} \wedge V_{t_n}$ one needs to recall the discussion from Section 7. The deterministic orbits when $u = v$ consist of heteroclinic orbits connecting the fix points at $(0, 0, \sqrt{u})$ and $(0, 0, -\sqrt{u})$. Since the time to reach the fix points on these orbits in infinite, it is not surprising that for $|u-v|$ small the periodic orbits 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 (7.3) concentrates around $\theta \sim \pi/2, 3\pi/2$, which corresponds to the fix 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 fix points since the time to reach the neighborhood of the fix point is small relative to the time it will take to leave that neighborhood once it has arrived. This idea will be made quantitative below.

Hence we define

$$\hat{A}_{\varepsilon,\rho} = \delta \sum_{n=0}^{N_\delta - 1} \chi_\rho(U^\varepsilon_{t_n} - V^\varepsilon_{t_n})\,(U^\varepsilon_{t_n} \wedge V^\varepsilon_{t_n}) \quad \text{and} \quad \Theta_{\varepsilon,\rho} = A_{\varepsilon,\rho} - \hat{A}_{\varepsilon,\rho}$$

$$\hat{B}_{\varepsilon,\rho} = \delta \sum_{n=0}^{N_\delta - 1} \bar{\chi}_\rho(U^\varepsilon_{t_n} - V^\varepsilon_{t_n})\gamma_\varepsilon(U^\varepsilon_{t_n}, V^\varepsilon_{t_n})\mathcal{A}(z^2)(U^\varepsilon_{t_n}, V^\varepsilon_{t_n}) \quad \text{and} \quad \Upsilon_{\varepsilon,\rho} = B_{\varepsilon,\rho} - \hat{B}_{\varepsilon,\rho}$$

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

$$\hat{\tau}_{\beta,\rho} = \min\Big\{\tau(u,v) : |u - v| \le \rho, \beta \le u \wedge v \le u \vee v \le M\Big\},$$

$$\Psi_{\alpha,\rho} = 4\sqrt{M}\int_0^{1-\alpha} \frac{1}{|1 - a^2|} da\,.$$

The utility of Ψ is the following which can be deduced from Section 7.3

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

Now

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

On the other hand, since

$$\Upsilon_{\varepsilon,\rho} = \varepsilon \sum_{n=0}^{N_\delta - 1} \bar{\chi}_\rho(U^\varepsilon_{t_n} - V^\varepsilon_{t_n}) \int_{\frac{t_n}{\varepsilon} + \left[\frac{\delta}{\varepsilon\tau^\varepsilon_n}\right]\tau^\varepsilon_n}^{\frac{t_{n+1}}{\varepsilon}} (Z^{(n)}_r)^2 dr\,,$$

we have the inequality

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

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

Now applying Lemma 9.9 below, we see that for all $\rho > 0$, as $\varepsilon \to 0$ one has

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

Now notice that as $\rho \to 0$, $\hat{A}_\rho + \hat{B}_\rho$ 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 7.2, 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 $\rho \to 0$.

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

$$F(\hat{A}_{\varepsilon,\rho} + \hat{B}_{\varepsilon,\rho} - K_{\varepsilon,\rho,\alpha,\beta} - L_{\varepsilon,\delta}) \le F(A_{\varepsilon,\rho} + B_{\varepsilon,\rho}) \le F(\hat{A}_{\varepsilon,\rho} + \hat{B}_{\varepsilon,\rho} + K_{\varepsilon,\rho,\alpha,\beta} + L_{\varepsilon,\delta})$$

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

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

where

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

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

It remains to exploit Lemma 9.10 below to deduce that

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

as $\varepsilon \to 0$. □

Lemma 9.9. *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 $n \to \infty$, $F_n \to F$ uniformly on each compact subset of $\mathcal{X}$. Then $F_n(X_n) \Rightarrow F(X)$, as $n \to \infty$.*

Proof of Lemma 9.9. 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

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

The first term of the righthand side can be made arbitrarily small by choosing ε 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 $n \to \infty$. □

Lemma 9.10. *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*

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

for any continuous, bounded and increasing function F.

Proof of Lemma 9.10. It is plain that the condition of the Lemma implies that $\mathbf{E}[F(X_n)] \to \mathbf{E}[F(X)]$ 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

$$\limsup_{n \to \infty} \mathbf{P}(X_n > M) \leq \lim_{n \to \infty} \mathbf{E}[F_M(X_n)] \\ = \mathbf{E}[F_M(X)].$$

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 $\mathbf{E}[F(X_n)] \to \mathbf{E}[F(X)]$ 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. □

10. Convergence of the invariant measures

Recall that for each $\varepsilon > 0$ μ^ε denotes the unique invariant measure of ξ^ε, and that λ denotes the unique invariant measure of the diffusion process (U_t, V_t) or equivalently of its semigroup Q_t.

Theorem 10.1. *If either one of the two assumptions of Theorem 6.5 is satisfied, then μ^ε converges weakly to μ as ε tends to 0, where μ is the unique invariant measure of the deterministic dynamics prescribed by (2.1), which satisfies both $\mu = \mu\mathfrak{s}_\pm^{-1}$, and $\mu\Phi^{-1} = \lambda$.*

Proof. The fact that any accumulation point of the collection $\{\mu^\varepsilon, \varepsilon > 0\}$ satisfies $\mu = \mu\mathfrak{s}_\pm^{-1}$ follows from Proposition 7.1. Corollary 7.4 states that at most one invariant measure of (2.1) satisfies both $\mu = \mu\mathfrak{s}_\pm^{-1}$ and $\mu\Phi^{-1} = \lambda$.

From Corollary 5.4 we know the collection $\{\mu^\varepsilon, \varepsilon > 0\}$ is tight. Consequently, there exists a sequence $\varepsilon_n \to 0$ and a measure $\tilde{\mu}$, such that $\mu^{\varepsilon_n} \Rightarrow \tilde{\mu}$.

Fix an arbitrary $t > 0$. If we initialize ξ^{ε_n} with its invariant measure μ^{ε_n}, then both marginal laws of the pair $(\xi_0^{\varepsilon_n}, \xi_t^{\varepsilon_n})$ equal μ^{ε_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 ξ.

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 μ^{ε_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 λ^{ε_n}. If we define $\tilde{\lambda} = \Phi^{-1}\tilde{\mu}$, then since $((U_0^\varepsilon, V_0^\varepsilon), (U_t^\varepsilon, V_t^\varepsilon)) \Rightarrow ((U_0, V_0), (U_t, V_t))$, Φ is continuous and from Lemma 9.3 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 measure for Q_t. Since from Theorem 6.8, Q_t has the unique invariant measure λ, we conclude that $\tilde{\lambda} = \lambda$.

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

Acknowledgements: The authors are indebted to A. Veretennikov, who pointed out the results in [GM01]. 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.

[FKS87] S. W. Fomin, I. P. Kornfel′d, and J. G. Sinai. *Teoria ergodyczna*. Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1987. Translated from the Russian by Jacek Jakubowski.

[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 Szücs.

[GM01] István Gyöngy and Teresa Martínez. On stochastic differential equations with locally unbounded drift. *Czechoslovak Math. J.*, 51(126)(4):763–783, 2001.

[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.

[Kuk10a] Sergei B. Kuksin. Damped-driven KdV and effective equations for long-time behaviour of its solutions. *Geom. Funct. Anal.*, 20(6):1431–1463, 2010.

[Kuk10b] Sergei B. Kuksin. Damped-driven KdV and effective equations for long-time behaviour of its solutions. *Geom. Funct. Anal.*, 20(6):1431–1463, 2010.

[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.

[Rue82] David Ruelle. Small random perturbations of dynamical systems and the definition of attractors. *Comm. Math. Phys.*, 82(1):137–151, 1981/82.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[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.

[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.