Instantons as a Means to Probe Chaotic Attractors

Andre Souza

October 15, 2015

1 Introduction

In systems with chaos one is often left bewildered on how to make sense of its dynamics. The extreme sensitivity to initial conditions renders our quantitative predictions useless, and yet there are often qualitative features that are robust to our ignorance. Even though chaotic dynamics are described by deterministic procedures, its unpredictability in the long run forces us to look at statistical quantities of interest; means, variances, correlations, or even the distribution of the state variables.

Normally one calculates these chaotic statistics by running long simulations, a brute force approach. Ideally one would like a faster method of obtaining statistics and, more ambitiously, understanding the structures that lead to the observed chaotic statistics.

Lately there are programs that try to exploit small noise limits and large deviation theory in order to provide insight into the equations of motion [4]. Here one first formulates a stochastic version of the state equations of interest, thus recasting the problem as one of stochastic differential equations. Fokker-Planck equations and path integrals now come into play as tools of investigation.

The goal of this WHOI: GFD 2015 project is to examine the stochastic version of chaotic deterministic systems in order to see whether or not the noiseless limit may be exploited to further understand the underlying deterministic dynamics. We will look at chaotic systems and present instanton calculations as well as their interpretation. In light of the results we comment on the applicability of the instanton formulation to turbulent flows.

2 Background

The necessary background to understand the instanton approach requires an understanding of random variables, multivariable calculus, recurrence relations, differential equations, asymptotics, and calculus of variations. The details are technical, but absolutely necessary to grasp the instanton formulation and interpretation. The excellent review article by Grafke et al. [3] covers the basics, but we shall go over them in more detail.

2.1 Path integrals

During the 2015 summer at the WHOI: GFD program we learned how to make sense of stochastic differential equations with delta correlated Gaussian white noise,

$$\dot{x} = f(x) + \epsilon\xi,$$

where $\epsilon \in \mathbb{R}$ is the noise strength, by considering it as the $N \to \infty$ limit of the Euler recursion relation

$$X^n = X^{n-1} + \Delta t f(X^{n-1}) + \epsilon\sqrt{\Delta t} G^n.$$

Here $\Delta t = T/N$, T is the "endtime", and each G^n for $n = 1, 2, ..., N$ is a normally identically distributed Gaussian random variable with mean zero and variance one. We think of $\vec{X} = (X^1, ..., X^N)$ as the path, an element X^n as a position, and X^0 specifically as the starting position. Although each step of our recursion relation is a Gaussian random variable with a mean (or drift) given by the deterministic trajectory $X^{n-1} + \Delta t f(X^{n-1})$ and variance $\epsilon\sqrt{\Delta t}$, the statistics of X^n are highly influenced by its history and the form of f, possibly leading to deviations from Gaussian statistics.

For a given stochastic process we are generally interested in observables that depend in some way on the "path" $\vec{X}$, for example, the distribution of the position at the endtime X^N. We would like an expression for the density of $\vec{X}$ in order to more conveniently calculate such quantities. This may be done by observing that the recursion relation is a change of variables from Gaussian random variables $\vec{G} = (G^1, G^2, ..., G^N)$ to the path $\vec{X} = (X^1, X^2, ..., X^N)$ given by

$$G^n = \left(\frac{X^n - X^{n-1}}{\Delta t} - f(X^{n-1})\right)\frac{\sqrt{\Delta t}}{\epsilon}$$

The Jacobian of the transformation is a lower triangular matrix

$$\left[\frac{\partial \vec{G}}{\partial \vec{X}}\right]_{ij} = \begin{cases} \frac{1}{\epsilon\sqrt{\Delta t}} & \text{if } i = j \\ -\frac{1}{\Delta t} - \partial_{x^{i-1}} f & \text{if } i - 1 = j \\ 0 & \text{otherwise} \end{cases}$$

from whence we can calculate the determinant as the product of the diagonals

$$\det\left[\frac{\partial \vec{G}}{\partial \vec{X}}\right] = \left(\frac{1}{\epsilon\sqrt{\Delta t}}\right)^N.$$

We can now leverage our knowledge of Gaussian distributions and use our change of variables calculation to give an expression for the probability density ρ in terms of the path,

$$\rho(\vec{g})dV = e^{-\frac{1}{2}\sum_{n=1}^N (g^n)^2} \prod_{n=1}^N \left[\sqrt{\frac{1}{2\pi}} dg^n\right]$$

$$\Leftrightarrow$$

$$\rho(\vec{g}(\vec{x}))dV = e^{-\frac{1}{2\epsilon^2}\sum_{n=1}^N \left(\frac{x^n - x^{n-1}}{\Delta t} - f(x^{n-1})\right)^2 \Delta t} \prod_{n=1}^N \left[\sqrt{\frac{1}{2\pi\Delta t\epsilon^2}} dx^n\right].$$

We use the lower case to denote a specific realization of the random variable. An interesting feature of this reformulation is that the deterministic trajectory, given by $x^{n+1} = x^n + \Delta t f(x^n)$ is the path that is given the most amount of weight and where the deviations from determinism are penalized by an exponentially weighted factor that is inversely proportional to the square of the noise strength ϵ. In the small noise limit this implies that nondeterministic paths are highly unlikely.

The path integral is the "$N \to \infty$" limit of our finite path space integral, leading to the density

$$e^{-\frac{1}{2\epsilon^2}\sum_{n=1}^{N}\left(\frac{x^n - x^{n-1}}{\Delta t} - f(x^{n-1})\right)^2 \Delta t} \prod_{n=1}^{N}\left[\sqrt{\frac{1}{2\pi\Delta t\epsilon^2}}dx^n\right] \text{“}\overset{\lim_{N\to\infty}}{=}\text{”} e^{-\frac{1}{2\epsilon^2}\int_0^T (\dot{x} - f(x))^2 dt}\mathcal{D}[x(t)]$$

where the differential element $\mathcal{D}[x(t)]$ has the normalization factor buried in it. The action in the argument of the exponential is known as the Friedlen-Wentzell action and will be the central object of concern for the calculations in this document. This will be expanded upon later.

If we are only interested in the distribution of our path at the endtime $x(T)$, we may formally obtain it by considering

$$\begin{aligned}
\rho(x(T))dx(T) &= \lim_{N\to\infty}\left[\rho(x^N, T, N)\sqrt{\frac{1}{2\pi\epsilon^2\Delta t}}dx^N\right] \\
\rho(x^N, T, N) &= \prod_{n=1}^{N-1}\left[\sqrt{\frac{1}{2\pi\Delta t\epsilon^2}}\int_{\mathbb{R}} dx^n\right] e^{-\frac{1}{2\epsilon^2}\sum_{n=1}^{N}\left(\frac{x^n - x^{n-1}}{\Delta t} - f(x^{n-1})\right)^2 \Delta t}.
\end{aligned}$$

We shall see that we can sometimes get away with performing a simpler calculation but at the cost losing the normalization factor. This formulation for the distribution at the endtime can be directly compared to the usual Fokker-Planck evolution for the density,

$$\partial_t \rho = -\partial_x(f\rho) + \frac{\epsilon^2}{2}\partial_{xx}\rho.$$

The Fokker-Panck equation states that the evolution of the density is one that is advected by the determistic equations of motion and diffused due to the noise, whereas the path integral states that the distribution at the endtime comes from an exploration of all possible paths weighted most heavily be the deterministic trajectories. These are two different but complementary interpretations for the evolution of the density.

We will never make use of the limit definition in order to calculate the endtime density, but it is illuminating to see what such a calculation would entail. If one attempts to calculate the density at the endtime using the limit definition of the path integral formulation, the following integral for the first time-step arises

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\epsilon^2}\left(\frac{x^1 - x^0}{\Delta t} - f(x^0)\right)^2 - \frac{1}{2\epsilon^2}\left(\frac{x^2 - x^1}{\Delta t} - f(x^1)\right)^2\right] dx^1.$$

The role of f manifests itself in this case. Although we are interested in the $\Delta t \to 0$ limit, nonlinearity in f may dominate the integral, changing the Gaussian statistics. Depending

on the system of interest this may or may not be an impediment to progress for analytic calculations. This is an explicit manifestation of how a path is dependent on its history. Thus we are led to different ways of calculating or estimating the integral.

2.2 Large deviation theory

Traditional large deviation theory concerns itself with the probability that sums of independent identically distributed random variables deviate from the mean by a large value. Namely, let $X^1, ..., X^N$ be independent identically distributed random variables and let S_N denote their sum. If the moment generating function $M(t) = \mathbb{E}e^{tX}$ is finite within some neighborhood of $t = 0$ and $0 < \text{var}(X)$, then for $a > \mathbb{E}X$

$$\frac{1}{N} \log \mathbb{P}(S_N > Na) \to -\mathcal{I}(a) \text{ as } N \to \infty$$

where $\mathcal{I}(a) = \sup_s [sa - \log(M(t))]$. The $\mathcal{I}(a)$ object is called the rate function.The log of the moment generating function is called the cumulant generating function.

In stochastic differential equations our random variables are no longer independent since they satisfy a Markov property; however, it is still possible that a large deviation principle may be satisfied. We say that a density satisfies a large deviation principle if

$$\rho(a) \sim \exp\left(-\frac{1}{\epsilon^2}\mathcal{I}(a)\right)$$

for some rate function $\mathcal{I}(a)$ in the limit $\epsilon \to 0$. Similar to what happens in the independent identically distributed case the cumulant generating function may be related to the rate function. In the $\epsilon \to 0$ limit the calculation goes as follows

$$\begin{aligned}
\epsilon^2 \log\left\langle e^{\frac{\lambda}{\epsilon^2}X(T)}\right\rangle &\sim \epsilon^2 \log\left[\int_{-\infty}^{\infty} da \exp\left(\frac{\lambda}{\epsilon^2}a + \ln\rho(a)\right)\right] \\
&= \epsilon^2 \log\left[\int_{-\infty}^{\infty} da \exp\left(\frac{\lambda}{\epsilon^2}a - \frac{1}{\epsilon^2}\mathcal{I}(a)\right)\right] \\
&\approx \epsilon^2 \log\left(D \exp\left[\frac{1}{\epsilon^2}\left(\sup_a [\lambda a - \mathcal{I}(a)]\right)\right]\right) \\
&= \epsilon^2 \log(D) + \sup_a [\lambda a - \mathcal{I}(a)] \\
&\approx \sup_a [\lambda a - \mathcal{I}(a)]
\end{aligned}$$

where in the first approximation we used Laplace's method to estimate the integral and picked up an extra constant D, and in the second approximation we assumed that $\epsilon^2 \log(D)$ goes to zero in the limit. As is usual in the case of asymptotics we expect the formula to be very good for small but finite ϵ, even though we formally did the calculation for the limit. Succinctly we may say that the log of the moment generating function is the Fenchel-Legendre Transform of the rate function.

On the other hand we can repeat the same calculation with the path integral formulation. In this case we have that

$$\begin{aligned}\epsilon^2 \log\left\langle e^{\frac{\lambda}{\epsilon^2}x(T)}\right\rangle &= \epsilon^2 \log\left[\int \mathcal{D}[x(t)] \exp\left(\frac{1}{\epsilon^2}\left[\lambda x(T) - A[x]\right]\right)\right] \\ &\approx \epsilon^2 \log\left[D_1 \exp\left(\frac{1}{\epsilon^2}\sup_{x(t)}\left[\lambda x(T) - A[x]\right]\right)\right] \\ &\approx \sup_{x(t)}\left[\lambda x(T) - A[x]\right]\end{aligned}$$

where $A[x] = \frac{1}{2}\int_0^T \left(\dot{x} - f(x)\right)^2 dt$, the first approximation came from using Laplace's method on the functional, D_1 is the constant that comes from our cavalier use of path integrals and Laplace's method, and the last approximation comes from assuming that $\epsilon^2 \log\left(D_1\right) \to 0$ as $\epsilon \to 0$.

Assuming that all the approximations are valid, we may put our two calculations together to arrive at the following relation

$$\sup_a\left[\lambda a - \mathcal{I}(a)\right] = \sup_{x(t)}\left[\lambda x(T) - A[x]\right]$$

when $\epsilon \to 0$. Again we do not expect exact equality for non-zero epsilon but we do expect this expression to be approximately valid.

We now further make the claim that the rate function $\mathcal{I}$ is directly related to the action. Justifying that it is the case follows under the assumptions of Friedlen-Wentzell theory, but here we will give a heuristic argument. Suppose that both sides admit a unique minimizer for some number a^* and some path x^*, then we have that

$$\lambda(x^*(T) - a^*) + \mathcal{I}(a^*) = A[x^*].$$

Furthermore, if $x^*(T) = a^*$, then

$$\mathcal{I}(a^*) = A[x^*].$$

It may be possible for this to occur if $A[x^*]$ is convex and the rate function $\mathcal{I}(a)$ is convex in which case the Fenchel-Legendre Transforms are invertible. From whence we see that it must be the case that $\mathcal{I}(a^*) = A[x^*]$, that is, $x^*(T) = a^*$. If neither are convex then the most we can say is that their convex envelopes are equivalent to one another.

Although here we focused on the distribution at the endtime $x(T)$ we may choose any other observable and follow the same procedure to get a relation between an observable and the minimizer of an action. For example we could choose $x(T)^2$ or the average value of the trajectory $T^{-1}\int_0^T x(t)dt$ in the time interval as our observable. Regardless of the exact choice, the rate function would be related to the minimizer of a functional subject to a constraint.

2.3 Instantons

In the previous section we saw that the minimizers of the Friedlen-Wentzell action play a direct role in determining the probability distribution function of a random variable under

a large deviation assumption. We call the minimizer of this action the instanton. It has the interpretation of being the "most likely path" of a stochastic trajectory conditioned on the starting and ending value.

Although a given realization of a stochastic process bears no resemblance to the instanton, it is still the most likely path in the following sense: If one generates a large ensemble of stochastic trajectories and filters out all the ones that reach within an epsilon window of the target value of an observable (for example all trajectories such that $x(T) \in [a-\epsilon, a+\epsilon]$ for some number a and positive number epsilon), then the instanton trajectory corresponds to the locations in space (for each time) in which the most number of trajectories pass through. Said differently, we divide up space and time into a bunch of little squares and we tally the number of times a square has a trajectory that passes through it. The squares with the most number of tallies is the instanton trajectory, the most likely path.

However we do not (in this document) use this stochastic formulation to calculate instantons, rather, we concentrate purely on finding the infimum of the Friedlen-Wentzell action. To find the infimum of the functional we employ calculus of variations machinery. Although setting a derivative equal to zero only yields a local minimizer, it is often the only way we can make progress in obtaining potential global minimizers. We will impose constraints into the minimization procedure, things like demanding that the final value of our trajectory attains a certain value or perhaps the average value. Both the Lagrangian and Hamiltonian formulations have their uses and in this document we will employ both.

We will now concentrate our efforts on determining trajectories that minimize the action

$$A[x] = \frac{1}{2}\int_0^T \|\dot{x} - f(x)\|^2 \, dt$$

subject to $x(0) = a$ and $x(T) = b$, where $x : \mathbb{R} \to \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^n$, and $\|\cdot\|$ is the usual Euclidean norm. This is the multidimensional form of the action in the path integral that was derived in the last section. It is straightforward albeit somewhat tedious to arrive at this expression from first principles.

As per usual we try to minimize our actions by calculating derivatives and setting it equal to zero. Variations of the action A with respect to the path x yield

$$\begin{aligned}\frac{\delta A}{\delta x} &= -\frac{d}{dt}\left(\dot{x} - f(x)\right) - [\nabla f]^T\left(\dot{x} - f(x)\right) \\ &= -\ddot{x} + \left(\nabla f - [\nabla f]^T\right)\dot{x} + [\nabla f]^T f(x).\end{aligned}$$

Upon setting the variation equal to zero we derive the Lagrangian form of the equations

$$\ddot{x} = \left(\nabla f - [\nabla f]^T\right)\dot{x} + [\nabla f]^T f(x).$$

A solution to this set of equations with the given boundary conditions is our instanton. We have not yet mentioned how to solve such an equation, but this will come shortly.

We may also solve the Hamiltonian form of the equations of motion

$$\begin{aligned}\dot{x} &= f + p \\ \dot{p} &= -[\nabla f]^T p.\end{aligned}$$

These will be what we will refer to as the instanton equations. The conjugate momenta p can be thought of as the necessary stochastic forcing to drive the state variables away from deterministic trajectories. Note that $p = 0$ corresponds to the deterministic trajectory.

The Hamiltonian form of the equations can be thought of as directly coming from the action

$$B[x, p] = \int_0^T \langle p, \dot{x} - f(x) \rangle - \frac{1}{2}\|p\|^2.$$

Going from the A action to the B action is called the Hubbard-Stratonovich transformation. This can be derived directly from the path integral representation. The one dimensional equivalent of this transformation is the identity

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx e^{-\frac{1}{2}x^2 + ixy}.$$

If we want to build in constraints we can introduce Lagrange multipliers to again reduce the problem to one of unconstrained optimization. For example, suppose that we would like to enforce the endpoint condition that $x(T) = b$. Although we typically set up the calculus of variations problem with this kind of constraint built in, we can also think of it in terms of finding the critical points of the augmented actions

$$\begin{aligned} A[\lambda, x] &= -\langle \lambda, x(T) - b \rangle + A[x] \\ B[\lambda, x, p] &= -\langle \lambda, x(T) - b \rangle + B[x, p] \end{aligned}$$

For our purposes here we will concentrate on the latter action. If one performs the usual calculation[1] on this object one sees that the conjugate momenta come equipped with an endpoint condition of the from $p(T) = -\lambda$. This may be derived several ways. One way is to consider variations of x and p that don't vanish at the endpoints in which case, for integration by parts to hold, it must be that $p(T) = -\lambda$. The conjugate momenta equations (as we shall see) are naturally evolved backwards, thus letting us avoid the awkwardness associated with solving boundary value problem via a shooting method or higher order method such as Newton-Kantorovich iteration.

Suppose that we, instead would like to consider a constraint on the average value of our state. Then the augmented actions look like

$$\begin{aligned} A[\lambda, x] &= -\lambda \left(\frac{1}{T} \int_0^T x(t) dt - b \right) + A[x] \\ B[\lambda, x, p] &= -\lambda \left(\frac{1}{T} \int_0^T x(t) dt - b \right) + B[x, p]. \end{aligned}$$

From whence the Hamiltonian form of the instanton equations get modified to

$$\begin{aligned} \dot{x} &= f + p \\ \dot{p} &= -[\nabla f]^T p - \frac{\lambda}{T}. \end{aligned}$$

[1] Calculating the derivative by considering variations of trajectories and using integration by parts when necessary.

Hence we see that constraints on the state variables manifest themselves as conditions on the conjugate momenta. Here $p(T) = 0$ at the endtime is the natural boundary condition. This natural boundary condition can be derived in two ways. The first is by requiring integration by parts to hold and treating the endpoint variations as nonvanishing. The other derivation assumes that the endpoint variations vanish and then maximizes over all endpoint conditions for p. In the case that we are enforcing a bulk integral constraint we add an inhomogenous term to the p equation and in the case of an endpoint condition we gain endpoint conditions for p.

In the sections that follow we will show how to use the instanton equations to solve for the probability densities of observables. Finally, we comment that is not necessarily the case that the instanton equations offer a unique solution. With multiple solutions to choose from we must pick out the one that minimizes the action as corresponding to the "true" instanton.

2.4 Example 1: Brownian motion

We will now put together all the theory and perform a few calculations. The simplest one is finding the probability distribution of Brownian motion in one dimension. Specifically the system that we will be looking at is

$$\dot{x} = \epsilon\xi$$

where ξ is delta correlated Gaussian white noise and $\epsilon \neq 0$ is our "noise strength". Our observable of interest will be $x(T)$, the distribution of x at the final time T.

We can readily obtain this distribution by solving The Fokker-Planck equation

$$\partial_t \rho = -\frac{\epsilon^2}{2}\partial_{xx}\rho.$$

We will assume that the trajectory starts at $x(0) = 0$, meaning that the initial density is $\rho(b, 0) = \delta(b)$. Given this initial condition the probability distribution for a later time T is calculated to be

$$\rho(b, T) = \frac{1}{\sqrt{2\pi\epsilon^2}} e^{-\frac{1}{2\epsilon^2}\frac{b^2}{T}}.$$

We will now arrive at the same probability distribution via the instanton approach. First note that the probability distribution for the position at the final time satisfies a large deviation principle. Furthermore, Laplace's method is exact for Gaussian distributions, hence we expect the instanton approach to yield very good answers.

Given that our observable is the trajectory at the final time $x(T) = b$ the Hamiltonian form of the instanton equations are

$$\dot{x} = p$$
$$\dot{p} = 0$$

with $x(0) = 0$, $p(T) = -\lambda$, and $x(T) = b$, which can be readily solved to yield

$$\begin{aligned} x^*(t) &= -\lambda t \\ p^*(t) &= -\lambda \\ \lambda &= -\frac{b}{T}. \end{aligned}$$

To solve this system of equations we do not necessarily need to specify where the trajectory ends up beforehand (the $x(T) = b$ condition). Indeed if we specified λ at the outset, this would have implicitly defined an endpoint $x(T)$. The Lagrange multiplier λ implicitly enforces this constraint. In nonlinear systems a choice of λ will often **not** lead to a unique final value for $x(T)$, but for nondegenerate linear systems we expect uniqueness. Furthermore, we expect that for a given final value there always exists a λ such that $x(T) = b$ for an arbitrary b. Heuristically, this comes from the fact that we can imagine noise driving our system to any point in phase space. This is not true if there are regions where the noise is zero.

The solution $x^*(t) = \frac{b}{T}t$ is the instanton for the Brownian motion system and it also happens to be the global minimizer of the action. It says that the most likely path of an observable that starts at $x(0) = 0$ and reaches $x(T) = b$ is a straight line. Again, this does not mean that a Langevin trajectory will look like this, but rather that an ensemble of paths pass through this straight line with more likelihood than other points in spacetime. Now that we have our instanton we can calculate the probability distribution via the large deviation assumption

$$\begin{aligned} \rho(b, T) &\sim \exp\left(-\frac{1}{2\epsilon^2} A[x^*]\right) \\ &= \exp\left(-\frac{b^2}{2\epsilon^2 T}\right) \end{aligned}$$

which is proportional to the exact probability distribution obtained from the Fokker-Planck equation. The reason that the normalization factor is lost is a consequence of the path integral formulation and Laplace's method. Recapitulating, we have solved a continuum of ODE's to arrive at the same density as the solution to the Fokker-Planck PDE.

We chose our observable to be the state of the system at the final time, but there is no reason why we cannot consider different objects, for example the average value of the path $\frac{1}{T}\int_0^T x(t)dt$, or even the square of the state at the final time $\frac{1}{2}x(T)^2$. The former presents no difficulties, but the latter brings up some interesting issues. There the final condition of p is $p(T) = -\lambda x(T)$ and when one solves the equations one see that $\lambda = -T^{-1}$, it is independent of b. Furthermore the instanton is exactly the same as was the case for the observable at the final value case. First of all this cannot be correct because $x(T)^2$ cannot take negative values, thus these equations can break down and it pays to be wary of their limit.

2.5 Example 2: Ornstein-Uhlenbeck process

A slightly more complicated example of using the instanton equations to obtain the probability distribution comes from examining Ornstein-Uhlenbeck processes. We will repeat the

same calculation as the previous section for this system. The stochastic ode is of the form

$$\dot{x} = -\gamma x + \epsilon\xi$$

where ξ is again taken to be Gaussian delta correlated white noise and $\gamma > 0$. The Fokker-Planck equation in this case is

$$\partial_t \rho = -\gamma \partial_x (x\rho) + \frac{\epsilon^2}{2}\partial_{xx}\rho$$

whose solution for $\rho(b,0) = \delta(b-a)$ is

$$\rho(b,t) = \sqrt{\frac{\gamma}{\pi\epsilon^2\left(1-e^{-2\gamma t}\right)}}\exp\left(-\frac{\gamma}{\epsilon^2}\left[\frac{\left(b-ae^{-\gamma t}\right)^2}{1-e^{-2\gamma t}}\right]\right).$$

Again a large deviation principle is satisfied, and the solution is a Gaussian, hence we expect that the instanton equation will yield the exact answer in this case. Choosing our observable to $x(T)$ we get the instanton equations

$$\begin{aligned}\dot{x} &= -\gamma x + p\\ \dot{p} &= \gamma p\end{aligned}$$

with $x(0) = a$, $p(T) = -\lambda$, and $x(T) = b$. The solution to these equations are

$$\begin{aligned}x(t) &= ae^{-\gamma t} - e^{-\gamma T}\frac{\lambda}{2\gamma}\left(e^{\gamma t} - e^{-\gamma t}\right)\\ p(t) &= -\lambda e^{\gamma(t-T)}\\ b &= ae^{-\gamma T} - \frac{\lambda}{2\gamma}\left(1-e^{-2\gamma T}\right)\end{aligned}$$

Plugging the instanton into to the action yields

$$\begin{aligned}\rho(b,T) &\sim \exp\left(-\frac{1}{2\epsilon^2}A[x]\right)\\ &= \exp\left(-\frac{1}{2\epsilon^2}\int_0^T p(t)^2 dt\right)\\ &= \exp\left(-\frac{\gamma}{\epsilon^2}\left[\frac{\left(b-ae^{-\gamma t}\right)^2}{1-e^{-2\gamma T}}\right]\right).\end{aligned}$$

Again we see that we get the same result as before, but without the normalization factor.

2.6 Numerically solving the instanton equations

Although there are more cases that can be handled analytically for more complicated systems we must fall back on computing the solutions numerically. To do so we find the Hamiltonian formulation the easiest "deterministic" way to compute the instantons. In the

multidimensional setting the equations of motion for the distribution of a state variable at the endtime is

$$\dot{x} = f(x) + p$$
$$\dot{p} = -[\nabla f]^T p$$

with boundary conditions $x(0) = a$, $p(T) = \lambda$, $x(T) = b$. If our interest is in, let's say, just the $i'th$ component of the vector x_i, then $\lambda_j = 0$ for $j \neq i$. This is equivalent to computing the probability distribution of x_i with all the other variables integrated out. The adjoint operator $-[\nabla f]^T$ may be calculated by hand.

The algorithm to solve the instanton equations goes as follows:

1. Given the conjugate momenta $p^{(n)}$, evolve the state equation equation forward in time using the initial condition $x(0) = a$ to generate a new state $\tilde{x}^{(n)}$.

2. Evolve the adjoint equation backwards using $p(T) = -\lambda$ and $\tilde{x}^{(n)}$ to generate a new momenta $\tilde{p}^{(n)}$.

3. Update x and p via

$$x^{(n+1)} = (1 - s)x^{(n)} + s\tilde{x}^{(n)}$$
$$p^{(n+1)} = (1 - s)p^{(n)} + s\tilde{p}^{(n)}$$

   for some $s \in (0, 1]$. This is the relaxation step.

4. Repeat until both x and p stop changing.

To initialize the procedure one may take $p^{(0)} = 0$ for small λ. Once small λ solutions are calculated one may proceed to the large λ case by numerically continuing, using the p solution from the smaller λ as an initial guess for the higher λ. The case $s = 1$ in the algorithm corresponds to a fixed point iteration and $s \in (0, 1)$ may be though of as a relaxation type procedure. One may attempt to choose s such that the residual of the equations are lower at each iteration. There is no guarantee that the algorithm will converge, but it has been seen to work for a lot of cases considered for this work. It has however, also failed. Numerically this would correspond to the new search direction given by $x^{(n)} - \tilde{x}^{(n)}$ as being inadequate, leading to $s \to 0$ as n gets larger.

To evolve the equations forward in time is a problem of numerical integration, of which there are a large variety of choices; however, one must be careful in choosing a method. Since both the instanton and the momenta must be known at each point in time to solve the equations of motion, using a Runge-Kutta scheme necessitates the use of interpolation to get intermediate values. This added complexity is why we opted for the simpler multi-step schemes. Heun's method, followed by third order Adam's Bashforth is a perfectly adequate globally third order scheme. Furthermore one must be careful in starting the time integration scheme. There is no point in using a higher order scheme if the first few time steps don't have the same local order as the global order. Hence why we start off with two steps of Heun's method and then use Third Order Adam's Bashforth on the rest.

But how does one verify that one indeed has the solution to the instanton equations? Doing one fixed point iteration and checking that the answer has not changed is one method.

A method that corresponds to checking how close our discrete numerical solution is to the continuum is to check to see that the Hamiltonian is conserved at each point in time. This is a completely separate check than the one to verify that the discrete equation are satisfied. Although the continuous system has a Hamiltonian that is conserved at each point in time, the discrete system need not conserve the discrete Hamiltonian. Generally the Hamiltonian will not be conserved, but will have slight variations in it that get smaller as one decreases the time-step. It is very important to check to see that the answer does not change as Δt gets smaller.

For systems with a large number of state variables memory requirements start to be a rate limiting factor. Depending on the choice of noise one can reduce the requirements by only storing a few states of p and setting the rest to zero. Thus the only thing that needs to be stored is the state at the beginning time, and the value of p at all times. This is the algorithm that has been developed by Grafke et al. [3]. However, there is a third option that eliminates the need of storing x and p at all points in time. Here one uses both the Lagrangian and Hamiltonian formulation. In this document this method was not implemented, but was developed just in case it was necessary. The checkpointing method cannot be used on the Hamiltonian form of the instanton equations, but using both the Hamiltonian and the Lagrangian formulation, one may employ the checkpoint method from optimal control theory.

There are other algorithms that could be used as well: for example second order methods (Newton-Kantorovich iteration), a spectral discretization, Heun's method, etc. This is in addition to the direct method, which is done by running the stochastic system directly. We opted for the simplest (and most standard) method to solve the equations. Although we did run into difficulty with this simple method, we do not think that this is a consequence of the method, as will be explained later.

3 Instantons in Chaotic Systems

All the examples that we have talked about so far have been for linear systems that satisfy a large deviation principle. We would now like to carry this program into the nonlinear regime. We have seen that instantons can tell us about deviations from deterministic dynamics, but can it tell us anything about a chaotic attractor itself? The initial motivation for this project was a paper by Grafke et al. in which they calculated the probability distribution for the velocity gradient in Burger's turbulence [4]. In that work the initial condition was taken to be the origin and the final time T was taken to be infinity. At the end they were able to show excellent agreement with the calculated probability distribution via the instanton approach and the one obtained from Monte Carlo simulations of the stochastic system. The authors then conjectured that the instanton approach should be a viable approach towards the study of the Navier-Stokes equation.

One of the important things to note is that the deterministic dynamics of Burger's equation has no chaos. What is known as Burger's Turbulence is the stochastically forced Burger's equation. Furthermore Burger's equation is integrable via the Cole-Hopf transformation in which it can be related to the heat equation. This by no means says that the stochastically forced Burger's equation is similar to the stochastically forced heat equation, but rather says that the underlying deterministic dynamics are non-chaotic.

This observation prompts one to look at the use of instantons in chaotic systems, in hopes of calculating similar objects. In Navier-Stokes the dream would be to calculate the probability distribution of the dissipation $\|\nabla\vec{u}\|^2$ on the chaotic attractor or perhaps of a component of the velocity field $\vec{u}$ at a choice point in space. There is an important caveat here. The probability distribution is constructed from a histogram of the signal $d(t) = \|\nabla\vec{u}\|^2$ in the long time limit. This is a deterministic object, completely independent of noise but hopefully related to the noiseless limit of a stochastic forced Navier-Stokes system. Hence again we want to understand objects in the noiseless limit, exactly where the large deviation theory and instanton approach shines the most.

Just because we would like to understand objects in the noiseless limit does not mean that the instanton is a viable method to understand the invariant measure. The main issue here is that the invariant measure comes from a long time limit. It is not necessarily the case that the $\epsilon rightarrow 0$ and $T \to \infty$ limits commute. This is relevant because the instanton equations make use of the $\epsilon \to 0$ limit first.

On the other hand adding noise to the system allows one to explore the entirety of state space in a finite amount of time via sufficiently large noises. Since deviations away from deterministic dynamics cost more noise it may be the case that one can explore the invariant measure. Said differently perhaps the noise makes it easier to access regions of state space corresponding to the invariant set while penalizing deviations away from the invariant set. We would want the end result to be independent of what starting point we chose on the attractor and ideally we would like like to take a long time limit. Unfortunately as we will see and explain (later), both will be impossible on a chaotic attractor.

However, there is another feature of instantons that make it an interesting tool to use on chaotic systems: its ability to find "most likely" paths from one point to another. Although typically this is done in the context of transitions from one stable point to another in systems that admit a potential function for the forcing term, it may be possible for the instanton to find "minimal paths" from one exceptional state of the system to another: for example transitions from one unstable fixed point to another. If this is applicable in simple chaotic systems it may be the case that new fixed points (coherent structures) may be discovered in Navier-Stokes equation using the instanton approach.

3.1 Lorenz

The first chaotic system that we will look at are the celebrated Lorenz equations,

$$\begin{aligned}\dot{x} &= \sigma(-x + y)\\ \dot{y} &= -y + (r - z)x\\ \dot{z} &= -bz + xy,\end{aligned}$$

where $\sigma \in (0, \infty), r \in (0, \infty)$ and $b \in (0, 4)$ [6]. The canonical parameter values for the chaotic regime are $(\sigma, r, b) = \left(10, 28, \frac{8}{3}\right)$. These equations are a prototypical model of continuous time dynamical systems that exhibit chaos. They were originally derived as a truncation of Rayleigh's problem, which is itself a model of thermal convection [5].

Figure 1 shows the trajectory in phase space for a typical initial condition on the attractor. Here one can see the delicate spirals and and low dimensionality of the attracting

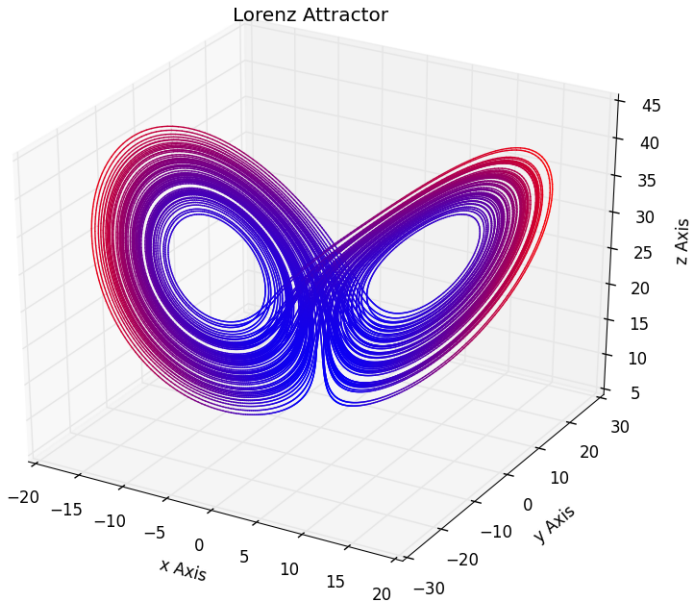

Figure 1: The phase plot of the Lorenz attractor at $r = 28$, $\sigma = 10$, and $b = 8/3$. The colors indicate the relative speed of a particle on the trajectory, where red is "fast" and blue is "slow".

set[2]. The holes in the wings are where the fixed points of the systems are located and the colors represent the relative speed on the attractor where red is fast and blue is slow.

There are several reasons why the Lorenz equations are an attractive testing ground for the instanton approach. One is that the deterministic dynamics remain bounded for all time. Secondly the Lorenz equations exhibit chaos, and this is exactly the regime in which we would like to test some of the instanton ideas. Third there are well defined quantities of interest that we would like to understand. The observable that we will concentrate on here is the long time average of the state variables, for example

$$\begin{aligned} \langle xy \rangle &= \limsup_{T\to\infty} \frac{1}{T} \int_0^T xy dt \\ &= b\langle z \rangle \end{aligned}$$

where the last line come from integrating the z equation for a long time, making use of the fact that the system is bounded for all time. The long time correlation of the x and y state variables are related to "heat transport" (the Nusselt number) in Rayleigh's original model.

[2]Using periodic orbit theory Viswanath estimated the Hausdorff dimension of the set to be approximately 2.06 [8]

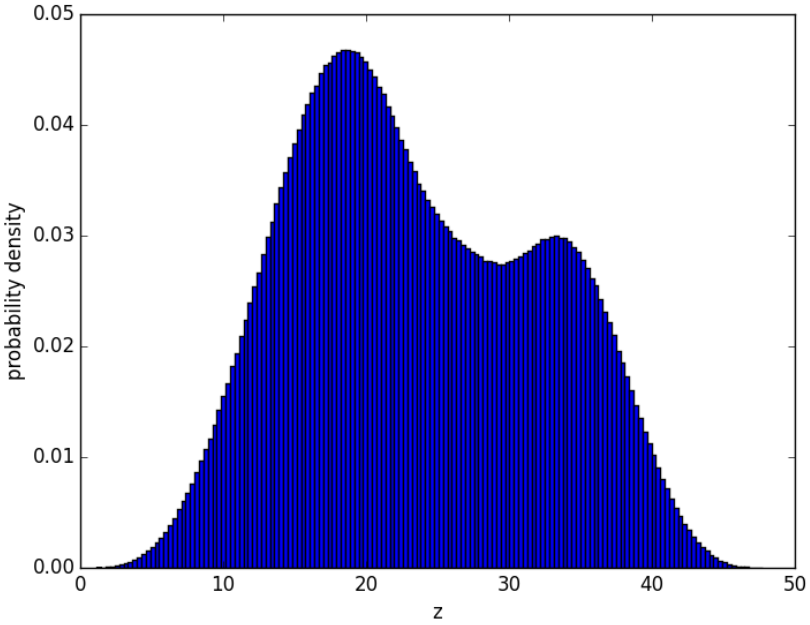

Figure 2: Histogram of Lorenz state variable z on the chaotic attractor for parameter values $r = 28$, $\sigma = 10$, and $b = 8/3$.

This quantity has gained some recent attention where one can prove upper bounds in both the deterministic case and as well as the noisy case [7] [1].

In figure 2 the histogram of $z(t)$ on the chaotic attractor is shown. The average value of z for parameter values (r, σ, b) appears to be about $z(t) \approx 23.5 \pm 0.1$. One can see that this distribution has compact support (as it must be since z is bounded on the attractor), is highly non-Gaussian, and is strictly positive. If one refines the partition of bins it seems that the distribution becomes more and more complex, leading to the conjecture that it is probably fractal.

One expects that this distribution is related to the Lorenz system with a small amount of noise in the steady state limit. For any amount of noise (however small), we also expect that the distribution of z becomes infinitely smooth and extends to $\pm\infty$ (in contrast to our compact support for the noiseless case). This conjecture is supported by numerical evidence by B. Marston et al. in an unpublished (as of October 2015) work. In that study the steady state Fokker-Planck equation for the Lorenz system was solved numerically and then directly compared to the long time statistics, albeit for different parameter values than the canonical ones. Very good agreement was found for a range of "small" noise strength values.

One of our goals with the instanton formulation is to check whether or not it is possible to replicate some features of the histogram in 2. For example we would like to test whether or not the instanton equations are able to pick out the mean value, higher order statistics such as the variance, or in what way the compact support of the distribution manifests

itself. We know that the instanton equations keep track of a single trajectory and tells us deviations away from the determinism whereas the statistics of a chaotic trajectory only manifest themselves in the infinite time limit. Solving the instanton equations in the infinite time limit for a point that starts off on the chaotic attractor is not feasible, but it is possible to solve a finite time computation. Thus we will check these questions for finite time intervals. Since noise allows one to explore all of phase space in a finite amount of time[3] and typically don't have distributions with compact support, one may wonder if adding noise accelerates convergence to steady state distribution. Said differently perhaps noise lets one peek farther into the future than the deterministic equations.

To this end we will obtain the instanton equations

$$\begin{aligned}\dot{x} &= x + p \\ \dot{p} &= -[\nabla f]^T p\end{aligned}$$

for the stochastic Lorenz system with isotropic Gaussian white noise

$$\begin{aligned}\dot{x} &= \sigma(-x + y) + \epsilon\xi_1 \\ \dot{y} &= -y + (r - z)x + \epsilon\xi_2 \\ \dot{z} &= -bz + xy + \epsilon\xi_3.\end{aligned}$$

To get the instanton equations we must first calculate ∇f, which is

$$\nabla f = \begin{bmatrix} -\sigma & \sigma & 0 \\ r - z & -1 & -x \\ y & x & -b \end{bmatrix}$$

$$\Rightarrow$$

$$-[\nabla f]^T = \begin{bmatrix} \sigma & -(r - z) & -y \\ -\sigma & 1 & -x \\ 0 & x & b \end{bmatrix}.$$

From this we obtain the following set of coupled nonlinear differential equations,

$$\begin{aligned}\dot{x} &= \sigma(-x + y) + p_x \\ \dot{y} &= -y + (r - z)x + p_y \\ \dot{z} &= -bz + xy + p_z \\ \dot{p}_x &= \sigma p_x - p_y(r - z) - p_z y \\ \dot{p}_y &= p_y - \sigma p_x - p_z x \\ \dot{p}_z &= b p_z + p_y x.\end{aligned}$$

As was stated previously the $p_x(t) = p_y(t) = p_z(t) = 0$ at each time t case corresponds to the deterministic evolution. The end condition here is naturally $p_x(T) = p_y(T) = p_z(T) = 0$. There is one nontrivial stochastic solution[4] that we can determine exactly from these set

[3] This is a consequence of distributions coming from stochastic ode's with Gaussian white noise.

[4] We can also calculate the fixed points of the deterministic system, but this is not a solution that has nonzero values for the conjugate momenta.

of equations which corresponds to the initial condition $x(0) = y(0) = p_x(T) = p_y(T) = 0$ and $z(0) = a$ and $p_z(T) = b$. These final conditions correspond to choosing $z(t)$ as the observable of interest. The solution is an Ornstein-Uhlenbeck process along the z-axis with $z = 0$ being the "stable point". Experience tells us that this cannot be the solution that we are looking for though. Since we added isotropic Gaussian white noise we expect probability to leak from the sides and get wrapped up in the attractor for any finite amount of time.

This brings us to our first departure from the examples that were considered earlier: we expect multiple solutions to the instanton equations. Given that we are dealing with nonlinear equations this is perhaps not unexpected, but it is surprising since there is no mention of it in the literature. An example of this phenomena is summarized by the phase space plot in 3. Here the initial condition was taken to be the origin and the final condition[5] for the conjugate momenta was taken to be $p_x(T) = p_y(T) = 0$ and $p_z(T) = \lambda$. The figure displays three solutions corresponding to the same final condition for z. The red straight line is the Ornstein-Uhlenbeck process solution, while the blue and green curves are two alternative solutions that achieve the same final value of z. The dots in the figure represent the fixed points of the Lorenz attractor.

One can see from the figure that the blue and green solutions seem to be converging to the unstable fixed points of the attractor. Both of the solutions taken together wrap around the outside of the attractor and appear to be related to the heteroclinic connections between the origin and the fixed points. The oscillatory nature of the convergence to the fixed point made taking the long time limit numerically intractable. Furthermore solutions for larger as well as smaller values of $|\lambda|$ were found to be very difficult to compute given the procedure outlined in section 2.6, thus the rate function corresponding to these solutions were not computed. However, given that the infinite time limit seems to be evolving towards the fixed point one would not expect the corresponding probability distribution to resemble that of Figure 2.

To calculate the different numerical solutions one had to generate different initial guesses for the starting conjugate momenta p_x, p_y, p_z. For the Ornstein-Uhlenbeck process it was sufficient to choose $p_x(t) = p_y(t) = p_z(t) = 0$ as the initial guess and use the procedure described in Section 2.6. For the other two solutions we used a numerical continuation procedure. First the problem with the final condition $p_x(T) = p_y(T) = \delta$ and $p_z(T) = \lambda$ was solved for a small δ (again using the zero solution for p as an initial guess) and then this solution was fed into the algorithm as the starting guess for the solution to the $p_x(T) = p_y(T) = 0$ and $p_z(T) = \lambda$ boundary conditions. Attempts were made to find more solutions, but none were found.

So far we have only talked about an initial condition that starts on the origin, which also happens to be a fixed point for the Lorenz system. We also looked at other initial values, for example

1. random points on the chaotic attractor,

[5]Since in the figure we chose a fixed final $z(t)$ this means that λ was different depending on which solution was being computed, the green and the blue curves had the same $\lambda \approx 10^{-5}$ while the red line had a much higher λ chosen so that the final value of z was the same. We could have chosen the same boundary condition $\lambda \approx 10^{-5}$ for the red curve but this would not appear on the graph since it would be absorbed in the red dot.

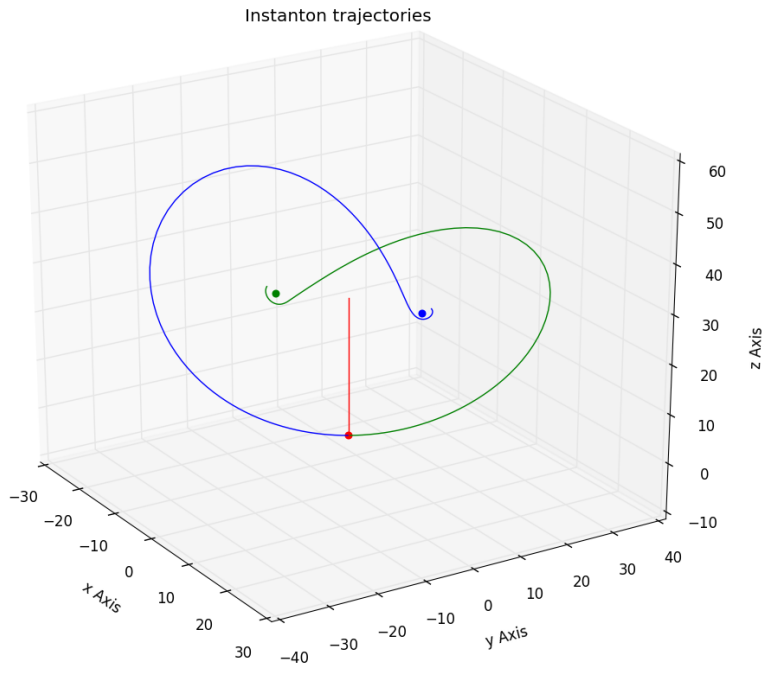

Figure 3: Multiple solutions for Lorenz system instanton trajectories with an initial condition starting at the origin. The blue and green curves appear to be related to heteroclinic connection from the origin (the red dot) to the two fixed points (blue and green dots). The red line is the Ornstein-Uhlenbeck solution to the Lorenz instanton equations corresponding to the same final value of z as the green and blue curves.

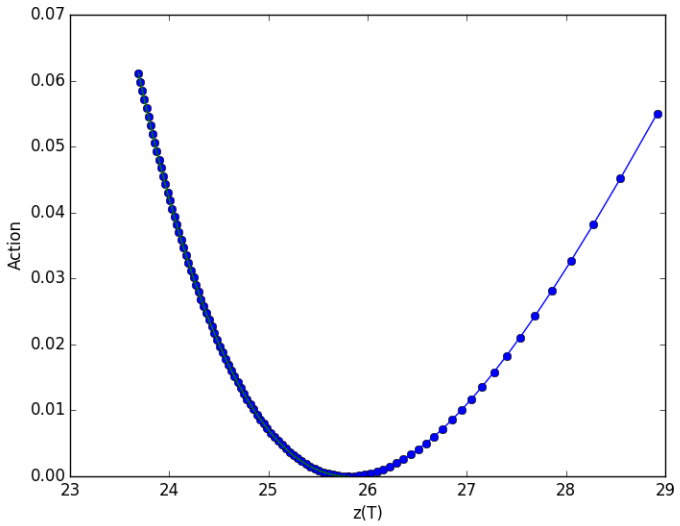

Figure 4: Cost versus final value of z for an initial condition on the attractor and time $T = 1.5$. The observable is taken to be z at the final time.

2. initial condition leading to a periodic orbit,

3. and other fixed points.

for different time periods. In all cases it was found that calculating trajectories for long time periods was not possible due to extreme ill-conditioning. However, modest values for time $T \approx 5$ were possible to compute, but finding the global minimum tended to be a challenge.

A representative result is summarized by figure 4. This shows the value of the action for an initial point on the attractor with a time $T = 1.5$. Here the observable was the final value of z. The minimum of the action is 0, which corresponds to the deterministic trajectory. Each dot is a different instanton solution. The solutions were obtained by starting with the deterministic trajectory $\lambda = 0$ and numerically continuing to higher λ. Attempts were made to go further but a few numerical issues prevented this. One can see that the quadratic behavior of the left and the right are different. Numerically continuing to smaller values of $z(T)$ were not a problem but when attempting to continue to higher values one starting finding different branches of solutions corresponding to the same final value for the conjugate momenta.

Similar computations were performed for a variety of different initial points, time periods, and observables, but all of them had the same parabolic structure. This means that all probability densities that were computed were essentially similar to Gaussian distributions and had a dependence on ϵ. This is in stark contrast to Figure 2 where there is no ϵ dependence. However it was observed that larger times required smaller values of cost to reach a larger final value of z. Hence it is suspected in the infinite time limit the cost for reaching any point of the attractor goes to zero.

Furthermore no hint to the compact support of the distribution was found, that is to say, the probability density computed from instanton solutions did not decay faster for

trajectories outside of the attractor. In light of these results it does not seem that we can interpret the instanton calculations as telling us more than the probability of deviating away from determinism. However, the calculation represented in Figure 3 gives hope of the instanton formulation being used to calculate both heteroclinic connections and perhaps being used to find unsteady fixed points.

3.2 Kuramoto-Sivashinsky

The second chaotic system that we will examine in this document is the Kuramoto-Sivashinksy equations KS equations)

$$\partial_t u + \partial_{xxxx} u + \partial_{xx} u + u \partial_x u = 0$$

which is periodic $x \in [0, \Gamma]$. This system is a hallmark of spatio-temporal chaos, and its dynamics are essentially confined to a finite dimensional dynamical system, even though it is ostensibly an infinite dimensional system. The ∂_{xxxx} term introduces dissipation into the system while the ∂_{xx} injects energy and in the long time limit these terms balance. The advective term transfers energy from the injective scale to the dissipative scale, guaranteeing that the solutions remain bounded. In this system the aspect ratio Γ serves as a measure of the possible complexity of the flow, where larger Γ implies more complexity.

The interpretation of the fourth and second derivative terms is most easily seen by multiplying the KS equations by u, integrating over space, and making use of periodicity to arrive at

$$\frac{1}{2}\partial_t \int_0^\Gamma u^2 dx = \int_0^\Gamma (\partial_x u)^2 \, dx - \int_0^\Gamma (\partial_{xx} u)^2 \, dx.$$

In the long time limit, since u is bounded, we have a balance between the average rate of injection and average rate of dissipation

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T \int_0^\Gamma (\partial_{xx} u)^2 \, dx dt = \limsup_{T\to\infty} \frac{1}{T} \int_0^T \int_0^\Gamma (\partial_x u)^2 \, dx dt.$$

The advective term interpretation may be seen by taking the spatial Fourier transform of the equations to get

$$\partial_t \hat{u}_n + (k_n^4 - k_n^2)\hat{u}_n + \widehat{u \partial_x u}_n = 0,$$

where $k_n = n\frac{2\pi}{\Gamma}$. The advective term is a convolution in Fourier space, meaning that each Fourier mode is intimately coupled to one another. Furthermore in this formulation it is much easier to see why Γ is a measure of the complexity. If one looks at the $k_n^4 - k_n^2$ term one can see that larger Γ allows more modes to be excited by the k_n^2 term.

We will be applying the instanton formalism to the KS equations and look at many of the same things that were done for the Lorenz equations. Before embarking on this journey we will take a brief moment to discuss what it means to add noise to a PDE and what kind of information we would like to extract. In the ODE case when we add noise to the system we wanted to understand the probability distribution of the state at each point in time. To this end a Fokker-Plank equation or Path Integral was employed to calculate such

a quantity. Since PDE's can be thought of as infinite dimensional ODE's it seems that the Fokker-Plank approach is out of the question since it would be a PDE with infinitely many "spatial" derivatives. This does not mean that the problem is completely intractable. If we add noise to a PDE we can still ask questions such as "What is the probability distribution of u at the origin?" or "What is the probability distribution of the first Fourier mode u_1?". In contrast to the ODE case in PDEs one must be very careful how noise is added to the system.

This is perhaps easiest to understand if we look at the KS equations in the Fourier mode representation. We cannot add uniform Gaussian white noise to each Fourier mode. The heuristic reasoning for this is that all scales will be excited by uniform Gaussian white noise, thus the connection with the deterministic equation is lost. The"energy" in some sense will be infinite. The typical way around this is to consider spatially correlated noise and leave it white in time. In Fourier space the equations of motion for the noisy KS equations could be chosen to be as follows

$$\begin{aligned} \partial_t \hat{u}_n + (k_n^4 - k_n^2)\hat{u}_n + \widehat{u\partial_x u}_n &= \epsilon w_n \xi_n \\ w_n &= k_n e^{-\frac{1}{2}k_n^2} \end{aligned}$$

where ξ_n is white noise in time. The w_n term is chosen so that the mean frequency is not excited and the decay term is chosen so that the smallest amplitudes are not "overly excited" by the noise.

We may go through the same path integral discretization procedure as before to arrive at the Friedlen-Wentzell action

$$\begin{aligned} A[u] &= \sum_n \left[(w_n)^{-2} \int_0^T \left(\partial_t \hat{u}_n + (k_n^4 - k_n^2)\hat{u}_k + \widehat{u\partial_x u}_n \right)^2 dt \right] \\ &= \int_0^T dt \left\| \partial_t u + \partial_{xxxx} u + \partial_{xx} u + u\partial_x u \right\|_\chi \end{aligned}$$

where the $\|\cdot\|_\chi$ norm is the norm associated with the first line and χ denotes the spatial correlation. Note that the higher Fourier modes have a much higher stochastic cost associated with them. The $n = 0$ mode will be taken to be zero throughout this work.

The instanton equations for Kuramoto-Sivashinsky are

$$\begin{aligned} \partial_t u + \partial_{xxxx} u + \partial_{xx} u + u\partial_x u &= \chi * p \\ -\partial_t p + \partial_{xxxx} p + \partial_{xx} p - u\partial_x p &= 0 \end{aligned}$$

where $\chi * p$ is a convolution of the conjugate momenta p with the spatial correlation χ. We may formally obtain it by simply observing that the operator that governs the backwards evolution for p will always be the adjoint operator and that the stochastic forcing term is modified by the correlation function. If we wish to consider averaged quantities, i.e. the average let's say, energy of the system

$$\frac{1}{LT} \int_0^T \int_0^L u^2 dx dt$$

the the instanton equations get modified as before with an inhomogenous term for the conjugate momenta p. In Fourier space the instanton equations are

$$\partial_t \hat{u}_n + (k_n^4 - k_n^2)\hat{u}_n + \widehat{u\partial_x u}_n = (w_n)^2 p_n$$
$$-\partial_t \hat{p}_n + (k_n^4 - k_n^2)\hat{p}_n - \widehat{u\partial_x p}_n = 0$$

Here the aspect ratio Γ is taken to be 22. At this aspect ratio it suffices to have $N = 128$ modes to represent the flow in the chaotic regime.

Many of the difficulties and insights from the Lorenz system carried over to the KS equations. Again it was found that taking the long time limit was intractable and lead to problems with convergence. Furthermore it seems that they can only tell us about deviations away from determinism. A sample result is displayed in Figure 5. Here the KS instanton equations[6] were taken from two initial conditions: the one on the left was from the initial condition $u = 0$ while the one on the right is a random initial point on the attractor. The horizontal axis is space, the vertical axis is time, and the colors represent whether or not the flow field is positive or negative. The final state of the evolution from zero is distinct from the evolution on the chaotic attractor and has a much simpler evolution. This calculation shows that the initial condition plays a huge role in the evolution.

A hypothesis that has not been tested yet is to check whether or not the instanton equations may be used to easily calculate unstable fixed points of the system. If this is possible the instanton equations offer an exciting alternative to the usual methods for calculating fixed points for PDEs that allow for convergence from much farther away than usual since it is related to a gradient ascent type of procedure. This will be tested in future work.

3.3 Generic insights and speculation

Many of the numerical difficulties are perhaps insurmountable in this project. The instanton equations are inherently nonlinear boundary value problem in the case that f is a nonlinear function. Although in previous studies one was able to take the long time limit, in the case of trajectories on a chaotic attractor this not possible. Typically one expects that chaotic trajectories are not entire functions (in the complex variable sense) of time, meaning that a rescaling of time would not ameliorate any problems. The extreme sensitivity to initial conditions renders the last state of a time integration meaningless in the long time limit. This manifests itself numerically as an ill-conditioning of the boundary value problem. For longer times things get exponentially worse, eventually rendering any amount of careful integration meaningless.

Another problem arose as well. The Lagrange multiplier λ became increasingly smaller as time got larger to reach the same point in state space. The heuristic reason for this goes as follows: In the long time limit there are many ways for a trajectory to get on one point of the attractor to another. The extreme sensitivity to initial conditions allows one to jump

[6]The spatial correlation was chosen to be $\chi_n = k_n e^{-k_n^2}$ and the observable was chosen to be the value of the velocity field at the origin. In Fourier space this means that $\lambda_n = 1$ for all conjugate momenta final conditions.

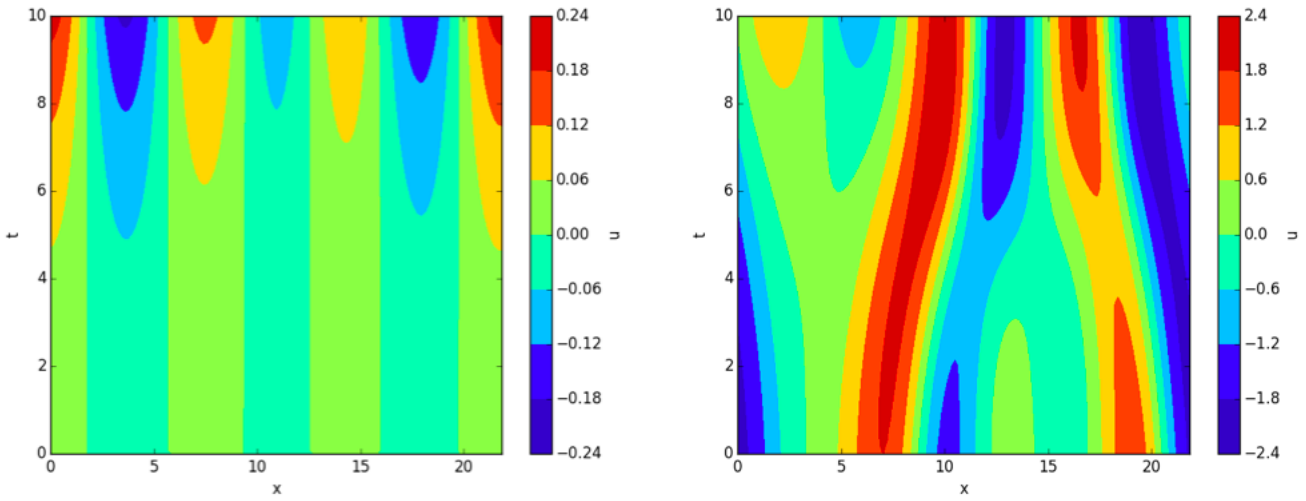

Figure 5: A contour plot of the instanton trajectories for the KS equations. The horizontal axis is the physical spatial coordinate, x, and the vertical axis is time. The color represents the value of the instanton solution u. In both cases the observable at the final time is taken to be $\partial_{xx}u(0,0)$ but the initial condition for the left is starting from the origin and the initial condition on the right starts from a point on the chaotic attractor.

onto a deterministic trajectory that gets to the endpoint just as easy. The longer the time we wait the more candidates there are for reaching a given final state, all of which may look completely different.

This same "problem" may also explain what leads to multiple solutions in the instanton equations. One way of organizing the framework is by thinking of things in the context of periodic orbit theory. For Lorenz the smallest period is of the order $T \approx 1.6$ and, as time increases, exponentially more periodic orbits come into existence. Each are embedded within the attractor and offer a viable candidate to get from one point in state space to another for a given λ. Hence as time grows we expect more and more solutions to the instanton equations. Once one has multiple solutions pruning which ones matter and which ones don't becomes more of an art and increasingly less quantitative. The infimum becomes essentially hopeless to calculate and unless one already knows the answer, it is relegated merely to a matter of guesswork. There could always be some solution lurking out in function space that is missed.

As we saw for the Lorenz example, three solutions to the instanton equations could be calculated for the initial condition that started at the origin. No solutions seemed to sweep itself up into the attractor, meaning that a lot of solutions could have been possibly missed. This poses a danger when trying to calculate the probability distribution at the end since the global infimum is the only critical point that matters in the $\epsilon \to 0$ limit.

Normally the instanton equations are calculated in the long time limit. If the instanton trajectory is simple then such a calculation can be rendered tractable. However, if the evolution has, for example, oscillations towards a final state then such a limit may be

rendered computationally infeasible. There are perhaps ways to get around this difficulty for particular systems, but these infinite horizon problems can be notoriously difficult to deal with.

There is an additional numerical issue that arose in attempting to solve the instanton equations. If the deterministic dynamics of the system allow for solution to blow up, so do the instanton equations. Such a scenario occurs with the Rössler system. Even though the Rössler system has a chaotic attractor (for the right choice of parameters) this does little good if enough noise can knock a trajectory off of the attractor and into a "blow up region" where all solutions quickly run away. This will manifest itself in the instanton equations by choosing a large enough end condition (λ in this document). This is not necessarily a bad feature if we would like to probe whether or not a system, i.e. Navier-Stokes, does exhibit blow up.

With regards to the applicability of the instanton formulation to Navier-Stokes, it seems likely that it may be able to find new fixed points as long as there are heteroclinic connections between the fixed points and the initial condition. If the initial condition is chosen "poorly" it may result in one being on the chaotic attractor which would lead to hopeless numerical difficulties. It seems, however, extremely unlikely that one could calculate the tail ends of probability distributions in turbulence given this method. As we saw with Lorenz and Kuramoto-Sivashinsky the tale end of the distribution is completely unrelated to the instanton equations. As long as the deterministic dynamics is dominant, as it seems to be the case in the turbulent regime, the role of noise is secondary and cannot be exploited in the context of instantons.

It is possible that the inability of the instanton equations to capture the distribution associated with the chaotic regime comes from an incompatibility between the $T \to \infty$ and $\epsilon \to 0$ limits. If one solves the steady state Fokker-Plank equation what one is doing is calculating $T \to \infty$ first for a fixed ϵ. One can then study $\epsilon \to 0$ limit of the distribution. It is this order that the limits must be taken in order to have a correspondence with the chaotic attractor. With the instanton equations one focuses on the $\epsilon \to 0$ limit first and then takes the $T \to \infty$ limit afterwards. Hence it seems like the instanton equations are fundamentally incompatible with calculating chaotic properties. This observation has been pointed out before [2].

4 Summary and Conclusions

The instanton equations come from the minimization of an action occurring in the path integral. Under a large deviation assumption they allow one to calculate the tail of probability distributions and even obtain the most likely trajectories that lead to such an extreme state. Although instantons can say a lot about deviations away from determinism this does little good if the deterministic part is the majority of the information as is the case with chaos.

The instanton equations were implemented in systems with chaotic dynamics of which this document focused on two: the Lorenz equations and Kuramoto-Sivashinsky equations. The instanton equations where solved for various initial conditions and lengths of time and the resulting probability densities were calculated. The instanton densities were found to be completely unrelated to those of the deterministic dynamics and it seems to be the case

that they have to be unrelated.

However, it seems plausible for the instanton equations to find new coherent structures that may be missed by conventional approaches. This was seen in the Lorenz equations where an initial condition starting at the origin was able to evolve towards the unstable fixed points. This leads to the belief that the instanton method may be a viable approach to finding unstable fixed points of a dynamical system as long as there exists a heteroclinic connection between them.

5 Acknowledgments

I would like to thank John Wettlaufer and Oliver Bühler for organizing an incredibly stimulating summer. The excellent lectures by Charlie Doering and Henk Dijkstra were fundamental in grounding me with the proper background to undertake this summer project. Their care in presenting the material is highly appreciated. I am also grateful for the many engaging discussions with Greg Chini, Predrag Cvitanović, Charlie Doering, David Goluskin, John Wettlaufer, as well as all of my fellow fellows and several visiting students. And finally I would like to thank Rich Kerswell for his guidance and encouragement throughout the project.

References

[1] S. Agarwal and J. Wettlaufer, *Maximal stochastic transport in the lorenz equations*, Physics Letters A, (2015), pp. –.

[2] M. Berry, *Asymptotics, singularities and the reduction of theories*, 134 (1995), pp. 597 – 607.

[3] T. Grafke, R. Grauer, and T. Schäfer, *The instanton method and its numerical implemenation in fluid mechanics*. `http://arxiv.org/abs/1506.08745`, 2015.

[4] T. Grafke, R. Grauer, T. Schäfer, and E. Vanden-Eijnden, *Relevance of instantons in burgers turbulence*, EPL (Europhysics Letters), 109 (2015), p. 34003.

[5] Lord Rayleigh, *On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side*, Philosophical Magazine, 32 (1916), pp. 529–546.

[6] E. N. Lorenz, *Deterministic nonperiodic flow*, Journal of the Atmospheric Sciences, 20 (1963), pp. 130–141.

[7] A. N. Souza and C. R. Doering, *Maximal transport in the Lorenz equations*, Physics Letters A, 379 (2015), pp. 518–523.

[8] D. Viswanath, *Symbolic dynamics and periodic orbits of the lorenz attractor*, Nonlinearity, 16 (2003), p. 1035.