WHOI-2016-05

2015 Program of Study: Stochastic Processes in Atmospheric & Oceanic Dynamics

by

John Wettlaufer and Oliver Bühler

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

November 2016

Technical Report

Funding was provided by the National Science Foundation Grant #OCE-1332750

Approved for Distribution:

Albert J. Plueddemann, Chair

Department of Physical Oceanography

Preface

Stochastic Processes in Atmospheric & Oceanic Dynamics was the theme at the 2015 GFD Program. Professors Charlie Doering (University of Michigan) and Henk Dijkstra (University of Utrecht) were the principal lecturers. Their lectures were collectively two-pronged. The first prong was launched by Charlie, who laid down the mathematical foundations of random variables, stochastic processes and the nature and analysis of stochastic differential equations. In the second, Henk took us through the many places in the Atmosphere, Ocean and Climate system where the infrastructure from the first prong plays out.

John Wettlaufer and Oliver Bühler were the stochastic co-directors. In keeping with the theme, the Cottage was in constant motion with many visitors and long-term staff members. Following the thematic principal lectures, the seminar room was busy all summer, with talks spanning an impressive range of topics that we are typically fortunate to experience in Walsh Cottage. Importantly, some of the newer staff ably jumped into the supervision of fellows projects – directly or indirectly. The fellows pursued a rich range of projects and have produced a fine set of reports. As usual they came from far and wide:

- Yana Bebieva, Yale University
- Tom Beucler, Massachusetts Institute of Technology
- Tom Eaves, University of Cambridge
- Giovanni Fantuzzi, Imperial College London
- Anna FitzMaurice, Princeton University
- Florence Marcotte, Inst. Physique du Globe de Paris
- Gunnar Peng, University of Cambridge
- Cesar Rocha, University of California, San Diego
- Andre Souza, University of Michigan
- Chris Spalding, California Institute of Technology

As the second week wound down, the 2015 GFD Sears Public Lecture was given by Susan Solomon (*MIT*) in Redfield Auditorium. She took us through a fascinating success story of how science impacted policy and eventually the global economy in her talk "Ozone Depletion: A Science and Policy Success Story." As a key player in the *Ozone Hole Saga*, Susan's talk drew a large and engaged audience, who asked many questions.

Some important acknowledgements: Support from Anders Jensen in the laboratory was appreciated, particularly as the going with some sensors got tough. Annie Doucette, Janet Fields and Julie Hildebrandt formed the administrative team that ensured the program ran with admirable efficiency. Srikanth Toppaladoddi, a 2012 fellow, was of great help in corralling the disparate piece of the proceedings. We are indebted to *WHOI*

Academic Programs Office, who continue to provide an ideal atmosphere.

Finally, we note that the passing of Louis Howard, who impacted the program in countless ways, marked the summer in a different manner. On 13 August 2015, a memorial for Louis at Carriage House was organized by George Veronis. Those of us who were fortunate to attend were emboldened by the many and varied stories shared by the participants. His smile, encouragement and scientific sage will be deeply missed. A detailed obituary is found at the following link:
https://math.mit.edu/about/history/obituaries/howard.php

Table of Contents

2015 Geophysical Fluid Dynamics Participants

STAFF AND VISITORS

Akers, Benjamin	Air Force Institute of Technology
Alexakis, Alexandros	École Normale Supérieure
Anderson, James	Stevens Institute of Technology
Balmforth, Neil	University of British Columbia
Boland, Emma	British Antarctic Survey
Bolton, Edward	Yale University
Bühler, Oliver	New York University
Bush, John	Massachusetts Institute of Technology
Caulfield, Colm-cille	University of Cambridge
Cessi, Paola	University of California, San Diego
Chini, Gregory	University of New Hampshire
Cvitanovic, Predrag	Georgia Institute of Technology
Del Sordo, Fabio	Yale University
Dijkstra, Hendrik	Utrecht University
Doering, Charles	University of Michigan
Flierl, Glenn	Massachusetts Institute of Technology
Gallet de St.-Aurin, Basile	CEA, Saclay
Garaud, Pascale	University of California, Santa Cruz
Goldenfeld, Nigel	University of Illinois at Urbana-Champaign
Goluskin, David	University of Michigan
Guarnieri, Filippo	Yale University
Helfrich, Karl	Woods Hole Oceanographic Institution
Keller, Joseph	Stanford University
Kerswell, Richard	University of Bristol
Lecoanet, Daniel	University of California, Berkeley
Llewellyn Smith, Stefan	University of California, San Diego
Mahadevan, Amala	Woods Hole Oceanographic Institution
Mancarella, Francesco	Nordic Institute for Theoretical Physics
Manucharyan, Georgy	Yale University
Marshall, John	Massachusetts Institute of Technology
Marston, John	Brown University
Mashayek, Ali	Massachusetts Institute of Technology
Mitra, Dhrubaditya	Nordic Institute for Theoretical Physics
Moon, Woosok	University of Cambridge
Morrison, Philip	University of Texas at Austin
O'Rourke, Amanda	Princeton University
Pedlosky, Joseph	Woods Hole Oceanographic Institution
Phillips, Rob	California Institute of Technology
Spall, Michael	Woods Hole Oceanographic Institution
Spiegel, Edward	Columbia University
Sreenivasan, Katepalli	New York University

Thiffeault, Jean-Luc	University of Wisconsin, Madison
Timmermans, Mary-Louise	Yale University
Toppaladoddi, Srikanth	Yale University
Veronis, George	Yale University
Von Herzen, Brian	The Climate Foundation
Wagner, Gregory	University of California, San Diego
Weiss, Jeffrey	University of Colorado, Boulder
Wells, Andrew	University of Oxford
Wettlaufer, John	Yale University
Whitehead, Jack	Woods Hole Oceanographic Institution
Wunsch, Carl	Massachusetts Institute of Technology
Yakhot, Victor	Boston University
Yecko, Philip	Cooper Union for the Advancement of Science & Art
Young, William	Scripps Institution of Oceanography

GUEST STUDENTS

Agarwal, Sahil	Yale University
Burns, Keaton	Massachusetts Institute of Technology
Miles, Christopher	University of Michigan
Simha, Akarsh	University of Texas at Austin

2015 Principal Lecturers

Henk Dijkstra

Charlie Doering

2015 Geophysical Fluid Dynamics Summer School Participants

First Row (L-R, sitting): Tom Beucler, Florence Marcotte, Gunnar Peng, Chris Spalding, Giovanni Fantuzzi, Yana Bebieva, Anna FitzMaurice, Tom Eaves, Cesar Rocha, Andre Souza

Second Row (L-R): Ed Spiegel (standing). Sitting: Charlie Doering, John Wettlaufer, Greg Chini, David Goluskin, Phil Morrison, Joe Keller, George Veronis, Jean-Luc Thiffeault, Henk Dijkstra, Jeff Weiss, Filippo Guarnieri, Neil Balmforth, Oliver Bühler (standing)

Third Row (L-R): Gregory Wagner, Basile Gallet, Christopher Miles, Keaton Burns, Akarsh Simha, Colm Caulfield, John Marshall, Bill Young, Dhruba Mitra, Glenn Flierl, Joe Pedlosky, Srikanth Toppaladoddi, Andrew Wells, Georgy Manucharyan, Sahil Agarwal, Woosok Moon, Amanda O'Rourke, Rich Kerswell, Fabio DelSordo

Not pictured: Ben Akers, Alexandros Alexakis, Jim Anderson, Emma Boland, Edward Bolton, John Bush, Paola Cessi, Predrag Cvitanovic, Pascale Garaud, Nigel Goldenfeld, Karl Helfrich, Daniel Lecoanet, Stefan Llewellyn Smith, Amala Mahadevan, John Marston, Ali Mashayek, Mike Spall, Katepalli Sreenivasan, Mary-Louise Timmermans, Brian von Herzen, Jack Whitehead, Carl Wunsch, Phil Yecko

GFD Lecture Schedule

PRINCIPAL LECTURES

June 15 - June 19: Charlie Doering, Univ. of Michigan (substitute lecturer Oliver Bühler, June 18)

Mathematical Foundations of Stochastic Processes (Lectures 1-5)

June 22 - June 26: Henk Dijkstra, Utrecht University

Climate Variability and its Null Hypothesis (Lecture 6)
The El Niño Phenomenon (Lecture 7)
The El Niño Phenomenon (cont'd) (Lecture 8)
Dansgaard-Oeschger Events (Lecture 9)
Dansgaard-Oeschger Events (cont'd) & Ocean Western Boundary Current Variability (Lecture 10)

SEMINARS

June 29 - July 3:

Monday, June 29: Oliver Bühler, New York University
Particle dispersion by linear and nonlinear stochastic waves
Tuesday, June 30: Rich Kerswell, University of Bristol
Boundary inflow, Taylor-Couette flow and accretion disks
Wednesday, July 1: Fabio Del Sordo, Yale University
Understanding stellar activity to search for Earth-like exoplanets
Thursday, July 2: Colm-cille Caulfield, University of Cambridge
Stirring solutions: Mixing industrial challenges with mathematics
Friday, July 3: Daniel Lecoanet, University of California, Berkeley
Internal wave excitation by turbulent convection

July 6 - July 10

Monday, July 6: Dhrubaditya Mitra, Nordic Institute for Theoretical Physics
Fluid dynamics with (large and small) balls
Tuesday, July 7: Charlie Doering, University of Michigan
Heat transport in turbulent (and not-so-turbulent) convection

Wednesday, July 8: Gregory Wagner, University of California, San Diego

Coupled evolution of near-inertial waves and quasi-geostrophic flow

Thursday, July 9: Greg Chini, University of New Hampshire

The pure Langmuir turbulence regime

Friday, July 10: Jean-Luc Thiffeault, University of Wisconsin, Madison

Random kicks due to swimming organisms

July 13 - July 17

Monday, July 13: David Goluskin, University of Michigan

Canonical convection configurations

Tuesday, July 14: Andrew Wells, University of Oxford

Layers, ripples, and channels: Shocks and surprises with ice-ocean interaction

Wednesday, July 15: Brad Marston, Brown University

Statistics of stochastically driven jets

Thursday, July 16, a.m.: Alexandros Alexakis, Ecole Normale Supérieure

Rotating Taylor-Green flow

Thursday, July 16, p.m.: Nigel Goldenfeld, University of Illinois at Urbana-Champaign

The emergence of collective modes, ecological collapse and directed percolation at the laminar-turbulence transition in pipe flow

Friday, July 17: Glenn Flierl, Massachusetts Institute of Technology

Arms race, seesaw or speciation – mathematics of evolution

July 20 – July 24

Monday, July 20: Woosok Moon, University of Cambridge

A stochastic Arctic sea ice model: Toward a smart gambling

Tuesday, July 21: Stefan Llewellyn Smith, University of California, San Diego

Tsunami propagation using a consistent acoustic-gravity wave formulation

Wednesday, July 22: John Bush, Massachusetts Institute of Technology

Pilot-wave hydrodynamics

Thursday, July 23: Carl Wunsch, Massachusetts Institute of Technology

Circulation and tides of an ice-covered earth

Friday, July 24: Rob Phillips, California Institute of Technology

Do surfers know things that geophysical fluid dynamicists don't?

July 27 – July 31

Monday, July 27: Ali Mashayek, Massachusetts Institute of Technology
Ocean mixing: From small-scale turbulence to large-scale meridional overturning circulation
Tuesday, July 28: Jack Whitehead, Woods Hole Oceanographic Institution
The continental drift convection cell – continuing on from Lou Howard's formulation
Wednesday, July 29: Edward Bolton, Yale University
Implications of the nonlinear seawater equation of state on salt fingers and cabbeling
Thursday, July 30: Srikanth Toppaladoddi, Yale University
Tailoring boundary geometry to optimize heat transport in turbulent convection
Friday, July 31: Victor Yakhot, Boston University
Small-scale universality, intermittency and turbulence modeling

August 3 – August 7

Monday, August 3: Bill Young, Scripps Institution of Oceanography
Swell problems
Tuesday, August 4, a.m.: Predrag Cvitanovic, Georgia Institute of Technology
Noise is your friend, or: The best possible resolution of state space in presence of noise
Tuesday, August 4, p.m.: Jim Anderson, Stevens Institute of Technology
Happy birthday general relativity: What has happened in the last century
Wednesday, August 5: Emma Boland, British Antarctic Survey
Mixing by meso-scale eddies in the southern ocean: Two different perspectives
Thursday, August 6, a.m.: Ed Spiegel, Columbia University
Beyond Navier-Stokes
Thursday, August 6, p.m.: Francesco Mancarella, Nordic Institute for Theoretical Physics
Using fluids to stiffen solids
Friday, August 7: Katepalli Sreenivasan, New York University
Confessions of ignorance on turbulent convection

FELLOWS' REPORTS

August 17 – August 20

Monday, August 17:

Gunnar Peng, University of Cambridge
The diffusion fish

Tom Eaves, University of Cambridge

Noisy homoclinic pulse dynamics: Back to the origin

Andre Souza, University of Michigan

Instantons in the presence of chaos: Is noise really your friend?

Tuesday, August 18:

Giovanni Fantuzzi, Imperial College London

Bounds for deterministic and stochastic dynamical systems using sum-of-squares programming

Tom Beucler, Massachusetts Institute of Technology

Large-scale advection, condensation and diffusion of water vapor

Wednesday, August 19:

Cesar Rocha, University of California, San Diego

Scaling ocean stratification: Coupled asymptotic model equations for strongly stratified flows

Chris Spalding, California Institute of Technology

The most catastrophic catastrophe: Population dynamics under random extreme events

Yana Bebieva, Yale University

Punctuated plume penetration: Entrainment dynamics of the layered filling box

Thursday, August 20:

Florence Marcotte, Institut de Physique du Globe de Paris

Fast cooling of a hot disc

Anna Fitzmaurice, Princeton University

A stochastic approach to examining the predictability of Arctic sea ice

Lecture 1 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Tom Beucler & Andre Souza

June 15, 2015

We will talk about stochastic processes in general but focusing on differential equations with white noise coefficients in them. We will try to give an intuitive feeling for this field. There is actually some sense in the whole thing. By the end of the lectures you should have a clue of how to model systems with white noise.

When we are dealing with many-body nonlinear systems, there are too many variables. They are too high dimensional to extract useful information, forcing us to use reduced models. In applied math there is a great tradition of getting reduced models from large or small parameters in equations or a separation of scales in time or length. We also model things with noise. How do we know the difference between random and complicated? Well we don't. This is a difficult concept.

The great victory of this approach is statistical mechanics. Settling for bulk quantities of interest as opposed to 10^{23} degrees of freedom was tractable and has stimulated people in the physics community to try to take this concept to the extreme. There are a lot of specific successes and examples, but no general theory.

Specifically what we be talking about:

1. Markov diffusion processes
2. Brownian motion
3. Gaussian white noise
4. Stochastic differential equations
5. Fokker-Planck equations or Forward Kolmogorov equations
6. Mean first passage times

SDEs are what you write down and the Fokker-Planck equations are what you can actually solve. The latter equations give you the evolution for a probability density function.

A random variable X is characterized by its cumulative distribution function (CDF) which is the probability that a random variable is below some scalar. In symbols ($\mathbb{P}(X \leq x)$). It is a monotonically increasing function with values between 0 and 1. It is "continuous from the right". The probablity distribution function is the derivative of the probability with respect to x.

$$\text{PDF} = \frac{d\text{CDF}}{dx}$$

Another way to say it is PDF $= \mathbb{P}(x \leq X \leq x + dx)/dx$ where dx is an infinitesimal or

$$\mathbb{P}(a < x < b) = \int_a^b \text{PDF}(x)dx$$

Stochastic Processes are a random functions of an "index set" which we will call time.

$$X(t) = \text{Random variable}$$

We can plot $X(t)$ (for a given realization of the random variable), which we will assume to be continuous. We can ask the question "what is the probability that $X(t)$" falls in some window. We will now introduce some notation

$$\text{PDF of } X(t) = \rho(x, t)$$

This is not enough to answer all the statistical questions that we would like to pose. We also need the joint distribution functions ($\rho(x_1, t_1; x_2, t_2)$ 2-time) and $\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1)$ n-time. A property of white noise is that $\rho(x_1, t_1)\rho(x_2, t_2)$. This type of process does not have enough structure for us to do modeling. We don't need an uncountable number of joint distributions to have a well-defined probability space. Smooth things tend to have a memory associated with them, thus white noise won't be continuous. The condition

$$1 = \int_{-\infty}^{\infty} \rho(x, t)dx$$

says that "I exist" and the compatibility condition is

$$\rho(x_1, t_1) = \int_{-\infty}^{\infty} \rho(x_1, t_1; x_2, t_2)dx_2.$$

This says that the probability of going through one window is the same as going through the same window as well as an infinitely large window. See Figure 1.

To answer sensible questions about the random process we need all of the joint probability functions. Suppose that we just give the joint probabilities, then we need to check the compatibility conditions.

$$\rho(x_1, t_1) \tag{1}$$
$$\rho(x_1, t_1; x_2, t_2) \tag{2}$$
$$\rho(x_1, t_1; x_2, t_2, ; x_3, t_3) \tag{3}$$
$$\vdots \tag{4}$$

We need to define expectations, averages, and moments of the random variables. The expectation of a random variable is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho(x)dx = \langle x \rangle = \bar{X}.$$

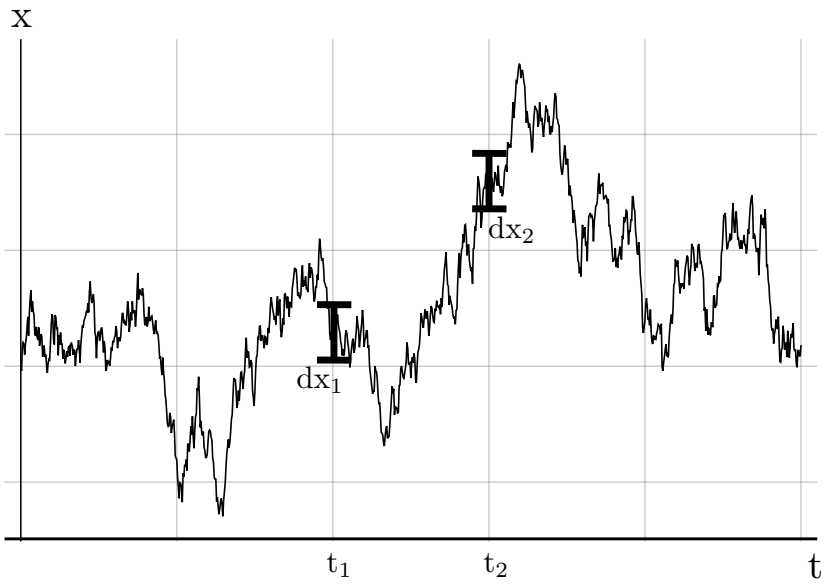

Figure 1: $\rho(x_1, x_2; t_1, t_2)dx_1dx_2$ is the probability that the process $X(t)$ passes through windows of sizes dx_1 and dx_2 at times t_1 and t_2 respectively.

We can also calculate

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(x)\rho(x)dx$$

and get moments

$$\mathbb{E}(X^N) = \int_{-\infty}^{\infty} x^N \rho(x)dx.$$

Furthermore we have

$$\mathbb{E}\left(\prod_{j=1}^{N} X(t_j)\right) = \int \cdots \int \prod_{j=1}^{N} x_j \rho(x_1, t_1; ...; x_N, t_N)dx_1 \cdots dx_N$$

which are known as the n-point correlation functions. The two point correlation function is sometimes known as THE correlation function. The moments DO NOT determine the probability distribution (in general). If it is the case that

$$\mathbb{E}(X(t)X(s)) = \mathbb{E}(X(t))\mathbb{E}(X(s))$$

we say the variables are uncorrelated. This does not imply independence, but independence does imply uncorrelated.

One more thing. Let's talk about this idea of independence. Suppose that two events A and B happen. We can look at $\mathbb{P}(A), \mathbb{P}(B)$ and $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$. If the variables are independent then $\mathbb{P}(A|B) = \mathbb{P}(A)$. For stochastic processes we may want to know things like "given that my random variable went through window 1, what is the probability that it goes through window 2?" We write this as $\rho(x_2, t_2|x_1, t_1) = \rho(x_2, t_2; x_1, t_1)/\rho(x_1, t_1)$.

We have all these joint probability functions

$$\rho(x_n, t_n|x_{n-1}, t_{n-1}; ...; x_1, t_1) \equiv \frac{\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1)}{\rho(x_{n-1}, t_{n-1}; ...; x_1, t_1)}$$

which is the probability of going through my latest window given that I went through all the other windows. We are now in a position to define Markov processes. If

$$\rho(x_n, t_n|x_{n-1}, t_{n-1}; ...; x_1, t_1) = \rho(x_n, t_n|x_{n-1}, t_{n-1}),$$

we can reconstruct the n-point distribution function

$$\rho(x_n, t_n; x_{n-1}, t_{n-1}; ...; x_1, t_1) = \rho(x_1, t_1) \prod_{j=2}^{j=N} \rho(x_j, t_j|x_{j-1}, t_{j-1}).$$

A Markov process is independent of the past, given the present. An example of a Markov process is a first order ODE.

Brownian motion (which is the same as Wiener process) is our next topic. These are random functions of time denoted by

$$W(t) = W_t.$$

The probability density at $t = 0$ is a delta function, $\rho(\omega, 0) = \delta(\omega)$. The transition density is

$$\rho(\omega, t|\omega', t') = \frac{1}{\sqrt{2\pi(t - t')}} e^{-\frac{1}{2}\frac{(w-w')^2}{t-t'}}$$

We have

$$\rho(\omega, t) = \int_{-\infty}^{\infty} \rho(\omega, t|\omega', 0)\rho(\omega', 0)d\omega' = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{\omega'^2}{t}}$$

There are some properties of Brownian motion that are absolutely essential to understanding white noise which we will talk about next time.

Lecture 2 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Gunnar Peng & Chris Spalding

June 16, 2015

1 Examples for last lecture

The moments do not determine the distribution. Consider the log-normal distribution and perturbed versions, whose PDFs are

$$f_0(x) = (2\pi)^{1/2} x^{-1} e^{-(\log x)^2/2}, \qquad f_a(x) = f_0(x)[1 + a\sin(2\pi \log x)] \text{ with } |a| \le 1. \tag{1}$$

It can be verified that these have the same moments – the nth moment is $e^{n^2/2}$ – even though the distributions are different. (Roughly speaking, the moments do determine the distribution if they do not grow too quickly, and evidently $e^{n^2/2}$ is too quick.)

Uncorrelation does not imply independence. The random variables

$$X \sim \text{Uniform}[-1, 1] \qquad \text{and} \qquad Y = X^2 \tag{2}$$

are clearly not independent. However, if we consider their mutual covariance,

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0, \qquad \mathbb{E}[X] = 0, \qquad \mathbb{E}[Y] = \mathbb{E}[X^2] = 1/3 \qquad \Rightarrow \qquad \mathbb{E}[XY] = 0 = \mathbb{E}[X]\mathbb{E}[Y], \tag{3}$$

it becomes apparent that the variables are uncorrelated.

2 Brownian motion / Wiener process (continued)

Recall. The Wiener process $W(t) = W_t$ has transition probabilities and initial condition

$$\rho(w, t \mid w', t') = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{1}{2}\frac{(w-w')^2}{t-s}\right) \text{ for } t \ge s, \qquad \rho(w, 0) = \delta(w), \tag{4}$$

which is illustrated in Figure 2. From this *continuous* distribution we can recover the *discretized*, n-time, probability density

$$\rho(w_n, t_n; \ldots; w_1, t_1) = \rho(w_n, t_n \mid w_{n-1}, t_{n-1}) \ldots \rho(w_2, t_2 \mid w_1, t_1)\rho(w_1, t_1) \qquad \text{for} \qquad t_n > \cdots > t_1. \tag{5}$$

Note that this is a natural factorization into independent increments, reflecting the Markovian property that each future increment is independent of the past.

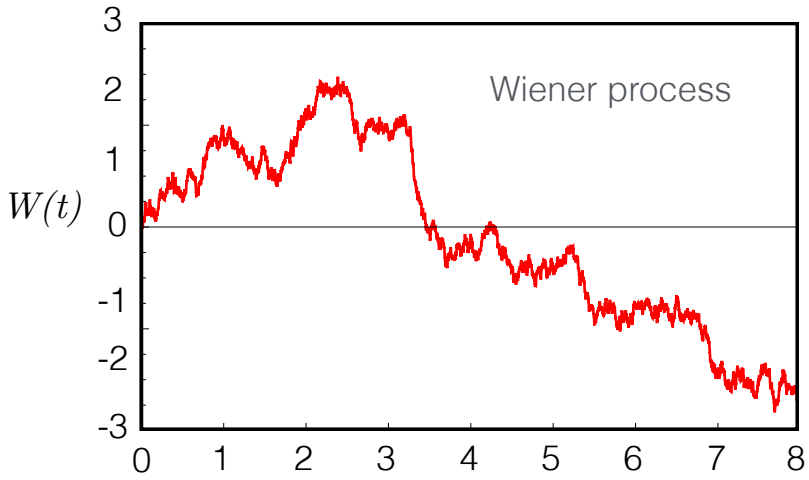

Figure 1: A typical realization of a Wiener process beginning at $W(0) = 0$. The step size used was $\Delta t = 0.001$.

To simulate the Wiener process on a computer, we discretize time from, say, $t_0 = 0$ to t_N, and start with $w_0 = 0$. Given the current value w_n, we repeatedly draw the next random value w_{n+1} using the probability distribution $\rho(w_{n+1}, t_{n+1} \mid w_n, t_n)$ given by Equation 4. This yields a single realization of the process; an example is shown in figure 1.

For comparison, Figure 3 displays both the continuous distribution (Equation 4) and a discrete distribution, obtained by sampling multiple realizations of the Wiener process (a typical example of which is shown in Figure 1). As is apparent, the continuous and discrete distributions look very similar, at least with the sample size of 10,000 used here. Note that the size of the time steps used to obtain realizations of the process does not affect the distribution, since the distribution obtained is given by equation 5, which is not an approximation.

Moments. The odd moments of a Wiener process vanish by symmetry, while the even moments may be calculated through integration by parts:

$$\mathbb{E}\left[W(t)^{2n+1}\right] = \int_{-\infty}^{\infty} w^{2n+1}\rho(w,t)\,\mathrm{d}w = 0, \tag{6}$$

$$\mathbb{E}\left[W(t)^{2n}\right] = \int_{-\infty}^{\infty} w^{2n}\rho(w,t)\,\mathrm{d}w = \frac{(2n)!}{2^n n!}t^n$$

$$= (2n-1)\cdot(2n-3)\cdot\cdots\cdot 3\cdot 1\, t^n, \tag{7}$$

$$\tag{8}$$

$$\text{e.g.} \qquad \mathbb{E}[W(t)^2] = t, \qquad \mathbb{E}[W(t)^4] = 3t^2, \qquad \mathbb{E}[W(t)^6] = 15t^3. \tag{9}$$

Correlation: It can be shown that, for $t \geq s$,

$$\mathbb{E}[W(t)W(s)] \equiv \int_{-\infty}^{\infty}\mathrm{d}w \int_{-\infty}^{\infty}\mathrm{d}w'\, w\, w'\, \rho(w,t;w',s) = s, \tag{10}$$

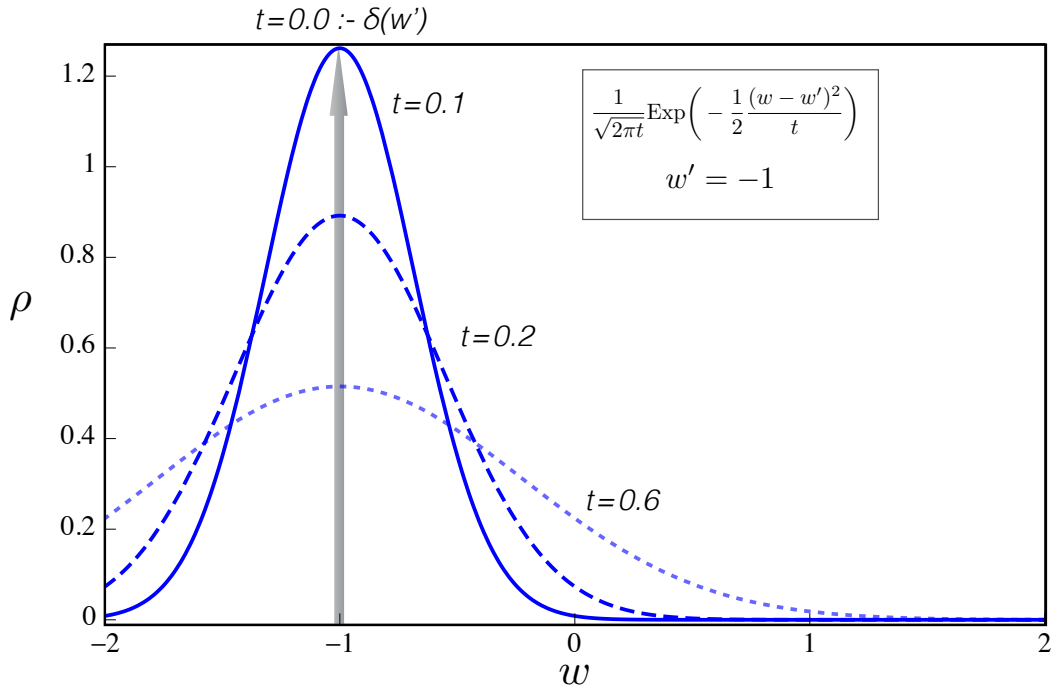

Figure 2: Illustration of the transition probability (density) appropriate for a Wiener process as a function of time. The initial distribution, set at time $t = s = 0$, is a delta function centered on $w' = -1$.

and hence, in general $\mathbb{E}[W(t)W(s)] = \min(t, s)$.

Exercise: Prove it the old-fashioned way (i.e., by changing variables and integrating)! One may get the answer through the slicker method below:

$$\begin{aligned}\mathbb{E}[W(t)W(s)] &= \mathbb{E}[\{W(t) - W(s)\}W(s) + W(s)^2] \\ &= \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s)] + s = s,\end{aligned} \tag{11}$$

where we used that the increment $W(t) - W(s)$ is independent of the past $W(s)$ provided that $t \geq s$.

2.1 Discrete analysis

Consider a discrete increment $\Delta W(t) = W(t + \Delta t) - W(t)$. It has mean and variance

$$\mathbb{E}[\Delta W(t)] = \underbrace{\mathbb{E}[W(t + \Delta t)]}_{0} - \underbrace{\mathbb{E}[W(t)]}_{0} = 0, \tag{12a}$$

$$\mathbb{E}[\Delta W(t)^2] = \underbrace{\mathbb{E}[W(t + \Delta t)^2]}_{t+\Delta t} - 2\underbrace{\mathbb{E}[W(t + \Delta t)W(t)]}_{t\,\text{(the earlier time)}} + \underbrace{\mathbb{E}[W(t)^2]}_{t} = \Delta t. \tag{12b}$$

This suggests that $\Delta W(t) = O(\sqrt{\Delta t})$. Hence $\Delta W/\Delta t = O(1/\sqrt{\Delta t})$ which will not converge as we take the continuum limit. Hence, $W(t)$ is not differentiable. However, there is no

reason we can't simply tabulate the random variable $\Delta W(t)/\Delta t$ for many time-steps of finite width $\Delta t > 0$. We call this Gaussian white noise (GWN). From realizations of $W(t)$ (such as the one illustrated in Figure 1), we can explicitly calculate $\Delta W/\Delta t$ (see Figure 4), provided that Δt is a multiple of the time step used to generate $W(t)$. The GWN thus created remains confined to the 'x'-axis - reflecting its statistically steady nature. Deviations from zero occur with a typical magnitude that is of order $1/\sqrt{\Delta t}$ where Δt^{-1} is the rate at which the derivative is sampled.

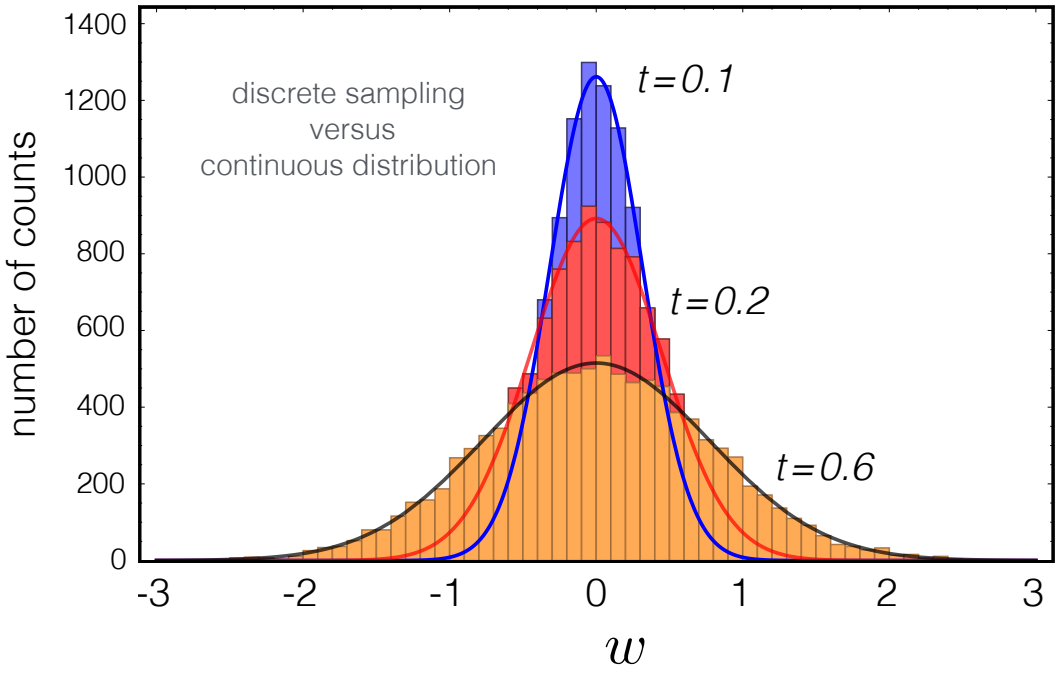

Figure 3: A comparison between the continuous distribution (as plotted in Figure 2) and the discrete distribution obtained by sampling multiple realizations of the Wiener Process (such as the one shown in Figure 1) starting with $W = 0$ at $s = 0$. Specifically, we recorded the value of $W(t)$ for $N = 10,000$ realizations of the Wiener process at the three values of time illustrated in Figure 2 ($t = 0.1, 0.2$ and 0.6). The resulting histograms, binned in groups of width 0.1, are presented. The respective continuous distributions, with mean $w' = 0$, are superimposed for comparison.

2.2 Continuous analysis

We (formally) define Gaussian white noise as

$$\xi(t) \equiv \frac{dW}{dt}. \tag{13}$$

This is a *distribution*-valued process, in contrast to a *function*-valued process. Hence, we investigate it through integration against smooth test functions (with compact support). If

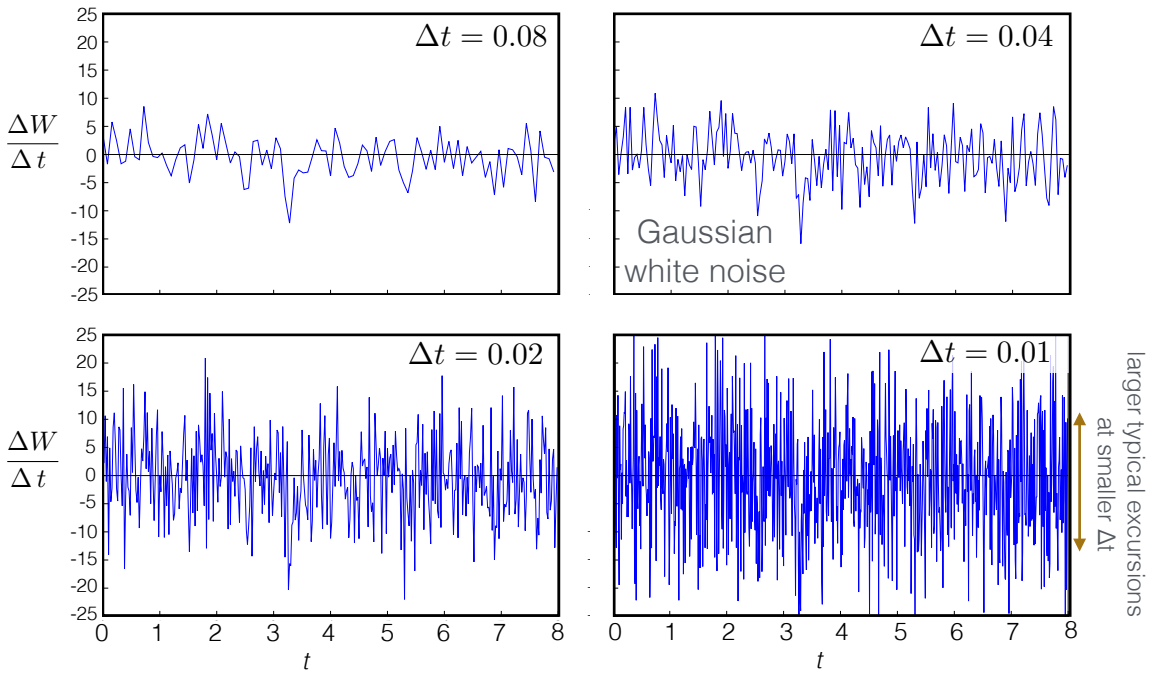

Figure 4: Numerical calculation of $\Delta W(t)/\Delta t$ sampled from the specific realization of $W(t)$ that is illustrated in Figure 1. We sample with 4 different values of the discrete time interval Δt (0.08, 0.04, 0.02 and 0.01). Gaussian white noise (GWN) is defined through such an operation. The GWN arising out of the Wiener process is statistically steady and confined to the 'x'-axis. Fluctuations from zero occur with a typical magnitude of the same order as $\sigma^{-1} = (\Delta t)^{-1/2}$. When compared to Figure 1 just after $t = 3$, for example, $W(t)$ begins to fall at the point where a large negative excursion occurs in $\Delta W/\Delta t$.

$\varphi(t)$ is such a function, then we define

$$\xi[\varphi] \equiv \int_0^\infty \xi(t)\varphi(t)\,\mathrm{d}t \equiv -\int_0^\infty W(t)\varphi'(t)\,\mathrm{d}t \qquad \text{(using integration by parts)}. \tag{14}$$

This is a Gaussian random variable with mean $\mathbb{E}[\xi[\varphi]] = -\int \mathbb{E}[W(t)]\varphi'(t)\,\mathrm{d}t = 0$. (We assume that integrals and expectations commute when necessary.) For the variance, we compute

$$\mathbb{E}\left[\xi[\varphi]^2\right] = \mathbb{E}\left[\int_0^\infty W(t)\varphi'(t)\,\mathrm{d}t \int_0^\infty W(s)\varphi'(s)\,\mathrm{d}s\right] = \int_0^\infty \int_0^\infty \underbrace{\mathbb{E}\left[W(t)W(s)\right]}_{\min(t,s)} \varphi'(t)\varphi'(s)\,\mathrm{d}t\,\mathrm{d}s. \tag{15}$$

We simplify the s-integral first, by integrating by parts:

$$\int_0^\infty \varphi'(s)\min(t,s)\,\mathrm{d}s = -\int_0^t \varphi(s)\,\mathrm{d}s \qquad \Rightarrow \qquad \mathbb{E}\left[\xi[\varphi]^2\right] = -\int_0^\infty \varphi'(t)\left(\int_0^t \varphi(s)\,\mathrm{d}s\right)\mathrm{d}t, \tag{16}$$

where the first equality came from noticing that $\min(t,s)$ as a function of s has a gradient of unity for $s<t$ and zero elsewhere.

A second integration by parts yields the variance,

$$\boxed{\mathbb{E}\left[\xi[\varphi]^2\right] = \int_0^\infty \varphi(t)^2\,\mathrm{d}t = \int_0^\infty\int_0^\infty \varphi(t)\varphi(s)\delta(t-s)\,\mathrm{d}t\,\mathrm{d}s.} \tag{17}$$

Comparing the final expression with the formal calculation

$$\mathbb{E}\left[\xi[\varphi]^2\right] = \mathbb{E}\left[\int_0^\infty \xi(t)\varphi(t)\,\mathrm{d}t\int_0^\infty \xi(s)\varphi(s)\,\mathrm{d}s\right] = \int_0^\infty\int_0^\infty \mathbb{E}\left[\xi(t)\xi(s)\right]\varphi(t)\varphi(s)\,\mathrm{d}t\,\mathrm{d}s \tag{18}$$

suggests the formal result

$$\boxed{\mathbb{E}[\xi(t)\xi(s)] = \delta(t-s).} \tag{19}$$

A similar calculation yields $\mathbb{E}\left[\xi[\varphi]\xi[\psi]\right] = \int_0^\infty \varphi(t)\psi(t)\,\mathrm{d}t$.

We note that white noise is stationary, i.e. its statistics are independent of time. For any stationary process, the covariance is

$$C(t,s) \equiv \mathbb{E}[X(t)X(s)] - \mathbb{E}[X(t)]\mathbb{E}[X(t)] = \mathbb{E}[X(t-s)X(0)] - \mathbb{E}[X(0)]^2 = c(t-s), \tag{20}$$

i.e. a function of $t-s$ only. We define the power spectrum as the Fourier transform

$$S(\omega) = \int_{-\infty}^\infty c(t)e^{-i\omega t}\,\mathrm{d}t. \tag{21}$$

For Gaussian white noise,

$$c(t-s) = \mathbb{E}[\xi(t)\xi(s)] = \delta(t-s) \qquad \Rightarrow \qquad S(\omega) = \int_{-\infty}^\infty \delta(t)e^{-i\omega t}\,\mathrm{d}t = 1. \tag{22}$$

The noise is called "white" because it has a flat power spectrum, i.e. the same amount of energy at each frequency.

3 Stochastic differential equations, a.k.a. Langevin equations

A stochastic differential equation (SDE) is one that contains noise terms $\xi(t)$.

Example. What is the solution $X(t)$ to the equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \xi(t), \tag{23}$$

with $X(0)=0$? It's clearly $X(t)=W(t)$!

Now, each $X(t)$ is but one random realization of the variable X described by the transition density ρ. In order to describe the spatial and temporal evolution of an ensemble of such realizations of $X(t)$, it is appropriate to ask how ρ itself evolves. For a pure Wiener

process **(Exercise)**, show that the transition density satisfies the diffusion equation and the initial condition

$$\rho(x,t \mid y,s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{1}{2}\frac{(x-y)^2}{t-s}\right) \qquad \Rightarrow \qquad \frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2} \quad \text{for} \quad t > s,$$
$$\rho(x,s) = \delta(x-y). \tag{24}$$

Since distributions can only be manipulated linearly, the most general SDE (without time delay) has the form

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = f(X(t),t) + g(X(t),t)\xi(t), \tag{25a}$$

which some mathematicians write (in a futile attempt to avoid differentiating non-differentiable things) as

$$\mathrm{d}X_t = f(X_t,t)\,\mathrm{d}t + g(X_t,t)\,\mathrm{d}W_t. \tag{25b}$$

The first term (with the function f) represents a deterministic, 'drift', term such as may be found in a regular dynamical systems equation. The second term describes stochastic noise with amplitude g.

The equivalent integral forms may be written as

$$X(t) = X(0) + \int_0^t f(X(s),s)\,\mathrm{d}s + \int_0^t g(X(s),s)\xi(s)\,\mathrm{d}s \tag{26a}$$

$$X_t = X_0 \quad + \int_0^t f(X_s,s)\,\mathrm{d}s \quad + \int_0^t g(X_s,s)\,\mathrm{d}W_s. \tag{26b}$$

(Note that the limits on the last integral indicate the range of s, not W_s.) The last integrals, involving the stochastic terms $\xi(s)$ or $\mathrm{d}W_s$, are ambiguous, and we will find that they may give different answers depending on how they are interpreted.

We mainly work with the Itō interpretation, in which (25) is defined as the continuous-time limit of the discrete-time system

$$\Delta X(t) \equiv X(t+\Delta t) - X(t) = f(X(t),t)\,\Delta t + g(X(t),t)\,\Delta W(t), \tag{27}$$

where $\Delta W(t) = W(t+\Delta t) - W(t) \sim N(0,\sqrt{\Delta t})$ (with $X(0)$ determined by an initial condition).

3.1 Conditional expectations

Recall. For $t \geq s$, the conditional expectation

$$\mathbb{E}\left[F(X(t)) \mid X(s) = y\right] = \int_{-\infty}^{\infty} F(x)\rho(x,t \mid y,s)\,\mathrm{d}x. \tag{28}$$

For example,

$$\mathbb{E}\left[W(t) \mid W(s) = w'\right] = \int_{-\infty}^{\infty} w\,\rho(w,t \mid w',s)\,\mathrm{d}w = w'. \tag{29}$$

We now compute the expectation $\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right]$ in two different ways. Firstly, in terms of the density:

$$\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right] = \mathbb{E}\left[X(t+\Delta t) \;\middle|\; X(t)=x\right] - \underbrace{\mathbb{E}\left[X(t) \;\middle|\; X(t)=x\right]}_{x} = \tag{30a}$$

$$= \int x'\,\rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x' - x \underbrace{\int \rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x'}_{\text{1 in disguise}} = \tag{30b}$$

$$= \int (x'-x)\rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x'. \tag{30c}$$

Secondly, making use of (27):

$$\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right] = \mathbb{E}\left[f(X(t),t) \;\middle|\; X(t)=x\right]\,\Delta t + \mathbb{E}\left[\underbrace{g(X(t),t)}_{\text{independent}}\Delta W(t) \;\middle|\; X(t)=x\right] = \tag{31a}$$

$$= f(x,t)\,\Delta t + \underbrace{\mathbb{E}\left[g(X(t),t) \;\middle|\; X(t)=x\right]}_{g(x,t)}\underbrace{\mathbb{E}\left[\Delta W(t) \;\middle|\; X(t)=x\right]}_{0} = f(x,t)\,\Delta t \tag{31b}$$

Hence, we conclude that

$$\boxed{\int_{-\infty}^{\infty} (x'-x)\,\rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x' = f(x,t)\,\Delta t.} \tag{32}$$

Similarly, we compute the expectation $\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right]$ in two different ways:

$$\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right] = \int (x'-x)^2 \rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x' \quad \text{as before, and} \tag{33a}$$

$$\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right] = \mathbb{E}_W \int \left[f(x',t)\,\Delta t + g(x',t)\,\Delta W(t)\right]^2 \rho(x', t+\Delta t|x, t)\,\mathrm{d}x' = \tag{33b}$$

$$= f(x,t)^2\,\Delta t^2 + g(x,t)\underbrace{\mathbb{E}_W[\Delta W(t)]}_{0} + g(x,t)^2\underbrace{\mathbb{E}_W[\Delta W(t)^2]}_{\Delta t}. \tag{33c}$$

Hence,

$$\boxed{\int_{-\infty}^{\infty} (x'-x)^2\,\rho(x', t+\Delta t \;\middle|\; x, t)\,\mathrm{d}x' = f(x,t)^2\,\Delta t^2 + g(x,t)^2\,\Delta t.} \tag{34}$$

Lecture 3 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Florence Marcotte & Cesar B. Rocha

June 17, 2015

Recalling from lecture 2 that we are studying the general first-order linear stochastic ordinary differential equation

$$\Delta X = f(X(t),t) + g(X(t),t)\xi(t)\Delta t\,, \tag{1}$$

where $\xi(t)$ is a Gaussian white noise, and

$$\Delta X(t) = X(t+\Delta t) - X(t)\,, \tag{2}$$

with

$$X(0) = X_0\,. \tag{3}$$

The first and second moments of the transition density are

$$\int (y-x)\rho(y,t+\Delta t|x,t)\mathrm{d}y = \mathbb{E}(\Delta X(t)|X(t)=x) = f(X,t)\Delta t\,, \tag{4}$$

$$\int (y-x)^2\rho(y,t+\Delta t|x,t)\mathrm{d}y = \mathbb{E}(\Delta X(t)^2|X(t)=x) = f(X,t)^2\Delta t^2 + g(X(t),t)^2\Delta t\,, \tag{5}$$

The higher order moments ($n>2$) are smaller than $\mathcal{O}(\Delta t)$:

$$\mathbb{E}(\Delta X(t)^n|X(t)=x) = \wr(\Delta t)\,. \tag{6}$$

The Fokker-Planck equation

The Chapman-Kolmogorov relation for the Markovian probability density ρ gives the transition density from state y at time s to state x at time t (where $s<t$) through an intermediate state z at time u, by integrating over all the possible intermediate states z:

$$\rho(x,t|y,s) = \int \rho(x,t|z,u)\rho(z,u|y,s)\mathrm{d}z\,. \tag{7}$$

This relation will be used here to interpret a stochastic differential equation as a continuous limit of the discrete process above. Let $\varphi(X)$ a smooth test-function.

$$\begin{aligned}\int \varphi(x)\rho(x,t+\Delta t|y,s)\mathrm{d}x &= \int \mathrm{d}x\varphi(x)\int \rho(x,t+\Delta t|z,t)\rho(z,t|y,s)\mathrm{d}z \\ &= \int \mathrm{d}z\rho(z,t|y,s)\int \mathrm{d}x\varphi(x)\rho(x,t+\Delta t|z,t)\,.\end{aligned} \tag{8}$$

Expanding $\varphi(x)$ about $\varphi(z)$ for small Δt we obtain

$$\varphi(x) = \varphi(z) + (x-z)\varphi'(z) + \frac{(x-z)^2}{2}\varphi''(z) + \dots \,. \tag{9}$$

Hence, using the moments formulae (4) through (6), we have

$$\int \varphi(x)\rho(x,t+\Delta t|z,t)\mathrm{d}x = \varphi(z)\underbrace{\int \rho(x,t+\Delta t|z,t)\mathrm{d}x}_{=1} + \varphi'(z)\underbrace{\int (x-z)\rho(x,t+\Delta t|z,t)\mathrm{d}x}_{=f(z,t)\Delta t} \tag{10}$$

$$+ \frac{1}{2}\varphi''(z)\underbrace{\int (x-z)^2\rho(x,t+\Delta t|z,t)\mathrm{d}x}_{=f(z,t)^2\Delta t^2 + g(z,t)^2\Delta t} + \mathcal{O}(\Delta t)\,. \tag{11}$$

Thus

$$\int \varphi(z)\left(\frac{\rho(z,t+\Delta t|y,s) - \rho(z,t|y,s)}{\Delta t}\right)\mathrm{d}z = \tag{12}$$

$$\int \left(\varphi'(z)f(z,t) + \frac{1}{2}\varphi''(z)g(z,t)^2\right)\rho(z,t|y,s)\mathrm{d}z + \mathcal{O}(1)\,. \tag{13}$$

Since $\varphi(z)$ is arbitrary, in the continuous limit we obtain the following partial differential equation (PDE) for the Markovian transition density $\rho(x,t|y,s)$;

$$\frac{\partial}{\partial t}\rho(x,t|y,s) = \frac{\partial}{\partial x}\left[\left(-f(x,t) + \frac{1}{2}\frac{\partial}{\partial x}g(x,t)^2\right)\rho(x,t|y,s)\right]\,, \tag{14}$$

known as the *Fokker-Planck* (or forward Kolmogorov) equation. To solve (14) we need initial data and boundary conditions. Here we will only worry about the former, which is given by

$$\lim_{t\to s}\rho(x,t|y,s) = \delta(x-y)\,, \tag{15}$$

where the limit above is taken from the right (i.e., $t > s$). The derivation of the Fokker-Planck equation can be generalized to higher dimensions. Consider the n-dimensional stochastic process

$$\mathbf{X}(t) = (X_1(t), X_2(t), \dots) = X_i(t)\,. \tag{16}$$

The stochastic differential equation is

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = f_i(\mathbf{x},t) + g_{ij}(\mathbf{x},t)\xi_j(t)\,, \tag{17}$$

where the summation over repeated indices is implicit. The associated Fokker-Planck equation for the Markovian transition density is

$$\frac{\partial}{\partial t}\rho(\mathbf{x},t|\mathbf{y},s) = \frac{\partial}{\partial x_i}\left[\left(-f_i + \frac{1}{2}\frac{\partial}{\partial x_j}g_{ik}g_{jk}\right)\rho(\mathbf{x},t|\mathbf{y},s)\right]\,, \tag{18}$$

where we can introduce the symmetric positive, semi-definite diffusion matrix $D_{ij} \stackrel{\text{def}}{=} g_{ik}g_{jk}$. At this point it is important to remark that different interpretations of the stochastic differential equation lead to different partial differential equations on the transition density. We will take up this important fact later.

Itô's Lemma

Let X_t be a stochastic variable whose evolution is governed by the following SDE, according to Itô's interpretation:

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t\,, \tag{19}$$

and whose transition probability density satisfies the following Fokker-Planck (or Kolmogorov Forward) equation

$$\frac{\partial \rho_X}{\partial t} = \frac{\partial}{\partial X}\left[\left(-f + \frac{1}{2}\frac{\partial}{\partial X}g(X)^2\right)\rho_X\right]\,. \tag{20}$$

We emphasize that the subscripts above are labels – they do not represent partial differentiation.

Now consider a function of the random process $X(t)$ with a well defined inverse

$$Y = F(X) \Longleftrightarrow X = G(Y)\,. \tag{21}$$

We want to find a stochastic differential equation for Y and thus we write

$$\rho_Y \mathrm{d}Y = \rho_X \mathrm{d}X\,, \tag{22}$$

and

$$\frac{\partial}{\partial X} = F'\frac{\partial}{\partial Y} \qquad \text{or} \qquad G'\frac{\partial}{\partial X} = \frac{\partial}{\partial Y}\,, \tag{23}$$

from which we obtain

$$\frac{\partial \rho_Y}{\partial t} = \frac{\partial}{\partial Y}\left[\left(-f + \frac{1}{2}\frac{\partial}{\partial X}g^2\right)F'\rho_Y\right] = \frac{\partial}{\partial Y}\left[\left(-F'f + \frac{1}{2}\frac{\partial}{\partial X}\frac{1}{F'}(gF')^2\right)\rho_Y\right]\,. \tag{24}$$

Finally, the equation for the Markovian transition density is then

$$\frac{\partial \rho_Y}{\partial t} = \frac{\partial}{\partial Y}\left[\left(-F'f - \frac{1}{2}F''g^2 + \frac{1}{2}\frac{\partial}{\partial Y}(gF')^2\right)\rho_Y\right]\,. \tag{25}$$

Using our "recipe", we obtain the associated stochastic differential equation

$$\mathrm{d}Y = \left(F'f + \frac{1}{2}F''g^2\right)\mathrm{d}t + gF'\mathrm{d}W\,, \tag{26}$$

$$\mathrm{d}F(X) = F'(X)\mathrm{d}X + \frac{1}{2}F''(X)(\mathrm{d}X)^2\,, \tag{27}$$

plugging in the equation for $\mathrm{d}X$, using $(\mathrm{d}X)^2 = g^2\mathrm{d}t + \wr(\Delta t)$, we obtain

$$\mathrm{d}F(X) = \left(F'(X)f + \frac{1}{2}F''(X)g^2\right)\mathrm{d}t + gF'\mathrm{d}W\,. \tag{28}$$

Equation (28) is known as Itô's change of variables formula or Itô's lemma (note the extra-term $\frac{1}{2}F''(X)g^2$, due to the presence of the noise).

A linear example: The Ornstein–Uhlenbeck equation

Consider the following model

$$\frac{\mathrm{d}U}{\mathrm{d}t} = -\gamma U + \sigma\xi(t)\,, \tag{29}$$

where γ and σ are constants, which is Langevin equation with a linear damping term (first term on the right hand side), the second term being a fluctuation forcing. This could model for example the acceleration of a solid body immersed in a fluid, resulting on the combination of the Stokes drag (if the fluid happens to have a mean flow) and the constant battering of molecules around the body (white noise). Using Itô's interpretation, the associated Fokker-Planck equation is

$$\frac{\partial}{\partial t}\rho(u,t|v,s) = \frac{\partial}{\partial u}\left[\left(\gamma u + \frac{\sigma^2}{2}\frac{\partial}{\partial u}\rho(u,t|v,s)\right)\right]\,, \tag{30}$$

with the initial condition

$$\lim_{t\to s}\rho(u,t|v,s) = \delta(u-v)\,. \tag{31}$$

The solution is then

$$\rho(u,t|v,s) = \frac{1}{\sqrt{2\pi\Sigma(t-s)}}\exp\left[-\frac{1}{2}\frac{(u-v\mathrm{e}^{-\gamma(t-s)})^2}{\Sigma(t-s)}\right]\,, \tag{32}$$

with

$$\Sigma(t-s) \stackrel{\text{def}}{=} \frac{\sigma^2}{2\gamma}\left(1-\mathrm{e}^{-2\gamma(t-s)}\right)\,. \tag{33}$$

To obtain an equation for the expectation of u, we just take the expectation of (29), to obtain

$$\frac{\mathrm{d}\mathbb{E}(U)}{\mathrm{d}t} = -\gamma\mathbb{E}(U)\,, \tag{34}$$

whose solution is

$$\mathbb{E}(U) = v\mathrm{e}^{-\gamma(t-s)}\,. \tag{35}$$

To derive an equation for the second moment, we first obtain an equation for U^2. Using Itô's change of variables (28) we have

$$F = u^2\,, \qquad F' = 2u\,, \qquad \text{and} \qquad F'' = 2\,, \tag{36}$$

we obtain

$$\mathrm{d}U^2 = -2\gamma U^2\mathrm{d}t + 2U\sigma\mathrm{d}W + \sigma^2\mathrm{d}t\,. \tag{37}$$

or

$$\mathrm{d}U^2 = 2U\mathrm{d}U + \sigma^2\mathrm{d}t\,. \tag{38}$$

The last term in (38) arises from the presence of the noise term. Now take the expectation of (38) to obtain

$$\mathrm{d}\mathbb{E}(U^2) = -2\gamma\mathbb{E}(U^2)\mathrm{d}t + \sigma^2\mathrm{d}t\,, \tag{39}$$

or

$$\frac{\mathrm{d}\mathbb{E}(U^2)}{\mathrm{d}t} = -2\gamma\mathbb{E}(U^2) + \sigma^2\,, \tag{40}$$

with the initial condition

$$\mathbb{E}(U^2) = v^2\,, t \to s\,. \tag{41}$$

The Markovian density equilibrates to a stationary state

$$\lim_{t\to\infty} \rho(u,t|v,s) = \sqrt{\frac{\gamma}{\pi\sigma^2}} \mathrm{e}^{-\frac{\gamma}{\sigma^2}u^2}\,, \tag{42}$$

which is in the form

$$\rho^{stat}(u,t|v,s) = \rho(u,t|v,s)\rho^{stat}(v)\,. \tag{43}$$

For the stationary state, the covariance is

$$\mathbb{E}^{stat}(U(t)U(s)) = \int\int uv\rho(u,t;v,s)\mathrm{d}u\mathrm{d}v = \frac{\sigma^2}{2\gamma}\mathrm{e}^{-\gamma|t-s|}\,. \tag{44}$$

The power spectrum is simply the Fourier transform of the autocovariance $C(t)$

$$S(\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} C(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\,. \tag{45}$$

Hence we obtain the Lorenzian spectrum

$$S(\omega) = \int_{-\infty}^{\infty} \frac{\sigma^2}{2\gamma}\mathrm{e}^{-\gamma|t|-\mathrm{i}\omega t}\mathrm{d}t = \frac{\sigma^2}{\gamma^2+\omega^2}\,. \tag{46}$$

Note that with (46) we can recover a flat spectrum in the white noise limit $\gamma \to \infty$ with appropriately rescaled noise amplitude.

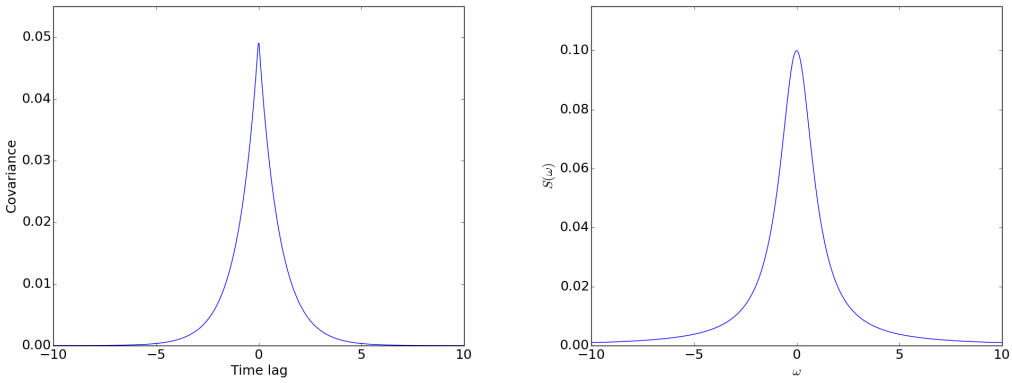

Figure 1: The stationary covariance (left) and spectrum (right) for the Ornstein–Uhlenbeck process with $\gamma = 1$ and $\sigma^2 = 0.1$.

A nonlinear example: The logistic equation

As an example of nonlinear equation, we consider the logistic equation

$$\mathrm{d}X = (\mu X - X^2)\mathrm{d}t\,, \tag{47}$$

where μ represents the ratio of birth rate to death rate, with initial condition $X(0) = X_0 > 0$. This is also known as the *Verhulst model* for population dynamics. The behavior of the deterministic equation (47) is well known; at large t, the solution approaches the fixed point μ (this is sometimes referred to as the carrying capacity of the system).

We now consider a stochastic logistic equation by adding a white noise term to the ratio of birth rate to death rate:

$$\mu(t) = \bar{\mu} + \sigma\xi(t)\,, \tag{48}$$

with $\bar{\mu}$ a constant. The stochastic differential equation is then

$$\mathrm{d}X = (\bar{\mu}X - X^2)\mathrm{d}t + \sigma X\xi(t)\mathrm{d}t\,, \tag{49}$$

with the associated Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial X}\left[\left(X^2 - \bar{\mu}X + \frac{\sigma^2}{2}\frac{\partial}{\partial X}X^2\right)\rho\right]\,. \tag{50}$$

Figure (2) shows numerical simulations for the stochastic logistic equation with two different initial conditions and various levels of noise. As discussed above, the deterministic behavior is well known; the solution initially grows ($X(0) < \mu$) or decays ($X(0) > \mu$) and asymptotically approaches the steady state, or the carrying capacity of the system, μ. For small noise ($\sigma < 0.1$), the solutions oscillate about the deterministic solution. For moderate noise levels ($\sigma^2 \sim \mathcal{O}(1)$), the solutions become very intermittent. For even larger levels of noise, the solutions show a tendency towards extinction.

If there is a stationary state then it must satisfy

$$0 = \frac{\partial}{\partial X}\left[\left(X^2 - \bar{\mu}X + \frac{\sigma^2}{2}\frac{\partial X^2}{\partial X}\right)\rho^{stat}\right]\,. \tag{51}$$

In general we can write

$$0 = \frac{\partial}{\partial X}\left[\left(-f + \frac{1}{2}\frac{\partial}{\partial X}g(X)^2\right)\rho^{stat}\right]\,, \tag{52}$$

from which we find

$$\rho^{stat}(X) = \frac{\mathcal{N}}{g(X)^2}\exp\left[2\int^X \frac{f(\xi)}{g(\xi)^2}\mathrm{d}\xi\right]\,. \tag{53}$$

With the explicit f and g from the stochastic version of the Verhulst model, we obtain

$$\rho^{stat}(X) = \mathcal{N}X^{2(\bar{\mu}/\sigma^2-1)}\mathrm{e}^{-\frac{2}{\sigma^2}X}\,. \tag{54}$$

The normalization constant is defined as

$$\int \rho^{stat}(X)\mathrm{d}X = 1 \Rightarrow \mathcal{N}\left(\frac{\sigma^2}{2}\right)^{(2\bar{\mu}/\sigma^2-1)}\Gamma\left(\frac{2\bar{\mu}}{\sigma^2} - 1\right) = 1\,, \tag{55}$$

where $\Gamma(x)$ is the gamma function. Note that for $\bar{\mu}/\sigma^2 \leq 1/2$ the normalization constant defined above is not bounded. Thus there is no stationary state with support on $x > 0$. For every small $\epsilon > 0$, the probability for X_t to fall above $\epsilon > 0$ as time goes to ∞ is 1 (convergence in probability).

Figure 3 shows a comparison of the stationary probability density (53) with empirical probability densities at $t = 50$ based on 10000 numerical simulations, similar to those presented in figure 2. With small noise, the probability density function almost Gaussian. As the level of noise increases, the probability density becomes highly skewed (the solution becomes highly intermittent). Moreover, there is a qualitative change in the behavior of the stationary density functions for $\sigma^2 > 1.0$ in (53). With a large level of noise, the probability density function is compressed near $x = 0$ (the probability of extinction is very high).

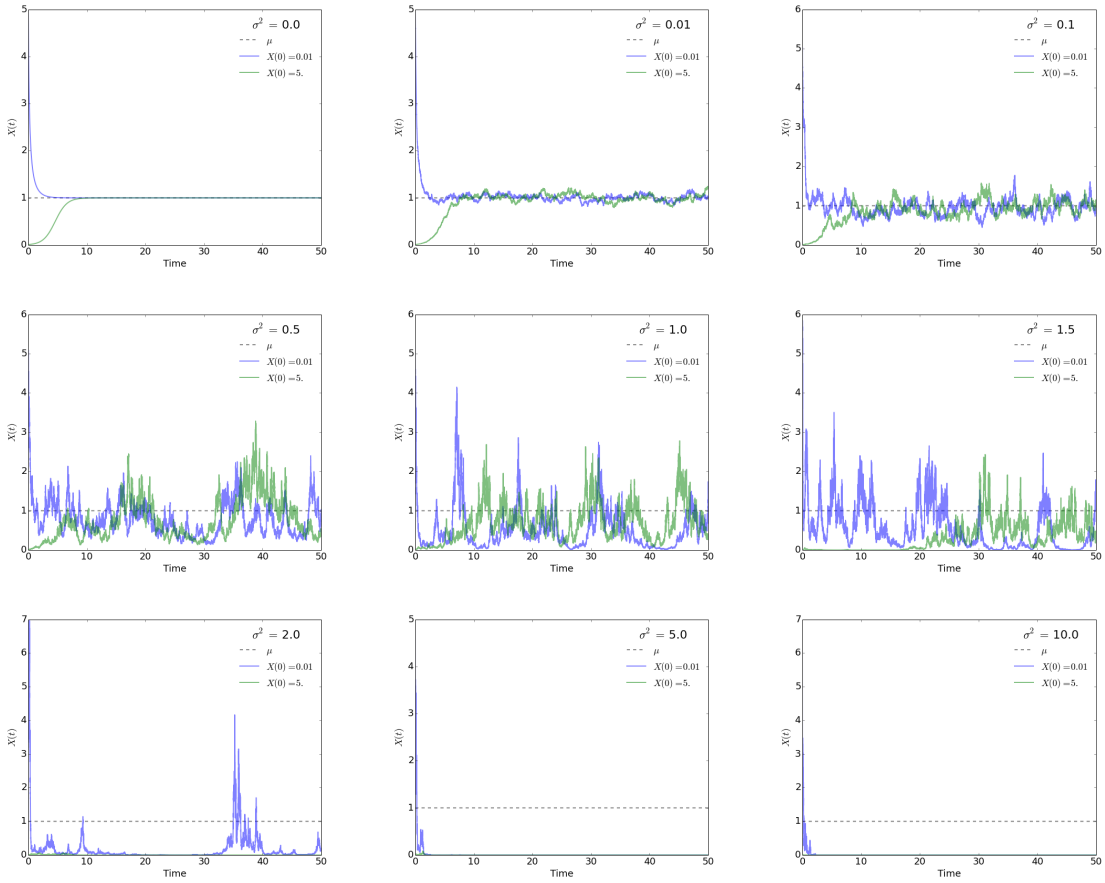

Figure 2: Numerical solutions to the stochastic logistic equation with $\bar{\mu} = 1$ for various levels of noise. Note that $\sigma^2 = 0$ is simply the deterministic solution. For high levels of noise ($\sigma^2 > 1$), one observes a tendency towards extinction.

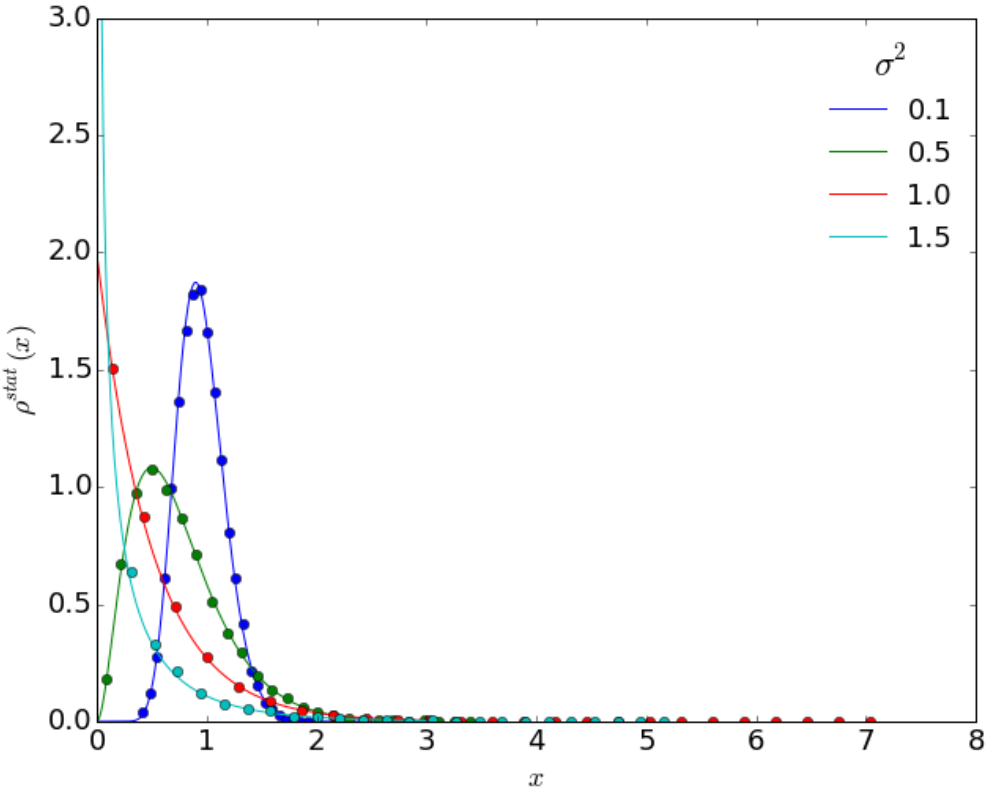

Figure 3: A comparison between the analytical stationary probability density function Eq. (53) (solid lines) against the empirical probability function (dots) at $t = 50$ based on 10000 simulations of the stochastic logistic equation with $\bar{\mu} = 1$. Note the qualitative change at $\sigma^2 \geq 1$. (In these figure labels X from Eq. (53) is written as x.)

Lecture 4 - Mathematical Foundations of Stochastic Processes (substitute lecturer Oliver Bühler)

Oliver Bühler; notes by Tom Eaves & Anna FitzMaurice

June 18, 2015

1 Itō calculus

Recall that for the stochastic differential equation

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t \tag{1}$$

we have the important relations

$$\mathbb{E}(g(X_t)\mathrm{d}W_t) = 0 \tag{2}$$

$$\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t \tag{3}$$

$$\mathbb{E}(\mathrm{d}W(t_1)\mathrm{d}W(t_2)) = \delta(t_1 - t_2)\mathrm{d}t_1\mathrm{d}t_2 \tag{4}$$

We would like to reconcile the fact that we have terms of order $\mathrm{d}t$ and terms of order $\sqrt{\mathrm{d}t}$ in the same equation. Essentially, $\mathrm{d}W_t$ is large and incoherent, whereas $\mathrm{d}t$ is small but coherent, and they act together to result in equal contributions. The fact that $\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t$ means that when attempting to work with the chain rule when changing variables, we need to evaluate more derivatives than expected in order to complete the stochastic differental equation to the correct order. For example,

$$\mathrm{d}(F(X_t)) = F'(X_t)\mathrm{d}X_t + F''(X_t)\mathrm{d}X_t\mathrm{d}X_t/2 + o(\mathrm{d}X_t\mathrm{d}X_t). \tag{5}$$

Additionally, $\mathrm{d}W_t\mathrm{d}W_t = \mathrm{d}t$ requires careful interpretation. Recall that for *finite* increments in the Weiner process,

$$\mathbb{E}(\Delta W^2) = \Delta t \tag{6}$$

and so the infinitesimal statement should be interpreted as any errors associated with approximating $\Delta W^2 \approx \Delta t$ are canceled in the limit of infinitesimal increments which are then summed over as an integral, and this process works because we are summing a family of independent Gaussian-distributed random variables.

With the Itō calculus rule (5) we may re-examine the examples already considered without reference to the Fokker–Planck equation.

1.1 Ornstein-Uhlenbeck Equation

Consider again the Ornstein-Uhlenbeck stochastic differential equation

$$\mathrm{d}U_t = -\gamma U_t \mathrm{d}t + \sigma \mathrm{d}W_t. \tag{7}$$

We can in fact integrate this exactly by multiplying through by $e^{\gamma t}$ to get

$$\mathrm{d}(e^{\gamma t} U_t) = e^{\gamma t} \sigma \mathrm{d}W_t, \tag{8}$$

and so

$$U_t = U_0 e^{-\gamma t} + \sigma e^{-\gamma t} \int_0^t e^{\gamma s} \mathrm{d}W_s, \tag{9}$$

which gives the expectation

$$\mathbb{E}(U_t) = e^{-\gamma t} \mathbb{E}(U_0) \tag{10}$$

and variance

$$\mathbb{E}(U_t^2) = \sigma^2 e^{-2\gamma t} \int_0^t \int_0^t e^{\gamma(s_1+s_2)} \mathbb{E}(\mathrm{d}W(s_1)\mathrm{d}W(s_2)). \tag{11}$$

We could instead obtain these results directly from the stochastic differential equation by forming an equation for $\mathrm{d}(U_t^2)$ using Itō calculus. From the Itō formula (5) we have

$$U_t \mathrm{d}U_t = \frac{\mathrm{d}(U_t^2)}{2} - \frac{\mathrm{d}U_t \mathrm{d}U_t}{2}, \tag{12}$$

and from (7) and the relation $\mathrm{d}W_t \mathrm{d}W_t = \mathrm{d}t$,

$$\mathrm{d}U_t \mathrm{d}U_t = \sigma^2 \mathrm{d}t + o(\mathrm{d}t), \tag{13}$$

which gives

$$\frac{\mathrm{d}(U_t^2)}{2} - \frac{\sigma^2}{2}\mathrm{d}t = -\gamma U_t^2 \mathrm{d}t + U_t \sigma \mathrm{d}W_t. \tag{14}$$

In steady state $\mathbb{E}^s(\mathrm{d}) = 0$ and so

$$\frac{\sigma^2}{2} = \mathbb{E}^s(\gamma U_t^2), \tag{15}$$

which is the fluctuation-dissipation relation for this process.

1.2 Linear population model

Consider the stochastic differential equation

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t. \tag{16}$$

This equation can be interpreted as a random interest rate model.

We have

$$\mathbb{E}(\mathrm{d}X_t) = \mu \mathbb{E}(X_t \mathrm{d}t), \tag{17}$$

and so

$$\mathbb{E}(X_t) = X_0 e^{\mu t}. \tag{18}$$

We also have

$$\mathrm{d}(X_t^n) = nX_t^{n-1}\mathrm{d}X_t + \frac{n(n-1)X_t^{n-2}\mathrm{d}X_t\mathrm{d}X_t}{2} \tag{19}$$

$$= \left(n\mu + \frac{n(n-1)\sigma^2}{2}\right) X_t^n \mathrm{d}t + n\sigma X_t^n \mathrm{d}W_t, \tag{20}$$

and so the n-th moment is

$$\mathbb{E}(X_t^n) = X_0^n \exp\left[\left(n\mu + \frac{n(n-1)\sigma^2}{2}\right)t\right]. \tag{21}$$

Alternatively, using the Itō calculus formula (5), we note that in the absence of noise, we would be interested in $\mathrm{d}(\log(X))$, and so we compute

$$\mathrm{d}(\log(X_t)) = \frac{\mathrm{d}X_t}{X_t} - \frac{\mathrm{d}X_t\mathrm{d}X_t}{2X_t^2} = \frac{\mathrm{d}X_t}{X_t} - \frac{\sigma^2\mathrm{d}t}{2} \tag{22}$$

$$= \mu\mathrm{d}t - \frac{\sigma^2}{2}\mathrm{d}t + \sigma\mathrm{d}W_t. \tag{23}$$

Hence, the solution is

$$X_t = X_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right]. \tag{24}$$

We can reconcile the fact that at first glance the results $\mathbb{E}(W_t) = 0$ and $\mathbb{E}(X_t) = X_0 e^{\mu t}$ appear incompatible with this solution by recognising that the occasions for which $W_t > 0$ and $W_t < 0$ do not contribute equally after exponentiating.

From this solution we see that if $\sigma^2 > 2\mu$, then extinction is guaranteed almost surely, as for the nonlinear population model discussed in a previous lecture. This is since near extinction, X is small, and so the linearised approximation is accurate.

We can find the transition density $\rho(x, t|x_0, 0)$ by solving the Fokker–Planck equation for $\rho_Y(y, t|y_0, 0)$ for the variable $Y_t = \log(X_t)$, since

$$\frac{\partial\rho_Y}{\partial t} = \left(\frac{\sigma^2}{2} - \mu\right)\frac{\partial\rho_Y}{\partial y} + \frac{\sigma^2}{2}\frac{\partial^2\rho_Y}{\partial y^2}, \tag{25}$$

which can be solved with Fourier transforms to give

$$\rho_Y = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{\left(y - y_0 - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right], \tag{26}$$

and so Y_t is normally distributed, meaning that X_t is log-normally distributed, with

$$\rho(x, t|x_0, 0) = \frac{1}{x\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{\left(\log x - \log x_0 - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right]. \tag{27}$$

Now let $\epsilon > 0$. Then,

$$\mathbb{P}(X_t > \epsilon | X_0 = x_0) = \int_\epsilon^\infty \rho(x, t|x_0, 0)\mathrm{d}x = \frac{1}{\sqrt{\pi}}\mathrm{erfc}\left(\frac{\log\epsilon - \log x_0 - (\mu - \frac{\sigma^2}{2})t}{\sqrt{2\sigma^2 t}}\right), \quad (28)$$

and so if $\sigma^2 > 2\mu$, we have that $\mathbb{P}(X_t > \epsilon | X_0 = x_0) \to 0$ as $t \to \infty$, despite the fact that $\mathbb{E}(X_t) = X_0 \exp(\mu t)$. Almost all trajectories decay eventually, but the moments of the distribution grow rapidly, and so in an ensemble we expect an occasional 'success'. We can see that the distribution becomes more shifted towards $x = 0$ as time t increases in Figure 1, in which the distribution for $\mu = 1$, $\sigma = 2$ and $x_0 = 1$ is plotted at various times.

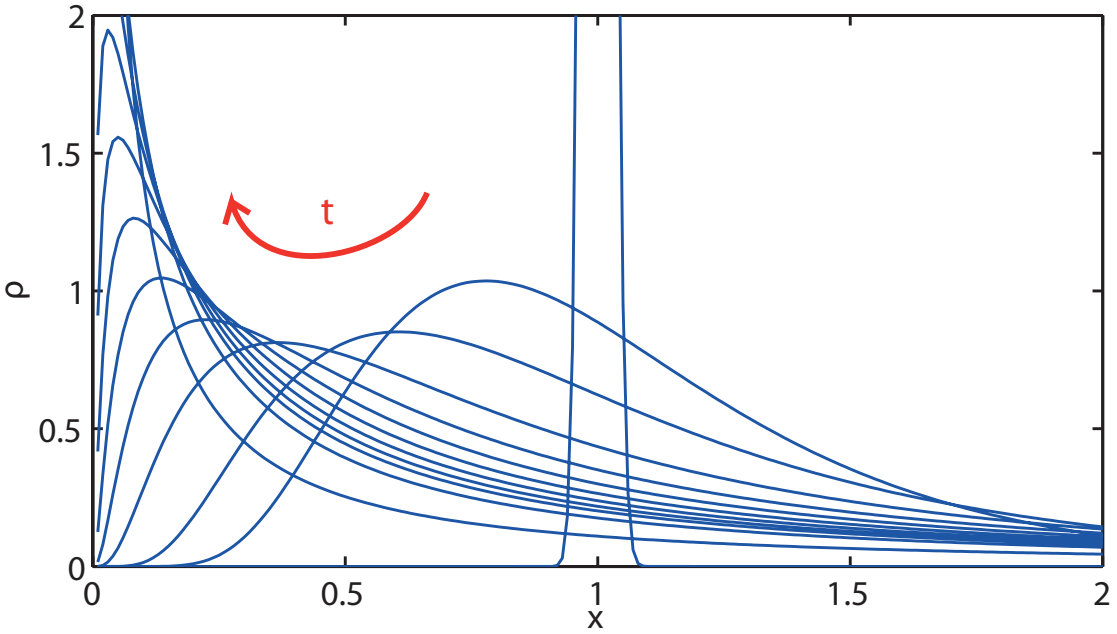

Figure 1: The transition density $\rho(x, t|1, 0)$ for the linear population growth model with $\mu = 1$ and $\sigma = 2$ at times $t = 0.0001$, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1 and 2.

1.3 Nonlinear population model

Consider the stochastic differential equation

$$\mathrm{d}X_t = (\mu X_t - X_t^2)\mathrm{d}t + \sigma X_t \mathrm{d}W_t. \quad (29)$$

Let $Y_t = X_t^{-1}$. Then, the Itō calculus formula (5) gives

$$\mathrm{d}Y_t = -\frac{\mathrm{d}X_t}{X_t^2} + \frac{\mathrm{d}X_t \mathrm{d}X_t}{X_t^3} \quad (30)$$

$$= -\frac{(\mu X_t - X_t^2)\mathrm{d}t + \sigma X_t \mathrm{d}W_t}{X_t^2} + \frac{\sigma^2 X_t^2 \mathrm{d}t}{X_t^3} \quad (31)$$

$$= (1 - (\mu - \sigma^2)Y_t)\mathrm{d}t - \sigma Y_t \mathrm{d}W_t, \quad (32)$$

which is linear in Y_t.

Then,

$$\mathbb{E}(\mathrm{d}Y_t) = -(\mu - \sigma^2)\mathbb{E}(Y_t \mathrm{d}t), \tag{33}$$

which gives

$$\mathbb{E}(Y_t) = Y_0 e^{-(\mu-\sigma^2)t} \to \infty \text{ as } t \to \infty \text{ if } \mu < \sigma^2, \tag{34}$$

as expected from a previous lecture.

2 Probability currents and steady states

Recall the Fokker-Planck equation for $\mathbf{x} \in \mathbf{R}^n$ and noise $\mathbf{w} \in \mathbf{R}^m$ subject to

$$\mathrm{d}X_i = f_i \mathrm{d}t + g_{ij}\mathrm{d}W_j, \tag{35}$$

namely

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{f}) = \frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{D_{ij}}{2}\rho\right), \tag{36}$$

where $D_{ij} = g_{ik}g_{jk}$ is an $n \times n$ matrix.

We can introduce a *probability current* $\mathbf{J}$ by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{37}$$

and so

$$J_i = f_i \rho - \frac{\partial}{\partial x_j}\left(\frac{D_{ij}}{2}\rho\right) \tag{38}$$

To have a steady state we need $\nabla \cdot \mathbf{J} = 0$, which can be achieved in two ways.

1. $\mathbf{J} = 0$ corresponds to equilibrium solutions, or detailed balance solutions in which each point of any boundary has zero net flux across it.

2. $\mathbf{J} \neq 0$ corresponds to solutions with flux.

First consider $\mathbf{J} = 0$. Then, write $\rho = e^{-\phi} > 0$. The condition $\mathbf{J} = 0$ becomes

$$\frac{D_{ij}}{2}\frac{\partial \phi}{\partial x_j} = -(f_i + v_i), \tag{39}$$

where

$$v_i = -\frac{\partial}{\partial x_j}\left(\frac{D_{ij}}{2}\right), \tag{40}$$

and so provided that D_{ij} is invertible the solution is obtained from

$$\frac{\partial \phi}{\partial x_j} = -2D^{-1}_{ij}(v_i + f_i). \tag{41}$$

Given that the left hand side of this equation is $\nabla\phi$, we have a compatibility condition for the existence of such a solution,

$$\mathrm{curl}(\mathbf{D}^{-1} \cdot (\mathbf{v} + \mathbf{f})) = 0. \tag{42}$$

We now show some examples of density currents satisfying $\nabla \cdot \mathbf{J} = 0$.

2.1 Uniform noise

This is an example of $\nabla \cdot \mathbf{J} = 0$ achieved through $\mathbf{J} = 0$. Let $D_{ij} \propto \delta_{ij}$. Then (41) becomes $\mathbf{f} = \nabla\phi$, which is a *gradient drift* solution, and for compatibility we require $\nabla \times \mathbf{f} = 0$.

2.2 Gradient flow plus Hamiltonian flow

In two dimensions, let $\mathbf{f} = -\sigma^2\nabla\phi + \nabla \times (H\hat{\mathbf{x}}_3)$ and $D_{ij} = 2\sigma^2\delta_{ij}$.

The corresponding deterministic equation would be

$$\dot{x_1} = -\phi_{x_1} - H_{x_2}, \tag{43}$$
$$\dot{x_2} = -\phi_{x_2} + H_{x_1}, \tag{44}$$

i.e. the sum of a gradient flow ϕ and a Hamiltonian flow H.

An example of such a stochastic differential equation would be noisy rotating decay

$$\mathrm{d}U_t = -\gamma U_t \mathrm{d}t + fV_t \mathrm{d}t + \sigma \mathrm{d}W_{t_1}, \tag{45}$$
$$\mathrm{d}V_t = -\gamma V_t \mathrm{d}t - fU_t \mathrm{d}t + \sigma \mathrm{d}W_{t_2}. \tag{46}$$

For this $\mathbf{f}$ and $\mathbf{D}$, try the solution $\rho = e^{-\phi}$ to get

$$\mathbf{J} = e^{-\phi}\mathbf{f} + \sigma e^{-\phi}\nabla\phi. \tag{47}$$

When taking the divergence, many terms cancel, and we are left with

$$\nabla \cdot \mathbf{J} = -e^{-\phi}\nabla\phi \cdot \nabla \times (H\hat{\mathbf{x}}_3), \tag{48}$$

and so we obtain $\nabla \cdot \mathbf{J} = 0$ provided that

$$J(\phi, H) = 0, \tag{49}$$

where $J(\cdot\,,\cdot)$ is the Jacobian.

2.3 Forced harmonic oscillator

The stochastic differential equation

$$\mathrm{d}X_t = Y_t \mathrm{d}t, \tag{50}$$
$$\mathrm{d}Y_t = -X_t \mathrm{d}t - \gamma Y_t \mathrm{d}t + \sigma \mathrm{d}W_t, \tag{51}$$

has stationary solution

$$\rho^s = \mathcal{N} \exp\left(-\frac{\gamma}{\sigma^2}(x^2 + y^2)\right). \tag{52}$$

3 Kolmogorov Backward Equation

We have so far considered the Fokker-Planck equation, which tells us the evolution forwards in time of a probability distribution for a given SDE from a corresponding initial condition. We now derive the Kolmogorov Backward Equation (KBE), which can be thought of as the PDE governing the evolution of a distribution backwards in time, and will subsequently demonstrate the application of this equation to a variety of problems of interest.

3.1 Derivation of the KBE

Starting from the Chapman-Kolmogorov equation (which follows simply from the Markov property for all Markovian processes)

$$\rho(x,t|y,s) = \int_{-\infty}^{\infty} \rho(x,t|x',t')\rho(x',t'|y,s)\,\mathrm{d}x', \tag{53}$$

differentiate with respect to t' to obtain

$$0 = \int_{-\infty}^{\infty} \left[\frac{\partial\rho}{\partial t'}(x,t|x',t')\rho(x',t'|y,s) + \rho(x,t|x',t')\frac{\partial\rho}{\partial t'}(x',t'|y,s)\right]\,\mathrm{d}x'. \tag{54}$$

Now substituting for $\frac{\partial\rho}{\partial t'}(x',t'|y,s)$ using the Fokker-Planck (forward) equation, and using integration by parts

$$\begin{aligned}
0 = \int_{-\infty}^{\infty} \mathrm{d}x' \;& \frac{\partial\rho}{\partial t'}(x,t|x',t')\rho(x',t'|y,s) \\
&+ \rho(x,t|x',t')\frac{\partial}{\partial x'}\left(-f(x') + \frac{1}{2}\frac{\partial}{\partial x'}g(x')^2\right)\rho(x',t'|y,s) \\
0 = \int_{-\infty}^{\infty} \mathrm{d}x' \;& \rho(x',t'|y,s)\left[\frac{\partial\rho}{\partial t'}(x,t|x',t') + f(x')\frac{\partial\rho}{\partial x'}(x,t|x',t') + \frac{1}{2}g(x')^2\frac{\partial^2\rho}{\partial x'^2}(x,t|x',t')\right].
\end{aligned}$$

If we now let the time interval $|t'-s| \to 0$, then $\rho(x',t'|y,s) \to \delta(x'-y)$, so we are left with

$$0 = \frac{\partial\rho}{\partial s}(x,t|y,s) + f(y)\frac{\partial\rho}{\partial y}(x,t|y,s) + \frac{1}{2}g(y)^2\frac{\partial^2\rho}{\partial y^2}(x,t|y,s), \tag{55}$$

or

$$\frac{\partial\rho}{\partial s}(x,t|y,s) = \left(f(y)\frac{\partial}{\partial y} + \frac{1}{2}g(y)^2\frac{\partial^2}{\partial y^2}\right)\rho(x,t|y,s), \tag{56}$$

which is known as the Kolmogorov Backward Equation. Note that the operator $\mathcal{L} \equiv f\partial_x + \frac{g^2}{2}\partial_x^2$ is the formal adjoint of the forward Fokker-Planck operator $\mathcal{L}^\dagger \equiv \partial_x(-f + \partial_x\frac{g^2}{2})$.

3.2 Survival times and first passage times

The power of the KBE becomes transparent if we consider the problem of a random process on some specified domain, and wish to make statements about the time taken for the process to stray outside the domain (variously known as the first passage time, the exit time, the escape time, the stopping time, or the hitting time of the process), or if we wish to determine the region of the boundary through which the process exits the domain.

3.2.1 Survival time

Consider a 1D process $X(t)$ on $x \in (x_a, x_b)$, and impose absorbing boundary conditions $\rho(x_a,t|x_0,t_0) = \rho(x_b,t|x_0,t_0) = 0$. It is of interest to compute the *survival probability* $S(t|x_0,t_0) \equiv \mathbb{P}(x_a < X(u) < x_b \;\forall u < t)$. By definition, S is a monotonically decreasing function of t, with $S(t_0|x_0,t_0) = 1$ for $x_0 \in (x_a,x_b)$ and $S(t|x_0,t_0) \to 0$ as $t \to \infty$, so

probability can be thought of as "leaking" over the edge of the domain as time progresses. Thus probability density is not conserved, and it can be seen that $S(t|x_0, t_0)$ is given by

$$S(t|x_0, t_0) = \int_{x_a}^{x_b} \rho(x, t|x_0, t_0) \, \mathrm{d}x. \tag{57}$$

Consequently we may obtain a PDE for S by integrating the KBE for the process, as follows

$$\int_{x_a}^{x_b} \left[\frac{\partial \rho}{\partial t_0}(x, t|x_0, t_0) = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) \rho(x, t|x_0, t_0) \right] \, \mathrm{d}x \tag{58}$$

$$-\frac{\partial}{\partial t_0} S(t|x_0, t_0) = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) S(t|x_0, t_0), \tag{59}$$

which can be solved for survival time S given boundary conditions $S(t|x_a, t_0) = S(t|x_b, t_0) = 0$ and initial condition $S(t_0|x_0, t_0) = 1$ for $x_a < x_0 < x_b$.

3.2.2 First passage time

For the 1D process above, define random variable t_{exit} as the first time at which $X = x_a$ or $X = x_b$. Then the mean exit time for a process starting at (x_0, t_0) is, by definition

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = \int_{t_0}^{\infty} (t - t_0) p(t|x_0, t_0) \, \mathrm{d}t, \tag{60}$$

where $p(t|x_0, t_0) \equiv -\frac{\mathrm{d}}{\mathrm{d}t} S(t|x_0, t_0)$ is the probability of absorption at time t. Integrating by parts,

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = -\Big[(t - t_0) S(t|x_0, t_0) \Big]_{t_0}^{\infty} + \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t, \tag{61}$$

and it can be seen that the boundary terms vanish provided $S(t|x_0, t_0) \sim o(t^{-1})$ as $t \to \infty$, which holds provided the mean survival time is well-defined, so

$$\mathbb{E}_{x_0}(t_{\text{exit}} - t_0) = \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t. \tag{62}$$

To get an equation for the mean exit time, we then integrate equation (59) between (t_0, ∞) to give

$$-\frac{\partial}{\partial t_0} \int_{t_0}^{\infty} S(t|x_0, t_0) \, \mathrm{d}t - S(t_0|x_0, t_0) = \int_{t_0}^{\infty} \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) S(t|x_0, t_0) \, \mathrm{d}t, \tag{63}$$

using Leibniz's rule. Now noting that the mean exit time is independent of t_0 for an autonomous system, and that $S(t_0|x_0, t_0) = 1$, we have

$$-1 = \left(f(x_0) \frac{\partial}{\partial x_0} + \frac{1}{2} g(x_0)^2 \frac{\partial^2}{\partial x_0^2} \right) \mathbb{E}_{x_0}(t_{\text{exit}} - t_0), \tag{64}$$

which may be solved for mean exit time $\mathbb{E}_{x_0}(t_{\text{exit}} - t_0)$.

3.3 Alternative derivation: Change of variables

An alternative derivation of the KBE to that above is to consider the change of variables $Y = u(X,t)$ in the SDE $\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W$, for some function u. Using Itō calculus, the change of variables becomes

$$\mathrm{d}Y = u_t \mathrm{d}t + u_X \mathrm{d}X + \frac{1}{2} u_{XX} \mathrm{d}X \mathrm{d}X \tag{65}$$

$$= (u_t + f u_X + \frac{g^2}{2} u_{XX})\mathrm{d}t + u_X g \mathrm{d}W, \tag{66}$$

on substituting for $\mathrm{d}X$ using the governing SDE, and noting that $\mathrm{d}W\mathrm{d}W = \mathrm{d}t$. We then have

$$\mathrm{d}Y = (u_t + \mathcal{L}u)\mathrm{d}t + u_X g \mathrm{d}W \tag{67}$$

for operator $\mathcal{L} \equiv f\partial_x + \frac{g^2}{2}\partial_x^2$, as before. The KBE is precisely the equation $u_t + \mathcal{L}u = 0$, and (as a backward heat equation) is well-posed when conditions are specified on some final time $t = T > t_0$. It can be seen from integrating equation (67) that the solution to the homogeneous problem $u_t + \mathcal{L}u = 0$ with condition $u(X,T) = \phi(X)$ generates the expectation $u(X,T) = \mathbb{E}(\phi(X(T))|X(t) = x)$, so for this reason $\mathcal{L}$ is sometimes referred to as the *generator*.

Lecture 5 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Yana Bebieva & Giovanni Fantuzzi

June 19, 2015

1 Stratonovich Interpretation of an SDE

There are different ways of interpreting stochastic differential equations (SDE). We know Itō's interpretation that gives the evolution equation for the transition density

$$\frac{\partial}{\partial t}\rho(x,t|y,s) = \frac{\partial}{\partial x_i}\left(-f_i + \frac{1}{2}\frac{\partial}{\partial x_i}g_{ik}g_{jk}\right)\rho = -\frac{\partial}{\partial x_i}J_i.$$

Here, $\vec{J}$ is a probability current vector field and g_{ij} are functions of x and t (i.e. $g = g(x,t)$, but for simplicity we will suppress the arguments).

The transition density satisfies an evolution equation when we differentiate with respect to the initial time, i.e. the Kolmogorov backward equation

$$-\frac{\partial}{\partial s}\rho(x,t|y,s) = \underbrace{\left(f_i\frac{\partial}{\partial y_i} + \frac{1}{2}g_{ik}g_{jk}\frac{\partial}{\partial y_j}\frac{\partial}{\partial y_i}\right)}_{\text{Generator of the process}}\rho.$$

We are still in a white noise limit and we should still get Markov process as the solution of the SDE (i.e. $X(t)$).

Now, we consider two examples where two different kinds of white noise limits give the same answer. Consider the process (in one dimension for simplicity)

$$\dot{X} = f(X) + g(X)\times \text{ (“approximate white noise”).}$$

White noise has a spectrum that is flat, meaning that correlation function is the delta-function. However if there is a very short time correlation (that we just can not resolve well enough), then this noise can be considered as an approximate white noise (note it is still a real noise). We take a very specific example of approximate white noise as a Gaussian process $g(X)$, that is statistically stationary with a very short correlation time τ. For example, we may consider the Ornstein - Uhlenbeck process $\zeta(t)$, whose SDE is

$$d\zeta(t) = -\frac{1}{\tau}\zeta(t)dt + \frac{1}{\sqrt{\tau}}dW.$$

We pick the amplitude of the noise to be $1/\sqrt{\tau}$, when the relaxation time gets smaller and smaller, the noise influence gets bigger and bigger. The Fokker - Planck equation for $\zeta(t)$ becomes

$$\frac{\partial}{\partial t}\rho(z,t|z_0,t_0) = \frac{\partial}{\partial z}\left(\frac{1}{\tau}z + \frac{1}{2}\frac{1}{\tau}\frac{\partial}{\partial z}\right)\rho = \frac{1}{\tau}\frac{\partial}{\partial z}\left(z + \frac{1}{2}\frac{\partial}{\partial z}\right)\rho,$$

and we see that τ is a time scale for the evolution transition density. The stationary solution of this equation is the Gaussian distribution

$$\rho^{stat}(z) = \frac{1}{\sqrt{\pi}}e^{-z^2} \sim \mathcal{N}(0,1).$$

Moreover, the stationary correlation function is an exponential decay

$$\mathbb{E}(\zeta(t)\zeta(s)) = \frac{1}{2}e^{-\frac{|t-s|}{\tau}},$$

so that the power spectrum of the process is a Lorentzian spectrum

$$S(w) = \frac{\tau}{1+w^2\tau^2}.$$

When τ goes to zero, the spectrum widens but the amplitude decreases to zero. In order to prevent this the amplitude from vanishing, we rescale

$$\zeta(t) \to \frac{1}{\sqrt{\tau}}\zeta(t),$$

so that the spectrum becomes

$$S(w) = \frac{1}{\tau}\frac{\tau}{1+w^2\tau^2}.$$

The amplitude of the spectrum is now 1 when $\tau \to 0$.

The arguments above lead us to conclude that the system of SDEs

$$\frac{dX}{dt} = f(X) + g(X)\frac{1}{\sqrt{\tau}}\zeta(t), \tag{1}$$

$$\frac{d\zeta}{dt} = -\frac{1}{\tau}\zeta + \frac{1}{\sqrt{\tau}}\xi(t), \tag{2}$$

in the limit of short correlation time τ behaves similarly to an SDE for X where the noise is approximately white (note the functions f and g could have explicit time dependence). The combination (1) and (2) is a vector-valued Markov process, and the Fokker - Planck equation for its transitions density is

$$\frac{\partial}{\partial t}\rho(x,z,t|x_0,z_0,t_0) = \left[\frac{\partial}{\partial x}\left(-f - \frac{1}{\sqrt{\tau}}zg\right) + \frac{1}{\tau}\frac{\partial}{\partial z}\left(z + \frac{1}{2}\frac{\partial}{\partial z}\right)\right]\rho \tag{3}$$

We would like to deduce from this an evolution equation for the distribution of X alone, i.e. the marginal (reduced) distribution

$$r(x,t|x_0,t_0) = \int dz\left(\int \rho(x,z,t|x_0,z_0,t_0)\rho^{stat}(z_0)dz_0\right),$$

in the limit $\tau \to 0$.

To this end, let $\epsilon = \sqrt{\tau}$ and rearrange equation (3) by grouping terms of order ϵ^n. We obtain

$$0 = \left(\frac{1}{\epsilon^2} F_0 + \frac{1}{\epsilon} F_1 + F_2 \right) \rho, \tag{4}$$

where

$$\begin{aligned} F_0 &= \frac{\partial}{\partial z}(z + \frac{1}{2}\frac{\partial}{\partial z}) \quad \text{– the Ornstein - Uhlenbeck operator,} \\ F_1 &= -z\frac{\partial}{\partial x} g, \\ F_2 &= -(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} f). \end{aligned}$$

Our ansatz is

$$\rho(x, z, t | x_0, z_0, t_0) = \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \dots,$$

i.e. the subscript of each ρ_i in the expansion indicates the corresponding power of ϵ. Now we plug this expression for ρ into (4) and group the terms according to the order of ϵ to obtain

$$\mathcal{O}(\epsilon^{-2}): \qquad 0 = F_0\rho_0 \tag{5a}$$

$$\mathcal{O}\left(\epsilon^{-1}\right): \qquad 0 = F_0\rho_1 + F_1\rho_0 \tag{5b}$$

$$\mathcal{O}(\epsilon^0): \qquad 0 = F_0\rho_2 + F_1\rho_1 + F_2\rho_0 \tag{5c}$$

We keep in mind that we want to derive from this an evolution equation for the reduced distribution

$$r(x, t | x_0, t_0) = \int dz \left(\int \rho(x, z, t | x_0, z_0, t_0) \rho^{stat}(z_0) dz_0 \right) = r_0 + \epsilon r_1 + \epsilon^2 r_2 + \dots \tag{6}$$

We know all properties of the operator F_0; in particular, we can solve the eigenvalue problem

$$F_0 p_n(z) = -n p_n(z),$$

to find the eigenfunctions

$$p_n(z) = H_n(z) p^{stat}(z) \equiv H_n(z) p_0(z),$$

where the Hermite polynomials H_n are defined as

$$\begin{aligned} H_n(z) &= (-1)^n e^{z^2} \frac{d^n}{dz^n} e^{-z^2}, \\ H_0(z) &= 1, \\ H_1(z) &= 2z, \\ H_2(z) &= 2(2z^2 - 1). \end{aligned}$$

We may compute all Hermite polynomials using the recursion relation

$$zH_n(z) = \frac{1}{2}H_{n+1}(z) + nH_{n-1}(z).$$

Now we are ready to solve (5a) to find

$$0 = F_0\rho_0 \Rightarrow \rho_0(x,z,t) = p_0(z)r_0(x,t)$$

where $p_0(z)$ can be multiplied by any function of x and t (it plays the role of the amplitude of the eigenfunction p_0 at each point in space and time). In this case this function is $r_0(x,t)$ because if we integrate $\rho = \rho_0 + \epsilon\rho_1 + ...$ over z to obtain r, the leading term is exactly r_0.

To order ϵ^{-1}, equation (5b) now reads

$$F_0\rho_1 = -F_1\rho_0 = zp_0(z)\frac{\partial}{\partial x}gr_0(x,t) = \frac{1}{2}p_1(z)\frac{\partial}{\partial x}gr_0(x,t)$$

where we used the recursion relation to replace $zp_0 = p_1/2$. Here, we deal with a linear inhomogeneous differential equation and the general solution of this equation is a particular solution plus a general solution of the homogeneous equation. Therefore

$$\rho_1 = \underbrace{-\frac{1}{2}\frac{\partial}{\partial x}gr_0(x,t)p_1(z)}_{\text{particular solution}} + \underbrace{r_1(x,t)p_0(z)}_{\text{general solution}}.$$

Finally, upon substitution of ρ_0 and ρ_1, equation (5c) becomes

$$F_0\rho_2 = z\frac{\partial}{\partial x}g\left(-\frac{1}{2}\frac{\partial}{\partial x}gr_0(x,t)p_1(z) + r_1(x,t)p_0(z)\right) + \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}f\right)r_0(x,t)p_0(z).$$

Let us now use $zp_1 = 2z^2p_0 = \frac{1}{2}p_2 + p_0$ and $zp_0 = \frac{1}{2}p_1$, to show that

$$\begin{aligned} F_0\rho_2 &= -\left(\frac{1}{4}p_2(z) + \frac{1}{2}p_0(z)\right)\frac{\partial}{\partial x}g\frac{\partial}{\partial x}gr_0(x,t) + \frac{1}{2}p_1(z)\frac{\partial}{\partial x}gr_1(x,t) + (\frac{\partial}{\partial t} + \frac{\partial}{\partial x}f)r_0(x,t)p_0(z) \\ &= p_0(z)\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial x}f - \frac{1}{2}\frac{\partial}{\partial x}g\frac{\partial}{\partial x}g\right]r_0(x,t) + p_1(z)\left[\frac{1}{2}\frac{\partial}{\partial x}gr_1(x,t)\right] + p_2(z)\left[-\frac{1}{4}\frac{\partial}{\partial x}g\frac{\partial}{\partial x}gr_0(x,t)\right]. \end{aligned} \tag{7}$$

Again, we need to add a particular solution and the solution of the homogeneous equation, i.e. $\rho_2 = \rho_{part} + r_2(x,t)p_0$. In order for a particular solution for this equation to exist the right hand side should be orthogonal to null space of the operator F_0. We know that $F_0p_1(z) = -p_1(z)$ and $F_0p_2(z) = -2p_2(z)$, but we can not invert F_0 on $p_0(z)$. Thus, in order to solve (7) the coefficient of $p_0(z)$ has to vanish and this gives a condition on $r_0(x,t)$

$$\frac{\partial}{\partial t}r_0(x,t|x_0,t_0) = \frac{\partial}{\partial x}\left(-f + \frac{1}{2}g\frac{\partial}{\partial x}g\right)r_0(x,t)$$

If we find $r_0(x,t)$ that satisfies this, we can get an explicit equation for $r_1(x,t)$, $r_2(x,t)$ etc. So if start with a real noise that is very fast ($\tau \to 0$, or $\epsilon \to 0$ in (6)), equation (1) becomes an SDE with an approximate white noise, $r(x,t) \to r_0(x,t)$ as $r_0(x,t)$ is the leading order term and we conclude that $\rho(x,t|x_0,t_0)$ of the process X satisfies the Fokker - Planck equation (a.k.a. Forward Kolmogorov Equation)

$$\underbrace{\frac{\partial}{\partial t}\rho = \frac{\partial}{\partial x}\left(-f + \frac{1}{2}g\frac{\partial}{\partial x}g\right)\rho}_{\text{Stratonovich Fokker - Planck}}. \tag{8}$$

We remark that Itō's interpretation of the SDE yielded the Fokker - Planck equation

$$\underbrace{\frac{\partial}{\partial t}\rho = \frac{\partial}{\partial x}\left(-f + \frac{1}{2}\frac{\partial}{\partial x}g^2\right)\rho}_{\text{Itō's the Fokker - Planck}}. \tag{9}$$

Thus if g is not a function of x (an additive noise), the transition density satisfying (8) is the same as that obtained from Itō's interpretation.

In the case of multiplicative noise, i.e. $g = g(x,t)$, can rewrite equation (8) as

$$\frac{\partial}{\partial t}\rho = \left(-f - \frac{1}{2}gg' + \frac{\partial}{\partial x}g^2\right)\rho \tag{10}$$

and then this equation is in the Itō form but with the modified drift. So if we let $f \mapsto f + \frac{1}{2}gg'$ in the SDE (9), then equation (10) describes the evolution of X according to the Itō's SDE.

2 Interpretation of an SDE: Itō vs Stratonovich

As remarked at the end of the previous section, in Itō's calculus the solution $X_t = X(t)$ of the SDE

$$dX_t = f(X_t,t)\,dt + g(X_t,t)\,dW_t \tag{11}$$

is the Markov process X_t whose transition density $\rho(x,t|x_0,t_0)$ satisfies the Itō's Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(-f + \frac{1}{2}\frac{\partial}{\partial x}g^2\right)\rho \tag{12}$$

(note the convention that an operator acts on all terms to its right). As previously explained, this followed from interpreting (11) as the continuous-time limit of the discrete-time process $X(t+\Delta t) - X(t) = f[X(t),t]\,\Delta t + g[X(t),t]\,\Delta W$, where $\Delta W = W(t+\Delta t) - W(t)$ is the jump of a Wiener process over the time interval Δt.

In the Stratonovich interpretation of (11), instead, the stochastic forcing in the SDE is obtained as the white-noise limit of a coloured stochastic process, such as the Ornstein-Uhlenbeck process. In this case, it is customary to write the SDE as

$$dX_t = f(X_t,t)\,dt + g(X_t,t) \circ dW_t, \tag{13}$$

where the " $\circ$ " sign indicates that dW_t should be interpreted as the white-noise limit of a coloured noise. The difference with Itō's interpretation is that the solution to (13) is the

Markov process X_t whose transition density $\rho(x,t|x_0,t_0)$ satisfies the Stratonovich Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(-f + \frac{1}{2} g \frac{\partial}{\partial x} g\right)\rho. \tag{14}$$

Clearly, the differential operators in equations (12) and (14) differ unless $g(x,t)$ is a constant, and the evolution of the PDF ρ depends on which interpretation of the SDE is chosen. However, an Itō SDE can easily be reformulated as a Stratonovich SDE (and vice versa). In fact, the solution of (11) is the same (in the sense of its statistics) as that of the Stratonovich SDE

$$dX_t = \left[f - \frac{1}{2} g \frac{\partial g}{\partial x}\right] dt + g \circ dW_t, \tag{15}$$

since the corresponding Fokker-Planck equation, computed from (14) using the modified drift $f - \frac{1}{2} g \frac{\partial g}{\partial x}$, can be rearranged to obtain (12). Similarly, one can see that the solution of (13) is the same (again, in the sense of its statistics) as the solution of the Itō SDE

$$dX_t = f\, dt + \frac{1}{2} g \frac{\partial g}{\partial x} + g\, dW_t. \tag{16}$$

This is particularly convenient since many physical systems are modelled by a coloured noise with very fast dynamics, which corresponds to a Stratonovich interpretation of the SDE; however, Itō's formulation is easier to implement numerically to simulate the system (for example, via the simple Euler-Maruyama or Milstein discretisation schemes).

Finally, we note that the Stratonovich interpretation maintains the standard rules of calculus for the differential of the random variable $Y_t = F(X_t)$, i.e.

$$dY_t = dF(X_t) = F'(X_t) \circ dX_t. \tag{17}$$

The symbol "$\circ$" indicates that X_t obeys a Stratonovich SDE. Equation (17) can be shown for one-dimensional processes if we assume that F is invertible, with inverse G. Then, the transition density $\rho_Y(y,t|y_0,t_0)$ of Y_t is related to $\rho_X(x,t|x_0,t_0)$ by

$$\rho_Y(y,t|y_0,t_0) = \rho_X[G(y),t|G(y_0),t_0)]G'(y), \tag{18}$$

Moreover, noticing that $G'(\cdot)F'(\cdot) = 1$ and $\frac{\partial}{\partial y} = G' \frac{\partial}{\partial x}$ one obtains

$$\begin{aligned}
\frac{\partial}{\partial t}\rho_Y &= G' \frac{\partial}{\partial t}\rho_X \\
&= \frac{\partial}{\partial x}\left(-f + \frac{1}{2} g \frac{\partial}{\partial x} g\right)\rho_Y \\
&= \frac{\partial}{\partial y}\left(-F'f + \frac{1}{2} F' g \frac{\partial}{\partial x} g \frac{F'}{F'}\right)\rho_Y \\
&= \frac{\partial}{\partial y}\left(-F'f + \frac{1}{2} F' g \frac{\partial}{\partial y} g F'\right)\rho_Y.
\end{aligned} \tag{19}$$

The claimed result follows from the Fokker-Planck equation for a Stratonovich SDE, which implies $dY_t = F'(f\, dt + g \circ dW_t) = F'(X_t) \circ dX_t$.

Example 1. (Stratonovich Logistic Equation) Consider the logistic equation

$$dX_t = (\bar{\mu} X_t - X_t^2)dt + \sigma\, X_t \circ dW_t$$

where $X_t \geq 0$ is the size of a population at time t and $\bar{\mu}$ and σ are given constants. The corresponding Fokker-Planck equation is

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}\left(x^2 - \bar{\mu}x - \frac{\sigma^2}{2}x + \frac{\sigma^2}{2}\frac{\partial}{\partial x}x^2\right)\rho,$$

and the corresponding stationary distribution can be calculated as

$$\rho^{\text{stat}}(x) = N\, x^{\left(\frac{2\bar{\mu}}{\sigma^2}-1\right)}\, \exp\left(-\frac{2\,x}{\sigma^2}\right),$$

$$N = \left(\frac{\sigma^2}{2}\right)^{-\frac{2\bar{\mu}}{\sigma^2}} \left[\Gamma\left(\frac{2\bar{\mu}}{\sigma^2}\right)\right]^{-1}.$$

Moreover, the exact solution of the Stratonovich Logistic equation can be found using an appropriate change of variables. Dividing (1) by $X_t^2 dt$ and applying the chain rule (17), we have

$$-\frac{d}{dt}\left(\frac{1}{X_t}\right) = \frac{\bar{\mu}}{X_t} + \frac{\sigma}{X_t}\frac{dW_t}{dt} - 1.$$

Letting $Y_t = (X_t)^{-1}$, we obtain the differential equation

$$\frac{dY_t}{dt} + [\bar{\mu} + \sigma\xi(t)]\, Y_t = 1,$$

which can be solved to find

$$Y_t = Y_0\, e^{-\bar{\mu}t - \sigma W(t)} + e^{-\bar{\mu}t - \sigma W(t)} \int_0^t e^{\bar{\mu}s + \sigma W(s)} ds$$

$$\therefore \quad X_t = \frac{X_0\, e^{\bar{\mu}t + \sigma W(t)}}{1 + \int_0^t e^{\bar{\mu}s + \sigma W(s)} ds}.$$

Sample realisation of the SDE are shown in Figure 1, while Figure 2 illustrates the stationary transition density functions. For any noise amplitude σ, the process is driven by the exponential growth $e^{\bar{\mu}t}$, until saturation. The qualitative difference with Itō's interpretation of the same SDE (the solution of which can be found by substituting $\bar{\mu} \mapsto \bar{\mu} - \frac{1}{2}\sigma^2$, cf. Lecture 3) is remarkable: in the Itō's interpretation a stationary probability distribution ceases to exist when $\sigma^2 > 2\bar{\mu}$, i.e. when the exponential term $e^{\left(\bar{\mu} - \frac{1}{2}\sigma^2\right)t}$ decays in time. In terms of the population dynamics described by the SDE, this means that Itō's formulation predicts extinction (at least in the infinite-time limit) when $\sigma^2 > 2\bar{\mu}$, while the individuals are always alive for any σ according to the Stratonovich solution (see the case $\sigma = 2$ in Figure 1).

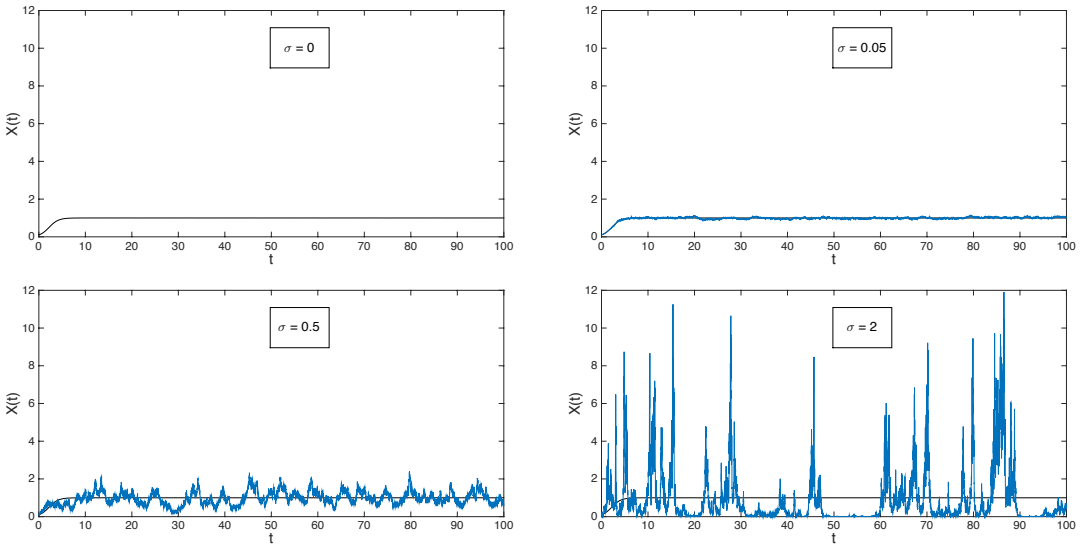

Figure 1: Sample realisation of the stochastic logistic equation, compared to the deterministic version (no noise) for $\bar{\mu} = 1$, $x_0 = 0.1$ and increasing noise amplitude σ. Clockwise (starting top-left): $\sigma = 0$, $\sigma = 0.05$, $\sigma = 0.5$, $\sigma = 2$.

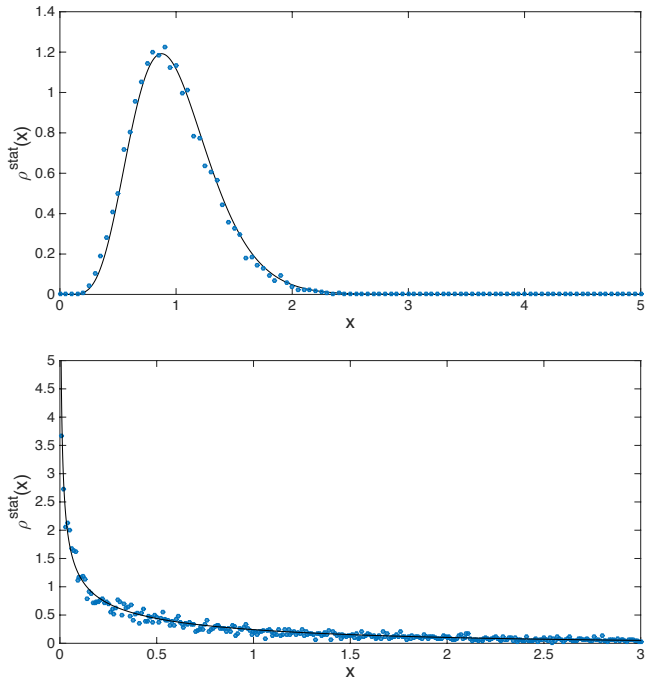

Figure 2: Comparison between the analytical transition density $\rho^{stat}(x)$ and the density function computed over 10^4 realisations of the SDE for $\sigma = 0.5$ (top) and $\sigma = 2$ (bottom).

3 Stratonovich SDEs for Vector-Valued Processes

The results presented in the previous section can be generalised to vector-valued processes. If $\mathbf{X}(t) = (X_1(t), X_2(t), ...)$ satisfies the set of SDEs

$$dX_i(t) = f_i(\mathbf{X}(t), t)dt + g_{ij}(\mathbf{X}(t), t) \circ dW_j(t), \quad i = 1, 2, ..., N, \quad j = 1, 2, ..., M, \tag{20}$$

then its transition density $\rho(\mathbf{x}, t|\mathbf{x}_0, t_0)$ can be computed with the Fokker-Planck equation

$$\frac{\partial}{\partial t}\rho = \frac{\partial}{\partial x_i}\left(-f_i(\mathbf{x}, t) + \frac{1}{2}g_{ik}(\mathbf{x}, t)\frac{\partial}{\partial x_j}g_{jk}(\mathbf{x}, t)\right)\rho. \tag{21}$$

Finally, we can translate this into the Itō's formulation (and vice versa) by modifying the drift in the same way as for the one-dimensional case.

4 White-noise Limit of a Dichotomous Markov Process

A symmetric dichotomous (a.k.a. two-step) Markov process $I(t)$ ($t \geq 0$) can take two values A and $-A$ (see Figure 3), the transition between one state and the other taking place at a constant rate α. The time intervals between state transitions are thus exponentially distributed and the probabilities $p_+(t) = \mathbb{P}[I(t) = A]$ and $p_-(t) = \mathbb{P}[I(t) = -A]$ evolve according to the master equation

$$\frac{d}{dt}\begin{pmatrix} p_+(t) \\ p_-(t) \end{pmatrix} = \begin{pmatrix} -\alpha & \alpha \\ \alpha & -\alpha \end{pmatrix}\begin{pmatrix} p_+(t) \\ p_-(t) \end{pmatrix}. \tag{22}$$

This equation can be solved to find

$$\begin{pmatrix} p_+(t) \\ p_-(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 + e^{-2\alpha t} & 1 - e^{-2\alpha t} \\ 1 - e^{-2\alpha t} & 1 + e^{-2\alpha t} \end{pmatrix}\begin{pmatrix} p_+(0) \\ p_-(0) \end{pmatrix}, \tag{23}$$

where $p_+(0)$ and $p_-(0)$ are the probabilities that the process starts at A or $-A$ respectively, with $p_+(0) + p_-(0) = 1$. The stationary distributions are immediately found as

$$\begin{pmatrix} p_+^{\text{stat}} \\ p_-^{\text{stat}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \tag{24}$$

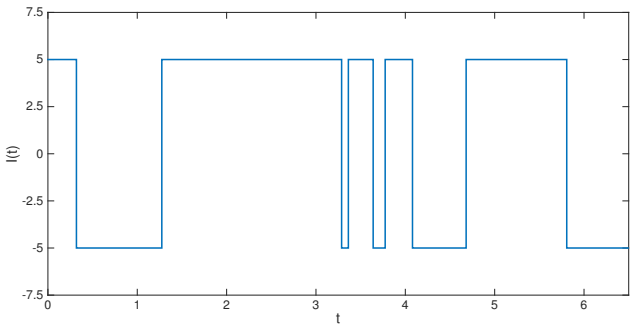

Figure 3: Sample realisation of $I(t)$ for $\alpha = 1$, $A = 5$.

and the conditional distributions $\mathbb{P}[I(t) = A|I(0) = \pm A]$, $\mathbb{P}[I(t) = -A|I(0) = \pm A]$ are obtained by setting $p_+(0)$ and $p_-(0)$ to 1 and 0 in turn. Specifically, one has

$$\begin{aligned}\mathbb{P}[I(t) = A|I(0) = \pm A] &= \frac{1}{2}\left(1 \pm e^{-2\alpha t}\right), \\ \mathbb{P}[I(t) = -A|I(0) = \pm A] &= \frac{1}{2}\left(1 \mp e^{-2\alpha t}\right).\end{aligned} \tag{25}$$

The conditional expectations can be used to compute the correlation

$$\begin{aligned}\mathbb{E}[I(t)I(0)] &= \sum_{n,m\in\{A,-A\}} n\,m\,\mathbb{P}[I(t) = n; I(0) = m] \\ &= \sum_{n,m\in\{A,-A\}} n\,m\,\mathbb{P}[I(t) = n|I(0) = m]\,\mathbb{P}[I(0) = m] \\ &= A^2\,e^{-2\alpha t},\end{aligned} \tag{26}$$

for $t \geq 0$, which can be generalised to $\mathbb{E}[I(t)I(s)] = A^2\,e^{-2\alpha|t-s|}$ for any two time instants t and s. This means that the dichotomous process has the Lorenzian power spectrum

$$S(\omega) = \int_{-\infty}^{+\infty} A^2\,e^{-2\alpha|t|}e^{-i\omega t}dt = \frac{A^2}{\alpha}\frac{4\alpha^2}{4\alpha^2+\omega^2} \quad . \tag{27}$$

Note that, as shown in Figure 4, $S(\omega)$ tends to a white spectrum in the limit $\alpha \to \infty$ if $A = \sqrt{\alpha}$.

Let us now consider a stochastic process X_t evolving according to

$$dX_t = f(X_t, t)dt + g(X_t, t)I(t). \tag{28}$$

The transition density of this process can be computed as $\rho(x, t|x_0, t_0) = \rho_+(x, t|x_0, t_0) + \rho_-(x, t|x_0, t_0)$, where the transition densities ρ_+ and ρ_- correspond to the mutually exclusive cases $I(t) = A$ and $I(t) = -A$ and satisfy the master equation

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho_+ \\ \rho_-\end{pmatrix} = \begin{pmatrix}-\alpha - \frac{\partial}{\partial x}(f + Ag) & \alpha \\ \alpha & -\alpha - \frac{\partial}{\partial x}(f - Ag)\end{pmatrix}\begin{pmatrix}\rho_+ \\ \rho_-\end{pmatrix}. \tag{29}$$

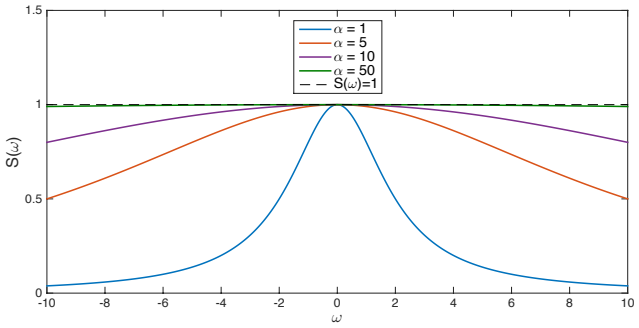

Figure 4: Correlation spectrum of $I(t)$ for $A = \sqrt{\alpha}$ and increasing values α.

Letting $\alpha = A^2$ in order to obtain the correct scaling of the noise spectrum, the equations for $\rho = \rho_+ + \rho_-$ and for $q = \rho_+ - \rho_-$ then read

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ q\end{pmatrix} = \begin{pmatrix}-\frac{\partial}{\partial x}\left(f\rho + Agq\right)\\ -2A^2 q - \frac{\partial}{\partial x}\left(fq + Ag\rho\right)\end{pmatrix}. \tag{30}$$

When the noise amplitude A is very large, we argue that ρ and q can be expanded as

$$\begin{aligned}\rho(x,t) &= \rho_0(x,t) + \frac{1}{A}\rho_1(x,t) + \text{higher order terms}\\ q(x,t) &= q_0(x,t) + \frac{1}{A}q_1(x,t) + \text{higher order terms}\end{aligned} \tag{31}$$

so that the evolution of the random variable X_t in the white-noise limit is described by ρ_0. Substituting into (30) and collecting terms of the same order yields

$$O(A^2): \quad 2q_0 = 0, \tag{32a}$$

$$O(A): \quad \begin{cases} 2q_1 + \dfrac{\partial}{\partial x}(g\,\rho_0) = 0,\\ \dfrac{\partial}{\partial x}(g\,q_0) = 0,\end{cases} \tag{32b}$$

$$O(1): \quad \begin{cases} \dfrac{\partial q_0}{\partial t} + 2q_2 + \dfrac{\partial}{\partial x}(f q_0 + g\,\rho_1) = 0,\\ \dfrac{\partial q_0}{\partial t} + \dfrac{\partial}{\partial x}(f\rho_0 + g\,q_1) = 0,\end{cases} \tag{32c}$$

from which the following Fokker-Planck equation can be derived for ρ_0:

$$\frac{\partial \rho_0}{\partial t} = \frac{\partial}{\partial x}\left(-f + \frac{1}{2}g\frac{\partial}{\partial x}g\right)\rho_0. \tag{33}$$

This equation is the same as the Stratonovich Fokker-Planck equation. We conclude that the solution to the SDE

$$\frac{dX_t}{dt} = f(X_t,t) + g(X_t,t)\circ\xi(t), \tag{34}$$

where $\xi(t)$ is the white-noise limit of a dichotomous Markov process (a.k.a. dichotomous noise or DMN), is the process X_t whose transition density satisfies the Stratonovich Fokker-Planck equation.

Lecture 6 - Climate Variability and its Null Hypothesis

Henk A. Dijkstra; notes by Yana Bebieva & Giovanni Fantuzzi

June 22, 2015

1 Introduction

The Earth's system is very non-linear and complicated. For instance, if a periodic forcing, as an input signal, is coming into the Earth's system (e.g. variability in a solar constant), then very complicated signal is coming out and we want to understand what process can explain this transformation. The ice core oxygen isotope records, that is usually linked to the local temperature, (Figure 1) shows various oscillation and variation in temperature of $5°C$. The other example of a noisy signal is El Niño/Southern Oscillation. Temperature anomaly with respect to the mean in the equatorial Pacific is shown in Figure 2.

Now, we are going to analyze these two cases by choosing the special time and spatial scale of phenomenon. For example, for El Niño it would be Pacific Basin and few years (inter-annual) correspondingly. And we try to understand that phenomenon using deterministic type of model and all unresolved processes we consider as a noise. In this sense stochastic dynamical system is obtained (Figure 3. We start with the simple example where underlying dynamical system is linear.

2 The Null Hypothesis

Consider stochastic climate model developed by Hasselmann in 1976 (Figure 4). It is a layer of the ocean (mixed layer) that has the heat flux coming from the atmosphere (Q_{oa}) and we are interested how mixed layer temperature (T) evolves in time. Write down the heat balance

$$\rho C_p \frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial z^2}$$

where C_p is the specific heat of seawater, ρ is density, λ is mixing coefficient, t is time, and z is vertical coordinate. Boundary conditions are

$$z = 0 : \lambda \frac{\partial T}{\partial z} = Q_{oa}$$
$$z = -h : \frac{\partial T}{\partial z} = 0,$$

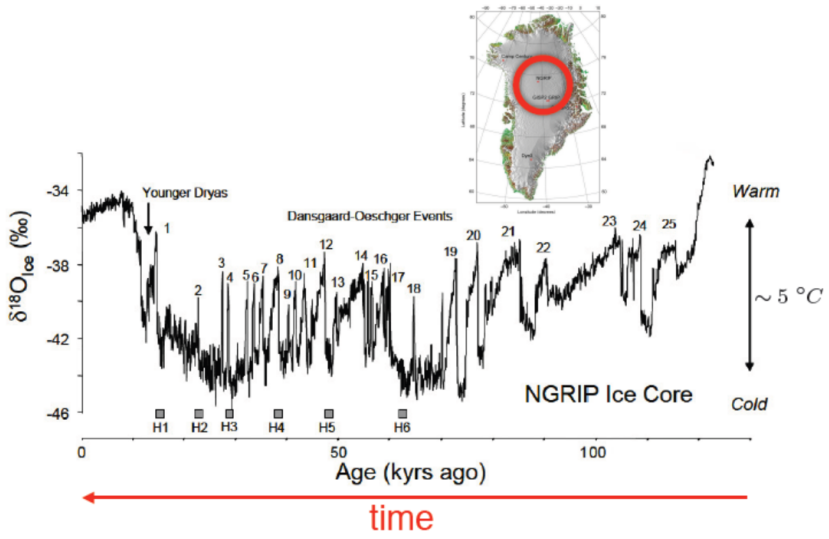

Figure 1: Ice core oxygen isotope record (ratio of ^{18}O to ^{16}O) taken from the NGRIP ice core at Greenland.

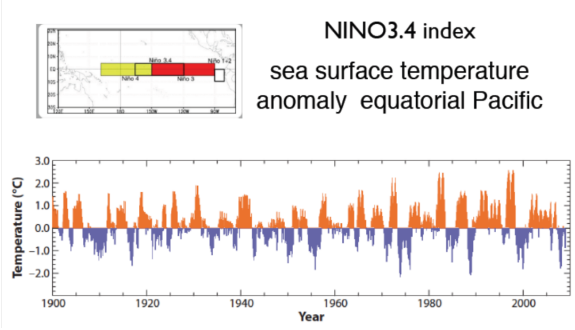

Figure 2: Temperature anomaly with respect to the mean in the equatorial Pacific in the Nino 3.4 region.

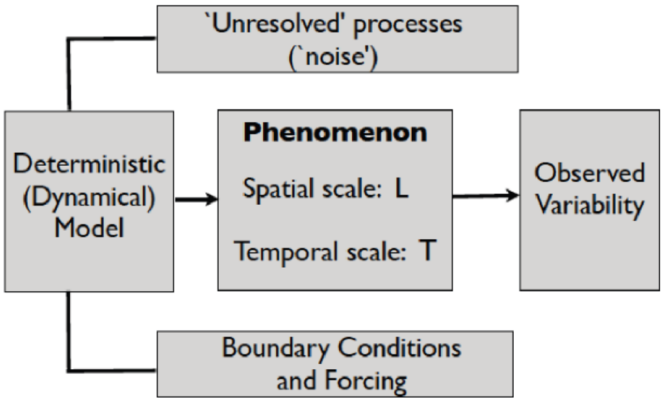

Figure 3: Stochastic dynamical systems approach scheme.

where h is depth of the layer. Now, we define depth averaged temperature

$$\overline{T} = \frac{1}{h}\int_{-h}^{0} T dz$$

This implies

$$\rho C_p h \frac{\partial \overline{T}}{\partial t} = Q_{oa} = \alpha\left(T_a - \overline{T}\right),$$

where T_a is atmospheric temperature.

We eventually get the equation

$$\begin{aligned}
\frac{\partial \overline{T}}{\partial t} &= \frac{\alpha}{\rho C_p h}\left(T_a - \overline{T}\right) \\
\gamma &= \frac{\alpha}{\rho C_p h} \\
\frac{\partial \overline{T}}{\partial t} &= -\gamma \overline{T} + \gamma T_a.
\end{aligned}$$

The time scale $1/\gamma$ is roughly 100 days. The question is how to represent γT_a. We look at a Fast-Slow systems with fast variable x and slow variable y.

$$\begin{aligned}
\frac{dx}{dt} &= f(x,y) \\
\frac{dy}{dt} &= g(x,y).
\end{aligned}$$

Looking at data measurements, one finds that variations of atmospheric temperature are much faster than ocean temperature fluctuations. Let y_0 be an initial condition, $\Delta y = y - y_0$, $t \ll \tau_y$ and take ensemble average. Assuming ergodicity

$$\begin{aligned}
\langle \frac{\partial y}{\partial t} \rangle = \frac{d}{dt}\langle y \rangle = \langle g(x,y) \rangle &= \frac{d\Delta y}{dt} \\
\langle \Delta y \rangle &= \langle g(x,y) \rangle t \\
\tilde{y} &= \Delta y - \langle \Delta y \rangle \\
\frac{d\tilde{y}}{dt} &= g(x,y) - \langle g(x,y) \rangle = \tilde{g}(x,y) \\
\langle \tilde{g}(x,y) \rangle &= 0.
\end{aligned}$$

If x is stationary, so is $\tilde{g}$. Then, $\tilde{g}$ looks like white noise and the model is

$$\frac{d\tilde{y}}{dt} = \sigma \xi(t).$$

In total we have

$$\frac{d\overline{T}}{dt} = -\gamma \overline{T} + \sigma \xi(t),$$

or

$$d\overline{T}_t = -\gamma \overline{T}_t dt + \sigma dW_t.$$

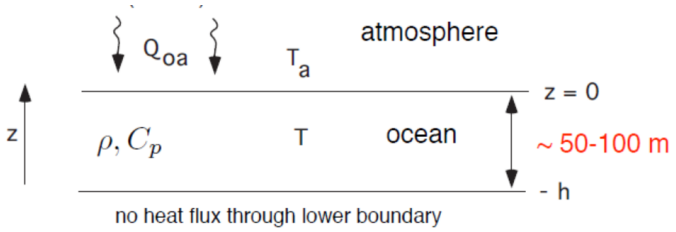

Figure 4: Scheme of Hasselmann's stochastic climate model. See text for definition of all parameters.

3 Analysis of the SDE

Let $X_t = \overline{T}_t$ be a stochastic process that satisfies the equation

$$dX_t = -\gamma X_t dt + \sigma dW_t$$
$$X_t = X_0 - \gamma \int_0^t X_s ds + \int_0^t \sigma dW_s.$$

How do we define a stochastic integral $\int_0^t \sigma dW_s$? The Itô stochastic integral definition is

$$\int_0^T h(t) dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j) \left(W(t_{j+1}) - W(t_j)\right)$$

and the Stratonovich stochastic integral is

$$\int_0^T h(t) \circ dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h\left(\frac{t_j + t_{j+1}}{2}\right) \left(W(t_{j+1}) - W(t_j)\right).$$

Where the limits exist in the L_2 norm (mean square) sense. So find 1) $\int_0^\tau dW_t = W_\tau$ (apply definition and the fact that $W_0 = 0$) and 2) $\int_0^\tau W_t dW_t$. Start with Taylor expansion using $(dW_t)^2 = dt$

$$\begin{aligned} h(W_t + dW_t) - h(W_t) &= h'(W_t)dW_t + \frac{1}{2}h''(W_t)(dW_t)^2 + ... \\ &= h'(W_t)dW_t + \frac{1}{2}h''(W_t)dt + ... \end{aligned}$$

Integrate both sides

$$h(W_T) - h(W_0) = \int_0^T h'(W_t)dW_t + \int_0^T \frac{1}{2}h''(W_t)dt.$$

We obtained the Itô's lemma.

Now, choosing $h(W_t) = W_t^2$ (consequently $h' = 2W_t$ and $h'' = 2$) we find using the Itô's lemma

$$W_T^2 - W_0^2 = \int_0^T 2W_t dW_t + \frac{1}{2}\int_0^T 2dt.$$

Again using that $W_0 = 0$ we obtain

$$\int_0^T W_t dW_t = \frac{1}{2}\left(W_T^2 - T\right).$$

Thus, any linear 1-D stochastic differential equation can be solved analytically via the Itô lemma.

$$dX_t = A_t dt + B_t dW_t$$
$$f = f(x,t), f_1(x,t) = \frac{\partial f}{\partial t}, f_2(x,t) = \frac{\partial f}{\partial x}, f_{22}(x,t) = \frac{\partial^2 f}{\partial x^2}.$$

Now, Taylor expand

$$f(t+dt, x_t + dX_t) - f(t, X_t) = f_1 dt + f_2 dX_t + \frac{1}{2}(f_{11}(dt)^2 + 2f_{12}dtdX_t + f_{22}(dX_t)^2) + ...$$

Then plug SDE into this and find

$$f(t, X_t) - f(0, X_0) = \int_0^t \left(f_1 + f_2 A_s + \frac{1}{2} f_{22} B_s^2\right) ds + \int_0^t f_2 B_s dW_s.$$

For our SDE model, choose

$$\begin{aligned} f(x,t) &= e^{\gamma t} x \\ f_1(x,t) &= \gamma e^{\gamma t} x \\ f_2(x,t) &= e^{\gamma t} \\ f_{22}(x,t) &= 0. \end{aligned}$$

Using this with our stochastic ODE and applying the Itô lemma

$$e^{\gamma t} X_t - X_0 = \int_0^t \left(\gamma e^{\gamma s} X_s - \gamma e^{\gamma s} X_s\right) ds + \sigma \int_0^t e^{\gamma s} dW_s.$$

And we have solved the SDE exactly as

$$X_t = e^{-\gamma t}\left[X_0 + \int_0^t \sigma e^{\gamma s} dW_s\right].$$

4 Numerical Solution of SDEs

One often requires a numerical solution of the SDE (interpreted in Itō's sense)

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t \tag{1}$$

over the time interval $[0, T]$. A basic approach to this problem is to consider discrete time instants $t_n = n\Delta t$, for a given time-step Δt, and approximate (1) with the *Euler-Maruyama scheme*

$$\tilde{X}_{n+1} = \tilde{X}_n + f(t_n, \tilde{X}_n)\Delta t + g(t_n, \tilde{X}_n)\Delta W_{n+1}. \tag{2}$$

Here, we have denoted $\tilde{X}_n = \tilde{X}(t_n)$ the approximate solution of (1) and $\Delta W_{n+1} = W_{n+1} - W_n$ is the jump of a Wiener process over the time interval Δt. The jumps ΔW are independent, Gaussian random variables $\Delta W \sim \mathcal{N}(0, \Delta t)$ and can therefore be generated at each iteration with an appropriate random number generator (e.g. `DW = sqrt(Dt)*randn(1)` in MATLAB). We also remark that Itō's interpretation of an SDE consists of the continuous-time limit of the Euler-Maruyama scheme.

Of course, more sophisticated schemes can be derived. An example is the *Milstein scheme*, which approximates the solution $X(t)$ of (1) as

$$\tilde{X}_{n+1} = \tilde{X}_n + f(t_n, \tilde{X}_n)\Delta t + g(t_n, \tilde{X}_n)\Delta W_{n+1} + \frac{1}{2}\left[(\Delta W_{n+1})^2 - \Delta t\right]. \tag{3}$$

The main difference between the various numerical schemes regards their convergence properties in the limit $\Delta t \to 0$. In the context of numerical schemes for SDEs, there are two notions of convergence. The first, known as *weak convergence*, considers convergence of the expectations of the approximation error $|X(t_n) - \tilde{X}_n|$; that is, a numerical scheme is weakly convergent if

$$\mathbb{E}\left(|X(t_n) - \tilde{X}_n|\right) \leq c\Delta t^{\eta}. \tag{4}$$

Here, c and η are constants (dependent on the type of scheme and the SDE considered); η is the convergence rate of the scheme.

Similarly, we say that a numerical scheme is *strongly convergent* if the expectations $\mathbb{E}[X(t_n)]$ and $\mathbb{E}(\tilde{X}_n)$ converge, i.e.

$$|\mathbb{E}[X(t_n)] - \mathbb{E}(\tilde{X}_n)| \leq c\Delta t^{\eta}. \tag{5}$$

Example 1 (**Ornstein-Uhlenbeck Process**)**.** Consider the Ornstein-Uhlenbeck process $dX_t = -\gamma X_t dt + \sigma dW_t$, discretised with the Euler-Maruyama (EM) scheme as

$$X_{n+1} = X_n - \gamma X_n \Delta t + \sigma \Delta W \quad = (1 - \gamma\Delta t)X_n + \sigma\Delta W. \tag{6}$$

We know that the stationary state solution $X^{\text{stat}}(t)$ is a Gaussian random variable with zero mean ($\mathbb{E}[X(t)] = 0$). In order to check the strong convergence of the EM scheme, we can take the expectation of (6), obtaining

$$\mathbb{E}(X_{n+1}) = (1 - \gamma\Delta t)\mathbb{E}(X_n) = (1 - \gamma\Delta t)^{n+1} X_0 \tag{7}$$

As $n \to \infty$, the expectation $\mathbb{E}(X_{n+1})$ converges to the analytic stationary result $\mathbb{E}[X(t)] = 0$ only if

$$|1 - \gamma\Delta t| \leq 1 \Rightarrow \Delta t \leq \frac{2}{\gamma}, \tag{8}$$

that is the EM scheme converges strongly if $\Delta t \leq \frac{2}{\gamma}$.

5 Applicability of the Hasselmann's Model to the SST Anomaly

Hasselmann's linear model for the SST anomaly, derived in Section 2, resulted in the SDE

$$\frac{d\bar{T}}{dt} = -\gamma\bar{T} + \sigma\xi(t) \tag{9}$$

which is the Ornstein-Uhlenbeck process. Thus, according to Hasselmann's model, the transition density ρ of $\bar{T}$ satisfies the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \frac{\partial(\gamma x \rho)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial x^2}, \tag{10}$$

and there exists a stationary distribution

$$\rho^{\text{stat}}(x) = \sqrt{\frac{\gamma}{\pi\sigma^2}}\, e^{-\frac{\gamma x^2}{\sigma^2}}. \tag{11}$$

Moreover, we recall that the stationary autocorrelation and the spectrum of the Ornstein-Uhlenbeck solution are, respectively,

$$\mathbb{E}(X_s X_t) = \frac{\sigma^2}{\gamma}\, e^{-|t-s|}, \tag{12}$$

$$S(\omega) = \frac{\sigma^2}{\gamma^2 + \omega^2}. \tag{13}$$

The main question to be answered at this point is how this model can be tuned using experimental observations of the SST anomaly (say, $\{\bar{T}_0^{\text{ex}}, \bar{T}_1^{\text{ex}}, \ldots, \bar{T}_N^{\text{ex}}\}$), so that any predictions based on Hasselmann's model can be trusted. In fact, we would like the analytic solution of the continuous time Ornstein-Uhlenbeck process, i.e.

$$\bar{T}(t) = e^{-\gamma t}\bar{T}_0 + \sigma e^{-\gamma t}\int_0^t e^{\gamma s} dW_s, \tag{14}$$

to reproduce the statistics of the time-series of measurements $\bar{T}_0^{\text{ex}}, \bar{T}_1^{\text{ex}}, \ldots, \bar{T}_N^{\text{ex}}$. To this purpose, we discretise (14) by looking at the time instants t_n and t_{n+1}, i.e. we consider

$$\bar{T}_{n+1} = e^{-\gamma t_{n+1}}\bar{T}_0 + \sigma e^{-\gamma t_{n+1}}\int_0^{t_{n+1}} e^{\gamma s} dW_s, \tag{15a}$$

$$\bar{T}_n = e^{-\gamma t_n}\bar{T}_0 + \sigma e^{-\gamma t_n}\int_0^{t_n} e^{\gamma s} dW_s. \tag{15b}$$

We can multiply (15b) by $e^{-\gamma(t_{n+1}-t_n)} = e^{-\gamma\Delta t}$ and subtract it from (15a) to obtain

$$\bar{T}_{n+1} = e^{-\gamma\Delta t}\bar{T}_n + \sigma e^{-\gamma t_{n+1}}\int_{t_n}^{t_{n+1}} e^{\gamma s} dW_s. \tag{16}$$

When Δt is small, we can approximate the integral in the last expression by $c\Delta W_{n+1}$, where c is a suitable constant. So, the last term can be approximated with a Gaussian random variable Z_n with zero mean and variance $\tilde{\sigma}^2$ (chosen appropriately). Then, we may approximate the solution of the Ornstein-Uhlenbeck process by the discrete stochastic process

$$\bar{T}_{n+1} = \alpha\bar{T}_n + Z_{n+1}, \tag{17}$$

where $\alpha = e^{-\gamma\Delta t} \in (0,1)$. This type of discrete process is known as the *AR(1) process* (where AR stands for auto-regressive), or as a *red noise process*. We can now compute

the spectrum of the AR(1) process and compare it to the spectrum obtained from the experimental data set $\{\bar{T}_0^{\rm ex}, \bar{T}_1^{\rm ex}, \ldots, \bar{T}_N^{\rm ex}\}$ to compute appropriate values of α and $\tilde{\sigma}$. To this purpose, we first note that

$$\begin{aligned} \bar{T}_k &= \alpha \bar{T}_{k-1} + Z_k \\ &= \alpha(\alpha \bar{T}_{k-2} + Z_{k-1}) + Z_k \\ &\vdots \\ &= \alpha^k Z_0 + \alpha^{k-1} Z_1 + \ldots + \alpha Z_{k-1} + Z_k \end{aligned} \tag{18}$$

so that

$$\begin{aligned} c_0 \stackrel{\rm def}{=} \mathbb{E}(\bar{T}_k^2) &= \left[1 + \alpha^2 + \left(\alpha^2\right)^2 + \ldots + \left(\alpha^2\right)^k\right] \tilde{\sigma}^2 \\ &\stackrel{k\to\infty}{=} \frac{\tilde{\sigma}^2}{1-\alpha^2} \end{aligned} \tag{19}$$

Note that we have used the relation $\mathbb{E}(Z_n Z_m) = \tilde{\sigma}^2 \delta_{nm}$ and the fact that the geometric series converges since $0 < \alpha < 1$. Thus, recalling that the random variable Z_n is independent of any past realisation $\bar{T}_m$, the correlation of the discrete process becomes

$$\begin{aligned} c_k \stackrel{\rm def}{=} \mathbb{E}(\bar{T}_i \bar{T}_{k+i}) &= \frac{1}{N} \sum_{i=0}^{N} \bar{T}_i \bar{T}_{k+i} \\ &= \frac{1}{N} \sum_{i=0}^{N} \bar{T}_i \left(\alpha \bar{T}_{k-1+i} + Z_{k+i}\right) \\ &= \frac{\alpha}{N} \sum_{i=0}^{N} \bar{T}_i \bar{T}_{k-1+i} \\ &= \alpha\, c_{k-1} \\ &= \alpha^k\, c_0 \end{aligned} \tag{20}$$

so as $k \to \infty$ (i.e. the discrete time series becomes infinitely long) we obtain

$$c_k = \frac{\alpha^k\, \tilde{\sigma}^2}{1-\alpha^2} = \frac{\tilde{\sigma}^2}{1-\alpha^2}\, e^{-\gamma t_k}, \tag{21}$$

where we have used $\alpha = e^{-\gamma \Delta t}$ (constant for a given Δt) and that, for uniform time-steps, $k\Delta t = t_k$. Taking the continuous-time version $c(t)$ of (21), we can compute the spectrum of the AR(1) process as

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} c(t) e^{-i\omega t} dt \\ &= \int_{-\infty}^{\infty} \frac{\tilde{\sigma}^2}{1-\alpha^2}\, e^{-\gamma|t|} e^{-i\omega t} dt \\ &= \frac{2\gamma}{1-\alpha^2} \frac{\tilde{\sigma}^2}{\gamma^2 + \omega^2} \end{aligned} \tag{22}$$

The absolute value was introduced to maintain the negativity of the argument of the exponential term in $c(t)$. Note that, up to a normalisation constant, this spectrum is the same as for the continuous-time Ornstein-Uhlenbeck process. Thus, one can tune the AR(1) to fit the measured data in the following way: compute the spectrum of the discrete measurements $\{\bar{T}_0^{\text{ex}}, \bar{T}_1^{\text{ex}}, \ldots, \bar{T}_N^{\text{ex}}\}$, then fit appropriate values of α and $\tilde{\sigma}$ in (22). Then, the corresponding AR(1) process reproduces the measured statistics, and can be used to estimate the statistical properties of the SST anomaly according to Hasselman's model.

Lecture 7: The El Niño Phenomenon

Henk A. Dijkstra; notes by Tom Eaves & Anna FitzMaurice

June 23, 2015

1 Phenomenology of El Niño Southern Oscillation

In its mean state, there is strong zonal asymmetry in the equatorial Pacific. The climatological mean easterly trade winds pile up warm water in the western Pacific, whilst cool water is upwelled in the east due to Ekman divergence there, as depicted in Figure 1. Corresponding to this zonal temperature gradient is a sea surface height (SSH) gradient, with high SSH in the west and low SSH in the east. The SSH gradient is compensated at depth by a depressed thermocline in the west, and a shallower thermocline in the east. We have good measurements of the equatorial Pacific from the TAO/TRITON array of approximately 70 moored ocean buoys, which have been monitoring surface and subsurface temperatures, wind speed and direction, and precipitation since 1994.

Every four to seven years, the western Pacific warm pool spreads eastward in a phenomenon known as *El Niño*. This sea surface temperature (SST) anomaly is associated with the eastward propagation of a subsurface temperature anomaly, which in turn is related to a shoaling of the western Pacific thermocline and a depression of eastern Pacific thermocline. These temperature anomalies are coincident with a weakening of the Walker circulation (the zonal circulation cell over the equatorial Pacific), and a shift of the region of highest precipitation to follow the maximum SSTs. The opposite of the El Niño state (i.e. a heightened "normal" state) is referred to as La Niña conditions, and the largescale temperature, pressure, and precipitation, anomalies associated with transitions between El Niño and La Niña states are together known as El Niño Southern Oscillation (ENSO). The El Niño and La Niña phases are shown schematically in Figure 2.

The index used to describe ENSO is an average of the SST anomaly over a region of the equatorial Pacific. Figure 3 shows a timeseries of the ENSO index for the NINO3 region of the eastern Pacific, which spans the region 150W to 90W, between 5S and 5N. A power series decomposition of this timeseries reveals a spectral peak centered about a period of 3 years.

2 The Zebiak & Cane model

To capture the oscillatory behavior of ENSO dynamically, a coupled atmosphere-ocean model is required, which admits feedbacks between perturbations to the equatorial easterlies, the thermocline depth, and equatorial SSTs, and as such allows the spontaneous growth of anomalies. We shall also see that oceanic wave dynamics are important to the development and decay of El Niños, and so necessary in a minimal model of ENSO.

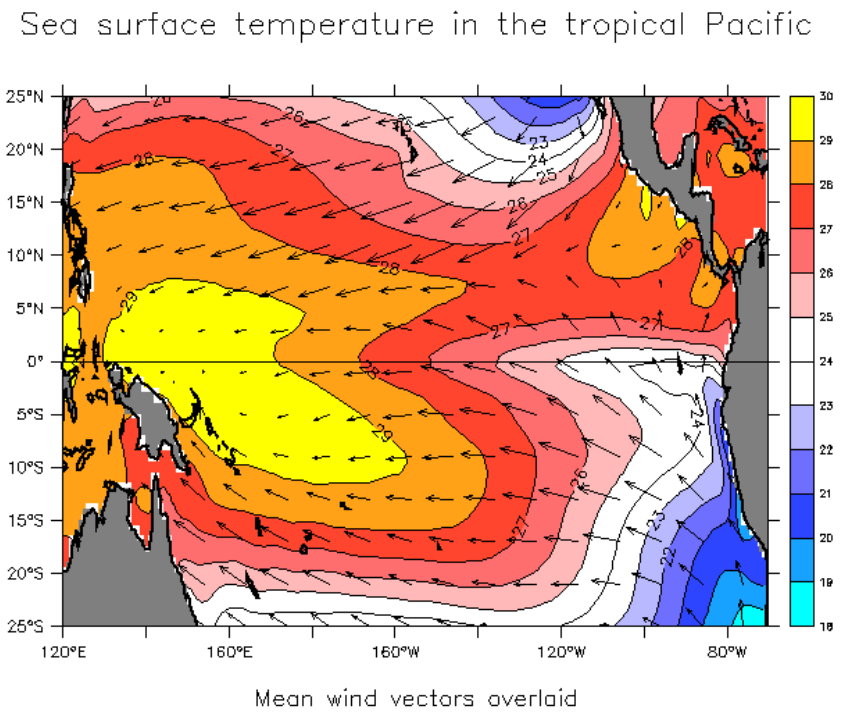

Figure 1: The climatological mean SST and wind stress in the tropical Pacific. [Reproduced from faculty.washington.edu/kessler.]

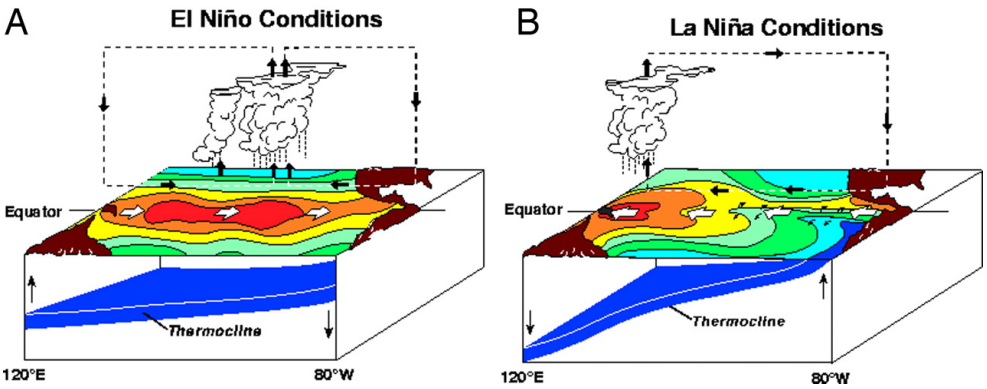

Figure 2: A schematic of (A) El Niño and (B) La Niña states of the tropical Pacfic. [Reproduced from www.noaa.gov.]

2.1 Model formulation

Zebiak & Cane (1987), hereon ZC, consider a $1\frac{1}{2}$-layer reduced gravity ocean (depicted in Figure 4) below a constant-depth mixed layer of temperature T, which feels a temperature-

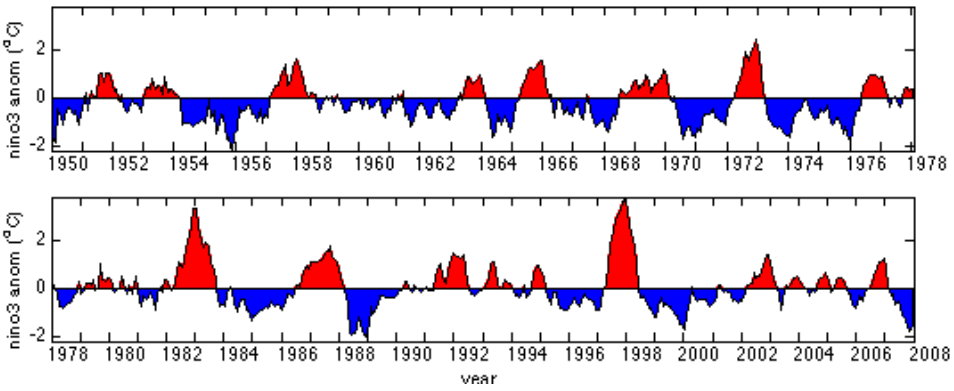

Figure 3: A timeseries of the NINO3 SST anomaly, with El Niño events colored in red and La Niña events colored in blue. [Reproduced from http://www.seas.harvard.edu/climate.]

dependent wind stress

$$\tau^x = \tau^x_{\text{ext}} + \mu\mathcal{A}(T - T_0), \tag{1}$$

for some atmospheric operator $\mathcal{A}$, and coupling parameter μ (with $\mu = 0$ corresponding to the entirely uncoupled case, and $\mu = 1$ describing "normal" coupling).

The reduced gravity ocean model equations for the horizontal velocities (u, v) and depth h are

$$\frac{\partial u_1}{\partial t} - \beta_0 y v_1 = -g' \frac{\partial h}{\partial x} + \frac{\tau^x}{\rho h}, \tag{2}$$

$$\beta_0 y u_1 = -g' \frac{\partial h}{\partial y}, \tag{3}$$

$$\frac{\partial h}{\partial t} = -H\left(\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y}\right), \tag{4}$$

$$x = 0 : \int_{-\infty}^{\infty} u_1(y)dy = 0, \tag{5}$$

$$x = L : u_1 = 0, \tag{6}$$

$$y \to \pm\infty : u_1,\ v_1,\ h, \text{ bounded}, \tag{7}$$

in the upper layer, and zero velocities in the lower layer.

The evolution of the mixed layer temperature T is governed by an advection-diffusion equation with relaxation back to some atmospheric temperature T_0, and relaxation to a specified subsurface temperature profile $T_s(h)$ in the presence of upwelling $w > 0$, as follows

$$\frac{\partial T}{\partial t} + u_1\frac{\partial T}{\partial x} + v_1\frac{\partial T}{\partial y} + w_1\mathcal{H}(w_1)\frac{T - T_s(h)}{H} + \alpha_T(T - T_0) - \kappa_H\nabla^2 T = 0, \tag{8}$$

where $\mathcal{H}$ is the Heaviside function, along with boundary conditions

$$x = 0,\ L : \frac{\partial T}{\partial x} = 0, \tag{9}$$

$$y \to \pm\infty : T \text{ bounded}. \tag{10}$$

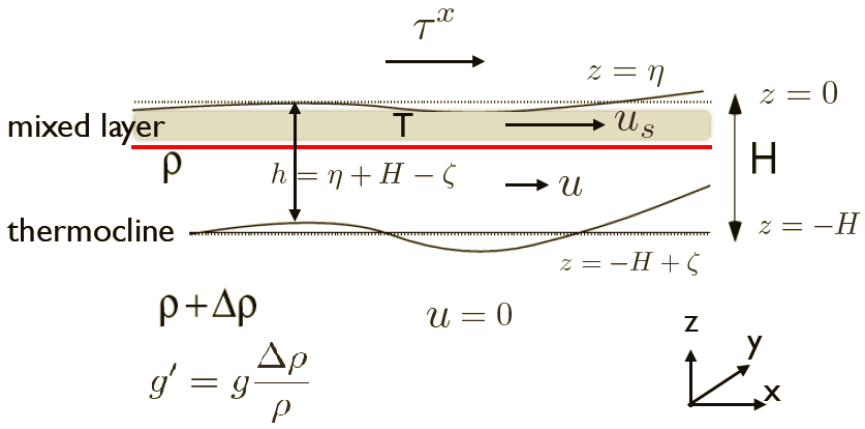

Figure 4: The ocean component of the ZC model. A thermocline of density ρ overlies a denser stationary layer of density $\rho + \Delta\rho$. The ocean feels atmospheric wind stress and temperature through constant depth mixed layer of temperature T.

2.2 Wave dynamics in the ZC ocean model

Consider free waves with $\tau_x = 0$, corresponding to no wind stress input, and no coupling with the mixed layer temperature field T. Let $u = \hat{u}(y)e^{i(kx-\sigma t)}$ and define v and h similarly. Then,

$$-i\sigma\hat{u} - \beta_0 y\hat{v} = -ikg'\hat{h}, \tag{11}$$

$$\beta_0 y\hat{u} = -g'\hat{h}', \tag{12}$$

$$-i\sigma\hat{h} + H(ik\hat{u} + \hat{v}') = 0. \tag{13}$$

Look first for solutions with $\hat{v} = 0$. Then,

$$-\sigma\hat{u} = -g'k\hat{h}, \tag{14}$$

$$-\sigma\hat{h} + Hk\hat{u} = 0, \tag{15}$$

$$\tag{16}$$

which has a non-zero solution only if

$$\sigma^2 = k^2 g' H, \tag{17}$$

and so

$$\frac{\sigma}{k} = \pm\sqrt{g'H} \equiv \pm c_0. \tag{18}$$

We can also solve for $\hat{h}$ and $\hat{u}$, since

$$\beta_0 y\hat{u} = \frac{\beta_0 y g' k\hat{h}}{\sigma} = -g'\hat{h}', \tag{19}$$

and so

$$\hat{h}(y) = \hat{h}(0)\exp\left[-\frac{\beta_0 k}{2\sigma}y^2\right] \equiv \hat{h}(0)\exp\left[-\frac{1}{2}\left(\frac{y}{\lambda_0}\right)^2\right], \tag{20}$$

where $\lambda_0 = \sqrt{c_0/\beta_0}$ is the Rossby deformation radius. Note that for $\hat{h}(y)$ to be bounded as $y \to \pm\infty$ we have set $\sigma/k = +c_0$.

We can solve for the general case $\hat{v} \neq 0$ by using the Hermite polynomials H_n, in which case

$$\hat{u}_j(y) = \frac{1}{2\sqrt{2}} \left(\frac{\psi_{j+1}(y)}{\sqrt{j+1}} - \frac{\psi_{j-1}(y)}{\sqrt{j}} \right), \tag{21}$$

$$\hat{h}_j(y) = \frac{1}{2\sqrt{2}} \left(\frac{\psi_{j+1}(y)}{\sqrt{j+1}} + \frac{\psi_{j-1}(y)}{\sqrt{j}} \right), \tag{22}$$

$$\hat{v}_j(y) = \psi_j(y), \tag{23}$$

where

$$\psi_j(y) = \frac{H_j(y/\lambda_0)\exp[-(y/\lambda_0)^2/2]}{(2^j j! \pi^{1/2})^{1/2}}, \tag{24}$$

and the corresponding phase speeds are

$$c_j = -\frac{c_0}{2j+1}. \tag{25}$$

2.3 Possible feedbacks in the ZC model

To explore the feedbacks possible in the full model, consider a perturbation (denoted by hats) to some steady state (denoted by overbars) of the simplified temperature equation

$$\frac{\partial T}{\partial t} = -w\frac{T - T_s(h)}{H}. \tag{26}$$

Linearizing about the steady state, this becomes

$$\frac{\partial \hat{T}}{\partial t} = -\hat{w}\frac{\bar{T} - T_s(\bar{h})}{H} - \bar{w}\frac{\hat{T} - T_s'(\bar{h})\hat{h}}{H}. \tag{27}$$

Now it can be seen that if there is a warm anomaly in the mixed layer (i.e. $\hat{T} > 0$) giving rise to a deepening of the thermocline (i.e. $\hat{h} > 0$), the second term on the right of equation 27 will be positive, leading to more warming. This positive feedback is known as the *thermocline feedback*. Similarly, from the first term on the right of equation 27, it can be deduced that a positive temperature anomaly, associated with a reduction in upwelling ($\hat{w} < 0$), will likewise enhance the positive temperature anomaly, acting as a positive feedback. This is called the *upwelling feedback*.

An analogous treatment of the zonal advection terms of the temperature equation

$$\frac{\partial T}{\partial t} = -u\frac{\partial T}{\partial x} \tag{28}$$

yields linearized equation

$$\frac{\partial \hat{T}}{\partial t} = -\bar{u}\frac{\partial \hat{T}}{\partial x} - \hat{u}\frac{\partial \bar{T}}{\partial x}. \tag{29}$$

Now a positive temperature anomaly will produce a zonal velocity anomaly $\hat{u}$ that acts to enhance this anomaly by the advection of the mean temperature field $\bar{T}$. This is the *zonal advection feedback*.

2.4 ZC model equilibria, stability, and bifurcations

After discretisation, we can reduce the ZC model to the general form

$$\frac{\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mu). \tag{30}$$

Then, equilibria $\mathbf{x} = \mathbf{x}^*$ are found by solving

$$\mathbf{f}(\mathbf{x}^*, \mu) = 0. \tag{31}$$

Suppose that we know one solution $\mathbf{x}^*(\mu = 0)$. As μ is varied, we expect $\mathbf{x}^*(\mu)$ to vary continuously (at least for a while, generically). Such a branch of solutions traces out a curve in the $\mathbf{x}$-μ plane. A convenient way to solve such a system is to utilise this continuous changing of the solutions by viewing the arclength s along this curve as a parameter and solve instead

$$\mathbf{f}(\mathbf{x}^*(s), \mu(s)) = 0, \tag{32}$$

$$\mathcal{N}\left(\frac{\mathbf{x}^*}{ds}, \frac{\mu}{ds}\right) = 0, \tag{33}$$

where $\mathcal{N}$ is a normalisation constraint on the arclength.

Given an equilibria $\mathbf{x}^*$, we are interested in its stability. For this purpose, write $\mathbf{x} = \mathbf{x}^* + \mathbf{x}'$. Then, linearising in $\mathbf{x}'$, the equation becomes

$$\frac{\mathbf{x}'}{dt} = J\mathbf{x}', \tag{34}$$

where $J_{ij} = \partial f_j / \partial x_j(\mathbf{x}^*(\mu))$ is the Jacobian of the dynamical system evaluated at $\mathbf{x} = \mathbf{x}^*(\mu)$.

We can determine the stability of a given solution $\mathbf{x}^*$ by writing $\mathbf{x}' = \hat{\mathbf{x}} e^{\sigma t}$, and then σ satisfies the eigenvalue problem

$$J\hat{\mathbf{x}} = \sigma\hat{\mathbf{x}}. \tag{35}$$

We say that $\mathbf{x}^*$ is a stable equilibria if $\mathbb{R}(\sigma) < 0$ and an unstable equilibria if $\mathbb{R}(\sigma) > 0$.

As μ is varied, there are a number of possibilities for the behaviour of such equilibria. In general there will be particular values of $\mu = \mu^*$, called *bifurcation* points, at which a given solution changes its stability properties, or ceases to exist altogether, as μ is varied through its bifurcation value μ^*. There a four canonical types of bifurcations found in dynamical systems, namely

- **Saddle–node bifurcation:** A saddle–node bifurcation is the bifurcation in which there are no equilibria for $\mu < \mu^*$ and a pair of equilibria with opposite stability properties for $\mu > \mu^*$. Through a change of variables, any one-dimensional system with a saddle–node bifurcation can be mapped, sufficiently close to its bifurcation point, onto the form

  $$\dot{x} = \mu - x^2. \tag{36}$$

  We see that for $\mu < 0$ there are no equilibria, whilst for $\mu > 0$ there are two equilibria $x^* = \pm\sqrt{\mu}$. The Jacobian is just $-2x$, and so the solution $x^* = \sqrt{\mu}$ is stable, whilst the solution $x^* = -\sqrt{\mu}$ is unstable.

- **Pitchfork bifurcation:** A pitchfork bifurcation is the bifurcation in which there is a single equilibria (which may or may not be stable) for $\mu \lessgtr \mu^*$ (the case $<$ is called supercritical, and the case $>$ is called subcritical), and three equilibria for $\mu \gtrless \mu^*$ in which the original equilibria point swaps its stability, and the two new equilibria have the stability of the original solution. The normal form for a pitchfork bifurcation is

$$\dot{x} = \mu x - x^3 \quad \text{(supercritical)}, \tag{37}$$

$$\dot{x} = \mu x + x^3 \quad \text{(subcritical)}. \tag{38}$$

  In the first case, the solution $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$, and the solutions $x = \pm\sqrt{\mu}$ only exist for $\mu > 0$, and are stable there. In the second case, the solution $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$, and the solutions $x = \pm\sqrt{-\mu}$ only exist for $\mu < 0$ and are unstable there. The transformation $t \mapsto -t$ reverses the stability of all the solutions, but the systems still retain their original labels 'supercritical' and 'subcritical'.

- **Transcritical bifurcation:** A transcritical bifurcation is the bifurcation in which there are two solutions for each of $\mu \lessgtr \mu^*$, but at $\mu = \mu^*$ they coincide, and swap their stability. The normal form for a transcritical bifurcation is

$$\dot{x} = \mu x - x^2. \tag{39}$$

  We see that the equilibria are always $x = 0$ or μ. However, the Jacobian is $\mu - 2x$, and so the solution $x = 0$ is stable for $\mu < 0$ and unstable for $\mu > 0$ whereas the solution $x = \mu$ is unstable for $\mu < 0$ and stable for $\mu > 0$.

- **Hopf bifurcation:** A Hopf bifurcation is the bifurcation in which there is one equilibria for $\mu \lessgtr \mu^*$ (which we again call super/subcritical), and a single equilibria for $\mu \gtrless \mu^*$ with the opposite stability, and a periodic orbit that coincides with the equilibria when $\mu = \mu^*$. The normal form for a Hopf bifurcation is

$$\dot{x} = \mu x - \omega y - x(x^2 + y^2), \tag{40}$$

$$\dot{y} = \mu y + \omega x - y(x^2 + y^2). \tag{41}$$

  It is clear that $(x, y) = (0, 0)$ is an equilibria for this system. The Jacobian at $(0, 0)$ is

$$J(0,0) = \begin{pmatrix} \mu & -\omega \\ \omega & \mu \end{pmatrix}, \tag{42}$$

  which has eigenvalues $\sigma = \mu \pm i\omega$, and so the equilibria $(x, y) = (0, 0)$ is stable for $\mu < 0$ and unstable for $\mu > 0$.

  To demonstrate that this system has a periodic orbit for $\mu > 0$ it is convenient to use polar co-ordinates $(x, y) = r(\cos\theta, \sin\theta)$, in which case

$$\dot{r} = \mu r - r^3, \tag{43}$$

$$\dot{\theta} = \omega, \tag{44}$$

and so provided $\omega \neq 0$, there is a stable periodic orbit solution $r = \sqrt{\mu}$ and $\theta = \omega t + \theta_0$ when $\mu > 0$. Note that in the degenerate case $\omega = 0$, we obtain a circle of fixed points that are marginally stable in the angular direction, and stable in the radial direction. Since the equation for r is nearly pitchfork-like (we don't allow $r < 0$), changing the signs of terms in the above equations yields super/subcritical Hopf bifurcations with the same convention as for the pitchfork case.

A Hopf bifurcation can be viewed physically as feedbacks amplifying to give an oscillatory signal.

There are a whole range of other bifurcations which are typically degenerate cases in which the first order nonlinearity occurs at even higher orders in x. We typically don't see these, and indeed typically don't see pitchfork or transcritical bifurcations. The reason for this is that, for example, if we have made some error in modeling whatever physical system we are interested in, then perhaps each of the equations for these bifurcations should have an extra constant ϵ added to the right hand side. Then we see that the saddle–node bifurcation remains a saddle–node bifurcation, the Hopf bifurcation remains a Hopf bifurcation, but the pitchfork bifurcation separates into an isolated non-bifurcating solution and a saddle–node bifurcation, and the transcritical bifurcation separates into a pair of saddle–node bifurcations.

For higher dimensional dynamical systems, we may invoke the centre manifold theorem (which is quite technical) to see that near a bifurcation point, the dynamics of the dynamical system collapses onto a low dimensional (often one or two) manifold on which the reduced dynamics generically take the form of one of the bifurcations discussed above.

2.5 Hopf bifurcation in the ZC model

As the ocean–atmosphere coupling parameter μ in the ZC model is increased from zero, there has been shown to be a Hopf bifurcation when $\mu \approx 0.525$, and the period of the resulting periodic orbit is approximately 4 years. This observation has been used as a first order explanation of El Niño. Additionally, it is known that μ scales with the square of ocean basin size, and so, for example, given that the size of the Atlantic ocean basin is approximately one third of the size of the Pacific ocean basin, we have $\mu_{\text{Atlantic}} \approx \mu_{\text{Pacific}}/9$, and so the lack of an El Niño event in the Atlantic could be explained by the fact that $\mu_{\text{Atlantic}} \lesssim 0.525 \lesssim \mu_{\text{Pacific}}$, and so there does not exist a periodic orbit solution of the ZC model.

For the ZC model, the Hopf bifurcation corresponds to an amplifying feedback of geometrically confined Rossby and Kelvin basin modes with SST modes.

3 Physical Mechanisms for ENSO

We have seen from the ZC model that with idealized ocean-atmosphere coupling and oceanic wave dynamics it is possible to find oscillatory solutions in certain parts of parameter space that resemble ENSO in amplitude and period. Below we heuristically describe two mechanisms that might give rise to such oscillatory behavior.

3.1 Wave oscillator

Consider a positive temperature anomaly at the equator in the Pacific ocean, which corresponds to a positive SSH anomaly on-equator, with compensatory negative SSH anomalies off-equator to the north and south. We have seen that such a signal may propagate eastwards as an equatorial Kelvin wave on the equator, which may be interpreted as the eastward propagating and growth of an El Niño. Meanwhile, the off-equator signal will propagate westward as a Rossby wave and, on reaching the westerward basin boundary, may be reflected as an equatorial Kelvin wave. This reflected wave signal has the possibility of interfering with and killing the original positive temperature anomaly, ending the El Niño. Whilst this delayed oscillator mechanism of El Niño undoubtedly influences ENSO dynamics, a consideration of the timescales involved (from the Kelvin and Rossby wave speeds) does not explain the observed ENSO period of four to seven years.

3.2 Recharge oscillator

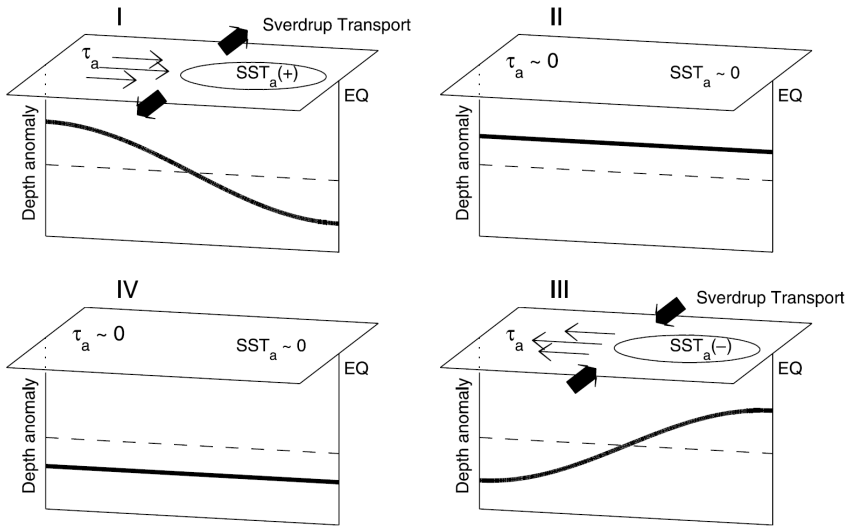

Figure 5: A schematic of the stages of the recharge oscillator mechanism for ENSO.

An alternative mechanism that produces longer timescale variability comes from considering the overall basin adjustment. A positive SST anomaly in the eastern Pacific will produce a westerly wind stress anomaly. The wind stress acts to change the thermocline slope, piling up water and so depressing the thermocline in the east, whilst shoaling the thermocline in the west. Such a perturbation to the thermocline slope will enhance the SST perturbation, acting as a positive feedback. As the positive temperature anomaly strengthens, there is a divergent transport of heat off-equator by the ocean, which shoals the thermocline, suppressing the SST anomaly and so reducing the westerly wind anomaly. The shoaling of the thermocline eventually carries the system into the opposite phase, with a negative SST anomaly in the east, and so an easterly wind stress anomaly. This causes the convergent transport of heat to the equator, resulting in the "recharge" of the ocean

heat content there. As such, this mechanism is known as the *recharge oscillator* view of ENSO. This process is shown schematically in Figure 5.

Lecture 8: The El Niño Phenomenon (continued)

Henk A. Dijkstra; notes by Florence Marcotte & Cesar Rocha

June 24, 2015

1 Deterministic Chaos

So far we have seen that the oscillatory behavior of El Nino Southern Oscillation (ENSO) mechanism was related to the saturation of a unstable mode above the threshold of a Hopf bifurcation, corresponding to a given critical coupling strength between the atmosphere and the ocean. An interesting point in the record of the ENSO signal is the signature of some mean seasonal cycle. For the western tropical Pacific ocean, negative anomalies in the records of zonal winds occur around April, whereas positive anomalies occur around December. Furthermore, sea surface temperature (SST) anomalies are observed at the same periods: positive SST anomalies are associated with negative zonal wind anomalies, and negative SST anomalies are associated with positive zonal wind anomalies. It is important to note that ENSO events and the seasonal cycle are sensitive to the same environmental factors such as wind forcing and the ocean circulation. ENSO's non-linear interaction with the seasonal cycle is characterized by a tendency to synchrony in periodic, subharmonic oscillation. At this point we saw an interesting movie: illustration of the 5 unsynchronized oscillators metronoms. Once they are coupled together through a moving plate, their oscillation period tend to synchronize. Once the plate is removed (the coupling is broken), the periods desynchronize again.

In order to illustrate the tendency for phase-locking of anomalous events, we pick up a constant driving frequency Ω and a starting point x_0, then iteratively compute x_{n+1} from x_n through the non-linear relationship:

$$x_{n+1} = x_n + \Omega - \frac{k}{2\pi} \sin\left(2 * \pi x (mod 1)\right) \tag{1}$$

The map of the interaction between anomalies and the driving cycle is obtained by displaying all the computed x in the (Ω, x) plane, for all the described Ω. As we increase the parameter k in the non-linear forcing, an increasing number of "windows" are opening and widening in the (Ω, x) space. These "windows" are the orbital periods of limit circles encountered in the iterative process, and correspond to rational multiples of the driving frequency Ω onto which the system is locking. The frequency ratio of the model to the driving frequency describe a "devil's staircase" as the number and width of frequency-locked steps (corresponding to the windows) increases infinitely.

The Zebiak-Cane model is the first coupled atmosphere-ocean model taking into account the interaction between the seasonal cycle and the Hopf bifurcation oscillatory frequency

by relating the seasonal frequency Ω to the atmospheric/oceanic coupling strength parameter μ and the upwelling feedback parameter δ through a "Devil's terrace" (we recover a Devil's staircase with two non-linear forcing terms, equivalent of a k_1 and a k_2 in previous description). As the coupling strength is increased, both the amplitude and the time scale of the oscillations are enhanced. By tuning the parameters μ and δ, Zebiak-Cane model predicts 3 ENSO events over 10 years, which is in good agreement with observational data.

This ends the deterministic part of the discussion. Stochasticity will be now included by adding some noise into ENSO models.

2 Effect of Noise on the Hopf Bifurcation

Unresolved fast and short scales can be integrated into ENSO models by adding noise. As an example, the westerly wind bursts (WWB) events are characterized by velocities above 7 m s^{-1}, with a typical duration of a few days. These unresolved processes are known to trigger the propagation of perturbations in the form of equatorial Kelvin and Rossby waves. The correlation between this events can be verified in using a singular value decomposition analysis of the SST-Wind covariance matrix.

If we assume that the effect of WWB is a noise in the system, then what is the response of the model ?

The response of Zebiak-Cane model to white versus red noise is represented in the subcritical and supercritical regimes. A remarkable result is that red noise can trigger a response even before the critical point for the Hopf bifurcation is reached, that is, while still in the subcritical regime. Adding red noise in the model can thus result in lowering the bifurcation threshold.

We consider the following normal form:

$$\dot{X} = \lambda X - \omega Y - X(X^2 + Y^2), \tag{2}$$

$$\dot{Y} = \lambda Y + \omega X - Y(X^2 + Y^2). \tag{3}$$

In our model, X would be the temperature anomaly on the East Pacific coast (at some fixed longitude, say 30E) and Y the thermocline depth on the West coast. We derive the stochastic extension of this normal form by adding some noise in the equation (where dW_1 and dW_2 are independent, Gaussian noises).

$$dX = (\lambda X - \omega Y - X(X^2 + Y^2))dt + \sigma dW_1, \tag{4}$$

$$dY = (\lambda Y + \omega X - Y(X^2 + Y^2))dt + \sigma dW_2. \tag{5}$$

In polar coordinates ($r = \sqrt{X^2 + Y^2}$ and $\theta = \arctan \frac{X}{Y}$), the system reads (using Ito's formula for change of variables):

$$
\begin{aligned}
dR =& \frac{\partial r}{\partial x}dX + \frac{\partial r}{\partial y}dY + \frac{1}{2}\big(\frac{\partial^2 r}{\partial x^2}(dX)^2 + \frac{\partial^2 r}{\partial y^2}(dY)^2\big) + ... \\
=& \frac{X}{r}\sigma dW_1 + \frac{Y}{r}\sigma dW_2 + \frac{1}{2}\big(\frac{Y^2}{r^3}(\sigma)^2 + \frac{X^2}{r^3}(\sigma)^2\big)dt + ... \\
=& \underbrace{\sigma(\cos\theta dW_1 + \sin\theta dW_2)}_{\text{noise term}} + \underbrace{\frac{\sigma^2}{2r}dt}_{\text{additional drift term}} \quad , \\
d\theta =& -\frac{y}{r^2}\sigma dW_1 + \frac{x}{r^2}\sigma dW_2 + \frac{1}{2}\big(\frac{2XY}{r^4}(\sigma)^2 - \frac{2XY}{r^4}(\sigma)^2\big)dt + ... \\
=& \underbrace{\sigma(\frac{\cos\theta}{r}dW_2 - \frac{\sin\theta}{r}dW_1)}_{\text{additional drift term}}.
\end{aligned}
$$

We now make the following transformation

$$
\begin{aligned}
dX &= \sigma dW_1, \\
dY &= \sigma dW_2.
\end{aligned}
$$

$$
dR = (\lambda r - r^3 + \frac{\sigma^2}{2r})dt + \sigma(\cos\theta dW_1 + \sin\theta dW_2), \tag{6}
$$

$$
d\theta = \sigma(\frac{\cos\theta}{r}dW_2 - \frac{\sin\theta}{r}dW_1) + \omega dt. \tag{7}
$$

The stationary probability density function is

$$
\rho_S(r) = N\exp\big(\frac{\lambda r^2}{\sigma^2} - \frac{r^4}{2\sigma^2}\big), \tag{8}
$$

To prove (8) we first derive a system of coupled Fokker-Planck equations associated with the stochastic system above. We obtain:

$$
\underline{f} = \begin{pmatrix} \lambda r - r^3 + \frac{\sigma^2}{2r} \\ \omega \end{pmatrix} \tag{9}
$$

$$
\underline{\underline{g}} = \sigma \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{\sin\theta}{r} & \frac{\cos\theta}{r} \end{pmatrix} \tag{10}
$$

The diffusion operator D_{ij} is then

$$
\underline{\underline{D}} = \underline{\underline{g}}\underline{\underline{g}}^T = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2} \end{pmatrix} \tag{11}
$$

We look for a stationary density function $\rho_S(r)$, that is, $\partial_t \rho_S(r) = 0$). Thus

$$
0 = -\frac{1}{r}\big((\lambda r - r^3 + \frac{\sigma^2}{2r})r\rho^S\big)' + \frac{\sigma^2}{2r}(r\rho_S')
$$

$$\frac{\rho'^S}{\rho^S} = -\frac{2}{\sigma^2}\lambda r - 2\frac{r^3}{\sigma^2} + \frac{1}{r}$$
$$\ln \rho^S = -\frac{\lambda}{\sigma^2} r^2 - \frac{r^4}{2\sigma^2} + \ln r$$

Finally

$$\rho^S(r) = N \exp\big(\frac{\lambda r^2}{\sigma^2} - \frac{r^4}{2\sigma^2}\big) \tag{12}$$

3 Stochastic Optimals

What forcing pattern is maximizing the variability of the system under subcritical conditions? In order to address this question, we introduce white noise (delta-correlated noise in time) with a spatial correlation given by the covariance matrix C:

$$\frac{d\Psi}{dt} = A(t)\Psi + f(t) \tag{13}$$

$$\Psi_n = \Psi_{n-1} + A_{n-1}\Psi_{n-1}dt + \underbrace{\sqrt{dt}\zeta_{n-1}}_{\text{forcing}} \tag{14}$$

$$\mathbb{E}(\zeta_i) = 0, \tag{15}$$
$$\mathbb{E}(\zeta_i\zeta_j) = \delta_{ij}C \tag{16}$$

Now rewrite:

$$\begin{aligned}\Psi_n =& (1 + A_{n-1}dt)\Psi_{n-1} + \sqrt{dt}\zeta_{n-1} \\ =& (1 + A_{n-1}dt)\big((1 + A_{n-2}dt)\Psi_{n-2} + \sqrt{dt}\zeta_{n-1}\big) + \sqrt{dt}\zeta_{n-2} \\ =& ...\end{aligned}$$

By recurrence, we find:

$$= R_{0,n}\Psi_0 + \sqrt{(dt)}\sum_{k=0}^{n-2} R_{k+1,n}\zeta_k, \tag{17}$$

where we have introduced the "propagator" $R_{n-1,n}$ such that

$$\Psi_n = (1 + A_{n-1}dt)\Psi_{n-1} = R_{n-1,n}\Psi_{n-1} \tag{18}$$

Hence the mean variance is given by

$$\mathbb{E}(\Psi_n) = \mathbb{E}(R_{0,n}\Psi_0) + 0 \tag{19}$$

$$\mathbb{E}(< \Psi_n - \mathbb{E}(\Psi_n), \Psi_n - \mathbb{E}(\Psi_n) >) = dt\sum_{k=0}^{n-2}\sum_{j=0}^{n-2}\mathbb{E}(\underbrace{< R_{k+1,n}\zeta_k, R_{j+1,n}\zeta_j >}_{<R^T_{j+1,n}, R_{k+1,n}\zeta_k, \zeta_j>}) \tag{20}$$

Now let

$$B = dt \sum_{k=0}^{n-2} \sum_{j=0}^{n-2} R^T_{j+1,n} R_{k+1,n} \delta j, kdt \,, \tag{21}$$

so that the total covariance is

$$N = tr(BC) = \sum_{i,j} \underbrace{\lambda_i \mu_j}_{\text{eigenvalues of B (resp C)}} | < \underbrace{v_i, w_j}_{\text{eigenvectors of B (resp C)}} > |^2 \tag{22}$$

where C is the covariance matrix. The first eigenvector of B is called the stochastic optimum. The use of this eigenvector as a forcing pattern triggers the maximum response from the model.

Lecture 9 - Dansgaard-Oeschger Events

Henk A. Dijkstra; notes by Gunnar Peng & Chris Spalding

June 25, 2015

1 Dansgaard-Oeschger Events

In previous lectures, we discussed relatively short-term climatic fluctuations in the form of the El Niño Southern Oscillation (ENSO). It is worth asking whether the global climate system possesses similar, apparently stochastic oscillations over longer timescales, so long that no human has lived through an entire cycle and so did not bother to call it an oscillation. In order to seek such climatic modes we must turn to paleoclimatological data. Here, we highlight evidence from ice cores. Specifically, one can drill into the ice on Greenland and extract a core of ice, which has been inexorably laying down layer after layer of annual ice for millennia, recording its local environment as it does so. Bubbles record the paleoatmosphere and this has been used to deduce that, as suspected from its thermal absorption properties, CO_2 has gone done when the ice has gone up, consistent with its greenhouse gas properties. However, it is difficult to use CO_2 as an accurate thermometer. A better thermometer exists in the form of the ratio of "heavy" ^{18}O to "light" ^{16}O isotopes within the water molecules of the ice.

The standard way to measure isotope ratios is via the δ-notation defined for oxygen as

$$\delta^{18}\mathrm{O} = \frac{[^{18}\mathrm{O}]/[^{16}\mathrm{O}]_{\mathrm{sample}} - [^{18}\mathrm{O}]/[^{16}\mathrm{O}]_{\mathrm{standard}}}{[^{18}\mathrm{O}]/[^{16}\mathrm{O}]_{\mathrm{standard}}} \times 1000 \tag{1}$$

or, in other words, the measured deviation of the isotope ratio from a given standard multiplied by a thousand, yielding "units" of permil. The lighter isotope of oxygen evaporates more easily and so rain clouds are generally depleted in ^{18}O relative to seawater. Furthermore, as these rainclouds move towards the poles, where they deposit ice, they rain out even more heavy oxygen [1] until, by the time they form ice, they are depleted by 10s of permil relative to their source water. The degree of water/vapour fractionation is temperature dependent. Qualitatively, in colder periods, the fractionation is greater, leading to more negative δ^{18}O values in ice caps. Additionally, in colder climates, the poles tend to have more ice, which sequesters much of the ocean's light oxygen, making the ocean heavier in general and the ice caps lighter. Cumulatively, these processes make for *lighter ice caps in colder periods.*

With this theory under our belt, we can look at the measured ice isotope data from Greenland ice cores (Figure 1). Notice first that the record spans about 120,000 years, just long enough to resolve a full "glacial cycle", i.e., the global oscillation from generally higher

[1] This process is known as "Rayleigh Distillation".

temperatures to generally low temperatures. Records going back much further have revealed that glacial cycles follow such a ∼100,000 year periodicity for at least the previous four cycles. The cause of such a periodicity in glacial cycles is still an active area of research, but here we focus on shorter-scale features, occurring within glacial periods (periods of light ice caps and lower temperatures). Specifically, the ice cores record rapid changes in temperature from cold periods (known as stadiums) to warm periods (interstadials). Between about 70-20 kyr ago, the climate swung abruptly between these states with a periodicity of about 1470 years, with temperature changes up to about 10°C. These transitions have been named Dansgaard-Oeschger (DO) Events and possess a peculiar asymmetry in that the warming phas

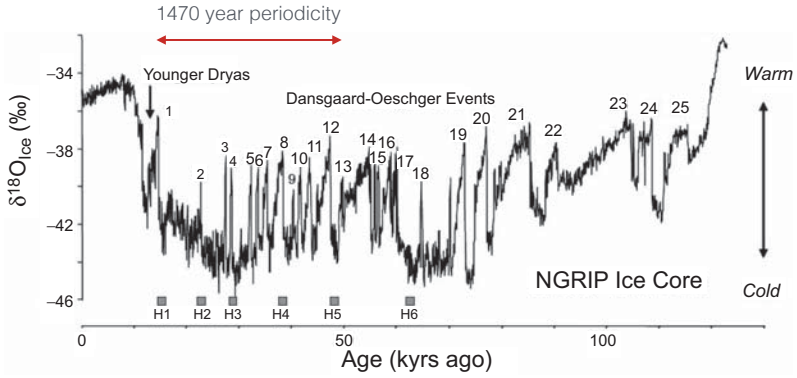

Figure 1: Plot of the δ^{18}O data from the NGRIP ice core record. The numbers refer to the interstadials (warm periods, high δ^{18}O). So-called Heinrich events are marked with the labels H1 to H6. The red arrow denotes the temporal range within which a strong 1470-year periodicity is observed in the δ^{18}O signal.

One mysterious feature of DO events is that they don't appear to correspond to any known natural frequency in climatic forcing, as opposed to the glacial cycles themselves which have been linked (with varying degrees of success) to Milankovich Cycles (the Earth's spin-axis precession, obliquity cycles and orbital eccentricity oscillations). The literature exploring the potential sources of DO-events is vast and we do not cover them here. However, what we present is more of a cautionary tale arising out of stochastic theory, namely, that stochastic forcing upon a dynamical system may amplify a periodic signal that you might have otherwise thought negligible, through the process of "Stochastic Resonance". In order to arrive at the key result, we first lay down a theory which allows for multiple stable states within the global overturning and then show how stochastic forcing might cause a system to shift between these states - analogous to DO-events.

1.1 Ocean circulation

Before attempting a dynamical systems description for DO-events, we provide a very brief background on the (Atlantic) Meridional Overturning Circulation, or (A)MOC. Looking at

the Atlantic as whole, there are two main sources of "deep water", meaning water that more or less descends to the ocean floor by way of negative buoyancy forcing. The first source is Antarctic Bottom Water (AABW), forming, unsurprisingly, near Antarctica. We shall not describe this source in much more detail except to note that it forms the deepest water in the Atlantic by way of extensive cooling in and around the edges of the Antarctic ice sheet. The second source, North Atlantic Deep Water (NADW), is our primary concern here, and more often supposed to be important in driving DO events.

NADW forms mostly in the Greenland and Labrador Seas. Its mechanism of formation is roughly as follows. Strong evaporation at low northern latitudes in the Atlantic increases the surface salinity of the water which is then carried northward into the regions of deep-water formation. The high salinity comes with a high density, which "pre-conditions" the water for convection by eroding its stable stratification. Subsequent cooling in the winter is usually able to destabilise the stratification, leading to deep convection and the formation of deep water. Owing to its high salinity, NADW is actually denser than southern-source AABW *at the surface* of the ocean. However, an interesting property of seawater is that the thermal expansion coefficient increases with pressure, making temperature more important for density at the bottom of the ocean than at the surface. What this amounts to is the coldness of the AABW leading to it being denser than NADW once both have reached the bottom of the ocean. The MOC is often described in terms of a zonally (longitudinally) averaged stream function

$$\Psi_{\,\mathrm{MOC}}(y,z,t) = \int_{z}^{0} \int_{x_w}^{x_e} v(x,y,z',t)\,dx\,dz', \tag{2}$$

where we have chosen $\Psi_{\,\mathrm{MOC}} = 0$ at the ocean surface and consider a basin with western margin situated at longitudinal position x_w and eastern margin at x_e. The variable y is meridional (latitudinal) distance along the surface of the Earth. Notice that $\Psi_{\,\mathrm{MOC}}$ is a volume transport and is usually measured in Sverdrups (Sv), where $1\,\mathrm{Sv} = 10^{\,6}\,\mathrm{m}^3\,\mathrm{yr}^{-1}$. Of interest is the heat carried by this volume, which depends upon the temperature of the northern-ward moving surface water. In general, stronger AMOC correlates with greater heat transport within recent, temporally-limited observations, but it is uncertain to extend this conclusion to longer timescales.

Direct measurements of the MOC are sparse, with a detailed evaluation of $\Psi_{\,\mathrm{MOC}}$ often drawn from numerical models. However, the MOC is now routinely monitored at 23^oN using the RAPID-MOCHA array. Essentially, what these observations have revealed is an extremely variable MOC strength, with total transport even appearing to reverse in sign briefly near the beginning of 2010. Paleoproxy evidence has been used to deduce changes to the AMOC in past climate regimes. In particular, the ratio of ^{231}Pa to ^{230}Th in Atlantic sediments has been used to deduce a dramatic and sudden switching off of the AMOC roughly 18 kyr ago, coinciding with Heinrich Stadial 1 (H1, figure 1). The AMOC appears to have abruptly switched back on again about 14,700 kyr ago, at the start of a period of warming known as the Bølling-Allerød interstadial (McManus et al. 2004). We do not have evidence for such transitions during all DO events, but the knowledge that the AMOC can change so dramatically warrants further investigation into whether multiple stable states might be inherent to the system.

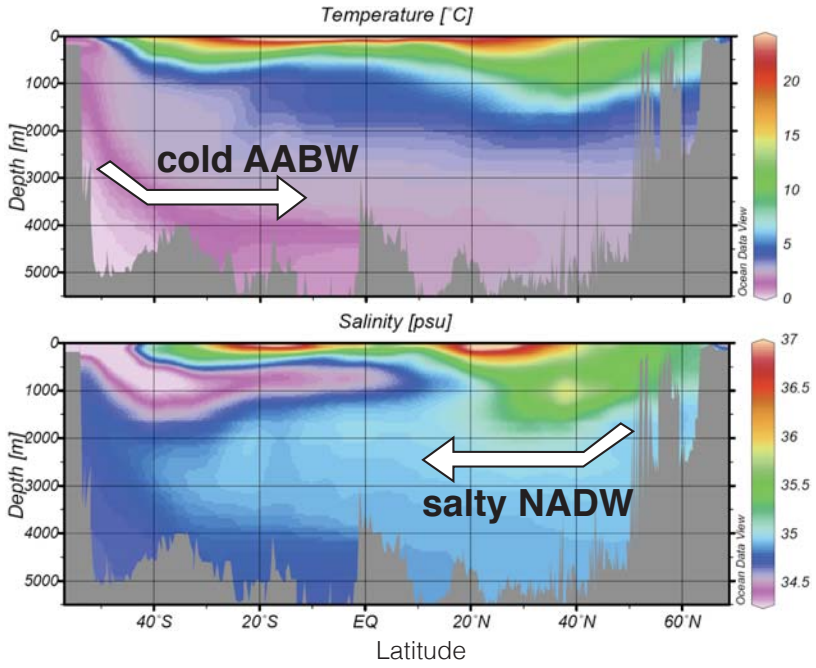

Figure 2: A section along the Atlantic. The top diagram shows the measured salinity and the lower depicts (potential) temperature. Water sinks to deep levels in the North and the South, as can be seen in the sections as a salty tongue descending from the north and a cold tongue from the south. At the ocean floor, the southern-source AABW is denser, by virtue of its low temperature than the northern-sources NADW, which is warmer. Variability in the associated heat transport from the northern regions has been implicated in the origin of DO events.

1.2 The salt-advection feedback

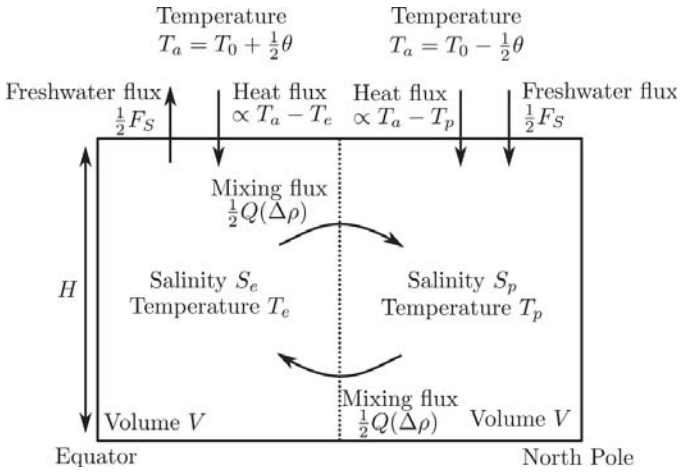

Figure 3: Stommel two-box model of salt-advection feedback.

The simplest picture one can imagine that captures the key aspects of the AMOC can be traced back to Stommel (1961). It is reasonable to suppose that the degree of mixing between equatorial and polar water reservoirs depends upon their mutual density difference $\Delta\rho$. A physical reason for this is that denser polar water is more pre-conditioned to convect to the ocean floor, enhancing meridional overturning and presumably pole-equator mixing. Accordingly, we begin the formulation of our simple model by supposing there to exist two reservoirs of water, one representing the poles and the other the equator, with temperatures and salinities T_p, S_p and T_e, S_e respectively (see figure 3).

We approximate the density of seawater as following a simple linear dependence upon T and S,

$$\rho = \rho_0 - \alpha_T(T - T_0) + \alpha_s(S - S_0), \tag{3}$$

where the thermal expansivity α_T and salinity coefficients α_s are assumed constant. We may then express the density difference between the two reservoirs as

$$\Delta\rho = -\alpha_T(T_p - T_e) + \alpha_s(S_p - S_e), \tag{4}$$

which in turn governs the mixing rate $Q(\Delta\rho)$.

These two reservoirs not only interact with each other, but are individually forced at their boundaries. Specifically, we suppose the temperature to relax, over a timescale t_r, to the local atmospheric temperature T_a. In the interest of symmetry, we suppose the polar box to possess $T_{a,p} = T_0 - \theta/2$ and the equatorial box to relax to $T_{a,e} = T_0 + \theta/2$. Salinity, too, is forced. However, a crucial aspect of atmosphere-ocean interactions is that, whereas colder water will have greater tendency to draw in heat than warmer water, salty water does not stimulate the atmosphere to rain on it! Consequently, salinity forcing is poorly modelled as a relaxation to some equilibrium value. We adopt a more physical form for the forcing whereby a prescribed flux $F_s/2$ of fresh water enters the polar ocean (in the form of rain, meltwater, etc.), with an equal volume (for simplicity) leaving at the equator by evaporation. As S_0 is the typical value of salinity in the ocean, the result of the freshwater flux is a decrease in salinity in the polar box with rate proportional to $F_S S_0$ and an equivalent increase in the equatorial box.

We may now write out the equations governing the two-basin system (see Cessi 1994):

$$\dot{T}_e = -\frac{1}{t_r}[T_e - (T_0 + \tfrac{1}{2}\theta)] - \frac{Q(\Delta\rho)}{2}(T_e - T_p), \qquad \dot{S}_e = +\frac{F_s}{2H}S_0 - \frac{Q(\Delta\rho)}{2}(S_e - S_p) \tag{5a}$$

$$\dot{T}_p = -\frac{1}{t_r}[T_p - (T_0 - \tfrac{1}{2}\theta)] - \frac{Q(\Delta\rho)}{2}(T_p - T_e), \qquad \dot{S}_p = -\frac{F_s}{2H}S_0 - \frac{Q(\Delta\rho)}{2}(S_p - S_e) \tag{5b}$$

where H is the ocean depth. We can now see that in the form written above, $Q(\Delta\rho)$ must be positive. The reason for this is that although Q is physically the advection of water between two reservoirs, this advection is closed, with as much going in as is coming out for each reservoir. If you reverse the direction of circulation the quantity of polar water moving into the equator and vice versa remain unchanged. With this in mind, considering the simplicity of the model, we are free to choose a functional form for Q that depends only on the magnitude of $\Delta\rho$. For definiteness, we choose

$$Q(\Delta\rho) = \frac{1}{t_d} + \frac{q}{\rho_0^2 V}(\Delta\rho)^2, \tag{6}$$

where V is the volume of each reservoir, q is a dimensional transport coefficient and t_d is the timescale of diffusive mixing between the two reservoirs that would occur in the absence of a density difference.

We are interested in obtaining the possible steady state solutions to the system of equations above and so it is convenient to define the temperature and salinity differences

$$\Delta T \equiv T_e - T_a, \qquad \Delta S \equiv S_e - S_a \tag{7}$$

and work in terms of these variables. From equations 5, we obtain the time evolution of the temperature and salinity differences:

$$\frac{d\Delta T}{dt} = -\frac{1}{t_r}(\Delta T - \theta) - Q(\Delta\rho)\Delta T, \tag{8a}$$

$$\frac{d\Delta S}{dt} = \frac{F_s}{H}S_0 - Q(\Delta\rho)\Delta S. \tag{8b}$$

We now introduce appropriate scales with which to reduce the dynamical variables ΔT and ΔS, together with time t, to their respective dimensionless forms. Appropriate choices are as follows

$$x \equiv \frac{\Delta T}{\theta}, \qquad y \equiv \frac{\alpha_s \Delta S}{\alpha_T \theta}, \qquad t' \equiv \frac{t}{t_d}. \tag{9}$$

Once scaled, the dynamical equations for $x(t')$ and $y(t')$ read

$$\dot{x} = -\alpha(x-1) - x\big[1 + \mu^2(x-y)^2\big], \tag{10a}$$

$$\dot{y} = F - y\big[1 + \mu^2(x-y)^2\big], \tag{10b}$$

where

$$\alpha = \frac{t_d}{t_r}, \qquad \mu^2 = \frac{q t_d (\alpha_t \theta)^2}{V}, \qquad F = \frac{\alpha_s S_0 t_d}{\alpha_t \theta H} F_s. \tag{11}$$

The parameter α is the ratio of the diffusive timescale to the timescale over which temperature would exponentially decay to the local atmospheric value. The parameter μ measures the strength of the buoyancy-driven convection between the two basins relative to the diffusive mixing. The parameter F measures the amount of freshwater forcing.

Parameter	Meaning	Value	Unit
t_r	temperature relaxation timescale	25	days
H	mean ocean depth	4,500	m
t_d	diffusion time scale	180	years
t_a	advection time scale	29	years
q	transport coefficient	1.92×10^{12}	$\mathrm{m^3 s^{-1}}$
V	ocean volume	$300 \times 4.5 \times 8,250$	$\mathrm{km^3}$
α_T	thermal expansion coefficient	10^{-4}	$\mathrm{K^{-1}}$
α_S	haline contraction coefficient	7.6×10^{-4}	–
S_0	reference salinity	35	$\mathrm{g\,kg^{-1}}$
θ	meridional temperature difference	25	K

Table 1: Parameters of the stochastic salt advection model.

We may simplify the equation above by noting that for parameters typical of the real ocean (see table) $\alpha \gg 1$, which means that the reservoirs will equilibrate with their local forcing temperatures much more rapidly than they are likely to mix each other's temperatures. Therefore, we may suppose that x remains close to 1 which reduces the problem to an ODE in $y(t)$ alone (where we drop the primes on t' for convenience):

$$\frac{dy}{dt} = F - y\left[1 + \mu^2(1-y)^2\right]. \tag{12}$$

If we suppose for now that $F = \bar{F}$ is independent of time, we can represent the time evolution of y using a potential function $V(y)$:

$$\frac{dy}{dt} = -V'(y), \qquad \text{where } V(y) = -\bar{F}y + \frac{1}{2}y^2 + \mu^2\left(\frac{1}{4}y^4 - \frac{2}{3}y^3 + \frac{1}{2}y^2\right), \tag{13}$$

and its derivative with respect to y is denoted by the prime. We illustrate $V(y)$ using $\bar{F} = 1.1$ and $\mu^2 = 6.2$ in Figure 4. As can be seen, $V(y)$ is a double-welled potential with two stable minima and an unstable maximum. In order to transition from one potential well to the other, a finite amplitude "kick" in y is required.

Recalling that y is simply the dimensionless salinity difference, we immediately see that the two reservoirs can remain in a stable state with either a large salinity difference or a small one. Physically, these correspond to the following. The poles are colder and fresher than the equator. If we freshen the poles, we increase ΔS, but because temperature drives the convection, this freshening reduces $\Delta\rho$ and so the MOC weakens. Therefore, the higher (lower) value of y is usually referred to as the off (on) state of circulation. Another way to look at it is that in order to balance the freshwater forcing at a large ΔS we need less mixing between the reservoirs than if we have a smaller ΔS. Ultimately, the conclusion here is that the meridional overturning circulation can jump between the on and off states impulsively, given a finite-amplitude forcing, such as a particularly large ice-melt event.

Of course, freshwater forcing F is unlikely to be constant in reality. Next, we consider F to vary stochastically, perhaps modelling rainstorms, or ice-sheet collapses, which create a set of random kicks of freshwater flux which we model as white noise with amplitude σ such that $F = \bar{F} + \sigma\xi(t)$. This leads to the stochastic Itô equation

$$dY_t = -V'(Y_t)\,dt + \sigma\,dW_t. \tag{14}$$

Note here that the result of adding fluctuations to F is additive noise in the equation for Y, rather than noise in the potential $V(y)$.

As we have seen in previous lectures, we can write down the forward Fokker-Planck equation in order to solve for the probability density function $p(y,t)$ that generates a given trajectory in Y_t

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y}\left(V'(y)p\right) + \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial y^2}. \tag{15}$$

Now, in the deterministic case before, we sought time-independent solutions for y. Of course, it makes no sense to look for truly time-independent solutions for the random variable Y_t, but a *statistically steady* solution may be found by setting $\partial p/\partial t = 0$ and

solving for the function $p^{\text{stat}}(y)$ satisfying stationary statistics. The solution is relatively straightforward and we simply state the result,

$$p^{\text{stat}}(y) = Ce^{-\frac{2}{\sigma^2}V(y)}, \text{ where } C = \left(\int_{-\infty}^{\infty} e^{-\frac{2}{\sigma^2}V(y)}\, dy\right)^{-1} \tag{16}$$

is the normalization coefficient and we have used the boundary condition that $p \to 0$ as $y \to \pm\infty$.

Some numerical results for equations 14 and 15 are shown in figure 4. The histograms and probability densities are initially peaked at the well near which the system was launched, indicating that the peak at $y = y_b$ is difficult to cross. They do eventually spread out, though, and attain the steady state given by equation 16. In this state, the system typically fluctuates around in one of the two wells and randomly transitions between them, while spending more time overall in the deeper well.

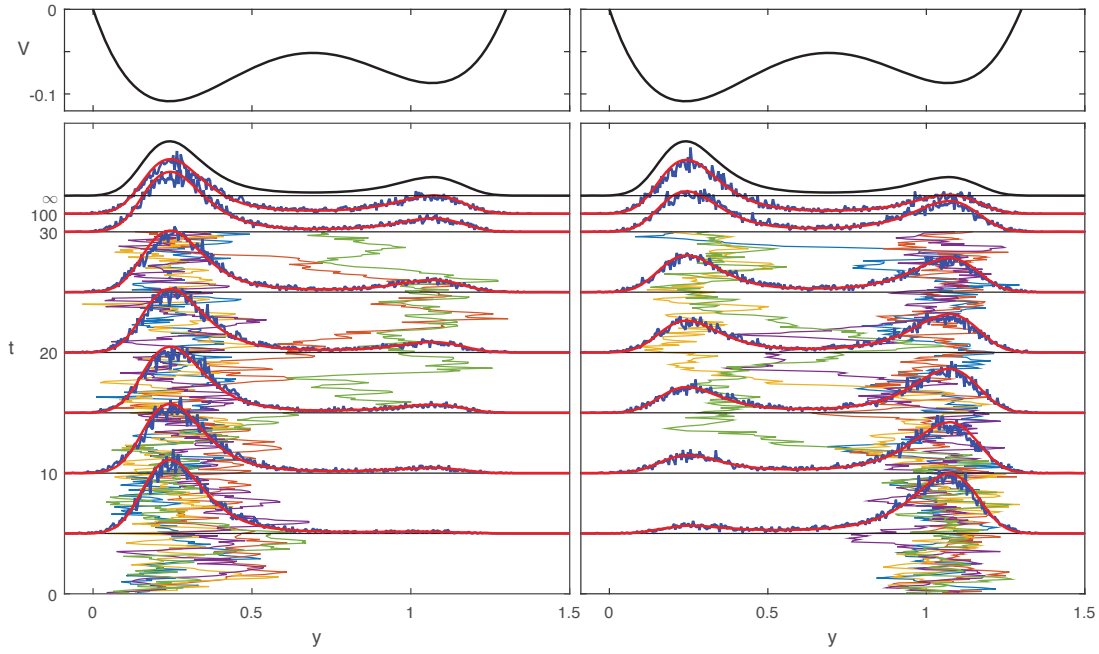

Figure 4: Motion in the double-well potential $V(y)$ from equation 13 with $\bar{F} = 1.1$ and $\mu^2 = 6.2$. Top: Potential $V(y)$. Bottom: Stochastic motion (equation 14) with noise amplitude $\sigma = 0.2$ starting from $Y_0 = 0$ (left) or $Y_0 = 1$ (right). The time evolution of five realizations are shown, as well as histograms (blue) from 10 000 realizations and the probability density (red) obtained from numerical solution of the corresponding Fokker–Planck equation (15). The distribution labelled $t = \infty$ is the steady-state distribution $p^{\text{stat}}(y)$ from equation 16.

1.3 Escape time

We discussed in a previous lecture the concept of mean escape times. In this context, suppose we are in the "on"-state $y = y_a$ of meridional overturning but subject the system to given, stochastic freshwater forcing. How long is it likely to take for the system to flip into the other ("off") state $y = y_c$?

The expected time $\bar{T}(y)$ required to escape to y_c when starting from y satisfies an equation related to the backward Kolmogorov equation:

$$-1 = -V'\bar{T}' + \tfrac{1}{2}\sigma^2\bar{T}'', \qquad \text{with} \qquad \bar{T}(y_c) = 0, \qquad \bar{T}'(-\infty) = 0, \tag{17}$$

where the boundary conditions state that it takes no time to reach y_c when starting from y_c, and that the escape time varies very little for y far below the potential well at y_a since the restoring deterministic drift is very strong there.

The equation is a linear first-order equation for $\bar{T}'(y)$ which we solve by multiplying by the integrating factor $\exp(-2V(y)/\sigma^2)$:

$$-e^{-\frac{2}{\sigma^2}V} = e^{-\frac{2}{\sigma^2}V}\left(-V'\bar{T}' + \tfrac{\sigma^2}{2}\bar{T}''\right) = \tfrac{\sigma^2}{2}\left(e^{-\frac{2}{\sigma^2}V}\bar{T}'\right)'. \tag{18}$$

Integration of both sides and using the boundary condition $\bar{T}'(-\infty) = 0$ yields

$$\bar{T}'(y) = e^{\frac{2}{\sigma^2}V(y)}\int_{-\infty}^{y} -\tfrac{2}{\sigma^2}e^{-\frac{2}{\sigma^2}V(s)}\,ds = -\tfrac{2}{\sigma^2}\int_{-\infty}^{y} e^{\frac{2}{\sigma^2}[V(y)-V(s)]}\,ds. \tag{19a}$$

A second integration using $\bar{T}(y_c) = 0$ yields

$$\bar{T}(y) = -\tfrac{2}{\sigma^2}\int_{z=y_c}^{y}\int_{s=-\infty}^{z} e^{\frac{2}{\sigma^2}[V(z)-V(s)]}\,ds\,dz. \tag{20}$$

Hence the mean escape time from the "on" state $y = y_a$ to the "off" state $y = y_c$ is

$$\boxed{\bar{T}(y_a) = \frac{2}{\sigma^2}\int_{z=y_a}^{y_c}\int_{s=-\infty}^{z}\exp\left(\frac{2}{\sigma^2}[V(z) - V(s)]\right)ds\,dz.} \tag{21}$$

1.3.1 Asymptotic approximation using Laplace's method

We can obtain an asymptotic approximation to the above integral in the limit of small noise, where σ^2 is much smaller than the typical variation $V(y_b) - V(y_a)$ of the potential, so that we can treat $M = 2/\sigma^2$ as a large parameter. In this case, the main contribution to the integral in equation 21 comes from the region where the exponent $M[V(z) - V(s)]$ is maximal, i.e. $z \approx y_b$ and $s \approx y_a$. The contributions from any other regions are exponentially small and can be ignored. We can thus approximate the result as

$$\bar{T}(y_a) \approx M\int_{y_b-\epsilon}^{y_b+\epsilon} e^{MV(z)}\,dz\int_{y_a-\epsilon}^{y_a+\epsilon} e^{-MV(s)}\,ds, \tag{22}$$

where $\epsilon > 0$ is small.

After a change of variables $z = y_b + x$ or $s = y_a + x$, the two integral factors in equation (22) have the form

$$I \equiv \int_{-\epsilon}^{\epsilon} e^{Mf(x)}\, dx, \tag{23}$$

where $M \gg 1$ and $f(x) = V(y_b + x)$ or $f(x) = -V(y_a + x)$ has a maximum at $x = 0$. We have argued that almost all of the contribution to the integral I comes from the region near this maximum, so we may Taylor expand $f(x)$ as $f(x) \approx f(0) + f''(0)x^2/2$, where no linear term is present and $f''(0) < 0$ since $x = 0$ is a maximum. After the expansion, we can extend the limits to infinity, again because the contributions from regions away from the exponential maximum near $x = 0$ are negligible, and hence

$$\int_{-\epsilon}^{\epsilon} e^{Mf(x)}\, dx \approx e^{Mf(0)} \int_{-\epsilon}^{\epsilon} e^{-\frac{1}{2}M|f''(0)|x^2}\, dx \tag{24a}$$

$$\approx e^{Mf(0)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}M|f''(0)|x^2}\, dx \tag{24b}$$

$$\approx e^{Mf(0)} \sqrt{\frac{2\pi}{M|f''(0)|}}, \tag{24c}$$

where we have made use of the standard result $\int_{-\infty}^{\infty} e^{-\alpha x^2}\, dx = \sqrt{\pi/\alpha}$.

The two integral factors in equation 22 are thus

$$\int_{y_b-\epsilon}^{y_b+\epsilon} e^{MV(z)}\, dz \approx \sqrt{\frac{2\pi}{M|V''(y_b)|}} e^{MV(y_b)}, \tag{25a}$$

$$\int_{y_a-\epsilon}^{y_a+\epsilon} e^{-MV(s)}\, ds \approx \sqrt{\frac{2\pi}{M|V''(y_a)|}} e^{-MV(y_a)}, \tag{25b}$$

and hence the mean escape time from the "on" state $y = y_a$ to the "off" state $y = y_c$ is approximately

$$\boxed{\bar{T}(y_a) = 2\pi \sqrt{\frac{1}{|V''(y_a)|\,|V''(y_b)|}} \exp\left(\frac{2}{\sigma^2}\left[V(y_b) - V(y_a)\right]\right).} \tag{26}$$

From the calculations, we can see that this escape time is the same from any state in the well near $y = y_a$ over the peak $y = y_b$ to any state in the well near $y = y_c$. This is in line with our intuition that, for weak noise, the deterministic drift quickly drives the system to the bottom of the well $y = y_a$ where it fluctuates until eventually a large enough random perturbation kicks the system over the crest $y = y_b$ and it falls into the other well $y = y_c$.

1.4 Periodic forcing

Within the autonomous framework above, the system will jump between on and off states stochastically, but will not display any periodic behaviour, as is observed for DO events.

We therefore augment the previous model with a periodic modulation to the deterministic part of the freshwater forcing, so that

$$F = \bar{F} + \sigma\xi(t) + A\sin\left(2\pi\frac{t}{T}\right), \tag{27}$$

where A is the amplitude of periodic forcing and T is the dimensionless period of forcing (as we are still working with dimensionless variables). The governing equation is thus $dy/dt = -dV/dy + \sigma\xi(t)$, where the potential $V(y,t)$ can be chosen as

$$V(y,t) = -\bar{F}y + \frac{1}{2}y^2 + \mu^2\left(\frac{1}{4}y^4 - \frac{2}{3}y^3 + \frac{1}{2}y^2\right) - A\sin\left(2\pi\frac{t}{T}\right)(y - 0.7). \tag{28}$$

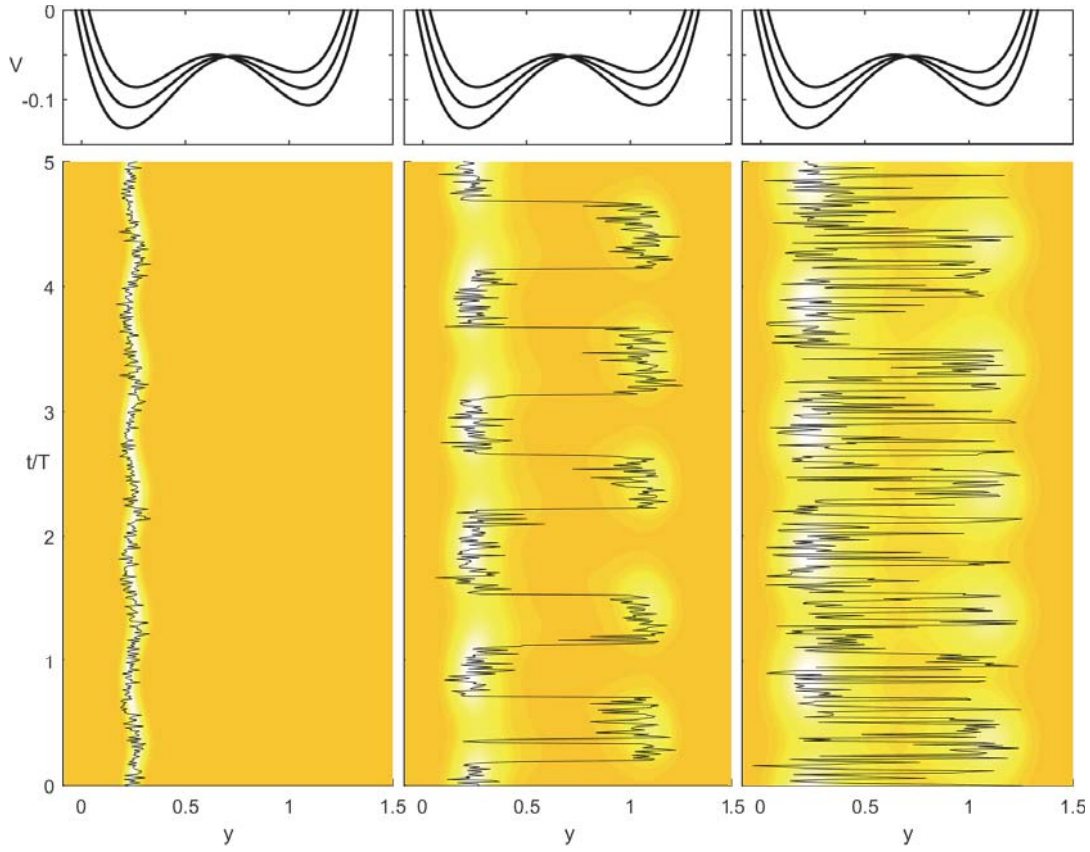

Figure 5: Motion in a time-periodic double-well potential (equation 28 with $\bar{F} = 1.1$, $\mu^2 = 6.2$ and $A = 0.05$). Top: The potential V at $t = -T/2,\ 0,\ T/2$. Bottom: Stochastic motion with noise amplitude $\sigma = 0.05$ (left), $\sigma = 0.15$ (middle), $\sigma = 0.25$ (right). The time evolution of one realization is shown (black curve), as well as the probability density (heat map) obtained from evolving the corresponding Fokker–Planck equation forward until a time-periodic state is reached. The period T chosen corresponds to 100 000 years.

In Figure 5, we show what happens for a small perturbation ($A = 0.05$) to the mean forcing $\bar{F}$ for various values of the noise amplitude σ. For small noise, the system remains

in the deeper well most of the time as expected. For large noise, the probability density system frequently transitions between the two wells, almost as if the middle peak at $y = y_b$ did not exist, and the periodicity is quite weak. However, for an intermediate value of noise strength, we recover periodic behaviour on the timescale T. The response is not a small perturbation, but a jump between on and off states every cycle. We have ended up with a system exhibiting so-called "**stochastic resonance**", whereby the noise is just large enough to switch between states almost every time the background forcing oscillates.

It is unclear whether the DO events are in fact generated by such a mechanism (the addition of a $\sim$1500 year periodicity in freshwater forcing is *ad hoc* – we know of no such forcing in reality), but it nonetheless constitutes a fascinating result that ordered behaviour may come out of the addition of white noise.

Lecture 10: Dansgaard-Oeschger Events (continued) & Ocean Western Boundary Current Variability

Henk A. Dijkstra; notes by Tom Beucler & Andre Souza

June 26, 2015

1 Stochastic Resonance

It is possible to synchronize transitions when you have a periodic forcing on a double potential well:

$$V(x) = -\frac{x^2}{2} + \frac{x^4}{4} - \epsilon x \cos(\Omega\tau)$$

Note that the fixed points location (defined by $V'(x) = 0$ and $V''(x) > 0$) do not strongly vary with τ if the amplitude ϵ is small (in which case the fixed points are given by $x_\pm \approx \pm 1$), unlike the value of the potential V at this fixed points. For $\epsilon \ll 1$, these two values are approximately given by:

$$V(x_\pm) \approx V(\pm 1) = -[\frac{1}{4} \pm \epsilon \cos(\Omega\tau)]$$

Using the Laplace's approximation, the transition times are approximately given by:

$$\boxed{< t_{-1\to 1} > \approx 2\pi \sqrt{\frac{1}{-V''(0)V''(-1)}} \exp\{\frac{2[V(0) - V(-1)]}{\sigma^2}\} \approx \sqrt{2}\pi \exp[\frac{1 - 4\epsilon\cos(\Omega\tau)}{2\sigma^2}]}$$

$$\boxed{< t_{1\to -1} > \approx 2\pi \sqrt{\frac{1}{-V''(0)V''(1)}} \exp\{\frac{2[V(0) - V(1)]}{\sigma^2}\} \approx \sqrt{2}\pi \exp[\frac{1 + 4\epsilon\cos(\Omega\tau)}{2\sigma^2}]}$$

The transition times vary with τ as the potential changes shape. Because the variance in the transition time is very small compared to the transition time itself, the transition occurs over a small time-interval. As a consequence, the Fourier spectrum has a strong peak at the forcing frequency Ω. In the case of a small periodic forcing ϵ, the synchronization can occur for moderate values of σ.

For example, if we take the small amplitude to be $\epsilon = 0.1$, the shape of the potential is very close to a double well. If we suppose that at $\tau = 0$, the state of the system is near $x_+ \approx 1$, the transition time $< t_{1\to -1} >$ at this τ is maximal, as the potential well is deepest. If $\frac{\pi}{\Omega} \ll < t_{1\to -1} > (\tau = 0)$, then the well will change shape and the system will almost surely exit the well at $\tau = \frac{\pi}{\Omega}$ where the mean escape time $< t_{1\to -1} >$ is minimal:

The same reasoning can be applied when the system starts near $x_- \approx -1$ at $\tau = \frac{\pi}{\Omega}$. Thus, for small amplitude, the transitions of the system approximately occur when τ is a

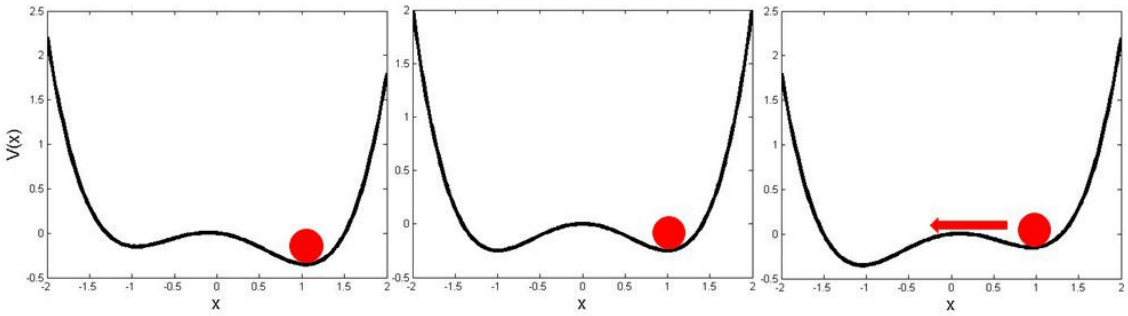

Figure 1: $V(x)$ vs x for $\tau = 0$ (left) $\tau = \frac{\pi}{2\Omega}$ (middle) and $\tau = \frac{\pi}{\Omega}$ (right). The state of the system is represented by the red dot.

multiple of $\frac{\pi}{\Omega}$, and the system stochastically resonates with the forcing of angular frequency Ω:

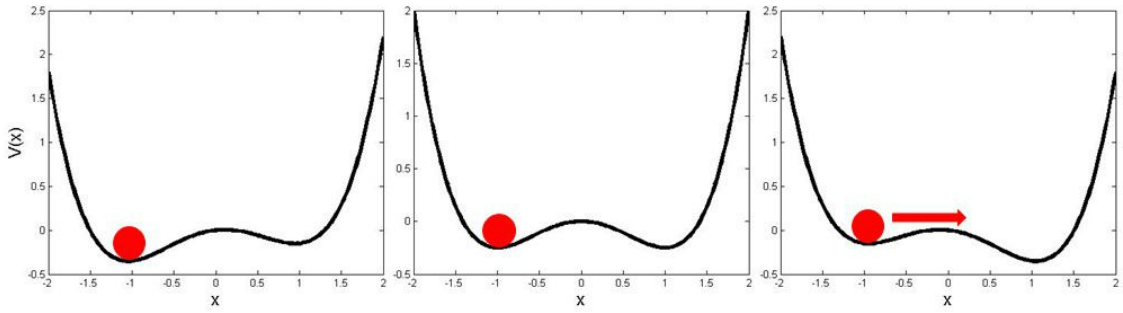

Figure 2: $V(x)$ vs x for $\tau = \frac{\pi}{\Omega}$ (left) $\tau = \frac{3\pi}{2\Omega}$ (middle) and $\tau = \frac{2\pi}{\Omega}$ (right). The state of the system is represented by the red dot.

The synchronization of transitions has applications in climatic models: If a periodic forcing is imposed, it is possible to have synchronization in the transitions, which allows for the existence of multiple equillibria. However, attributing this signal to stochastic resonance is a controversial topic. For instance, stochastic resonance has been suggested as a mechanism for the Dansgaard-Oeschger events, but many questions/doubts remain:

- Did multiple states of the Meridional Overturning Circulation exist during glacial times?
- What is the origin of the 1500 years period in the freshwater forcing?

2 Ocean Western Boundary Current Variability

2.1 Chaotic behavior of the Kuroshio current

We are now interested in systems that exhibit chaotic behaviors, and ask ourselves what to do when the circulation is so hard to model directly? Looking at the Sea Surface Height

(SSH) variability, we can see that it is predominant in the Western boundaries of the basins. In this lecture, we will focus on the Kuroshio current in Japan. Qiu et al. (2005) have observed a strong interannual-decadal time scale transitions between different Kuroshio paths (which look like spaghetti on the SSH contours), referred as path transitions. Looking at the variation of the path length in time, you have a strong variability: we start with a low path length, and then suddenly see a much bigger path lengths, corresponding to a meandric current. In this case, due to the lack of data, it is hard to test for red noise; instead they use a very simple reduced gravity shallow-water model for the wind:

$$\frac{\partial \underline{u}}{\partial t} + (\underline{u} \cdot \underline{\nabla})\underline{u} + f\underline{k} \wedge \underline{u} = -g'\underline{\nabla \eta} + A_H \nabla^2 \underline{u} + \frac{\underline{\tau}}{\rho h} - \gamma |\underline{u}| \underline{u}$$

$$\frac{\partial h}{\partial t} + \underline{\nabla} \cdot (h\underline{u}) = 0$$

Taking a few points along the coast line, the wind profile $\underline{\tau}$ is mimicked. As it is very difficult to measure/determine the lateral friction A_H, it is taken as a control parameter. For $A_H = 220 m^2.s^{-1}$, Pierini et al. (2009) notice a strong transition in the behavior of the model, which does not correspond to the observations anymore, if metrics such as SSH are considered. The strong transition in the behavior of the system leads us to think that a nonlinear transition occurs when A_H becomes large enough.

2.2 Deterministic quasi-geostrophic barotropic model

To model this transition, we adopt a Quasi-geostrophic barotropic model, giving us an evolution equation for the streamfunction ψ:

$$\boxed{\frac{\partial \zeta}{\partial t} + \underline{u} \cdot \underline{\nabla \zeta} + \beta \frac{\partial \psi}{\partial x} = \frac{\nabla^2 \zeta}{Re} + \alpha \underline{k} \cdot (\underline{\nabla} \wedge \underline{\tau})}$$

where the horizontal velocity $\underline{u}$ and the relative vorticity ζ are related to the streamfunction ψ by:

$$\underline{u} = \underline{k} \wedge \underline{\nabla \psi}$$

$$\zeta = \nabla^2 \psi$$

We choose a double-gyre wind stress:

$$\underline{\tau} = -\cos(2\pi y)\underline{i}$$

so that we have eastwards wind in the Northern part of the domain and westwards wind in the Southern part of the domain. The control parameters is now the Reynolds number of the flow:

$$Re = \frac{UL}{A_H} \propto A_H^{-1}$$

Note that we have chosen the wind field to be symmetrical about the mean axis of the domain, which implies meridional symmetry of the equations of motion about this axis. We thus observe a Pitchfork bifurcation as the Reynolds number exceeds its critical value, which is the only co-dim 1 bifurcation leading to symmetry breaking. The flow can indeed break from a double-gyre configuration to:

- Steady-state streamfunction patterns.
- A jet-down configuration.
- A jet-up configuration.

At low-frequency, the two eigenmodes of the system (P and L) can merge to give interesting periods depending on the choice of the parameters in the model; for instance, the period of the gyre mode at Hopf bifurcation ($Re \sim 80$) is approximately 1.5y whereas the period of other Rossby-basin modes is much shorter, of order 2-4 months. Increasing Re and looking at the bifurcation diagram in (Re, $\frac{\psi_{min}+\psi_{max}}{|\psi_{maix}|}$) space, the deterministic system undergoes in this order:

- A Pitchfork bifurcation.
- A Hopf bifurcation.
- Gyre modes then appear.
- A homoclinic orbit then appears, connecting back the system to its initial fixed point.

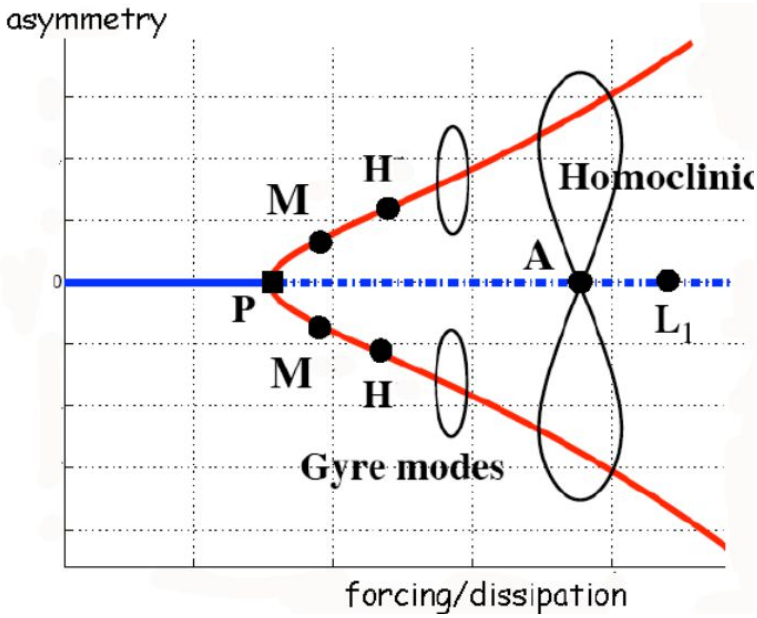

Figure 3: Schematic bifurcation diagram for the double-gyre model

The central question of this lecture is then: What happens if noise is added to the system? In the case of a chaotic system, such as the previous one, the PDEs are too hard to tackle directly, and we thus need to derive a low-order model. We:

- Neglect the diffusive term $\frac{1}{Re}\nabla^2\zeta$ and replace it with a Rayleigh drag $-\mu\zeta$.
- Choose the wind-stress strength as the control parameter.

We then truncate the system by focusing on the 4 first modes of the basin in the meridional direction:

$$\boxed{\psi(x,y,t) = \sum_{k=1}^{4} A_k(t) G(x) \sin(ky)}$$

where:

$$G_s(x) = \exp(-sx) \cdot \sin x$$

and $(x,y) \in [0,\pi]^2$. If we Galerkin project the equations for ψ, we obtain a low-order model, consisting of 4 ODEs for the amplitudes $A_k(t)$:

$$\frac{dA_1}{dt} = c_1(A_1A_2 + A_2A_3 + A_3A_4) - A_1$$

$$\frac{dA_2}{dt} = 2c_2(A_1A_3 + A_2A_4) - c_2A_1^2 - A_2 + c_5\alpha$$

$$\frac{dA_3}{dt} = c_3A_1(A_4 - A_2) - A_3$$

$$\frac{dA_4}{dt} = -c_4A_2^2 - 2c_4A_1A_3 - A_4$$

A wind-stress amplitude $\tau_0 = 0.1Pa$ gives $\alpha = 20$. It is possible to prove that this low-order model exhibits a transition sequence as α increases:

- First, we have a Pitchfork bifurcation.
- Then, we have a Hopf bifurcation.
- Finally, a homoclinic orbit appears.

Numerically, we show that the previous system has a steady attractive set with a first Lyapunov exponent $\lambda_1 > 0$.

2.3 Stochastic low-order models for chaotic systems

We are now able to add noise in the system, by transforming the previous system of ODEs in a system of SDEs:

$$dA_1 = [c_1(A_1A_2 + A_2A_3 + A_3A_4) - A_1]dt$$

$$dA_2 = [2c_2(A_1A_3 + A_2A_4) - c_2A_1^2 - A_2]dt + c_5\alpha(dt + \sigma \circ dW)$$

$$dA_3 = [c_3A_1(A_4 - A_2) - A_3]dt$$

$$dA_4 = [-c_4A_2^2 - 2c_4A_1A_3 - A_4]dt$$

Note that adding noise to a dynamical system changes the definition of what we call an attractor. We study the very simple 1D example:

$$\begin{cases} \dot{x} = -x + t \\ x(s) = x_0 \end{cases}$$

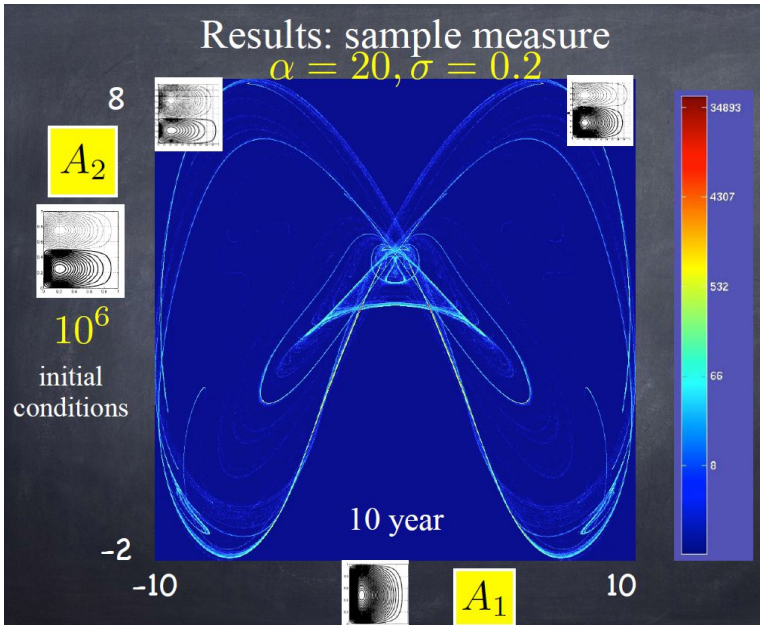

Figure 4: Numerical simulation of the low-order dynamical system with $N = 10^6$ initial conditions for a wind-stress $\tau = 0.1hPa$

Adding noise to this equation will not change its monotony. The solution of this system is given by:

$$x(t) = \exp(s - t) \cdot [x_0 - s + 1] + t - 1$$

If we wanna look at the attractor in terms of asymptotic limit, we start by noticing that all the trajectories eventually go to $t - 1$. If you consider the following flow:

$$\varphi(t, s)[x] = \exp(s - t) \cdot [x - s + 1] + t - 1$$

you are forced to go backwards in time and define an attractor differently.

$$\boxed{\lim_{s \to -\infty} |\varphi(t, s)[x] - A(t)| \to 0}$$

where:

$$A(t) = t - 1$$

Indeed, the noise might make the system change attractor, which prevents us of going forward in time to define the attractor. Similarly to what we did for the previously derived low-order model, we can add noise to the Lorenz model:

$$dX = s(Y - X)dt + \sigma X dW$$

$$dY = (rX - Y - XZ)dt + \sigma Y dW$$

$$dZ = (-bZ + XY)dt + \sigma Z dW$$

Numerically, we initiate the system with many initial conditions (typically $N \sim 10^6 - 10^9$) and perturb each initial condition with a single noise realization. We then look at the density of trajectories in phase space. To study the effects of noise in the wind-stress forcing on the intrinsic variability in this PDE model, three methods exist:

1. Study the local PDF through linearized dynamics (cf Kuehn et al. 2012)
2. Use dynamical orthogonal field theory (cf Sapsis et al. 2009 and Sapsis et al. 2013)
3. Use non-Markovian model reduction techniques (cf Chekroun et al. 2015)

We will study the second method and apply it to our stochastic barotropic QG model:

$$\frac{\partial \underline{u}}{\partial t} = -\underline{\nabla} p + \frac{\nabla^2 \underline{u}}{Re} - (\underline{u} \cdot \underline{\nabla})\underline{u} - f\underline{k} \wedge \underline{u} + \underline{\tau_W}(\underline{x}, t)$$

$$\underline{\nabla} \cdot \underline{u} - 0$$

where we separate the wind in its deterministic double-gyre part and a stochastic part:

$$\underline{\tau_W} = \tau_{DG}(y)\underline{i} + \sigma \tau_{stochastic}(x, y, t)\underline{i}$$

2.4 The dynamical orthogonal field method

They dynamical orthogonal field equation can be derived following Sapsis and Lermusiaux (2009). Starting from a general SDE of the form:

$$\boxed{\frac{\partial \underline{u}(\underline{x}, t, \omega)}{\partial t} = \underline{\mathcal{L}}[\underline{u}(\underline{x}, t, \omega), \omega]}$$

where:

$$(\underline{x}, t, \omega) \in \mathcal{D} \times \mathcal{T} \times \Omega$$

We define the mean of the field $\underline{u}$ as:

$$\overline{u}(\underline{x}, t) = \mathbb{E}^{\omega}[\underline{u}(\underline{x}, t, \omega)] = \int_{(\Omega)} \underline{u}(\underline{x}, t, \omega) d\mathbb{P}(\omega)$$

where $\mathbb{P}$ is a probability measure on ω. From the covariance matrix:

$$\underline{\underline{C_{\underline{u}(\cdot,t)\underline{u}(\cdot,t)}}}(\underline{x}, \underline{y}) = \mathbb{E}^{\omega}[(\underline{u}(\underline{x}, t, \omega) - \overline{\underline{u}}(x, t))(\underline{u}(\underline{y}, t, \omega) - \overline{\underline{u}}(\underline{y}, t))^T]$$

we can define the integral operator:

$$\underline{\underline{\mathcal{T}_C}}(\varphi) = \int_{(\mathcal{D})} \underline{\underline{C_{\underline{u}(\cdot,t)\underline{u}(\cdot,t)}}}(\underline{x}, \underline{y}) \cdot \underline{\varphi}(\underline{x}, t) d\underline{x}$$

and prove that it is compact, self-adjoint and positive. It follows that any random field $\underline{u}$ has a Karhuven-Loeve expansion at a given time t:

$$\underline{u}(\underline{x}, t, \omega) = \overline{\underline{u}}(\underline{x}, t) + \sum_{i=1}^{+\infty} Y_i(t, \omega)\underline{u_i}(\underline{x}, t)$$

where:

- $\underline{u_i}(\underline{x},t)$ are the eigenfunctions of $\underline{\underline{\mathcal{T}_C}}$
- $Y_i(t,\omega)$ are zero-mean stochastic processes with variance $\mathbb{E}^\omega[Y_i^2(t,\omega)]$ which are eigenvalues of the following eigenvalue problem:

$$\underline{\underline{\mathcal{T}_C}}[\underline{u_i}(\underline{x},t)] = \mathbb{E}^\omega[Y_i^2(t,\omega)]\underline{u_i}(\underline{y},t)$$

In many physical problems of interest (including ours), $\mathbb{E}^\omega[Y_i^2(t,\omega)] \sim \exp(-ci)$ where $c>0$, so that to a good approximation, we can truncate the Karhuven-Loeve expansion to a finite number of terms:

$$\boxed{\underline{\overline{u}}(\underline{x},t,\omega) = \underline{\overline{u}}(\underline{x},t) + \sum_{i=1}^{s} Y_i(t,\omega)\underline{u_i}(\underline{x},t)}$$

We can see that the variation of the stochastic coefficients Y_i can express exclusively the evolution of the uncertainty within the stochastic space $V_S = span(\underline{u_i}|i \in [1,s])$. However, the evolution of the stochastic basis $\underline{u_i}$ itself allows the uncertainty to cover V_S and $V_S^\perp$. To avoid redundancy in the evolution of the uncertainty, we impose that the evolution of the basis $\underline{u_i}$ stays in $V_S^\perp$, ie:

$$\frac{dV_S}{dt} \perp V_S \;\Leftrightarrow < \frac{\partial}{\partial t}\underline{u_i}|\underline{u_j} >= 0$$

This is the dynamically orthogonal condition (DO condition). We can now derive the DO field equations by inserting the DO representation in the initial evolution equation:

$$\frac{\partial}{\partial t}\underline{\overline{u}} + \frac{dY_i}{dt}\underline{u_i} + Y_i\frac{\partial}{\partial t}\underline{u_i} = \underline{\mathcal{L}}[\underline{u}(\underline{x},t,\omega),\omega]$$

Applying $\mathbb{E}^\omega$ to the previous equation, we obtain an evolution equation for the mean part of the representation:

$$\boxed{\frac{\partial}{\partial t}\underline{\overline{u}} = \mathbb{E}^\omega\{\mathcal{L}[\underline{u}(\underline{x},t,\omega),\omega]\}}$$

Taking the inner product of the evolution equation with $\underline{u_j}$, applying the orthonormality condition of the $\underline{u_i}$, and the DO condition, we obtain:

$$\frac{dY_j}{dt} =< \frac{\partial}{\partial t}\underline{\overline{u}}|\underline{u_j} >=< \underline{\mathcal{L}}(\underline{u})|\underline{u_j} >$$

Applying $\mathbb{E}^\omega$ to the previous equation and using the evolution equation for the mean part of the representation, it is possible do derive an equation for the zero-mean stochastic processes Y_i:

$$\boxed{\frac{dY_i}{dt} =< \underline{\mathcal{L}}[\underline{u},\omega] - \mathbb{E}^\omega\{\underline{\mathcal{L}}[\underline{u},\omega]\}|\underline{u_i} >}$$

Finally, it is also possible to derive an equation for the basis vectors:

$$\boxed{\frac{\partial}{\partial t}\underline{u_i} = \underline{\Pi_{V_S^\perp}}[\mathbb{E}^\omega\{\underline{\mathcal{L}}[\underline{u},\omega]Y_j\}] \cdot \underline{\underline{C^{-1}_{Y_i(t)Y_j(t)}}}}$$

where the operator $\underline{\Pi_{V_S^\perp}}$ is defined as:

$$\underline{\Pi_{V_S^\perp}}[\underline{F}(\underline{x})] = \underline{F}(\underline{x}) - < \underline{F}(\underline{x})|\underline{u_k}(\underline{x},t) > \underline{u_k}(\underline{x},t)$$

2.5 Application to the quasi-geostrophic barotropic model

We come back to our BT QG model and apply the DO analysis following Sapsis and Dijkstra (2013). We first expand the horizontal velocity and the pressure:

$$\underline{u}(\underline{x},t,\omega) = \underline{\overline{u}}(\underline{x},t) + \sum_{i=1}^{s} Y_i(t,\omega)\underline{u_i}(\underline{x},t)$$

$$p = p_0 + Y_i p_i - Y_i Y_j p_{ij} + Z_r b_r$$

where Z_r are the coefficients of the noisy part of the wind:

$$\sigma\tau_a(x,y,t) = \sum_{k=1}^{s} Z_k(t,\omega)\sigma_k(\underline{x},t)$$

The DO mean equations can be written:

$$\boxed{\frac{\partial \underline{\overline{u}}}{\partial t} = -\underline{\nabla p_0} + \frac{\nabla^2 \underline{\overline{u}}}{Re} - (\underline{\overline{u}}\cdot\underline{\nabla})\underline{\overline{u}} - f\underline{k}\wedge\underline{\overline{u}} + \underline{\tau_d}(\underline{x},t) - \underline{\underline{C_{Y_iY_j}}}\cdot[-\underline{\nabla p_{ij}} + \frac{1}{2}(\underline{u_i}\cdot\underline{\nabla})\underline{u_j} + \frac{1}{2}(\underline{u_j}\cdot\underline{\nabla})\underline{u_i}]}$$

$$0 = \underline{\nabla}\cdot\underline{\overline{u}}$$

Projecting on the DO modes, it is possible to obtain a solvable system of (s+1) PDEs. The stochastic wind stress forcing follows the bulk formula for the momentum flux:

$$\underline{\tau_{stochastic}} = \rho_{air} C_D |\underline{u}'|\underline{u}'$$

where ρ_{air} is the air density, $\underline{u}'$ the near-surface wind's velocity and C_D the drag coefficient. The near-surface wind is taken to be stochastic:

$$\underline{u}' = f(x,y)\underline{\eta}(t)$$

where $\underline{\eta}(t)$ is a white/colored noise vector depending on the experiment, with mean 0 and variance σ. The weight function f parametrizes the spatial structure of the atmospheric variability with a Gaussian shape, whose origin is placed at the center of the basin:

$$f(x,y) = \alpha[\pi\lambda_x\lambda_y \mathrm{erf}(\frac{L_x}{2\lambda_x})\mathrm{erf}(\frac{L_y}{2\lambda_y})]^{-\frac{1}{2}} \exp(\frac{x^2}{2\lambda_x^2} + \frac{y^2}{2\lambda_y^2})$$

Looking at the resulting 3D contours of the PDF, we can see that the main effect of the noise is to allow for multiple equillibria of the system. Looking at the effect of the noise in more detail, we can see significant differences if the noise is chosen to be white or colored:

- If we choose a white-noise excitation $\underline{\eta}(t) \propto \underline{dW}(t)$ where $\underline{W}$ is a vectorial Wiener process, then the stochastic excitation has zero effect on the instantaneous evolution of the mean field, and the shape of the stochastic subspace V_S. As a consequence, it will not influence the general statistics of the double-gyre flow.

- More generally, we can choose a colored noise excitation, for instance through the Ornstein-Uhlenbeck process:

$$\tau \underline{d\eta}(t) = -\underline{\eta}(t) + \sqrt{2\tau}\underline{dW}(t)$$

where τ is the decorrelation time scale ($\tau \gg 1$ corresponds to the deterministic case and $\tau \ll 1$ to the white-noise case). This does not only allow to internally transfer energy between the DO modes, but it also allows the stochastic modes to directly absorb energy from the stochastic forcing. Depending on the value of τ, colored noise can either destabilize some DO modes and push the system in a statistically steady regime (small memory $\Leftrightarrow$ small τ), or reduce the complexity of the system by bringing it to a single unstable mode (long memory $\Leftrightarrow$ large τ).

3 Conclusion

Is summary, it is possible to add noise to a chaotic system and still solve for a low order equivalent system. We can see that the noise adds a lot of variability, helping us explore new flavors of the climatic system. To study the effect of the noise on a PDE system with more precision, we have introduced the method of the dynamical-orthogonal field. Under certain condition, this method reduces the analysis of a general continuous stochastic field to a finite number of orthonormal mode, which define a stochastic subspace where the solution lives. In the special case of the BT QG model , we have seen that white-noise has a minimal effect on the reduced dynamics, whereas colored noise can significantly change the behavior of the system.

The Diffusion Fish

Gunnar Peng

October 15, 2015

1 Introduction

The study of stratified fluids, i.e. bodies of e.g. water or air in which density varies with height, is of great importance for the understanding of the oceans and the atmosphere. As is well known, a system with dense fluid overlying light fluid is unstable, and if left to evolve will spontaneously generate a flow that interchanges and mixes the fluids until a stable stratification is reached, where the density decreases with increasing height. A less well-known fact is that flow can also be spontaneously generated in a stably stratified fluid, and that is the focus of this report. The spontaneous flow is called diffusion-driven flow, and is due to a combination of diffusion and buoyancy effects.

For simplicity, in this report we consider a liquid whose density varies due to a varying concentration of salt, although the discussion applies to thermal stratifications as well. (If the stratification is due to a combination of multiple salts and/or heat with different diffusive properties, then double-diffusive effects may come into play and we do not consider this case here.) In a stable stratification, salt will spontaneously diffuse up from the salty and dense fluid below to the fresh and light fluid above. This raises the centre of gravity of the fluid, and hence increases the potential energy of the system (at the expense of thermodynamic free energy), which can then be harnessed to drive a flow.

The classical example of diffusion-driven flow due to the presence of sloping insulating boundaries was described by Phillips [13] and Wunsch [18] in 1970 (see figure 1(a)). We assume that the stable background stratification is uniform, so that salt diffuses upward at a constant rate. Thus, away from the wall, at each point (e.g. A in the figure) the amount of salt leaving upward due to diffusion equals the amount of salt arriving from below, and hence the concentration does not change. Just above the wall (e.g. B in the figure), however, the salt that diffuses up and away is not replenished from below due to the presence of the insulating wall, and hence (in the absence of flow) the fluid near the wall becomes lighter. Put another way, the pycnoclines (surfaces of constant density) must meet the insulating wall at right angles, and do so by bending down, as shown in the figure, resulting in the fluid near the wall becoming lighter.

The lighter fluid wants to rise (relative to the ambient fluid away from the wall), and hence flows up along the boundary in a "buoyancy layer". A steady flow is achieved with the velocity profile shown, and the salt leaving point B by diffusion is continually replenished by advection from below (the dashed arrow). Similarly, fluid near a sloping upper boundary becomes heavy, and hence flows down the slope. This buoyancy-layer flow is called Phillips–Wunsch flow, and will be discussed more quantitatively in §2.2.

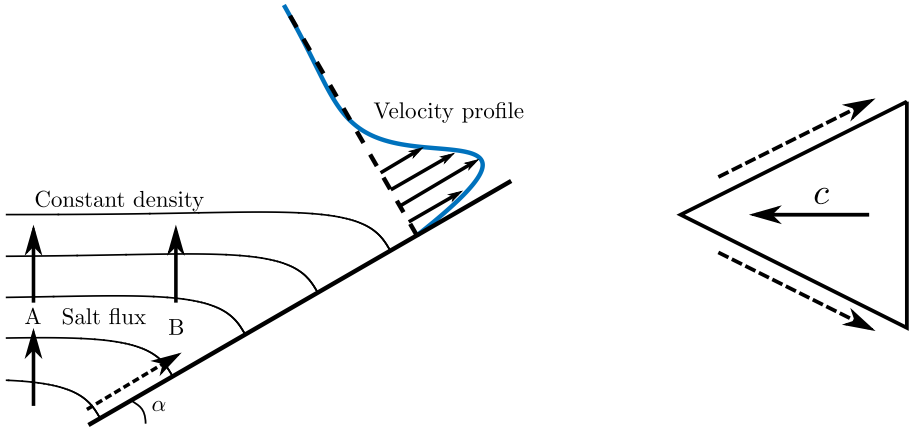

Figure 1: (a) Schematic of diffusion-driven (Phillips–Wunsch) flow near a sloping insulating boundary in stably stratified fluid (adapted from [13, 12]). (b) Schematic of diffusion-driven propulsion. The Phillips–Wunsch flow (dashed arrows) along the sloping surfaces propels the wedge in the opposite direction.

More recent studies of Phillips–Wunsch flow include an experimental verification by Peacock, Stocker and Aristoff in 2004 [12] and a few years later theoretical and numerical studies of the diffusion-driven flow inside containers with sloping walls by Page & Johnson [10, 11] and Page [8, 9].

Around the same time as the original papers by Phillips and Wunsch, the motion of bodies through stably stratified fluids was extensively studied. The existence of an exact analogy between the governing equations for two-dimensional viscous stratified flow and two-dimensional viscous rotating flow (see e.g. Veronis [15, 16, 17]) led to parallel developments in both fields, such as the calculation of the drag on a body moving horizontally in stratified fluid (Foster & Saffman [2]) and axially in a rotating fluid (Moore & Saffman [6, 7]).

Although the Phillips–Wunsch flow on sloping boundaries and the motion of insulating bodies through stratified fluid were thoroughly studied around 1970, these two ideas where not combined until 2010, when Allshouse, Barad and Peacock [1] showed that the Phillips–Wunsch flow could be harnessed by asymmetric bodies for horizontal propulsion. They placed a wedge in stably stratified fluid and found that it moves (see figure 1(b)) at a constant speed c which depends on the parameters of the problem. No studies so far have given a theoretical explanation for the propulsion or prediction for the propulsion speed. We will do so for a variety of two-dimensional cases in this report.

In §2, we describe the assumptions and approximations made and derive the main governing equations. We then investigate three main cases, where the wedge is placed in either a box with insulating walls (§3), a box with fixed-buoyancy walls (§4), or a large or infinite box (§5). Finally, we summarize our results in §6 and discuss possible extensions to this work.

2 Problem Setup

We consider a fluid with constant dynamic viscosity μ, whose density ρ depends on the salinity $S(\boldsymbol{x}, t)$. We assume that the variations in S about a constant reference value S_0 are sufficiently small that the equation of state can be linearized, i.e.

$$\rho = \rho_0(1 + \beta_S(S - S_0)), \tag{2.1}$$

where β_S is the coefficient of saline contraction and ρ_0 is the reference value for ρ. The (upward) buoyancy force can then be written as

$$-\rho g = \rho_0(-g + B), \tag{2.2}$$

where g is the gravitational acceleration and we have defined the buoyancy (or reduced gravity) by

$$B = g\frac{\rho - \rho_0}{\rho_0} = g\beta_S(S_0 - S). \tag{2.3}$$

The salt is advected by the velocity field $\boldsymbol{u}(\boldsymbol{x}, t)$ and diffuses with diffusivity κ, leading to the equation

$$\dot{S} + \boldsymbol{u}\cdot\boldsymbol{\nabla}S = \kappa\nabla^2 S \qquad \Rightarrow \qquad \dot{B} + \boldsymbol{u}\cdot\boldsymbol{\nabla}B = \kappa\nabla^2 B, \tag{2.4a,b}$$

where overdot denotes the time derivative and $\boldsymbol{\nabla}$ is the gradient operator.

We further assume that the Boussinesq approximation holds, so we can ignore any effects due to density variations, apart from the buoyancy force. We thus obtain the incompressible Navier–Stokes equations

$$\rho_0\left(\dot{\boldsymbol{u}} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}\right) = -\boldsymbol{\nabla}P + \mu\nabla^2\boldsymbol{u} + \rho_0(-g + B)\,\boldsymbol{e}_z, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{u} = 0, \tag{2.5a,b}$$

where $P(\boldsymbol{x}, t)$ is the pressure and $\boldsymbol{e}_z$ is the unit vector in the vertical direction. As the kinematic viscosity $\nu = \mu/\rho_0$ of the salt solution is much larger than the diffusivity κ of the salt (the Schmidt number is typically $\nu/\kappa \sim 10^3$), the inertial terms (left-hand side) of equation (2.5a) can be neglected for diffusion-driven flow. (We confirm this claim in §2.2.) We are left with the Stokes equations

$$\boldsymbol{0} = -\boldsymbol{\nabla}(P/\rho_0 + gz) + \nu\nabla^2\boldsymbol{u} + B\,\boldsymbol{e}_z, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{u} = 0. \tag{2.6a,b}$$

We will solve the main governing equations (2.4b, 2.6) in a domain between the wedge and an outer bounding box, and require boundary conditions on $\boldsymbol{u}$ and B. As the boundaries are rigid, the fluid satisfies the no-slip condition

$$\boldsymbol{u} = \boldsymbol{U}_b, \tag{2.7}$$

where $\boldsymbol{U}_b$ is the velocity of the boundary. For the buoyancy, we consider either insulating conditions or fixed-buoyancy conditions

$$B_n = 0 \qquad \text{or} \qquad B = B_b, \tag{2.8}$$

where the subscript n denotes the derivative in the normal direction pointing into the fluid and B_b is the prescribed value. We specify the detailed geometry and choice of condition (2.8) in each separate section.

2.1 Two-dimensional governing equations

We consider the case when the wedge is sufficiently wide (in the y-direction) that the flow can be approximated as being two-dimensional in the x- and z-directions with no variation in the y-direction. This allows introduction of the stream function $\psi(x,z,t)$, defined by $\boldsymbol{u} = (u,w) = (\psi_z, -\psi_x)$, where subscripts denote differentiation and we omit the y-component of any vector. Incompressibility (2.6b) is then automatically satisfied, while taking the curl of the momentum equation (2.6a) yields the vorticity equation

$$B_x = \nu\nabla^4\psi, \tag{2.9a}$$

which describes a balance between the generation of vorticity $\nabla^2\psi$ by horizontal variations in buoyancy and the dissipation of vorticity by viscous effects. The advection–diffusion equation takes the form

$$\dot{B} + \psi_z B_x - \psi_x B_z = \kappa\nabla^2 B. \tag{2.9b}$$

We consider mainly the case when the buoyancy field has a uniform background stratification $N^2 z$, and the perturbations $b = B - N^2 z$ to the buoyancy field are small compared with the background. The gradient N^2 of the background field is the square of the Brunt–Väisälä (buoyancy) frequency. After the change of variables from B to b, the governing equations (2.9) become

$$b_x = \nu\nabla^4\psi, \qquad \dot{b} + \psi_z b_x - \psi_x b_z - N^2\psi_x = \kappa\nabla^2 b. \tag{2.10a,b}$$

Typically, the term $-N^2\psi_x$ in (2.10b), which expresses advection of the background stratification N^2 by the vertical velocity $-\psi_x$, dominates the remaining terms on the left-hand side. In this case, we can eliminate either ψ or b from (2.10) to obtain

$$b_{xx} + \frac{\kappa\nu}{N^2}\nabla^6 b = 0 \qquad \text{or} \qquad \psi_{xx} + \frac{\kappa\nu}{N^2}\nabla^6\psi = 0. \tag{2.11}$$

This reveals an inherent length scale

$$L_0 = \left(\frac{\kappa\nu}{N^2}\right)^{1/4} \tag{2.12}$$

of the governing equations. The equations (2.10) or (2.11) are typically too complicated to solve exactly, so we will restrict ourselves to cases where the other length scales of the problem (such as the size of the wedge and the size of the domain) are much larger than L_0, and use the method of matched asymptotic expansions to obtain approximate analytical solutions.

Although equations (2.10) suffice to describe the flow and buoyancy fields, we will need the pressure in order to calculate the force on the wedge. Hence, we retain the momentum equation (2.6a) as well, in the form

$$p_x = \nu(\psi_{zxx} + \psi_{zzz}), \qquad p_z = -\nu(\psi_{xxx} + \psi_{xzz}) + b, \tag{2.13}$$

where have defined the rescaled pressure perturbation $p = P/\rho_0 + gz - N^2 z^2/2$.

2.2 Solutions near non-horizontal walls

As discussed in the introduction, diffusion-driven flow is generated by the interaction of stratified fluid with sloping boundaries. We investigate the details of this in more detail, by considering an infinite wall with angle α to the horizontal placed in a stratified fluid with linear ambient stratification $B = N^2 z$ and no ambient flow (see figure 1(a)). The solutions obtained here will serve as boundary-layer solutions for later calculations, and so the aim in particular is to understand the far-field behaviour of these solutions.

We seek steady solutions for the buoyancy perturbation b and the stream function ψ which depend only on the distance η from the wall, and hence are independent of the distance along the wall. The governing equations (2.10) reduce to

$$-\sin\alpha\, b_\eta = \nu\, \psi_{\eta\eta\eta\eta}, \qquad N^2 \sin\alpha\, \psi_\eta = \kappa b_{\eta\eta}. \tag{2.14}$$

As the wall is stationary, the no-slip condition (2.7) yields

$$\psi = 0, \qquad \psi_\eta = 0 \qquad \text{at } \eta = 0, \tag{2.15}$$

where we have chosen the arbitrary additive constant for the stream function such that $\psi = 0$ on the wall.

If the wall is insulating (which is the case for Phillips–Wunsch flow), then the condition (2.8) for no perpendicular gradient of total buoyancy $B = N^2 z + b$ yields

$$b_\eta = -N^2 \cos\alpha \qquad \text{at } \eta = 0. \tag{2.16}$$

The solution to equations (2.14, 2.15, 2.16) that does not grow exponentially as $\eta \to \infty$ is

$$\psi = \kappa \cot\alpha \, \left[1 - (\cos\gamma\eta + \sin\gamma\eta)\, e^{-\gamma\eta}\right], \qquad b = N^2 \frac{\cos\alpha}{\gamma} \cos\gamma\eta\, e^{-\gamma\eta} + b_\infty, \tag{2.17}$$

where b_∞ is a constant of integration and

$$\gamma^{-1} = \left(\frac{4\kappa\nu}{N^2 \sin^2\alpha}\right)^{1/4} = \sqrt{\frac{2}{\sin\alpha}} L_0 \tag{2.18}$$

is the length scale on which the flow decays away from the wall. We conclude that there is a flow confined to a boundary layer of thickness $O(L_0)$ near the wall, and the net flux of fluid up the slope is the far-field value $\kappa\cot\alpha$ of the stream function ψ. Although in this case we would set the constant of integration b_∞ to be zero, to recover the ambient stratification with $b \to 0$ as $\eta \to \infty$, we note that this is in general not necessary and the far-field buoyancy perturbation could have a non-zero value.

We note that the Phillips–Wunsch flux $\kappa\cot\alpha$ is zero for vertical walls ($\alpha = \pi/2$). In this case, the solution is in fact trivial ($\psi = 0$, $b = b_\infty$) and there is no boundary-layer flow. As the slope α decreases, the flux increases and eventually diverges to infinity as α approaches zero and the wall becomes horizontal. We thus assume for the rest of this report that α is not too small, so that $\cot\alpha = O(1)$. In our analyses that follow, the velocity is largest in the Phillips–Wunsch boundary layer, so the Reynolds number is everywhere smaller than the local estimate $\kappa\cot\alpha/\nu$. Since $\nu \gg \kappa$, the use of the Stokes equations (2.6) is indeed appropriate.

If instead the wall has prescribed buoyancy (i.e. salinity) $B = N^2 z$, then

$$b = 0 \qquad \text{at } \eta = 0, \tag{2.19}$$

and the solution to (2.14, 2.15, 2.19) that does not grow exponentially as $\eta \to \infty$ is

$$\psi = -\frac{\kappa b_\infty}{N^2} \frac{\gamma}{\sin\alpha} \left[1 - (\cos\gamma\eta + \sin\gamma\eta)\, e^{-\gamma\eta}\right], \qquad b = b_\infty \left[1 - \cos\gamma\eta\, e^{-\gamma\eta}\right] \tag{2.20}$$

This result states that the difference b_∞ in buoyancy between the far field and the wall drives a flux proportional to b_∞ in an $O(L_0)$ boundary layer near the wall.

3 Wedge in a Box with Insulating Walls

We first consider the case when the wedge is placed in a box with insulating walls. On the top and bottom of the box, we assume that the buoyancy is held fixed at two different constant values, which would generate a uniform stable stratification $B = N^2 z$ in the absence of a wedge.

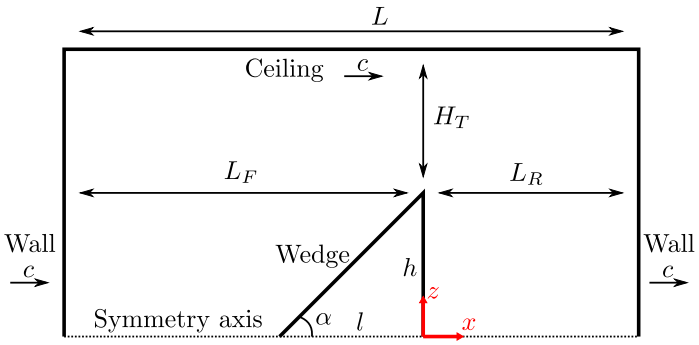

Figure 2: Schematic geometry for a wedge placed inside a box.

The wedge is placed with its apex pointing to the left (see figure 2), which turns out to be its direction of motion. We work in the frame of reference moving with the wedge, with the origin $(0, 0)$ at the midpoint of the back of the wedge at all times. We assume that the level of neutral buoyancy of the wedge is vertically centred in the box, and that the initial conditions are symmetric about $z = 0$. Hence, the top-down symmetry will be preserved throughout the evolution, and we need only consider the upper half $z \geq 0$ of the system.

The wedge has length l and (half-)height h, so that its corners are at $(-l, 0)$ and $(0, h)$. Its apex half-angle is α, so that

$$\cot\alpha = \frac{l}{h}. \tag{3.1}$$

The box has length $L = L_F + L_R$, where $L_F(t)$ and $L_R(t)$ are the distances from the front (left) wall and the rear (right) wall to the base of the wedge and evolve according to

$$c = -\dot{L}_F = \dot{L}_R, \tag{3.2}$$

where $c(t)$ is the leftward speed of the wedge (and hence, in this reference frame, the rightward speed of the bounding box). The height of (the upper half of) the box is $h + H_T$.

We assume that the wedge has a length scale L_1 which is much greater than the natural scale L_0 of the flow given in (2.12). (In the experiments [1], typically $L_1 = O(10\text{ mm})$ and $L_0 = O(0.1\text{ mm})$.) We thus obtain a small parameter

$$\varepsilon = L_0/L_1 \tag{3.3}$$

that we can exploit using the method of matched asymptotic expansions. We also take the height H_T of the box to be $O(L_1)$, but in this section we assume that the horizontal dimensions L_F and L_R of the box are $O(\varepsilon^{-1}L_1)$, i.e. much larger than the dimensions of the wedge.

At every instant in time, we seek to calculate the flow field (via the stream function $\psi(x,z,t)$), the buoyancy distribution $b(x,z,t)$, and the net "propulsive" left-ward force $F(t)$ on (the upper half of) the wedge as a function of the unknown speed $c(t)$ of the wedge. Requiring that $F = 0$ then determines the actual value $c(t)$.

We non-dimensionalize the variables as follows. Length is scaled by L_1, except for L_F and L_R which are scaled by $\varepsilon^{-1}L_1$. Given that the flow is driven by the Phillips–Wunsch flux $\kappa\cot\alpha$ (see §2.2), we scale the stream function by κ and velocity by κ/L_1. We scale time by by $\varepsilon^{-1}L_1^2/\kappa$, as this is the time scale of the overall evolution due to the horizontal motion of the wedge. (We neglect the fast initial adjustment at the start of the experiment.) Buoyancy perturbations are scaled by $\varepsilon N^2 L_1$, for reasons which will become clear later (and thus we scale the pressure and force by $\varepsilon N^2 L_1^2$ and $\varepsilon N^2 L_1^3$ respectively). The rescaled governing equations (2.10) are thus

$$b_x = \varepsilon^3\nabla^4\psi, \qquad \varepsilon^2\dot{b} + \varepsilon(\psi_z b_x - \psi_x b_z) - \psi_x = \varepsilon\nabla^2 b. \tag{3.4a,b}$$

We choose the stream function ψ to be zero on the centreline $z = 0$. Hence, ψ (as well as b) are odd functions of z, and we obtain the symmetry conditions

$$\psi = \psi_{zz} = 0, \qquad b = 0 \qquad \text{on symmetry axis} \qquad z = 0. \tag{3.5a}$$

The boundary conditions on the insulating wedge (analogous to (2.16)) are

$$\psi = \psi_n = 0, \quad b_n = -\tfrac{1}{\varepsilon}\cos\alpha \quad \text{on wedge front} \quad x = -l + z\cot\alpha, \tag{3.5b}$$
$$\psi = \psi_n = 0, \quad b_n = 0 \quad \text{on wedge rear} \quad x = 0. \tag{3.5c}$$

On the bounding box, which moves to the right at speed $c(t)$, we have the no-slip condition and insulating or prescribed-buoyancy conditions:

$$\psi = cz, \quad \psi_x = 0, \quad b_x = 0 \quad \text{on walls} \quad x = -\varepsilon^{-1}L_F,\ \varepsilon^{-1}L_R, \tag{3.5d}$$
$$\psi = cz, \quad \psi_z = c, \quad b = 0 \quad \text{on ceiling} \quad z = h + H_T. \tag{3.5e}$$

For a given value of $c(t)$, ψ and b are determined uniquely by the equations (3.4) and boundary conditions (3.5). Equation (2.13), which in non-dimensional form is

$$p_x = \varepsilon^3(\psi_{zxx} + \psi_{zzz}), \qquad p_z = -\varepsilon^3(\psi_{xxx} + \psi_{xzz}) + b, \tag{3.6a,b}$$

then determines the pressure up to an arbitrary additive constant, which we can choose without loss of generality such that $p = 0$ at $(x, z) = (0, h + H_T)$. The horizontal leftward force on the wedge is given by

$$F(t) = \int_0^h \left[(-p + \varepsilon^3\, 2\psi_{xz}) + \varepsilon^3(\psi_{xx} - \psi_{zz})\cot\alpha)\right]_{x=-l+z\cot\alpha} + \left[p - \varepsilon^3\, 2\psi_{xz}\right]_{x=0} dz, \quad (3.7)$$

and the correct value of $c(t)$ is the one that makes $F(t)$ vanish.

3.1 Numerical results using the finite-element method

In order to inform, and later validate, our asymptotic analysis, we developed a simple code in FreeFem++ [4] that solves the governing equations (3.4, 3.5). However, for simplicity we do not simulate the full time-evolution of the system with moving boundaries. Instead, we neglect the time derivative $\dot{b}$ in (3.4b), as it is of higher order than the other terms. This quasi-static approximation leaves us with an instantaneous problem with no time derivatives, that is readily solved by our program.

For our numerical calculations, we use the following parameters:

$$\varepsilon = \frac{L_0}{L_1} = \frac{1}{100}, \quad h = H_T = 1, \quad L = 1, \quad (L_F, L_R) = \left(\frac{1}{4}, \frac{3}{4}\right), \left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{3}{4}, \frac{1}{4}\right), \quad (3.8)$$

with a focus on the symmetric case $L_F = L_R = 1/2$ with $\alpha = 45°$. Figure 3 shows results for this particular case.

We observe that a Phillips–Wunsch boundary layer develops on the sloping surface of the wedge, and that fluid arrives in a boundary layer near $z = 0$ in front of the wedge (to the left), and leaves in a boundary layer near $z = h$ behind the wedge (to the right). The buoyancy perturbations are approximately uniform in x both in front of, behind and above the wedge. As we shall see, the buoyancy field is key to calculating the force on the wedge, and we will show more detailed results later in §3.3.

3.2 Asymptotic calculation

We divide the domain into the regions shown schematically in figure 4(b). The bulk of the fluid is divided into three "outer" regions which we call the front region, the rear region, and the top region. These regions are joined by horizontal boundary layers at $z = h$, and in addition there are boundary layers at $z = 0$.

3.2.1 Boundary layers on the insulating surfaces

Near the sloping surface of the wedge (but away from the corners), we expect to find a boundary layer with the Phillips–Wunsch solution discussed in §2.2. Indeed, if we define rescaled coordinates (η, χ) perpendicular and parallel to the slope by

$$x = -l + \chi\cos\alpha - \varepsilon\eta\sin\alpha, \qquad z = \chi\sin\alpha + \varepsilon\eta\cos\alpha, \quad (3.9)$$

then the governing equations (3.4) and boundary conditions (3.5b) on the wedge become

$$-\sin\alpha\, b_\eta + \varepsilon\cos\alpha\, b_\chi = b_{\eta\eta\eta\eta} + \varepsilon^2\, 2b_{\eta\eta\chi\chi} + \varepsilon^4\, b_{\chi\chi\chi\chi}, \quad (3.10a)$$

$$\varepsilon^3\dot{b} + \varepsilon(\psi_\eta b_\chi - \psi_\chi b_\eta) + \sin\alpha\,\psi_\eta - \varepsilon\cos\alpha\,\psi_\chi = b_{\eta\eta} + \varepsilon^2 b_{\chi\chi}, \quad (3.10b)$$

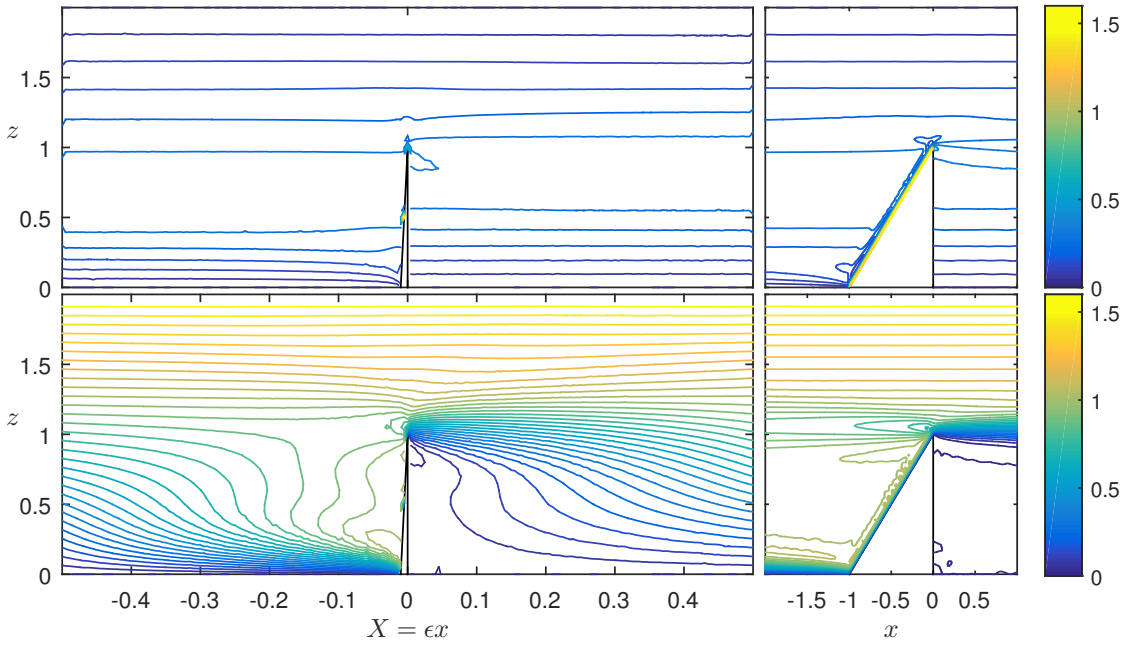

Figure 3: Contour plots of buoyancy perturbation b (top row) and stream function ψ (bottom row, flow from left to right in the frame of reference of the wedge) showing the whole box with stretched coordinates (left) and a close-up of the wedge (right), for the parameters (3.8), $L_F = L_R = 1/2$, $\alpha = 45°$. The contour spacing is 0.05. (The irregularities in the contours are artefacts of the plotting tool and are not present in the original data.)

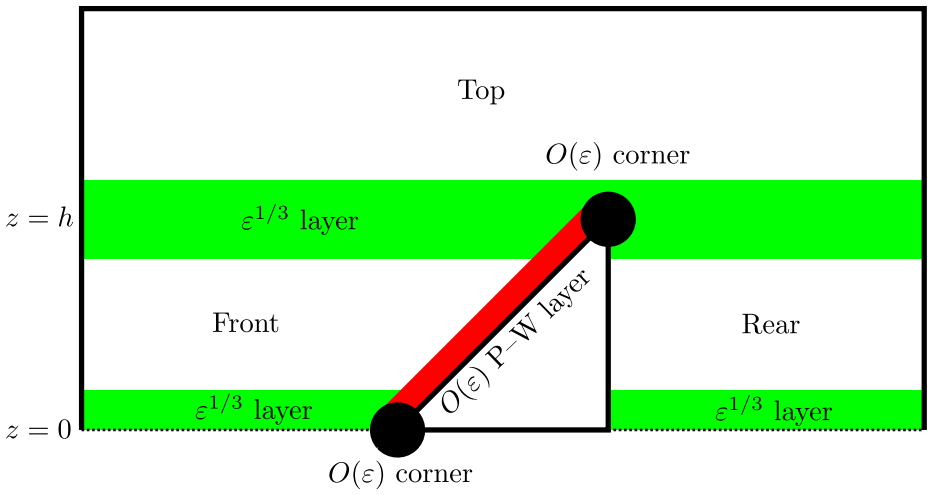

Figure 4: Schematic of asymptotic regions for a wedge in a box with insulating walls.

$$\psi = \psi_\eta = 0, \quad b_\eta = -\cos\alpha \qquad \text{on the wedge } \eta = 0. \tag{3.10c}$$

These equations are analogous to the equations (2.14, 2.15, 2.16) for Phillips–Wusch flow at leading order, and hence the solution (which does not grow exponentially as $\eta \to \infty$) is

$$\psi = \cot\alpha \left[1 - (\cos\gamma\eta + \sin\gamma\eta)e^{-\gamma\eta}\right] + O(\varepsilon), \tag{3.11a}$$

$$b = \cot\alpha \sqrt{2\sin\alpha}\, \cos\gamma\eta\, e^{-\gamma\eta} + b_\infty(\chi) + O(\varepsilon), \tag{3.11b}$$

where $\gamma = \sqrt{\sin\alpha/2}$, analogously to (2.17).

The far-field behaviour of the solution (3.11) is to be matched to the outer solution for the front region. We find that

$$\psi \to \cot\alpha + O(\varepsilon), \quad b \to b_\infty(\chi) + O(\varepsilon) \qquad \text{as } \eta \to \infty. \tag{3.12}$$

For the buoyancy field, the matching determines the constant of integration $b_\infty(\chi)$ to be the effective value $b^F(-l + \chi\cos\alpha, \chi\sin\alpha)$ of the outer buoyancy field b^F on the wedge, but imposes no constraint on b^F. However, the outer stream function ψ^F ahead of the wedge must satisfy the effective condition

$$\psi^F = \cot\alpha + O(\varepsilon) \qquad \text{on the wedge } x = -l + z\cot\alpha. \tag{3.13a}$$

Since the back of the wedge is vertical, the corresponding boundary layer there is trivial at leading order (as discussed in §2.2). Hence the boundary condition (3.5c) for the stream function on the rear of the wedge becomes the effective condition on the leading-order stream function ψ_0^R in the rear region:

$$\psi^R = 0 + O(\varepsilon) \qquad \text{on the wedge } x = 0. \tag{3.13b}$$

The insulating front and rear wall behave like the insulating back of the wedge, and hence also have trivial boundary layers. The conditions (3.5d) on the stream function then translate to effective conditions on the outer solutions, including ψ^T for the top region:

$$\psi^F = cz + O(\varepsilon) \qquad \text{at } x = -\varepsilon^{-1}L_F, \qquad \psi^R = cz + O(\varepsilon) \qquad \text{at } x = \varepsilon^{-1}L_R, \tag{3.13c}$$

$$\psi^T = cz + O(\varepsilon) \qquad \text{at } x = -\varepsilon^{-1}L_F, \qquad \psi^T = cz + O(\varepsilon) \qquad \text{at } x = \varepsilon^{-1}L_R. \tag{3.13d}$$

3.2.2 Outer solution

Using the conditions (3.13), we can now start obtaining solutions for the outer regions. As these regions have length $O(\varepsilon^{-1})$, we introduce a scaled horizontal coordinate $X = \varepsilon x$. The governing equations (3.4) become

$$b_X = \varepsilon^2(\psi_{zzzz} + \varepsilon^2\psi_{XXzz} + \varepsilon^4\psi_{XXXX}), \tag{3.14a}$$

$$\varepsilon(\dot{b} + \psi_z b_X - \psi_X b_z) - \psi_X = b_{zz} + \varepsilon^2 b_{XX}. \tag{3.14b}$$

At leading order, these equations simplify to

$$b_X = 0 + O(\varepsilon), \qquad -\psi_X = b_{zz} + O(\varepsilon). \tag{3.15a,b}$$

Equation (3.15a) describes how the flow is too weak to sustain horizontal variations in buoyancy, and thus we obtain $b = b(z,t) + O(\varepsilon)$ in every outer region. Equation (3.15b)

describes how the vertical diffusion of buoyancy is balanced by advection of the background buoyancy field by the vertical velocity $-\psi_X$. By integrating this equation in X, and applying the conditions (3.13), we obtain equations describing the balance of vertical flux in each outer region:

$$L_F b^F_{zz} = \psi^F|_{X=-L_F} - \psi^F|_{X=O(\varepsilon)} + O(\varepsilon) = cz - \cot\alpha + O(\varepsilon), \tag{3.16a}$$

$$L_R b^R_{zz} = \psi^R|_{X=0} \quad - \psi^R|_{X=L_R} + O(\varepsilon) = -cz \quad + O(\varepsilon), \tag{3.16b}$$

$$L b^T_{zz} = \psi^T|_{X=-L_F} - \psi^T|_{X=L_R} + O(\varepsilon) = 0 \quad + O(\varepsilon). \tag{3.16c}$$

We also obtain the stream function directly as

$$\psi_F = -X\frac{cz}{L_F} + \cot\alpha\frac{L_F + X}{L_F} + O(\varepsilon), \tag{3.17a}$$

$$\psi_R = X\frac{cz}{L_R} \quad + O(\varepsilon), \tag{3.17b}$$

$$\psi_T = 0 \quad + O(\varepsilon). \tag{3.17c}$$

The equations (3.16) are ordinary differential equations for $b^{F,R,T}$ depending on the variable z, and require boundary and matching conditions at $z = 0,\ h,\ h + H_T$. At the ceiling $z = h + H_T$, the condition (3.5e) that b vanishes applies directly. However, the conditions at $z = 0$ and $z = h$ are obtained from consideration of horizontal boundary layers. These calculations are quite complicated and not very enlightening, so we will deal with them later in §3.4.

We assert for now that b is continuous to leading order at the symmetry axis, and in fact

$$b^{F,R} = 0 + O(\varepsilon^{2/3}) \text{ on symmetry axis } z = 0. \tag{3.18}$$

Applying this condition and the ceiling condition yields the results

$$b^F = \quad c\frac{z^3 - h^2 z}{6L_F} - \cot\alpha\frac{z^2 - hz}{2L_F} + A_F\frac{z}{h} \quad + O(\varepsilon^{2/3}), \tag{3.19a}$$

$$b^R = -c\frac{z^3 - h^2 z}{6L_R} \quad + A_R\frac{z}{h} \quad + O(\varepsilon^{2/3}), \tag{3.19b}$$

$$b^T = \quad A_T\frac{h + H_T - z}{H_T} + O(\varepsilon^{2/3}), \tag{3.19c}$$

where $A_{F,R,T}$ are constants of integration to be determined by matching at $z = h$.

At $z = h$, the boundary-layer solutions from §3.4 yield continuity of b at leading order and continuity of the total vertical diffusive buoyancy flux (with $O(\varepsilon^{1/3})$ errors),

$$b^F = b^R = b^T, \qquad L_F b^F_z + L_R b^R_z = L b^T_z \qquad \text{at } z = h. \tag{3.20a,b}$$

(We note that full continuity of b_z would be two conditions rather than one, and hence too many conditions to impose.) Applying these conditions to the solutions (3.19) determines the constants

$$A_F,\ A_R,\ A_T = \cot\alpha\,\frac{h^2}{2L(1 + h/H_T)} + O(\varepsilon^{1/3}), \tag{3.21}$$

which are substituted back into (3.19) to yield the solution.

3.2.3 Force calculation

Having calculated the stream function and buoyancy field in the main outer regions, we now have sufficient information to determine the leading-order force F on (the upper half of) the wedge. Since forces balance for Stokes flow and no horizontal body forces are present, the net horizontal force from the fluid on the wedge is equal and opposite to the horizontal force from the fluid on the bounding box, which is simpler to calculate.

As we can see from (3.6b), the pressure is approximately hydrostatic, $p_z \approx b$, in the outer region (and it is straightforward to check that this also holds true in the horizontal boundary layers). In addition, the forces on the bounding box are due to pressure only at leading order. Hence, the leading-order force is given by the difference in pressure on the front and rear walls.

As pressure is only defined up to an additive constant, we are free to choose the pressure to be zero at the point $(0, h+H_T)$ on the ceiling directly above the base of the wedge. From (3.6a), we thus find that $p = 0 + O(\varepsilon^3)$ along the top wall. Instead of obtaining the pressure field from separate asymptotic expansions in each region, we make use of the approximate hydrostatic relationship to write

$$p = -\int_z^{h+H_T} b\,dz + O(\varepsilon), \tag{3.22}$$

and hence the force on either wall is

$$\tilde{F} = \int_0^{h+H_T} p\,dz + O(\varepsilon) = -\int_0^{h+H_T} b\,z\,dz + O(\varepsilon), \tag{3.23}$$

where the latter expression is obtained by using (3.22) and interchanging the order of the two integrals.

The net force is the difference between (3.23) for the front and rear walls. In the top region, the outer solution (3.19c) is independent of X, and hence contributes equally to both integrals and has no effect. The leading-order leftward force on the wedge (rightward force on the wall), obtained from the outer solutions (3.19a,b), is thus

$$F = \int_0^h (b^F - b^R) z\,dz = \cot\alpha \frac{h^4}{24 L_F} - c\frac{h^5}{45}\left(\frac{1}{L_F} + \frac{1}{L_R}\right) + O(\varepsilon^{1/3}). \tag{3.24}$$

The first term, which is proportional to $\cot\alpha$, describes the propulsive force on the wedge due to the diffusion-driven flow. The second term, which is proportional to c, describes the drag on the wedge and was originally calculated by Foster & Saffman [2]. For a free wedge, the two forces balance, and hence the wedge moves at speed

$$c = \frac{\cot\alpha}{h}\frac{15 L_R}{8L} + O(\varepsilon^{1/3}). \tag{3.25}$$

(The dimensional version of this result simply has an additional factor κ on the right-hand side.)

Finally, we note that the leading-order result for the speed (3.25) does not depend on the height of the bounding box. In fact, this result also applies for other types of

boundary conditions on the top and bottom surfaces, such as a fixed-flux condition. This is because, although such changes would affect the resulting values of the constants $A_{F,R,T}$, the expression (3.24) for the forces depends on these only through the difference $(A_F - A_R)$, which is always prescribed by the matching condition (3.20) to be zero at leading order.

3.3 Summary and comparison between asymptotic and numerical results

Our asymptotic analysis has revealed the physical mechanisms behind diffusion-driven propulsion: The sloping surface of the wedge induces a Phillips–Wunsch flow up the slope. For a stationary wedge, this flux is balanced by a uniform downwelling in the front region with the same flux, which advects buoyant liquid downward and hence reduces the hydrostatic pressure in front of the wedge. The pressure difference between the front and the rear regions results in a leftward propulsive force. When the wedge moves forward, the downwelling is reduced since part of the fluid volume removed by the Phillips–Wunsch flux is balanced by the front region shrinking instead. A force-free wedge moves at the speed (3.25), for which the net force is zero.

Figure 5(a) shows a comparison between the asymptotically predicted and numerically calculated buoyancy profiles (using the same parameters as in §3.1). We find that there is good agreement between the two, confirming the validity of the asymptotic analysis. Figure 5(b) shows the dependence of the propulsion speed c on the slope of the wedge and the distances L_F and L_R to the front and rear walls. Again, a good agreement is achieved between the asymptotic and numerical results. (Also shown are a set of results calculated in §3.4, which include $O(\varepsilon^{1/3})$ corrections to the leading-order result calculated so far.)

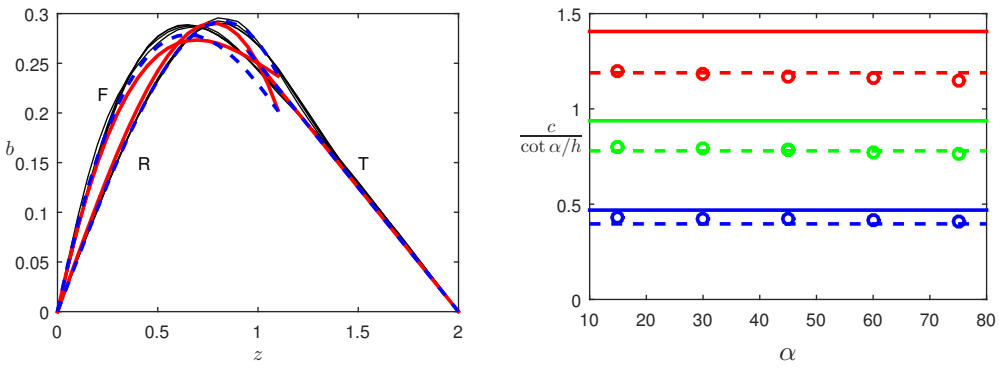

Figure 5: Comparison between asymptotic and numerical results. (a) Buoyancy distribution b as a function of vertical position z for the case shown in figure 3. The thin black curves are numerical data from vertical slices $X = \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4$ in the forward and rear regions. The thick curves are asymptotic results (3.19), with the leading-order coefficients (3.21, 3.25) (solid red curves) or the corrected ones (3.41, 3.43) below (dashed blue curves). (b) Propulsion speed c as a function of wedge slope α for the parameters (3.8), showing numerical results (circles), leading-order asymptotic results (3.25) (solid lines) and corrected asymptotic results (3.43) below (dashed lines). The values of (L_F, L_R) are $(1/4, 3/4)$ (top red series), $(1/2, 1/2)$ (middle green series) and $(3/4, 1/4)$ (bottom blue series).

The reason for choosing to work with a large, i.e. $O(\varepsilon^{-1} L_1)$, bounding box is now clear:

The upward Phillips–Wunsch flux, which in dimensional terms is $O(\kappa)$, is balanced by a downwelling with velocity $O(\kappa/L_F)$ in the forward region. The resulting advective flux of buoyancy $O(N^2 L_1 \kappa/L_F)$ is in turn balanced by diffusion of the buoyancy perturbations, which must be $b = O(N^2 L_1^2/L_F)$. For $L_F \gg L_1$, the buoyancy perturbation is small compared with the background buoyancy $B = O(N^2 L_1)$, and hence the non-linear advective terms and time derivative in the governing equation (2.10b) could be neglected.

If the dimensions of the box are $O(L_1)$, then the buoyancy perturbations are of the same order as the background stratification. In addition, the change in geometry of the system due to the motion of the wedge occurs on a faster time scale L_1^2/κ due to the reduced length of the box. Thus, both the time derivative and the non-linear terms become important, and the problem becomes more complicated. However, the mechanism of propulsion and the qualitative flow structure can be expected to remain the same.

3.4 The $\varepsilon^{1/3}$ layers and $O(\varepsilon^{1/3})$ corrections

We complete our analysis by calculating the boundary-layer solutions near $z = 0$ and $z = h$ that will yield the conditions, such as (3.20), asserted earlier in §3.2.2. These calculations are mostly a technicality, which is why we have left them to this separate subsection.

We first deal with the boundary layer near $z = 0$ in the forward region $-L_F < X < O(\varepsilon)$. The scaling $z \sim \varepsilon^{1/3}$ yields a new balance in the governing equations (3.14), so we define the rescaled vertical coordinate $\zeta = z/\varepsilon^{1/3}$. The governing equations (3.14) then become

$$b_X = \varepsilon^{2/3}\psi_{\zeta\zeta\zeta\zeta} + O(\varepsilon^3), \qquad -\varepsilon^{2/3}\psi_X = b_{\zeta\zeta} + O(\varepsilon). \tag{3.26}$$

The domain under consideration is (up to $O(\varepsilon)$) an infinite strip $-L_F < X < 0$, $-\infty < \zeta < \infty$. The appropriate effective boundary conditions in the X-direction are obtained by revisiting the calculations for the Phillips–Wunsch boundary-layer solution with a rescaled vertical coordinate ζ:

$$\psi = \varepsilon^{1/3}\zeta + O(\varepsilon^{2/3}) \text{ at } X = -L_F, \qquad \psi = \cot\alpha + O(\varepsilon^{2/3}) \text{ at } X = 0. \tag{3.27}$$

In the ζ-direction, we have the symmetry-axis conditions (3.5a) at $\zeta = 0$, and must also match to the outer solution as $\zeta \to \infty$.

We assume that the outer solution has an expansion

$$b^F = b_0^F + \varepsilon^{1/3} b_1^F + \varepsilon^{2/3} b_2^F + O(\varepsilon) \tag{3.28a}$$

with the generic behaviour (for $i = 0, 1, 2$)

$$b_i^F = C_i^F + D_i^F z + E_i^F z^2 + O(z^3) \qquad \text{as } z \to 0. \tag{3.28b}$$

We expand the boundary-layer solution as

$$b = b_0 + \varepsilon^{1/3} b_1 + \varepsilon^{2/3} b_2 + O(\varepsilon), \qquad \psi = \psi_2 + O(\varepsilon^{1/3}), \tag{3.29a}$$

and find that matching to the outer solutions (3.28b) requires

$$b_0 \sim C_0^F, \qquad b_1 \sim C_1^F + D_0\zeta, \qquad b_2 \sim C_2^F + D_1\zeta + E_0\zeta^2 \qquad \text{as } \zeta \to \infty. \tag{3.29b}$$

(It is straightforward to verify a posteriori that a correct matching of b also yields the correct matching of ψ, so we do not consider the latter here.)

At $O(1)$ and $O(\varepsilon^{1/3})$, the governing equations (3.26) yield that b_0 and b_1 are linear functions of ζ. Imposing the matching conditions (3.29b) and symmetry-axis conditions (3.5a) yields

$$b_0 = 0, \qquad b_1 = D_0\zeta, \qquad C_0^F = C_1^F = 0. \tag{3.30}$$

This confirms the condition (3.18) for the forward region, and the same analysis applies to the rear region as well.

The adjustment of ψ from a non-zero value (3.17a) in the outer solution to zero on the symmetry axis (3.5a) occurs here at $O(\varepsilon^{2/3})$. The equations

$$b_{2X} = \psi_{2\zeta\zeta\zeta\zeta}, \qquad -\psi_{2X} = b_{2\zeta\zeta}, \tag{3.31a,b}$$

are to be solved with boundary conditions (3.5a), (3.27) and (3.29b). A solution can be found using e.g. Fourier transforms and yields a condition on C_2 (i.e. the value of the outer solution b^F on the symmetry axis at $O(\varepsilon^{2/3})$), but we do not present this here.

Instead, we turn to the boundary layer near $z = h$, for which we define the rescaled coordinate $\zeta = (z-h)/\varepsilon^{1/3}$. In addition, it turns out that the stream function must be $O(\varepsilon^{-1/3})$ to deal with leading-order differences in b_z, so we define a rescaled stream function $\Psi = \varepsilon^{1/3}\psi = O(1)$. We follow the same steps as for the boundary layer at $z = 0$, but the analysis is more complicated.

The rescaled governing equations are

$$b_X = \varepsilon^{1/3}\Psi_{\zeta\zeta\zeta\zeta} + O(\varepsilon^3), \qquad -\varepsilon^{1/3}\Psi_X = b_{\zeta\zeta} + O(\varepsilon), \tag{3.32}$$

and the domain is again an infinite strip $-L_F < X < L_R$, $-\infty < \zeta < \infty$ but with a cut at $X = 0$, $-\infty < \zeta \le 0$ representing the wedge whose thickness $X = O(\varepsilon^{4/3})$ can be neglected at leading order. The appropriate boundary conditions in the X-direction are

$$\Psi = \varepsilon^{1/3} ch \quad + O(\varepsilon^{2/3}) \qquad \text{on walls} \qquad X = -L_F,\ L_R, \quad -\infty < \zeta < \infty \tag{3.33a}$$

$$\Psi = \varepsilon^{1/3}\cot\alpha + O(\varepsilon^{2/3}) \qquad \text{on wedge} \qquad X = 0^-, \quad \zeta < 0 \tag{3.33b}$$

$$\Psi = 0 \quad + O(\varepsilon^{2/3}) \qquad \text{on wedge} \qquad X = 0^+, \quad \zeta < 0. \tag{3.33c}$$

We expand the outer solutions as

$$b^{F,R,T} = b_0^{F,R,T} + \varepsilon^{1/3} b_1^{F,R,T} + \varepsilon^{2/3} b_2^{F,R,T} + O(\varepsilon), \tag{3.34a}$$

$$b_i^{F,R,T} = C_i^{F,R,T} + D_i^{F,R,T}(z-h) + E_i^{F,R,T}(z-h)^2 + O(z-h)^3 \quad \text{as } z \to h, \tag{3.34b}$$

and expand the boundary-layer solution as

$$b = b_0 + \varepsilon^{1/3} b_1 + \varepsilon^{2/3} b_2 + O(\varepsilon), \qquad \Psi = \Psi_1 + \varepsilon^{1/3}\Psi_2 + O(\varepsilon). \tag{3.35a}$$

The resulting matching conditions are

$$b_0 \sim C_0^{F,R,T}, \tag{3.35b}$$

$$b_1 \sim C_1^{F,R,T} + D_0^{F,R,T}\zeta, \tag{3.35c}$$

$$b_2 \sim C_2^{F,R,T} + D_1^{F,R,T}\zeta + E_0^{F,R,T}\zeta^2, \tag{3.35d}$$

as $|\zeta| \to \infty$ in each corresponding region.

At leading order, the governing equations (3.32) reveal that b_0 is a linear function of ζ, but the matching conditions (3.35b) prevent b_0 from growing linearly with ζ, so

$$b_0 = \text{constant} = C_0^F = C_0^R = C_0^T. \tag{3.36}$$

This yields the result (3.20a) stated above.

At $O(\varepsilon^{1/3})$, the governing equations (3.32) take the form (3.31), which together with the boundary conditions (3.33) and (3.35c) can be solved using the Wiener–Hopf method. The calculations and explicit solution are given in Appendix A of Moore and Saffman [7], and we do not repeat them here. The result is

$$L_F D_0^F + L_R D_0^R = L D_0^T, \tag{3.37}$$

from which we obtain the matching condition (3.20b) stated above. However, further inspection of the solution reveals that

$$C_1^F - C_1^T = \Delta \frac{L_R}{L}(D_0^R - D_0^F), \qquad C_1^R - C_1^T = \Delta \frac{L_F}{L}(D_0^F - D_0^R), \tag{3.38a}$$

$$\text{where} \qquad \Delta = -\frac{2\,\zeta(1/3)}{\pi^{1/3}}(L_F^{1/3} + L_R^{1/3} - L^{1/3}) > 0 \tag{3.38b}$$

and $\zeta(\cdot)$ denotes the Riemann ζ-function. We will use this result to get $O(\varepsilon^{1/3})$ corrections to the leading-order results calculated in §3.2.

At $O(\varepsilon^{2/3})$, the governing equations (3.32) again take the form (3.31b). This can again be solved using the Wiener–Hopf method, but the analysis would be very complicated. Instead, we integrate the analogue of equation (3.31b) over the region $-L_F < X < L_R$, $|\zeta| < M$ for some large constant M. We simplify the resulting left-hand side using the conditions (3.33) and the right-hand side using the conditions (3.35d), and obtain

$$-M \cot\alpha = 2M(L E_0^T + L_F E_0^F + L_R E_0^R) + L D_1^T - L_F D_1^F - L_R D_1^R. \tag{3.39}$$

The terms involving M cancel (by virtue of (3.19)), and we are left with a condition analogous to (3.37).

Thus, the matching conditions (3.20) at $z = h$ can be extended by (3.38, 3.39) to

$$b^F - b^T = 0 + \varepsilon^{1/3}\Delta \frac{L_R}{L}(b_z^R - b_z^F) + O(\varepsilon^{2/3}), \tag{3.40a}$$

$$b^R - b^T = 0 + \varepsilon^{1/3}\Delta \frac{L_F}{L}(b_z^F - b_z^R) + O(\varepsilon^{2/3}), \tag{3.40b}$$

$$L_F b_z^F + L_R b_z^R - L b_z^T = 0 \qquad + O(\varepsilon^{2/3}). \tag{3.40c}$$

and we obtain adjusted values of the constants:

$$A_F = \cot\alpha \frac{h^2}{2L(1+h/H_T)} + \varepsilon^{1/3}\Delta\left[\cot\alpha \frac{hL_R}{2L_F L} - c\frac{h^2}{3L_F}\right] + O(\varepsilon^{2/3}), \tag{3.41a}$$

$$A_R = \cot\alpha \frac{h^2}{2L(1+h/H_T)} + \varepsilon^{1/3}\Delta\left[-\cot\alpha \frac{h}{2L} + c\frac{h^2}{3L_R}\right] + O(\varepsilon^{2/3}), \tag{3.41b}$$

$$A_T = \cot\alpha \frac{h^2}{2L(1+h/H_T)} \qquad + O(\varepsilon^{2/3}). \tag{3.41c}$$

These can be directly substituted into the solutions (3.19) to yield the corrected results shown in figure 5(a), which do indeed agree better with the numerical results than the leading-order asymptotic results. The corrected force is

$$F = \cot\alpha \frac{h^4}{24L_F}\left(1+\varepsilon^{1/3}\frac{4\Delta}{h}\right) - c\frac{h^5}{45}\left(\frac{1}{L_F}+\frac{1}{L_R}\right)\left(1+\varepsilon^{1/3}\frac{5\Delta}{h}\right)+O(\varepsilon^{2/3}), \qquad (3.42)$$

and the resulting corrected speed

$$c = \frac{\cot\alpha}{h}\frac{15L_R}{8L}\left(1-\varepsilon^{1/3}\frac{\Delta}{h}\right)+O(\varepsilon^{2/3}), \qquad (3.43)$$

(where Δ is given in (3.38b)) also agrees well with the numerical results (see figure 5(b)).

4 Wedge in a Box with Fixed-buoyancy Walls

We now consider the case when the buoyancy B is prescribed to be equal to the background stratification $N^2 z$ on the walls of the bounding box, rather than the walls being insulating with no buoyancy flux through them. We focus on the case when the dimensions of the box and the wedge both have the same scale $O(L_1)$.

We non-dimensionalize lengths using the scale L_1, stream function using κ, velocity using κ/L_1, time using L_1^2/κ, and buoyancy perturbation using $\varepsilon N^2 L_1$. The resulting governing equations are

$$b_x = \varepsilon^3\nabla^4\psi, \qquad \varepsilon(\dot{b}+\psi_x b_z - \psi_z b_x) - \psi_x = \varepsilon\nabla^2 b, \qquad (4.1a,b)$$

and the boundary conditions are

$$\psi = \psi_{zz} = 0, \quad b = 0 \quad \text{on symmetry axis} \quad z = 0. \qquad (4.2a)$$
$$\psi = \psi_n = 0, \quad b_n = -\tfrac{1}{\varepsilon}\cos\alpha \quad \text{on wedge front} \quad x = -l + z\cot\alpha, \qquad (4.2b)$$
$$\psi = \psi_n = 0, \quad b_n = 0 \quad \text{on wedge rear} \quad x = 0, \qquad (4.2c)$$
$$\psi = cz,\ \psi_x = 0, \quad b = 0 \quad \text{on walls} \quad x = -L_F,\ L_R, \qquad (4.2d)$$
$$\psi = cz,\ \psi_z = c, \quad b = 0 \quad \text{on ceiling} \quad z = h + H_T. \qquad (4.2e)$$

4.1 Numerical results

In this section, we use the parameters

$$\varepsilon = \frac{L_0}{L_1} = \frac{1}{100}, \quad h = H_T = 1, \quad (L_F, L_R) = (2,1),(2,2),(2,3), \qquad (4.3)$$

and focus on the particular case $(L_F, L_R) = (2,1)$ and $\alpha = 45°$. Figure 6 shows numerical results obtained using the finite-element method as described in §3.1.

We again find that a Phillips–Wunsch boundary-layer flow develops on the sloping surface of the wedge, and that there are horizontal boundary layers at $z = 0$ and $z = h$. However, in addition there are also boundary layers on the front and back walls. The buoyancy perturbations remain mainly independent of x as before.

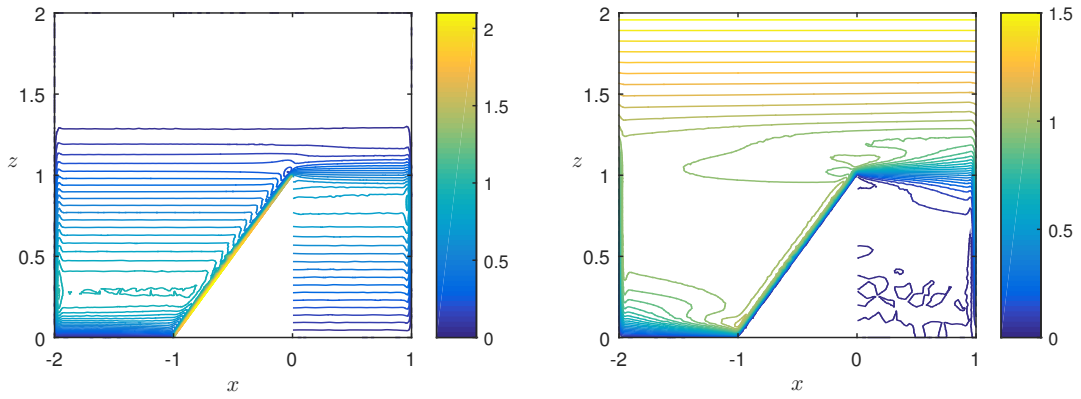

Figure 6: Contour plots of buoyancy perturbation b (left) and stream function ψ (right, flow from left to right in the frame of reference of the wedge) for the parameters (4.3), $(L_F, L_R) = (2, 1)$, $\alpha = 45°$. The contour spacing is 0.05.

4.2 Asymptotic calculation

The asymptotic calculation is again similar to the one in §3.2, although there are more boundary layers involved. The structure of these boundary layers is shown in figure 7, and their relevance will become clear as we proceed with the calculations.

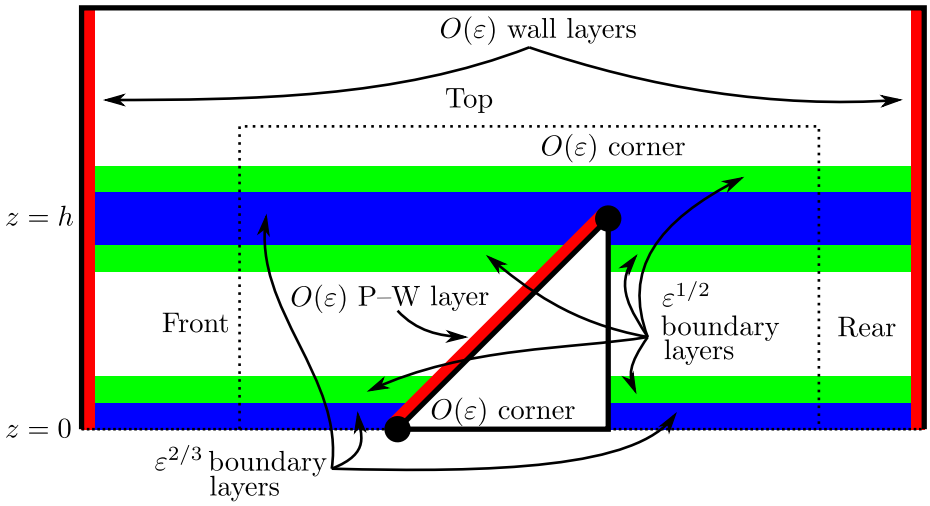

Figure 7: Schematic of asymptotic regions for a wedge in a box with fixed-buoyancy walls. We calculate the force on the dotted box.

4.2.1 Boundary layers on the wedge and walls

We again begin by seeking effective boundary conditions for the outer solutions $\psi^{F,R,T}$ in the x-direction. On the wedge, the calculations from §3.2.1 apply, and we obtain

$$\psi^F = \cot\alpha + O(\varepsilon) \text{ at wedge } x = -l + z\cot\alpha, \qquad \psi^R = 0 + O(\varepsilon) \text{ at wedge } x = 0, \quad (4.4)$$

describing how a Phillips–Wunsch flux of magnitude $\cot\alpha$ is driven up the slope.

For the front wall, we obtain a boundary-layer solution analogous to (2.20) by defining a rescaled variable $\xi = (x + L_F(t))/\varepsilon$ and solving the resulting equations

$$b_\xi = \psi_{\xi\xi\xi\xi} + \varepsilon^2\psi_{\xi\xi zz} + \varepsilon^4\psi_{zzzz}, \quad \varepsilon^2\dot{b} + \varepsilon(-cb_\xi + \psi_z b_\xi - \psi_\xi b_x) - \psi_\xi = b_{\xi\xi} + \varepsilon^2 b_{zz}, \quad (4.5a)$$

$$\psi = cz, \quad \psi_\xi = 0, \quad b = 0 \qquad \text{at the wall } \xi = 0. \quad (4.5b)$$

The result is

$$\psi = cz + \frac{b_\infty(z)}{\sqrt{2}}\left[1 - (\cos\tfrac{\eta}{\sqrt{2}} + \sin\tfrac{\eta}{\sqrt{2}})e^{-\eta/\sqrt{2}}\right] + O(\varepsilon), \quad (4.6a)$$

$$b = b_\infty(z)\left[1 - \cos\tfrac{\eta}{\sqrt{2}}e^{-\eta/\sqrt{2}}\right] + O(\varepsilon). \quad (4.6b)$$

Again, matching the far-field behaviour to the outer solution determines the constant of integration $b_\infty(z)$, and yields an effective condition on the outer stream function in terms of the outer buoyancy

$$\psi^F = cz + \tfrac{1}{\sqrt{2}}b^F + O(\varepsilon) \qquad \text{on the front wall } x = -L_F, \quad (4.7a)$$

A similar analysis of the rear wall yields

$$\psi^R = cz - \tfrac{1}{\sqrt{2}}b^R + O(\varepsilon) \qquad \text{on the rear wall } x = L_R. \quad (4.7b)$$

As for the solution (ψ^T, b^T) in the top region $z \geq h$, we similarly find

$$\psi^T = cz + \tfrac{1}{\sqrt{2}}b^T + O(\varepsilon) \qquad \text{on the front wall } x = -L_F, \quad (4.7c)$$

$$\psi^T = cz - \tfrac{1}{\sqrt{2}}b^T + O(\varepsilon) \qquad \text{on the rear wall } x = L_R. \quad (4.7d)$$

4.2.2 Leading-order outer solution

We now have sufficient information to calculate the leading-order outer solutions. The governing equations (4.1) yield

$$\psi^F_x,\ b^F_x,\ \psi^R_x,\ b^R_x,\ \psi^T_x,\ b^T_x = 0 + O(\varepsilon). \quad (4.8)$$

These equations describe, as before, how the flow is too weak to sustain horizontal gradients in buoyancy. In addition, a vertical downwelling is suppressed because the buoyancy perturbation is too weak for its diffusion to balance the resulting advection. Thus, ψ and b are functions of z only and we immediately find, from the effective conditions (4.4, 4.7),

$$\psi^F = \cot\alpha + O(\varepsilon), \qquad b^F = \sqrt{2}\,(\cot\alpha - cz) + O(\varepsilon), \quad (4.9a)$$

$$\psi^R = 0 \quad + O(\varepsilon), \qquad b^R = \sqrt{2}\,cz \quad + O(\varepsilon), \quad (4.9b)$$

$$\psi^T = cz \quad + O(\varepsilon), \qquad b^T = 0 \quad + O(\varepsilon). \quad (4.9c)$$

4.2.3 Horizontal $\varepsilon^{1/2}$ boundary layers

We expect, as before, to find boundary layers near $z = 0$ and $z = h$. Assuming that x remains $O(1)$ while z scales with some positive power of ε, two possible balances are found in the governing equations (4.1), corresponding to two boundary-layer thicknesses $\varepsilon^{1/2}$ and $\varepsilon^{2/3}$. These boundary layers are analogous to the outer region (§3.2.2) and $\varepsilon^{1/3}$ boundary layers (§3.4) from the insulating case, respectively. We calculate the solutions for the $\varepsilon^{1/2}$ layer here, while the calculations for the $\varepsilon^{1/3}$ layer are exactly identical to those in §3.4 (apart from a change $X \to x$ and $\varepsilon \to \varepsilon^{1/2}$) so we will use those results directly here.

Near $z = 0$, in the front region, we define a rescaled coordinate $\zeta = z/\varepsilon^{1/2}$, and obtain the governing equations

$$b^F_x = 0 + O(\varepsilon), \qquad -\psi^F_x = b^F_{\zeta\zeta} + O(\varepsilon^{1/2}), \tag{4.10a,b}$$

whose form we recognize from equation (3.15) for the outer region in the insulating-wall case.

At the left end of this $\varepsilon^{1/2}$ boundary layer, i.e. where it meets the wall, we must seek a corner solution with $x + L_F \sim \varepsilon$ and $z \sim \varepsilon^{1/2}$. The analysis of this region is identical to the wall analysis from §4.2.1, except that the vertical length scale is $O(\varepsilon^{1/2})$ rather than $O(1)$. The resulting effective condition, analogous to (4.7a), is

$$\psi^F = \tfrac{1}{\sqrt{2}} b + O(\varepsilon^{1/2}) \qquad \text{on the front wall } x = -L_F. \tag{4.11a}$$

Similarly, at the right end of the $\varepsilon^{1/2}$ boundary layer, i.e. where it meets the wedge, it is straightforward to verify that the Phillips–Wunsch solution from §3.2.1 is recovered, with result

$$\psi^F = \cot\alpha + O(\varepsilon^{1/2}) \qquad \text{on the wedge } x = -l + \varepsilon^{1/2}\zeta\cot\alpha. \tag{4.11b}$$

We proceed, as in §3.2.2, by integrating (4.10b) in x (using the fact that b is independent of x to leading order) and obtain

$$\hat{L}_F b^F_{\zeta\zeta} = \tfrac{1}{\sqrt{2}} b^F - \cot\alpha + O(\varepsilon^{1/2}), \qquad \text{where } \hat{L}_F = L_F - l. \tag{4.12}$$

In matching with the outer layer, i.e. $\zeta \to \infty$, we require that b does not grow exponentially. At $\zeta = 0$, matching with the $\varepsilon^{2/3}$ layer (cf. §3.4) yields $b^F = 0 + O(\varepsilon^{1/3})$. The solution is

$$b^F = \sqrt{2}\cot\alpha \left[1 - \mathrm{e}^{-\zeta/\sqrt{\sqrt{2}\hat{L}_F}}\right] + O(\varepsilon^{1/3}), \tag{4.13a}$$

$$\psi^F = \cot\alpha \left[1 + \frac{x+l}{\hat{L}_F}\mathrm{e}^{-\zeta/\sqrt{\sqrt{2}\hat{L}_F}}\right] + O(\varepsilon^{1/3}). \tag{4.13b}$$

Similarly to in §3.4, the nested $O(\varepsilon^{2/3})$ boundary layer ensures that ψ satisfies the condition $\psi = 0$ from (4.2a) on the symmetry axis.

A similar analysis applies near the symmetry axis in the rear region, but in this case the resulting boundary-layer solutions are trivial since the outer solution (4.9b) satisfies the symmetry-axis conditions (4.2a).

Near $z = h$, we instead define the rescaled coordinate $\zeta = (z - h)/\varepsilon^{1/2}$. The rescaled leading-order governing equations (4.10) remain the same, but the conditions (4.4, 4.7) on the wedge and walls become

$$\psi^F = ch + \tfrac{1}{\sqrt{2}} b^F \qquad \text{at } x = -L_F, \qquad \psi^F = \cot\alpha \qquad \text{at } x = \varepsilon^{1/2}\zeta\cot\alpha, \tag{4.14a}$$

$$\psi^R = 0 \qquad \text{at } x = 0, \qquad \psi^R = ch - \tfrac{1}{\sqrt{2}} b^R \qquad \text{at } x = L_R, \tag{4.14b}$$

$$\psi^T = ch + \tfrac{1}{\sqrt{2}} b^T \qquad \text{at } x = -L_F, \qquad \psi^T = ch - \tfrac{1}{\sqrt{2}} b^T \qquad \text{at } x = L_R, \tag{4.14c}$$

with errors $O(\varepsilon^{1/2})$. Integration of (4.10b) yields

$$L_F b^F_{\zeta\zeta} = \tfrac{1}{\sqrt{2}} b^F + ch - \cot\alpha + O(\varepsilon^{1/2}), \tag{4.15a}$$

$$L_R b^R_{\zeta\zeta} = \tfrac{1}{\sqrt{2}} b^R - ch + O(\varepsilon^{1/2}), \tag{4.15b}$$

$$L_A b^T_{\zeta\zeta} = \tfrac{1}{\sqrt{2}} b^T + O(\varepsilon^{1/2}), \tag{4.15c}$$

where we have defined the average

$$L_A = \frac{L_F + L_R}{2} = \frac{L}{2}. \tag{4.16}$$

Requiring that the solutions b do not grow exponentially in the matching with the outer solutions (i.e. as $\zeta \to -\infty$ for the front and rear regions and $\zeta \to +\infty$ for the top region) yields

$$b^F = \sqrt{2}\left[\, \cot\alpha - ch + A_F \mathrm{e}^{\zeta/\sqrt{\sqrt{2}L_F}} \,\right] + O(\varepsilon^{1/2}), \tag{4.17a}$$

$$b^R = \sqrt{2}\left[\, ch + A_R \mathrm{e}^{\zeta/\sqrt{\sqrt{2}L_R}} \,\right] + O(\varepsilon^{1/2}), \tag{4.17b}$$

$$b^T = \sqrt{2}\left[\, A_T \mathrm{e}^{-\zeta/\sqrt{\sqrt{2}L_A}} \right] + O(\varepsilon^{1/2}), \tag{4.17c}$$

where the three constants of integration $A_{F,R,T}$ will be determined by matching to the $\varepsilon^{2/3}$ layer. (The corresponding solutions for ψ are easily obtained from (4.10b).)

The nested $\varepsilon^{2/3}$ layer yields the following matching conditions, analogous to (3.40),

$$b^F - b^T = 0 + \varepsilon^{1/6}\Delta\frac{L_R}{L}(b^R_\zeta - b^F_\zeta) + O(\varepsilon^{1/3}), \tag{4.18a}$$

$$b^R - b^T = 0 + \varepsilon^{1/6}\Delta\frac{L_F}{L}(b^F_\zeta - b^R_\zeta) + O(\varepsilon^{1/3}), \tag{4.18b}$$

$$L_F b^F_\zeta + L_R b^R_\zeta - L b^T_\zeta = 0 + O(\varepsilon^{1/3}), \tag{4.18c}$$

where Δ is given by (3.38b). Applying these conditions to the solutions (4.17) determines

the constants

$$A_F = \frac{1}{D}\left[-\cot\alpha\left(\sqrt{L_R}+2\sqrt{L_A}\right)+2ch\left(\sqrt{L_R}+\sqrt{L_A}\right)\right]+$$
$$-\frac{\varepsilon^{1/6}E}{\sqrt{\sqrt{2}L_F}}\left[\sqrt{L_A}+\sqrt{L_R}\right]+O(\varepsilon^{1/3}), \tag{4.19a}$$
$$A_R = \frac{1}{D}\left[\cot\alpha\sqrt{L_F}-2ch\left(\sqrt{L_F}+\sqrt{L_A}\right)\right]+$$
$$+\frac{\varepsilon^{1/6}E}{\sqrt{\sqrt{2}L_R}}\left[\sqrt{L_A}+\sqrt{L_F}\right]+O(\varepsilon^{1/3}), \tag{4.19b}$$
$$A_T = \frac{1}{D}\left[-\cot\alpha\sqrt{L_F}+ch\left(\sqrt{L_F}-\sqrt{L_R}\right)\right]+$$
$$+\frac{\varepsilon^{1/6}E}{\sqrt{\sqrt{2}L_A}}\left[\frac{\sqrt{L_R}-\sqrt{L_F}}{2}\right]+O(\varepsilon^{1/3}), \tag{4.19c}$$

where

$$D = 2\sqrt{L_A}+\sqrt{L_F}+\sqrt{L_R}, \qquad E = \frac{2\Delta}{D}\left[-\cot\alpha\frac{\sqrt{L_A}+\sqrt{L_R}}{D}+ch\right]. \tag{4.20}$$

We can combine the solutions (4.9, 4.13, 4.17) to form the following composite solutions, which are valid in both the outer regions and the $\varepsilon^{1/2}$ boundary layers:

$$b^F = \sqrt{2}\left[\cot\alpha\left(1-\exp\frac{-z}{\sqrt{\sqrt{2}\varepsilon\hat{L}_F}}\right)-cz+A_F\exp\frac{-(h-z)}{\sqrt{\sqrt{2}\varepsilon L_F}}\right]+O(\varepsilon^{1/3}), \tag{4.21a}$$
$$b^R = \sqrt{2}\left[cz+A_R\exp\frac{-(h-z)}{\sqrt{\sqrt{2}\varepsilon L_R}}\right]+O(\varepsilon^{1/3}), \tag{4.21b}$$
$$b^T = \sqrt{2}\left[A_T\exp\frac{-(z-h)}{\sqrt{\sqrt{2}\varepsilon L_A}}\right]+O(\varepsilon^{1/3}). \tag{4.21c}$$

4.2.4 Force calculation

As in 3.2.3, the pressure is approximately hydrostatic in the outer regions. Rather than calculating the net force on either the wedge or the bounding box (which are now both lined with boundary layers), we use an intermediate surface (shown in figure 7) which intersects the outer regions away from any boundary layers (apart from the ones at $z = h$ which cannot be avoided). The net force on (the top half of) the wedge is then given by a formula similar to (3.24), and we use the composite solution (4.21) to obtain

$$F = \int_0^h (b^F - b^R)z\,dz + O(\varepsilon^{5/6}) = \tag{4.22a}$$
$$= \sqrt{2}\left[\cot\alpha\frac{h^2}{2}-\frac{2ch^3}{3}+\varepsilon^{1/2}2^{1/4}h\left(\sqrt{L_F}A_F-\sqrt{L_R}A_R\right)\right]+O(\varepsilon^{5/6}), \tag{4.22b}$$

where the main error is due to the $O(\varepsilon^{1/3})$ corrections in the $\varepsilon^{1/2}$ layer and the neglect of $O(\varepsilon^{1/6})$ variations in the $\varepsilon^{2/3}$ layer near $z = h$. Hence the speed of the wedge is

$$c(t) = \frac{3}{4}\frac{\cot\alpha}{h}\left[1 + \frac{\varepsilon^{1/2}2^{5/4}}{Dh}\left(\sqrt{L_F L_R} - \tfrac{1}{2}\sqrt{L_F L_A} + \tfrac{3}{2}\sqrt{L_R L_A}\right) + \right.$$
$$\left. + \frac{\varepsilon^{2/3}2\Delta}{Dh}\left(-\tfrac{3}{2}\sqrt{L_F} + \tfrac{1}{2}\sqrt{L_R} - \sqrt{L_A}\right)\right] + O(\varepsilon^{5/6}), \qquad (4.23)$$

as this is the value for which $F = 0$.

4.3 Summary and comparison with numerical results

Our analysis of this case with fixed-buoyancy conditions on the bounding walls bears many similarities to the previous analysis with insulating walls (see §3.3). The main difference is the appearance of front and rear outer regions with new behaviour (4.9), namely no leading-order flow ($\psi_x = \psi_z = 0$) relative to the wedge. This phenomenon is called "blocking" and is the stratified analogue of Taylor columns in rotating systems. The Phillips–Wunsch flux up the slope is then not supplied from a uniform downwelling in the front region, but rather from a strong downward current confined to the front wall (and via a symmetry-axis boundary-layer jet).

Near $z = 0$ and $z = h$, there are $\varepsilon^{1/2}$ boundary layers which behave like the outer regions in §3.2.2, and nested inside these are $\varepsilon^{2/3}$ boundary layers which are identical to the $\varepsilon^{1/3}$ layers in §3.4. These two types of boundary layer are the stratified analogues of Stewartson $E^{1/4}$ and $E^{1/3}$ layers for rotating flows.

We compare our composite (outer and $\varepsilon^{1/2}$-layer) solutions (4.21) with numerical results in figure 8a, and find that they agree well. However, the results for the propulsion speed c are less convincing in the case $\varepsilon = 1/100$ (see figure 8b), and so we have included the case $\varepsilon = 1/400$ as well (figure 8c). In the latter case, we can see a clear improvement between the leading-order result, the $O(\varepsilon^{1/2})$ correction, and finally the $O(\varepsilon^{2/3})$ correction.

5 Wedge in a Very Large or Infinite Domain

Having investigated the cases with a wedge placed in boxes whose size are comparable to or slightly larger than the size of the wedge, we finally consider the case when the box is very much larger than the size of the wedge. We approach this case by considering what happens as the horizontal dimensions of the fixed-buoyancy box in §subsec are increased.

As we can see from e.g. (4.21), the dimensional thickness of the $\varepsilon^{1/2}$ boundary layers is $O((L_0 L)^{1/2})$, while the $\varepsilon^{2/3}$ boundary layers can be seen to have thickness $O((L_0^2 L)^{1/3})$ from e.g. (2.11). As L grows to $O(\varepsilon^{-1}L_1)$, the $\varepsilon^{1/2}$ boundary layers near $z = 0$ and $z = h$ invade and replace the outer region. (Explicit asymptotic solutions can be found in this case, but end up being complicated expressions involving hyperbolic trigonometric functions from the solution of equations similar to (4.15), so we do not report them here.) As L grows further to $O(\varepsilon^{-2})$, the $\varepsilon^{2/3}$ boundary layers fill the domain, and equations like (3.26) need to be solved.

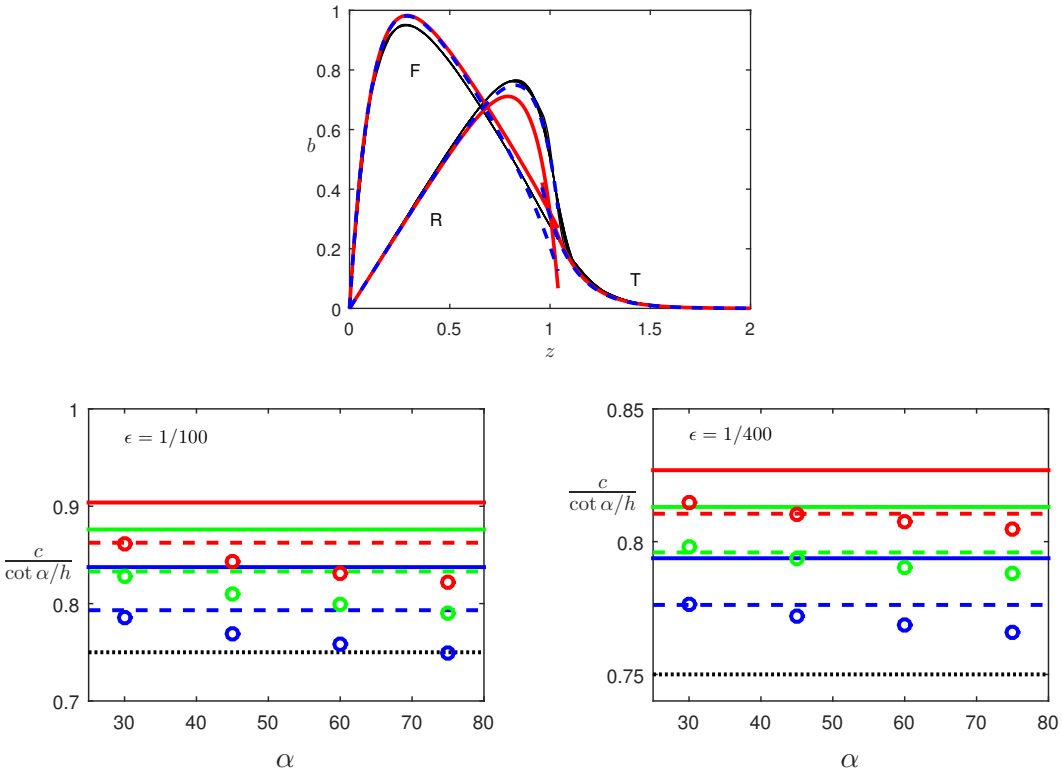

Figure 8: Comparison between asymptotic and numerical results. (a) Buoyancy distribution b as a function of vertical position z for the case shown in figure 6. The thin black curves are numerical data from vertical slices at $-1.9 \le x \le -1.1$ and $0.1 \le x \le 0.9$ with spacing 0.1 in the forward and rear regions. The thick curves are asymptotic composite solutions (4.21), with the leading-order (solid red curves) or corrected (dashed blue curves) coefficients (4.19, 4.23). (b,c) Propulsion speed c as a function of wedge slope α for the parameters (3.8), showing numerical results (circles) and asymptotic results (4.23) with the leading-order term only (black dotted line), $O(\varepsilon^{1/2})$ corrections (solid lines) and $O(\varepsilon^{2/3})$ corrections (dashed lines). The values of (L_F, L_R) are $(2,3)$ (top red series), $(2,2)$ (middle green series) and $(2,1)$ (bottom blue series).

Finally, for $L \gg \varepsilon^{-2}L_1$ (and $H_T \gg L_1$), we still have to solve the $\varepsilon^{2/3}$-layer equations, but since the box is much larger than the natural length scales

$$z \sim L_1, \qquad x \sim \varepsilon^{-2}L_1, \tag{5.1}$$

we can treat the domain as being infinite, which simplifies the analysis greatly.

5.1 Asymptotic analysis

Based on the scaling (5.1), we non-dimensionalize lengths by L_1 and introduce a stretched horizontal coordinate $X = \varepsilon^2 x$. We scale the stream function by κ, velocity by κ/L_1, buoyancy by $\varepsilon^2 N^2 L_1$, pressure by $\varepsilon^2 N^2 L_1^2$ and force by $\varepsilon^2 N^2 L_1^3$. We proceed immediately

with calculating the outer solution, valid throughout the domain away from any boundaries. Unlike in §3 and §4, we consider the whole domain rather than just the upper half $z \geq 0$. Also, we work with the stream function $\Psi = \psi - cz$ in the reference frame of the ambient fluid to simplify the decay conditions (but retain the use of a coordinate system moving with the wedge so that the geometry does not evolve with time).

The governing equations (2.9) simplify to

$$b_X = \Psi_{zzzz}, \qquad -\Psi_X = b_{zz}, \tag{5.2}$$

with errors of size $O(\varepsilon^2)$ or smaller. As we treat the domain as being infinite, we impose decay conditions in the far field,

$$\Psi,\ b \to 0 \qquad \text{as} \qquad X,\ z \to \pm\infty, \tag{5.3a}$$

and the conditions on the bounding box, whether they have fixed buoyancy or no buoyancy flux, have no effect at leading order.

In the horizontally stretched coordinate system, the width of the wedge is $O(\varepsilon^2)$ and hence negligible at leading order. Thus, we can treat it as a cut at $X = 0$, $|z| \leq h$. The Phillips–Wunsch flow on the wedge yields the effective conditions

$$\Psi = \cot\alpha - cz \quad \text{at} \quad X = 0^-, \qquad \Psi = -cz \quad \text{at} \quad X = 0^+ \qquad \text{for } 0 < z < h \tag{5.3b}$$

on the upper portion of the wedge, and the corresponding antisymmetric conditions ($\Psi = -\cot\alpha + cz$ and $\Psi = cz$) on the lower portion $-h < z < 0$, with errors of size $O(\varepsilon)$.

We exploit the linearity of the leading-order equations (5.2) to decompose the boundary conditions (5.3b), which describe the redistribution of fluid due to Phillips–Wunsch flow, into two parts. The first part, with Ψ anti-symmetric in X,

$$\Psi = \tfrac{1}{2}\cot\alpha \quad \text{at} \quad X = 0^-, \qquad \Psi = -\tfrac{1}{2}\cot\alpha \quad \text{at} \quad X = 0^+ \qquad \text{for } 0 < z < h, \tag{5.4a}$$

describes the effect of moving fluid from the centreline $z = 0$ to the heights $z = \pm h$, and will be solved by a distribution of sources and sinks on the wedge. However, this gives rise to a pressure distribution that is symmetric in X, and hence yields no net horizontal force. The second part, with Ψ symmetric in X,

$$\Psi = \tfrac{1}{2}\cot\alpha - cz \quad \text{at} \quad X = 0^- \text{ and } X = 0^+ \qquad \text{for } 0 < z < h, \tag{5.4b}$$

describes the effect of moving fluid from one side of the wedge to the other, and will be solved by a distribution of force singularities on the wedge. The sum of the two solutions form the solution to the original boundary conditions.

Although the point source and point force solutions have been described many times before (see e.g. [5, 7, 3]), we rederive them briefly here for completeness.

5.1.1 Antisymmetric part

We first consider the flow due to a unit point source at the origin. The flow can not be described by a continuous stream function, as it does not satisfy the continuity equation

(2.6b) at the origin (and indeed the conditions (5.4a) are discontinuous on the wedge). However, we can still define

$$\Psi^s(X,z) = \int_0^z u(X,z')\,dz' \qquad \Rightarrow \qquad u = \Psi^s_z, \tag{5.5}$$

and the continuity equation (2.6b) with a point source yields

$$u_X + w_z = \delta(X)\delta(z) \qquad \Rightarrow \qquad w = -\Psi^s_X + \delta(X)\frac{\mathrm{sgn}(z)}{2}, \tag{5.6}$$

where δ is the Dirac δ-function and sgn is the signum function. Including the point source in the governing equations (5.2) and eliminating b yields

$$\Psi^s_{XX} + \Psi^s_{zzzzzz} = \delta'(X)\frac{\mathrm{sgn}(z)}{2}. \tag{5.7}$$

We take a Fourier transform in the z-direction and obtain

$$\widetilde{\Psi}^s_{XX} = k^6\widetilde{\Psi}^s + \frac{1}{ik}\delta'(X) \qquad \Rightarrow \qquad \widetilde{\Psi}^s = \mathrm{sgn}(X)\frac{1}{2ik}e^{-|k^3X|}, \tag{5.8}$$

after application of the decay boundary conditions in the X-direction. We only require the solution on the z-axis, so we set $X = 0^{\pm}$ and invert the Fourier transform to find

$$\Psi^s(0^{\pm},z) = \pm\frac{\mathrm{sgn}(z)}{4}. \tag{5.9}$$

Hence, as may be expected, the original antisymmetric conditions (5.4a) are satisfied by the distribution

$$\Psi(X,z) = \cot\alpha\,[\Psi^s(X,z-h) - 2\Psi^s(X,z) + \Psi^s(X,z+h)] \tag{5.10}$$

of two point sources of strength $\cot\alpha$ at $z = \pm h$ and a point sink of double the strength at $z = 0$.

5.1.2 Symmetric part

We now consider the flow due to a horizontal rightward unit point force at the origin (corresponding to the fluid imparting a unit leftward force on the wedge). This introduces the term $\delta(X)\delta(z)$, on the right-hand side of the approximate horizontal momentum equation $p_X = \Psi_{zzz}$. Modifying the governing equation (5.2) yields

$$\Psi^f_{XX} + \Psi^f_{zzzzzz} = -\delta(X)\delta_{zzz}(z). \tag{5.11}$$

As in §5.1.1, it is straightforward to solve the equation using a Fourier transform,

$$\widetilde{\Psi}^f_{XX} = k^6\widetilde{\Psi}^f + ik^3\delta(X) \qquad \Rightarrow \qquad \widetilde{\Psi}^f = -\frac{i\,\mathrm{sgn}(k)}{2}e^{-|k^3X|}, \tag{5.12}$$

and inverting the transform at $X = 0$ yields

$$\Psi^f(0,z) = \frac{1}{2\pi z}. \tag{5.13}$$

An unknown force distribution $f(z)$ (symmetric in z) located on the wedge $-h < z < h$ gives rise to the flow

$$\Psi(X,z) = \int_{-h}^{h} f(z')\Psi^f(X, z-z')\,dz', \tag{5.14}$$

so the condition (5.4b) yields

$$\frac{1}{2\pi}\fint_{-h}^{h} \frac{f(z')}{z-z'}\,dz' = \operatorname{sgn}(z)\tfrac{1}{2}\cot\alpha - cz, \tag{5.15}$$

where the horizontal bar indicates that the Cauchy principal value is taken for the integral. The inversion formula for finite-range Hilbert transforms is given by [14] as

$$g(z) = \frac{1}{2\pi}\fint_{-h}^{h} \frac{f(z')}{z-z'}\,dz' \quad \Rightarrow \quad \pi\sqrt{h^2-z^2}f(z) = F - 2\fint_{-h}^{h} \frac{g(z')\sqrt{h^2-z'^2}}{z-z'}\,dz', \tag{5.16}$$

where $F = \int_{-h}^{h} f(z')\,dz'$. Hence, consideration of $z = h$ yields the net force as

$$F = 2\int_{-h}^{h} \left[\operatorname{sgn}(z)\tfrac{1}{2}\cot\alpha - cz\right] \frac{\sqrt{h^2-z^2}}{h-z}\,dz = 2h\cot\alpha - ch^2\pi, \tag{5.17a}$$

and the force distribution is given by

$$f(z) = \frac{2}{\pi}\cot\alpha\,\ln\frac{h+\sqrt{h^2-z^2}}{|z|} - 2c\sqrt{h^2-z^2}. \tag{5.17b}$$

The flow field is then determined by (5.14).

As the antisymmetric solution in §5.1.1 has no net force. Hence, we deduce that the net leftward force on the wedge in the original problem is also given by (5.17a), and the propulsion speed is thus

$$c = \frac{2}{\pi}\frac{\cot\alpha}{h} + O(\varepsilon). \tag{5.18}$$

6 Summary and Discussion

We have investigated the two-dimensional diffusion-driven flow that gives rise to propulsion of an insulating wedge in a stably stratified fluid. Using the method of matched asymptotic expansions, we have calculated the flow field, the buoyancy distribution and the propulsion speed in the case when the dimensions of the wedge are much larger than the natural length scale $L_0 = (\kappa\nu/N^2)^{1/4}$ of the flow.

Whether the outer boundaries are insulating (§3), have prescribed buoyancy (§4), or are very far away (§5), we find that the root cause of the propulsion is the Phillips–Wunsch flux of magnitude $\kappa\cot\alpha$ up the sloping upper surface of the wedge (and down the sloping lower surface, by symmetry). This removal of fluid at the apex of the wedge is balanced by vertical flow towards the symmetry axis ahead of the wedge, either throughout the forward region (§3, §5) or along a boundary layer on the front wall (§4). The resulting vertical advection of buoyancy lowers the hydrostatic pressure ahead of the wedge, resulting in a propulsive

force. When the wedge moves at speed c, the $O(ch)$ redistribution of fluid generates a drag force by the same mechanism.

Since the forces on the wedge must balance, we can expect the scaling

$$c \sim \frac{\kappa \cot \alpha}{h} \tag{6.1}$$

to hold. Our calculations confirm this result, with various $O(1)$ numerical prefactors, for all three types of outer boundary condition (3.43, 4.23, 5.18). Somewhat surprisingly, the scaling (6.1) depends on neither the the strength N^2 of the background stratification nor the viscosity ν of the fluid. (However, N^2 and ν do affect ϵ and hence the thicknesses of the various boundary layers and the corrections to the leading-order velocity.)

6.1 Comparison with experiments

We compare our theoretical results to the experimental results by Allshouse et al. [1]. In their experiments, the natural length scale was typically $L_0 = O(0.1 \text{ mm})$ while the wedge had size $L_1 \sim h = O(10 \text{ mm})$, so there was indeed a reasonable separation of scales with $\varepsilon = L_0/L_1 \ll 1$.

However, since they used wedges whose width-to-length ratio was typically 1/4 (and never larger than 2), our two-dimensional analysis which requires the wedges to be much wider than they are long does not apply. Also, many of their results are for a wedge with slope $\alpha = 5°$, which is very close to horizontal and again excluded by our analysis. Hence, we can only make a qualitative comparison with their results.

Overall, the experimental results indicate that the velocity c scales with the boundary-layer velocity $c_0 = \kappa \cot \alpha / L_0$, which is $1/\varepsilon$ times our scaling (6.1). This is evidenced by figures 3(a,b) in [1], which show that c does depends on the strength of the stratification, closely following the power law $c \propto N^{1/2}$, and that c/c_0 does not vary with the Schmidt number κ/ν. However, we note that the latter figure indicates that the numerical prefactor of the scaling is very small. In addition, the range of velocities in the former figure is limited to 1–4 μm/s, so the scaling result is not conclusive. We note that this range is comparable to the values 0.8–2.5 μm/s given by our scaling (6.1). Hence, further study of the three-dimensional case and the near-horizontal case is needed to conclusively decide which scaling is correct.

6.2 Extensions

Our analyses extend relatively easily to two-dimensional objects of other shapes. If the shape of the object is described by $S_-(z) < x < S_+(z)$, then the effective leading-order conditions $\psi = \cot \alpha$ on the wedge slope and $\psi = 0$ on the wedge back are replaced by $\psi = S'_\pm(z)$ at $x = S_\pm(z)$. The variation in slope allows fluid to be entrained or expelled continually along the surface of the object, rather than only at the corners of the wedge, but this is not a problem for the analyses in §3, §4 or §5.

If the object is not top-down symmetric, then the boundary conditions for the stream function will involve an unknown constant corresponding to the unknown circulation around the object. However, this constant is straightforwardly determined by requiring that the

hydrostatic pressure is continuous on both sides of the object. There may also be issues with rotation due to non-zero torque.

A more serious problem with objects of arbitrary shape is that they may have near-horizontal surfaces. In particular, a smooth object which does not have corners at the top and bottom will have horizontal surfaces there. As stated in §2.2, Phillips–Wunsch flow breaks down on surfaces that are too close to horizontal. Whether such top and bottom regions can be ignored for rounded shapes, just like we could ignore the corners of the wedge a leading order, remains to be seen.

As for extensions into the third dimension, a radically different theoretical approach would be required. This is because we have been relying on calculating the force on the wedge using the hydrostatic pressure, which is only possible due to the two-dimensional wedge acting as a dam between the front and rear regions. When fluid is allowed to flow around the wedge in the third dimension, a large pressure difference between the front and back wedge would simply drive such a flow around the wedge until the pressure force is balanced by viscous drag. Hence, the forces on the wedge are due to both pressure forces and viscous forces at the same order. Thus, it seems that a solution would have to involve solving a complicated equation in a complicated domain.

Nevertheless, based on our two-dimensional analysis we can hypothesize that the three-dimensional system is amenable to simplification by asymptotic decomposition. The Phillips–Wunsch boundary layer would be replaced by a line of sinks at the front of the wedge and two lines of sources at the back, and even though the outer equations have a complicated geometry and require numerical solution the computational cost would be greatly reduced due to the boundary layers not needing to be resolved.

Acknowledgements

I would like to thank Neil Balmforth and Bill Young for proposing this project and for our time together plodding through endless asymptotic expansions on the blackboards.

References

[1] M. R. Allshouse, M. F. Barad, and T. Peacock, *Propulsion generated by diffusion-driven flow*, Nature Physics, 6 (2010), pp. 516–519.

[2] M. R. Foster and P. G. Saffman, *The drag of a body moving transversely in a confined stratified fluid*, Journal of Fluid Mechanics, 43 (1970), pp. 407–418.

[3] D. D. Freund and R. E. Meyer, *On the mechanism of blocking in a stratified fluid*, Journal of Fluid Mechanics, 54 (1972), pp. 719–744.

[4] F. Hecht, *New development in freefem++*, Journal of Numerical Mathematics, 20 (2012), pp. 251–265.

[5] R. C. Y. Koh, *Viscous stratified flow towards a sink*, Journal of Fluid Mechanics, 24 (1966), pp. 555–575.

[6] D. W. Moore and P. G. Saffman, *The rise of a body through a rotating fluid in a container of finite length*, Journal of Fluid Mechanics, 31 (1968), pp. 635–642.

[7] ——, *The structure of free vertical shear layers in a rotating fluid and the motion produced by a slowly rising body*, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 264 (1969), pp. 597–634.

[8] M. A. Page, *Combined diffusion-driven and convective flow in a tilted square container*, Physics of Fluids (1994-present), 23 (2011).

[9] ——, *Steady diffusion-driven flow in a tilted square container*, The Quarterly Journal of Mechanics and Applied Mathematics, 64 (2011), pp. 319–348.

[10] M. A. Page and E. R. Johnson, *On steady linear diffusion-driven flow*, Journal of Fluid Mechanics, 606 (2008), pp. 433–443.

[11] ——, *Steady nonlinear diffusion-driven flow*, Journal of Fluid Mechanics, 629 (2009), p. 299.

[12] T. Peacock, R. Stocker, and J. M. Aristoff, *An experimental investigation of the angular dependence of diffusion-driven flow*, Physics of Fluids (1994-present), 16 (2004), pp. 3503–3505.

[13] O. M. Phillips, *On flows induced by diffusion in a stably stratified fluid*, Deep Sea Research and Oceanographic Abstracts, 17 (1970), pp. 435–443.

[14] Heinz Söhngen, *Die lösungen der integralgleichung* $g(x) = (1/2\pi)\int_{-a}^{a} f(\xi)/(x-\xi)\,d\xi$ *und deren anwendung in der tragflügeltheorie*, Mathematische Zeitschrift, 45 (1939), pp. 245–264.

[15] G. Veronis, *Analogous behavior of homogeneous, rotating fluids and stratified, non-rotating fluids*, Tellus, 19 (1967), pp. 326–336.

[16] ——, *Analogous behavior of rotating and stratified fluids*, Tellus, 19 (1967), pp. 620–634.

[17] ——, *The analogy between rotating and stratified fluids*, Annual Review of Fluid Mechanics, 2 (1970), pp. 37–66.

[18] C. Wunsch, *On oceanic boundary mixing*, Deep Sea Research and Oceanographic Abstracts, 17 (1970), pp. 293–301.

Noisy Homoclinic Pulse Dynamics

Tom Eaves

October 15, 2015

1 Near-homoclinic Pulse Dynamics

Near-homoclinic dynamics arise in a number of physically relevant problems in fluid dynamics, forexample, wherever there are coherent solitary waves, pulses in the signal of some important system variable, or bursts of turbulence in boundary layers. Near-homoclinic dynamics involve trajectories through phase space of a given dynamical system at parameter values close to a bifurcation at which there exits a homoclinic orbit, that repeatedly visit the vicinity of an unstable saddle point, at the origin say, followed by large excursions through the space, closely shadowing the saddle point's homoclinic orbit. If some variable like the energy or the distance from the saddle point is plotted against time, the resulting time series appears as a sequence of widely separated pulses.

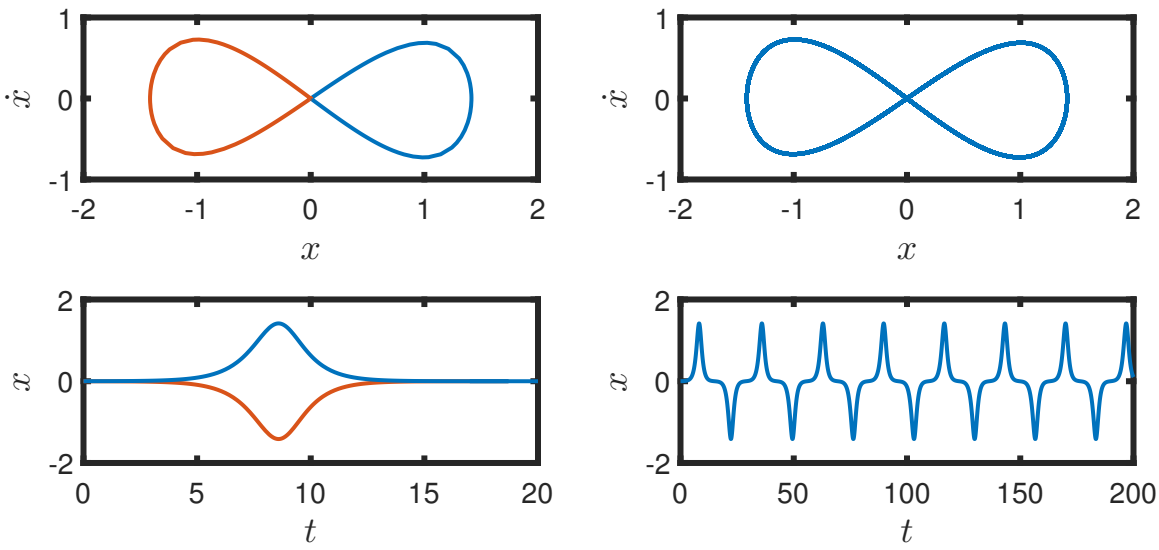

Figure 1: A deterministic two-dimensional system with a homoclinic orbit and its symmetric solution $x \mapsto -x$ (left) and a trajectory at nearby parameters (right). Top, trajectory in state space $(x(t), \dot{x}(t))$. Bottom, the homoclinic solution $x(t)$ and a sample train of pulses $x(t)$.

The exact form of such a 'pulse train' depends heavily on the type of saddle point that forms the start and end points of the homoclinic trajectory. In two dimensions, such

as the system plotted in Figure 1, both eigenvalues of the saddle point are real, and the time interval between pulses for the long-time asymptotic solution is constant. For a three-dimensional system like the Shimizu–Morioka system [14], which is a Lorenz-like system, for which the saddle point has three real eigenvalues $\lambda_2 < \lambda_1 < 0 < \lambda_3$, the strong contraction due to λ_2 can be ignored at leading order, and the dynamics appear at first sight to be similar to the two-dimensional case. However, the influence of the third stable direct acts to 'fold' trajectories together during their evolution through the vicinity of the origin, and chaotically distributed time intervals between pulses arises. The Shimizu–Morioka system is plotted in Figure 2.

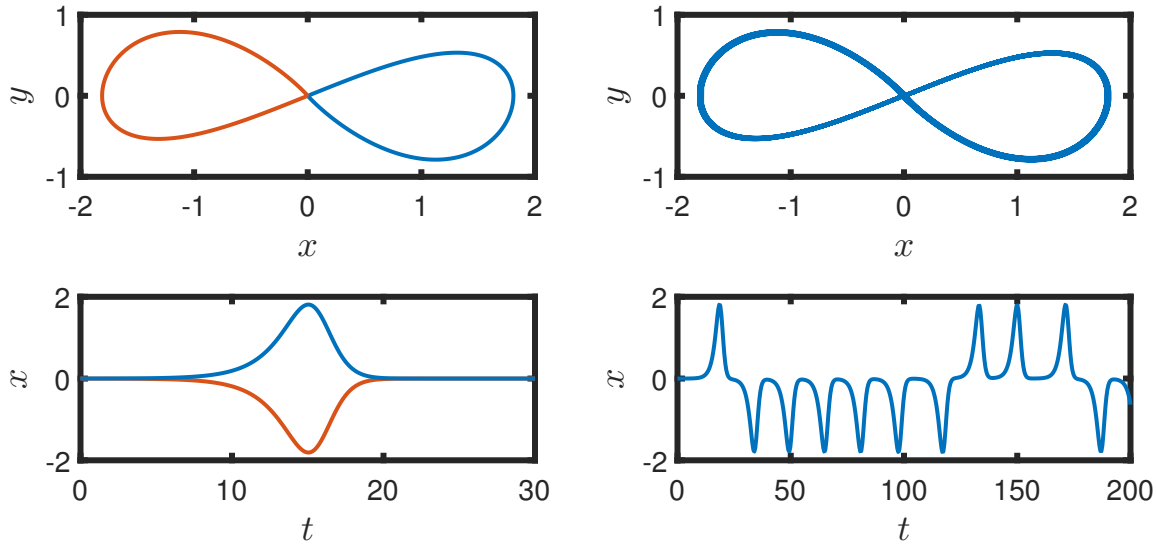

Figure 2: A deterministic three-dimensional Shimizu–Morioka system with a homoclinic orbit and its symmetric solution $(x, y) \mapsto (-x, -y)$ (left) and a trajectory at nearby parameters (right). Top, two-dimensional projection of trajectory in state space $(x(t), y(t))$. Bottom,the homoclinic solution $x(t)$ and a sample train of pulses $x(t)$..

Another canonical type of saddle point found in such systems is one for which the unstable direction has dimension one, and hence a single real positive eigenvalue, whilst the stable direction consists of a complex pair of eigenvalues. Their dynamics consist of trajectories spiraling into the saddle point before diverging away from the saddle point around its unstable direction. If this is combined with a strong contraction in one direction whilst following the homoclinic trajectory through state space, we call such systems 'Shilnikov' systems. These systems also display chaotically distributed time intervals between pulses, but in this case the chaos arises through the dynamics near the origin 'mixing up' trajectories that approach the origin through different locations, and thus completing a different number of spirals before leaving the vicinity of the saddle point. One such system is plotted in Figure 3.

Both Lorenz and Shilnikov type behaviour are observed in many physical systems. As was discussed in this year's GFD lectures by H. Dijkstra, the quasi-geostrophic double-gyre circulation undergoes bifurcations in an asymmetry variable, which measures the relative

occurrence of coherent structures above and below the symmetry line of forcing, to either Lorenz or Shilnikov phenomena, depending upon the parameters in the problem [9, 15].

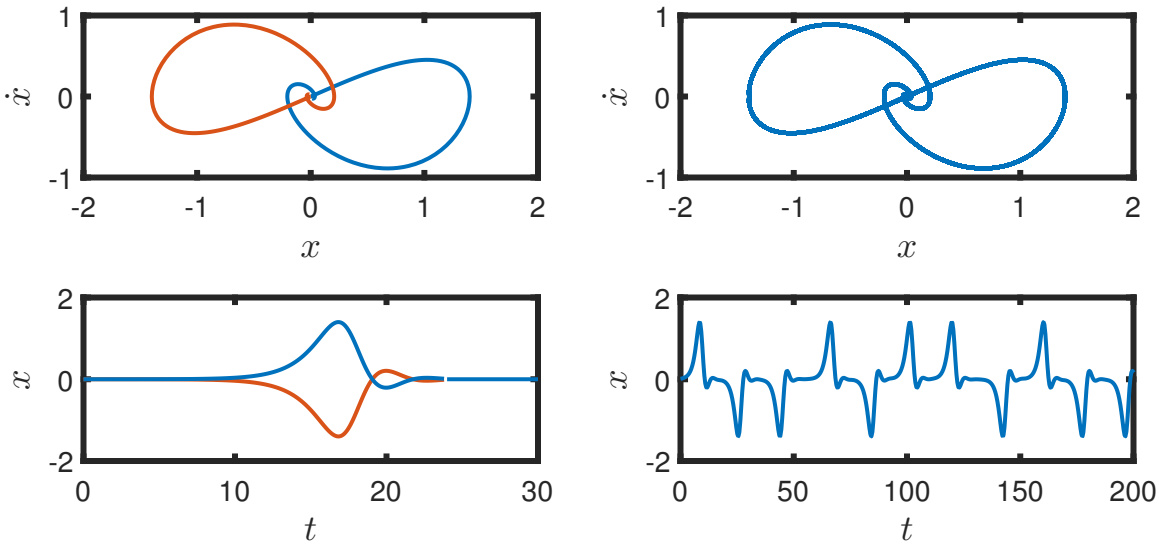

Figure 3: A deterministic three-dimensional Shilnikov system with a homoclinic orbit and its symmetric solution $x \mapsto -x$ (left) and a trajectory at nearby parameters (right). Top, two-dimensional projection of trajectory in state space $(x(t), \dot{x}(t))$. Bottom, the homoclinic solution $x(t)$ and a sample train of pulses $x(t)$.

The time intervals between pulses in all such systems depend most sensitively on how the trajectories evolve through the origin, as the closer a trajectory is to the stable manifold when approaching the origin, the longer it takes to leave the origin, and this time increases exponentially as the distance to the stable manifold decreases. In contrast, the 'outer' flow away from the origin takes an approximately constant time for all trajectories. The time intervals are therefore controlled by the position at which trajectories enter the vicinity of the origin.

Two successful approaches have been suggested to estimate the distribution of time intervals between pulses, or equivalently the mapping between trajectories leaving the origin and re-entering the origin. The approach taken by Shilnikov [13], after whom Shilnikov systems are named, is to solve the linearised system near the origin exactly, and to assume a linear mapping between points on a Poincaré section of trajectories leaving the origin to points on a Poincaré section of trajectories returning to the origin. This represents stong contraction in one of the spatial dimensions during the outer flow, and an outer flow that closely follows the homoclinic orbit, and is explained in more intuitive detail by Glendinning and Sparrow [7]. This can then be written as a single map between points on the entrance Poinaré section as

$$Z_{n+1} = C + BZ_n^{\delta} \cos\left(\Omega \log\left(\frac{Z_0}{Z_n}\right) + \Phi\right), \tag{1}$$

for some constant B, and C is a constant that measures the distance from a bifurcation point at which the homoclinic orbit exists, Φ is a constant phase, Ω is related to the unstable

eigenvalue and the imaginary part of the stable eigenvalue, and δ is the ratio of the real part of the stable eigenvalue to the unstable eigenvalue. When $C = 0$ and $\delta < 1$ this map has an infinite number of unstable fixed points, and exhibits chaos for most parameter values.

The approach taken by Balmforth *et al.* [4] and summarised a review paper [3], is to suppose that a solution to the equations in near homoclinic conditions can be written as a sum of homoclinic trajectories with an error term of $O(\epsilon)$, where ϵ measures the distance from homoclinicity. These homoclinic trajectories must then be sufficiently widely separated in time so that the interaction of adjacent homoclinic orbits is of size $O(\epsilon)$. An equation for the error term is derived, and a secularity condition deduced from the requirement that the error term is small. This secularity condition involves interactions of adjacent homoclinic trajectories, and hence can be interpreted as a mapping between successive spacings in time of the homoclinic trajectories. Under further approximation, this map can be reduced to the Shilnikov Z_n map. This method is the primary method generalised here to stochastic systems, and its details are left until later in the text.

Stochastic near-homoclinic systems have been considered in a variety of contexts, from noise-driven excitable systems [11, 5], to exponential tails in the timing of turbulence bursting events [18] and intermittent switching between cycles in a heteroclinic network [1]. The only major theoretical work comes from Stone and Holmes [17] and extensions [18, 16]. This work deals with the stochastic dynamics near the origin only, and solves the Ornstein–Uhlenbeck processes there, followed by assumptions of long residency time near the origin and small amplitude noise to find simple expressions for the density of points leaving and arriving at the origin, along with expected residency times. The key assumption in all of this work is that since the noise amplitude is small, its only significant effect is near to the origin, and that once a trajectory is closely following the homoclinic orbit away from the origin, the effects of noise can be neglected, and the distribution of points leaving the origin may be linearly mapped to a distribution of points arriving at the origin via the deterministic linear mapping, as in Shilnikov's approach for deterministic systems.

The work presented here demonstrates that this assumption is in fact unfounded, and we show that for the three different homoclinic systems discussed above, the primary effect of noise on a homoclinic trajectory is the influence during the orbit away from the origin producing a significant difference in the distribution of points arriving at the origin, and that this effect produces variances in the position and timing of trajectories that are at least an order of magnitude larger than the effects near the origin. Section 2 deals with the two-dimensional Duffing system originally considered in the theoretical paper of Stone and Holmes [17]. Section 3 deals with the three-dimensional Shimizu–Morioka system [14], which can be interpreted as the Lorenz system at high Rayleigh number and exhibits Lorenz-like behaviour. Section 4 deals with a three-dimensional equation that arises as the normal form of a co-dimension three bifurcation with a reflection symmetry [2] in which three eigenvalues of an equilibrium point simultaneously have zero real part, and exhibits Shilnikov-like behaviour. We draw our conclusions in Section 5.

2 A Stochastic Duffing Equation

Consider the following deterministic Duffing equation for $x(t)$

$$\ddot{x} = x - x^3 - \epsilon\gamma\dot{x} + \epsilon\beta x^2\dot{x}. \tag{2}$$

A sample trajectory with $\gamma = 0.08$ and $\beta = 0.1$ is shown in Figure 1, along with the nearby homoclinic trajectory at $\gamma \approx 0.080012$.

This equation has a fixed point at the origin, which is a saddle, with eigenvalues

$$\lambda = \frac{1}{2}(-\epsilon\gamma \pm \sqrt{\epsilon^2\gamma^2 + 4}) = \pm 1 - \frac{\epsilon\gamma}{2} + O(\epsilon^2), \tag{3}$$

with corresponding eigenvectors

$$\boldsymbol{v}_\pm = \begin{pmatrix} \pm 1 - \frac{\epsilon\gamma}{2} + O(\epsilon^2) \\ 1 \end{pmatrix}. \tag{4}$$

The system also has two addition fixed points at $(x, \dot{x}) = (\pm\sqrt{1 - \epsilon\gamma}, 0)$, which are unstable/stable sinks if $\beta \gtrless 0$.

Consider the addition of a white noise process to this equation, namely

$$\dot{x} = y + \epsilon\xi_x, \tag{5}$$

$$\dot{y} = x - x^3 - \epsilon\gamma y + \epsilon\beta x^2 y + \epsilon\xi_y, \tag{6}$$

where

$$\mathbb{E}(\xi_{x,y}(t)) = 0, \tag{7}$$

$$\mathbb{E}(\xi_{x,y}(t)\xi_{x,y}(s)) = \delta(t - s), \tag{8}$$

and ξ_x and ξ_y are independent.

Let δ be such that $\epsilon \ll \delta \ll 1$ and consider the small domain about the origin $\mathcal{D} = \{(x, \dot{x}) \cdot \hat{\boldsymbol{v}}_\pm \leq \delta\}$ where $\hat{\boldsymbol{v}}_\pm$ are the normalised eigenvectors of the saddle point at the origin. Within $\mathcal{D}$ we can approximate the dynamics, in local saddle coordinates, as

$$\dot{x}_1 = -\mu x_1 + \epsilon\xi_{x_1}, \tag{9}$$

$$\dot{x}_2 = \lambda x_2 + \epsilon\xi_{x_2}, \tag{10}$$

where $-\mu$ and λ are respectively the stable and unstable eigenvalues of the saddle point at the origin.

We can then consider three cases for the action of noise, namely

1. $\xi_{x_1} \neq 0$, $\xi_{x_2} \neq 0$ and $\xi = 0$ outside of $\mathcal{D}$ (noisy origin)
2. $\xi_{x_1} = 0$, $\xi_{x_2} = 0$ and $\xi \neq 0$ outside of $\mathcal{D}$ (deterministic origin)
3. $\xi_{x_1} \neq 0$, $\xi_{x_2} \neq 0$ and $\xi \neq 0$ outside of $\mathcal{D}$ (noise everywhere)

In addition, we may split case (1.) into two, namely $\xi_{x_1} \neq 0$ and $\xi_{x_2} = 0$ (noisy stable direction) and $\xi_{x_1} = 0$ and $\xi_{x_2} \neq 0$ (noisy unstable direction). This is done in Appendix A.

Before considering each of the three cases in turn, we first solve the deterministic behaviour within $\mathcal{D}$.

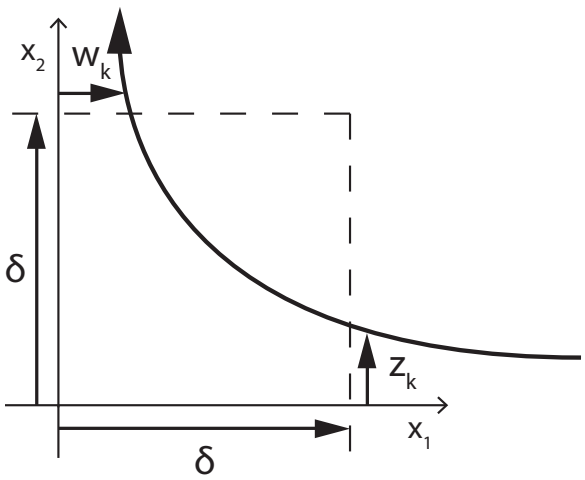

Figure 4: Sketch of the dynamics through the domain $\mathcal{D}$ about the origin with entry point z_k and leaving point w_k.

2.1 Deterministic origin dynamics

When a trajectory enters $\mathcal{D}$ for the kth occasion, at time $t = t_0$, we have $(x_1, x_2) = (\delta, z_k)$, see Figure 4. The trajectory then satisfies

$$x_1 = \delta e^{-\mu(t-t_0)}, \tag{11}$$

$$x_2 = z_k e^{\lambda(t-t_0)}. \tag{12}$$

The trajectory then leaves $\mathcal{D}$ when $(x_1, x_2) = (w_k, \delta)$, which occurs after a time interval

$$t_k \equiv t_{\text{leave}} - t_0 = \frac{1}{\lambda} \log\left(\frac{\delta}{z_k}\right), \tag{13}$$

from which we can obtain w_n, the x_1 position at time of exit,

$$w_k = \delta^{1-\mu/\lambda} z_k^{\mu/\lambda}. \tag{14}$$

We then make the assumption that the dynamics outside of $\mathcal{D}$ act as a linear mapping between the point at which $\mathcal{D}$ is left for the kth time to the point at which $\mathcal{D}$ is entered for the $(k+1)$th time, as did Shilnikov [13], i.e. that

$$z_{k+1} = \alpha w_k + c = \alpha \delta^{1-\mu/\lambda} z_k^{\mu/\lambda} + c, \tag{15}$$

where $c = 0$ when there exists a homoclinic orbit.

Since $z_k < \delta < 1$, we see that when $\mu/\lambda > 1$, we have the limit $z_k \to c$ as $k \to \infty$. We can also solve for the time intervals between entering events. First note that

$$t_{k+1} = -\frac{1}{\lambda} \log\left(\frac{z_{k+1}}{\delta}\right) = -\frac{1}{\lambda} \log(\alpha \delta^{-\mu/\lambda} z_k^{\mu/\lambda} + c/\delta) = -\frac{1}{\lambda} \log(\alpha e^{-\mu t_k} + c/\delta). \tag{16}$$

Assuming that the time between exit and re-entrance of $\mathcal{D}$ is T, the time interval between entering events $\Delta_k = T + t_k$ satisfies

$$\Delta_{k+1} = T - \frac{1}{\lambda}\log(\alpha e^{-\mu\Delta_k} e^{\mu T} + c/\delta). \tag{17}$$

When there exists a homoclinic orbit, $c = 0$, and so

$$\Delta_{k+1} = T - \frac{\log(\alpha)}{\lambda} - \frac{\mu T}{\lambda} + \frac{\mu}{\lambda}\Delta_k, \tag{18}$$

and so if $\mu/\lambda > 1$, we have the limit $\Delta_k \to \infty$ as $n \to \infty$, and so the homoclinic orbit is attracting.

2.2 Noisy origin, $\xi = 0$ outside $\mathcal{D}$

The analysis in this section is a summary of the work of Stone and Holmes [17] and their approach to determining the dynamics for noise near the origin. In Appendix A we adopt the same approach as Stone and Holmes to investigate the effect of noise near the origin only in either the stable or unstable direction of the saddle. We wish to solve the stochastic differential equations

$$\dot{x}_1 = -\mu x_1 + \epsilon\xi_{x_1}, \tag{19}$$

$$\dot{x}_2 = \lambda x_2 + \epsilon\xi_{x_2}, \tag{20}$$

where λ and $-\mu$ are the unstable and stable eigenvalues at the origin.

First, we define the Gaussian function for the normal probability density function

$$\mathcal{N}_x(a, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]. \tag{21}$$

Then, the known solutions to the Ornstein–Uhlenbeck processes (19-20), given an initial known position are

$$\rho(x_1, t|\delta, 0) = \mathcal{N}_{x_1}\left(\delta e^{-\mu t}, \frac{\epsilon^2}{2\mu}(1 - e^{-2\mu t})\right) \tag{22}$$

$$\rho(x_2, t|z_n, 0) = \mathcal{N}_{x_2}\left(z_k e^{\lambda t}, \frac{\epsilon^2}{2\lambda}(e^{2\lambda t} - 1)\right) \tag{23}$$

Observe that the means satisfy the equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x_1\rangle = -\mu\langle x_1\rangle \tag{24}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle x_2\rangle = \lambda\langle x_2\rangle, \tag{25}$$

and so we might expect that if the deterministic system spends a large amount of time in $\mathcal{D}$, then so will the stochastic system. This allows us to make the approximation that the x_1-process becomes stationary before exiting $\mathcal{D}$, i.e. that $e^{-\mu t} \ll 1$, and so we have

$$\rho(x_1, t|\delta, 0) \sim \mathcal{N}\left(0, \frac{\epsilon^2}{2\mu}\right) \quad \text{as } t \to \infty, \tag{26}$$

and so the leaving point w_k has a distribution independent of z_k,

$$\rho(w_k) = \mathcal{N}\left(0, \frac{\epsilon^2}{2\mu}\right). \tag{27}$$

After leaving $\mathcal{D}$, the motion is deterministic, and so we recall the linear mapping

$$z_{k+1} = \alpha w_k + c, \tag{28}$$

giving the density of return points z_{k+1},

$$\rho(z_{k+1}|z_k) = \rho(z_{k+1}) = \mathcal{N}\left(c, \frac{\epsilon^2\alpha^2}{2\mu}\right), \tag{29}$$

using standard results for linear combinations of normally distributed random variables.

To find the timings t_k, we need to solve the unstable Y process. It is also known that the solution to the Ornstein–Uhlenbeck process (20) given an initial known *normal distribution* is

$$\rho(x_2, t|\mathcal{N}(x_2^0, \sigma^2), 0) = \mathcal{N}\left(x_2^0 e^{\lambda t}, \sigma^2 e^{2\lambda t} + \frac{\epsilon^2}{2\lambda}(e^{2\lambda t} - 1)\right), \tag{30}$$

and so, we see that

$$\rho\left(x_2, t\left|\mathcal{N}\left(ce^{\lambda t}, \frac{\epsilon^2\alpha^2}{2\mu}\right), 0\right.\right) = \mathcal{N}\left(ce^{\lambda t}, \frac{\epsilon^2\alpha^2}{2\mu}e^{2\lambda t} + \frac{\epsilon^2}{2\lambda}(e^{2\lambda t} - 1)\right) \tag{31}$$

$$= \mathcal{N}\left(ce^{\lambda t}, \frac{\epsilon^2}{2\lambda}(e^{2\lambda s} - 1)\right) \tag{32}$$

$$\equiv \rho(x_2, s|s = 0), \tag{33}$$

where $s = t + t'$ and

$$t' = \frac{1}{2\lambda}\log\left(1 + \frac{\alpha^2\lambda}{\mu}\right). \tag{34}$$

Then, the mean passage time $\mathbb{E}(t_k)$ satisfies

$$\mathbb{E}(t_k) = \int_0^\infty \mathbb{P}(t_k > t)\ \mathrm{d}t \tag{35}$$

$$= \int_{t'}^\infty \mathbb{P}(t_k > s)\ \mathrm{d}s \tag{36}$$

$$= \int_{t'}^\infty \int_{-\delta}^{\delta} \rho(x_2, s|s = 0)\ \mathrm{d}x_2\ \mathrm{d}s \tag{37}$$

$$\sim \begin{cases} \frac{1}{\lambda}\log\left(\frac{\delta}{\epsilon}\right) & \text{for } c \ll \epsilon \ll \delta \ll 1, \\ \frac{1}{\lambda}\log\left(\frac{\delta}{c}\right) & \text{for } \epsilon \ll c \ll \delta \ll 1. \end{cases} \tag{38}$$

The main result here is that for nearly homoclinic conditions in which $c \ll \epsilon$, noise acts to effectively push a trajectory away from the stable axis to a distance ϵ, after which the mean trajectory is essentially deterministic, whereas far from homoclinicity in which $\epsilon \ll c$, noise is unimportant, and the mean trajectory is identical to its deterministic version. The

extra condition that $\epsilon, c \ll \delta$ ensures that trajectories remain in $\mathcal{D}$ for a long time, and so the assumption $e^{-\mu t_k} \ll 1$ is valid.

Parameter values considered by Stone and Holmes [17] are $\gamma = 0.08$, $\beta = 0.1$ and $\epsilon = 0.0006$. Using these values, we can compare the results above to a numerical solution of the stochastic differential equations with $\xi = 0$ outside of $\mathcal{D}$, defined by $\delta = 0.1$. Figure 5 shows a comparison of the distributions for $\rho(w_k)$ and the distributions for $\rho(z_k)$ for both a numerical integration of the SDEs and the derived results (27,29) given by Stone and Holmes [17] with $c = 0$. There is a clear discrepancy in these results, and so we also plot the sum of two normal distributions with means $\pm\delta \exp(-\mu \mathbb{E}(t_k))$ for w_n where $\mathbb{E}(t_n)$ is taken from the numerical integration, and an approximate offset $c = \pm 2.763 \times 10^{-4}$ for z_n taken from the numerical integration. The mean time through $\mathcal{D}$ for the numerical solution is 5.767 and the mean time according to (38) with $c \ll \epsilon$ is 5.325. There remains an error in the comparison of the results for $\rho(w_k)$, likely due to the fact that the approximation for $\rho(w_k)$ is independent of z_k.

We note that we can break from the approach of Stone and Holmes in this section and Appendix A. Rather than making approximation about long residency times to simplify the relevant probability densities, we can instead keep the exact solutions to the Ornstein–Uhlenbeck processes and iteratively produce a sample trajectory. To obtain z_{k+1} from z_k, we choose a t_k and w_k from the relevant densities, and then map w_k to z_{k+1} using the linear mapping assumption. This would give a reliable, if low-tech way of producing the correct comparisons to the SDEs.

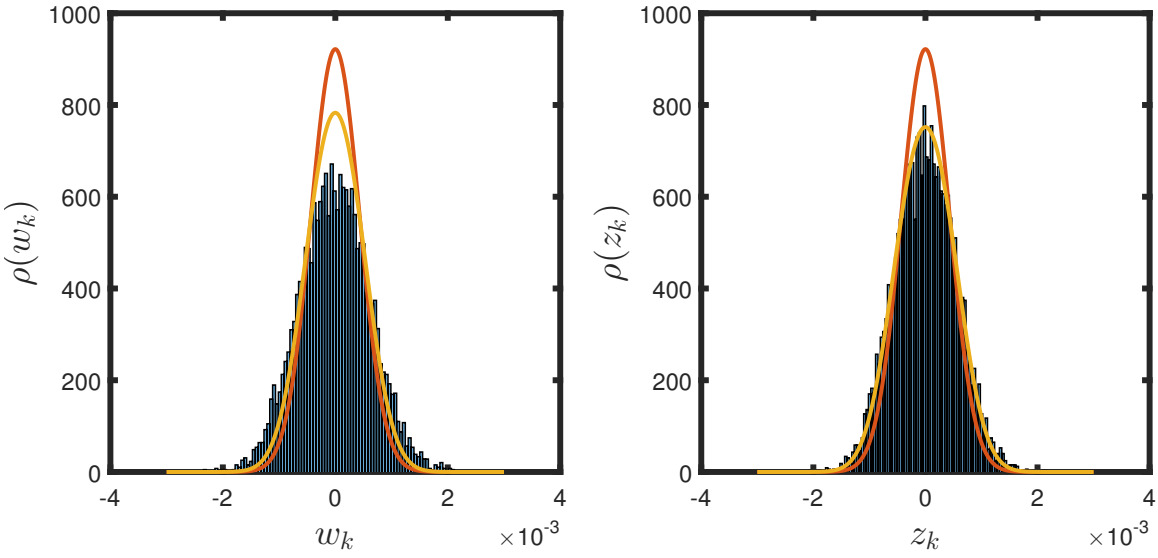

Figure 5: Probability density of leave points w_n (left) and return (right) points z_n to $\mathcal{D}$ with $\delta = 0.1$ and noise strength $\epsilon = 0.0006$ acting only outside $\mathcal{D}$, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for a direct numerical simulation with 7500 returns to $\mathcal{D}$ (blue histogram), the approximate solution (27,29) from Stone and Holmes [17] (red line), and the sum of two normal distributions with means $\pm\delta \exp(-\mu \mathbb{E}(t_n))$ for w_n and $\pm 2.763 \times 10^{-4}$ for z_n (orange line).

2.3 Deterministic origin, $\xi \neq 0$ outside $\mathcal{D}$

Consider the equation

$$\ddot{x} = x - x^3 - \epsilon\gamma\dot{x} + \epsilon\beta x^2\dot{x} + \epsilon\xi(t) \tag{39}$$

outside the domain $\mathcal{D} \approx \{|\dot{x}+x| \leq \delta,\ |\dot{x}-x| \leq \delta\}$. Write the solution as $x = x_0+\epsilon x_1+O(\epsilon^2)$. Then, at $O(1)$ we have

$$\ddot{x}_0 = x_0 - x_0^3, \tag{40}$$

which has first integral

$$\frac{1}{2}\dot{x}_0^2 = \frac{1}{2}x_0^2 - \frac{1}{4}x_0^4 + E, \tag{41}$$

where we can interpret the constant of integration E as an energy parameter. The case $E = 0$ gives the homoclinic orbits

$$x_0 = \pm\sqrt{2}\mathrm{sech}(t - t_0) \tag{42}$$

for some t_0. We will concentrate on the positive solution, $x_0 = \sqrt{2}\mathrm{sech}(t - t_0)$.

Next, at $O(\epsilon)$ we obtain

$$\ddot{x}_1 - x_1 + 3x_0^2x_1 = -\gamma\dot{x}_0 + \beta x_0^2\dot{x}_0 + \xi(t). \tag{43}$$

Multiply through by $\dot{x}_0$ and integrate to get

$$\left[\dot{x}_0\dot{x}_1 - x_0x_1 + x_0^3x_1\right]_{t_a}^{t_b} = -\gamma\int_{t_a}^{t_b}\dot{x}_0^2\ \mathrm{d}t + \beta\int_{t_a}^{t_b}x_0^2\dot{x}_0^2\ \mathrm{d}t + \int_{t_a}^{t_b}\xi(t)\dot{x}_0\ \mathrm{d}t, \tag{44}$$

where t_a and t_b are defined respectively by

$$\dot{x} + x = \sqrt{2}\delta, \tag{45}$$

$$\dot{x} - x = -\sqrt{2}\delta. \tag{46}$$

For $\epsilon \ll \delta \ll 1$, we have $x_0 \approx 2\sqrt{2}e^{-|\hat{t}|}$, where $\hat{t} = t - t_0$. This then gives $\dot{x}_0 \approx -2\sqrt{2}\ \mathrm{sgn}(\hat{t})e^{-|\hat{t}|}$, and so the leading order condition on $\partial\mathcal{D}$ determines t_a and t_b, namely

$$\dot{x}_0(t_a) = x_0(t_a) = \frac{\sqrt{2}\delta}{2}, \tag{47}$$

$$\dot{x}_0(t_b) = -x_0(t_b) = -\frac{\sqrt{2}\delta}{2}, \tag{48}$$

$$t_b = -t_a = \log\left(\frac{4}{\delta}\right). \tag{49}$$

At second order we obtain

$$\dot{x}_1(t_a) = -x_1(t_a), \tag{50}$$

$$\dot{x}_1(t_b) = x_1(t_b), \tag{51}$$

and so

$$\left[\dot{x}_0\dot{x}_1 - x_0x_1 + x_0^3x_1\right]_{t_a}^{t_b} = \sqrt{2}\delta(x_1(t_a) - x_1(t_b)) + O(\delta^3). \tag{52}$$

We also have the exact relations

$$-\gamma \int_{-\infty}^{\infty} \dot{x}_0^2 = -\frac{4\gamma}{3}, \tag{53}$$

$$\beta \int_{-\infty}^{\infty} x_0^2 \dot{x}_0^2 \, \mathrm{d}t = \frac{16\beta}{15}, \tag{54}$$

and so, given that $-t_a, t_b \gg 1$, the deterministic orbit is homoclinic at leading order if these two integrals are equal, i.e. $\beta = 5\gamma/4$. Assuming homoclinic conditions, we obtain

$$\sqrt{2}\delta(x_1(t_a) - x_1(t_b)) = \int_{-\infty}^{\infty} \xi(t)\dot{x}_0 \, \mathrm{d}t. \tag{55}$$

It is a known result [6] that for deterministic functions of time $f(t)$,

$$\int_a^b f(t)\xi(t) \, \mathrm{d}t \sim \mathcal{N}\left(0, \int_a^b f(t)^2 \, \mathrm{d}t\right). \tag{56}$$

We can use this result to deduce that

$$x_1(t_b) \sim \mathcal{N}\left(x_1(t_a), \frac{1}{2\delta^2}\int_{-\infty}^{\infty} \dot{x}_0^2 \, \mathrm{d}t\right) \tag{57}$$

$$= \mathcal{N}\left(x_1(t_a), \frac{2}{3\delta^2}\right). \tag{58}$$

The leading order return map to $\mathcal{D}$ for this stochastic Duffing equation in homoclinic conditions is then,

$$z_{k+1} = \left.\frac{\dot{x} + x}{\sqrt{2}}\right|_{t_b} \tag{59}$$

$$= \sqrt{2}\epsilon x_1(t_b) \tag{60}$$

$$= \mathcal{N}\left(\sqrt{2}\epsilon x_1(t_a), \frac{4\epsilon^2}{3\delta^2}\right) \tag{61}$$

$$= \mathcal{N}\left(w_k, \frac{4\epsilon^2}{3\delta^2}\right) \tag{62}$$

$$= \mathrm{sgn}(z_k)\delta^{1-\mu/\lambda}|z_k|^{\mu/\lambda} + \frac{\epsilon}{\delta}\sqrt{\frac{4}{3}}\eta_k, \tag{63}$$

where $\eta_k \sim \mathcal{N}(0,1)$ is a zero mean, unit variance, Gaussian random variable. Hence,

$$m = \mathbb{E}(z_{k+1}|z_k) = \mathrm{sgn}(z_k)\delta^{1-\mu/\lambda}|z_n|^{\mu/\lambda} \tag{64}$$

$$\mathbb{E}((z_{k+1} - m)^2|z_k) = \frac{4\epsilon^2}{3\delta^2}. \tag{65}$$

Recall that the variance in the re-entry point for noise only within $\mathcal{D}$ scaled as ϵ^2. Given $\epsilon \ll \delta \ll 1$, we see that the variance associated with this outer flow is much larger.

We then have the transition density

$$\rho(z_{k+1}|z_k) = \mathcal{N}\left(\mathrm{sgn}(z_k)\delta^{1-\mu/\lambda}|z_k|^{\mu/\lambda}, \frac{4\epsilon^2}{3\delta^2}\right), \tag{66}$$

and so

$$\rho(z_{k+1}) = \int_{-\infty}^{\infty} \rho(z_{k+1}|z_k)\rho(z_k)\mathrm{d}z_k. \tag{67}$$

We can look for the stationary distribution $\rho(z_k) = \rho^s(z)$ that the mapping approaches as $n \to \infty$. This is given by the integral equation

$$\rho^s(z) = \int_{-\infty}^{\infty} \frac{\delta}{\epsilon}\sqrt{\frac{3}{8\pi}} \exp\left[-\frac{3\delta^2(z - \mathrm{sgn}(s)\delta^{1-\mu/\lambda}|s|^{\mu/\lambda})^2}{8\epsilon^2}\right] \rho^s(s)\ \mathrm{d}s. \tag{68}$$

Figure 6 shows the return distribution for noise outside of $\mathcal{D}$ only and noise uniformly everywhere in the system respectively. Also shown is the mapping (63). For noise outside of $\mathcal{D}$ we also plot the estimated stationary distribution via the method described below, and for the 'noise everywhere' calculation we also plot the result from Stone and Holmes.

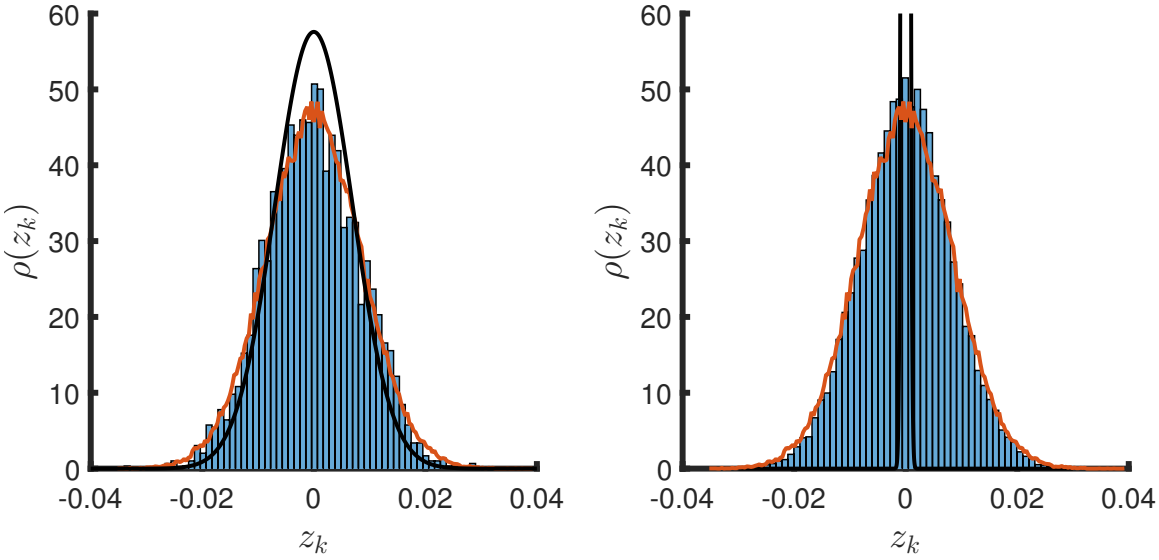

Figure 6: Left: Probability density of return points to $\mathcal{D}$ with $\delta = 0.1$ and noise strength $\epsilon = 0.0006$ acting only outside $\mathcal{D}$, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for a direct numerical simulation with 3000 returns to $\mathcal{D}$ (blue histogram), the return map (63) iterated 10^5 times (red line) and the asymptotic approximation to the stationary distribution of the return map (63) (black line). Right: Noise strength $\epsilon = 0.0006$ acting everywhere, and the result of Stone and Holmes [17] (black line).

Before computing additional properties of this solution, we need some results about stationary distributions about stable fixed points of maps.

2.4 Weak noise estimation of stationary distributions

We can estimate stationary distributions for stochastic mappings with stable deterministic fixed points or periodic orbits [10]. Suppose that we have the stochatic mapping

$$z_{n+1} = f(z_n) + \epsilon\eta_n, \tag{69}$$

where $\eta_n \sim \mathcal{N}(0,1)$. Then we know that

$$\rho(z_{n+1}|z_n) = \mathcal{N}(f(z_n), \epsilon^2), \tag{70}$$

and so the stationary distribution statisfies

$$\rho^s(z) = \int \rho(z|s)\rho^s(s)\ \mathrm{d}s. \tag{71}$$

Suppose in addition that (for the simplest possible case) the deterministic part of the mapping has a singe stable fixed point, $f(z_*) = z_*$ with $|f'_*| \equiv |f'(z_*)| < 1$. Then we might suppose that for small noise strength ϵ, the stationary distribution is centred about z_* with some variance σ^2 which in general is different from, but related to, ϵ, and should depend on the local rate of contraction of $f(z)$ about the fixed point.

To this end, try the ansatz

$$\rho^s(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-z_*)^2}{2\sigma^2}\right] \tag{72}$$

$$= \int \frac{1}{2\pi\epsilon\sigma} \exp\left[-\frac{(z-f(s))^2}{2\epsilon^2} - \frac{(s-z_*)^2}{2\sigma^2}\right] \mathrm{d}s \tag{73}$$

Now let $s = z_* + \epsilon S$ and $z = z_* + \epsilon Z = f(z_*) + \epsilon Z$. Then, defining $\alpha^2 = \epsilon^2/\sigma^2$, we obtain

$$\rho^s(z) = \int \frac{1}{2\pi\sigma} \exp\left[-\frac{(Z - f'_* S)^2}{2} - \frac{\alpha^2 S^2}{2}\right] \mathrm{d}S \tag{74}$$

$$\equiv \int \frac{1}{2\pi\sigma} \exp(\phi(S))\ \mathrm{d}S. \tag{75}$$

Now, the function $\phi(S)$ has a minima at

$$S = S_0 = \frac{f'_*}{\alpha^2 + f'^2_*} Z, \tag{76}$$

and we may approximate

$$\phi(S) = \phi(S_0) + \frac{(S-S_0)^2}{2}\phi''(S) \tag{77}$$

$$= -\frac{\alpha^2}{\alpha^2 + f'^2_*}\frac{Z^2}{2} - (f'^2_* + \alpha^2)\frac{(S-S_0)^2}{2}. \tag{78}$$

Substituting this back into the integral and evaluating the resulting Gaussian, we obtain

$$\rho^s(z) = \frac{1}{2\pi\sigma}\sqrt{\frac{2\pi}{f_*'^2+\alpha^2}}\exp\left[-\frac{\alpha^2}{\alpha^2+f_*'^2}\frac{Z^2}{2}\right] \tag{79}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2(f_*'^2+\alpha^2)}}\exp\left[-\frac{\alpha^2}{\alpha^2+f_*'^2}\frac{(z-z_*)^2}{2\epsilon^2}\right], \tag{80}$$

which is consistent with the original ansatz if $\sigma^2 = \epsilon^2/(1-f_*'^2)$, and this modified variance σ^2 is positive provided that the fixed point is stable. Hence, an approximate solution is

$$\rho^s(z) = \frac{1}{\sqrt{2\pi\epsilon^2/(1-f_*'^2)}}\exp\left[-\frac{(z-z_*)^2}{2\epsilon^2/(1-f_*'^2)}\right]. \tag{81}$$

Now consider a stable N-cycle $\{z_i\}_{i=1}^N$ distinct points with $f(z_i) = z_{i+1}$. Define $f_i' = f'(z_i)$. Then, we pose the ansatz for the stationary distribution

$$\rho^s(z) = \sum_{i=1}^N \frac{a_i}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{(z-z_i)^2}{2\sigma_i^2}\right], \tag{82}$$

with $a_i > 0$ and $\sum a_i = 1$ so that this represents a probability distribution. An analogous calculation to the one above gives

$$\rho^s(z) = \sum_{i=1}^N \frac{a_i}{\sqrt{2\pi\sigma_i^2(f_i'^2+\alpha^2)}}\exp\left[-\frac{\alpha_i^2}{\alpha_i^2+f_i'^2}\frac{(z-z_i)^2}{2\epsilon^2}\right], \tag{83}$$

where $\alpha_i^2 = \epsilon^2/\sigma_i^2$.

One of the compatibility conditions gives $a_i = a_{i+1}$, and so $a_i = 1/N$. The other compatibility condition gives

$$\sigma_i^2 = \epsilon^2\frac{1+\sum_{k=1}^{N-1}\prod_{j=1}^k f_{i-j}'^2}{1-\prod_{j=1}^N f_j'^2}, \tag{84}$$

where the subscripts are taken modulo N, which reduces to the fixed point case if $N = 1$.

For the Duffing equation considered above in homoclinic conditions, we have $z_* = 0$ and $f_*' = 0$, and so the estimated variance of the stationary distribution is the same as that for the noise term in the equation. This estimate is plotted in Figure 6 along with the numerical simulation of the full system, and the map iterated a large number of times (10^5). The agreement is not particularly good, but this is due to the fact that $z_* \neq 0$ for the parameter values chosen by Stone and Holmes [17]. We fix this later.

2.5 Timing map for deterministic origin, $\xi \neq 0$ outside $\mathcal{D}$

We can also compute the time between pulses, or equivalently the time between re-entries of $\mathcal{D}$. The total flight time is the sum of the time taken to pass through the origin plus the time taken between leaving and returning to $\mathcal{D}$,

$$\Delta_n = T_{\text{origin}} + T_{flight} = \frac{1}{\lambda}\log\left(\frac{\delta}{|z_n|}\right) + 2t_b = \frac{1}{\lambda}\log\left(\frac{\delta}{|z_n|}\right) + 2\log\left(\frac{4}{\delta}\right). \tag{85}$$

The timing map should then be independent of δ. Substituting Δ_n into the return map (63) we obtain

$$e^{-\lambda\Delta_{n+1}} = \delta^{2(\lambda-\mu)}16^{\mu-\lambda}e^{-\mu\Delta_n} + \frac{\delta^{2\lambda-2}\epsilon}{16^{\lambda}}\sqrt{\frac{4}{3}}\eta_n \tag{86}$$

$$= e^{-\mu\Delta_n} + \frac{\epsilon}{16}\sqrt{\frac{4}{3}}\eta_n, \tag{87}$$

at leading order, since $\lambda, \mu = 1 \pm \epsilon\gamma/2$, and so δ drops out. Note however that we must keep λ and μ different from one in the exponentials as contraction is necessary for bounded solutions. This can be seen be noting that if $\lambda = \mu = 1$ in the return map (63), then z_n satisfies a Wiener process with variance growing linearly with the number of iterations.

We can also demonstrate that the timing map has exponential tails. To see this, recall that the return map (63) has a stable fixed point at $z_n = 0$ and so the approximations of the previous section imply that we have the stationary distribution

$$\rho_z(z) \approx \mathcal{N}\left(0, \frac{4\epsilon^2}{3\delta^2}\right). \tag{88}$$

Then, the change of variables to Δ gives

$$\rho_\Delta(\Delta) \approx 16\lambda\sqrt{\frac{3}{2\pi}}\frac{\delta^{2-2\lambda}e^{-\lambda\Delta}}{\epsilon}\exp\left[-\delta^{4-4\lambda}\frac{96}{\epsilon^2}e^{-2\lambda\Delta}\right] \tag{89}$$

$$\approx 16\lambda\sqrt{\frac{3}{2\pi}}\frac{e^{-\lambda\Delta}}{\epsilon}\exp\left[-\frac{96}{\epsilon^2}e^{-2\lambda\Delta}\right] \tag{90}$$

$$\sim 16\lambda\sqrt{\frac{3}{2\pi}}\frac{e^{-\lambda\Delta}}{\epsilon} \quad \text{as } \Delta \to \infty \tag{91}$$

In Figure 7 we plot the distribution of the times between maxima for the same numerical simulation as for the return points plotted in 6 along with the estimate above, both with and without making the approximation $\delta^{1-\lambda} = 1$. In each case the exponential decay is correct, but the density without the assumption is more accurate. This is likely due to the fact that for these parameter values, λ and μ are in fact significantly different from one. Additionally, for this range of λ and μ, the excursion time away from the origin is not quite $2\log(4/\delta)$, providing an additional source of error. All of these approximation are fixed in the next section, where we generalise the homoclinic pulse expansion technique of ODE of Balmforth *et al.* [4] to stochastic systems.

In addition, we can estimate the expected time between maxima as

$$\mathbb{E}(\Delta) \approx \int_0^\infty 16\lambda\sqrt{\frac{3}{2\pi}}\frac{\Delta e^{-\lambda\Delta}}{\epsilon}\exp\left[-\frac{96}{\epsilon^2}e^{-2\lambda\Delta}\right]\,\mathrm{d}\Delta \tag{92}$$

$$= \int_0^{\sqrt{96}/\epsilon} 16\sqrt{\frac{3}{2\pi}}\frac{1}{\sqrt{96}\lambda}\log\left(\frac{\sqrt{96}}{\epsilon s}\right)e^{-s^2}\mathrm{d}s \tag{93}$$

$$\sim \frac{1}{\lambda}\log\left(\frac{\sqrt{96}}{\epsilon}\right) \quad \text{as } \epsilon \to 0. \tag{94}$$

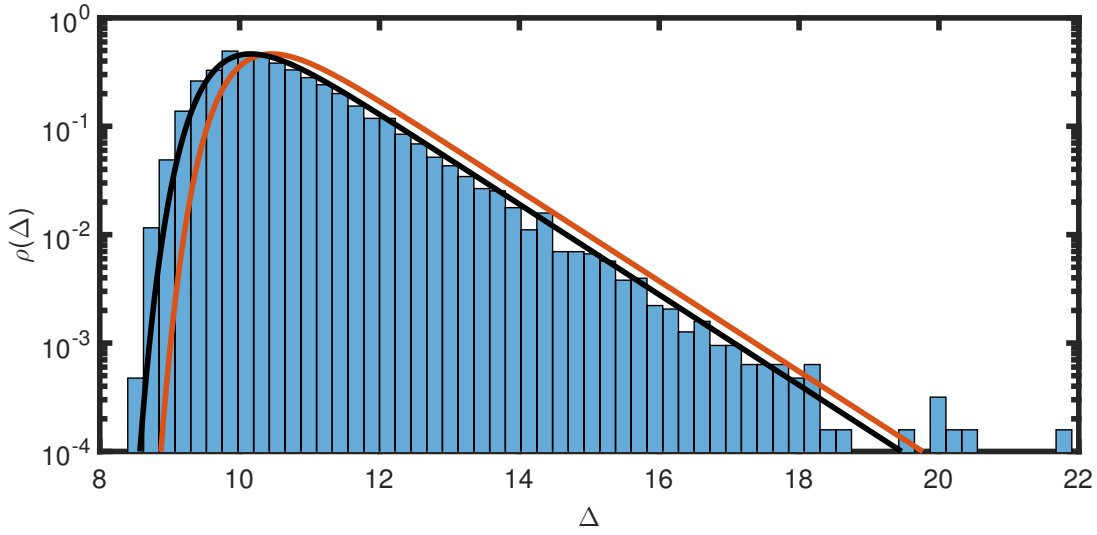

Figure 7: Probability density of times between maxima to with noise strength $\epsilon = 0.0006$ acting only outside $\mathcal{D}$, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for a direct numerical simulation with 3000 returns to $\mathcal{D}$ (blue histogram), the estimated solution with $\delta^{1-\lambda} = 1$ (red line) and the estimated solution with $\delta^{1-\lambda} \neq 1$ (black line).

For the numerical data shown in Figure 7 we have $\mathbb{E}(\Delta) = 10.695$, and the formula above with $\epsilon = 0.0006$ gives $\mathbb{E}(\Delta) = 10.097$. To compare to the work of Stone and Holmes, recall that the mean time through the origin was

$$\mathbb{E}(t_n) = \frac{1}{\lambda} \log\left(\frac{\delta}{\epsilon}\right), \tag{95}$$

and so the mean flight time is

$$\begin{aligned} \tau &= \mathbb{E}(t_n) + 2t_b && (96)\\ &= \frac{1}{\lambda} \log\left(\frac{\delta}{\epsilon}\right) + 2\log\left(\frac{4}{\delta}\right) && (97)\\ &= \frac{1}{\lambda} \log\left(\frac{\delta^{1-2\lambda} 4^{2\lambda}}{\epsilon}\right) && (98)\\ &\approx \frac{1}{\lambda} \log\left(\frac{16}{\epsilon\delta}\right), && (99) \end{aligned}$$

which, in addition to depending explicitly on δ, takes the value 13.003 for the parameters used here, which is clearly not a good approximation.

2.6 Homoclinic pulse expansion for the stochastic Duffing system

To move away from the assumption that $\gamma, \beta = O(\epsilon)$, and to correct the discrepancy with the timing map distribution requiring $\delta^{1-\lambda} = 1$, consider the equation

$$\ddot{x} - x + x^3 + \gamma\dot{x} - \beta x^2 \dot{x} = \epsilon\sigma\xi(t). \tag{100}$$

Let $\gamma = \gamma_0(\beta)$ be the parameter set for which there exists a homoclinic orbit, and write $\gamma = \gamma_0 + \epsilon\gamma_1$. To consider $\xi \neq 0$ *everywhere*, we instead make the ansatz for the full solution as

$$x(t) = \sum_k \theta_k H(t - t_k) + \epsilon R \equiv \sum \theta_k H_k + \epsilon R, \tag{101}$$

where $H(t)$ is the homoclinic solution, the sequence of times $\{t_k\}$ are sufficiently widely seperated so that $H_k H_{k\pm1} = O(\epsilon)$, the polarity $\theta_k = \pm 1$ accounts for the symmetry $x \mapsto -x$, and ϵR is the error made in making this assumption. This is the singular perturbation method for finding timing between homoclinic pulses of ODEs of Balmforth *et al.* [4], but here we can generalise it to stochastic systems.

In order to begin the asymptotic expansion, we need to explain what is meant by $O(\epsilon)$ interaction of neighbouring homoclinic orbits. Most importantly, we need to have an expansion for the nonlinear terms in (100). For correctly chosen times $\{t_k\}$ we have

$$\left(\sum_k \theta_k H_k\right)^3 = \sum_k \theta_k H_k^3 + 3\sum_k H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1}) + O(\epsilon^2), \tag{102}$$

for the cubic term, where the second term is $O(\epsilon)$. The term $\beta x^2 \dot{x}$ may be treated similarly.

The $O(\epsilon^0)$ equation is satisfied automatically, and the $O(\epsilon)$ equation in the vicinity of t_k gives

$$\mathcal{J}_k R = -\frac{3}{\epsilon}H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1}) + \frac{\beta}{\epsilon}\frac{\mathrm{d}}{\mathrm{d}t}[H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1})] + \sigma\xi - \gamma_1\theta_k\dot{H}_k \tag{103}$$

for each k, where

$$\mathcal{J}_k = \mathcal{L} + 3H_k^2 - \beta\frac{\mathrm{d}}{\mathrm{d}t}H_k^2. \tag{104}$$

We note that the correct operator acting on R should involved the sum over all k to include all pulses, but we note that this sum is highly peaked about each homlinic trajectory, and so we may approximate by splitting the sum up, and requiring R to satisfy a simpler equation for each k. The error made in this approximation is of higher order in ϵ and so we employ it here [4].

Next, define $N_k \neq 0$ by

$$\mathcal{J}_k^\dagger N_k = 0. \tag{105}$$

Then, multiply through by N_k and integrate to obtain

$$\begin{aligned} 0 = \int_{-\infty}^{\infty} N_k \Big[&-\frac{3}{\epsilon}H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1}) \\ &+ \frac{\beta}{\epsilon}\frac{\mathrm{d}}{\mathrm{d}t}[H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1})] + \sigma\xi - \gamma_1\theta_k H_k\Big]\,\mathrm{d}t \qquad (106) \\ = \int_{-\infty}^{\infty} \Big[&-\frac{3N_k + \beta\dot{N}_k}{\epsilon}H_k^2(\theta_{k+1}H_{k+1} + \theta_{k-1}H_{k-1}) + N_k(\sigma\xi - \gamma_1\theta_k\dot{H}_k)\Big]\,\mathrm{d}t \qquad (107) \end{aligned}$$

We have that

$$H_{k-1} \sim h_\infty e^{\mu(t-t_{k-1})} \quad \text{as } t - t_{k-1} \to \infty \tag{108}$$

$$H_{k+1} \sim h_0 e^{\lambda(t-t_{k+1})} \quad \text{as } t - t_{k+1} \to -\infty, \tag{109}$$

and so although strictly speaking we should only integrate the equations over the k-th pulse, we can approximate the neighbouring pulses in the integral, and integrate over the whole real line and obtain

$$\int_{-\infty}^{\infty}(3N_k+\beta\dot{N}_k)H_k^2H_{k+1}\,\mathrm{d}t \approx \int_{-\infty}^{\infty}(3N_k+\beta\dot{N}_k)H_k^2h_0e^{\lambda(t-t_{k+1})}\,\mathrm{d}t \tag{110}$$

$$= h_0e^{-\lambda(t_{k+1}-t_k)}\int_{-\infty}^{\infty}(3N_k+\beta\dot{N}_k)H_k^2e^{\lambda(t-t_k)}\,\mathrm{d}t \tag{111}$$

$$= h_0e^{-\lambda(t_{k+1}-t_k)}\int_{-\infty}^{\infty}(3N+\beta\dot{N})H^2e^{\lambda t}\,\mathrm{d}t \tag{112}$$

$$\equiv Ae^{-\lambda\Delta_{k+1}}, \tag{113}$$

where $\Delta_k = t_k - t_{k-1}$, and similarly,

$$\int_{-\infty}^{\infty}(3N_k+\beta\dot{N}_k)H_k^2H_{k-1}\,\mathrm{d}t \approx h_\infty e^{-\mu(t_k-t_{k-1})}\int_{-\infty}^{\infty}(3N+\beta\dot{N})H^2e^{-\mu t}\,\mathrm{d}t \equiv Be^{-\mu\Delta_k}. \tag{114}$$

We are then left with the timing map

$$\theta_{k+1}e^{-\lambda\Delta_{k+1}} = \theta_k\epsilon C + \theta_{k-1}De^{-\mu\Delta_k} + \epsilon\sigma M\eta_k, \tag{115}$$

where $\eta_k \sim \mathcal{N}(0,1)$, $C = -C_0/A$, $D = -B/A$, $M = -M_0/A$, and

$$C_0 = \int_{-\infty}^{\infty}\gamma_1 N\dot{H}\,\mathrm{d}t \tag{116}$$

$$M_0^2 = \int_{-\infty}^{\infty}N^2\mathrm{d}t. \tag{117}$$

We can convert the timing map (115) into a return-like map through the change of variables $z_k = \theta_k\theta_{k-1}\exp(-\lambda\Delta_k)$, which gives

$$z_{k+1} = \epsilon C + \mathrm{sgn}(z_k)D|z_k|^{\mu/\lambda} + \epsilon\sigma M\eta_k. \tag{118}$$

For small ϵ this equation has a deterministic fixed point $z = z_* = O(\epsilon)$ with $0 < |f_*'| \lesssim O(1)$ when λ and μ are not far from one. Hence, we can approximate the stationary distribution for z_k by

$$\rho_{z_k}(z) = \frac{1}{\sqrt{2\pi\epsilon^2\sigma^2M^2/(1-f_*'^2)}}\exp\left[-\frac{(z-z_*)^2}{2\epsilon^2\sigma^2M^2/(1-f_*'^2)}\right], \tag{119}$$

and so the stationary distribution for the timings Δ_k is given approximately by

$$\rho_{\Delta_k}(\Delta) = \frac{\lambda e^{-\lambda\Delta}}{\sqrt{2\pi\epsilon^2\sigma^2M^2/(1-f_*'^2)}}\left(\exp\left[-\frac{(e^{-\lambda\Delta}-z_*)^2}{2\epsilon^2\sigma^2M^2/(1-f_*'^2)}\right]\right.$$
$$\left.+\exp\left[-\frac{(e^{-\lambda\Delta}+z_*)^2}{2\epsilon^2\sigma^2M^2/(1-f_*'^2)}\right]\right) \tag{120}$$

$$\sim \frac{2\lambda e^{-\lambda\Delta}}{\sqrt{2\pi\epsilon^2\sigma^2M^2/(1-f_*'^2)}}\exp\left[-\frac{z_*^2}{2\epsilon^2\sigma^2M^2/(1-f_*'^2)}\right] \quad \text{as } \Delta\to\infty, \tag{121}$$

and so we obtain an exponential decay of the timing probability density function.

For the parameters $\gamma = 0.08$ and $\beta = 0.01$ with noise strength $\epsilon\sigma = 0.0006$, as used by Stone and Holmes [17], we have $\epsilon C = -1.362 \times 10^{-6}$ with fixed point $z_* = -2.576 \times 10^{-6}$ and $f'_* = 0.511$. Figure 8 shows the estimated densities $\rho_z(z)$ and $\rho_\Delta(\Delta)$ respectively, along with a numerical iteration of the timing map (115).

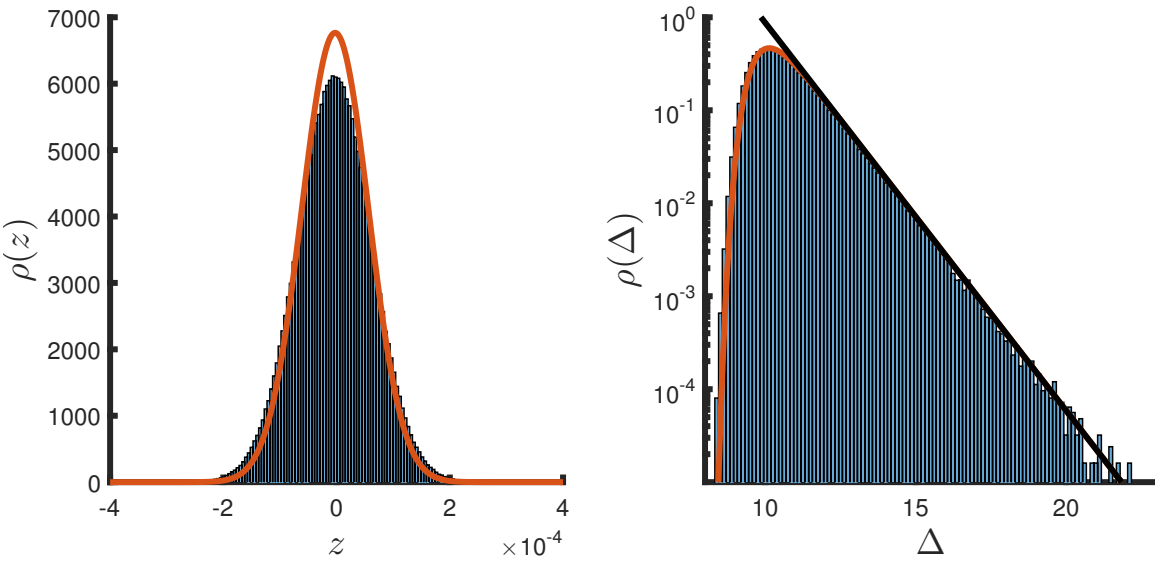

Figure 8: Left: Probability density of z_k, for parameters $\gamma = 0.08$ and $\beta = 0.1$ with noise strength $\epsilon\sigma = 0.0006$ for the derived asymptotic map (blue histogram), and the estimated solution with (red line). Right: Probability density of Δ_n, for parameters $\gamma = 0.08$ and $\beta = 0.1$ with noise strength $\epsilon\sigma = 0.0006$ for the derived asymptotic map (blue histogram), the estimated solution with (red line), and its large Δ exponential decay (black line).

We can also approximate the mean spacing $\mathbb{E}(\Delta)$ from

$$\mathbb{E}(\Delta) = \int_0^\infty \Delta \rho_{\Delta_k}(\Delta)\, \mathrm{d}\Delta, \tag{122}$$

which takes different forms in the two cases $|z_*| \ll a \ll 1$ and $a \ll |z_*| \ll 1$, where

$$a^2 = \epsilon^2\sigma^2 M^2/(1 - f_*'^2), \tag{123}$$

which is the effective variance of the z_n stationary distribution.

First, we look at $a \ll |z_*| \ll 1$, which we would expect to correspond to the case where noise is not strong enough to change $\mathbb{E}(\Delta)$ from its deterministic value. We have that the first exponential in the integral for $\mathbb{E}(\Delta)$ takes its maximum at $y = e^{-\lambda\Delta} = z_*$, which is well separated from the lower limit of integration, and the second exponential is asymptotically small. Hence,

$$\mathbb{E}(\Delta) \sim \frac{1}{\sqrt{2\pi a^2}} \frac{1}{\lambda} \log\left(\frac{1}{z_*}\right) \int_{-\infty}^{\infty} e^{-(y-z_*)^2/2a^2}\, \mathrm{d}y \tag{124}$$

$$= \frac{1}{\lambda} \log\left(\frac{1}{z_*}\right), \tag{125}$$

which is the deterministic value, as expected.

Next, for $|z_*| \ll a \ll 1$, we expect that the fixed point z_* does not greatly influence $\mathbb{E}(\Delta)$, and that the mean spacing should depend instead on a. In this limit, the two exponentials are comparible, and we may set $z_* = 0$ (though not $f_*'^2 = f_0'^2 = 0$), so that a remains unchanged. We then obtain

$$\mathbb{E}(\Delta) \sim \frac{2}{\sqrt{2\pi a^2}} \int_0^1 \frac{1}{\lambda} \log\left(\frac{1}{y}\right) e^{-y^2/2a^2} \, \mathrm{d}y \tag{126}$$

$$\sim \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{1}{\lambda} \log\left(\frac{1}{as\sqrt{2}}\right) e^{-s^2} \, \mathrm{d}s \tag{127}$$

$$= \frac{1}{\lambda} \log\left(\frac{1}{a}\right) + \frac{1}{2\lambda}(\gamma_e + \log 2) \tag{128}$$

$$\sim \frac{1}{\lambda} \log\left(\frac{1}{a}\right), \tag{129}$$

where γ_e is Euler's gamma constant.

Figure 9 shows the mean spacing $\mathbb{E}(\Delta)$ as a function of noise strength $\epsilon\sigma$ for the parameters $\gamma = 0.08$ and $\beta = 0.1$, for the timing map (115) and the original stochastic differential equation. Also plotted are the two asymptotic estimates above, for $a \ll |z_*| \ll 1$ and $|z_*| \ll a$. For the second case, $|z_*| \ll a$, we have $(\gamma_e + \log(2))/2\lambda \approx 0.661$, which is around 5% of the typical mean spacing in this range. For this reason, we plot both the leading order asymptotic result, and its correction. It is clear that the first correction is needed in this case. There are also plotted dashed lines at the two locations $\sigma\epsilon = z_*$ and $a = a(\epsilon\sigma) = z_*$. A first guess might be that the transition in behaviour occurs when the strength of the noise $\epsilon\sigma$ is comparable with z_*, but we have shown above that it is in fact the effective noise strength a that controls the change in behaviour. The figure clearly shows this.

We see also in Figure 9 that the asymptotic approximation (129) begins to break down as $\epsilon\sigma$ increases from 10^{-3}. This is to be expected eventually since we require $a(\epsilon) \ll 1$ for the Laplace approximation of the integral for $\mathbb{E}(\Delta)$ to be valid. Additionally, as $\mathbb{E}(\Delta)$ decreases, it eventually approaches the flight time taken along the homoclinic orbit, and so the assumption of well spaced pulses breaks down.

This generalised homoclinic pulse expansion technique has is clearly able to accurately reproduce the results of the full stochastic differential equations. In stark contrast to the results of Stone and Holmes [17], we have been able to fully characterise the effect of noise on the system. It is clear that the results for the noise away from the origin problem and the noise everywhere case are essentially the same and the the noise near the origin case of Stone and Holmes is significantly different (see Figures 6 and 8). The dynamics can be reduced to a simple stochastic one dimensional map of the form $z_{k+1} = f(z_k) + \sigma\eta_k$ which can be interpreted as the system obeying the deterministic dynamics $f(z_k)$ except for a random 'kick' upon re-entry to the origin, which depends on the sensitivity of the homoclinic solution to noise perturbations along its whole length. The system is controlled most closely by this kick upon re-entry, and is a fully nonlinear phenomena of the coupling between noise and nonlinear dynamics.

In the following sections we use the same technique to investigate the two other homoclinic systems, and find in each case the simple reduction to a stochastic mapping of the same form.

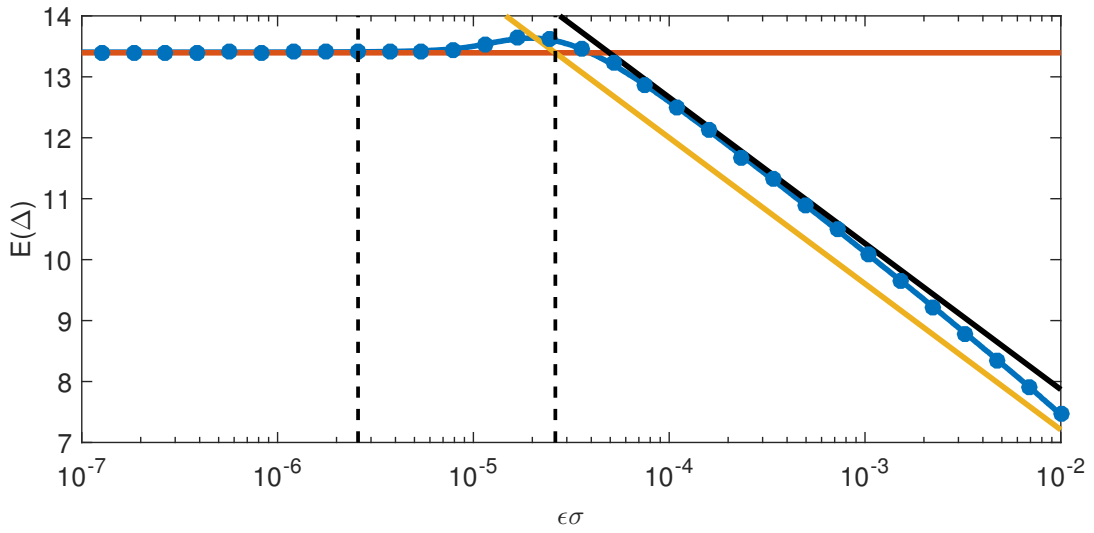

Figure 9: Mean spacing $\mathbb{E}(\Delta)$ as a function of noise strength $\epsilon\sigma$, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for the derived asymptotic map (blue line), and the original stochastic differential equation (blue dots). Also plotted is the asymptotic result for $a \ll |z_*| \ll 1$ (red line), the leading asymptotic result for $|z_*| \ll a$ (orange line) and its correction (solid black line). The leftmost dashed line shows $\epsilon\sigma = z_*$ and the rightmost dashed line shows $a = a(\epsilon\sigma) = z_*$.

3 Stochastic Shimizu–Murayama Model

The deterministic Shimizu–Murayama system [14] models the Lorenz equations at high Rayleigh number, and for a range of parameters exhibits near homoclinic behaviour [12]. The three-dimensional stochastic system is

$$\dot{x} = y + \epsilon\sigma\xi_x \tag{130}$$

$$\dot{y} = x(1 - z) - \lambda y + \epsilon\sigma\xi_y \tag{131}$$

$$\dot{z} = -\alpha(z - x^2) + \epsilon\sigma\xi_z, \tag{132}$$

or equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{f}(\boldsymbol{x}) + \epsilon\sigma\boldsymbol{\xi}, \tag{133}$$

where $\boldsymbol{x} = (x, y, z)$, $\boldsymbol{\xi} = (\xi_x, \xi_y, \xi_z)$, the matrix A is

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\alpha \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -\lambda_0 & 0 \\ 0 & 0 & -\alpha \end{pmatrix} + \epsilon \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\lambda_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \equiv A_0 + \epsilon A_1, \tag{134}$$

where $\lambda = \lambda_0 + \epsilon\lambda_1$ and $\lambda_0(\alpha)$ are the parameter values at which there is a homoclinic orbit, and the vector $\boldsymbol{f}(\boldsymbol{x})$ is

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} 0 \\ -xz \\ \alpha x^2 \end{pmatrix}. \tag{135}$$

We will consider cases with $\alpha, \lambda > 0$. A sample trajectory with $\alpha = 0.4$ and $\lambda = 1.1954$ is shown in Figure 2 along with its nearby homoclinic orbit at $\lambda \approx 1.2054$.

This system has fixed points at $(0,0,0)$ and $(\pm 1, 0, 1)$. The second of these is either a stable sink or unstable source. We are interested in the cases for which the origin is a saddle with two stable directions and one unstable directions. The eigenvalues and eigenvectors at the origin are

$$-\alpha \ : \ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{stable}, \tag{136}$$

$$-\mu_- = (-\lambda - \sqrt{4+\lambda^2})/2 \ : \ \begin{pmatrix} (\lambda - \sqrt{4+\lambda^2})/2 \\ 1 \\ 0 \end{pmatrix} \quad \text{stable}, \tag{137}$$

$$\mu_+ = (-\lambda + \sqrt{4+\lambda^2})/2 \ : \ \begin{pmatrix} (\lambda + \sqrt{4+\lambda^2})/2 \\ 1 \\ 0 \end{pmatrix} \quad \text{unstable}. \tag{138}$$

Let $\boldsymbol{H} = (H^x, H^y, H^z)$ be the homoclinic orbit that leaves the origin with $x > 0$. This system has the symmetry $(x, y, z) \mapsto (-x, -y, z)$, and so define $\phi\boldsymbol{H} = (\theta H^x, \theta H^y, H^z)$ for the polarity $\theta = \pm 1$. We then pose a solution of the form

$$\boldsymbol{x} = \sum_k \phi_k \boldsymbol{H}_k + \epsilon \boldsymbol{R}, \tag{139}$$

where $\boldsymbol{H}_k = \boldsymbol{H}(t - t_k)$ and the times t_k are well-separated so that $\boldsymbol{H}_k \cdot \boldsymbol{H}_{k\pm 1} = O(\epsilon)$. Then, at $O(\epsilon^0)$ we find the homoclinic solution, and at $O(\epsilon)$ we obtain

$$\mathcal{L}_{\phi_k} \boldsymbol{R} = \frac{1}{\epsilon} \boldsymbol{f}'(\phi_k \boldsymbol{H}_{\boldsymbol{k}}) \cdot (\phi_{k+1} \boldsymbol{H}_{k+1} + \phi_{k-1} \boldsymbol{H}_{k-1}) + A_1 \boldsymbol{H}_k + \sigma \boldsymbol{\xi}, \tag{140}$$

where

$$\mathcal{L}_{\phi_k} = \frac{\mathrm{d}}{\mathrm{d}t} - A_0 - \boldsymbol{f}'(\phi_k \boldsymbol{H}_k), \tag{141}$$

and

$$\boldsymbol{f}'(\phi\boldsymbol{H}) = \begin{pmatrix} 0 & 0 & 0 \\ -H^z & 0 & -\theta H^x \\ 2\alpha\theta H^x & 0 & 0 \end{pmatrix} \tag{142}$$

is the Jacobian matrix of the nonlinear part of the differential equations.

Now define $\mathcal{L}^\dagger_{\phi_k}$ by

$$\mathcal{L}^\dagger_{\phi_k} = -\frac{\mathrm{d}}{\mathrm{d}t} - A_0^\dagger - \boldsymbol{f}'(\phi_k \boldsymbol{H}_k)^\dagger, \tag{143}$$

and define $\boldsymbol{N}_{\phi_k} \neq 0$ by

$$\mathcal{L}^\dagger_{\phi_k} \boldsymbol{N}_{\phi_k} = 0. \tag{144}$$

Note that the equation satisfied by $\boldsymbol{N}_-$ is related to that satisfied by $\boldsymbol{N}_+$ through a symmetry, so that if $\boldsymbol{N}_+ = (N^x, N^y, N^z)$, then $\boldsymbol{N}_- = (N^x, N^y, -N^z)$. To this end, write instead $\boldsymbol{N}_{\phi_k} = \psi_k \boldsymbol{N}_k = (N^x_k, N^y_k, \theta_k N^z_k)$ and $\boldsymbol{N}_k$ satisfies

$$\mathcal{L}^\dagger_{+(k)} \boldsymbol{N}_k = 0. \tag{145}$$

Now we may take the dot product of the equation satisfied by $\boldsymbol{R}$ with $\psi_k \boldsymbol{N}_k$ and integrate to obtain

$$0 = \int \left[\frac{1}{\epsilon} \psi_k \boldsymbol{N}_k \cdot \boldsymbol{f}'(\phi_k \boldsymbol{H}_k) \cdot (\phi_{k+1} \boldsymbol{H}_{k+1} + \phi_{k-1} \boldsymbol{H}_{k-1}) + \psi_k \boldsymbol{N}_k \cdot (A_1 \phi_k \boldsymbol{H}_k + \sigma \boldsymbol{\xi}) \right] \, \mathrm{d}t. \tag{146}$$

In order to estimate the interaction integrals, we note that

$$\boldsymbol{H} \sim \begin{pmatrix} h_0^x \\ h_0^y \\ 0 \end{pmatrix} e^{\mu_+ t} \quad \text{as } t \to -\infty \tag{147}$$

$$\boldsymbol{H} \sim \begin{pmatrix} h_\infty^x \\ h_\infty^y \\ 0 \end{pmatrix} e^{-\mu_- t} + \begin{pmatrix} 0 \\ 0 \\ h_\infty^z \end{pmatrix} e^{-\alpha t} \quad \text{as } t \to \infty, \tag{148}$$

but $\mu_- > \alpha$ and so we may approximate

$$\boldsymbol{H} \sim \begin{pmatrix} 0 \\ 0 \\ h_\infty^z \end{pmatrix} e^{-\alpha t} \quad \text{as } t \to \infty, \tag{149}$$

which is equivalent to assuming a strong contraction in one of the stable directions, and so trajectories essentially become two-dimensional when passing nearby to the origin.

We may now approximate each of the interaction integrals to obtain

$$\int \psi_k \boldsymbol{N}_k \cdot \boldsymbol{f}'(\phi_k \boldsymbol{H}_k) \cdot \phi_{k+1} \boldsymbol{H}_{k+1} \, \mathrm{d}t \tag{150}$$

$$= \int \begin{pmatrix} N_k^x \\ N_k^y \\ \theta_k N_k^z \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ -H_k^z & 0 & -\theta_k H_k^x \\ 2\alpha \theta_k H_k^x & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \theta_{k+1} h_0^x \\ \theta_{k+1} h_0^y \\ 0 \end{pmatrix} e^{\mu_+ (t - t_{k+1})} \, \mathrm{d}t \tag{151}$$

$$= h_0^x e^{-\mu_+ (t_{k+1} - t_k)} \theta_{k+1} \int (2\alpha H^x N^z - H^z N^y) e^{\mu_+ t} \, \mathrm{d}t \tag{152}$$

$$\equiv \theta_{k+1} e^{-\mu_+ \Delta_{k+1}} A, \tag{153}$$

where $\Delta_k = t_k - t_{k-1}$, and similarly,

$$\int \psi_k \boldsymbol{N}_k \cdot \boldsymbol{f}'(\phi_k \boldsymbol{H}_k) \cdot \phi_{k-1} \boldsymbol{H}_{k-1} \, \mathrm{d}t \tag{154}$$

$$= -h_\infty^z e^{-\alpha (t_k - t_{k-1})} \theta_k \int H^x N^y e^{-\alpha t} \, \mathrm{d}t \tag{155}$$

$$\equiv -\theta_k e^{-\alpha \Delta_k} B_0. \tag{156}$$

Putting this together, we obtain the timing map

$$\theta_{k+1} e^{-\mu_+ \Delta_{k+1}} = \epsilon \theta_k C + \theta_k B e^{-\alpha \Delta_k} + \epsilon \sigma M \eta_k, \tag{157}$$

where $\eta_k \sim \mathcal{N}(0,1)$ and $B = B_0/A$, $C = C_0/A$ and $M = M_0/A$, where

$$C_0 = \int \lambda_1 N^y H^y \, \mathrm{d}t, \tag{158}$$

$$M_0^2 = \int \boldsymbol{N} \cdot \boldsymbol{N} \, \mathrm{d}t. \tag{159}$$

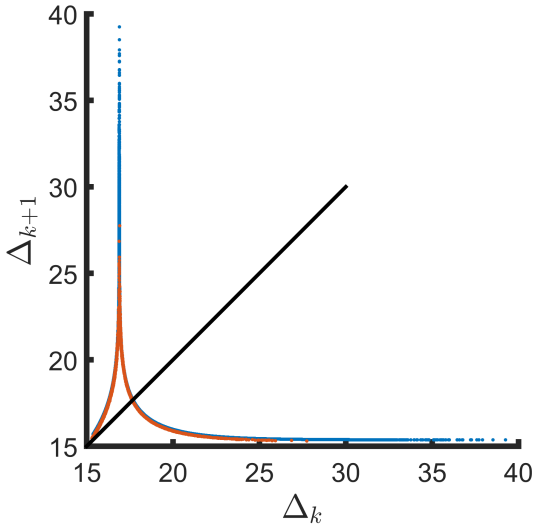

Figure 10: Timings for the Shimizu–Morioka system with no noise, $\sigma = 0$, and parameters $\alpha = 0.4$ and $\lambda = 1.1954$. Numerical simulation (red), and iterates of the derived timing map (blue).

Figure 10 shows a comparison between the above timing map and a numerical simulation of the system for no noise, $\sigma = 0$, $\alpha = 0.4$ and $\lambda = 1.1954$. The agreement is very good, and so we now concentrate only on the map, rather than the numerical simulations.

Figure 11 shows iterations of the derived timing map for noise strengths $\epsilon\sigma = 0$, 10^{-4}, 10^{-3} and 4×10^{-3} with $\alpha = 0.4$ and $\lambda = 1.1954$. We see that as the noise strength is increased, the deterministic structure is gradually broadened and smoothed out, before eventually being destroyed altogether. However, the shape of the probability distributions is largely unchanged, and are similar to the Duffing distributions for all noise strengths, as shown in Figure 12. Only at large noise strengths is there an appreciable shift in the peak of the distribution and a broadening of the tails.

4 A Stochastic Shilnikov System

Consider the ODE

$$\dddot{x} + \gamma\ddot{x} + \dot{x} - cx + x^3 = 0, \tag{160}$$

which is the normal form of a co-dimension three bifurcation [2] with the symmetry $x \mapsto -x$. Consider a stochastic equivalent of this equation

$$\dddot{x} + \gamma\ddot{x} + \dot{x} - cx + x^3 + \epsilon\sigma\xi(t) = 0, \tag{161}$$

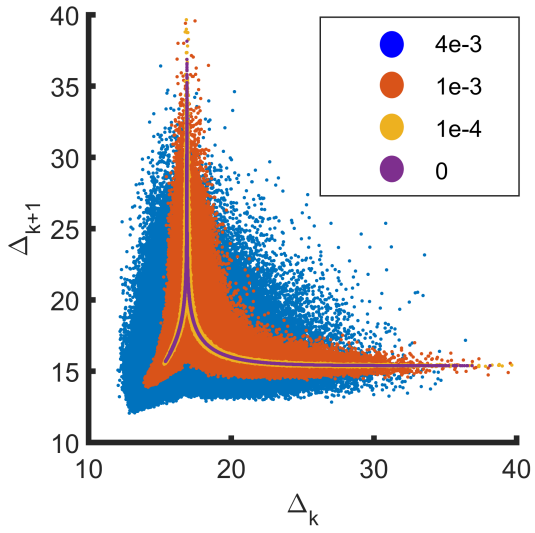

Figure 11: Timings for the Shimizu–Morioka system calculated from the derived map with noise strengths $\epsilon\sigma = 0$, 10^{-4}, 10^{-3} and 4×10^{-3} for parameters $\alpha = 0.4$ and $\lambda = 1.1954$.

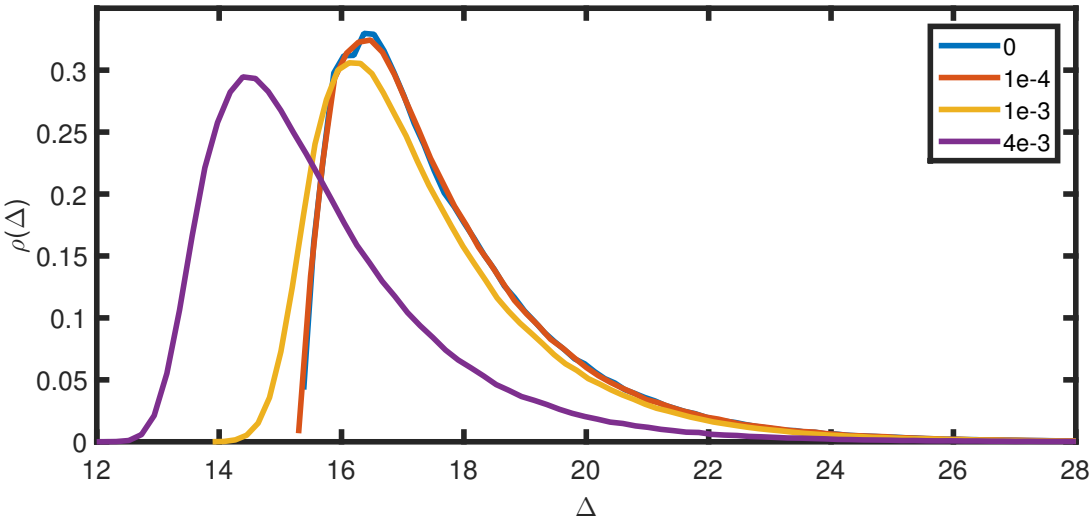

Figure 12: Probability density $\rho_\Delta(\Delta)$ of timings for the Shimizu–Morioka system calculated from the iterates derived map with noise strengths $\epsilon\sigma = 0$, 10^{-4}, 10^{-3} and 4×10^{-3} for parameters $\alpha = 0.4$ and $\lambda = 1.1954$.

where $\xi(t)$ is a white noise forcing, and σ is an $O(1)$ constant.

When $\sigma = 0$, this system has an unstable saddle at the origin, with a one-dimensional unstable direction and a two-dimensional unstable direction in which the dynamics are a focus. A sample trajectory for $\gamma = 0.7$ and $c = 1.108$ is shown in Figure 3 along with the nearby homoclinic orbit with $c \approx 1.107887$ which we label $H(t)$, and the time origin is such

that $H(t)$ takes its maxima at $t = 0$. We label the parameter values at which there exits a homoclinic orbit as $c = c_0(\gamma)$.

We again make the ansatz

$$x = \sum_k \theta_k H(t - t_k) + \epsilon R, \tag{162}$$

along with $c = c_0(\gamma) + \epsilon c_1$, for some sequence of widely separated times $\{t_k\}$ so that $H_k H_{k\pm1} = O(\epsilon)$ and the polarity $\theta_k = \pm 1$ accounts for the symmetry $x \mapsto -x$.

The $O(\epsilon^0)$ equation is satisfied automatically, and the $O(\epsilon)$ equation is

$$\mathcal{L}_k R = c_1 \theta_k H_k - \frac{3}{\epsilon} H_k^2 (\theta_{k+1} H_{k+1} + \theta_{k-1} H_{k-1}) - \sigma \xi(t), \tag{163}$$

where

$$\mathcal{L}_k = \frac{\mathrm{d}^3}{\mathrm{d}t^3} + \gamma \frac{\mathrm{d}^2}{\mathrm{d}t^2} + \frac{\mathrm{d}}{\mathrm{d}t} - c_0 + 3H_k^2. \tag{164}$$

Now define the adjoint operator to $\mathcal{L}_k$,

$$\mathcal{L}_k^\dagger = -\frac{\mathrm{d}^3}{\mathrm{d}t^3} + \gamma \frac{\mathrm{d}^2}{\mathrm{d}t^2} - \frac{\mathrm{d}}{\mathrm{d}t} - c_0 + 3H_k^2, \tag{165}$$

and the null adjoint solution $N_k(t) \neq 0$ by

$$\mathcal{L}_k^\dagger N_k = 0. \tag{166}$$

Multiply the equation for R through by $N_k(t)$ and integrate to obtain

$$0 = c_1 \theta_k A - \frac{3}{\epsilon} \int_{-\infty}^{\infty} N_k H_k^2 (\theta_{k+1} H_{k+1} + \theta_{k-1} H_{k-1}) \, \mathrm{d}t - \sigma \int_{-\infty}^{\infty} N_k \xi \, \mathrm{d}t, \tag{167}$$

where

$$A = \int_{-\infty}^{\infty} c_1 N_k H_k \, \mathrm{d}t. \tag{168}$$

We know that

$$H \sim \begin{cases} h_0 e^{\lambda t} & \text{as } t \to -\infty, \\ h_\infty e^{-\mu t} \cos(\omega t + \phi) & \text{as } t \to \infty, \end{cases} \tag{169}$$

where λ is the unstable eigenvalue of the origin, and $-\mu \pm i\omega$ is the stable eigenvalue, for some h_0, h_∞ and ϕ.

Since the pulse train of homoclinic orbits are widely separated, we have that

$$3 \int_{-\infty}^{\infty} N_k H_k^2 H_{k+1} \, \mathrm{d}t \sim \int_{-\infty}^{\infty} N_k H_k^2 h_0 e^{\lambda(t - t_{k+1})} \, \mathrm{d}t \tag{170}$$

$$= 3 h_0 e^{\lambda(t_k - t_{k+1})} \int_{-\infty}^{\infty} N_k H_k^2 e^{\lambda(t - t_k)} \, \mathrm{d}t \tag{171}$$

$$= 3 h_0 e^{-\lambda(t_{k+1} - t_k)} \int_{-\infty}^{\infty} N H^2 e^{\lambda t} \, \mathrm{d}t \tag{172}$$

$$\equiv e^{-\lambda(t_{k+1} - t_k)} D \tag{173}$$

and also

$$3\int_{-\infty}^{\infty} N_k H_k^2 H_{k-1}\,\mathrm{d}t \sim \int_{-\infty}^{\infty} N_k H_k^2 h_\infty e^{-\mu(t-t_{k-1})}\cos(\omega(t-t_{k-1})+\phi)\,\mathrm{d}t \tag{174}$$

$$= 3h_\infty e^{-\mu(t_k-t_{k-1})}\Bigg[\cos(\omega(t_k-t_{k-1})+\phi)\int_{-\infty}^{\infty} NH^2 e^{-\mu t}\cos(\omega t)\,\mathrm{d}t$$

$$-\sin(\omega(t_k-t_{k-1})+\phi)\int_{-\infty}^{\infty} NH^2 e^{-\mu t}\sin(\omega t)\,\mathrm{d}t\Bigg] \tag{175}$$

$$\equiv e^{-\mu(t_k-t_{k-1})}\cos(\omega(t_k-t_{k-1})+\Phi)B \tag{176}$$

Defining the time interval between pulses $\Delta_k \equiv t_k - t_{k-1}$, we obtain

$$\theta_{k+1}e^{-\lambda\Delta_{k+1}} = \epsilon\theta_k C + \theta_{k-1}Ee^{-\mu\Delta_k}\cos(\omega\Delta_k+\Phi) + \epsilon\sigma M\eta_k, \tag{177}$$

where $\eta_k \sim \mathcal{N}(0,1)$, and

$$C = \frac{A}{D}, \tag{178}$$

$$E = -\frac{B}{D}, \tag{179}$$

$$M^2 = \left(\frac{1}{D}\right)^2 \int_{-\infty}^{\infty} N^2\,\mathrm{d}t. \tag{180}$$

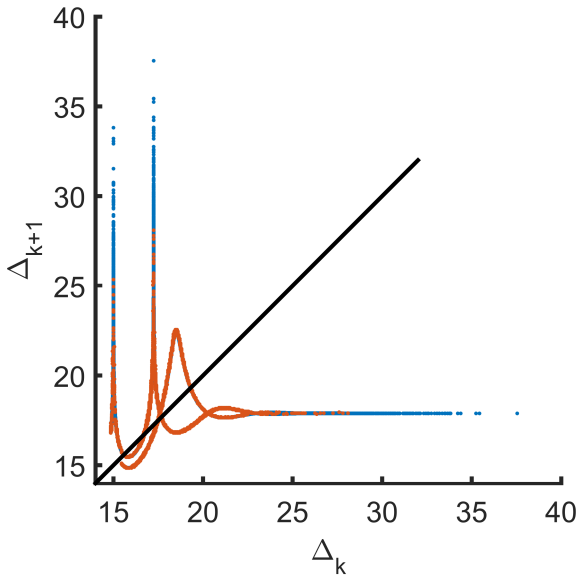

Figure 13: Times between pulses for the deterministic timing map (177) iterated 10^6 times (blue) and a full numerical simulation with 5500 pulses (red) for $c = 1.108$ and $\gamma = 0.7$.

To verify that we have obtained the correct timing map, we first remove the noise term, setting $\sigma = 0$. Figure 13 shows a comparison of the times between pulses for a full numerical

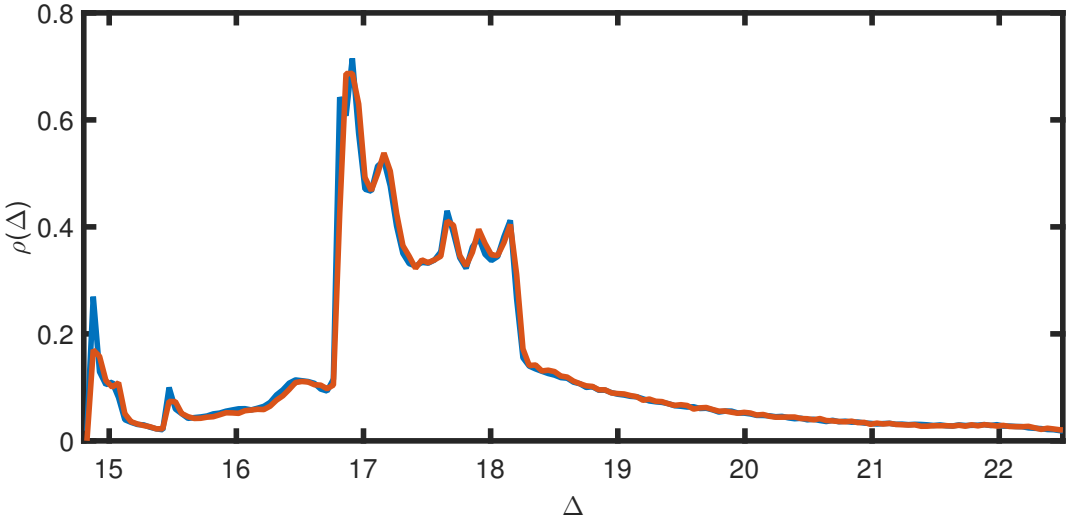

Figure 14: Probability density of times between pulses for the timing map (177) iterated 10^6 times (blue) and a full numerical simulation with 560 000 pulses (red) for $c = 1.108$ and $\gamma = 0.7$ with noise strength $\epsilon\sigma = 10^{-6}$.

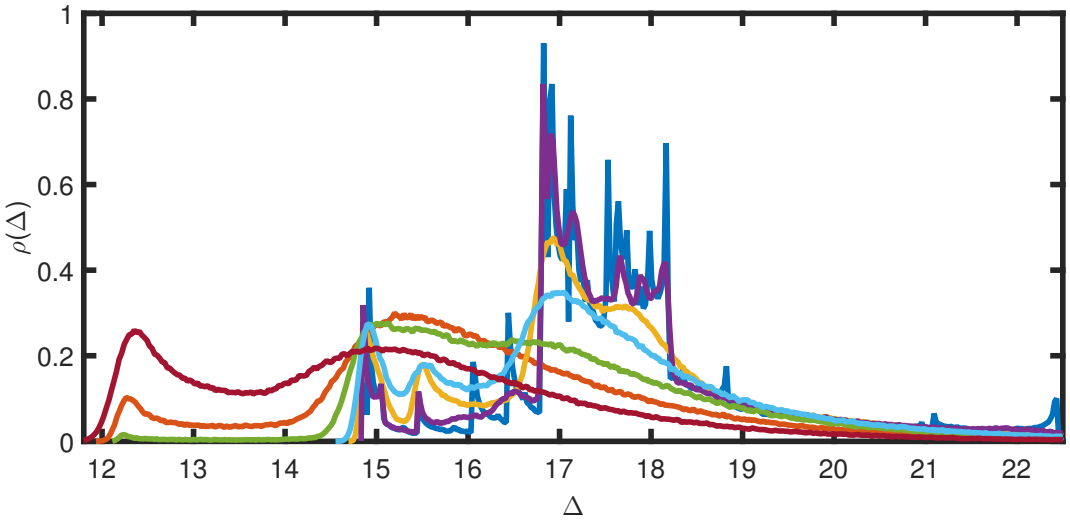

Figure 15: The pdf for the timing map (177) for noise strengths $\epsilon\sigma = 0$ (dark blue), 10^{-6} (purple), 10^{-5} (orange), 2×10^{-5} (light blue), 5×10^{-5} (green), 10^{-4} (red) and 2×10^{-4} (magenta).

simulation of (161) and the derived timing map (177) iterated 10^6 times for $c = 1.108$ and $\gamma = 0.7$. The agreement is excellent.

Figure 14 shows a comparison of the probability density of timings $\rho(\Delta)$ for noise strength $\epsilon\sigma = 10^{-6}$ for the timing map (177) iterated 10^6 times and a direct numerical

computation of (161) for 560 000 pulses. The agreement is excellent, and so we now deal only with the map.

Figure 15 shows the probability density of timings $\rho(\Delta)$ for a range of noise strengths, as well as the deterministic invariant measure normalised to be a probability density. For very small noise strengths, the peaks of the invariant measure are simply rounded off. For larger noise strengths, the peaks in the invariant measure are fully homogenised, and a new smooth peak begins to arise at smaller Δ. Appendix B shows images of the stationary distribution for a large range of ϵc_1 and $\epsilon\sigma$. The main features are that as noise strength is increased, it affects large Δ and small c_1 first. We see clearly that attractors with high Δ are rapidly smoothed out, and the densities begin to be centered around attractors at lower Δ. There is also evidence of exponential tails at large Δ. This is to be expected, since

$$\rho_\Delta(\Delta) = \lambda e^{-\lambda\Delta}\rho_z(e^{-\lambda\Delta}) \sim \lambda e^{-\lambda\Delta}\rho_z(0) \quad \text{as } \Delta \to \infty, \tag{181}$$

and so we obtain exponential tails provided that $\rho_z(0)$ is bounded. We can demonstrate this by observing that

$$\rho_z(0) = \frac{1}{\sqrt{2\pi a^2}} \int e^{-[f(s)]^2/2a^2} \rho_z(s)\, \mathrm{d}s \leq \frac{1}{\sqrt{2\pi a^2}} \int \rho_z(s)\, \mathrm{d}s = \frac{1}{\sqrt{2\pi a^2}}. \tag{182}$$

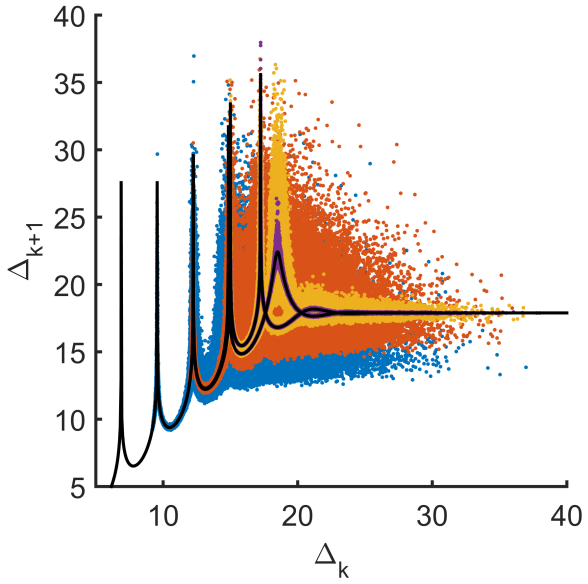

Figure 16: Times between pulses for the timing map (177) iterated 10^6 times for noise strength $\epsilon\sigma = 10^{-6}$ (purple), 10^{-5} (yellow), 10^{-4} (red) and 4×10^{-4} (blue) and the deterministic timing map (177) (black lines).

To investigate the emergence of the new peak in the probability density $\rho(\Delta)$, in Figure 16 we plot iterates of the timing map (177) for various noise strengths $\epsilon\sigma$ along with the full deterministic map. For the deterministic case, iterates of this map reside in the attracting set which has $\Delta \gtrsim 15$, see Figure 13. For very weak noise, the attracting set remains much

the same, but is simply smoothed out a little. Larger noise strengths blur the deterministic attracting set completely, but in addition begin to realise, and are attracted to, parts of the deterministic map at smaller Δ. Such phenomena have been observed by numerous authors, and are often referred to as 'noise induced boundary crises' [11, 8].

There is also a noise threshold beyond which a significant number of trajectories are attracted to the peak in the deterministic map just to the left of $\Delta = 10$ beyond which the map breaks down, and iterates diverge to $-\infty$. This can be rationalised by noticing that the time of flight for $H(t)$ is approximately 10, and so $\Delta < 10$ represents solutions that do not visit the origin at all, and the assumption of widely spaced pulses is no longer valid.

5 Conclusions

Through generalising the singular perturbation analysis of homoclinic pulse dynamics in ODEs [4] to stochastic systems, we have demonstrated that for a large class of near-homoclinic dynamical systems, the most significant effect of noise on the timing between homoclinic pulses and on the position of return of trajectories to the origin is due to a cumulative effect of noise on the trajectory away from the origin, rather than the details of the dynamics near to the origin as originally conjectured by Stone and Holmes [17]. This cumulative effect may be interpreted as providing a random kick sampled from a well-defined normal distribution to a deterministic trajectory just as it enters any domain of interest centred around and close to the origin, and that this kick is asymptotically large when compared to the effect of noise within this domain, in which its role is simply to provide a buffer region of size $O(\epsilon)$ about the stable manifold from which the stochastic trajectory is ejected, preventing trajectories from remaining within the domain.

We are able to make good analytical progress in the two-dimensional Duffing system since the resulting return map has a stable fixed point that we can expand around. For sufficiently small amplitude noise, the dynamics are controlled by the distance from homoclinicity, whilst for sufficiently large amplitude noise, the dynamics is controlled by the noise by creating an effective distance from homoclinicity based on the standard deviation of the resulting stationary distribution of the return map.

The derived maps for the Shimizu–Morioka model have nearly the same functional form as the Duffing system, but the parameter values here allow chaotic solutions. In this system we demonstrated that the singular perturbation analysis is easily generalisable to inherently multidimensional dynamics.

The derived maps for Shilnikov system have a rich and varied behaviour, as shown in Appendix B. Noise acts to smooth out the deterministic system, first at small deviations from homoclinicity and large spacings, and eventually at all parameter values as the noise amplitude is increased from zero. We also see noise induced boundary crises [11, 8] in which the presence of noise causes the stationary distribution to no longer be centered around the large spacing deterministic attractor, but to swap to another deterministic attractor at smaller spacing.

In all systems we observe exponential tails in the distribution of timing spacings as in Stone and Holmes [18], but note that this phenomenon is simply a manifestation of a change of variables from return points to timings.

We have developed a toolbox for investigating the effects of noise on homoclinic trajectories and have applied it to the two canonical homoclinic bifurcation scenarios, Lorenz and Shilnikov. Looking forward, we are now in a position to investigate more exotic systems like heteroclinic networks, bifocal orbits and excitable systems.

We also wish to apply the ideas contained withing this report to PDEs that contain traveling wave solutions that are of homoclinic or heteroclinic type. Near-homoclinic pulses can be interpreted as coherent structures in a number of physically relevant PDEs with traveling wave solutions [3]. A canonical example would be the real Ginzburg–Landau equation which has 'kink' solutions that connect two equilibria. It is know that an initial distribution of kinks that connect back and forth between the two equilibria will evolve in time to collide and annihilate each other in finite exponential time [3]. Given the results presented here, it seems likely that for stochastic dynamics, these annihilations will still occur, but instead because the kinks will begin to behave like Brownian random walkers.

References

[1] D. Armbruster, E. Stone, and V. Kirk, *Noisy heteroclinic networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 13 (2003), p. 71.

[2] A. Arneodo, P. H. Coullet, and E. A. Spiegel, *The dynamics of triple convection*, vol. 31, 1985.

[3] N. Balmforth, *Solitary Waves and Homoclinic Orbits*, Annual Review of Fluid Mechanics, 27 (1995), pp. 335–374.

[4] N. J. Balmforth, G. R. Ierley, and E. A. Spiegel, *Chaotic Pulse Trains*, SIAM Journal on Applied Mathematics, 54 (1994), pp. 1291–1334.

[5] M. Eguia and G. Mindlin, *Distribution of interspike times in noise-driven excitable systems*, Physical Review E, 61 (2000), pp. 6490–6499.

[6] C. W. Gardiner, *Handbook of stochastic methods*, Springer, 1985.

[7] P. Glendinning and C. Sparrow, *Local and global behavior near homoclinic orbits*, Journal of Statistical Physics, 35 (1984), pp. 645–696.

[8] R. E. Lee DeVille, E. Vanden-Eijnden, and C. B. Muratov, *Two distinct mechanisms of coherence in randomly perturbed dynamical systems*, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 72 (2005), pp. 1–10.

[9] B. T. Nadiga and B. P. Luce, *Global Bifurcation of Shilnikov Type in a Double-Gyre Ocean Model*, Journal of Physical Oceanography, 31 (2001), pp. 2669–2690.

[10] R. P. and P. Talkner, *Invariant densities for noisy maps*, Physical Review A, 44 (1991), pp. 6348–6363.

[11] A. Pikovsky and J. Kurths, *Coherence Resonance in a Noise-Driven Excitable System*, Physical Review Letters, 78 (1997), pp. 775–778.

[12] A. L. Shil'nikov, *On bifurcations of the Lorenz attractor in the Shimizu-Morioka model*, Physica D: Nonlinear Phenomena, 62 (1993), pp. 338–346.

[13] L. P. Shilnikov, *A Contribution To the Problem of the Structure of an Extended Neighborhood of a Rough Equilibrium State of Saddle-Focus Type*, Mathematics of the USSR-Sbornik, 10 (1970), pp. 91–102.

[14] T. Shimizu and N. Morioka, *On the bifurcation of a symmetric limit cycle to an asymmetric one in a simple model*, Physics Letters, 76 (1980), pp. 201–204.

[15] E. Simonnet, M. Ghil, and H. Dijkstra, *Homoclinic bifurcations in the quasi-geostrophic double-gyre circulation*, Journal of Marine Research, 63 (2005), pp. 931–956.

[16] E. Stone and D. Armbruster, *Noise and O(1) amplitude effects on heteroclinic cycles.*, Chaos (Woodbury, N.Y.), 9 (1999), pp. 499–506.

[17] E. Stone and P. Holmes, *Random Perturbations of Heteroclinic Attractors*, SIAM Journal on Applied Mathematics, 50 (1990), pp. 726–743.

[18] ——, *Unstable fixed points, heteroclinic cycles and exponential tails in turbulence production*, Physics Letters A, 155 (1991), pp. 29–42.

A Separating the directions of noise near the Duffing saddle

A.1 Noisy stable direction, $\xi_{x_1} \neq 0$ in $\mathcal{D}$

In this case, the deterministic results for x_2 still hold, and so the nth residence time t_n still satisfies

$$t_n = \frac{1}{\lambda} \log\left(\frac{\delta}{z_n}\right), \tag{183}$$

provided that z_n is known. Since we know that upon entrance we have $x_1 = \delta$, we wish to solve the Ornstein-Uhlenbeck process for x_1 for the transition probability density

$$\rho(x_1, t_n | \delta, 0). \tag{184}$$

The known solution to for x_1 is

$$\rho(x_1, t_n|\delta, 0) = \mathcal{N}_{x_1}\left(\delta e^{-\mu t_n}, \frac{\epsilon^2}{2\mu}(1 - e^{-2\mu t_n})\right) \tag{185}$$

$$= \mathcal{N}_{x_1}\left(\delta^{1-\mu/\lambda} z_n^{\mu/\lambda}, \frac{\epsilon^2}{2\mu}(1 - \delta^{-2\mu/\lambda} z_n^{2\mu/\lambda})\right) \tag{186}$$

This transition probability density is the the probability density of the exit points w_n.

Recalling the linear mapping

$$z_{n+1} = \alpha w_n + c, \tag{187}$$

along with standard results for linear combinations of normally distributed random variables, we see that

$$\rho(z_{n+1}|z_n) = \mathcal{N}_z\left(\alpha\delta^{1-\mu/\lambda} z_n^{\mu/\lambda} + c, \frac{\alpha^2\epsilon^2}{2\mu}(1 - \delta^{-2\mu/\lambda} z_n^{2\mu/\lambda})\right). \tag{188}$$

Finally, since $t_n = -\log(z_n/\delta)/\lambda$, we have

$$\rho(t_{n+1}|t_n) \approx \begin{cases} \lambda e^{-\lambda t}\mathcal{N}_{e^{-\lambda t}}\left(\alpha e^{-\mu t_n} + c/\delta, \frac{\alpha^2\epsilon^2}{2\mu\delta^2}(1 - e^{-2\mu t_n})\right) & \text{for } \delta \gg c \gg \epsilon, \\ 2\lambda e^{-\lambda t}\mathcal{N}_{e^{-\lambda t}}\left(\alpha e^{-\mu t_n}, \frac{\alpha^2\epsilon^2}{2\mu\delta^2}(1 - e^{-2\mu t_n})\right) & \text{for } c \ll \epsilon. \end{cases} \tag{189}$$

For long residency times with $\mu t_n \gg 1$, which is valid for δ small enough such that the deterministic trajectories are very close to the homoclinic orbit, we can make the approximation $e^{-\mu t_n} \approx 0$. This is formally equivalent to assuming that trajectories through $\mathcal{D}$ attain a statistically steady state before leaving $\mathcal{D}$. In this case, we obtain

$$\rho^s(w_n) = \mathcal{N}_w\left(0, \frac{\epsilon^2}{2\mu}\right), \tag{190}$$

$$\rho^s(z_n) = \mathcal{N}_z\left(c, \frac{\alpha^2\epsilon^2}{2\mu}\right), \tag{191}$$

$$\rho^s(t_n) = \begin{cases} \lambda e^{-\lambda t}\mathcal{N}_{e^{-\lambda t}}\left(\frac{c}{\delta}, \frac{\alpha^2\epsilon^2}{2\mu\delta^2}\right) & \text{for } \delta \gg c \gg \epsilon, \\ 2\lambda e^{-\lambda t}\mathcal{N}_{e^{-\lambda t}}\left(0, \frac{\alpha^2\epsilon^2}{2\mu\delta^2}\right) & \text{for } c \ll \epsilon. \end{cases} \tag{192}$$

In particular, $\mathbb{E}^s(w_n)$ and $\mathbb{E}^s(z_n)$ are the same as the deterministic result either under the assumption that $\mu t_n \gg 1$ or in the limit $n \to \infty$.

Also, for $\epsilon \ll c \ll \delta \ll 1$ we have

$$\mathbb{E}^s(t_n) = \int_0^\infty \frac{\lambda t e^{-\lambda t}}{\sqrt{\pi\alpha^2\epsilon^2/\mu\delta^2}} \exp\left[-\frac{(e^{-\lambda t} - c/\delta)^2}{\alpha^2\epsilon^2/\mu\delta^2}\right] \mathrm{d}t \tag{193}$$

$$\sim \frac{1}{\sqrt{\pi\alpha^2\epsilon^2/\mu\delta^2}} \int_{-\infty}^\infty \lambda t_* e^{-\lambda t_*} e^{-\lambda^2 c^2 \mu (t-t_*)^2/\alpha^2\epsilon^2} \mathrm{d}t \tag{194}$$

$$\sim \frac{\alpha\epsilon\sqrt{\pi} t_* e^{-\lambda t_*}}{c\sqrt{\mu\pi\alpha^2\epsilon^2/\mu\delta^2}} \tag{195}$$

$$\sim \frac{1}{\lambda} \log\left(\frac{\delta}{c}\right), \tag{196}$$

which matches the deterministic result for $c \neq 0$, where $t_* = \log(\delta/c)/\lambda$ is the stationary point of the term in the second exponential, and we have employed Laplace's method for approximating integrals, with large parameter δ^2/ϵ^2.

For $c \ll \epsilon$ we have

$$\mathbb{E}^s(t_n) = \int_0^\infty \frac{2\lambda t e^{-\lambda t}}{\sqrt{\pi\alpha^2\epsilon^2/\mu\delta^2}} \exp\left[-\frac{e^{-2\lambda t}}{\alpha^2\epsilon^2/\mu\delta^2}\right] \mathrm{d}t \tag{197}$$

$$= \int_0^1 \frac{2}{\lambda\sqrt{\pi\alpha^2\epsilon^2/\mu\delta^2}} \log\left(\frac{1}{y}\right) \exp\left[-\frac{y^2}{\alpha^2\epsilon^2/\mu\delta^2}\right] \mathrm{d}y \tag{198}$$

$$= \int_0^{\delta\sqrt{\mu}/\alpha\epsilon} \frac{2}{\lambda\sqrt{\pi}} \log\left(\frac{\delta\sqrt{\mu}}{\alpha\epsilon s}\right) e^{-s^2} \mathrm{d}s \tag{199}$$

$$\sim \int_0^\infty \frac{2}{\lambda\sqrt{\pi}} \log\left(\frac{\delta\sqrt{\mu}}{\alpha\epsilon s}\right) e^{-s^2} \mathrm{d}s \tag{200}$$

$$= \int_0^\infty \frac{2}{\lambda\sqrt{\pi}} \log\left(\frac{\delta}{\epsilon}\right) e^{-s^2} \mathrm{d}s + \int_0^\infty \frac{2}{\lambda\sqrt{\pi}} \log\left(\frac{\sqrt{\mu}}{\alpha s}\right) e^{-s^2} \mathrm{d}s \tag{201}$$

$$\sim \frac{1}{\lambda} \log\left(\frac{\delta}{\epsilon}\right), \tag{202}$$

since the second integral in (201) is just some $O(1)$ number. This is the same result as for noise uniformly every within $\mathcal{D}$.

Figure 17 shows a comparison between a direct numerical simulation for the return and leave probability densities $\rho(z_n)$ and $\rho(w_n)$ respectively, and their long residency time asymptotic stationary limit for $\gamma = 0.08$, $\beta = 0.01$, $\delta = 0.1$, $\epsilon = 0.0006$, $\alpha = 1$ and $c = 0$. There is clearly a very good match. The mean of $t_n = \log(\delta/|z_n|)/\lambda$ is 6.264, and the value of its asymptotic result $\log(\delta/\epsilon)/\lambda$ is 5.325.

A.2 Noisy unstable direction, $\xi_{x_2} \neq 0$ in $\mathcal{D}$

In this case we have $x_1 = \delta e^{-\mu t}$ and x_2 satisfies the unstable Ornstein-Uhlenbeck process with transition density

$$\rho(x_2, t|z_n, 0) = \mathcal{N}_{x_2}\left(z_n e^{\lambda t}, \frac{\epsilon^2}{2\lambda}(e^{2\lambda t} - 1)\right) \tag{203}$$

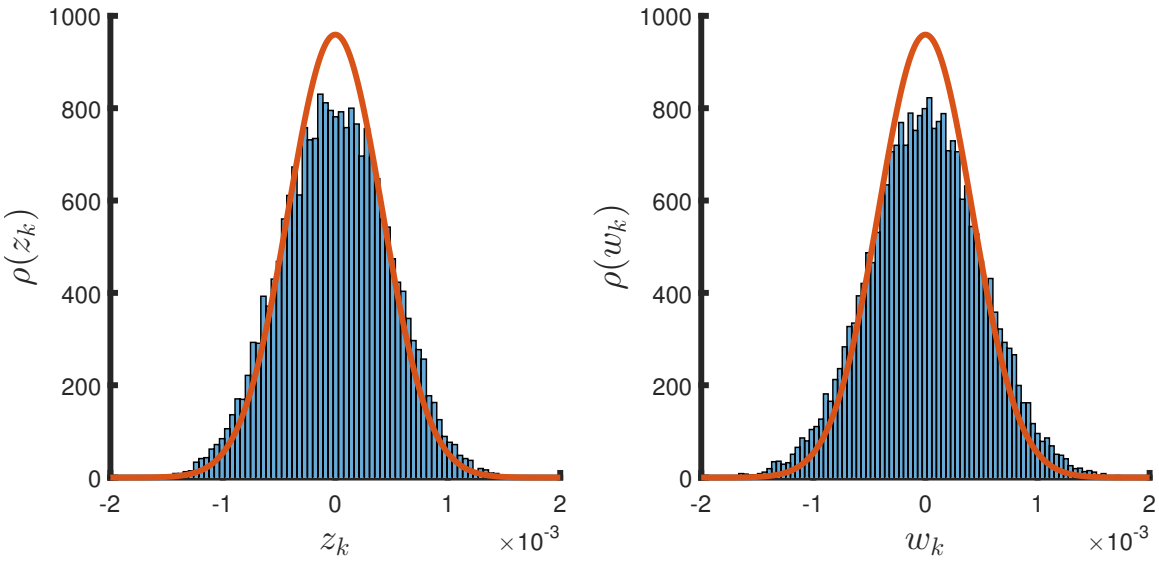

Figure 17: Left: Probability density of return points z_k to $\mathcal{D}$ with $\delta = 0.1$ and noise strength $\epsilon = 0.0006$ acting only in the stable direction of the saddle, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for a direct numerical simulation with 2250 returns to $\mathcal{D}$ (blue histogram) and the long residency time asymptotic stationary distribution (191) with $\alpha = 1$ and $c = 0$ (red line). Right: Leave points w_k.

In the case of large times, $\lambda t \gg 1$ we get

$$\rho(x_2, t|z_n, 0) \sim \mathcal{N}\left(z_n e^{\lambda t}, \frac{\epsilon^2}{2\lambda} e^{2\lambda t}\right) \tag{204}$$

In this limit we also have $w_n \approx 0$, and so $z_{n+1} \sim c$ provided that c is not too small.

In order to find the exit point w_n we need a distribution for the time taken for x_2 to leave $\mathcal{D}$. We have by definition

$$\mathbb{P}(t_n > t) = \int_{-\delta}^{\delta} \rho(x_2, t|z_n, 0)\, \mathrm{d}x_2 \tag{205}$$

$$= \int_{-\delta}^{\delta} \sqrt{\frac{\lambda}{\pi\epsilon^2(e^{2\lambda t} - 1)}} \exp\left[-\frac{\lambda(z - z_n e^{\lambda t})^2}{\epsilon^2(e^{2\lambda t} - 1)}\right] \mathrm{d}z \tag{206}$$

$$= \frac{1}{2}\left[\mathrm{erf}\left(\frac{\delta - z_n e^{\lambda t}}{\sigma(t)}\right) + \mathrm{erf}\left(\frac{\delta + z_n e^{\lambda t}}{\sigma(t)}\right)\right], \tag{207}$$

where $\sigma(t)^2 = \epsilon^2(e^{2\lambda t} - 1)/\lambda \sim \epsilon^2 e^{2\lambda t}/\lambda$. Note that under the assumption that $z_n \sim c$, the error functions have widely different expansions in all the cases $c \ll \epsilon \ll \delta$, $\epsilon \ll c \ll \delta$, $\epsilon \gg \delta e^{-\lambda t}$ and $c \gg \delta e^{-\lambda t}$, due to the exponential expansion casing rapid reordering of terms.

Define $b_\pm(t) = (\delta \pm z_n e^{\lambda t})/\sigma(t)$. Then,

$$\rho(t_n|z_n) = \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \mathbb{P}(t_n > t)\right) \tag{208}$$

$$= -\frac{1}{\sqrt{\pi}}\left[b'_+ e^{-b_+^2} + b'_- e^{-b_-^2}\right] \tag{209}$$

$$= \sqrt{\frac{\lambda^3}{\pi\epsilon^2(e^{2\lambda t}-1)^3}}\left[(\delta e^{2\lambda t} + z_n e^{\lambda t})\exp\left(-\frac{\lambda(\delta + z_n e^{\lambda t})^2}{\epsilon^2(e^{2\lambda t}-1)}\right)\right.$$
$$\left. + (\delta e^{2\lambda t} - z_n e^{\lambda t})\exp\left(-\frac{\lambda(\delta - z_n e^{\lambda t})^2}{\epsilon^2(e^{2\lambda t}-1)}\right)\right] \tag{210}$$

$$= \sqrt{\frac{\lambda^3}{\pi\epsilon^2(1-e^{-2\lambda t})^3}}\left[(\delta e^{-\lambda t} + z_n e^{-2\lambda t})\exp\left(-\frac{\lambda(\delta e^{-\lambda t} + z_n)^2}{\epsilon^2(1-e^{-2\lambda t})}\right)\right.$$
$$\left. + (\delta e^{-\lambda t} - z_n e^{-2\lambda t})\exp\left(-\frac{\lambda(\delta e^{-\lambda t} - z_n)^2}{\epsilon^2(1-e^{-2\lambda t})}\right)\right] \tag{211}$$

Then, since $w_n = \delta e^{-\mu t_n}$, we have

$$\rho(w_n|z_n) = \frac{1}{2\mu w_n}\sqrt{\frac{\lambda^3}{\pi\epsilon^2(1-(w_n/\delta)^{2\lambda/\mu})^3}}$$
$$\left[(\delta(w_n/\delta)^{\lambda/\mu} + z_n(w_n/\delta)^{2\lambda/\mu})\exp\left(-\frac{\lambda(\delta(w_n/\delta)^{\lambda/\mu} + z_n)^2}{\epsilon^2(1-(w_n/\delta)^{2\lambda/\mu})}\right)\right.$$
$$\left. + (\delta(w_n/\delta)^{\lambda/\mu} - z_n(w_n/\delta)^{2\lambda/\mu})\exp\left(-\frac{\lambda(\delta(w_n/\delta)^{\lambda/\mu} - z_n)^2}{\epsilon^2(1-(w_n/\delta)^{2\lambda/\mu})}\right)\right]. \tag{212}$$

Finally, since $z_{n+1} = \alpha w_n + c$, we obtain the transtion density

$$\rho(z_{n+1}|z_n) = \frac{1}{2\mu(z_{n+1}-c)}\sqrt{\frac{\lambda^3}{\pi\epsilon^2(1-((z_{n+1}-c)/\alpha\delta)^{2\lambda/\mu})^3}}$$
$$\left[(\delta((z_{n+1}-c)/\alpha\delta)^{\lambda/\mu} + z_n((z_{n+1}-c)/\alpha\delta)^{2\lambda/\mu})\right.$$
$$\exp\left(-\frac{\lambda(\delta((z_{n+1}-c)/\alpha\delta)^{\lambda/\mu} + z_n)^2}{\epsilon^2(1-((z_{n+1}-c)/\alpha\delta)^{2\lambda/\mu})}\right)$$
$$+ (\delta((z_{n+1}-c)/\alpha\delta)^{\lambda/\mu} - z_n(z_{n+1}-c)/\alpha\delta)^{2\lambda/\mu})$$
$$\left.\exp\left(-\frac{\lambda(\delta((z_{n+1}-c)/\alpha\delta)^{\lambda/\mu} - z_n)^2}{\epsilon^2(1-((z_{n+1}-c)/\alpha\delta)^{2\lambda/\mu})}\right)\right]. \tag{213}$$

Note that unless $\lambda/\mu = 1$, these distributions are not Gaussian, since for example, $\rho(w_n = 0|z_n) = \infty$. In general, we have $\lambda/\mu = 1 - \epsilon\gamma + O(\epsilon^2)$.

We can approximately find stationary distributions by the relation

$$\rho^s(z) = \int_{-\infty}^{\infty} \rho(z|s)\rho^s(s)\,\mathrm{d}s = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\rho(z|z_i), \tag{214}$$

where z_i is some predetermined sequence of known values of the stochastic process. For this, we can use DNS of the equations of motion to produce a sequence $\{z_i\}_1^n$ for some large n, and estimate ρ^s.

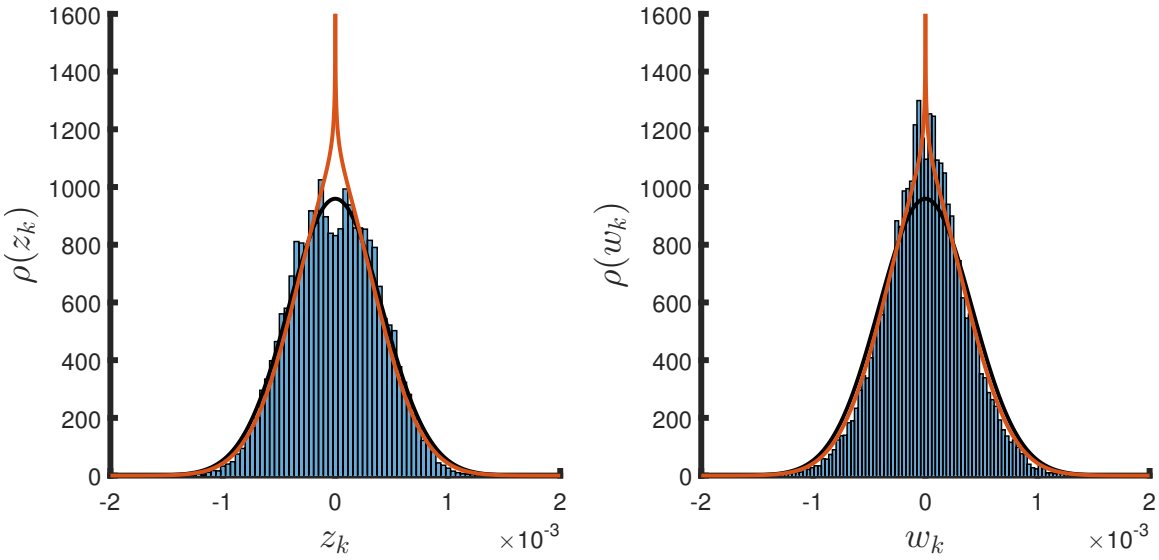

Figure 18: Probability density of return points z_k to $\mathcal{D}$ with $\delta = 0.1$ and noise strength $\epsilon = 0.0006$ acting only in the unstable direction of the saddle, for parameters $\gamma = 0.08$ and $\beta = 0.1$ for a direct numerical simulation with 2250 returns to $\mathcal{D}$ (blue histogram) and the long residency time asymptotic stationary distribution approximation (214) with $\alpha = 1$ and $c = 0$ (red line). Right: leave points w_k.

Figure 18 shows a comparison between a direct numerical simulation for the return and leave probability densities $\rho(z_n)$ and $\rho(w_n)$ respectively, and their stationary approximations using (214) for $\gamma = 0.08$, $\beta = 0.01$, $\delta = 0.1$, $\epsilon = 0.0006$, $\alpha = 1$ and $c = 0$. Plotted also is the Gaussian distribution for noise in the *stable* direction only. The leaving distribution $\rho(w_n)$ is well approximated by the result here, and is clearly non-Gaussian. The return distribution $\rho(z_n)$ is more closely Gaussian, and suggests that the linear approximation $z_{n+1} = \alpha z_n + c$ is not completely valid.

B Stationary probability densities $\rho_\Delta(\Delta)$ for the Shilnikov system

B.1 $\epsilon\sigma = 0$

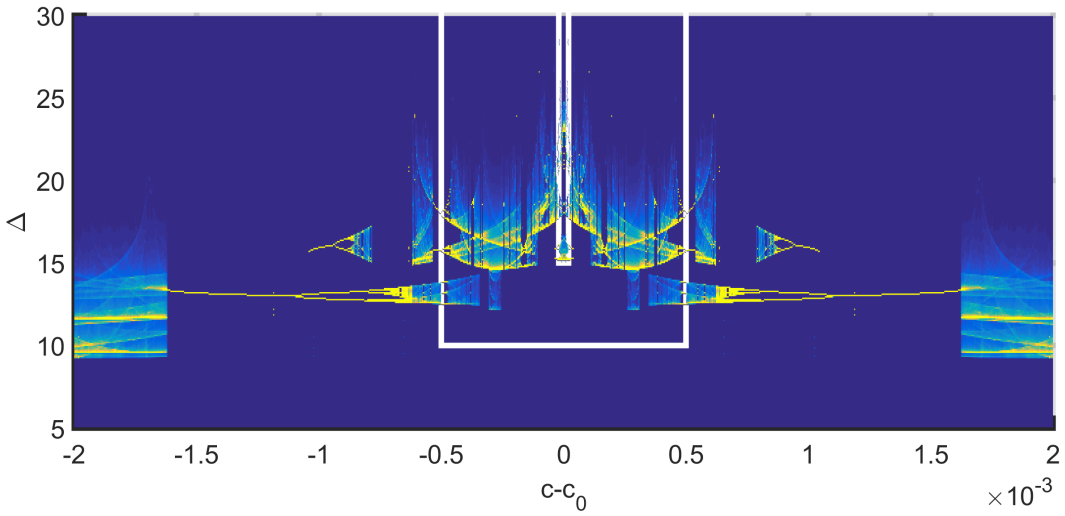

Figure 19: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 0$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 20 and 21.

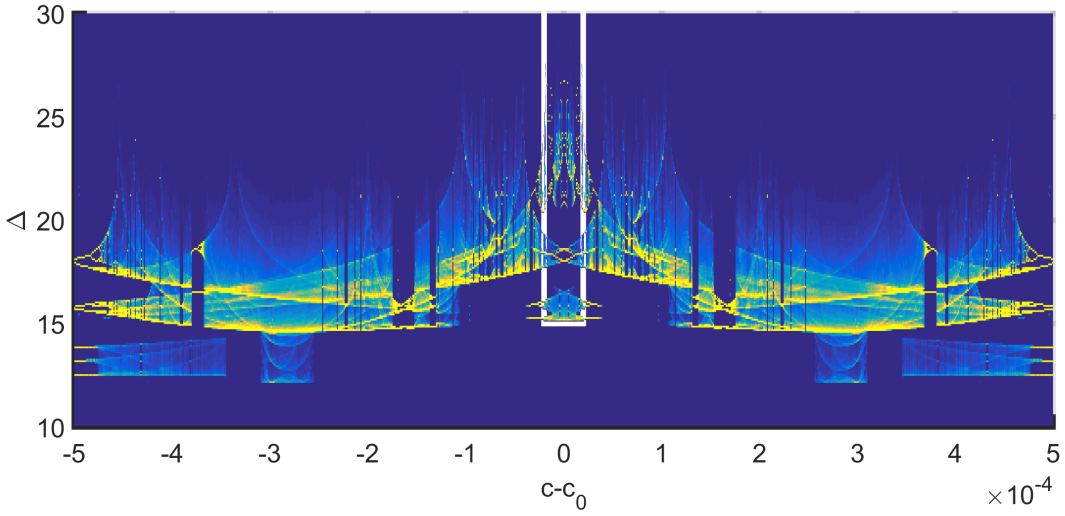

Figure 20: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 0$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 21.

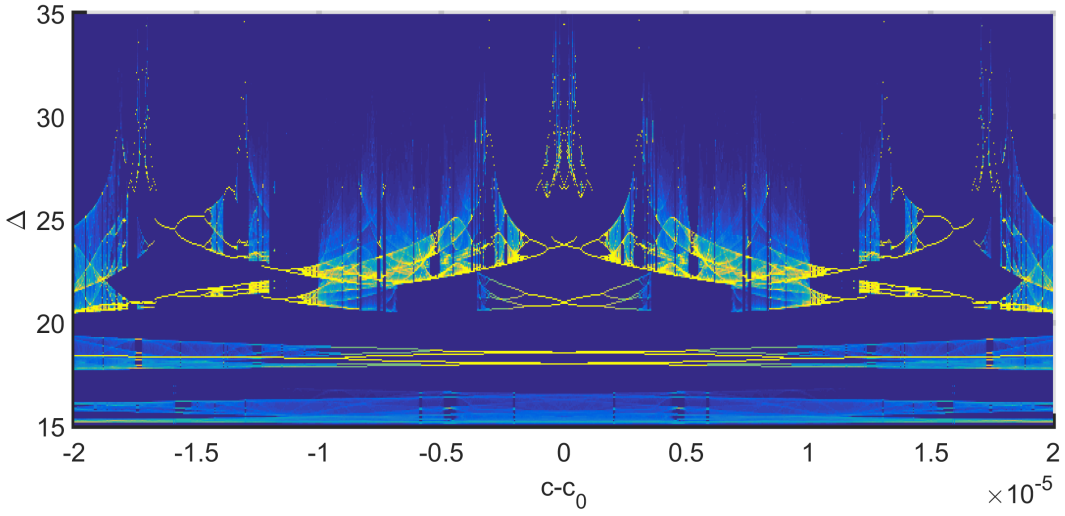

Figure 21: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 0$. Color from blue to yellow represents low to high values. Scale is arbitrary.

B.2 $\epsilon\sigma = 10^{-8}$

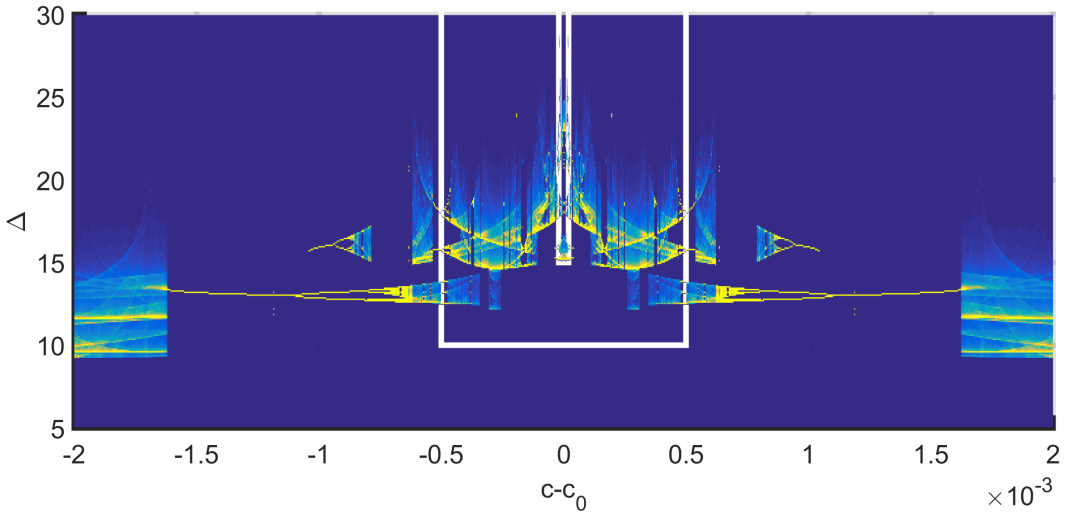

Figure 22: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-8}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 23 and 24.

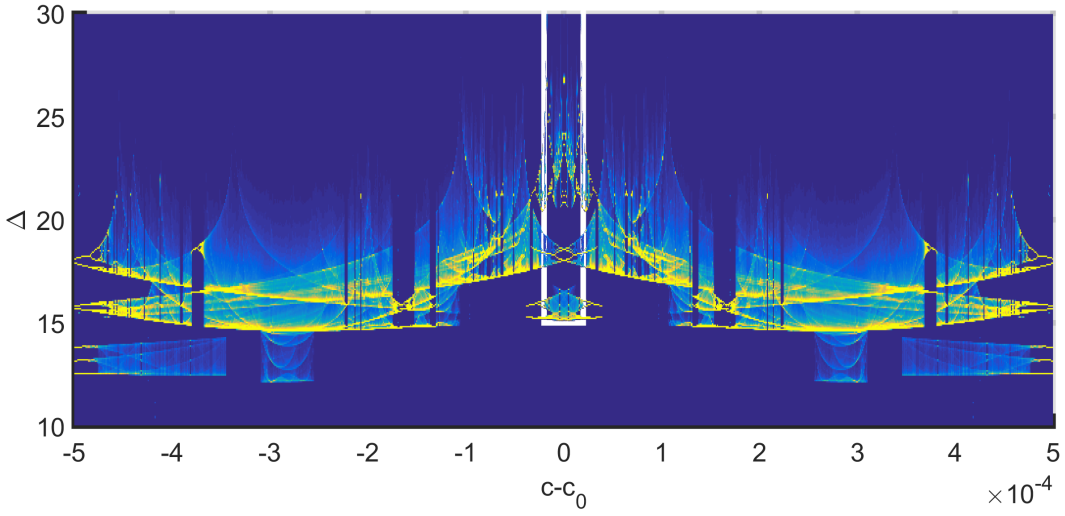

Figure 23: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-8}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 24.

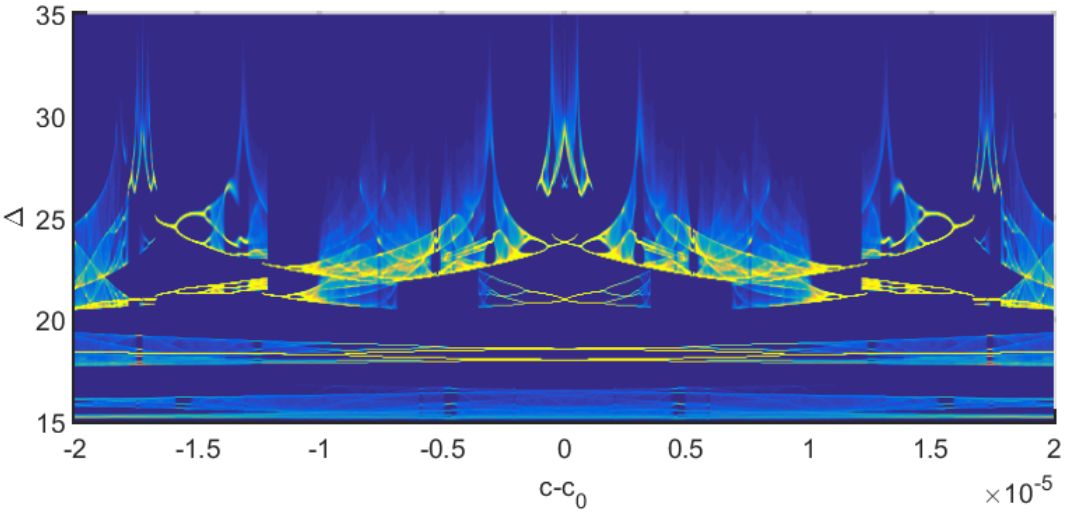

Figure 24: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-8}$. Color from blue to yellow represents low to high values. Scale is arbitrary.

B.3 $\epsilon\sigma = 10^{-7}$

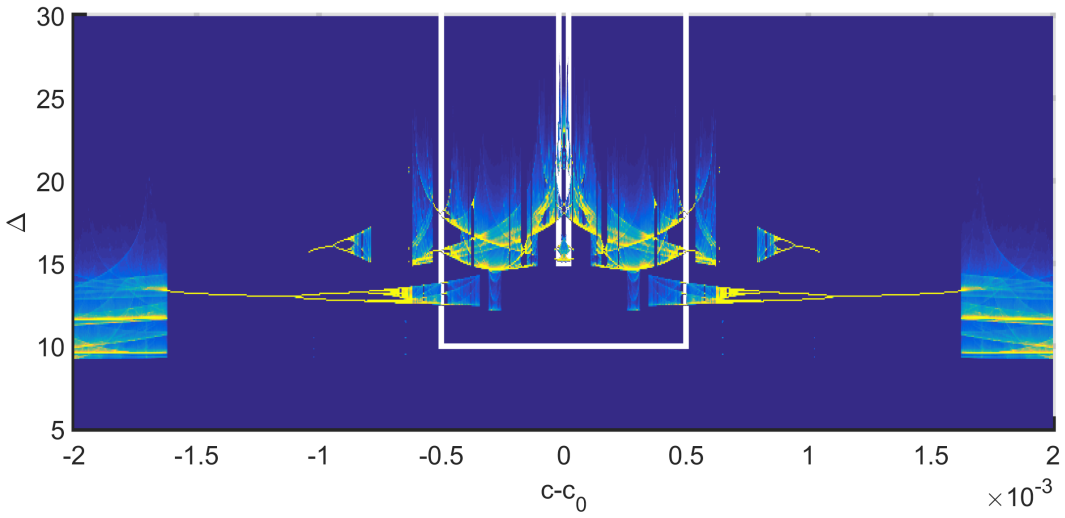

Figure 25: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-7}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 26 and 27.

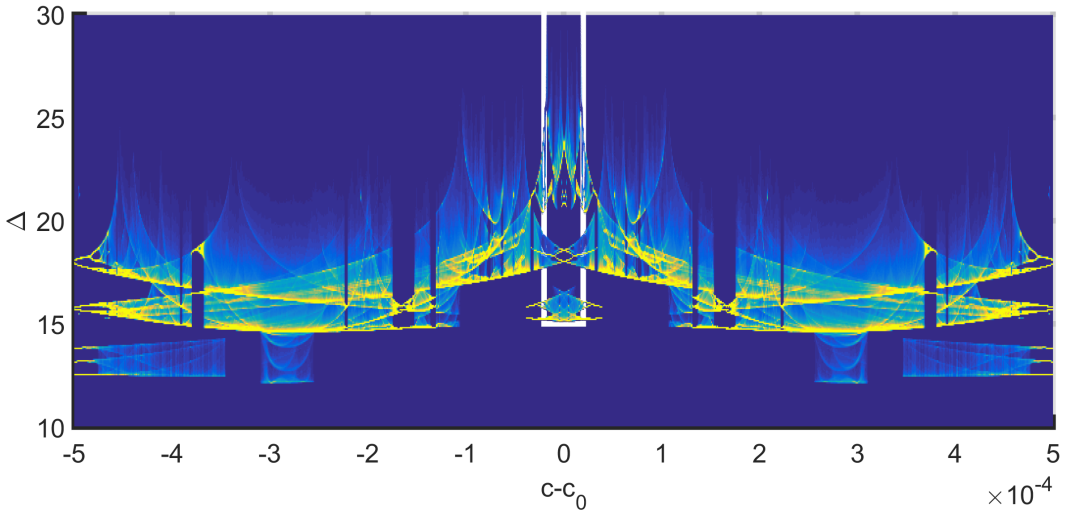

Figure 26: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-7}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 27.

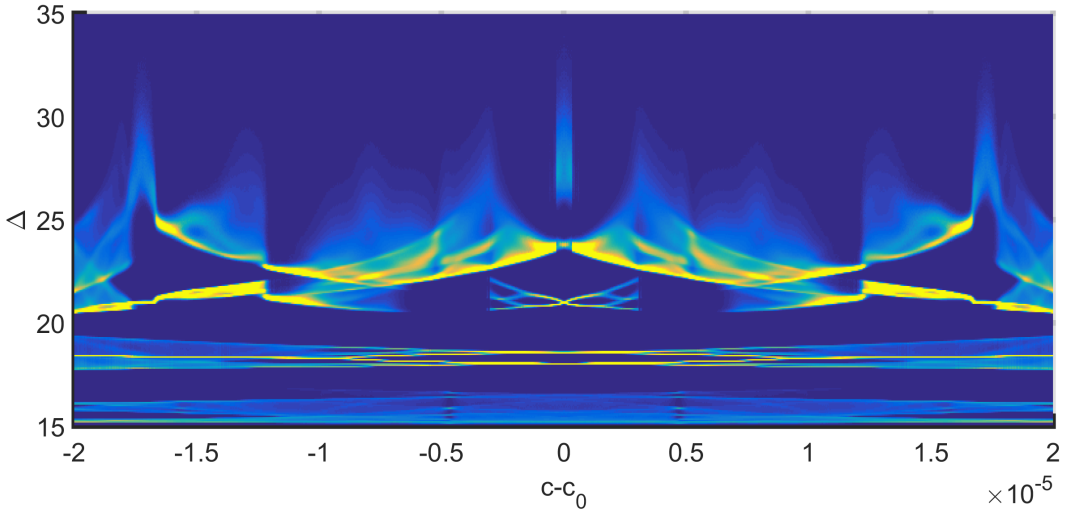

Figure 27: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-7}$. Color from blue to yellow represents low to high values. Scale is arbitrary.

B.4 $\epsilon\sigma = 10^{-6}$

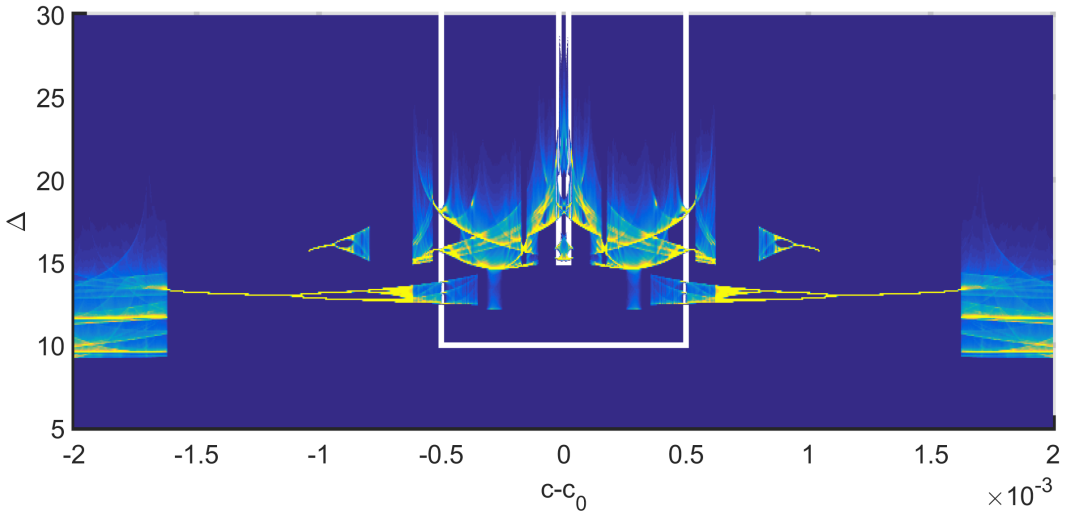

Figure 28: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-6}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 29 and 30.

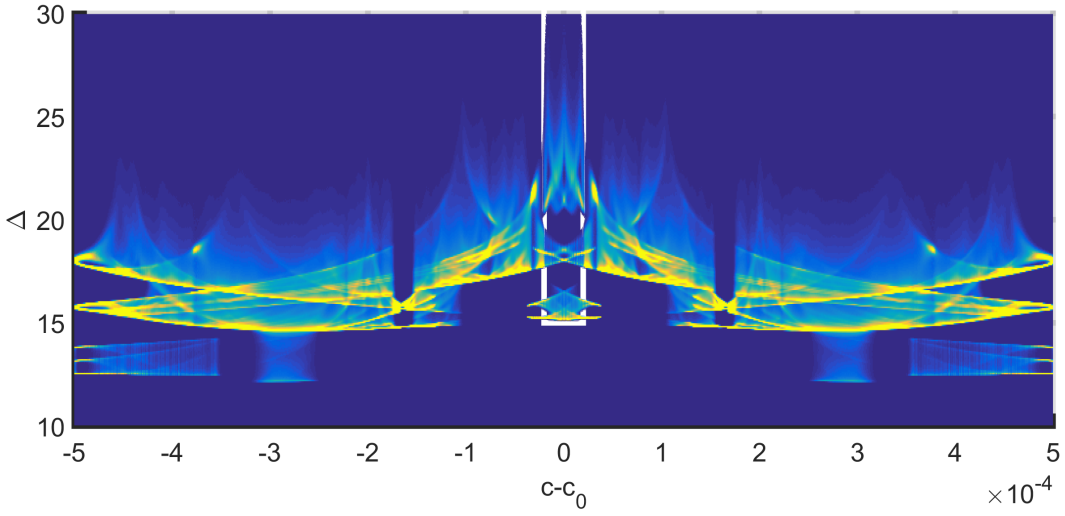

Figure 29: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-6}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 30.

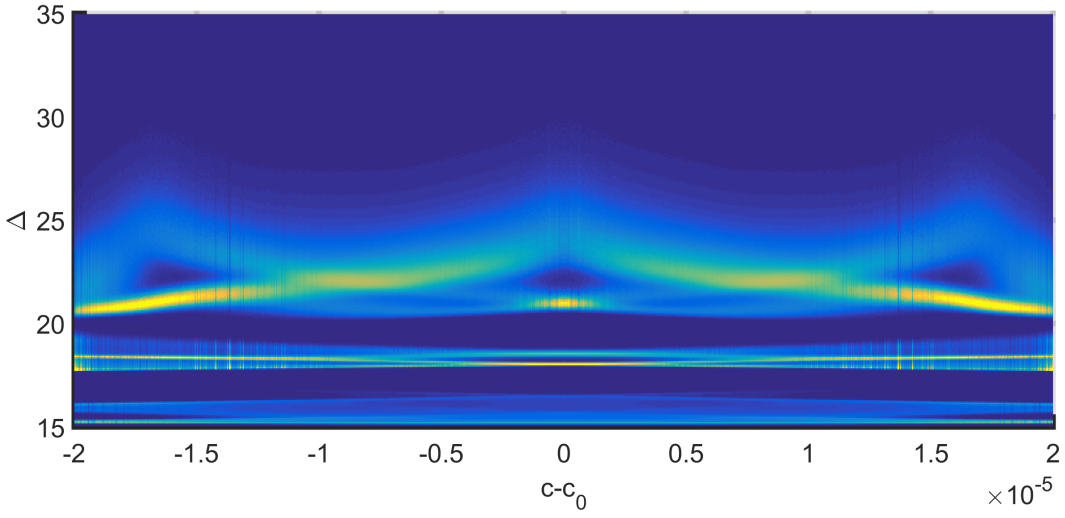

Figure 30: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-6}$. Color from blue to yellow represents low to high values. Scale is arbitrary.

B.5 $\epsilon\sigma = 10^{-5}$

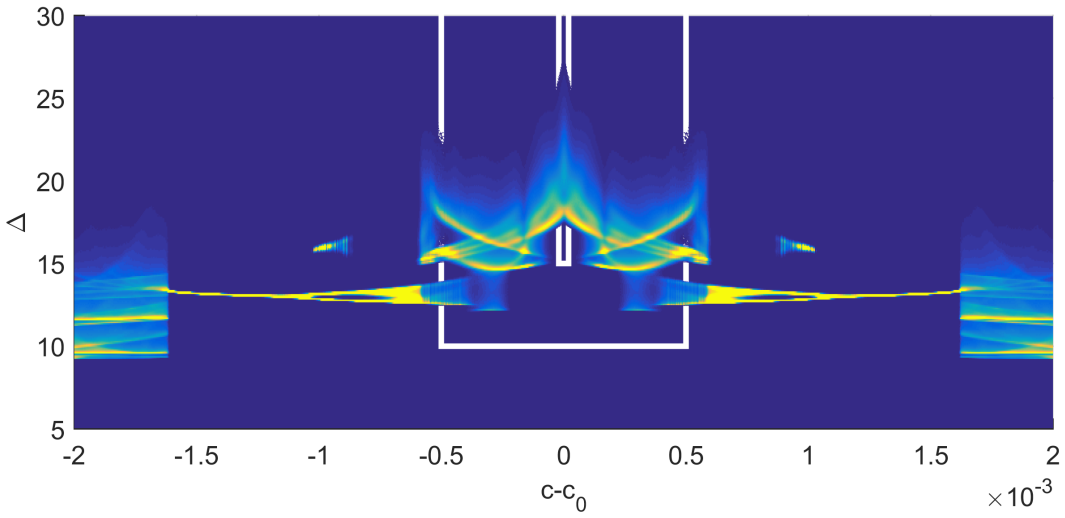

Figure 31: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-5}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 32 and 33.

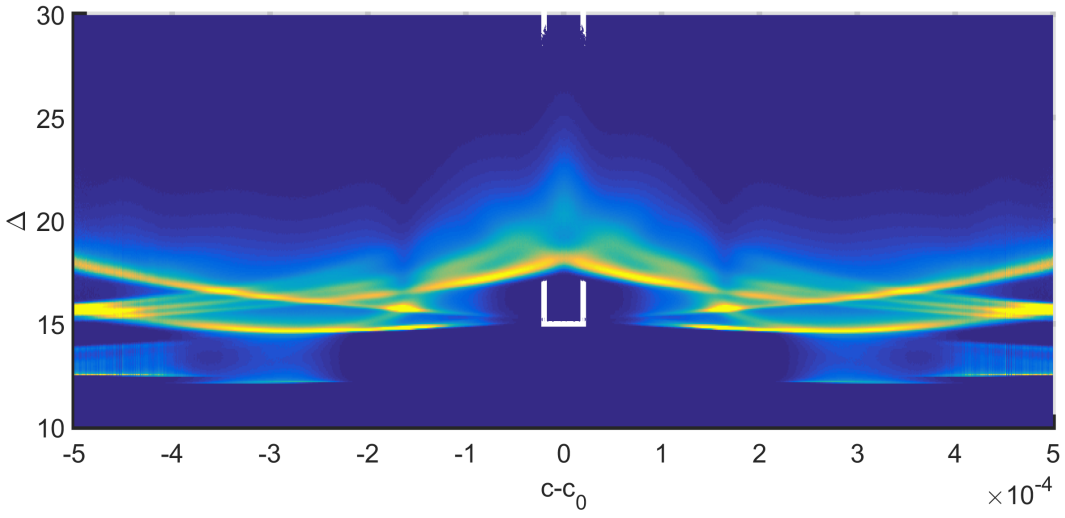

Figure 32: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-5}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 33.

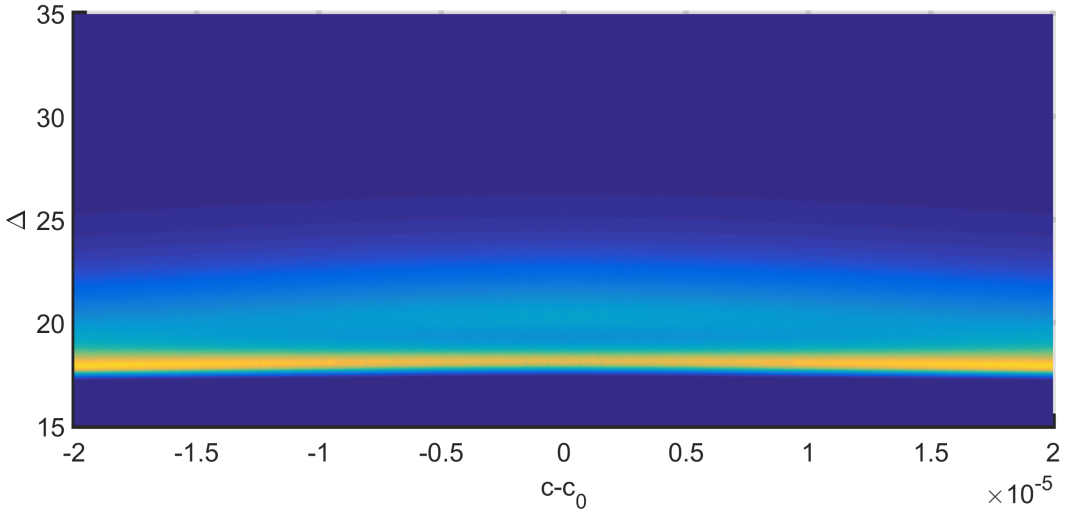

Figure 33: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-5}$. Color from blue to yellow represents low to high values. Scale is arbitrary.

B.6 $\epsilon\sigma = 10^{-4}$

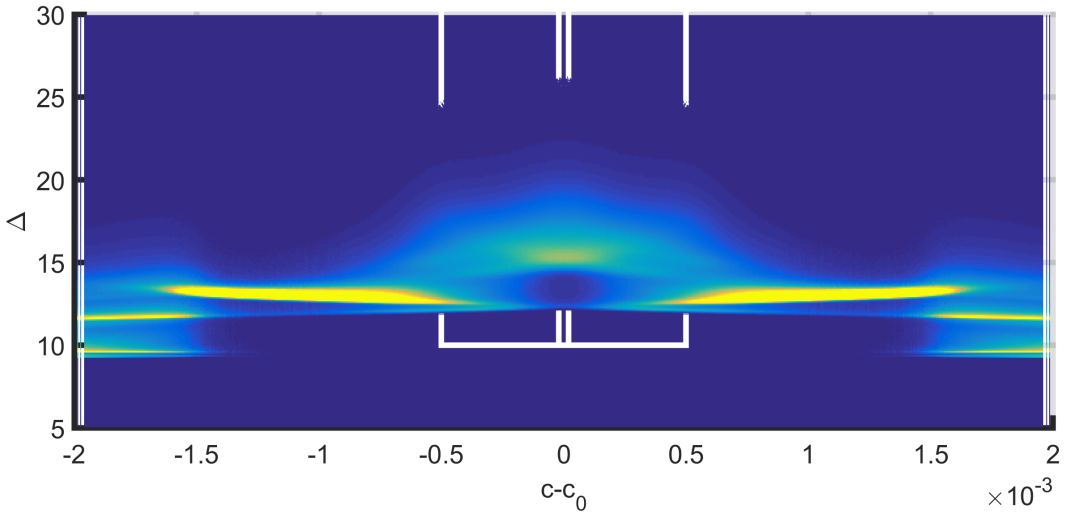

Figure 34: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-4}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show regions plotted in Figures 35 and 36.

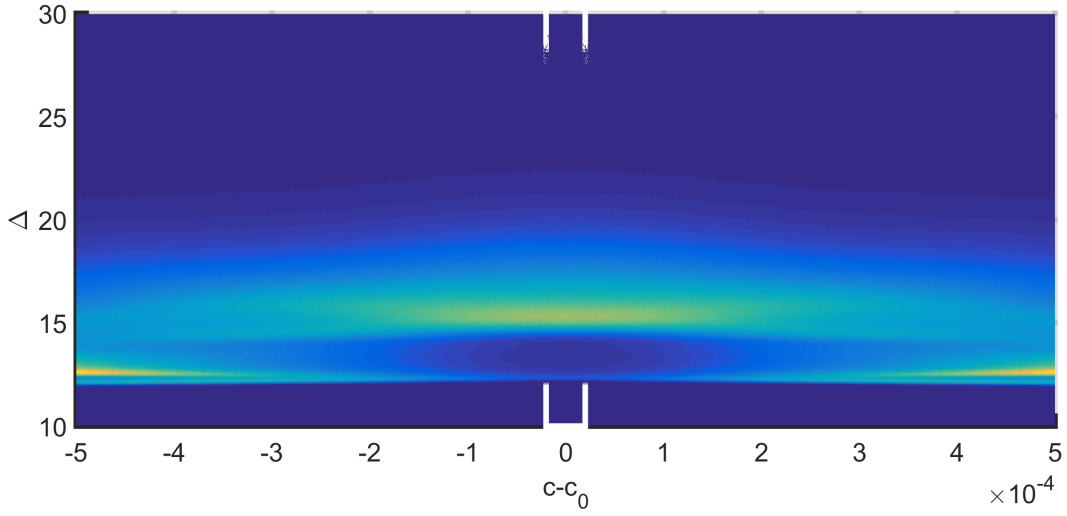

Figure 35: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-4}$. Color from blue to yellow represents low to high values. Scale is arbitrary. White lines show region plotted in Figure 36.

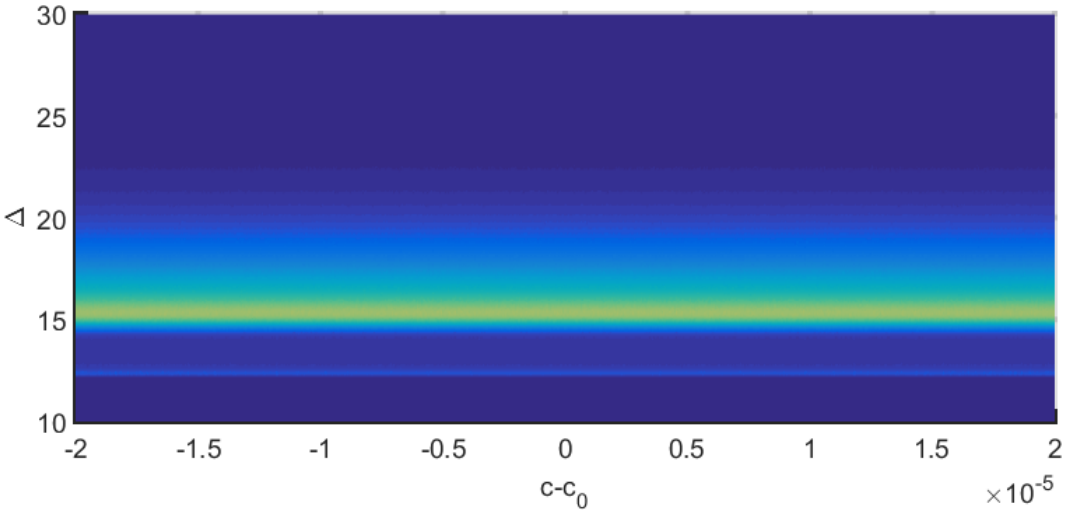

Figure 36: Stationary probability density $\rho_\Delta(\Delta)$ of the timing map (177) for the Shilnikov system for noise strength $\epsilon\sigma = 10^{-4}$. Color from blue to yellow represents low to high values. Scale is arbitrary.

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

$$\rho(x(T))dx(T) = \lim_{N\to\infty}\left[\rho(x^N, T, N)\sqrt{\frac{1}{2\pi\epsilon^2\Delta t}}dx^N\right]$$

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

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

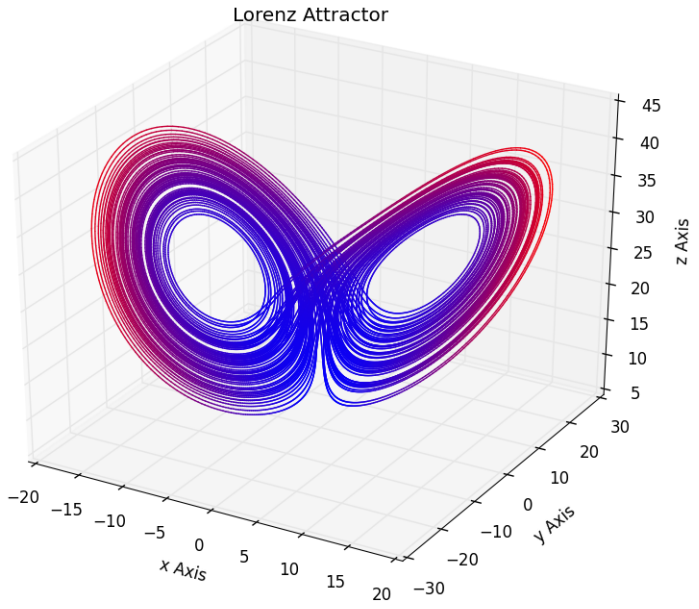

Figure 1: The phase plot of the Lorenz attractor at $r = 28$, $\sigma = 10$, and $b = 8/3$. The colors indicate the relative speed of a particle on the trajectory, where red is "fast" and blue is "slow".

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]

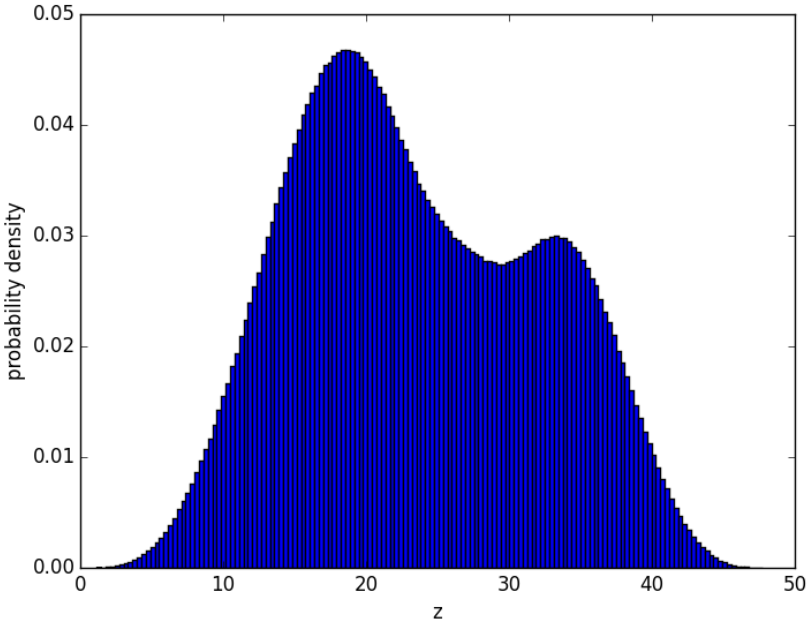

Figure 2: Histogram of Lorenz state variable z on the chaotic attractor for parameter values $r = 28$, $\sigma = 10$, and $b = 8/3$.

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.

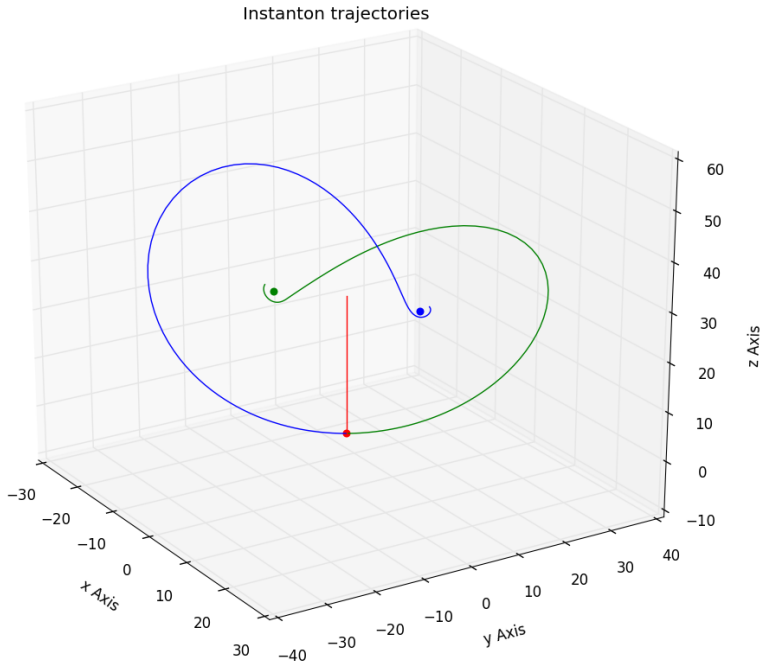

Figure 3: Multiple solutions for Lorenz system instanton trajectories with an initial condition starting at the origin. The blue and green curves appear to be related to heteroclinic connection from the origin (the red dot) to the two fixed points (blue and green dots). The red line is the Ornstein-Uhlenbeck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

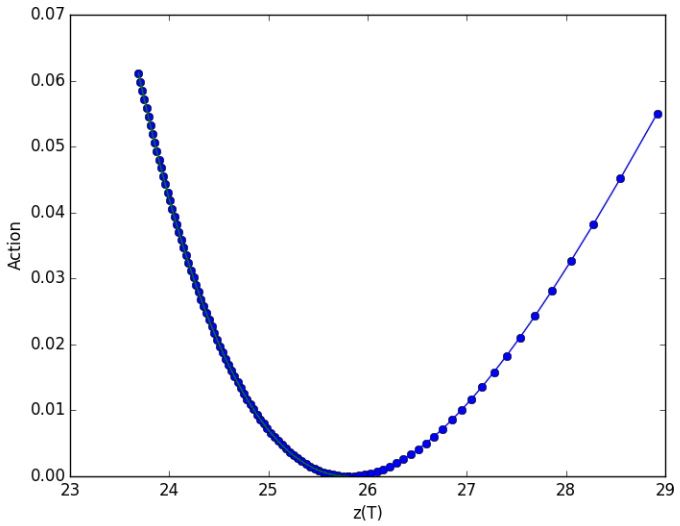

Figure 4: Cost versus final value of z for an initial condition on the attractor and time $T = 1.5$. The observable is taken to be z at the final time.

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.

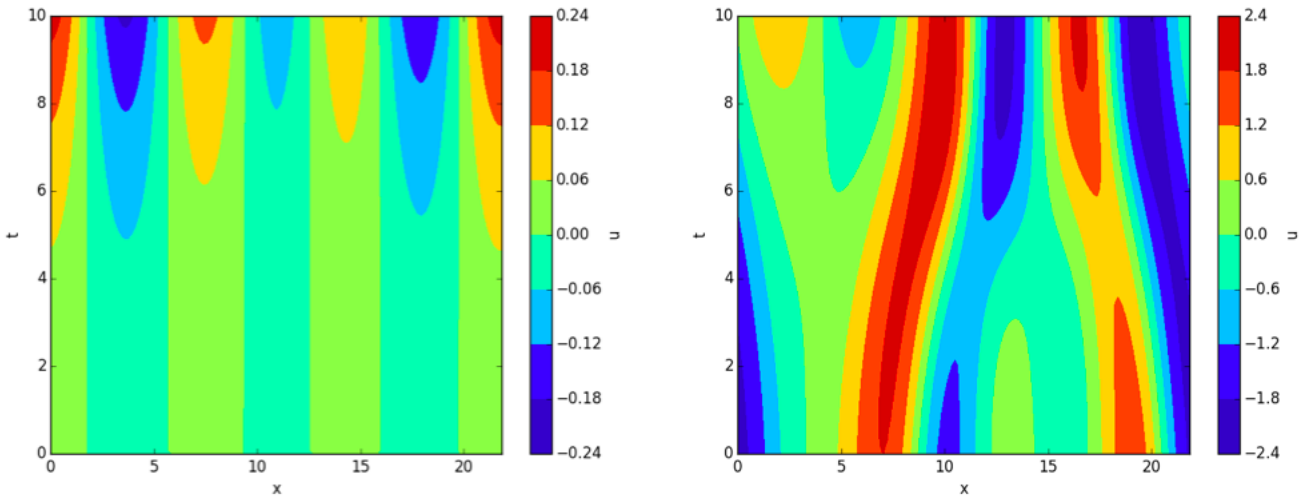

Figure 5: A contour plot of the instanton trajectories for the KS equations. The horizontal axis is the physical spatial coordinate, x, and the vertical axis is time. The color represents the value of the instanton solution u. In both cases the observable at the final time is taken to be $\partial_{xx}u(0,0)$ 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.

Bounds for Deterministic and Stochastic Dynamical Systems Using Sum-of-Squares Optimisation

Giovanni Fantuzzi

October 15, 2015

1 Introduction

One of the challenges in studying physical systems that exhibit complex temporal (or spatio-temporal) dynamics is to obtain rigorous quantitative predictions of the system's behaviour. Given that rigorous, closed-form solution of the governing equations are not generally available, an interesting problem is to quantify the average properties of the system, described by an observable function φ of the system's state $\mathbf{x}$. Note that $\mathbf{x}$ may be infinite-dimensional and depend on position $\mathbf{s}$ and time t. Specifically, one is interested in estimating the value of

$$\langle \varphi(\mathbf{x}) \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi[\mathbf{x}(t)]\, dt \tag{1}$$

or, for spatially-extended systems over a domain Ω with measure $\mu(\Omega)$,

$$\langle \varphi(\mathbf{x}) \rangle = \lim_{T\to\infty} \frac{1}{T}\frac{1}{\mu(\Omega)} \int_0^T \int_\Omega \varphi[\mathbf{x}(\mathbf{s},t)]\, d\mathbf{s}dt, \tag{2}$$

where we assume that the long-time limits exist. Assuming that the system has only one attracting set of dynamical interest (either a stable equilibrium, a periodic orbit or a strange attractor), this amounts to computing the time-average of φ as the system evolves on the attractor, irrespective of the specific initial condition.

Problems of this type have received increasing interest in recent years, focussing mainly on spatially extended systems described by partial differential equations (PDEs) and leading to the development of a variational technique known as the "background method" [4]. The method is based on the optimisation of a quadratic functional of the state variable, denoted by $V(\mathbf{x})$, that allows the derivation of rigorous bounds for $\langle\varphi\rangle$ (either upper or lower); in some cases, V can be interpreted as a Lyapunov function [2]. Classical applications of the background method include the estimation of the net turbulent heat transport in Rayleigh-Bénard convection [6, 7] and the computation of rigorous bounds on the energy dissipation in shear flows [5, 9]. Similar techniques can also be applied to finite-dimensional systems exhibiting chaotic behaviour, such as truncated low-order models of Rayleigh-Bénard convection including the well-known Lorenz system [14, 15].

An alternative approach to derive rigorous bounds for finite-dimensional systems with polynomial dynamics has been proposed recently [3]. The method, introduced in the context of fluid flows [3], is based on Sum-of-Squares (SOS) polynomial optimisation and allows

the construction of higher-than-quadratic Lyapunov-type functionals, generalising the background method.

Whilst the bounds obtained by these methods are rigorous and hold irrespectively of the system's initial condition, they do not generally represent the time averages observed in experiments or numerical simulations accurately. This is usually because the system possesses at least one unstable solution $\mathbf{x}_u(t)$— a fixed point or a periodic orbit — for which the rigorous bounds obtained with the aforementioned techniques are sharp; yet, $\mathbf{x}_u$ is never observed in practice since the real system is always subject to small perturbations. If a neighbourhood $\mathcal{U}$ of $\mathbf{x}_u$ is known to not belong to the system's attractor, a possible solution to this problem is to directly remove $\mathcal{U}$ from the analysis. Another solution, that in principle does not require any a priori knowledge of $\mathbf{x}_u$, is to model the physical disturbances by adding a stochastic forcing term of strength ε to the dynamical system [3].

In this work, we first review some of the ideas introduced in [3] in the context of a general dynamical system with polynomial dynamics. We then investigate how the influence of an unstable solution on the estimates of $\langle\varphi\rangle$ obtained with SOS optimisation can be removed. We will limit ourselves to unstable fixed points and develop the two approaches outlined above for systems with a repelling fixed point; we will not consider the case of saddle points or unstable limit cycles. We will also illustrate how these ideas work in practice by applying them to the well known Van der Pol oscillator.

2 Bounds Using SOS Optimisation: A Review

To make this work self-contained, we start by reviewing the ideas presented in [3]. Consider the dynamical system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n \tag{3}$$

and assume that the trajectories $\mathbf{x}(t)$ are uniformly bounded as $t \to \infty$ regardless of the initial condition $\mathbf{x}_0$. Suppose there exists a function $V[\mathbf{x}(t)]$, continuous in $\mathbf{x}$, and a constant L such that

$$\dot{V} + \varphi - L \geq 0 \tag{4}$$

for all possible values of the state $\mathbf{x}$. Since any trajectory $x(t)$ is uniformly bounded as $t \to \infty$, so is V; hence, time averaging the last expression we obtain

$$\langle\varphi\rangle \geq L. \tag{5}$$

An upper bound U can be found in a similar way by reversing the inequality sign, and we summarise the above in the following:

Proposition 2.1. *Let $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ be a dynamical system whose trajectories are bounded at all times and let $\varphi(\mathbf{x})$ be an observable. If there exist continuous functions $V_u(\mathbf{x})$, $V_l(\mathbf{x})$, and constants U, L such that*

$$\mathcal{D}_u(\mathbf{x}) := \mathbf{f}\cdot\nabla V_u + \varphi - U \leq 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \tag{6a}$$

$$\mathcal{D}_l(\mathbf{x}) := \mathbf{f}\cdot\nabla V_l + \varphi - L \geq 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \tag{6b}$$

then

$$L \leq \langle\varphi\rangle \leq U. \tag{7}$$

The functions V_u and V_l (which we will occasionally refer to as *storage functions*) that achieve given bounds U and L may not be unique; yet, as one could expect, constructing them is generally a challenging task. However the problem is greatly simplified when $\mathbf{f}$ and φ are polynomials of the states x_i, $i \in \{1, ..., n\}$. In fact, if V_u and V_l are chosen to be polynomials, so are $\mathcal{D}_u$ and $\mathcal{D}_l$, hence (6a) and (6b) amount to verifying the non-negativity of a polynomial expression. Whilst this is an NP-hard problem, the computational complexity can be significantly reduced by replacing the conditions $\mathcal{D}_l(\mathbf{x}) \geq 0$ and $-\mathcal{D}_u(\mathbf{x}) \geq 0$ (note the minus sign) with the stronger conditions that $\mathcal{D}_l$ and $-\mathcal{D}_u$ admit a SOS decomposition, i.e. that there exists families of polynomials $\{p_i(\mathbf{x})\}_{i=1}^M$ and $\{q_i(\mathbf{x})\}_{i=1}^N$ such that

$$\begin{aligned} \mathcal{D}_u(\mathbf{x}) &= \sum_{i=1}^{M} p_i(\mathbf{x})^2, \\ \mathcal{D}_l(\mathbf{x}) &= \sum_{i=1}^{N} q_i(\mathbf{x})^2. \end{aligned} \tag{8}$$

These conditions can be formulated in terms of linear matrix inequality (LMI) constraints, a particular type of convex constraint; a brief explanation is given in Appendix A. Optimisation problems with LMI constraints, known as semidefinite programmes (SDPs), can in turn be solved efficiently with a number of software packages, e.g. YALMIP [10] and SOSTOOLS [12]. Consequently, polynomial storage functions and the corresponding bounds U and L may be constructed systematically by solving the SoS optimisation problems

$$\begin{aligned} &\min_{V_u, U} \quad U \\ &\text{such that} \quad U - \mathbf{f} \cdot \nabla V_u - \varphi \in \Sigma \end{aligned} \tag{9}$$

and

$$\begin{aligned} &\max_{V_l, L} \quad L \\ &\text{such that} \quad \mathbf{f} \cdot \nabla V_l + \varphi - L \in \Sigma \end{aligned} \tag{10}$$

where Σ denotes the set of SOS polynomials and the optimisation is over the coefficients of the polynomials V_u and V_l.

3 Improved Bounds for Deterministic Systems

As mentioned in the introduction and as noted in [3], the existence of unstable invariant trajectories $\mathbf{x}_u(t)$ (equilibria and/or limit cycles) poses a problem if one is interested in bounds that accurately describe the time averages measured in experiments. This is because the bounds U and L obtained from Proposition 2.1 must hold for any possible trajectory of the system, including unstable invariant solutions. To illustrate the idea, let $\mathbf{x}_u$ be an unstable equilibrium, such that $\mathbf{f}(\mathbf{x}_u) = 0$; for definiteness, assume that $\varphi(\mathbf{x}_u)$ is much lower than the observed time average $\langle\varphi\rangle$. Then, any lower bound L cannot coincide with $\langle\varphi\rangle$ since evaluating (6b) at $\mathbf{x}_u$ yields

$$\mathcal{D}_l(\mathbf{x}_u) = \varphi(\mathbf{x}_u) - L \geq 0, \tag{11}$$

implying that $L \leq \langle \varphi(\mathbf{x}_u) \rangle = \varphi(x_u) < \langle \varphi \rangle$. Similarly, integrating both sides of the inequality $\mathcal{D}_l(\mathbf{x}_u) \geq 0$ along an unstable limit cycle $\mathbf{x}_u(t)$, the term $\mathbf{f} \cdot \nabla V_l$ vanishes by periodicity and one obtains

$$L \leq \langle \varphi[\mathbf{x}_u(t)] \rangle \tag{12}$$

i.e. the bound is constrained by the time average of φ over the periodic orbit. The same problem arises for any upper bound U if $\varphi(\mathbf{x}_u) > \langle \varphi \rangle$.

One possible solution is to enforce (6a) and (6b) everywhere except for a neighbourhood $\mathcal{U}$ of an unstable invariant solution $\mathbf{x}_u$. This relaxation can indeed be carried out rigorously and implemented if $\mathbf{x}_u$ is known and if all trajectories starting at the perturbed position $\mathbf{x}_0 = \mathbf{x}_u + \delta\mathbf{x}$ permanently leave $\mathcal{U}$ after a finite time $\tau = \tau(\mathbf{x}_0)$.

Let us proceed formally, and assume that the dynamical system (3) has global attractor $\mathcal{A}$ and an unstable solution $\mathbf{x}_u$. Then, $\mathcal{B} = \mathbb{R}^n \setminus \{\mathbf{x}_u\}$ is the basin of attraction of $\mathcal{A}$. Since all trajectories leave $\mathcal{U}$, moreover, $\mathcal{T} = \mathbb{R}^n \setminus \mathcal{U}$ is an absorbing domain. Clearly, $\mathcal{A} \subseteq \mathcal{T} \subseteq \mathcal{B}$ and for any trajectory starting inside $\mathcal{B}$ one has

$$\langle \varphi \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi[\mathbf{x}(t)] \, dt = \lim_{T \to \infty} \frac{1}{T} \int_\tau^T \varphi[\mathbf{x}(t)] \, dt \tag{13}$$

i.e. the time average of φ is completely determined by the dynamics inside $\mathcal{T}$. The same result applies if $\mathbf{x}_u$ is a saddle by letting $\mathcal{U}$ be a neighbourhood of the entire stable manifold $\mathcal{W}_s$ and $\mathcal{B} = \mathbb{R}^n \setminus \mathcal{W}_s$. A trivial extension of Proposition 2.1 is therefore

Proposition 3.1. *Let $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ be a dynamical system in $\mathbb{R}^n$, let $\mathcal{T}$ be a bounded absorbing domain containing an attractor $\mathcal{A}$ and let $\mathcal{B}$ be the basin of attraction of $\mathcal{A}$. If there exist continuous functions $V_u(\mathbf{x})$, $V_l(\mathbf{x})$, and constants U, L such that*

$$\mathcal{D}_u(\mathbf{x}) = \mathbf{f} \cdot \nabla V_u + \varphi - U \leq 0, \qquad \forall \mathbf{x} \in \mathcal{T}, \tag{14a}$$

$$\mathcal{D}_l(\mathbf{x}) = \mathbf{f} \cdot \nabla V_l + \varphi - L \geq 0, \qquad \forall \mathbf{x} \in \mathcal{T}, \tag{14b}$$

then for any initial condition $\mathbf{x}_0 \in \mathcal{B}$

$$L \leq \langle \varphi \rangle \leq U. \tag{15}$$

The problem of eliminating the influence of $\mathbf{x}_u$ on the bounds therefore reduces to that of finding a suitable absorbing domain for the attracting set. This is generally not a trivial task; however, when $\mathbf{f}(\mathbf{x})$ is polynomial an absorbing domain $\mathcal{T}$ may be constructed using SOS techniques [16]. Moreover, $\mathcal{T}$ is generally a semi-algebraic set; for clarity, let us assume that $\mathcal{T} = \{\mathbf{x} \,|\, g(\mathbf{x}) \geq 0\}$ for some polynomial g.

Equation (14b) then requires that $\mathcal{D}_l(\mathbf{x}) \geq 0$ when $g(\mathbf{x}) \geq 0$ (a similar argument holds for (14a) and will not be considered for brevity). It is easy to see that this condition is satisfied if there exists a non-negative polynomial $s(\mathbf{x})$ such that $\mathcal{D}_l(\mathbf{x}) - s(\mathbf{x})g(\mathbf{x}) \geq 0$. Whilst not necessary, this approximation — known as the generalised $\mathcal{S}$-procedure [16, Lemma 2.1] — allows the formulation of two SOS optimisation problems from Proposition 3.1 as

$$\begin{aligned} \min_{V_u, U, s} \quad & U \\ \text{such that} \quad & U - \mathbf{f} \cdot \nabla V_u - \varphi - s\, g \in \Sigma \\ & s \in \Sigma \end{aligned} \tag{16}$$

and

$$\begin{aligned} &\max_{V_l, L, s} \quad L \\ \text{such that} \quad &\mathbf{f} \cdot \nabla V_l + \varphi - L - s\, g \in \Sigma \\ &s \in \Sigma \end{aligned} \tag{17}$$

The $\mathcal{S}$-procedure generalises to more complicated semi-algebraic absorbing domains; more details and examples can be found in [16]. However, its applicability relies on the knowledge of an absorbing domain that does not contain unstable solutions and that is tractable using SOS techniques. This may not be the case for saddle point with a complicated stable manifold, or if the unstable trajectory cannot be separated from the attractor; an example combining both issues is the unstable saddle equilibrium at the origin in the well-known Lorenz system [17].

4 Bounds for Stochastically-Driven Systems

An alternative approach to eliminate the influence of unstable point on the bounds, proposed by Chernyshenko *et al.* [3], is to model the external disturbances that affect any real system with a small-amplitude stochastic forcing term. If the system is stochastically stable (in the sense of [18]), in fact, one can infer bounds for the original, unperturbed system by studying the vanishing-noise limit.

In Section 4.1, we extend the initial ideas of [3] by considering a stochastic dynamical system forced by finite-amplitude noise, and show how to determine bounds on its statistical properties using SOS programming. Our analysis applies not only in the small-noise limit, but to system which are inherently stochastic. In Section 4.2, we will then study the problem of computing rigorous bounds in the specific case of vanishing noise strength.

4.1 Bounds for system with finite noise

Consider the stochastic dynamical system driven by additive white noise

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \sqrt{2\varepsilon}\, \boldsymbol{\sigma} \boldsymbol{\xi}, \tag{18}$$

where $\boldsymbol{\xi}$ is a standard Wiener process and $\mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^n$. The constant matrix $\boldsymbol{\sigma} \in \mathbb{R}^{n \times n}$ describes the relative effect of each ξ_i on each state x_i, while the overall noise strength ε represents the balance between the deterministic and the stochastic dynamics.

This system can be interpreted as a stochastic perturbation of (3), and its state $\mathbf{x}$ (now a random variable) is described by the system's probability density function (PDF) ρ. We remind the reader that ρ must be a non-negative distribution such that $\|\rho\|_{L^1} = 1$. We assume that the system's trajectories remain bounded at all times, and that a statistical equilibrium is reached so the PDF satisfies the steady Fokker-Planck equation

$$\nabla \cdot (\varepsilon \mathbf{D} \nabla \rho - \mathbf{f} \rho) = 0, \tag{19}$$

where $\mathbf{D} = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$. The stationary expectation of an observable $\varphi(\mathbf{x})$ can be computed as

$$\langle \varphi \rangle_\varepsilon = \int_{\mathbb{R}^n} \rho(\mathbf{x}) \varphi(\mathbf{x}) d\mathbf{x} \tag{20}$$

where we have introduced a subscript ε to indicate that the expectation depends on the overall noise strength. Clearly, L is a lower bound for $\langle\varphi\rangle_\varepsilon$ if

$$\int_{\mathbb{R}^n} \rho\,(\varphi - L)\,d\mathbf{x} \geq 0 \tag{21}$$

Enforcing (19) explicitly with a Lagrange multiplier function $V_l(\mathbf{x})$ and integrating by parts we obtain

$$\begin{aligned}&\int_{\mathbb{R}^n} \rho\,[\nabla\cdot(\varepsilon\mathbf{D}\nabla V_l) + \mathbf{f}\cdot\nabla V_l + \varphi - L]\,d\mathbf{x}\\ &+ \lim_{R\to\infty}\int_{\|\mathbf{x}\|=R} (\varepsilon\,V_l\,\mathbf{D}\nabla\rho - \varepsilon\,\rho\,\mathbf{D}\nabla V_l - \rho\,V_l\,\mathbf{f})\cdot\boldsymbol{\nu}(\mathbf{x})\,dS \geq 0\end{aligned} \tag{22}$$

where $\boldsymbol{\nu}(\mathbf{x})$ is the outwards unit normal to the sphere $\|\mathbf{x}\| = R$ and dS is the surface element. Since we have assumed that the system's trajectories are bounded when $\varepsilon = 0$, it is reasonable to expect that ρ decays exponentially at infinity, so that the boundary term vanishes if V_l does not grow too quickly. Thus, one is left with the condition

$$\int_{\mathbb{R}^n} \rho\,[\nabla\cdot(\varepsilon\mathbf{D}\nabla V_l) + \mathbf{f}\cdot\nabla V_l + \varphi - L]\,d\mathbf{x} \geq 0. \tag{23}$$

Since ρ is non-negative and V_l is arbitrary (up to some controlled-growth conditions at infinity), one could prove that L is a lower bound for $\langle\varphi\rangle_\varepsilon$ if there exists V_l such that the term in brackets is everywhere non-negative. Applying the same argument after reversing the inequality sign gives sufficient conditions for an upper bound U on $\langle\varphi\rangle_\varepsilon$, and we conclude the following:

Proposition 4.1. *Let $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \sqrt{2\varepsilon}\,\boldsymbol{\sigma}\boldsymbol{\xi}$, with $\mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^n$ and $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$, be a stochastic system for which a steady PDF exists and let $\mathbf{D} = \boldsymbol{\sigma}^T\boldsymbol{\sigma}$. If there exist functions V_u and V_l such that*

$$\lim_{R\to\infty}\int_{\|\mathbf{x}\|=R} (\varepsilon V_u\mathbf{D}\nabla\rho - \varepsilon\rho\mathbf{D}\nabla V_u - \mathbf{f}\rho V_u)\cdot\boldsymbol{\nu}(\mathbf{x})\,dS(\mathbf{x}) = 0 \tag{24a}$$

$$\lim_{R\to\infty}\int_{\|\mathbf{x}\|=R} (\varepsilon V_l\mathbf{D}\nabla\rho - \varepsilon\rho\mathbf{D}\nabla V_l - \mathbf{f}\rho V_l)\cdot\boldsymbol{\nu}(\mathbf{x})\,dS(\mathbf{x}) = 0 \tag{24b}$$

and

$$\varepsilon\nabla\cdot(\mathbf{D}\nabla V_u) + \mathbf{f}\cdot\nabla V_u + \varphi - U \leq 0 \qquad \forall\mathbf{x}\in\mathbb{R}^n, \tag{25a}$$

$$\varepsilon\nabla\cdot(\mathbf{D}\nabla V_l) + \mathbf{f}\cdot\nabla V_l + \varphi - L \geq 0 \qquad \forall\mathbf{x}\in\mathbb{R}^n, \tag{25b}$$

then the stationary expectation of the random variable $\varphi(\mathbf{x})$ is bounded by

$$L \leq \langle\varphi\rangle_\varepsilon \leq U. \tag{26}$$

Note that the same result was derived in [3] using an alternative approach and fixing $\boldsymbol{\sigma}$ to be the identity matrix. For a given noise amplitude ε, a SOS relaxation of inequalities (25a) and (25b) yields the optimisation problems for the bounds

$$\begin{aligned}&\min_{V_u,U} \quad U\\ &\text{such that} \quad U - \varepsilon\nabla\cdot(\mathbf{D}\nabla V_u) - \mathbf{f}\cdot\nabla V_u - \varphi \in \Sigma\end{aligned} \tag{27}$$

and

$$\begin{aligned} &\max_{V_l, L} \quad L \\ &\text{such that} \quad \varepsilon\nabla\cdot(\mathbf{D}\nabla V_l) + \mathbf{f}\cdot\nabla V_l + \varphi - L \in \Sigma \end{aligned} \tag{28}$$

Note that we have assumed that the boundary terms vanish when V_u and V_l are polynomials, which is equivalent to the statement that all moments of the distribution ρ order up to the degree of the storage functions exist.

Furthermore, note that inequalities (25a) and (25b) are the same as (6a) and (6b), respectively, with the addition of a second order diffusive term. Consequently, a point $\mathbf{x}_u$ such that $\mathbf{f}(\mathbf{x}_u) = 0$ — corresponding to a fixed point of the deterministic system obtained for $\varepsilon = 0$ — does not constrain the bounds on $\langle\varphi\rangle_\varepsilon$ if V_u and V_l have large enough gradients.

4.2 Bounds for system with vanishing noise

Let us assume that a stochastic system is stable in the sense of [18]. Since stochastic bounds are not constrained by fixed points of the corresponding deterministic system (i.e.for $\varepsilon = 0$), the limit $\varepsilon \to 0$ can be studied to infer bounds on the corresponding deterministic system that are not affected by unstable equilibria.

Unfortunately, in practice polynomial storage functions and the SoS optimisation problems (27) and (28) give tight bounds only when the noise strength ε is relatively large. To illustrate the reason of this limitation, let us assume without loss of generality that $\mathbf{x}_u = 0$ and, for definiteness, consider $\varphi = \|\mathbf{x}\|^2$. To achieve a lower bound greater than the trivial result $L = 0$, say $L \sim \mathcal{O}(1)$, one needs

$$\varepsilon\nabla\cdot(\mathbf{D}\nabla V_l) \sim \mathcal{O}(1) \tag{29}$$

at least in a region near the origin, where $\mathbf{f}(\mathbf{x})$ and $\varphi(\mathbf{x})$ are almost negligible. Similar considerations can be made for V_u. When ε is small, large enough gradients can only be achieved if V_u and V_l are polynomials of very high degree, making the SOS problem numerically intractable.

One would therefore like a parametrisation of V_u and V_l with ε that satisfies (29) in a neighbourhood of the unstable states, and that is suitable for polynomial optimisation.

It turns out that an appropriate functional form to study the case $\varepsilon \to 0$ can be derived if the unstable solution is a repelling equilibrium. Henceforth, we will assume that the deterministic system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ has a repelling (focus or node) fixed point at $\mathbf{x} = 0$, i.e. all eigenvalues of the Jacobian $\mathbf{J}_0 \in \mathbb{R}^{n\times n}$ of $\mathbf{f}$ at the origin have positive real part. Without loss of generality, we will also assume that $\varphi(0) = 0$; this can always be achieved with an appropriate shift in φ. For definiteness, we will consider the problem of finding a lower bound when $\langle\varphi\rangle > 0$ and the bound is constrained by the unstable equilibrium at the origin; the analysis can be trivially extended to upper bounds.

Let us start by assuming that V_l is chosen so that $L = \langle\varphi\rangle_\varepsilon$ exactly. Repeating the derivation of Proposition 4.1 with an equality sign, we see that V_l satisfies

$$\varepsilon\,\nabla\cdot(\mathbf{D}\nabla V_l) + \mathbf{f}\cdot\nabla V_l + \varphi - \langle\varphi\rangle_\varepsilon = 0, \tag{30}$$

with boundary conditions described by (24b). When $\varepsilon \to 0$, this is a singularly perturbed boundary value problem, and the method of matched asymptotic expansions can be used

to study the behaviour of V_l. Rather than determining the solution V_l, however, we are interested in determining its scaling with ε and its approximate functional form near the origin, where we expect large gradients.

The appropriate "inner layer" coordinate stretching is $\mathbf{x} = \varepsilon^{1/2}\hat{\mathbf{x}}$, where $\hat{\mathbf{x}} \sim \mathcal{O}(1)$ as $\varepsilon \to 0$. We therefore expect that, near the origin, we can approximate $V_l = V_l(\varepsilon^{-1/2}\mathbf{x})$.

To determine a suitable functional form, instead, let us consider the intermediate region $\varepsilon^{1/2} \ll x_i \ll 1$, $i \in \{1, ..., n\}$, where, to leading order, equation (30) reduces to

$$\mathbf{x}^T \, \mathbf{J}_0^T \, \nabla V_l = \langle \varphi \rangle_\varepsilon. \tag{31}$$

Introducing a characteristic coordinate s such that

$$\frac{d\mathbf{x}}{ds} = \mathbf{J}_0 \, \mathbf{x}, \tag{32}$$

one has

$$V_l = \langle \varphi \rangle_\varepsilon \, s + const. \tag{33}$$

Moreover, if $\mathbf{v}_i$ and λ_i denote each of the n eigenvectors and eigenvalues of $\mathbf{J}_0$ (with $\Re\{\lambda_i\} > 0$ since the origin is a repellor), the solution of (32) can be written as

$$\mathbf{x} = \sum_{i=1}^{n} A_i \mathbf{v}_i e^{\lambda_i \, s} \tag{34}$$

for some constants A_i, suggesting that s should be some logarithmic function of $\mathbf{x}$.

Finally, since $\mathbf{x} = 0$ is a repelling point, it is not unreasonable to expect that there exists a linear coordinate transformation $\mathbf{x} \to \mathbf{u}(\mathbf{x})$ that makes the dynamics near the origin rotationally invariant; this situation is sketched in Figure 4.2 for a 2D system. In this case, one expects V_l to depend only on the (squared) radius $\zeta(\mathbf{x}) = \|\mathbf{u}(\mathbf{x})\|^2$. Note that ζ is a homogeneous, positive definite quadratic form of $\mathbf{x}$, i.e.

$$\zeta(\mathbf{x}) = \mathbf{x}^T \mathbf{Z} \mathbf{x}, \quad \mathbf{Z} \succ 0 \tag{35}$$

for a symmetric matrix $\mathbf{Z}$ to be chosen appropriately.

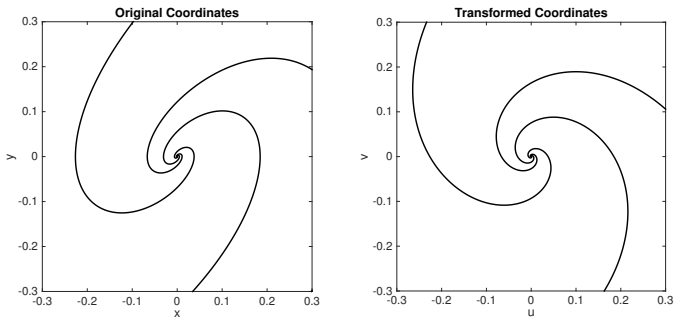

Figure 1: Sketch of trajectories for a 2D system in the original $\mathbf{x} = (x, y)$ coordinates (left) and the transformed $\mathbf{u}(x) = (u(\mathbf{x}), v(\mathbf{x}))$ coordinates.

Combining these heuristic arguments, we suggest that

$$V_l(\mathbf{x}) \approx \alpha \log[\zeta(\mathbf{x})] \tag{36}$$

in the intermediate layer, for some constant α. Consequently, we argue that an appropriate form for V_l is

$$V_l(\mathbf{x}) = \alpha \log[\varepsilon + \zeta(\mathbf{x})] + P_d(\mathbf{x}) \tag{37}$$

Here, ε has been added to the argument of the logarithm to regularise V_l at $\mathbf{x} = 0$ and maintain the correct balance of ε and $\mathbf{x}$ in the inner layer, while P_d is a polynomial of degree d that approximates the outer solution of (30). Moreover, this ansatz could be seen as the generalisation to multiple dimensions of the asymptotic results for one-dimensional systems, presented in Appendix B.

Despite not being polynomial, ansatz (37) is suitable for a SOS formulation. In fact we can substitute

$$\begin{aligned} \nabla V &= \frac{\alpha \nabla \zeta}{\varepsilon + \zeta} + \nabla P_d \\ \nabla \cdot (\mathbf{D} \nabla V) &= \alpha \frac{\nabla \cdot (\mathbf{D} \nabla \zeta)}{\varepsilon + \zeta} - \alpha \frac{\nabla \zeta \cdot (\mathbf{D}\, \nabla \zeta)}{(\varepsilon + \zeta)^2} + \nabla \cdot (\mathbf{D} \nabla P_d) \end{aligned} \tag{38}$$

into (30), multiply by $(\varepsilon + \zeta)^2$ and gather terms to obtain the polynomial inequality

$$\mathcal{L}(\mathbf{x}) := \mathcal{L}_0(\mathbf{x}) + \varepsilon \mathcal{L}_1(\mathbf{x}) + \varepsilon^2 \mathcal{L}_2(\mathbf{x}) + \varepsilon^3 \mathcal{L}_3(\mathbf{x}) \geq 0, \tag{39}$$

where

$$\begin{aligned} \mathcal{L}_0 &= \alpha\zeta\,(\mathbf{f} \cdot \nabla\zeta) + \zeta^2\,(\mathbf{f} \cdot \nabla P_d + \varphi - L)\,, \\ \mathcal{L}_1 &= \alpha\zeta\,\nabla \cdot (\mathbf{D}\nabla\zeta) - \alpha \nabla\zeta \cdot (\mathbf{D}\,\nabla\zeta) + \zeta^2\,\nabla \cdot (\mathbf{D}\nabla P_d) + \alpha \mathbf{f} \cdot \nabla\zeta + 2\zeta\,(\mathbf{f} \cdot \nabla P_d + \varphi - L)\,, \\ \mathcal{L}_2 &= \alpha \nabla \cdot (\mathbf{D}\nabla\zeta) + 2\zeta\,\nabla \cdot (\mathbf{D}\nabla P_d) + \mathbf{f} \cdot \nabla P_d + \varphi - L, \\ \mathcal{L}_3 &= \nabla \cdot (\mathbf{D}\nabla P_d). \end{aligned} \tag{40}$$

Consequently, a lower bound L on $\langle\varphi\rangle_\varepsilon$ can be calculated at a fixed, small ε by solving the optimisation problem

$$\begin{aligned} &\max_{P_d, \alpha, L, \mathbf{Z}} && L \\ \text{such that} \quad &\mathcal{L}(\mathbf{x}) \in \Sigma, \\ &\zeta(\mathbf{x}) = \mathbf{x}^T \mathbf{Z} \mathbf{x}, \quad \mathbf{Z} \succ 0. \end{aligned} \tag{41}$$

Note that when ε becomes small, the terms $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ represent small perturbations of $\mathcal{L}_0$, which is an augmented version of the inequality constraint (25b) for the $\varepsilon = 0$ case. In particular, the term $\alpha\zeta\,(\mathbf{f} \cdot \nabla\zeta)$ represents the contribution of the large gradients and allows us to improve the bounds as $\varepsilon \to 0$.

In fact, we can further develop this idea and derive an optimisation problem which is *rigorous* in the limit $\varepsilon \to 0$, i.e. such that the computed L is a lower bound for $\lim_{\varepsilon \to 0} \langle\varphi\rangle_\varepsilon$. To show this, let us prove that in this limit (23) holds if $\mathcal{L}_0 \geq \gamma\zeta^2$ for any strictly positive

and arbitrarily small constant γ. In fact, since ζ is quadratic in $\mathbf{x}$, it can be verified that $\mathcal{L}_0$ is the dominant term in $\mathcal{L}$ for any fixed $\mathbf{x} \neq 0$. Moreover,

$$\lim_{\varepsilon \to 0} \varepsilon^n \int_{\|x\| \geq r} \rho(\mathbf{x}) \frac{\mathcal{L}_n(\mathbf{x})}{(\varepsilon + \zeta)^2} d\mathbf{x} = 0, \qquad n \in \{1, 2, 3\} \tag{42}$$

for any finite radius r since we have assumed that ρ decays faster than any polynomial. Consequently, if $\mathcal{L}_0 \geq \gamma \zeta^2$ then $\mathcal{L}(\mathbf{x})$ is positive when $\varepsilon \to 0$ at least outside a ball B_R of radius $R \sim \varepsilon^{1/2-\eta}$ with $0 < \eta < 1/2$. We conclude that

$$\int_{\mathbb{R}^n \setminus B_R} \rho \left[\varepsilon \nabla \cdot (\mathbf{D} \nabla V_l) + \mathbf{f} \cdot \nabla V_l + \varphi - L \right] d\mathbf{x} = \int_{\mathbb{R}^n \setminus B_R} \frac{\rho \mathcal{L}}{(\varepsilon + \zeta)^2} \, d\mathbf{x} \geq 0 \tag{43}$$

as $\varepsilon \to 0$. Moreover, although the integrand develops a singularity at $\mathbf{x} = 0$ when $\varepsilon \to 0$ it is possible to show (see Appendix C) that

$$\int_{B_R} \rho \left[\varepsilon \nabla \cdot (\mathbf{D} \nabla V_l) + \mathbf{f} \cdot \nabla V_l + \varphi - L \right] d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0 \tag{44}$$

if ρ is bounded on B_R as $\varepsilon \to 0$ (a reasonable assumption since $\mathbf{x} = 0$ is an unstable equilibrium of the deterministic system). Therefore, a sufficient condition for L to be a valid lower bound on $\langle \varphi \rangle_\varepsilon$ in the limit of vanishing noise is that $\mathcal{L}_0 \geq \gamma \zeta^2$, i.e. (dropping a factor of ζ and rearranging)

$$\alpha \mathbf{f} \cdot \nabla \zeta + \zeta \left(\mathbf{f} \cdot \nabla P_d + \varphi \right) \geq (L + \gamma) \zeta. \tag{45}$$

Note that the role of γ is simply to decrease the tightest possible L by an arbitrarily small constant; consequently, we can drop it from the analysis and determine a rigorous bound L with the optimisation problem

$$\begin{aligned} &\max_{P_d, \alpha, L, \mathbf{Z}} \quad L \\ \text{such that} \quad &\alpha \mathbf{f} \cdot \nabla \zeta + \zeta \left(\mathbf{f} \cdot \nabla P_d + \varphi - L \right) \in \Sigma, \\ &\zeta(\mathbf{x}) = \mathbf{x}^T \mathbf{Z} \mathbf{x}, \quad \mathbf{Z} \succ 0. \end{aligned} \tag{46}$$

Finally, note that ζ is an unknown quadratic form, so the optimisation problem is bilinear. An optimal ζ (denoted by ζ^*) could be determined using a bilinear SDP solver. However, since the SOS constraint is homogeneous in ζ, any choice of $\zeta = \beta \zeta^*$ for $\beta > 0$ is optimal; this issue may be resolved by adding a constraint on the coefficients of ζ. A simpler solution is to fix ζ a priori according to the following observation. In a neighbourhood of the origin where $x, y \ll 1$, the SOS constraint becomes, to leading order in $\mathbf{x}$,

$$\alpha \tilde{\mathbf{f}} \cdot \nabla \zeta - L \zeta \geq 0 \tag{47}$$

where $\tilde{\mathbf{f}} = \mathbf{J}_0 \, \mathbf{x}$ denotes the linearised dynamics near the origin. Therefore, if L is to be positive we require

$$\alpha \tilde{\mathbf{f}} \cdot \nabla \zeta > 0, \tag{48}$$

i.e. that $\alpha \dot{\zeta}$ is positive near the unstable point.

A suitable ζ can be constructed if $\mathbf{J}_0$ can be diagonalised. Specifically, let $\mathbf{U}$ denote the matrix of eigenvectors of $\mathbf{J}_0$ and $\mathbf{\Lambda}$ be the usual diagonal matrix of eigenvalues, such that

$$\mathbf{U}^{-1}\mathbf{J}_0\mathbf{U} = \mathbf{\Lambda}, \tag{49}$$

and let $\mathbf{w} = \mathbf{U}^{-1}\mathbf{x}$. Then, an appropriate choice is

$$\zeta = \mathbf{w}^T\mathbf{w} = \mathbf{x}^T[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x}, \tag{50}$$

since near the origin we have

$$\begin{aligned}
\dot{\zeta} &= 2\dot{\mathbf{x}}^T[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x} \\
&= 2\mathbf{x}^T\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x} + \text{h.o.t.} \\
&= 2\mathbf{x}^T[\mathbf{U}\mathbf{U}^{-1}]^T\,\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x} + \text{h.o.t.} \\
&= 2\mathbf{w}^T[\mathbf{U}^{-1}\mathbf{J}_0\,\mathbf{U}]^T\mathbf{w} + \text{h.o.t.} \\
&= 2\mathbf{w}^T\mathbf{\Lambda}\mathbf{w} + \text{h.o.t.},
\end{aligned} \tag{51}$$

Neglecting the higher order terms near $\mathbf{x} = 0$ and recalling that $\mathbf{\Lambda}$ is positive definite since we are considering a repelling fixed point, we conclude that $\dot{\zeta} > 0$ and so (48) holds for $\alpha > 0$. Note, however, that this is not the only possible choice of ζ; see Appendix E for more examples.

4.3 Equivalence with the $\mathcal{S}$-procedure

Whilst the vanishing-noise formulation presented above and the $\mathcal{S}$-procedure of Section 3 have been obtained in completely separate ways, they are in fact related. To see this, consider the inequality

$$\alpha\,(\mathbf{f}\cdot\nabla\zeta) + \zeta\,(\mathbf{f}\cdot\nabla P_d + \varphi - L) \geq 0 \tag{52}$$

more carefully. The inequality is satisfied at $\mathbf{x} = 0$; for $\mathbf{x} \neq 0$, divide by ζ, rewrite the first term as a time-derivative and rearrange the terms to obtain

$$(\mathbf{f}\cdot\nabla P_d + \varphi - L) + \left(\frac{\alpha}{\zeta}\right)\dot{\zeta} \geq 0. \tag{53}$$

Having already noted that $\alpha > 0$ when ζ is as in (50), we recognise that this is a particular form of the more general $\mathcal{S}$-procedure

$$(\mathbf{f}\cdot\nabla P_d + \varphi - L) + s(\mathbf{x})\dot{\zeta} \geq 0, \tag{54}$$

i.e. we are imposing the inequality

$$\mathbf{f}\cdot\nabla P_d + \varphi - L \geq 0 \tag{55}$$

obtained from Proposition 2.1 *outside* a region $\mathcal{R}$ where $\dot{\zeta} > 0$. This represents a region of repulsion for the unstable fixed point, in which ζ acts as an "inverse" Lyapunov function. Consequently, (52) can be viewed as an application of Proposition 3.1 with $\mathcal{T} = \mathbb{R}^n \setminus \mathcal{R}$. Adding noise to a dynamical system with a repelling fixed point and using the logarithmic

ansatz (37) for the storage function V_l is therefore equivalent to carrying out an $\mathcal{S}$-procedure. Consequently, we expect that the larger the region of repulsion defined by the level sets of ζ, the better the lower bound L for a given degree of P_d in (37) — an observation that may assist the construction of a good ζ.

This equivalence could be expected since $\mathcal{L}_0$ in (39) does not inherit any noise-related terms from the full formulation. This is the result of the particular choice (37) of the form of V_l. If ε is taken as a small but finite value, however, the addition of noise is not equivalent to the $\mathcal{S}$-procedure. In this case, an alternative form of V_l has to be considered to keep the influence of noise when $\epsilon \to 0$.

5 Application to the Van der Pol Oscillator

Let us illustrate how the ideas presented so far can be applied in practice by considering the Van der Pol oscillator

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0, \tag{56}$$

or, in state-space representation,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} := \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{f}(\mathbf{x}) := \begin{pmatrix} y \\ \mu(1 - x^2)y - x \end{pmatrix}. \tag{57}$$

Here, $\mu > 0$ represents the strength of the nonlinear damping force. We are interested in finding upper and lower bounds for $\varphi = x^2 + \dot{x}^2 = x^2 + y^2 = \|\mathbf{x}\|^2$, a measure of the total (potential plus kinetic) energy in the system. As is well known, for any μ the equilibrium position $\mathbf{x} = 0$ is unstable, and the system settles into periodic oscillations for any initial perturbation (Figure 2).

5.1 Upper Bound for the deterministic oscillator

Upper bounds on $\langle\varphi\rangle$ were computed by solving the optimisation problem (9) for $0.1 \leq \mu \leq 5$ and for a range of polynomial degrees d. We used the SOS module of YALMIP [10] to transform (9) into a semidefinite program (SDP). Initial numerical experiments showed that the resulting SDP is ill-conditioned even for modest polynomial degrees, and cannot

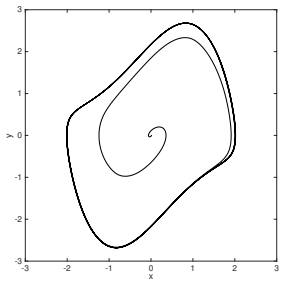

Figure 2: Sample state-space orbit of the Van der Pol oscillator for $\mu = 1$, starting near the unstable origin and converging to the periodic orbit.

be solved reliably by standard double-precision SDP solvers. We therefore resorted to the high-precision solver SDPA-GMP [8]; more details and comments on the numerical implementation can be found in Appendix D.

The upper bounds computed for $d = 6$, $d = 8$, $d = 10$ and $d = 12$ are shown in Figure 3, alongside the values of $\langle\varphi\rangle$ obtained by direct numerical integration of (56). As one would expect, the quality of the bound increases with d; the bounds are within approximately 5% of the simulated value for all values μ considered when $d = 10$, and almost sharp for $d = 12$. The contours of V_u, shown in Figure 4, suggest that better bounds are achieved when the storage function has negative peaks concentrated near the corners of the periodic orbit, where the system evolves slowly. This explains why higher polynomial degrees are necessary to achieve sharp bounds at large values μ, for which the periodic orbit becomes more elongated.

5.2 Lower bound for the deterministic oscillator

For the deterministic system, the trivial lower bound $\langle\varphi\rangle \geq 0$ cannot be improved using Proposition 2.1, since it is saturated by the equilibrium at the origin. In order to apply Proposition 3.1 and the $\mathcal{S}$-procedure to find a tight lower bound for trajectories attracted to the periodic orbit, we need to construct an absorbing domain that does not contain $\mathbf{x} = 0$. As already mentioned in Section 3, SOS optimisation can be used to construct absorbing domains that well approximate the periodic orbit [16]; however, this involves further complication and is beyond the scope of the present investigation. Instead, we will only consider the simple family of domains $\mathcal{T}_r = \{(x, y) \,|\, g(x, y) = x^2 + y^2 - r^2 \geq 0\}$ for $r \leq 1$. To show that these are indeed absorbing domains, let us reverse the direction of time in (57) and consider the energy $E = x^2 + y^2$. One has

$$\begin{aligned} \dot{E} &= 2x\dot{x} + \ y\dot{y} \\ &= -2xy - 2\mu y^2(1 - x^2) + 2xy \\ &= 2\mu y^2(x^2 - 1) \end{aligned} \tag{58}$$

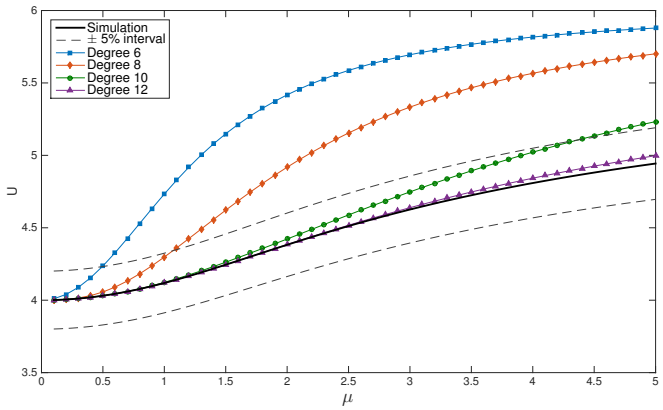

Figure 3: Upper bound for polynomial V_u of varying degree as a function of μ vs. the exact value of $\langle\varphi\rangle$ and the $\pm 5\%$ accuracy intervals.

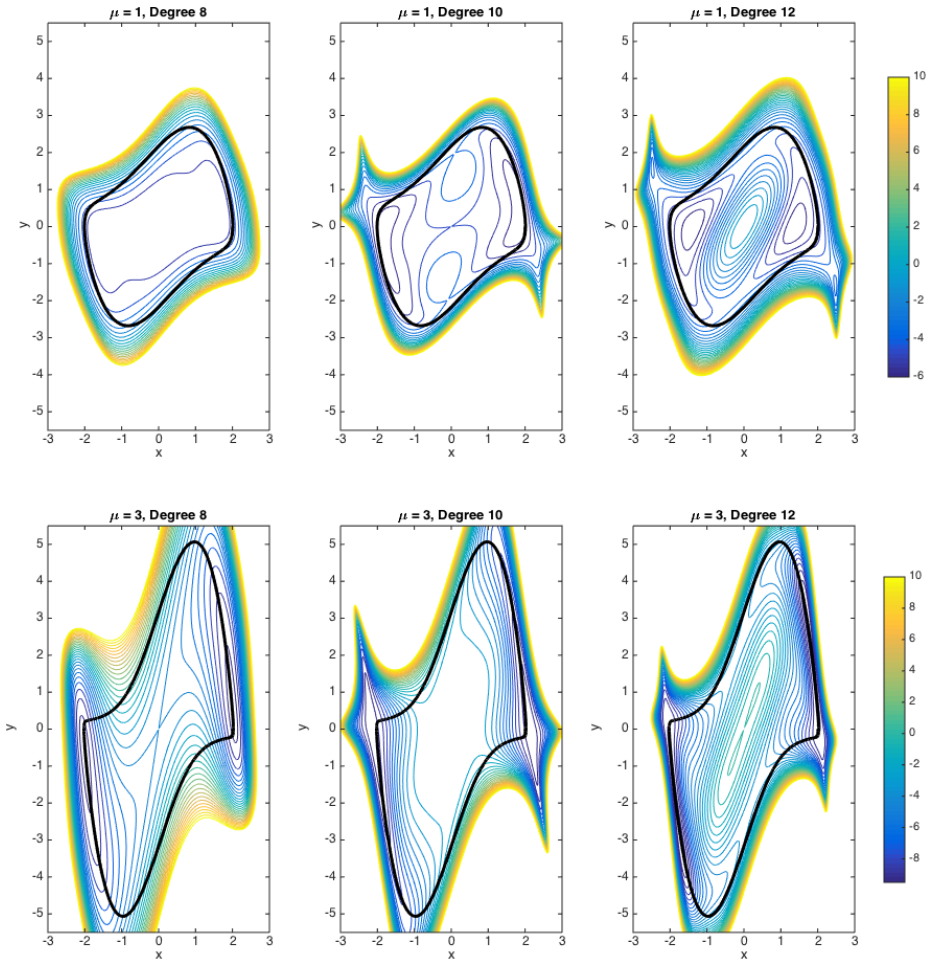

Figure 4: Contours of the optimal V_u for varying polynomial degree. Top: $\mu = 1$. Bottom: $\mu = 3$. The system's periodic orbit (thick black line) is also shown.

meaning that $E \leq 0$ when $|x| \leq 1$. Therefore, any contour of E which is contained in the strip $|x| \leq 1$ defines the boundary of a region of attraction of the origin for the time-reversed oscillator. One concludes that in the original system all orbits will leave the ball $x^2+y^2 < r^2$ if $r \leq 1$, i.e. $\mathcal{T}_r$ is an absorbing domain.

An immediate corollary of this simple proof is the lower bound $\langle\varphi\rangle \geq 1$. While this is already a step forward, more significant improvements and even sharp bounds on $\langle\varphi\rangle$ can be obtained by solving (17).

Initial numerical experiments revealed that it is sufficient to define V_l using monomials of even order only; this is a useful simplification, as it reduces the computational effort and improves the numerical conditioning of the SDP. In all cases, the degree of the $\mathcal{S}$-procedure multiplier s was fixed to be the same as V_l for simplicity.

Figure 5 illustrates the lower bounds computed for storage functions V_l of degree 8, 10 and 12 using the two different absorbing domains $\mathcal{T}_{0.5}$ and $\mathcal{T}_1$. Because $\mathcal{T}_1$ is a better

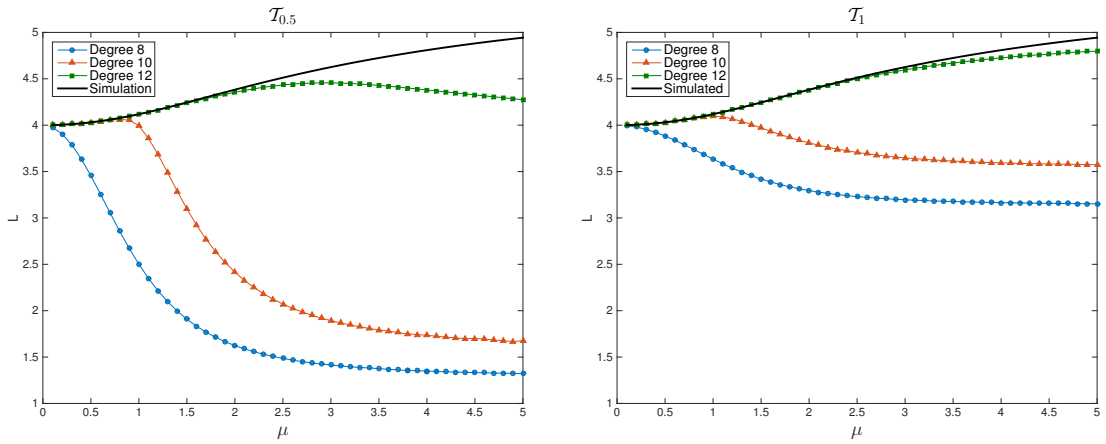

Figure 5: Lower bound computed with the $\mathcal{S}$-procedure for $d = 8, 10, 12$ using the absorbing domains $\mathcal{T}_{0.5}$ (left) $\mathcal{T}_1$ (right).

approximation to the periodic orbit than $\mathcal{T}_{0.5}$, it is not surprising that better bounds are obtained using the former for a given polynomial degree. For the same reason, the bounds worsen as μ increases since the periodic orbit becomes more elongated. We expect that if more sophisticated domain of attractions were computed, for example using the SOS techniques in [16], sharper bounds could be obtained for large μ. Alternatively, bounds of comparable accuracy could be obtained with a lower polynomial degree, reducing the cost of the optimisation.

5.3 Bounds for a stochastic Van der Pol oscillator

Let us now add a stochastic forcing term of strength $\sqrt{2\varepsilon}$ to the deterministic Van der Pol differential equation, i.e. consider

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = \sqrt{2\varepsilon}\,\xi. \tag{59}$$

where ξ denotes white noise. In state-space formulation, this becomes

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \sqrt{2\varepsilon}\begin{pmatrix} 0 \\ \xi \end{pmatrix} \tag{60}$$

Using the notation of Section 4, this corresponds to considering

$$\boldsymbol{\sigma} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{61}$$

First, let us try to compute upper and lower bounds assuming that V_u and V_l are polynomials. Figure 6 shows the results obtained after solving (27) and (28) for values of ε ranging from 10^{-6} to 1 and polynomial degrees 8, 10 and 12. All results were obtained for $\mu = 1$ and defining V_u and V_l using monomials of even order only. The Figure also shows some preliminary values for exact expectation, computed after solving the stationary

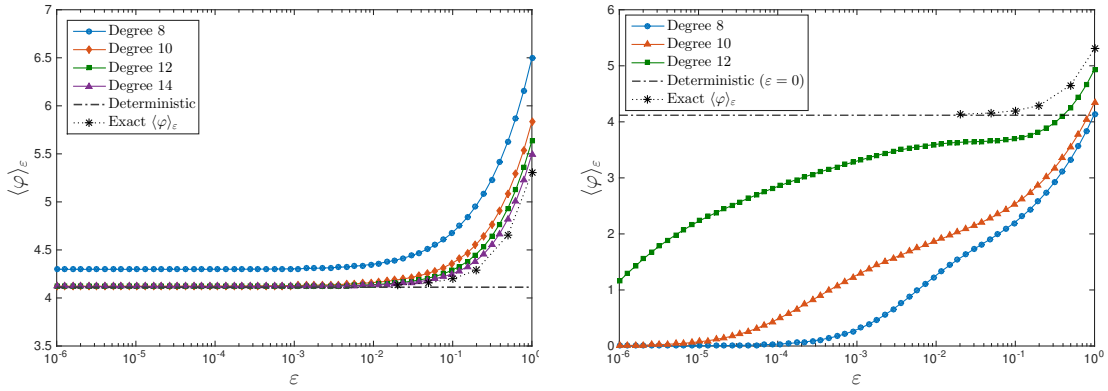

Figure 6: Bounds for the stochastic Van der Pol oscillator for fixed finite noise amplitude. The deterministic ($\varepsilon = 0$) value of $\langle\varphi\rangle$ is shown by the dashed line. Left: upper bound. Right: lower bound.

Fokker-Planck equation (19) using a low-order finite difference method; the steady state is achieved by time-stepping an initial distribution using an implicit Euler scheme, employing operator splitting between x and y derivatives (that is, half a time-step is taken ignoring any derivative in y, then another half time-step ignoring any derivative in x).

As far as the upper bounds are concerned, they are effectively indistinguishable from the value obtained for the deterministic oscillator in Section 5.1 for $\varepsilon < 10^{-3}$ approximately and for polynomial degrees larger than 8. This is not surprising, since (27) reduces to (9) as $\varepsilon \to 0$ when V_u is a polynomial. At larger ε, the bounds are reasonably accurate and capture the increase in $\langle\varphi\rangle_\varepsilon$. Such an increase is indeed consistent with the stronger effect the stochastic forcing has on the deterministic dynamics, suggesting that the well-defined, localised periodic orbit is "smeared" by the noise.

Regarding the lower bound, instead, the addition of noise is effective only when ε is relatively large; for all polynomial degrees, the bound L decreases to 0 as $\varepsilon \to 0$. This is consistent with the observation made in Section 4 that a polynomial V_l of fixed degree cannot have large enough gradients at $\mathbf{x} = 0$ to overcome the decrease in ε.

The numerical difficulties at small ε can be resolved if, instead of a polynomial, V_l is as in (37). For simplicity, rather than trying to determine an optimal ζ, we prescribed $\zeta = x^2 - xy + y^2$ so that the optimisation problem (41) is convex and can be solved using standard SDP solvers. It can be verified that this choice of ζ satisfies (48) for $\mu = 1$; more details can be found in Appendix E. Figure 7 shows the lower bounds computed using (41) for this choice of ζ, $\mu = 1$ and polynomials P_d of degree $d = 8$, $d = 10$ and $d = 12$. The improvement compared to the results obtained with a polynomial storage function are significant, and, for $d = 12$, L is indistinguishable from the deterministic bound when $\varepsilon < 10^{-3}$ approximately. Moreover, we expect that more accurate bounds could be obtained at large ε by increasing the degree of P_d.

Finally, we can compute lower bounds in the limit $\varepsilon \to 0$ by solving optimisation problem (46). We fixed the degree of P_d to 12, and considered different quadratic forms ζ, shown in Table 1. The quadratic forms ζ_2 and ζ_3 were constructed using the eigenvectors of the

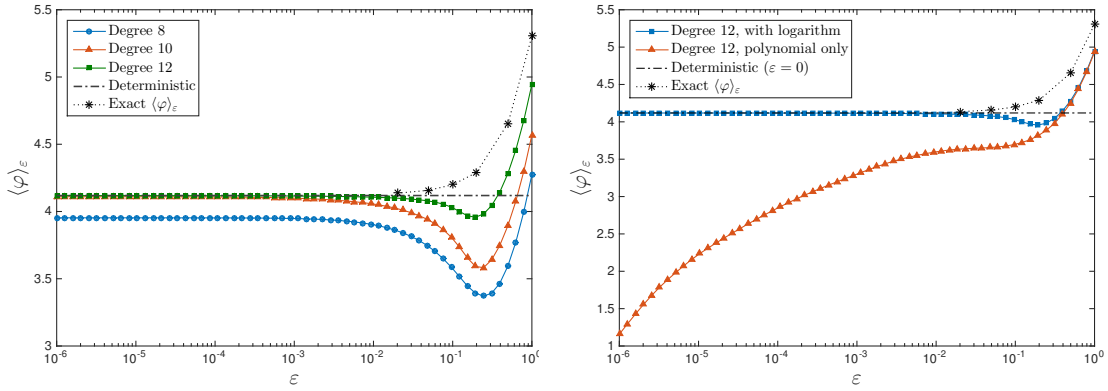

Figure 7: Lower bounds for the stochastic Van der Pol oscillator for fixed finite noise amplitude. Left: bounds for different degrees of P_d. Right: Comparison between the bounds computed from (28) and with (41) for $d = 12$.

Jacobian $\mathbf{J}_0$ (Appendix E); the different formulae for $\mu \leq 2$ and $\mu > 2$ are due to a change from complex to real eigenvectors. Instead, ζ_1 was obtained by arbitrarily fixing $\mu = 1$ in ζ_2. It can be verified that ζ_2 and ζ_3 satisfy (48) near the origin except when $\mu = 2$ (the "critical damping" condition), while ζ_1 satisfies (48) only for $4 - 2\sqrt{3} < \mu < 4 + 2\sqrt{3}$; see Appendix E for more details.

The lower bounds on $\langle\varphi\rangle$, computed as a function of μ, are shown in Figure 8. Overall,

Table 1: Choices of ζ for the Van der Pol oscillator.

	$\mu \leq 2$	$\mu > 2$
ζ_1	$x^2 - xy + y^2$	$x^2 - xy + y^2$
ζ_2	$x^2 - \mu xy + y^2$	$\mu x^2 - 4xy + \mu y^2$
ζ_3	$x^2 - \mu xy + y^2$	$(\mu^2 - 2)x^2 - 2\mu xy + 2y^2$

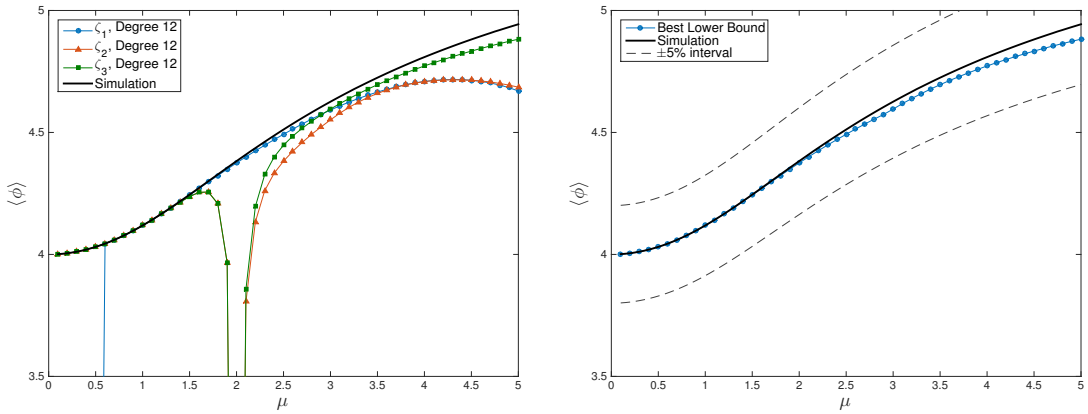

Figure 8: Left: lower bounds computed from (46) for the choices of ζ shown in Table 1 and $\deg(P_d) = 12$. Right: best lower bound compared to numerical integration of (57).

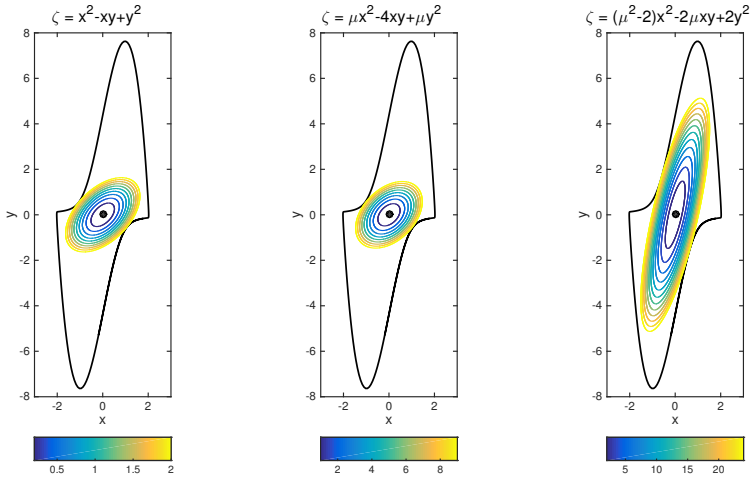

Figure 9: Contours of ζ_1 (left), ζ_2 (centre) and ζ_3 (right) for $\mu = 5$. The system's periodic orbit is also shown (thick black line).

our bounds are well within $\pm 5\%$ of the exact value of $\langle \varphi \rangle$; the poor performance of ζ_1 for $\mu \leq 4 - 2\sqrt{3} \approx 0.54$ and of ζ_2, ζ_3 near $\mu = 2$ was expected based on our previous comments. Moreover, ζ_3 significantly outperforms the other choices for $\mu > 3$ approximately. This can be understood by recalling the equivalence of the logarithmic ansatz and the $\mathcal{S}$-procedure: as shown in Figure 9, the contours of ζ_3 define a better region of attraction for the periodic orbit. Finally, the bounds worsen as μ increases for a fixed polynomial degree and for a fixed ζ, similarly to what was observed for the upper bounds of Section 5.1 and for the lower bounds obtained with the $\mathcal{S}$-procedure in Section 5.2. This could be resolved by increasing the degree of the polynomial P_n in (37) or by a more careful choice of ζ.

6 Further Comments

Although we have tried to keep our work as general as possible, we remark that the analysis of Section 4.2 is only appropriate to eliminate the influence of repelling fixed points; we have not considered the more common cases in which the bounds are constrained by saddle points or unstable limit cycles.

Unfortunately, many systems exhibiting interesting dynamics (such as the Lorenz system) possess unstable saddle points, and one cannot generally expect to successfully apply the techniques we have presented. For example, we expect that the logarithmic functional form we have proposed in Section 4 will not generally be suitable for systems with an unstable saddle point. The reason is that using the logarithmic ansatz is equivalent to implementing an $\mathcal{S}$-procedure, but a region of repulsion around a saddle point cannot generally be defined without including in it the entire stable manifold — normally, a convoluted set that cannot be easily approximated by polynomials. This was confirmed by a brief numerical investigation on the simple system

$$\begin{aligned} \dot{x} &= (x+y)(4-x^2-y^2) \\ \dot{y} &= y(2-x^2-y^2) \end{aligned} \tag{62}$$

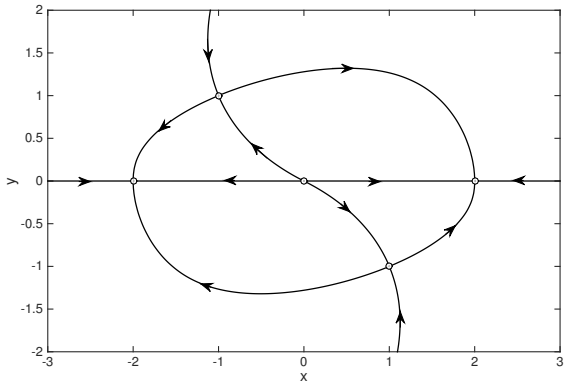

Figure 10: Fixed points of system (62) with their stable and unstable manifolds.

which has a repelling fixed point at the origin, two saddle points at $(\pm 1, \mp 1)$ and two stable equilibria at $(\pm 2, 0)$ as shown in Figure (10); tight bounds on $\varphi = y$ could not be obtained after adding noise to the system even with a logarithmic storage function.

Moreover, we expect that storage functions of impractically high degree will be needed even for noise of relatively large amplitude, making the SOS optimisation problem intractable. This is because the trajectories in the stable manifold will still try to approach the saddle point, forced by the deterministic component of the flow. Then, one expects that each of $\mathbf{f} \cdot \nabla V_u$ and $\mathbf{f} \cdot \nabla V_l$ in inequalities (25a) and (25b) will have opposite signs along the stable and unstable manifolds near the saddle point, unless ∇V_u and ∇V_l can change sign rapidly. This would require polynomial approximations of large degree.

Future work should therefore concentrate on determining an appropriate scaling and functional form for V_u and V_l for systems with saddle points and unstable periodic orbits.

Finally, we remark that the practical implementation of the SOS problems obtained throughout this work poses some technical challenges. Specifically, polynomials of high degree are required to compute relatively sharp bounds, significantly increasing the size of the SDP problems to be solved. This poses a limit on the dimension of the dynamical systems that one can study at a reasonable computational cost and time. Moreover, all SDP problems considered in this work were ill-conditioned, and the results we have presented could only be obtained using computationally expensive high-precision algorithms. Such numerical difficulties should be addressed more systematically in the future if, as it seems inevitable, the theoretical development of appropriate scaling arguments for the storage functions is to be reliably assisted by numerical investigations.

7 Conclusion

To summarise, we have demonstrated that bounds for long-time-averaged properties of systems with polynomial dynamics can be obtained by constructing suitable storage functions using SOS optimisation. Moreover, we have shown that the influence of unstable equilibria on the bounds can be removed via the $\mathcal{S}$-procedure (if a suitable absorbing domain can be defined), or, extending the ideas of [3], by adding noise to a system.

In particular, whilst the formulation of Section 4 holds for a general stochastic system with finite noise strength, a key development is the rigorous formulation of the optimisation problem in the vanishing noise limit for repelling fixed points. If the system is stochastically stable, rigorous bounds for a deterministic system can be inferred when Proposition 3.1 and/or the $\mathcal{S}$-procedure cannot be applied. In this context, we have demonstrated that simple polynomial storage functions are not appropriate to prove bounds that are insensitive to unstable solutions as $\varepsilon \to 0$, and a suitable asymptotic scaling of V_u and V_l with ε should be used.

Despite our successful application of the ideas we have presented to a simple example (the Van der Pol oscillator), some theoretical questions — such as whether it is possible to remove the influence of saddle points on the bounds — and some practical challenges in the implementation of the SOS optimisation remain unresolved. We anticipate that these issues will be the subject of future work, if rigorous bounds are to be obtained for physical systems of practical interest.

8 Acknowledgements

I would like to thank Dr. D. Goluskin for his thorough supervision and the constant support he provided. Moreover, I am grateful to Prof. G. Chini, Prof. S. I. Chernyshenko, Prof. C. R. Doering and Dr. A. Wynn for the numerous discussions and suggestions that helped to shape this work. Finally, I would like to thank all GFD fellows and staff for their patience and help throughout the duration of the programme.

A Introduction to SOS Optimisation

For simplicity, let us consider the problem of determining whether a polynomial $p(x)$ of degree $2N$, i.e.

$$p(x) = \sum_{n=0}^{2N} c_n \, x^n \tag{63}$$

is positive for any $x \in \mathbb{R}$ (the same argument can be generalised to multiple dimensions; for more details, see [13, 11, 1] and references therein). Clearly, a sufficient condition is that p admits a SOS decomposition, i.e. there exists a family of polynomials $\{g_i(x)\}_{i=0}^{M}$ such that

$$p(x) = g_0(x)^2 + g_1(x)^2 + \ldots + g_M(x)^2. \tag{64}$$

It can be shown that this is equivalent to the existence of a positive definite matrix $\mathbf{Q}$ (written as $\mathbf{Q} \succeq 0$) and a vector $\mathbf{z}(x)$ of monomials of x such that

$$p(x) = \mathbf{z}(x)^T \, \mathbf{Q} \, \mathbf{z}(x). \tag{65}$$

For example, if $p(x)$ has degree $2N$ one can take $\mathbf{z}(x)^T = (1, x, x^2, ..., x^N)$. Note that the matrix $\mathbf{Q}$ is generally not unique.

The problem of whether $p(x)$ admits a SOS decomposition can therefore be rewritten as the feasibility semidefinite problem

$$\begin{aligned} & \text{find} \quad \mathbf{Q} \\ \text{such that} \quad & p(x) - \mathbf{z}(x)^T \, \mathbf{Q} \, \mathbf{z}(x) = 0, \\ & \mathbf{Q} \succeq 0, \end{aligned} \tag{66}$$

where the equality constraint $p(x) - \mathbf{z}(x)^T \, \mathbf{Q} \, \mathbf{z}(x) = 0$ is interpreted as a set of equality constraints obtained by setting all coefficients of the difference $p(x) - \mathbf{z}(x)^T \, \mathbf{Q} \, \mathbf{z}(x)$ to zero. Note that these equality constraints are linear with respect to the coefficients c_n of $p(x)$, as well as with respect to the entries of $\mathbf{Q}$. Consequently, semidefinite programming can be used to find values of any unknown coefficients c_n such that $p(x)$ admits a SOS decomposition, or that minimise a linear function $f(c_0, c_1, \ldots, c_N)$ subject to $p(x)$ being a SOS polynomial. Again, more details can be found in [13, 11, 1].

B Asymptotic Analysis for 1D systems

Let us consider the one-dimensional dynamical system $\dot{x} = f(x) + \sqrt{2\varepsilon}\xi$, where $f(x)$ is a polynomial and ξ is a white noise process. Let $\rho_\varepsilon(x)$ be the stationary probability density function of the system, satisfying

$$\frac{\partial(f\rho_\varepsilon)}{\partial x} = \varepsilon \frac{\partial^2 \rho_\varepsilon}{\partial x^2}. \tag{67}$$

Integrating once we obtain

$$f\,\rho_\varepsilon = \varepsilon \frac{\partial \rho_\varepsilon}{\partial x}, \tag{68}$$

where the constant of integration has been set to zero since we must have $\rho \to 0$ as $|x| \to \infty$. This equation can be solved after letting

$$F(x) = \int f(x)\,dx \tag{69}$$

to find

$$\rho_\varepsilon(x) = \mathcal{N}\, e^{\frac{1}{\varepsilon}F(x)}, \qquad \mathcal{N} = \left[\int_{-\infty}^{+\infty} e^{\frac{1}{\varepsilon}F(x)}\,dx\right]^{-1}. \tag{70}$$

Consequently, the expectation of an observable φ can be computed as

$$\langle\varphi\rangle_\varepsilon = \int_{-\infty}^{+\infty} \varphi(x)\,\rho_\varepsilon(x)\,dx = \mathcal{N}\int_{-\infty}^{+\infty} \varphi(x)\, e^{\frac{1}{\varepsilon}F(x)}\,dx. \tag{71}$$

However, we are interested in computing $\langle\varphi\rangle_\varepsilon$ within the framework of Proposition 4.1, in the hope that we can gain some insight to tackle more complicated cases for which the Fokker-Planck equation cannot be solved as easily.

According to Proposition 4.1, $\langle\varphi\rangle_\varepsilon$ can be calculated by finding a function V what satisfies

$$\begin{aligned} \varepsilon\, V'' + f\, V' + \varphi - L_\varepsilon &= 0, \\ \lim_{|x|\to\infty} \left(\rho_\varepsilon\, V'\right) &= 0, \end{aligned} \tag{72}$$

where $(\cdot)'$ denotes differentiation with respect to x and the boundary term has been simplified using (68). As we will see in the following, a solution to this problem can only be found when $L_\varepsilon = \langle\varphi\rangle_\varepsilon$. Changing variable to $W = V'$, we can write an exact general solution to (72) for any value of ε as

$$W(x) = W_0 e^{-\frac{1}{\varepsilon}F(x)} + \frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_0^x \left[L_\varepsilon - \varphi(s)\right]\, e^{\frac{1}{\varepsilon}F(s)}\,ds. \tag{73}$$

The integration constant W_0 is determined by the boundary conditions, which using the expression for ρ_ε in (70) reduce to

$$W_0 + \frac{1}{\varepsilon}\int_0^{+\infty} \left[L_\varepsilon - \varphi(s)\right]\, e^{\frac{1}{\varepsilon}F(s)}\,ds = 0, \tag{74a}$$

$$W_0 - \frac{1}{\varepsilon}\int_{-\infty}^{0} \left[L_\varepsilon - \varphi(s)\right]\, e^{\frac{1}{\varepsilon}F(s)}\,ds = 0. \tag{74b}$$

Note that we have two boundary conditions for a first-order differential equation. In order to satisfy both, we let

$$L_\varepsilon = \langle \varphi \rangle_\varepsilon \tag{75}$$

such that, by definition of $\langle \varphi \rangle_\varepsilon$,

$$\int_{-\infty}^{+\infty} \left[L_\varepsilon - \varphi(x) \right] \, e^{\frac{1}{\varepsilon} F(x)} \, dx = 0. \tag{76}$$

Except from Section B.1 below, we will use (74a) and let

$$W_0 = \frac{1}{\varepsilon} \int_{-\infty}^{0} \left[L_\varepsilon - \varphi(s) \right] \, e^{\frac{1}{\varepsilon} F(s)} \, ds \tag{77}$$

so that W becomes

$$W(x) = \frac{1}{\varepsilon} \, e^{-\frac{1}{\varepsilon} F(x)} \int_{-\infty}^{x} \left(L_\varepsilon - s^2 \right) \, e^{\frac{1}{\varepsilon} F(s)} \, ds. \tag{78}$$

Note that the second boundary condition (74b) is also satisfied by virtue of (76). Alternatively, using (74b) to define W_0 one obtains the equivalent expression

$$W(x) = -\frac{1}{\varepsilon} \, e^{-\frac{1}{\varepsilon} F(x)} \int_{x}^{+\infty} \left(L_\varepsilon - s^2 \right) \, e^{\frac{1}{\varepsilon} F(s)} \, ds. \tag{79}$$

Unfortunately, neither (78) nor (79) give much information about the behaviour and scaling of W for a general dynamical system. We will therefore proceed by discussing some illustrative examples which allow us to draw some important conclusions about the applicability of SOS optimisation to determine sharp bounds for $\langle \varphi \rangle_\varepsilon$.

B.1 Case 1: $f(x) = x - x^3$, $\varphi(x) = x^2$

Let $f(x) = x - x^3$ and $\varphi(x) = x^2$ so that $F(x) = \frac{x^2}{2} - \frac{x^4}{4}$ is an even function with maxima at $x = \pm 1$ and a local minimum at $x = 0$, as shown in Figure 11. In this case, equation (72) can be solved using the method of matched asymptotic expansions (not show here) to find

$$W(x) \approx \underbrace{\frac{1}{x}}_{\text{outer sol.}} + \underbrace{\frac{1}{\sqrt{\varepsilon}} e^{-\frac{x^2}{2\varepsilon}} \int_0^{\frac{x}{\sqrt{\varepsilon}}} e^{\frac{s^2}{2}} \, ds}_{\text{inner sol.}} - \underbrace{\frac{1}{x}}_{\text{common part}}, \tag{80}$$

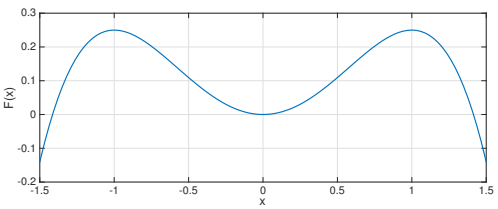

Figure 11: $F(x) = \frac{x^2}{2} - \frac{x^4}{4}$

where the inner solution is valid when $x \sim \sqrt{\varepsilon}\hat{x}$. In the intermediate layer $\sqrt{\varepsilon} \ll x \ll 1$ the solution reduces to the common part and one has

$$W(x) \sim \frac{1}{x} \Rightarrow V(x) = \int W(x)dx \sim \log|x| = \frac{1}{2}\log x^2. \tag{81}$$

This fact was used in Section 4 to justify the use of logarithmic ansatz for V in higher-dimensional systems.

The behaviour of W in the outer and inner regions could also be inferred from the expression for the exact solution. Rather than using (78) or (79), we note that φ is even and add the two conditions in (74) to deduce that $W_0 = 0$ is an appropriate choice for the integration constant in (73). Hence, we have

$$W(x) = \frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_0^x \left(L_\varepsilon - s^2\right)\, e^{\frac{1}{\varepsilon}F(s)}\, ds. \tag{82}$$

Moreover, subtracting (74b) with $W_0 = 0$ from this equation we find

$$W(x) = -\frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_x^{+\infty} \left(L_\varepsilon - s^2\right)\, e^{\frac{1}{\varepsilon}F(s)}\, ds, \tag{83}$$

which is the same as (79).

To study the asymptotic behaviour of W as $\varepsilon \to 0$, we note that W is an odd function so we restrict the attention to $x > 0$. First, we use Laplace's method to estimate

$$L_\varepsilon = \int_{-\infty}^{+\infty} x^2\, \rho_\varepsilon\, dx \sim 1 - \varepsilon + \mathcal{O}(\varepsilon^2). \tag{84}$$

When x is small, precisely $x = \varepsilon^{1/2}\hat{x}$ (where $\hat{x} \sim \mathcal{O}(1)$ and the scaling of ε is suggested by the term $\frac{1}{\varepsilon}F(x) \sim \frac{x^2}{2\varepsilon}$ when x is small) the leading order behaviour of L_ε can be used to estimate

$$W(x) \sim \frac{1}{\sqrt{\varepsilon}}\, e^{-\frac{x^2}{2\varepsilon}} \int_0^{\frac{x}{\sqrt{\varepsilon}}} e^{\frac{s^2}{2}}\, ds. \tag{85}$$

This is the same as the inner solution in (80).

When $x \sim \mathcal{O}(1)$ but $x < 1$, we can use Laplace's method to estimate the integral term in W, where the dominant contribution is given by the end-point x of the integration domain, since F is monotonically increasing over the interval $(0, 1)$ (cf. Figure 11). Recalling that $F' = f > 0$ for $x < 1$, we can show that

$$W(x) \sim \frac{1 - x^2}{x - x^3} = \frac{1}{x}, \tag{86}$$

which corresponds to the outer solution in (80). The same behaviour is found for $x > 1$ using (83).

Finally, we can estimate the behaviour at $x = 1$ to the leading order in ε as

$$\begin{aligned} W(1) &\sim \frac{1}{\varepsilon} e^{-\frac{1}{4\varepsilon}} \int_{-\infty}^{1} \left[1 - \varepsilon - 1 - 2(s-1)\right]\, e^{\left[\frac{1}{4\varepsilon} - \frac{1}{\varepsilon}(s-1)^2 + \ldots\right]}\, ds \\ &\sim 1 + \mathcal{O}(\sqrt{\varepsilon}), \end{aligned} \tag{87}$$

which is consistent with (86). A comparison between a direct numerical integration of (82) and its asymptotic expansion for $\varepsilon = 0.01$ is shown in Figure 12.

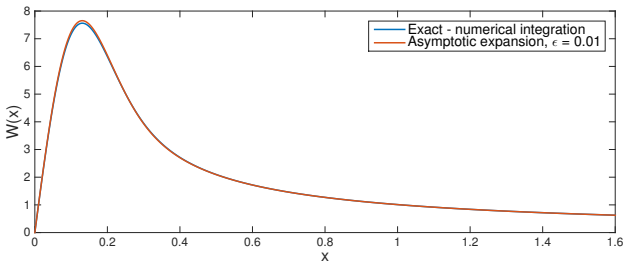

Figure 12: Comparison between the numerical integration and the asymptotic approximation of W, $\varepsilon = 0.01$.

B.2 Case 2: $f(x) = x - x^3$, $\varphi(x) = x$

A simple change from $\varphi = x^2$ to $\varphi = x$ implies a dramatic change in the behaviour of W. The functions $F(x)$ and $\rho_\varepsilon(x)$ are as in the previous example, and the symmetry of ρ_ε implies that

$$L_\varepsilon = \int_{-\infty}^{+\infty} x\, \rho_\varepsilon(x) dx = 0. \tag{88}$$

Equation (72) then becomes

$$\varepsilon W'(x) + f(x)W(x) + x = 0. \tag{89}$$

The method of matched asymptotic expansions fails in this case to produce an approximation to W if one assumes the usual outer solution

$$W_{\text{outer}} = \frac{L_\varepsilon - x}{x - x^3} = \frac{1}{x^2 - 1}, \tag{90}$$

since one cannot construct inner solutions at $x = \pm 1$ that satisfy the matching condition. Let us show this by trying to construct an inner solution at $x = 1$. The appropriate scaling for the inner variables is $x = 1 + \varepsilon^{1/2}\, \hat{y}$ and $W = \varepsilon^{-1/2}\hat{W}$, so the leading-order equation for $\hat{W}$ becomes

$$\hat{W}' - 2\hat{y}\hat{W} + 1 = 0, \quad \hat{y} = \frac{x-1}{\sqrt{\varepsilon}}. \tag{91}$$

Thus, we have

$$W = \frac{1}{\sqrt{\varepsilon}}\hat{W} = \frac{\sqrt{\pi}}{2\sqrt{\varepsilon}}\, e^{\hat{y}^2}\, [A - \operatorname{erf}(\hat{y})], \quad \hat{y} = \frac{x-1}{\sqrt{\varepsilon}}. \tag{92}$$

where erf is the standard error function and A is a constant of integration to be determined so as to match the assumed outer solution (90). Specifically, shifting coordinates $x = 1 + \sqrt{\varepsilon}\hat{y}$ in (90), we require

$$\lim_{\hat{y}\to+\infty} \frac{\sqrt{\pi}}{2\sqrt{\varepsilon}}\, e^{\hat{y}^2}\, [A - \operatorname{erf}(\hat{y})] \sim \frac{1}{2\sqrt{\varepsilon}\hat{y}}, \tag{93a}$$

$$\lim_{\hat{y}\to-\infty} \frac{\sqrt{\pi}}{2\sqrt{\varepsilon}}\, e^{\hat{y}^2}\, [A - \operatorname{erf}(\hat{y})] \sim \frac{1}{2\sqrt{\varepsilon}\hat{y}}. \tag{93b}$$

These conditions cannot be satisfied simultaneously, since the first one requires $A = 1$, while the second requires $A = -1$. Hence, matching is impossible.

This can be explained by considering the exact solution W, which according (78) and (79) can be written as

$$W(x) = \frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_{-\infty}^{x} s\, e^{\frac{1}{\varepsilon}F(s)}\, ds \tag{94a}$$

$$= -\frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_{x}^{+\infty} s\, e^{\frac{1}{\varepsilon}F(s)}\, ds. \tag{94b}$$

Note that this solution is even, so we only need to study its behaviour for $x > 0$.

When $x > 1$, W can be estimated from (94b) using Laplace's method, where the dominant contribution to the integral come from the end-point x of the domain of integration. We obtain

$$W(x) \sim \frac{1}{x^2 - 1}, \tag{95}$$

which corresponds to the usual outer solution (90). This, however, is not the appropriate outer solution when $0 < x < 1$; to show this, we again use Laplace's method on (94b), but this time the dominant contribution comes from $x = 1$. We obtain

$$W(x) \sim \sqrt{\frac{\pi}{\varepsilon}}\, e^{\frac{1}{4\varepsilon} - \frac{1}{\varepsilon}F(x)}, \tag{96}$$

and since $F(x) < \frac{1}{4}$ when $0 < x < 1$ (cf. Figure 11), this means that W behaves like a gaussian (cf. Figure 13).

This behaviour could have been derived from an asymptotic analysis of the original equation by applying the WKB method. If we assume that

$$W(x) = e^{\frac{1}{\varepsilon}\psi(x)} \left[W_0(x) + \varepsilon W_1(x) + ...\right], \tag{97}$$

equation (89) becomes

$$\left[\psi' W_0 + \varepsilon\psi' W_1 + \varepsilon W_0' + f\, W_0 + \varepsilon f\, W_1 + \mathcal{O}(\varepsilon^2)\right] e^{\frac{1}{\varepsilon}\psi} + x = 0. \tag{98}$$

Assuming that $\psi(x) \geq 0$, we can neglect the last term and impose

$$\mathcal{O}(\varepsilon^0): \qquad \psi' W_0 + f\, W_0 = 0 \tag{99}$$

$$\mathcal{O}(\varepsilon^1): \qquad \psi' W_1 + W_0' + f\, W_1 = 0 \tag{100}$$

In order to have a non-zero W_0, we must impose

$$\psi'(x) = -f(x) \implies \psi(x) = K - F(x). \tag{101}$$

Since $-F$ is bounded from below (cf. Figure 11), the integration constant K can indeed be chosen to satisfy $\psi \geq 0$ as originally assumed. It then follows from that (100) that W_0 =constant, and so

$$W(x) = W_0 e^{\frac{K}{\varepsilon}} e^{-\frac{F(x)}{e}} + \dots. \tag{102}$$

Incorporating the constant term $e^{\frac{K}{\varepsilon}}$ into W_0, the WKB outer solution for $0 < x < 1$ can be written to leading order as

$$W_{\text{outer}}(x) = W_0 e^{-\frac{1}{\varepsilon}F(x)}, \tag{103}$$

where W_0 has yet to be determined; note that we recover (96) if we choose

$$W_0 = \sqrt{\frac{\pi}{\varepsilon}}\, e^{\frac{1}{4\varepsilon}}. \tag{104}$$

This choice can indeed be motivated by matching the outer solution

$$W_{\text{outer}} = \begin{cases} W_0 e^{-\frac{1}{\varepsilon}F(x)}, & 0 < x < 1, \\ \dfrac{1}{x^2 - 1}, & x > 1. \end{cases} \tag{105}$$

with the inner solution near $x = 1$ given by (92). In particular, we choose $A = 1$ so that

$$W_{\text{inner}}(x) = \frac{\sqrt{\pi}}{2\sqrt{\varepsilon}}\, e^{\frac{(x-1)^2}{\varepsilon}} \left[1 - \operatorname{erf}\left(\frac{x-1}{\sqrt{\varepsilon}}\right)\right]. \tag{106}$$

It is easy to verify that the inner and outer solution match for $x > 1$, while matching is achieved for $x < 1$ if W_0 is as in (104). A comparison between the exact solution and its composite asymptotic expansion derived combining (105) and (106) is shown in Figure 13 for $\varepsilon = 0.02$ (the two curves are graphically indistinguishable).

B.3 General Third-Order Systems

From the examples above, we may conclude that an asymptotic solution of the equation

$$\varepsilon W'(x) + f(x)W(x) + \varphi(x) - L_\varepsilon = 0 \tag{107}$$

can only be achieved if one consider a more general outer solution than the "normal" outer solution

$$W_{\text{outer}} = \frac{L_\varepsilon - \varphi(x)}{f(x)}. \tag{108}$$

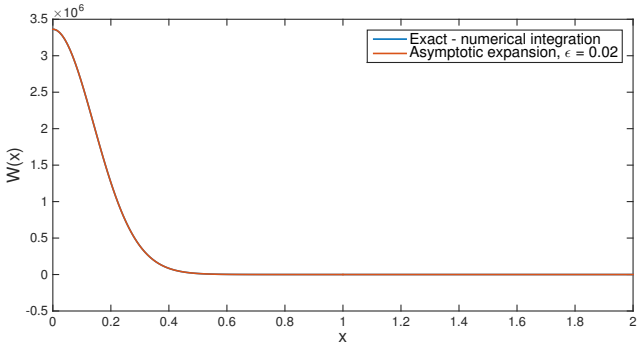

Figure 13: Comparison between the numerical integration and the asymptotic approximation of W, $\varepsilon = 0.02$ (the two curves are graphically indistinguishable).

To illustrate this concept, let us assume that f is a cubic polynomial with zeros at $x_{s,1} < x_u < x_{s,2}$ such that $x_{s,1}$ is the most stable fixed point of the deterministic system $\dot{x} = f(x)$. This means that F has a global maximum at $x_{s,1}$, a local minimum at x_u and a local maximum at $x_{s,2}$ (cf. Figure 14). In general, φ assumes different values at the fixed points and therefore one finds $L_\varepsilon = \varphi(x_{s,1}) + \mathcal{O}(\varepsilon)$. For example, one may consider $f(x) = 2x - x^2 - x^3$, $x_{s,1} = -2$ and $\varphi(x) = x^2$, in which case $L_\varepsilon = 4 + \mathcal{O}(\varepsilon)$.

Recall from (78) and (79) that if $L_\varepsilon = \langle \varphi \rangle_\varepsilon$ the solution of (107) can be written as

$$\begin{aligned} W(x) = & \quad \frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_{-\infty}^{x} [L_\varepsilon - \varphi(s)]\; e^{\frac{1}{\varepsilon}F(s)}\, ds \\ = & \quad -\frac{1}{\varepsilon}\, e^{-\frac{1}{\varepsilon}F(x)} \int_{x}^{+\infty} [L_\varepsilon - \varphi(s)]\; e^{\frac{1}{\varepsilon}F(s)}\, ds. \end{aligned} \tag{109}$$

Let us now study the behaviour of W for a different values of x. When $x \gg x_{s,2}$ an asymptotic analysis using Laplace's method shows that the "normal" outer solution (108) is an appropriate approximation to W (this is similar to our discussion in Section B.2). This approximation is valid as x decreases towards $x_{s,2}$.

Since F has a local maximum at $x_{s,2}$, when x is decreased past $x_{s,2}$ the behaviour of W changes; Laplace's method shows that

$$W(x) \sim \sqrt{\frac{2\pi}{|F''(x_{s,2})|\varepsilon}}\, [L_\varepsilon - \varphi(x_{s,2})]\; e^{\frac{1}{\varepsilon}[F(x_{s,2}) - F(x)]}, \tag{110}$$

i.e. W has an exponential behaviour. In particular, W reaches a maximum at x_u, when the difference $F(x_{s,2}) - F(x)$ is at its maximum (cf. Figure 14).

As x is decreased even further, one reaches a point x_0 at which $F(x_0) = F(x_{s,2})$. When $x < x_0$, the asymptotic behaviour changes again; one can show that, for $x < x_0$, W scales as in (108). This behaviour is maintained until $x = x_{s,1}$. Finally, under our assumption that $L_\varepsilon = \varphi(x_{s,1}) + \mathcal{O}(\varepsilon)$ it is can be shown that (108) holds for $x < x_{s,1}$ and, in fact, for $x = x_{s,1}$ (the analysis is analogous to that of Section B.1).

This analysis allows us to conclude that an asymptotic analysis of (107) should consider (108) as the outer solution for $x < x_0$ and $x > x_{s,2}$, while (110) should hold when

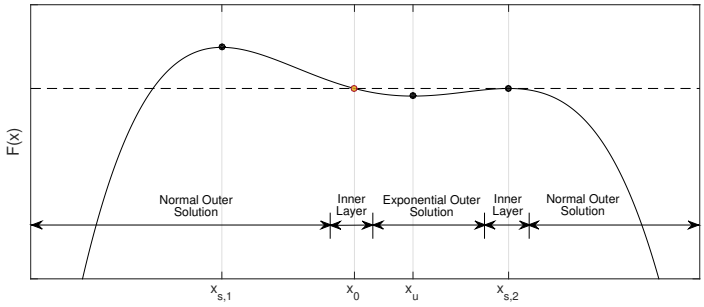

Figure 14: Validity region for different types of outer solutions for W. A typical profile for F satisfying our assumptions is shown. The location of inner layers where the behaviour transition smoothly is also sketched.

$x_0 < x < x_{s,2}$. Inner layers are required when $x \to x_0$ and $x \to x_{s,2}$ for a smooth transition between the different behaviours. The validity regions for each type of outer solutions are illustrated in Figure 14; the location of inner layers, where the behaviour smoothly transitions from one outer solution type to another, is also sketched.

The analysis for the inner layer near $x_{s,2}$ is identical to that carried out in Section B.2 for $x = x_{s,2} = 1$, so we will only consider the inner solution near $x \to x_0$. Letting $y = x - x_0$ and recalling our assumption that f is cubic, we may rewrite

$$f(x) = f(y + x_0) = -(a_0 + b_0 y + c_0 y^2 + d_0 y^3) \tag{111}$$

where the constants $a_0, ..., d_0$ depend on the value of x_0. In particular, $a_0 = -f(x_0) = -F'(x_0)$, and since we have assumed that $\mathbf{x}_{s,1} < x_0 < \mathbf{x}_u$, we conclude that $a_0 > 0$ (this can be seen from Figure 14). Equation (107) then becomes

$$\varepsilon W'(y) - \left(a_0 + b_0 y + c_0 y^2 + d_0 y^3\right) W(y) + \varphi(y + x_0) - L_\varepsilon = 0, \tag{112}$$

The appropriate coordinate stretching for this equation is $y = \varepsilon \hat{y}$, $W_{\text{inner}} = \hat{W}$; then, the inner solution satisfies the equation

$$\varepsilon W'_{\text{inner}} - a_0 W_{\text{inner}} + \varphi(x_0) - L_\varepsilon = 0 \tag{113}$$

and can be written explicitly as

$$W_{\text{inner}}(\hat{y}) = A\, e^{a_0 \hat{y}} - \frac{L_\varepsilon - \varphi(x_0)}{a_0}. \tag{114}$$

Recalling that $a_0 > 0$, it can be verified that this expression matches with (108) as $\hat{y} \to -\infty$ and with (110) as $\hat{y} \to +\infty$ if

$$A = \sqrt{\frac{2\pi}{|F''(x_{s,2})|\varepsilon}}\, [L_\varepsilon - \varphi(x_{s,2})]\; e^{-\frac{1}{\varepsilon}F(x_{s,2})}, \tag{115}$$

which is the required asymptotic behaviour (as illustrated in Figure 14).

B.4 Remarks & Implication for SOS Optimisation

In light of the examples discussed in the previous sections, we conclude that the appropriate form for the function W is highly dependent on both the system's dynamics $f(x)$ and the observable $\varphi(x)$. In general, there exist intervals in which W has an exponential growth of type $e^{\frac{1}{\varepsilon}(\cdot)}$. This behaviour is due to the existence of a stable fixed point x_s at which $L_\varepsilon - \varphi(x_s) \neq 0$ (to leading order in ε) such that standard outer solution (108) cannot be matched asymptotically to any appropriate inner solution. These considerations can be generalised to system with polynomial flows $f(x)$ of degree larger than 3.

The regions where W behaves exponentially, however, disappear when φ assumes the same value at *all* stable points (such as the example studied in Section B.1). In this special case, in fact, the "normal" outer solution does not become singular at any of the stable points, and no inner layers are required. An asymptotic solution similar to that proposed in Section B.1 can then be constructed.

For 1D system with bounded trajectories, one always has multiple stable points, hence these regions of exponential behaviour are generally unavoidable. Thus, one cannot usually approximate W with polynomials or rational functions that capture the correct scaling in the inner layers for arbitrary ε, as the degree of such polynomial approximations would have to be infinite.

The implication of these results is that, in general, one cannot hope to derive sharp lower bounds on $\langle\varphi\rangle_\varepsilon$ using SOS techniques that hold analytically as $\varepsilon \to 0$. Indeed, the computational effort and the order of polynomial approximations to the exact W increase so rapidly that SOS optimisation becomes impractical even when ε is fixed to a small value and not treated analytically.

C Vanishing Noise Limit: Proof of Negligible Contributions

In Section 4.2 we have let

$$V_l(\mathbf{x}) = \alpha \log[\varepsilon + \zeta(\mathbf{x})] + P_n(\mathbf{x})$$

and we have found a polynomial P_n such that

$$\int \rho \left[\varepsilon \nabla \cdot (\mathbf{D}\nabla V_l) + \mathbf{f} \cdot \nabla V_l + \varphi - L\right] d\mathbf{x} \geq 0$$

everywhere except from a ball B_R of radius $R \sim \varepsilon^{1/2-\eta}$ with $0 < \eta < 1/2$; for definiteness, let us write $R = C\varepsilon^{1/2-\eta}$ for some constant C. To complete the argument, we need to show that

$$\int_{B_R} \rho \left[\varepsilon \nabla \cdot (\mathbf{D}\nabla V_l) + \mathbf{f} \cdot \nabla V_l + \varphi - L\right] d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0. \tag{116}$$

Upon substitution of the ansatz, the integral becomes

$$\int_{B_R} \rho \left[\varepsilon \frac{\alpha \nabla \cdot (\mathbf{D}\nabla\zeta)}{\varepsilon + \zeta} - \varepsilon \frac{\alpha \nabla\zeta \cdot (\mathbf{D}\nabla\zeta)}{(\varepsilon+\zeta)^2} + \varepsilon \nabla \cdot (\mathbf{D}\nabla P_n) + \frac{\alpha \mathbf{f} \cdot \nabla\zeta}{\varepsilon + \zeta} + \mathbf{f} \cdot \nabla P_n + \varphi - L\right] d\mathbf{x}.$$

Let us proceed term by term and let us assume that ρ is bounded in B_R uniformly as $\varepsilon \to 0$. Since P and φ are continuous,

$$\left| \int_{B_R} \rho \left[\varepsilon \nabla \cdot (\mathbf{D}\nabla P_n) + \mathbf{f} \cdot \nabla P_n + \varphi - L\right] d\mathbf{x} \right| \leq \frac{4\pi}{3} R^3 \max_{B_R} \left\{\rho |\varepsilon \nabla \cdot (\mathbf{D}\nabla P_n) + \mathbf{f} \cdot \nabla P + \varphi - L|\right\}$$

so

$$\boxed{\int_{B_R} \rho \left[\varepsilon \nabla \cdot (\mathbf{D}\nabla P_n) + \mathbf{f} \cdot \nabla P + \varphi - L\right] d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0} \tag{117}$$

To study the other terms, we switch to polar coordinates, $(x_1, ..., x_n) \to (r, \theta_1, ..., \theta_{n-1})$ where $r \in [0, R]$, $\theta_1, ..., \theta_{n-2} \in [0, \pi]$, $\theta_{n-1} \in [0, 2\pi]$ and

$$d\mathbf{x} = r^{n-1} \sin^{n-2}(\theta_1) ... \sin(\theta_{n-2})\, dr d\theta_1 ... d\theta_{n-1}.$$

Since ζ is a homogeneous, positive definite quadratic form of $\mathbf{x}$ and $\mathbf{D}$ is positive semi-definite (recall that $\mathbf{D} = \boldsymbol{\sigma}^T \boldsymbol{\sigma}$) we can write

$$\zeta(\mathbf{x}) = r^2 F(\theta_1, ..., \theta_{n-1})$$
$$\nabla\zeta \cdot (\mathbf{D}\nabla\zeta) = r^2 G(\theta_1, ..., \theta_{n-1})$$

for some strictly positive function F and non-negative function G, while $\nabla \cdot (\mathbf{D}\nabla\zeta)$ is a real number. Moreover, let

$$F^* = \min_{\theta_1, ..., \theta_{n-1}} F(\theta_1, ..., \theta_{n-1}),$$
$$G^* = \max_{\theta_1, ..., \theta_{n-1}} G(\theta_1, ..., \theta_{n-1})$$

and

$$I = \int_{r=0}^{R} \int_{\theta_{n-1}=0}^{2\pi} \int_{\theta_{n-2}=0}^{\pi} ... \int_{\theta_1=0}^{\pi} \frac{r^{n-1} \sin^{n-2}(\theta_1) ... \sin(\theta_{n-2})}{\varepsilon + r^2 F(\theta_1, ..., \theta_{n-1})} \, dr d\theta_1 ... d\theta_{n-1}.$$

Then, we have

$$\left| \int_{B_R} \frac{\alpha\varepsilon \, \rho \, \nabla \cdot (\mathbf{D}\nabla\zeta)}{\varepsilon + \zeta} d\mathbf{x} \right| \leq \varepsilon \, |\alpha \nabla \cdot (\mathbf{D}\nabla\zeta)| \, \max_{B_R}(\rho) \, I$$

$$\leq \varepsilon \, |\alpha \nabla \cdot (\mathbf{D}\nabla\zeta)| \, \max_{B_R}(\rho) \, 2\,\pi^{n-1} \int_{r=0}^{R} \frac{r^{n-1}}{\varepsilon + r^2 F^*} dr$$

If $n = 2$, the last term can be integrated to give

$$\left| \int_{B_R} \frac{\alpha\varepsilon \, \rho \, \nabla \cdot (\mathbf{D}\nabla\zeta)}{\varepsilon + \zeta} d\mathbf{x} \right| \leq \varepsilon \left\{ |\alpha \nabla \cdot (\mathbf{D}\nabla\zeta)| \, \max_{B_R}(\rho) \, \frac{\pi^{n-1}}{F^*} \log\left(1 + C^2 \, F^* \varepsilon^{-2\eta}\right) \right\}$$

while when $n \geq 3$ we can estimate

$$\left| \int_{B_R} \frac{\alpha\varepsilon \, \rho \, \nabla \cdot (\mathbf{D}\nabla\zeta)}{\varepsilon + \zeta} d\mathbf{x} \right| \leq \varepsilon^{3/2-\eta} \left\{ 2C\pi^{n-1} |\alpha \nabla \cdot (\mathbf{D}\nabla\zeta)| \max_{B_R}(\rho) \max_{r\in[0,R]} \left(\frac{r^{n-1}}{\varepsilon + r^2 F^*} \right) \right\}.$$

It can be verified that the maximum of the last term is achieved at the endpoint $r = R = C\varepsilon^{1/2-\eta}$. Taking the limit shows that for all $n \geq 2$

$$\boxed{\int_{B_R} \frac{\alpha\varepsilon \, \rho \, \nabla \cdot (\mathbf{D}\nabla\zeta)}{\varepsilon + \zeta} d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0} \tag{119}$$

Similarly, we can show

$$\left| \int_{B_R} \frac{\alpha\varepsilon\rho \nabla\zeta \cdot (\mathbf{D}\nabla\zeta)}{(\varepsilon + \zeta)^2} d\mathbf{x} \right| \leq \varepsilon \, |\alpha| 2\pi^{n-1} \max_{B_R}(\rho) \int_{r=0}^{R} \frac{r^{n+1} G^*}{\varepsilon^2 + r^4 F^{*2}} dr$$

$$\leq \begin{cases} \varepsilon \left\{ \dfrac{|\alpha| \pi^{n-1} G^*)}{2F^{*2}} \max\limits_{B_R}(\rho) \, \log(1 + C^4 \, F^{*2} \varepsilon^{-4\eta}) \right\}, & n = 2 \\ \varepsilon^{3/2-\eta} \left\{ 2C|\alpha| \pi^{n-1} G^* \max\limits_{B_R}(\rho) \max\limits_{r\in[0,R]} \left(\dfrac{r^{n+1}}{\varepsilon^2 + r^4 F^{*2}} \right) \right\}, & n \geq 3 \end{cases}$$

where, again, the last maximum is achieved at $r = R = C\varepsilon^{1/2-\eta}$. Hence we deduce

$$\boxed{\int_{B_R} \frac{\alpha\varepsilon\rho \nabla\zeta \cdot (\mathbf{D}\nabla\zeta)}{(\varepsilon + \zeta)^2} d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0} \tag{120}$$

Finally, since $\mathbf{f}(0) = 0$ and $\nabla\zeta$ is linear, the term $\mathbf{f} \cdot \nabla\zeta$ is a polynomial of $\mathbf{x}$ that only contains monomials of degree 2 and higher. Consequently, we can write

$$\mathbf{f} \cdot \nabla\zeta = \sum_{m=1}^{\deg \mathbf{f}} r^{1+m} H_m(\theta)$$

for some continuous functions H_m such that their standard L^∞ norm $\|H_m\|_\infty$ is finite. Each term in this series can be considered separately; for each m we have

$$\left| \int_{B_R} \rho \frac{\alpha r^{1+m} H_m(\theta)}{\varepsilon + \zeta} d\mathbf{x} \right| \leq 2\pi^{n-1} |\alpha| \, \|H_m\|_\infty \max_{B_R}(\rho) \int_0^R \frac{r^{n+m}}{\varepsilon + r^2 F^*} dr$$
$$\leq \varepsilon^{1/2-\eta} \left\{ 2C\pi^{n-1} |\alpha| \, \|H_m\|_\infty \max_{B_R}(\rho) \max_{r\in[0,R]} \left(\frac{r^{n+m}}{\varepsilon + r^2 F^*} \right) \right\},$$

which tends to 0 as $\varepsilon \to 0$ (since $n \geq 2$, $m \geq 1$, and the last maximum is obtained at $r = R = C\varepsilon^{1/2-\eta}$). We therefore conclude that

$$\boxed{\int_{B_R} \rho \frac{\alpha \mathbf{f} \cdot \nabla \zeta}{\varepsilon + \zeta} d\mathbf{x} \to 0 \text{ as } \varepsilon \to 0} \tag{121}$$

Combining (117)-(121) proves (116). Note that the proof presented for estimates (119) and (120) is valid for systems of dimension $n \geq 2$ or higher. For one-dimensional systems, instead, one needs to consider the behaviour of ρ explicitly (cf. Appendix B).

D Remarks on Technical Implementation of SOS Problems

Initial numerical experiments revealed that the SOS problems used to compute upper and lower bounds are particularly ill-conditioned and cannot generally be solved by standard double-precision SDP solvers. From experience, this seems especially true for lower bound problems, and in general for polynomial degrees higher than approximately 4 or 6. The results presented in this work were obtained by pre-processing the SOS problem using the MATLAB toolbox YALMIP [10], and solving the resulting SDP with SDPA-GMP [8] with the parameter `precision` set to `200`. The procedure outlined below was followed for each individual optimisation (we assume the reader is familiar with at least the basic commands of the SOS optimisation module in YALMIP, such as `sossolve`):

1. Set up the SOS problem in YALMIP using `sdpvar` variables and YALMIP's command `sos` to create SOS constraints. For example, a polynomial storage function `V` and the vector of its coefficients `Vcoeffs` are defined with the commands

   ```
   >> [V,Vcoeffs] = polynomial(x,degreeV);
   ```

   In an attempt to reduce the problem size, all SOS polynomials required by the $\mathcal{S}$-procedure were defined using the commands

   ```
   >> z = monolist(x,degreeS);
   >> Sm = sdpvar(size(z,1));
   >> S = z'*Sm*z;
   >> Constraints = [Sm>=0];
   ```

   rather that by using YALMIP's `polynomial` and `sos` commands. This was recommended in [16].

2. Export the SDP to SDPA-GMP using a modified version of YALMIP's `solvesos` function combined with the `export` command. This was done in order to exploit YALMIP's pre-processing capabilities (e.g. Newton polytope and symmetry reductions, [11]).

3. Solve the SDP problem with SDPA-GMP and import the solution back into MATLAB/YALMIP. This is necessary because the output from the solver cannot be easily interpreted in terms of the polynomials and `sdpvar` objects defined by the user in Step 1 above. We remark here that this is likely to introduce numerical errors, due to the different numerical precision used by the solver and by MATLAB.

4. Re-define the problem using a reduced monomial basis, then repeat Steps 1-3. The reduced monomial basis can be obtained after removing unused or zero coefficients from `Vcoeffs`. This is a basic attempt to reproduce YALMIP's post-processing routines [11] and should improve the numerical conditioning. Note that this should not change the problem's solution, simply improve the numerical conditioning and make any computation more reliable. Any significant change in the solution should be interpreted as a warning for numerical problems.

5. Repeat steps 1-4 until the monomial basis cannot be reduced any further.

6. Check the feasibility of the solution by fixing all variables to their optimal value and calling `solvesos` again.

7. [Optional] Check if a certificate of strict positivity exists using Theorem 4 in [11].

This procedure was implemented successfully for most SDPs related to upper bound problems. For lower bound prolems, instead, the numerical solution obtained with SDPA-GMP could not be post-processed reliably with YALMIP, and the feasibility test in step 6 above was almost never successful. However SDPA-GMP never reported infeasibility or numerical problems when solving the original SOS optimisation; in fact, it terminated with the optimal flag `pdOPT`, indicating that an optimal solution could be found. Consequently, the upper and lower bounds computed by SDPA-GMP were considered acceptable.

Further tests were carried out by disabling YALMIP's pre-processing routines; the results were unchanged but the computation time increased significantly. This suggests that the preliminary problem manipulation carried out by YALMIP is beneficial and does not constitute a large source of error. Yet, the solution of the SDP produced by YALMIP requires high-precision. Clearly, one would need to maintain high numerical precision to correctly interpret the data returned by the solver and recover the coefficients of the polynomials as defined originally by the user. Unfortunately, the "dictionary" used by YALMIP to operate this "translation" is not immediately available to the user, and the solver data must be returned to the SOS module in YALMIP if further processing/checking is needed. This, in our opinion, is the main source of numerical errors that prevented us to successfully carry out a feasibility test as in step 6 above.

Finally, a comment on the solver. Whilst SDPA-GMP allowed us to compute the results presented in this report, it does not allow multi-threading and, consequently, we expect that larger SDPs cannot be solved in a reasonable computation time. This is a fundamental issue if higher-dimensional systems are to be studied, as the SDP associated with the SOS formulation of the upper/lower bound problems become very large even for modest polynomial degrees.

Clearly, future work should address these issues more carefully, especially if further analysis of the optimal solution and certificates of true positivity are needed. As a first step, a high-precision parser for SOS problems should be implemented to carry out the necessary pre-processing and to interpret the solution from SDPA-GMP in a user-friendly format. As a long-term suggestion, however, we recommend the development of a high-precision SOS parser with pre- and post-processing capabilities similar to those of YALMIP and of a multi-threaded high-precision SDP solver.

E Constructing ζ from Eigenvector Analysis

Remark: we construct suitable quadratic forms ζ in the context of the Van der Pol Oscillator. However, the same ideas and derivations apply to a general system with a repelling fixed point.

In the vicinity of the origin, the Van der Pol oscillator

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \mu(1-x^2)y - x \end{aligned} \tag{122}$$

can be expanded as

$$\dot{\mathbf{x}} = \mathbf{J}_0\mathbf{x} + \text{h.o.t.}, \qquad \mathbf{J}_0 = \begin{pmatrix} 0 & 1 \\ -1 & \mu \end{pmatrix}. \tag{123}$$

Letting $\mu = 2\nu$, the eigenvalues of $\mathbf{J}_0$ can be written as

$$\lambda_{1,2} = \nu \pm \sqrt{\nu^2 - 1} \tag{124}$$

and the matrix of normalised (i.e. unit norm) eigenvectors is

$$\mathbf{U} = \begin{pmatrix} A & B \\ A\nu + A\sqrt{\nu^2-1} & B\nu - B\sqrt{\nu^2-1} \end{pmatrix} \tag{125}$$

with

$$A = \left[2\nu^2 + 2\nu\sqrt{\nu^2-1}\right]^{-1/2}, \quad B = \left[2\nu^2 - 2\nu\sqrt{\nu^2-1}\right]^{-1/2}. \tag{126}$$

When $\mu = 2$ (i.e. $\nu = 1$), the two eigenvalues and eigenvectors coincide and $\mathbf{J_0}$ cannot be diagonalised; otherwise, we have

$$\mathbf{\Lambda} := \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \mathbf{U}^{-1}\,\mathbf{J}_0\,\mathbf{U}. \tag{127}$$

Note that $\mathbf{\Lambda}$ is a positive definite matrix.

E.1 Construction of ζ_1

We arbitrarily fix $\zeta_1 = x^2 - xy - y^2$, which is clearly positive definite. Moreover,

$$\begin{aligned} \dot{\zeta_1} &= 2x\dot{x} - \dot{x}y - x\dot{y} + 2y\dot{y} \\ &= x^2 - \mu xy + (2\mu - 1)y^2 + \text{h.o.t.} \end{aligned} \tag{128}$$

Neglecting the higher-order terms, the right hand side is positive in the vicinity of the origin if

$$\Delta = \mu^2 y^2 - 4(2\mu - 1)y^2 < 0, \tag{129}$$

which is satisfied for $4 - 2\sqrt{3} < \mu < 4 + 2\sqrt{3}$. Thus, ζ_1 satisfies the necessary condition (48) for $\alpha > 0$ if μ is in this range.

E.2 Construction of ζ_2

Let $\mathbf{w} = \mathbf{U}^{-1}\mathbf{x}$ and consider ζ_2 as

$$\zeta_2 \propto \|\mathbf{w}\|^2 = \mathbf{x}^T[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x}. \tag{130}$$

The (positive) proportionality constant can be chosen arbitrarily, since the constraint in (46) is homogeneous in ζ. In particular, one can chose such constants to write

$$\zeta_2 = \begin{cases} x^2 - \mu xy + y^2, & \mu < 2 \\ \mu x^2 - 4xy + \mu y^2, & \mu > 2 \end{cases} \tag{131}$$

Clearly, ζ_2 is positive definite. Moreover, it satisfies the necessary condition (48) near the origin. In fact,

$$\begin{aligned} \dot{\zeta_2} &\propto 2\mathbf{x}^T\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{U}^{-1}\mathbf{x} + \text{h.o.t.} \\ &= 2\mathbf{w}^T\mathbf{U}^T\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{w} + \text{h.o.t.} \\ &= 2\mathbf{w}^T[\mathbf{U}^{-1}\mathbf{J}_0\,\mathbf{U}]^T\mathbf{w} + \text{h.o.t.} \\ &= 2\mathbf{w}^T\mathbf{\Lambda}\mathbf{w} + \text{h.o.t.}. \end{aligned} \tag{132}$$

Neglecting the higher order terms in a neighbourhood of the origin and recalling that the matrix of eigenvalues is positive definite, we conclude that the last expression is positive. However, this expression cannot be used for $\mu = 2$, since $\mathbf{U}$ cannot be inverted.

E.3 Construction of ζ_3

In order to take into account the dynamics near the origin, and not only the geometric information contained in the eigenvectors, let

$$\zeta_3 \propto \mathbf{w}^T\mathbf{\Lambda}\mathbf{w}^T = \mathbf{x}^T[\mathbf{U}^{-1}]^T\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{x}. \tag{133}$$

Choosing appropriate proportionality constants to simplify the form of ζ_3, we can write

$$\zeta_3 = \begin{cases} x^2 - \mu xy + y^2, & \mu < 2, \\ \left(\mu^2 - 2\right)x^2 - 2xy + 2y^2, & \mu > 2. \end{cases} \tag{134}$$

Clearly, ζ_3 is positive definite. Moreover,

$$\begin{aligned} \dot{\zeta_3} &\propto 2\mathbf{x}^T\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{x} + \text{h.o.t.} \\ &= 2\mathbf{w}^T\mathbf{U}^T\mathbf{J}_0^T\,[\mathbf{U}^{-1}]^T\mathbf{\Lambda}\mathbf{w} + \text{h.o.t.} \\ &= 2\mathbf{w}^T[\mathbf{U}^{-1}\mathbf{J}_0\,\mathbf{U}]^T\mathbf{\Lambda}\mathbf{w} + \text{h.o.t.} \\ &= 2\mathbf{w}^T\mathbf{\Lambda}^T\mathbf{\Lambda}\mathbf{w} + \text{h.o.t.} \\ &= 2\mathbf{w}^T\begin{pmatrix} |\lambda_1|^2 & 0 \\ 0 & |\lambda_2|^2 \end{pmatrix}\mathbf{w} + \text{h.o.t.} \end{aligned} \tag{135}$$

Neglecting the higher order terms in a neighbourhood of the origin, we conclude that the last expression is positive and therefore (48) is satisfied. However, this expression cannot be used for $\mu = 2$, since $\mathbf{U}$ cannot be inverted.

References

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[2] J. C. Bronski and T. N. Gambill, *Uncertainty estimates and L_2 bounds for the Kuramoto-Sivashinsky equation*, Nonlinearity, 19 (2006), pp. 2023–2039.

[3] S. I. Chernyshenko, P. J. Goulart, D. Huang, and A. Papachristodoulou, *Polynomial sum of squares in fluid dynamics : a review with a look ahead*, Phylosophical Transactions of the Royal Society, A., 372 (2014).

[4] P. Constantin and C. R. Doering, *Variational bounds in dissipative systems*, Physica D: Nonlinear Phenomena, 82 (1995), pp. 221–228.

[5] C. R. Doering and P. Constantin, *Variational bounds on energy dissipation in incompressible flows: Shear flow*, Physical Review E, 49 (1994).

[6] ——, *Variational bounds on energy dissipation in incompressible flows. III. Convection*, Physical Review E, 53 (1996), pp. 5957–5981.

[7] C. R. Doering, F. Otto, and M. G. Reznikoff, *Bounds on vertical heat transport for infinite Prandtl number Rayleigh-Benard convection*, Journal of Fluid Mechanics, 560 (2006), pp. 229–241.

[8] K. Fujisawa, M. Fukuda, K. Kobayashi, M. Kojima, K. Nakata, M. Nakata, and M. Yamashita, *SDPA (SemiDefinite Programming Algorithm) and SDPA-GMP Users Manual Version 7.1.1*, tech. rep., Department of Mathematical and Computing Sciences Tokyo Institute of Technology, 2008.

[9] G. I. Hagstrom and C. R. Doering, *Bounds on Surface Stress-Driven Shear Flow*, Journal of Nonlinear Science, 24 (2014), pp. 185–199.

[10] J. Lofberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, in Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.

[11] ——, *Pre-and post-processing sum-of-squares programs in practice*, IEEE Transactions on Automatic Control, 54 (2009), pp. 1007–1011.

[12] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. A. Parrilo, *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, `http://arxiv.org/abs/1310.4716`, 2013. Available from `http://www.eng.ox.ac.uk/control/sostools`.

[13] P. a. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology, 2000.

[14] A. N. Souza and C. R. Doering, *Maximal transport in the Lorenz equations*, Physics Letters A, 379 (2014), pp. 1–13.

[15] A. N. Souza and C. R. Doering, *Transport bounds for a truncated model of Rayleigh-Bénard convection*, Physica D: Nonlinear Phenomena, 308 (2015), pp. 26–33.

[16] W. Tan, *Nonlinear Control Analysis and Synthesis using Sum-of-Squares Programming*, PhD Thesis, University of california, Berkeley, 2006.

[17] M. Viana, *Whats new on Lorenz strange attractors?*, The Mathematical Intelligencer, 22 (2000), pp. 6–19.

[18] L.-S. Young, *What are SRB measures, and which dynamical systems have them?*, Journal of Statistical Physics, 516 (2002), pp. 1–21.

A Correlated Stochastic Model for the Large-scale Advection, Condensation and Diffusion of a Passive Scalar

Tom Beucler

October 15, 2015

1 Introduction

1.1 Distribution of moisture in the atmosphere

At first glance, the distribution of atmospheric water vapor is highly inhomogeneous and varies significantly on very short distances. Lagrangian trajectories of each individual moist parcels in the Tropics strongly depend on the wind field and the intensity of tropical convection, and cloud microphysical processes will have a big influence on the parcel's water content [7]. However understanding the distribution of water vapor in the atmosphere is crucial for several reasons:

- **Its climatic importance**: Water vapor is the main greenhouse gas in the atmosphere, because of its molecular characteristics (eg a permanent dipole moment) and its abundance. Furthermore, the condensation of water produces clouds, which reflect, scatter and absorb solar and terrestrial radiation. Finally, the condensation of water produces latent heat which affects the global atmospheric circulation. Thus, it is impossible to quantitatively study the distribution of clouds [3, 11] or even climate change [19, 21] without understanding the atmospheric water content.
- **Its central meteorological role**: It is impossible to understand weather without knowing the distribution of water vapor and liquid water in the atmosphere. Important weather systems, such as cloud clusters, cyclones, or the Madden-Julian oscillation, are all examples of aggregates of water in the atmosphere. Understanding the self-aggregation of water vapor is a significant research challenge [25, 28], and knowledge of the large-scale advection/distribution of water vapor is required to understand the evolution of moist clusters in time.

The comparison of observations with simple numerical models gives us insights in how atmospheric water vapor is distributed:

- Understanding the mixing occurring on dry/moist isentropic surfaces between the moist Tropics, the dry subtropics, and the sub-saturated mid-latitudes is central to understanding the distribution of water vapor [4, 2, 1, 10, 16].
- The Probability Density Function (PDF) of water vapor is bimodal with a dry and a moist spike [15].

- There is a consensus that microphysical processes do not need to be known precisely to understand the distribution of water vapor, at least outside of the Tropics, and that focusing on computing the advecting wind with precision is more relevant in most cases [5, 20].

The last point justifies the advection-condensation paradigm, which assumes that the distribution of water vapor can be deduced from the velocities in the atmosphere and the saturation specific humidity profile, without a precise knowledge of the microphysical and convective processes producing this profile. It is an interesting alternative to models that need a precise knowledge of the atmospheric properties at each location [17], and was initially used to compute the short term evolution of a given water vapor field in the atmosphere using the observed wind field data. Pierrehumbert et al. [15] started using a stochastic advection-condensation model to understand the fundamentals of water vapor distribution. Assuming a one-dimensional Brownian motion for the air parcels and an exponentially decreasing saturation specific humidity profile, they were able to solve initial value problems and obtain the PDF of moisture numerically. O'Gorman and Schneider [12, 13] added correlation to the stochastic velocity field by assuming that it was governed by an Ornstein-Uhlenbeck process (Gaussian colored noise). They were able to analytically compute the moisture flux and the condensation rate in the Ballistic and Brownian limits; they computed the same quantities numerically in the 1D/2D case. Sukhatme and Young [23] were able to solve for the PDF, the moisture flux and the condensation rate in the general case (giving significant insight in the physics of the process), but they had to come back to a white noise process for the velocity field.

Thus, our first motivating questions are: **What are the physics of the advection-condensation model in the case where the velocity process has a finite time-correlation?** Is it possible to analytically approximate:

- The bimodal PDF of specific humidity?
- The moisture flux?

Understanding how the flux of a condensing scalar behaves in the case of a stochastic process correlated in time has broader applications and leads to new motivations, as we show in paragraph 1.2.

1.2 Diffusivity of a passive scalar

A passive scalar is a diffusive substance with no dynamical effect (such as a change in buoyancy) on the fluid flow that contains it [27]. Under certain conditions, anthropogenically introduced dyes, biological nutrients, and of course water vapor can be approximated as passive scalars. Here, we show how the Lagrangian evolution equation of a passive scalar q influences its diffusion, ie its movement down a given concentration gradient G (in m^{-1}). If the scalar has no sources nor sinks, it can be shown [24] that in a statistically steady state, its flux $\mathcal{F}_q$ verifies Fick's law:

$$\mathcal{F}_q \stackrel{\text{def}}{=} \overline{v'q'} = -DG \tag{1}$$

where $(\overline{X})$ is an appropriate ensemble average on the fluid's parcels, and the deviations from this average are denoted with a prime. v is the parcel's velocity, defined as the Lagrangian time-derivative of its displacement, and assumed to have a mean value equal to zero without loss of generality. In Fick's law, we define the coefficient D (in $\mathrm{m}^2.\mathrm{s}^{-1}$) to be the diffusivity of any passive scalar q. In the absence of sources and sinks:

$$D = \int_0^{+\infty} C(\tau) d\tau \tag{2}$$

where C is the steady covariance of the velocity process v:

$$C(\tau) \stackrel{\text{def}}{=} \overline{v(t=\tau)v(t=0)} \tag{3}$$

and t (in s) is a measure of the time. This simple result shows that the sole understanding of C can fully describe the diffusion of q. Now, let us consider the case where q is linearly damped at a frequency λ. The evolution of the perturbation of q from its gradient G is given by:

$$\frac{dq'}{dt} + \lambda q' + Gv = 0 \tag{4}$$

Integrating this equation in time, applying a proper ensemble average, and assuming that the statistically steady state has been reached, we find that $\mathcal{F}_q$ follows an adapted Fick's law, with a special diffusivity that depends on the damping rate λ:

$$D_\lambda = \int_0^{+\infty} \exp(-\lambda\tau) C(\tau) d\tau \tag{5}$$

In the limit of a null damping ($\lambda = 0$), or a white noise velocity process ($C(\tau) = D_q \delta^+(\tau)$), we recover the case without sources or sinks. In general, linear damping decreases $\mathcal{F}_q$ while keeping it proportional to G (ie Fickian), since the rate of damping λ is the same everywhere in the fluid. The previous generalization of Taylor's diffusivity has applications in biology, where biological properties can be advected by the the flow as well as consumed at a linear rate, and allows to compute the eddy fluxes of the biological properties at all time [9]. If we try to apply the same reasoning to a condensing scalar q, the evolution of the perturbation q' is given by:

$$\frac{dq'}{dt} + Gv + \lambda[q - q^*(y)]\mathcal{H}[q - q^*(y)] = 0 \tag{6}$$

where condensation brings back linearly the scalar q to its saturation value q^* when $q > q^*$. We have defined the Heaviside function $\mathcal{H}$ and the condensation time λ^{-1}. **A key motivation of this work is to find how $\mathcal{F}_q$ deviates from its Fickian upper bound $-DG$ in the case of a condensing scalar.** We immediately see that we need a positive damping ($\lambda > 0$) and a correlated process ($C(\tau > 0) \neq 0$), to avoid the case $D_\lambda = D$ which would not provide any insight on $\mathcal{F}_q$.

2 The Advection-Condensation Model for Specific Humidity

2.1 Specific humidity

We begin by defining the specific humidity q of a parcel as:

$$\boxed{q \stackrel{\text{def}}{=} \frac{m_v}{m_t}} \tag{7}$$

where m_v is the mass of water vapor in the parcel and m_t is the total mass of the parcel. We assume that q is governed by the following advection-condensation equation:

$$\boxed{\frac{Dq}{Dt} = \mathcal{S} - \mathcal{C}} \tag{8}$$

which involves:

- The material derivative on an isentropic surface:

$$\frac{D}{Dt} \stackrel{\text{def}}{=} \frac{\partial}{\partial t} + \overrightarrow{u} \cdot \overrightarrow{\nabla} \tag{9}$$

  where if we define the position vector on the isentropic surface $\overrightarrow{x}$, the velocity on the isentropic surface is defined to be:

$$\overrightarrow{u} \stackrel{\text{def}}{=} \frac{\partial \overrightarrow{x}}{\partial t} \tag{10}$$

  and the gradient is defined to be:

$$\overrightarrow{\nabla} \stackrel{\text{def}}{=} \frac{\partial}{\partial \overrightarrow{x}} \tag{11}$$

- A condensation sink $\mathcal{C}$ that instantly brings back the specific humidity q to its saturation value q^*:

$$\mathcal{C} = \lim_{\lambda \to +\infty} \lambda (q - q^*) \mathcal{H}(q - q^*) \;\Rightarrow\; \boxed{\forall t,\; q = \min(q, q^*)} \tag{12}$$

- A source term $\mathcal{S}$, which we will model in different ways throughout this report.

2.2 An example of saturation specific humidity profile

2.2.1 Saturation specific humidity

The saturation specific humidity q^* can be expressed as a function of the saturation water vapor partial pressure e^* and the environmental pressure p using the ideal gas law (eg [6]):

$$\boxed{q^* = \frac{\epsilon e^*}{p - (1 - \epsilon) e^*}} \tag{13}$$

where ϵ is the ratio of the specific gas constant of dry air r_d and water vapor r_v:

$$\epsilon \overset{\text{def}}{=} \frac{r_d}{r_v} \approx \frac{287 J.K^{-1}.kg^{-1}}{461 J.K^{-1}.kg^{-1}} \approx 0.621 \tag{14}$$

We can then use an integral version of the Clausius-Clapeyron equation if we want to express the saturation water vapor partial pressure e^* as a function of the environmental temperature T. Clausius-Clapeyron relation, valid for an ideal gas undergoing a phase change in thermodynamic equilibrium, can be written:

$$\frac{de^*}{dT} = \frac{L_v e^*}{r_v T^2} \tag{15}$$

where $L_v(T)$ is the latent heat of vaporization of water vapor. An approximate integral of the previous relation is given by Bolton's formula:

$$e^*(p,T) \approx 6.112 \cdot \exp(\frac{17.67 T[°C]}{T[°C] + 243.5}) \tag{16}$$

2.2.2 Exponentially decreasing profile

Let's assume that the temperature only varies with latitude, and that its meridional gradient is constant:

$$T \approx T_0 - \frac{\Delta T}{\Delta y} \cdot y \tag{17}$$

where $(T_0, \Delta T, \Delta y)$ are known parameters. Integrating Clausius-Clapeyron's 15 relation keeping e^*'s prefactor constant, we obtain:

$$e^* = e^*(T_0) \exp[\frac{L_v(T_0)}{r_v T_0^2}(T - T_0)] = e^*(T_0) \exp[-\frac{L_v(T_0)}{r_v T_0^2} \frac{\Delta T}{\Delta y} \cdot y] \tag{18}$$

We then neglect the reduced saturation water vapor pressure $(1-\epsilon)e^*$, compared to the total pressure p that we assume to be approximately constant $p \approx p_0$, so that the saturation specific humidity 13 can be written:

$$q^* = \frac{\epsilon e^*}{p - (1-\epsilon)e^*} \approx \frac{\epsilon}{p_0} e^* \approx \frac{\epsilon e^*(T_0)}{p_0} \exp[-\frac{L_v(T_0)}{r_v T_0^2} \frac{\Delta T}{\Delta y} \cdot y] \tag{19}$$

The previous relation 19 gives us a saturation specific humidity profile that exponentially decreases with latitude:

$$\boxed{q^*(y) \approx q_{max} \exp(-\alpha y) \;\Leftrightarrow\; y^*(q) \approx -\frac{1}{\alpha} \ln(\frac{q}{q_{max}})} \tag{20}$$

where we have introduced the following physical parameters:

$$q_{max} \overset{\text{def}}{=} \frac{\epsilon e^*(T_0)}{p_0} \approx 2.5\% \tag{21}$$

$$\alpha \overset{\text{def}}{=} \frac{L_v(T_0)}{r_v T_0^2} \frac{\Delta T}{\Delta y} \approx 4.10^{-7} m^{-1} \approx \frac{1}{2500 \text{km}} \tag{22}$$

where the following orders of magnitude have been used:

- For the planet Earth, the Equator's temperature annually averages at $T_0 \approx 30°C$ and the North Pole's temperature has an average around $T_0 - \frac{\Delta T}{\Delta y}(a\frac{\pi}{2}) \approx -20°C$, leading to a global temperature gradient of $\frac{\Delta T}{\Delta y} \approx 5.10^{-3}$°C.km^{-1}.
- The latent heat of vaporization of water: $L_v \approx 2.10^6$J.kg^{-1}.
- We can estimate the saturation water vapor pressure at T_0 using Bolton's formula: $e^*(T_0) \approx 42\text{hPa} \approx (4\%)p_0$.

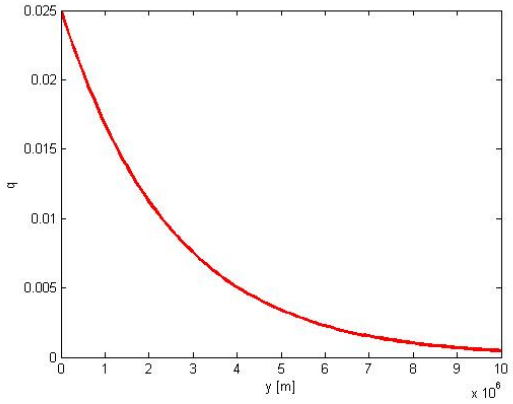

Figure 1: Approximate profile of q^* vs y for the Earth's atmosphere

2.3 Global distributions in the advection-condensation model

Under certain conditions, the global distributions of displacement and specific humidity have universal forms. Let's consider the following set of assumptions for an advection-condensation model defined by equation 8:

- We work on a latitudinal domain, extending from $y = 0$ to $y = L$.
- We assume that the saturation specific humidity is strictly decreasing from from $y = 0$ where $q^*(0) = q_{max}$ to $y = L$ where $q^*(L) = q_{min}$. We represent the domain in (q, y) space on figure 2 to make its boundaries more concrete to the reader.
- We assume that the system has reached a steady state so that its Probability Density function (PDF) $\rho(q, y, v)$ satisfies:
$$\frac{\partial \rho}{\partial t} = 0 \tag{23}$$
- We assume that the velocity v of the system is governed by an isotropic and homogeneous stochastic process.

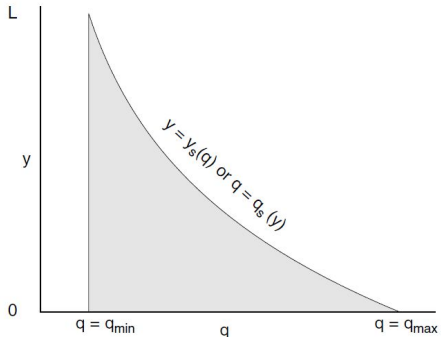

Figure 2: Domain in (q, y) space

Under the previous assumptions, we expect the global distribution of displacement:

$$P_Y(y) \overset{\text{def}}{=} \int_{\mathbb{R}} dv \int_{q_{min}}^{q^*(y)} dq \rho(q, y, v) \tag{24}$$

to have no gradients in y, ie:

$$\frac{dP_Y}{dy} = 0 \tag{25}$$

It is then possible to compute the constant P_Y from the normalization condition on the steady PDF ρ:

$$1 = \int_0^L dy \int_{\mathbb{R}} dv \int_{q_{min}}^{q^*(y)} dq \rho(q, y, v) = \int_0^L dy P_Y = L P_Y \tag{26}$$

From the knowledge of P_Y, we can rewrite the normalization condition on ρ in a way that does not involve any integral in y:

$$\boxed{\int_{\mathbb{R}} dv \int_{q_{min}}^{q^*(y)} dq \rho(q, y, v) = \frac{1}{L}} \tag{27}$$

This very powerful constraint will allow us to compute exact analytical solutions for the PDF in special cases. Furthermore, it determines the form of the global distribution of specific humidity, as we will show now. To simplify the proof, the marginalization over the variable v is always implied. From the fact that P_Y is constant (equation 26):

$$\forall y,\ P_Y(y) = \frac{P_Y(0) + P_Y(L)}{2} \tag{28}$$

Integrating 28 from $y' = 0$ to $y' = y$ yields:

$$\int_{q_{min}}^{q^*(y)} dq \int_0^y dy' \rho(q, y') dq = \int_0^y dy' P_Y(y') = \frac{y}{2}[P_Y(0) + P_Y(L)] \tag{29}$$

where we have used Fubini's theorem to exchange integrals. As $\rho_\infty[q > q^*(y)] = 0$, we can prove from 29 that:

$$\int_{q^*(y)}^{q_{max}} dq \int_0^y dy' \rho(q, y') dq = \frac{y}{2}[P_Y(0) + P_Y(L)] \tag{30}$$

We change the independent variable from y to q in equation 30, and carefully adapt the bounds of the integrals using figure 2:

$$\int_q^{q_{max}} \int_0^{y^*(q')} \rho(q', y')dq'dy' = \frac{y^*(q)}{2} \int_q^{q_{max}} [\rho(q', 0) + \rho(q', L)]dq' \tag{31}$$

Differentiating equation 31 with respect to q gives:

$$P_Q(q) = \frac{1}{2}\frac{d}{dq}\{y^*(q) \int_q^{q_{max}} [\rho(q', 0) + \rho(q', L)]dq'\} \tag{32}$$

where we have defined the global distribution of specific humidity:

$$P_Q(q) \stackrel{\text{def}}{=} \int_{\mathbb{R}} dv \int_0^{y^*(q)} \rho_\infty(q, y', v)dy' \tag{33}$$

Because of the condensation rule 12, there can only be parcels of specific humidity $q = q_{min}$ at $y = L$. Adding the normalization condition 27, it means that:

$$\rho(q', L) = \frac{\delta^+(q - q_{min})}{L} \tag{34}$$

Using 34 in equation 32 gives:

$$\frac{d}{dq}\{y^*(q) \int_q^{q_{max}} \rho_\infty(q', L)dq'\} = \frac{d}{dq}\{\frac{y^*(q)}{L} \int_q^{q_{max}} \delta^+(q' - q_{min})dq'\} = \frac{1}{L}\frac{dy^*}{dq}\delta_{qq_{min}} + \delta^+(q - q_{min})$$

where $\delta_{qq_{min}}$ is the Kronecker symbol for $q = q_{min}$. The first term is only nonzero when $q = q_{min}$, and is finite, which means it will be negligible compared to the second term in practice. As a consequence, we obtain the following general expression for the global distribution of specific humidity:

$$\boxed{P_Q(q) = \frac{\delta^+(q - q_{min})}{2} + \frac{1}{2}\frac{d}{dq}[y^*(q) \int_q^{q_{max}} \rho(q', 0)dq']} \tag{35}$$

Not only does this distribution show the presence of a dry peak in the PDF, but it also states something very fundamental about the parcel's history. Indeed, since the velocity process is isotropic and homogeneous:

- Half of the parcels have last hit the dry Northern boundary, and have been dried to $q = q_{min}$. They constitute the dry peak of $P_Q(q)$ which has a $\frac{1}{2}$ amplitude.
- Half of the parcels have last hit the moist Southern boundary, and have been re-moistened by the Southern boundary condition coming through $\rho(q', 0)dq'$ as well as condensed because of the saturation specific humidity profile, explaining the presence of $y^*(q)$.

We will see in section 4 where the advection-condensation model is solved for many different cases, that the form 35 of the global distribution of specific humidity universally applies as long as the assumptions stated at the beginning of this paragraph 2.3 apply.

3 Brownian Linearly Damped Motion on an Isentropic Surface

3.1 Preliminary assumptions

Before dealing with the complexity of condensation, we begin with a simple linear damping model for moisture, similarly to what has been tackled in the introduction (cf equation 4). We make the following simplifying assumptions:

1. We reduce our problem to a 1D Cartesian process on a line, parametrized by its ordinate y.

2. The domain we work on extends from the Equator ($y = 0$) to a given latitude $L > 0$.

3. We model the source term $\mathcal{S}$ following [23]. The model "resets" the specific humidity q to a random value chosen from a specified distribution $\Phi(q)$ when it encounters the Southern boundary of the domain ($y = 0$). Physically, we could think of the fictive domain $y \leq 0$ as the Tropics, that remoisten the dry parcels that are advected southward by the mid-latitude eddies. The normalization condition on Φ yields:

$$\int_{q_{min}}^{q_{max}} \Phi = 1 \tag{36}$$

4. The sink term is a linear damping with a typical decay frequency λ:

$$\mathcal{C}(q) = \lambda q \tag{37}$$

3.2 Stochastic differential equation and Fokker-Planck equation for the PDF

For simplicity sake, we assume that the moist parcels have a Brownian motion of diffusivity κ, leading to the following stochastic differential equation (SDE) for the position $Y(t)$ and the moisture $Q(t)$ of the moist parcels:

$$\boxed{\begin{cases} dY(t) = \sqrt{2\kappa} dW(t) \\ dQ(t) = \{\mathcal{S}[Y(t)] - \mathcal{C}[Q(t)]\} dt \end{cases}} \tag{38}$$

where $W(t)$ is a Wiener process. The corresponding steady Fokker-Planck equation (FPE) for the PDF ρ can be written by identifying the drift and the diffusion terms in the previous SDE 38:

$$\frac{\partial}{\partial q}[(\mathcal{S} - \mathcal{C})\rho] = \kappa \frac{\partial^2 \rho}{\partial y^2} \tag{39}$$

Note that integrating this equation from $q' = 0$ to $q' = q_{max}$ and using the normalization 27 gives:

$$-\lambda q_{max} \rho(q_{max}, y) = \kappa \int_0^{q_{max}} \frac{\partial^2 \rho}{\partial y^2} dq' = \kappa \frac{\partial^2}{\partial y^2} \int_0^{q_{max}} \rho(q', y) dq' = \kappa \frac{\partial^2}{\partial y^2} L^{-1} = 0$$

$$\boxed{\forall y,\ \rho(q_{max}, y) = 0} \tag{40}$$

In order to solve the FPE, we change variables:

$$\begin{cases} y \mapsto \sqrt{\frac{\lambda}{\kappa}} y \\ q \mapsto \ln(\frac{q_{max}}{q}) \end{cases} \tag{41}$$

We are then left with a diffusion-amplified equation in our new (q, y) space:

$$\boxed{(\frac{\partial}{\partial q} - 1 - \frac{\partial^2}{\partial y^2})\rho = 0} \tag{42}$$

Note that the normalization condition 27 in the new(q, y) space yields:

$$1 = L q_{max} \int_0^{+\infty} \exp(-q) \cdot \rho(q, y) dq \tag{43}$$

3.3 Solution to the steady FPE 42

First, we need to define 3 boundary conditions:

1. The resetting at the Southern boundary along with the normalization condition impose:
$$\rho(y = 0) = \frac{\Phi(q_{\text{dim}})}{L} = \frac{\Phi[q_{max} \exp(-q)]}{L} = \tilde{\Phi}(q) \tag{44}$$

2. We impose a no-flux boundary conditions at the Northern boundary:
$$\frac{\partial \rho}{\partial y}(y = L) = 0 \tag{45}$$
where we have introduced the dimensionless extent of the domain:
$$L \overset{\text{def}}{=} L_{\text{dim}} \sqrt{\frac{\lambda}{\kappa}} \tag{46}$$

3. With the new variables 41, equation 40 can be written:
$$\rho(q = 0) = 0 \tag{47}$$

The mixed boundary value problem on a finite domain has an exact analytic solution, given by:

$$\rho(q, y) = \frac{\pi}{L^2} \int_0^q dq' \tilde{\Phi}(q') \exp(q - q') \sum_{n \in \mathbb{N}} (2n + 1) \sin[\frac{\pi(2n + 1)y}{2L}] \exp[-\frac{\pi^2(2n + 1)^2(q - q')}{4L^2}] \tag{48}$$

We now focus on given remoistening profiles $\Phi(q)$ in two specific cases:

1. In case I, we consider a complete remoistening at the Equator:

$$\Phi_I(q_{\dim}) \stackrel{\text{def}}{=} \delta^-(q_{\dim} - q_{max}) \tag{49}$$

corresponding to $\tilde{\Phi}_I(q) = \frac{\delta^-(q)}{L_{\dim}}$. Coming back to the dimensional variables (ρ, q, y), the PDF in case I can be written:

$$\boxed{\rho_I(q,y) = \frac{\pi\kappa}{\lambda L^3} \sum_{n\in\mathbb{N}} (2n+1) \sin[\frac{\pi(2n+1)y}{2L}](\frac{q}{q_{max}})^{\frac{\pi^2\kappa(2n+1)^2}{4\lambda L^2}-1}} \tag{50}$$

2. In case II, we consider a uniform remoistening at the Equator:

$$\Phi_{II}(q_{\dim}) = \frac{1}{q_{max} - q_{min}} \tag{51}$$

corresponding to $\tilde{\Phi}_{II}(q) = \frac{1}{L_{\dim}(q_{max}-q_{min})}$. Coming back to the original variables (q, y), the PDF in case II can be written:

$$\rho_{II}(q,y) = \frac{\pi\kappa}{\lambda L^3(q_{max}-q_{min})} \sum_{n\in\mathbb{N}} \frac{(2n+1)4\frac{\lambda}{\kappa}L^2}{\pi^2(2n+1)^2 - 4\frac{\lambda}{\kappa}L^2} \sin[\frac{\pi(2n+1)y}{2L}][1-(\frac{q}{q_{max}})^{1-\frac{\pi^2\kappa(2n+1)^2}{4\lambda L^2}}] \tag{52}$$

3.4 Global steady PDF

From now on, we come back to dimensional variables. By definition, the global steady PDF is:

$$P_Q(q) = \int_0^L dy \rho(q,y)$$

1. In case I:

$$\boxed{P_Q(q) = \frac{2\kappa}{\lambda L^2} \sum_{n\in\mathbb{N}} (\frac{q}{q_{max}})^{\frac{\pi^2\kappa(2n+1)^2}{4\lambda L^2}-1}} \tag{53}$$

Note that P_Q can be expressed using an elliptic function:

$$P_Q(q) = \frac{2\kappa}{\lambda L^2} \frac{q_{max}}{q} \cdot \theta_2[0, (\frac{q}{q_{max}})^{\frac{\kappa}{\lambda}(\frac{\pi}{L})^2}]$$

where θ_2 is the Jacobi theta function of second type, defined by:

$$\theta_2[z,q] = \sum_{n\in\mathbb{Z}} q^{(n+\frac{1}{2})^2} \cdot \exp[(2n+1)iz]$$

2. In case II:

$$P_Q(q) = \frac{2\kappa}{\lambda L^2(q_{max}-q_{min})} \sum_{n\in\mathbb{N}} \frac{(2n+1)4\frac{\lambda}{\kappa}L^2}{\pi^2(2n+1)^2 - 4\frac{\lambda}{\kappa}L^2} [1 - (\frac{q}{q_{max}})^{1-\frac{\pi^2\kappa(2n+1)^2}{4\lambda L^2}}] \tag{54}$$

3.5 Averages and moments of the distribution

By definition, the average of a function $f(q)$ is:

$$\overline{f}(y) \stackrel{\text{def}}{=} L \int_0^{q_{max}} f(q)\rho(q,y)dq \tag{55}$$

Using the expression computed for ρ, it is possible (but tedious) to estimate $< f >$ for any well-behaved function. For our problem, the moments of the distribution $< q^n >$ are particularly interesting as they correspond to the average moisture ($n = 1$), the average square moisture ($n = 2$)... To compute the average moisture, we multiply the steady FPE 39 by q and integrate the result from $q = 0$ to $q = q_{max}$:

$$\int_0^{q_{max}} qdq \frac{\partial}{\partial q}(\lambda q \rho_\infty) + \kappa \int_0^{q_{max}} q \frac{\partial^2 \rho_\infty}{\partial y^2} = 0 \tag{56}$$

Integrating 56 by parts, noting that the boundary term is zero, and using the definition of $< q >$, we find a very simple equation for $< q >$:

$$\lambda \overline{q} = \kappa \frac{\partial^2 \overline{q}}{\partial y^2} \tag{57}$$

More generally, for $< q_n >$, multiplying the FPE 39 by q^n and integrating by parts gives:

$$\boxed{n\lambda \overline{q^n} = \kappa \frac{\partial^2 \overline{q^n}}{\partial y^2} \;\Leftrightarrow\; n\overline{q^n} = \frac{\partial^2 \overline{q^n}}{\partial y^2}} \tag{58}$$

We are thus left with a simple second order differential equation with constant coefficients to solve, where the two boundary conditions are given by:

1. The resetting condition at the Southern boundary:

$$\overline{q^n}(y = 0) = \int_0^{q_{max}} q^n \Phi(q) dq = \Phi_n \leq q_{max}^n$$

2. The no-flux condition at the Northern boundary:

$$\frac{\partial}{\partial y} \overline{q^n}(y = L) = 0$$

leading to the following solution:

$$\boxed{\overline{q^n}(y) = \Phi_n[\cosh(\sqrt{n}y) - \tanh(\sqrt{n}L)\sinh(\sqrt{n}y)]} \tag{59}$$

Note that in the specific case of a complete remoistening at the Equator 49, we obtain $\Phi_n = q_{max}^n$.

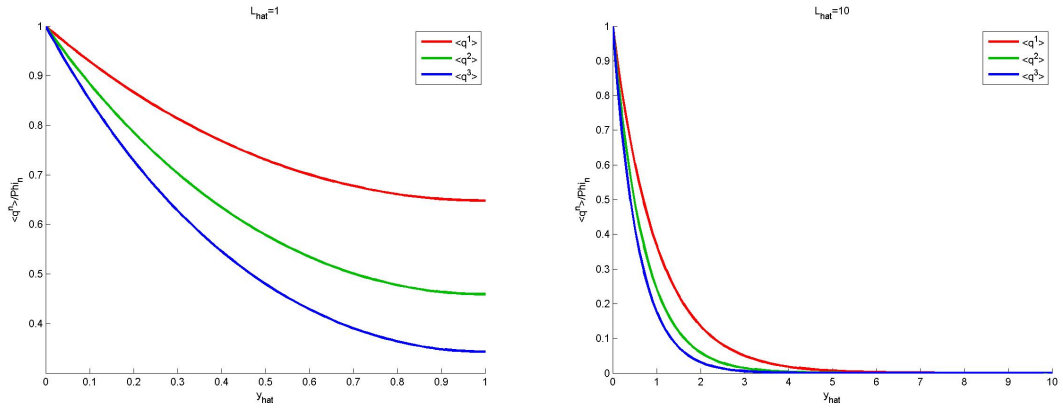

Figure 3: $\frac{<q^n>}{\Phi_n}$ for $n = 1$ (red) $n = 2$ (green) and $n = 3$ (blue) vs y for $L = 1$ (left) and $L = 10$ (right)

We can see that:

- The moments of q are damped faster throughout the domain as the dimensionless extent of the domain 46 (which varies proportionally to the damping frequency λ) increases.
- The higher moments of q are damped faster than the lower moments of q, which was to be expected with a linear damping, which effect is amplified when multiplied by q^n.

3.6 Diffusion in the linearly damped case

Let $y_0 \in]0, L[$: if we integrate the dimensional equation for $\overline{q}$ 57 from $y = y_0$ to $y = L$, we obtain:

$$\boxed{\lambda \int_{y_1}^{L} \overline{q} dy = -\kappa (\frac{\partial \overline{q}}{\partial y})_{y_0} = \mathcal{F}(y_0)} \tag{60}$$

This indicates that in steady state, the linear damping of moisture in the Northern part of the domain $y > y_0$is compensated by the Fickian diffusive flux of moisture $\mathcal{F}(y_0)$ from the Southern part of the domain $y < y_0$ to the Northern part. As shown in the more general case, we obtain that the diffusivity of the linearly damped moisture q is equal to κ in the case of a white noise velocity field.

3.7 Numerical experiments in the complete remoistening case $\Phi_I(q) = \delta^+(q - q_{min})$

To check the validity of our general solution in this specific case, we numerically simulate the SDE with the following dimensional values:

$$\begin{cases} N_{tot} = 3.10^4 \\ L = 1 \\ q_{max} = 1 \\ \kappa = 1 \end{cases}$$

and compare the numerical PDF to the marginal distributions $P_Y(y)$and $P_Q(q)$ for different values of the damping frequency λ:

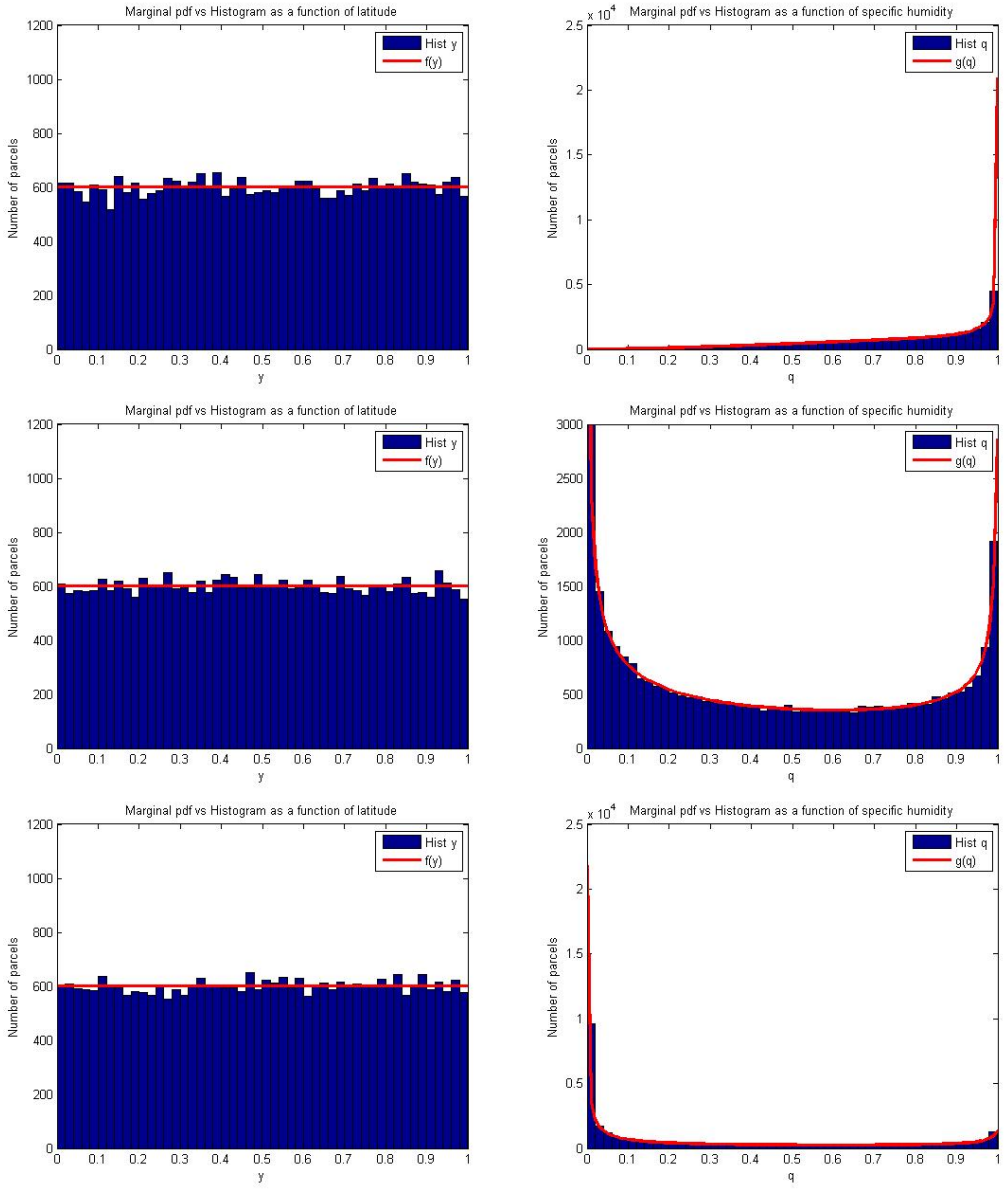

Figure 4: $P_Y(y)$ (left) and $P_Q(q)$ (right) for a square dimensionless length $L^2 \in \{1, 5, 10\}$ (top, middle, bottom)

To interpret the previous results, it is useful to remember that our system is governed by the dimensionless extent of the domain defined in 46:

- For small damping $L^2 \leq 1$:

$$P_Q(q) \propto \sum_{n\in\mathbb{N}^*} q^{(2n+1)^2 L^{-2}} \underset{(L\to 0)}{\to} \delta^-(q - q_{max})$$

  and the moist peak of the distribution near $q = q_{max}$ dominates. Physically, the moisture of the parcels do not have time to damp before the parcels encounters the Southern boundary and is completely remoistened again.

- For large damping $L^2 \geq 10$:

$$P_Q(q) \propto q^{-1} \underset{(L\to+\infty)}{\to} \delta^+(q)$$

  and the dry peak of the distribution near $q = 0$ dominates. Physically, the moisture of the parcels damps out very fast after the parcels have been remoistened, and it takes them a long time before they re-encounter the moistening Southern boundary of the domain.

- For intermediate $L^2 \in [1, 10]$, the distribution of the parcels is bimodal, with a moist and a dry peak.

4 Time-correlated Velocities

White noise does not exist in nature, and atmospheric data indeed show that velocities are correlated in time, the correlation time usually being the typical eddy turnover time.

4.1 The Ornstein-Uhlenbeck model

4.1.1 Gaussian colored noise

We replace the white noise in the previous part by Gaussian colored noise for the Lagrangian velocities. According to Doob's theorem, this noise can be obtained through an Ornstein-Uhlenbeck process for the velocities. If we consider 1D meridional motion, our new system of SDEs can be written:

$$\boxed{\begin{cases} dY(t) = V(t)dt \\ \tau dV(t) = -V(t)dt + \sqrt{2\kappa}dW(t) \\ dQ(t) = \{\mathcal{S}[Y(t)] - \mathcal{C}[Y(t), Q(t)]\}dt \end{cases}} \tag{61}$$

where $V(t) = \frac{dY}{dt}$ is the meridional Lagrangian velocity of the process, and κ its associated diffusivity. The Ornstein-Uhlenbeck process is a standard model of turbulent dispersion [26, 14, 18], and the novelty of the stochastic advection-condensation model is to introduce the condensing variable Q. The corresponding FPE for the PDF $\rho(t, q, y, v)$ can be written:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial q}[(\mathcal{S} - \mathcal{C})\rho] + \frac{\partial}{\partial y}(\rho v) = \frac{1}{\tau}\frac{\partial}{\partial v}(\rho v) + \frac{\kappa}{\tau^2}\frac{\partial^2 \rho}{\partial v^2} \tag{62}$$

It can be proven that the auto-correlation function for the Ornstein-Uhlenbeck process can be written:

$$\mathbb{E}[V(t_1)V(t_2)] = \frac{\kappa}{\tau}\exp(-\frac{|t_2 - t_1|}{\tau}) \neq 0 \tag{63}$$

which is what makes this noise "colored" in contrast to white.

4.1.2 Marginalized distribution

We make the Fokker-Planck equation non-dimensional by rescaling:

$$y \mapsto \frac{y}{\sqrt{\tau\kappa}} \tag{64}$$

$$v \mapsto \sqrt{\frac{\tau}{\kappa}}v \tag{65}$$

If we marginalize the distribution over q:

$$p(y, v) \stackrel{\text{def}}{=} \int_{q_{\min}}^{q^*(y)} \rho(t, q, y, v)dq \tag{66}$$

we recover the well-known FPE for the OU process from the FPE 62:

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}(vp) = \frac{\partial}{\partial v}(vp + \frac{\partial p}{\partial v}) \tag{67}$$

The solution of equation 67 along with the full normalization condition for the PDF ρ 26 is the well-known Maxwellian distribution:

$$p(y, v) = \frac{1}{\sqrt{2\pi}L}\exp(-\frac{v^2}{2}) \tag{68}$$

We have introduced the sole dimensionless parameter of our model, ie the dimensionless domain's extension:

$$L \stackrel{\text{def}}{=} \frac{L_{\text{dim}}}{\sqrt{\kappa\tau}} \tag{69}$$

In the ballistic limit $L \ll 1$, the domain is so small that the velocity of a given parcel is constant in time. In the Brownian limit $L \gg 1$, the domain is so large that the displacement of a given parcel is a white noise process. Both cases are studied in A.

4.1.3 Steady PDF in the general case

We transform the steady version of the FPE 67 by changing variable for the density ρ, using the solution 68:

$$\rho \mapsto \frac{p(y, v)}{\kappa}\rho \tag{70}$$

giving:

$$v\frac{\partial \rho}{\partial y} = \mathcal{D}\rho \stackrel{\text{def}}{=} \frac{\partial^2 \rho}{\partial v^2} - v\frac{\partial \rho}{\partial v} \tag{71}$$

Note that the normalization condition for ρ is now simply:

$$\forall(y,v) \int_{q_{min}}^{q^*(y)} \rho(q,y,v)dq = 1 \tag{72}$$

In order to solve 72, we first look for exponential solutions in y (or more rigorously we Laplace-transform the previous differential equation in y):

$$\rho(q,y,v) \propto \exp(\alpha y)\varphi_\alpha(v)$$

where the functions φ_α are the eigenfunctions of the following eigenvalue problem:

$$\boxed{\mathcal{D}\varphi_\alpha = \alpha v \varphi_\alpha} \tag{73}$$

It is shown in the appendix B that the functions φ_α for $\alpha^2 \in \mathbb{N}$, along with the generalized eigenfunction $v \mapsto v$, constitute an orthogonal basis in which we can expand any well-behaved function which verifies the initial differential equation. Solving a half-range problem is a hard mathematical problem, and adding the specific boundaries of our problem in (q,y) space makes the analytical solution too complicated to be tractable in a useful way. We will thus describe the four steps to solve this problem in the general case without doing the algebra, based on the fact that the PDF comprises a dry spike, a saturated spike, and a smooth part (cf appendix C):

1. First, we write ρ as a combination of the solutions to the differential equation:

   $$\rho(q,y,v) = B(q)[y-v] + \sum_{\alpha^2 \in \mathbb{N}} A_\alpha(q)\varphi_\alpha(v)\exp(\alpha y)$$

   where $\varphi_0(v) = 1$. The remoistening condition at the Equator can be written:

   $$\rho(q, y=0, v>0) = \Phi(q)$$

   giving a first relation between all the coefficients:

   $$\forall v > 0,\ -B(q)v + \sum_{\alpha^2 \in \mathbb{N}} A_\alpha(q)\varphi_\alpha(v) = \Phi(q) \tag{74}$$

   From this relation 74, it is possible to express the function B and half of the functions A_α using the function Φ and the other half of the functions A_α (we will call them C_α to avoid confusion). If the projection is done well, $\Phi(q)$ should only appear in $A_0(q)$ as it appears as a constant in y in the final solution.

2. Then, we decompose each function (C_α) in a dry and a smooth part:

   $$C_\alpha(q) = C_{\alpha,\text{dry}}\delta^+(q - q_{min}) + C_{\alpha,\text{smooth}}(q)$$

   which implies that all the functions of q : (A_α, B) have a dry and a smooth part. To find the coefficients $C_{\alpha,\text{dry}}$, we apply the normalization condition 72 for the southwards moving parcels at $y = L$ where $q = q^*(y) = q_{min}$. We can only apply it to the

southwards moving parcels because the northwards moving parcels have a saturated peak, as we prove in appendix C:

$$\forall v<0,\ 1=\int_{q_{min}}^{q_{min}}\rho(q,y=L,v<0)dq=B_{\text{dry}}[L-v]+\sum_{\alpha^2\in\mathbb{N}}C_{\alpha,\text{dry}}\varphi_\alpha(v)\exp(\alpha L) \tag{75}$$

3. We then need to compute the smooth part of the functions: $C_{\alpha,\text{smooth}}(q)$. It is also a half-range problem in v, as we can only apply the normalization condition 72 for all y to the southwards parcels which do not have a saturated peak:

$$\forall y,\forall v<0,\ 1=\int_{q_{min}}^{q^*(y)}\rho(q,y,v)dq$$

$$\forall y,\forall v<0,\ 1=[B_{\text{dry}}+\int_{q_{min}}^{q^*(y)}B_{\text{smooth}}(q)dq][L-y]+\sum_{\alpha^2\in\mathbb{N}}[C_{\alpha,\text{dry}}+\int_{q_{min}}^{q^*(y)}C_{\alpha,\text{smooth}}(q)dq]\varphi_\alpha(v)\exp(\alpha y)$$

Changing the independent variable from y to q and adapt the bounds of the integrals using 2:

$$\forall q,\forall v<0,\ 1=[B_{\text{dry}}+\int_{q_{min}}^{q}B_{\text{smooth}}(q')dq'][L-y^*]+\sum_{\alpha^2\in\mathbb{N}}[C_{\alpha,\text{dry}}+\int_{q_{min}}^{q}C_{\alpha,\text{smooth}}(q')dq']\varphi_\alpha(v)\exp(\alpha y^*) \tag{76}$$

and we can find the functions $C_{\alpha,\text{smooth}}$ by solving the previous equation 76 for each $\int_{q_{min}}^{q}C_{\alpha,\text{smooth}}(q')dq'$ and then differentiating with respect to q.

4. Finally, once all the coefficients have been determined, we need to add the saturated peak to the northwards PDF (cf C):

$$\rho(q,y,v)=B(q)[y-v]+\sum_{\alpha^2\in\mathbb{N}}C_\alpha(q)\varphi_\alpha(v)\exp(\alpha y)+\mathcal{W}(y,v>0)\delta^-(q-q^*)$$

where $\mathcal{W}(y,v)$ is an undetermined weight function such that $\mathcal{W}(y,v<0)=0$. We can find its positive part by applying the normalization condition 72 to the northwards PDF:

$$\forall y,\forall v>0,\ \mathcal{W}(y,v>0)=1-\int_{q_{min}}^{q^*(y)}\{B(q)[y-v]+\sum_{\alpha^2\in\mathbb{N}}C_\alpha(q)\varphi_\alpha(v)\exp(\alpha y)\}dq \tag{77}$$

4.1.4 Global distributions

For the Ornstein-Uhlenbeck process, the assumptions at the beginning of paragraph 2.3 are verified, giving us the global PDFs in y defined in 24and the global distribution in q defined in 33:

$$\boxed{P_Y(y)=\frac{1}{L}} \tag{78}$$

$$\boxed{2P_Q(q)=\delta^+(q-q_{min})-\frac{1}{L}\frac{d}{dq}[\Lambda y^*]} \tag{79}$$

The global PDF in v can be found by integrating 68 in y:

$$\boxed{P_V(v) \stackrel{\text{def}}{=} \int_0^L dy p(y) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{v^2}{2})} \tag{80}$$

To check the validity of the previous analytical expressions, we simulate the global PDFs in the classical case of the complete remoistening at the Southern boundary $\Phi_I(q) = \delta^-(q - q_{max})$ and an exponentially decreasing profile of saturation specific humidity $q^*(y) = q_{max} \exp(-\alpha y)$. We choose the following parameters:

$$\begin{cases} N_{tot} = 10^6 \\ q_{max} = 1 \\ q_{min} = q_{max} \exp(-\alpha L) = 0.1 \\ L = \frac{(L_{\text{dim}}=1)}{\sqrt{(\kappa_V=1)(\tau=1)}} = 1 \end{cases}$$

and compare the analytical expressions found for the global distributions $P_Y(y)$, $P_Q(q)$ and $P_V(v)$ to the numerical simulation of the Ornstein-Uhlenbeck SDEs:

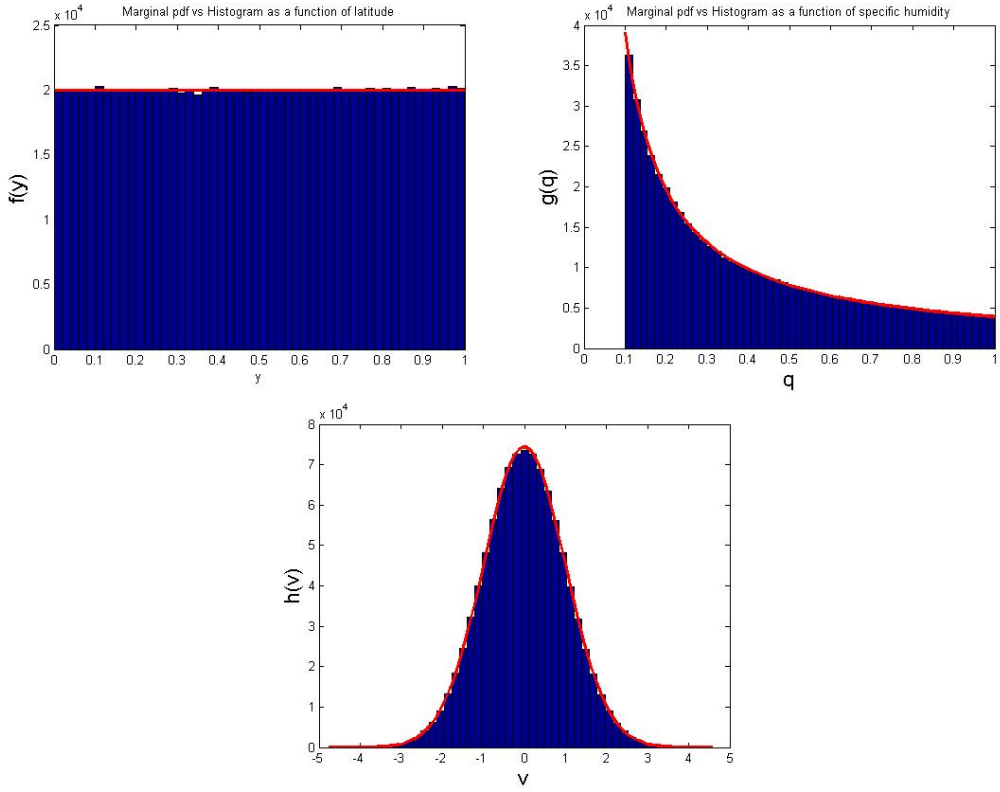

Figure 5: Numerical (blue) vs analytical (red) global PDFs for the Ornstein-Uhlenbeck model in case I of a complete remoistening

4.2 The two-stream model

4.2.1 Preliminary assumptions

A discrete version of the Ornstein-Uhlenbeck model (4.1) is the n-stream model (appendix D), where it is easier to compute exact analytical solutions for the PDF for small n. The simplest $n-$stream model is the $2-$stream model, defined by the following assumptions:

1. We work on a bounded domain $y \in [0, L]$, where the saturation specific humidity $q^*(y)$ is a strictly decreasing function of latitude.
2. The parcels are remoistened at the equator by a distribution $\Phi(q)$ normalized to 1.
3. We have two ensembles of moist parcels with an exchange frequency $\frac{\beta}{2}$:

- The parcels moving Northward, which have a velocity $+V$, and are described by the PDF $N(q, y)$.
- The parcels moving Southward, which have a velocity $-V$, and are described by the PDF $S(q, y)$.

The equations for (N, S) are thus:

$$\begin{cases} \frac{\partial N}{\partial t} + V\frac{\partial N}{\partial y} = \frac{\beta}{2}(S - N) \\ \frac{\partial S}{\partial t} - V\frac{\partial S}{\partial y} = \frac{\beta}{2}(N - S) \end{cases} \tag{81}$$

4.2.2 Steady southwards PDF

We now assume the northwards and southwards PDF (N, S) to be steady. Their normalization conditions can be deduced from 26:

$$\forall y, \int_{q_{min}}^{q^*(y)} N(q, y)dq = \int_{q_{min}}^{q^*(y)} S(q, y)dq = \frac{1}{2L} \tag{82}$$

We change variables to make our equations dimensionless:

$$y \mapsto \frac{\beta y}{V} \tag{83}$$

$$(N, S) \mapsto (\frac{V}{\beta})(N, S) \tag{84}$$

to obtain the steady dimensionless version of 81:

$$\begin{cases} \frac{\partial N}{\partial y} = \frac{\partial S}{\partial y} \\ \frac{\partial}{\partial y}(N + S) = S - N \end{cases} \tag{85}$$

Note that the system is governed by a single dimensionless number, the dimensionless extent of the domain:

$$L \overset{\text{def}}{=} \frac{\beta L_{\text{dim}}}{V} \tag{86}$$

The boundary condition along with the normalization condition 82 yields:

$$N(q,0) = \Phi(q) \int_{q_{min}}^{q_{max}} S(q,0)dq = \frac{\Phi(q)}{2L} \tag{87}$$

Integrating the previous system 85 using 87 yields:

$$\begin{cases} N(q,y) = yA(q) + \frac{\Phi(q)}{2L} \\ S(q,y) = (y+2)A(q) + \frac{\Phi(q)}{2L} \end{cases}$$

The PDF comprises a dry peak, a smooth peak and a saturated spike (cf appendix C). Here, we decompose A into a dry peak and a smooth part:

$$A(q) = A_{\text{smooth}}(q) + A_{\text{dry}}\delta^+(q - q_{min})$$

The normalization condition 82 for S at $y = L$ where $q = q_{min} = q^*(y)$ yields:

$$A_{\text{dry}} = \frac{1}{2L(2+L)}$$

The southwards PDF S satisfies the normalization condition for all $y \in [0, L]$, and does not have any saturated peak by definition (cf C):

$$1 = 2L \int_{q_{min}}^{q^*(y)} S(q,y)dq = (y+2)(\frac{1}{2+L} + 2L \int_{q_{min}}^{q^*(y)} A_{\text{smooth}}(q)dq) + \int_{q_{min}}^{q^*(y)} \Phi(q)dq$$

$$2L \int_{q_{min}}^{q^*(y)} A_{\text{smooth}}(q)dq = \frac{\int_{q^*(y)}^{q_{max}} \Phi(q)dq}{y+2} - \frac{1}{2+L} \tag{88}$$

Switching the independent variable from y to q , changing the bounds of the integrals following figure 2, and differentiating the previous equation 88 with respect to q, we obtain:

$$2LA_{\text{smooth}} = \frac{d}{dq} \frac{\Lambda(q)}{2 + y^*(q)}$$

which allows us to express the PDF S as a function of (Λ, y^*):

$$\boxed{2LS(q,y) = (y+2)[\frac{d}{dq} \frac{\Lambda(q)}{2 + y^*(q)} + \frac{\delta^+(q - q_{min})}{2+L}] - \frac{d\Lambda}{dq}} \tag{89}$$

In the two previous equations, we have introduced the Cumulative Distribution Function of the remoistening distribution Φ:

$$\Lambda(q) \overset{\text{def}}{=} \int_q^{q_{max}} \Phi(q')dq' \tag{90}$$

4.2.3 Steady northwards PDF

We incorporate a saturated spike $\mathcal{W}(y)\delta^-[q-q^*(y)]$ in the total northwards PDF $\mathcal{N}$:

$$\mathcal{N}(q,y) = N(q,y) + \mathcal{W}(y)\cdot\delta^-[q-q^*(y)]$$

The total normalization condition for $\mathcal{N}$ (82) can now be written:

$$1 = 2L\int_{q_{min}}^{q^*(y)} \mathcal{N}(q,y)dq = 2L\int_{q_{min}}^{q^*(y)} N(q,y)dq + 2L\mathcal{W}(y)$$

which allows us to compute the weight $\mathcal{W}(y)$ of the saturated spike in this specific case:

$$1 = 1 - \frac{2}{2+y}\Lambda[q^*(y)] + 2L\mathcal{W}(y) \;\Rightarrow\; 2L\mathcal{W}(y) = \frac{2}{2+y}\Lambda[q^*(y)]$$

The generalized northwards PDF is then:

$$\boxed{2L\mathcal{N}(q,y) = y[\frac{d}{dq}\frac{\Lambda(q)}{2+y^*(q)} + \frac{\delta^+(q-q_{min})}{2+L}] - \frac{d\Lambda}{dq} + \frac{2\Lambda[q^*(y)]}{2+y}\delta^-[q-q^*(y)]} \tag{91}$$

We can verify a posteriori that the value found for $\mathcal{W}$ in this paragraph agrees with the differential equation we derive for the saturated weight $\mathcal{W}$ in appendix 119:

$$2L(\frac{d\mathcal{W}}{dy} + \frac{\mathcal{W}}{2}) = \frac{2}{2+y}\frac{dq^*}{dy}\frac{d\Lambda}{dq}(q^*) + \frac{y\Lambda(q^*)}{(2+y)^2}$$

$$2LN(q^*,y) = -\frac{d\Lambda}{dq}(q^*) + y[\frac{\frac{d\Lambda}{dq}(q^*)}{2+y} - \frac{\frac{dy^*}{dq}\Lambda(q^*)}{(2+y)^2}]$$

$$-2L\frac{dq^*}{dy}N(q^*,y) = 2\frac{dq^*}{dy}\frac{\frac{d\Lambda}{dq}(q^*)}{2+y} + \frac{y\Lambda(q^*)}{(2+y)^2} = 2L(\frac{d\mathcal{W}}{dy} + \frac{\mathcal{W}}{2}) \tag{92}$$

Equation 92 is just the weight equation 119, multiplied by the dimensionless extent of the domain L, which proves that it indeed represents the amount of supersaturated parcels. Note that adding the saturated peak in the PDF might prevent it to satisfy the remoistening boundary condition at $y = 0$ 87 in the special case where:

$$\mathcal{W}(0) = \Lambda[q_{max}] = \int_{q_{max}}^{q_{max}} \Phi(q)dq \neq 0 \;\Leftrightarrow\; \Phi(q) = \delta^-(q-q_{max})$$

This special case is treated in appendix F of this report. However, if we are to simulate this special case numerically (which we will do in 4.2.8), we will have a peak of finite amplitude $\varepsilon \neq 0$, which will result in $\mathcal{W}_{numerical}(0) = 0$.

4.2.4 Steady global distributions of specific humidity

The global distribution of specific humidity are by definition:

- For the southwards parcels:

$$2Ls(q) \overset{\text{def}}{=} 2L \int_0^{y^*(q)} S(q,y)dy$$

Using equation 89:

$$2Ls(q) = \frac{y^*(4+y^*)}{2}\frac{\delta^+(q-q_{min})}{2+L} + \frac{y^*(4+y^*)}{2}\frac{d}{dq}\frac{\Lambda(q)}{2+y^*(q)} - y^*\frac{d\Lambda}{dq}$$

$$\boxed{4Ls(q) = \frac{L(L+4)}{2+L}\delta^+(q-q_{min}) - \frac{(y^*)^2}{2+y^*}\frac{d\Lambda}{dq} - \frac{dy^*}{dq}\frac{y^*(4+y^*)}{(2+y^*)^2}\Lambda} \tag{93}$$

- For the northwards parcels:

$$2Ln(q) \overset{\text{def}}{=} 2L \int_0^{y^*(q)} \mathcal{N}(q,y)dy$$

Using equation 91:

$$2Ln(q) = \frac{(y^*)^2}{2}\frac{\delta^+(q-q_{min})}{2+L} + \frac{(y^*)^2}{2}\frac{d}{dq}\frac{\Lambda(q)}{2+y^*(q)} - y^*\frac{d\Lambda}{dq} + |\frac{dy^*}{dq}|\frac{2\Lambda(q)}{2+y^*(q)}$$

$$\boxed{4Ln(q) = \frac{L^2}{2+L}\delta^+(q-q_{min}) - \frac{y^*(4+y^*)}{2+y^*}\frac{d\Lambda}{dq} - \frac{dy^*}{dq}\frac{8+4y^*+(y^*)^2}{(2+y^*)^2}\Lambda} \tag{94}$$

- For all the parcels, following definition 33, the global distribution of specific humidity is given by:

$$2LP_Q(q) \overset{\text{def}}{=} 2L(n+s)(q)$$

$$\boxed{2P_Q(q) = \delta^+(q-q_{min}) - \frac{1}{L}\frac{d}{dq}[\Lambda y^*]} \tag{95}$$

Result 95 is equal to the universal form of $P_Q(q)$ 35 found in paragraph 2.3, as the two-stream model verifies the assumptions listed at the beginning of this paragraph.

4.2.5 Averages

By definition the average of a function $f(q)$ is:

$$\overline{f}(y) \overset{\text{def}}{=} L \int_{q_{min}}^{q^*(y)} f(q)(N+S)dq \tag{96}$$

Using the expressions 91 and 89 we computed for (N,S):

$$L(N+S) = [y+1][\frac{d}{dq}\frac{\Lambda(q)}{2+y^*(q)} + \frac{\delta^+(q-q_{min})}{2+L}] - \frac{d\Lambda}{dq} + \frac{\Lambda[q^*(y)]}{2+y}\delta^-[q-q^*(y)]$$

We can use integration by parts on integral 96 to obtain $\overline{f}$ as a function of $(\Lambda, q^*(y), f, \frac{df}{dq})$:

$$\boxed{\overline{f}(y) = \{\Lambda f\}[q^*(y)] - (y+1)\int_{q_{min}}^{q^*(y)} \frac{\Lambda(q)}{2+y^*(q)}\frac{df}{dq}dq - \int_{q_{min}}^{q^*(y)} \frac{d\Lambda}{dq}f(q)dq} \tag{97}$$

The meridional gradient of average 97 can be computed using Leibniz's formula to differentiate integrals, giving a fairly simple expression once simplified:

$$\boxed{\frac{d\overline{f}}{dy} = \frac{dq^*}{dy}\frac{\{\Lambda\frac{df}{dq}\}[q^*]}{2+y} - \int_{q_{min}}^{q^*(y)} \{\frac{\Lambda}{2+y^*}\frac{df}{dq}\}(q)dq} \tag{98}$$

We can use expressions 97 and 98 to compute the average moisture content and its meridional gradient as functions of latitude:

$$\overline{q}(y) = q^*\Lambda(q^*) - (y+1)\int_{q_{min}}^{q^*(y)} \frac{\Lambda(q)}{2+y^*(q)}dq - \int_{q_{min}}^{q^*(y)} \frac{d\Lambda}{dq}qdq \tag{99}$$

$$\frac{d\overline{q}}{dy} = \frac{dq^*}{dy}\frac{\Lambda(q^*)}{2+y} - \int_{q_{min}}^{q^*(y)} \{\frac{\Lambda}{2+y^*}\}(q)dq \tag{100}$$

4.2.6 Diffusivity of parcels in the two-stream model

The diffusivity of parcels in this model can be most easily computed by coming back to the initial equations 4.2 and working with dimensional variables. We define the dimensional northward flux of parcels as:

$$\mathcal{F}_c \overset{\text{def}}{=} V(N-S) \tag{101}$$

From the initial equations 4.2, we can relate the meridional gradient of the concentration of parcels c to the flux $\mathcal{F}_c$ defined in 101:

$$V\frac{\partial}{\partial y}(N+S) = V\frac{\partial c}{\partial y} = \beta(S-N) = -\frac{\beta}{V}\mathcal{F}_c$$

giving Fick's law for the parcel's flux:

$$\mathcal{F}_c = -\kappa\frac{\partial c}{\partial y}$$

where we have defined the dimensional diffusivity of parcels:

$$\boxed{\kappa \overset{\text{def}}{=} \frac{V^2}{\beta}} \tag{102}$$

In this case, we can prove that the diffusivity is consistent with its more fundamental definition. First, let's compute the correlation function for the velocities of the system in two different ways to better understand the system. To avoid the trivial case with no/weak exchanges of parcels, we release all the parcels at $t = 0$ at the Southern boundary, which means they all initially have northwards velocities:

1. We can directly solve the transient regime between $t' = 0$ and $t' = t$:

$$\begin{cases} \frac{dN}{dt} = \frac{\beta}{2}(S - N) \\ \frac{dS}{dt} = \frac{\beta}{2}(N - S) \end{cases} \Rightarrow \begin{cases} 2N = 1 + \exp(-\beta t) \\ 2S = 1 - \exp(-\beta t) \end{cases}$$

where $1 = (N - S)(t = 0)$. The correlation at time t is then:

$$C(t) = V(0)V(t) = V^2(N - S)(0)(N - S)(t) = V^2 \exp(-\beta t)$$

2. We can also note that from the exchange equations, the probability of n signs reversal is given by the Poisson distribution of coefficient $\frac{\beta}{2}$:

$$\mathbb{P}(n) = \exp(-\frac{\beta t}{2})\frac{(\beta t)^n}{2^n n!}$$

The correlation at time t is the expected value of $V(0)V(t)$:

$$C(t) = \mathbb{E}[V(0)V(t)]$$

$$C(t) = V \cdot V[\mathbb{P}(n = 2k) - \mathbb{P}(n = 2k + 1)] = V^2 \exp(-\frac{\beta t}{2}) \sum_{k \in \mathbb{N}} \frac{(\beta t)^{2k}}{2^{2k} 2k!} - \frac{(\beta t)^{2k+1}}{2^{2k+1}(2k + 1)!}$$

$$C(t) = V^2 \exp(-\frac{\beta t}{2}) \cdot [\cosh(\frac{\beta t}{2}) - \sinh(\frac{\beta t}{2})] = V^2 \exp(-\beta t)$$

In both cases, the diffusivity of the system is then equal to Taylor's diffusivity defined in 2:

$$\boxed{\kappa = \int_{\mathbb{R}_+} C(t)dt = \int_{\mathbb{R}_+} V^2 \exp(-\beta t)dt = \frac{V^2}{\beta}} \tag{103}$$

which is consistent with the fact that the parcels are purely passive tracers.

4.2.7 Diffusivity of moisture in the two-stream model

If we now want to look at the diffusivity of moisture, we need to compute:

1. The dimensional northward flux of moisture:

$$\mathcal{F}_q \overset{\text{def}}{=} L \int_{q_{min}}^{q^*(y)} q(VN - VS)(q)dq$$

$$\frac{\mathcal{F}_q}{V} = \int_{q_{min}}^{q^*(y)} qL(N - S)(q)dq$$

We use the dimensionless expressions found for the PDF 91 and 89 so that integration by parts yields:

$$\frac{\mathcal{F}_q}{V} = \int_{q_{min}}^{q^*(y)} \frac{\Lambda(q)dq}{2 + \frac{\beta y^*}{V}}$$

2. The dimensionless mean meridional gradient of moisture computed in 100:

$$\frac{V}{\beta}\frac{d\overline{q}}{dy} = \frac{d\overline{q}}{d(\frac{\beta y}{V})} = \frac{V}{\beta}\frac{dq^*}{dy}\frac{\Lambda(q^*)}{2+\frac{\beta y}{V}} - \int_{q_{min}}^{q^*(y)} \frac{\Lambda(q)}{2+\frac{\beta y^*}{V}} dq$$

leading to a dimensional moisture flux which is lower than its Fick's value:

$$\boxed{\frac{\mathcal{F}_q}{\kappa} = -\frac{d\overline{q}}{dy} - |\frac{dq^*}{dy}|\frac{\Lambda(q^*)}{2+\frac{\beta y}{V}}} \tag{104}$$

$$\mathcal{F}_q < \mathcal{F}_{q,\text{Fickian}} = -\kappa\frac{d\overline{q}}{dy}$$

The difference between $\mathcal{F}_q$ and $\mathcal{F}_{q,\text{Fickian}}$ increases as the saturation specific humidity profile $\frac{dq^*}{dy}$ (and thus the mean specific humidity profile $\frac{d\overline{q}}{dy}$) becomes sharper and sharper. This difference is 0 for a flat profile $\frac{dq^*}{dy} = 0$, and the flux is perfectly Fickian in this limit.

4.2.8 Application to an exponentially decreasing $q^*(y)$ and a given remoistening $\Phi(q)$

We make the following assumptions:

1. We assume that the saturation specific humidity profile exponentially decreases with latitude:
$$q^*(y) \approx q_{max}\exp(-\alpha y) \;\Leftrightarrow\; y^*(q) \approx -\frac{1}{\alpha}\ln(\frac{q}{q_{max}})$$

2. We assume that the "Southern boundary resetting" produces complete saturation at $y = 0$ (case I of complete remoistening):
$$\Phi_I(q) = \delta^-(q - q_{max})$$
or remoistens the parcels from a uniform distribution on $[q_{min}, q_{max}]$ (case II of uniform remoistening):
$$\Phi_{II}(q) = \frac{1}{q_{max} - q_{min}}$$

To check the validity of our general solution in this specific case, we numerically simulate the SDE with:

$$\begin{cases} N_{tot} = 10^6 \\ q_{min} = 0.1 \\ q_{max} = 1 \\ L = \frac{(\beta=1)(L_{\text{dim}}=2)}{(V=1)} = 2 \end{cases}$$

and compare the numerical PDF to the marginal distributions $P_Y(y)$ and $(P_Q, n, s)(q)$:

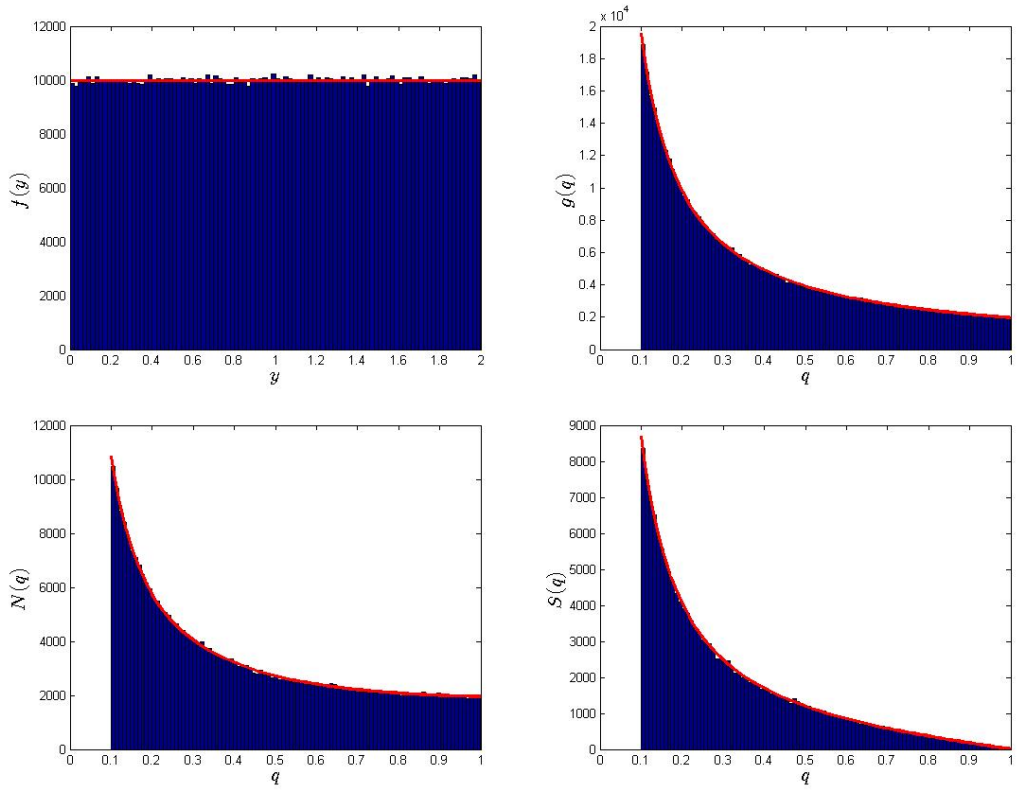

Figure 6: Numerical simulation of the SDE (blue) compared to the theoretical PDF (red) for $f(y)$ (top left), $g(q)$ (top right), $N(q)$ (bottom left) and $S(q)$ (bottom right) for a complete remoistening (case I)

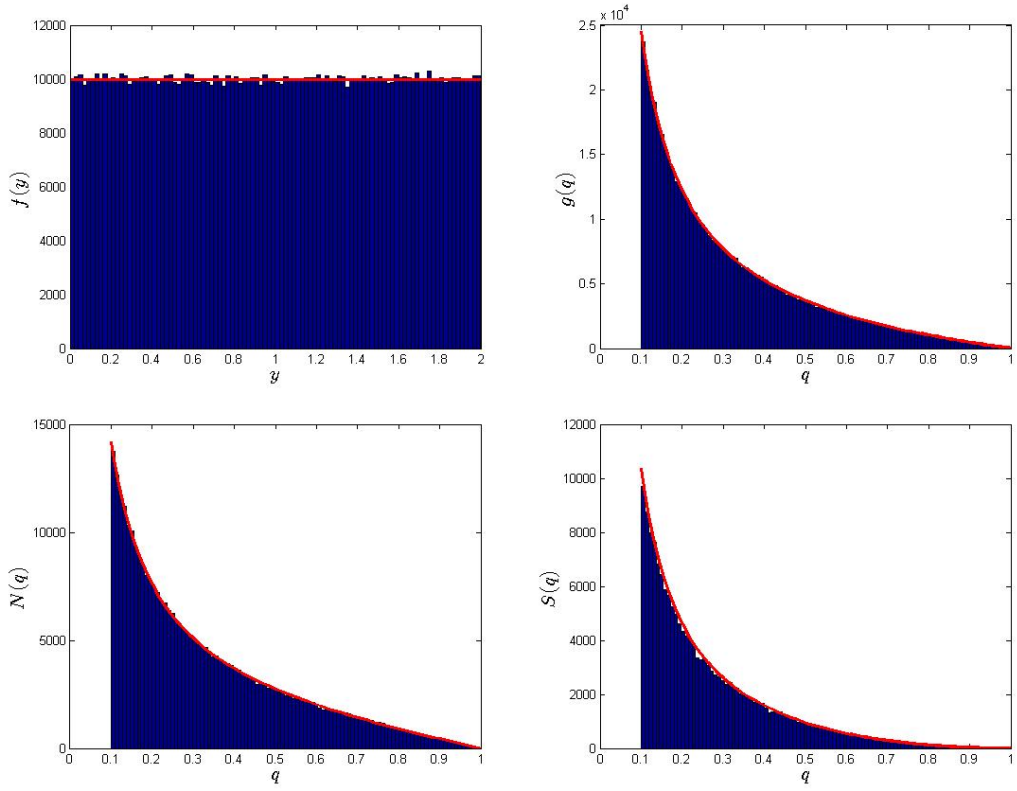

Figure 7: Numerical simulation of the SDE (blue) compared to the theoretical PDF (red) for $f(y)$ (top left), $g(q)$ (top right), $N(q)$ (bottom left) and $S(q)$ (bottom right) for a uniform remoistening (case II)

When L is increased, the SDE takes a longer time to reach a statistically steady state, as it takes longer for the parcels to reach the Southern and the Northern boundaries, making the numerical simulations of the steady PDF more costful. A good metric to see if a simulation has reached the statistically steady state, based on the steady expression of $P_Q(q)$, is based on the relative intensity of the dry peak/smooth part of $P_Q(q)$:

$$\frac{\|P_{Q,\text{smooth}}(q)\|}{\|P_Q(q)\|} = 1 - \frac{\|P_{Q,\text{dry}}(q)\|}{\|P_Q(q)\|}$$

At the beginning of the simulation, this ratio is close to 1, as not a lot of parcels have hit the Northern dry boundary. As the simulation reaches the steady state, we have proven in 95 that this ratio tends towards $\frac{1}{2}$. Using this ratio, we can see that if we want to obtain the steady PDF in the following case:

$$\begin{cases} N_{tot} = 10^4 \\ q_{min} = 0.1 \\ q_{max} = 1 \\ L = \frac{(\beta=10)(L_{\text{dim}}=2)}{(V=1)} = 20 \end{cases}$$

we need to run the simulation at least 10 times longer than in the previous case to obtain the steady PDF:

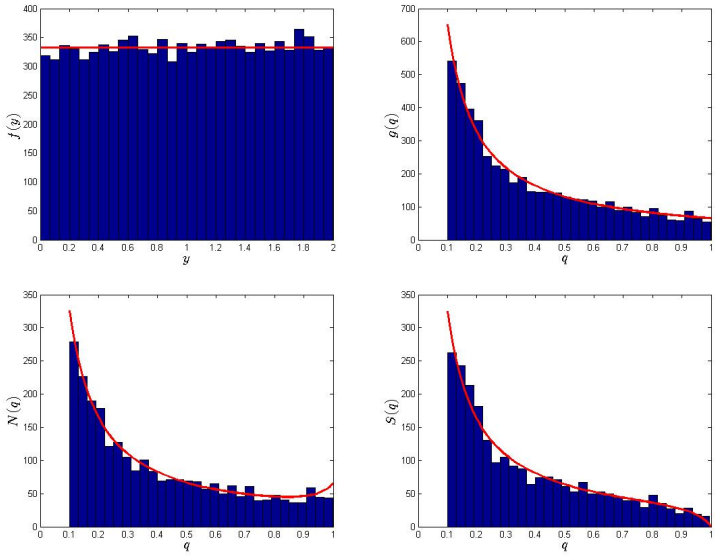

Figure 8: Numerical simulation of the SDE (blue) compared to the theoretical PDF (red) for $f(y)$ (top left), $g(q)$ (top right), $N(q)$ (bottom left) and $S(q)$ (bottom right) in case I (complete remoistening)

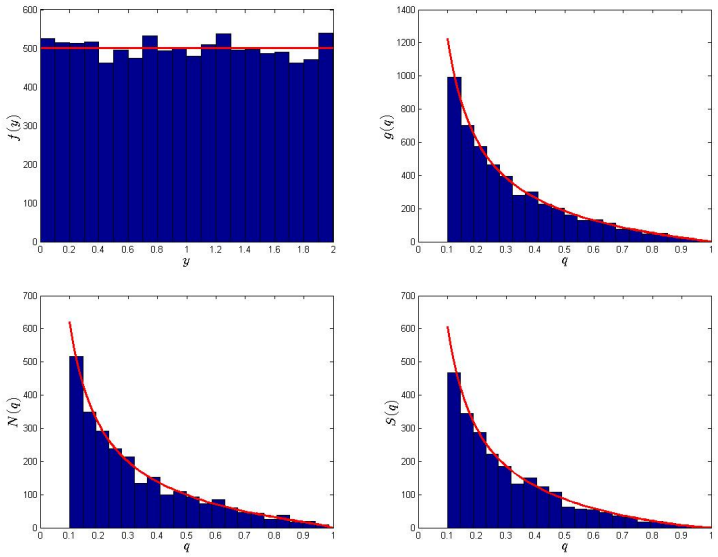

Figure 9: Numerical simulation of the SDE (blue) compared to the theoretical PDF (red) for $f(y)$ (top left), $g(q)$ (top right), $N(q)$ (bottom left) and $S(q)$ (bottom right) in case II (uniform remoistening)

Fortunately, in the case $L \to +\infty$, the system reaches the white noise limit (cf 4.2.9) and we can use the white noise SDEs (cf appendix A.2) to numerically simulate the steady PDF rather than waiting for an infinite amount of time. The global steady PDFs are very good quantities to check the validity of our analytical solutions because of their universal aspect, but because of this universality, they tell us very little about the two-stream model itself. The key quantities we are interested in are the degree of sub-saturation $\overline{q_{\text{sub}}} \overset{\text{def}}{=} \overline{q^* - q}$, the moisture flux $\mathcal{F}_q$, and the average condensation rate $\overline{\mathcal{C}} = -\frac{d\mathcal{F}_q}{dy}$. Their dimensional analytical expression in the two-stream model are respectively:

$$\boxed{\overline{q_{\text{sub}}}(y) = (y+1)\int_{q_{min}}^{q^*(y)} \frac{\Lambda(q)dq}{2+\frac{\beta y^*}{V}} - \int_{q_{min}}^{q^*(y)} \frac{d\Lambda}{dq} q_{sub} dq} \tag{105}$$

$$\boxed{\frac{\mathcal{F}_q}{\kappa} = -\frac{d\overline{q}}{dy} - |\frac{dq^*}{dy}| \frac{\Lambda(q^*)}{2+\frac{\beta y}{V}}} \tag{106}$$

$$\boxed{\frac{\overline{\mathcal{C}}(y)}{\kappa} = \frac{d^2 < q >}{dy^2} - \frac{1}{\sqrt{\tau\kappa}} \frac{dq^*}{dy} \frac{\Lambda(q^*)}{(2+\frac{\beta y}{V})^2} - (\frac{dq^*}{dy})^2 \frac{\frac{d\Lambda}{dq}(q^*)}{2+\frac{\beta y}{V}} - \frac{d^2 q^*}{dy^2} \frac{\Lambda(q^*)}{2+\frac{\beta y}{V}}} \tag{107}$$

We compare them in case I of a complete remoistening and for the numerical simulations with dimensionless extents $L \in \{2, 20\}$ studied previously:

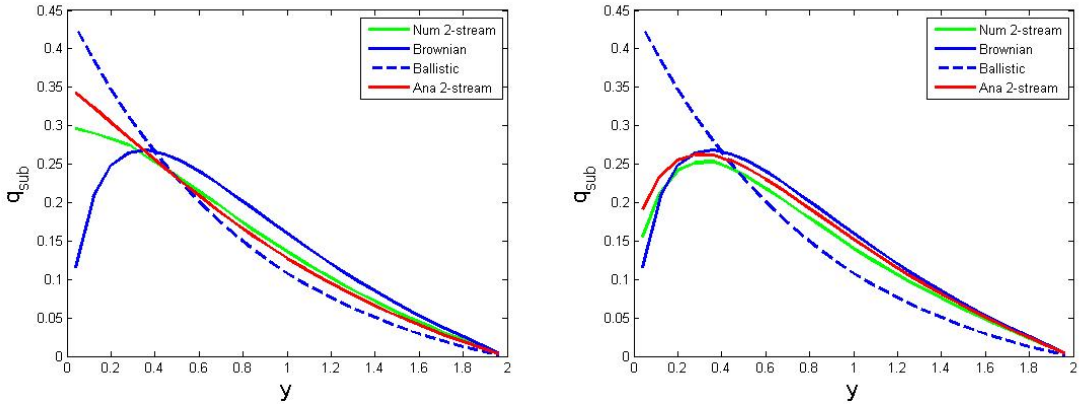

Figure 10: $< q_{sub} >$ vs y for $\hat{L} = 1$ (left) and $\hat{L} = 10$ (right); the analytical expression (red) is compared to the numerical flux (green)

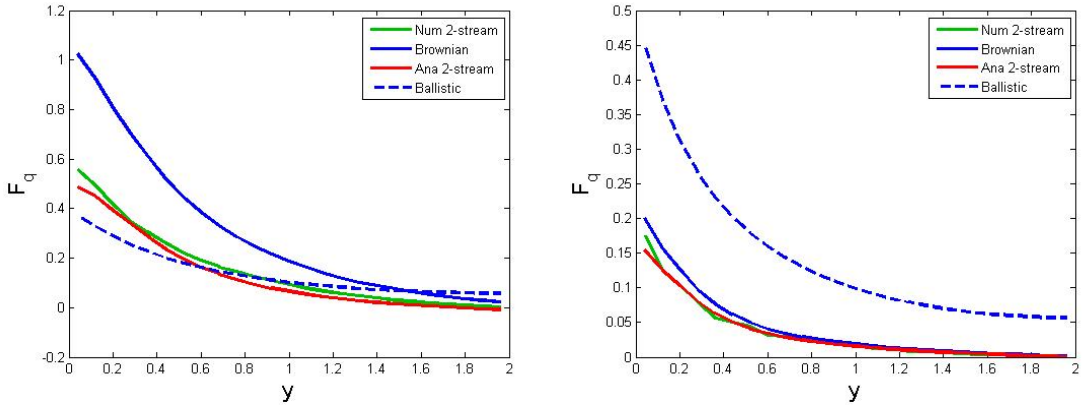

Figure 11: $\mathcal{F}_q$ vs y for $\hat{L} = 1$ (left) and $\hat{L} = 10$ (right); the analytical expression (red) is compared to the numerical flux (green)

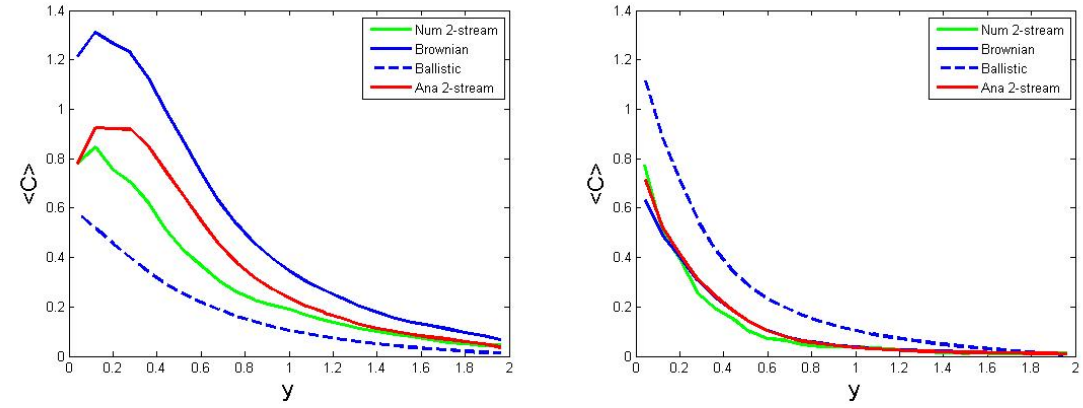

Figure 12: $< \mathcal{C} >$ vs y for $\hat{L} = 1$ (left) and $\hat{L} = 10$ (right); the analytical expression (red) is compared to the numerical flux (green)

Note that the analytical and numerical expressions of $\overline{\mathcal{C}}$ do not agree as well, as they imply the computation of second derivatives, which is not extremely precise due to the fact that we have only 25 discrete points to compute it in y.

4.2.9 White noise limit

The white noise limit can be obtained by keeping the diffusivity $\kappa = \frac{V^2}{\beta}$ and the dimensional length of the domain L_{dim} constant while letting $(V, \beta) \to +\infty$. As a consequence:

$$L = \frac{\beta L_{\text{dim}}}{V} = L_{\text{dim}} \sqrt{\frac{\beta}{\kappa}} \to +\infty$$

It follows that except very close to the equator (and the proportion of parcels with y close to 0 decreases very fast as $L \to +\infty$):

$$y \gg 1$$

The behavior of the PDF of the northwards and southwards parcels (91 and 89) follows:

$$2LS(q,y) = (y+2)[\frac{d}{dq}\frac{\Lambda(q)}{2+\frac{\beta y^*}{V}} + \frac{\delta^+(q-q_{min})}{2+L}] - \frac{d\Lambda}{dq} \sim_{L\to+\infty} y[\frac{d}{dq}\frac{\Lambda(q)}{y^*(q)} + \frac{\delta^+(q-q_{min})}{L}] - \frac{d\Lambda}{dq}$$

$$2L\mathcal{N}(q,y) = y[\frac{d}{dq}\frac{\Lambda(q)}{2+y^*(q)} + \frac{\delta^+(q-q_{min})}{2+L}] - \frac{d\Lambda}{dq} + \frac{2\Lambda[q^*(y)]}{2+\frac{\beta y}{V}}\delta^-[q-q^*(y)] \sim_{L\to+\infty} 2LS(q,y)$$

The total PDF is then asymptotically given by:

$$\boxed{L(\mathcal{N}+S) \sim_{L\to+\infty} y[\frac{d}{dq}\frac{\Lambda(q)}{y^*(q)} + \frac{\delta^+(q-q_{min})}{L}] - \frac{d\Lambda}{dq}} \tag{108}$$

which means that as expected, we recover the PDF found in the Brownian case (cf appendix A.2) assuming a white noise behavior in this limit.

4.3 Analytical approximation to the OU model

The n-stream model is the natural generalization of the two-stream model, where we consider n ensembles of parcels with n different velocities, that are exchanged at a rate proportional to $\frac{\beta}{2}$ (cf D). As $n \to +\infty$, the n-stream model converges to the OU model. Here, we use the analytical expressions derived for the average sub-saturation and the moisture flux in the case of the 2/3/4-stream model. We study how well they approximate the same quantities in the OU model, that we obtain from running Monte-Carlo simulations. The simulations are very similar to those described in 4.2.8 except that:

- Each parcel is advected by a velocity that follows an OU process with time-correlation τ.
- The parcels do not need to be exchanged between different processes anymore.

We have seen in 4.2.8 that the 2-stream model exactly reproduces the OU global distribution of moisture $P_Q(q)$. In figure 13, the average sub-saturation is very-well captured by the 2-stream model, except for the boundary layer near $y = 0$ which requires higher-order models, such as the 4-stream model. The values chosen for $(\kappa, \tau, L_{\text{dim}})$ give $L = 1$, and the OU model is quantitatively closer to the ballistic limit $L \ll 1$. As L increases, the OU average sub-saturation approaches the Ballistic limit, and a maximum appears in the distribution, corresponding to a zone of minimal relative humidity. In figure 14, the moisture flux is also well-approximated by the 2/3/4-stream model, confirming the important of taking into account the condensation in 1. Indeed, approximating the moisture flux as Fickian overestimates it by a factor ten near $y = 0$ for $L = 1$.

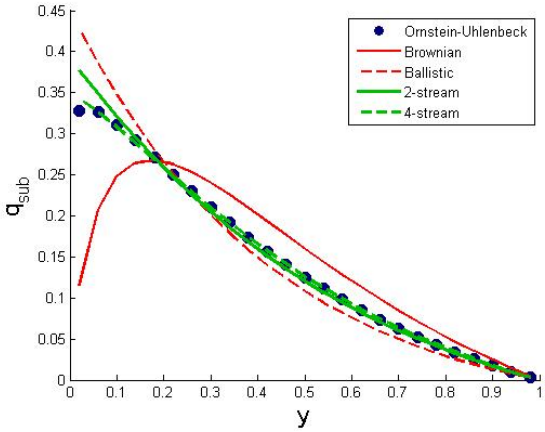

Figure 13: $\overline{q_{\text{sub}}}(y)$ vs y for $(\kappa, \tau, L_{\text{dim}}) = (1, 1, 1)$

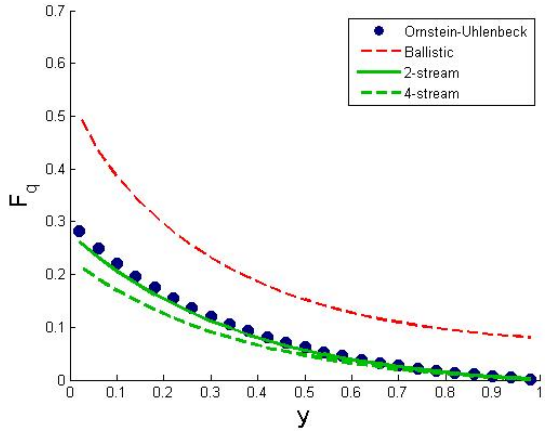

Figure 14: $\mathcal{F}_q(y)$ vs y for $(\kappa, \tau, L_{\text{dim}}) = (1, 1, 1)$

5 Conclusion

In conclusion, we have studied the main physics of the time-correlated advection-condensation model:

- The Ornstein-Uhlenbeck model for the velocity process is very well analytically approximated by the $n-$stream models.
- These models naturally produce a bimodal PDF of specific humidity, with a dry spike, a saturated spike, and a smooth part that decreases with q.
- The degree of sub-saturation and the condensation rate both decrease monotonically from the Equator to the Pole.
- In this model, time-correlation does not affect the global distribution for moisture $P_Q(q)$.

We have also obtained new insights in the diffusivity of a condensing scalar:

- The moisture flux is smaller than if Fick's law applied.
- The reduction of this moisture flux is proportional to the number of condensed parcels and the saturation specific humidity gradient to first approximation.

There are two important next steps for this project. First, it would be interesting to apply the 2/3/4-stream models to reanalysis data:

- Should latent heating be added to the model? The answer to this question would tell us when the moist parcels evolve on a dry or a moist isentropic surface.
- It is important to identify the limiting latitudes for this model: most likely, the model will not extend into the Tropics where microphysics and convection play a capital role in determining the distribution of specific humidity. As a consequence, the Tropics will most likely be taken as a Southern boundary condition for the model, producing a given remoistening distribution $\Phi(q)$.

In order to apply the model to reanalysis data, it is important to extent the model to a longitudinally symmetric sphere:

- An extension of the model to a longitudinally averaged sphere in the case of a white noise velocity process can be found in appendix E.
- It will be necessary to generalize the 2/3-stream model on a symmetric sphere.
- It would be ideal to also generalize the Ornstein-Uhlenbeck process to a symmetric sphere.

Finally, for geophysical applications, it will probably be important to generalize the model to a molecular diffusivity coefficient κ that varies with latitude.

A Limits of the OU model

A.1 The Ballistic limit

In the Ballistic limit $L \ll 1$, we rescale the latitude $Y = Ly$ to resolve the boundary layer, and obtain the following steady FPE in dimensionless form:

$$Lv\frac{\partial \rho}{\partial Y} = \frac{\partial}{\partial v}(v\rho + \frac{\partial \rho}{\partial v}) \tag{109}$$

To first approximation $L \to 0$; using 68 and 109:

$$\rho(q, y, v) \approx \frac{r(q, y, v)}{\sqrt{2\pi}L} \exp(-\frac{v^2}{2}) \tag{110}$$

where r is a function satisfying $\frac{\partial r}{\partial v} = 0$ almost everywhere. Physically, the velocity of a parcel remains unchanged from one boundary to another, which allows us write r as a sum of:

- A northwards part, understanding that the parcels saturate continuously as they move northwards:
$$\{\Phi(q)\mathcal{H}[q^*(y) - q] + \delta^-[q - q^*(y)]\Lambda(q)\}\mathcal{H}(v)$$
where we have used the definition 90.
- A southwards part, only constituted by a dry spike:
$$\delta^+(q - q_{\min})\mathcal{H}(-v)$$

In this limit, the average of a function $f(q)$ is defined by:

$$\overline{f}(y) \overset{\text{def}}{=} L\int_{-\infty}^{+\infty} dv \int_{q_{\min}}^{q_s(y)} f(q)\rho(q, y, v)dq \tag{111}$$

From 111 and 110, we can approximate:

- The sub-saturation:
$$\overline{q_{sub}}(y) \approx q^*(y) - q_{\min} - \frac{1}{2}\int_{q_{\min}}^{q^*(y)} \Lambda$$
- The moisture flux:
$$\mathcal{F}_q \approx \sqrt{\frac{2\kappa}{\pi\tau}}\overline{q}(y)$$

A.2 The Brownian limit

In the Brownian limit $L \gg 1$, the stochastic differential equations for the displacement collapses to the model developed in [23]:

$$dY(t) = \sqrt{2\kappa} dW(t)$$

making the FPE a Laplace equation for the PDF:

$$\frac{\partial^2 \rho}{\partial y^2} = 0$$

The detailed solution of the Brownian problem can be found in [Sukhatme]; the main results of interest in our case are:

- The PDF:
$$\rho(q, y) = \frac{1}{L}\{\Phi(q) + y[\frac{\delta^+(q - q_{\min})}{L} + \frac{d}{dq}\frac{\Lambda(q)}{y^*(q)}]\}$$
- The average sub-saturation:
$$\overline{q_{sub}}(y) = \int_{q_{\min}}^{q^*(y)} \Phi q_{sub} + y \int_{q_{\min}}^{q^*(y)} \frac{\Lambda(q) dq}{y^*(q)}$$
- The Fickian moisture flux:
$$\mathcal{F}_q = -\kappa \frac{d\bar{q}}{dy}$$

B The $\varphi_\alpha-$functions

B.1 Definition from the second order differential equation

We start with the following second order differential equation, which is an eigenvalue problem in α:

$$\alpha v \varphi_\alpha = \mathcal{D}\varphi_\alpha \stackrel{\text{def}}{=} \frac{d^2\varphi_\alpha}{dv^2} - v\frac{d\varphi_\alpha}{dv} \tag{112}$$

Note that $\mathcal{D}$ is a self-adjoint operator for the scalar product defined for two well-behaved functions (f, g) by:

$$< f|g > \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-\frac{v^2}{2}) f(v) g(v) dv \tag{113}$$

The eigenfunctions, ie the bounded solution of this differential equation for $\alpha^2 \in \mathbb{N}$, are a special kind of parabolic cylinder functions, that we call the $\varphi_\alpha-$functions, defined by:

$$\boxed{\varphi_\alpha(v) \stackrel{\text{def}}{=} \exp(-\alpha v - \alpha^2) He_{\alpha^2}(v + 2\alpha)} \tag{114}$$

where He_{α^2} is the α^2-th "probabilistic" Hermite polynomial in v. *To prove that the functions 114 are solutions to the eigenvalue problem 112, we can change variables from φ_α to ψ:*

$$\varphi_\alpha(v) = \psi(v) \exp(-\alpha v - \alpha^2)$$

$$\frac{d\varphi_\alpha}{dv} = [\frac{d\psi}{dv} - \alpha\psi]\exp(-\alpha v - \alpha^2)$$
$$\frac{d^2\varphi_\alpha}{dv^2} = [\frac{d^2\psi}{dv^2} - 2\alpha\frac{d\psi}{dv} + \alpha^2\psi]\exp(-\alpha v - \alpha^2)$$

leading to the differential equation:

$$\frac{d^2\psi}{dv^2} - (v + 2\alpha)\frac{d\psi}{dv} + \alpha^2\psi = 0$$

An obvious change of variable is $w = v + 2\alpha$, leading to:

$$\frac{d^2\psi}{dw^2} - w\frac{d\psi}{dw} + \alpha^2\psi = 0$$

which solutions can be express using the well-known "deterministic" Hermite polynomials of α^2-order:

$$\psi = He_{\alpha^2}(w)$$

where the "probabilistic" Hermite polynomials are defined by:

$$He_{\alpha^2}(v) \stackrel{\text{def}}{=} (-1)^{\alpha^2}\exp(\frac{v^2}{2})\frac{d^{\alpha^2}}{dv^{\alpha^2}}\exp(-\frac{v^2}{2}) = (v - \frac{d}{dv})^{\alpha^2}\cdot 1 \tag{115}$$
$$\psi = H_{\alpha^2}(\frac{w}{\sqrt{2}})$$

It is also possible to use "deterministic" Hermite polynomials, defined by:

$$H_{\alpha^2}(v) \stackrel{\text{def}}{=} (-1)^{\alpha^2}\exp(v^2)\frac{d^{\alpha^2}}{dv^{\alpha^2}}\exp(-v^2) = (2v - \frac{d}{dv})^{\alpha^2}\cdot 1 \tag{116}$$

The solution ψ is then given by:

$$\psi = 2^{-\frac{\alpha^2}{2}}H_{\alpha^2}(\frac{w}{\sqrt{2}})$$

Finally, to make sure we have solved the complete eigenvalue problem, we have to check for degenerate eigenvalues, which we do by differentiating the initial differential equation with respect to α:

$$\mathcal{D}\frac{\partial\varphi_\alpha}{\partial\alpha} = v\varphi_\alpha + \alpha v\frac{\partial\varphi_\alpha}{\partial\alpha}$$

Since $\mathcal{D}$ is self-adjoint for $< | >$:

$$0 =< \frac{\partial\varphi_\alpha}{\partial\alpha}|(\mathcal{D} - \alpha v)\varphi_\alpha >=< (\mathcal{D} - \alpha v)\varphi_\alpha|\varphi_\alpha >=< v\varphi_\alpha|\varphi_\alpha >$$

This equation is only satisfied for $\alpha = 0$, which means that we may take a generalized eigenfunction at this point:

$$\varphi_g(v) \stackrel{\text{def}}{=} v$$

Since:

$$\forall\alpha,\ <\varphi_\alpha|\varphi_g> \neq 0$$

by differentiating a second time with respect to α, we can show that the multiplicity of the eigenvalue $\alpha = 0$ is exactly 2. This proves that we only need φ_0 and φ_g to construct a general solution to the initial differential equation.

B.2 Orthogonality of the $\varphi_\alpha-$functions

Let φ_α and φ_β be 2 eigenfunctions of 112 with eigenvalues $\alpha \neq \beta$:

$$(\alpha - \beta) < v\varphi_\alpha|\varphi_\beta >=< \mathcal{D}\varphi_\alpha|\varphi_\beta > - < \varphi_\alpha|\mathcal{D}\varphi_\beta >= 0$$

so that $(\varphi_\alpha, \varphi_\beta)$ satisfy the orthogonality relation:

$$< v\varphi_\alpha|\varphi_\beta >= \delta_{\alpha\beta} \tag{117}$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol for (α, β). From the properties of Hermite polynomials, it can be proven that:

$$< v\varphi_0|\varphi_g >= 1$$

$$< v\varphi_\alpha|\varphi_\beta >= -2\alpha(\alpha^2!)\delta_{\alpha\beta} \cdot 1\{\alpha\beta > 0\}$$

$$< v\varphi_0|\varphi_0 >=< v\varphi_g|\varphi_g >=< v\varphi_0|\varphi_{\alpha\neq 0} >=< v\varphi_g|\varphi_{\alpha\neq 0} >= 0$$

meaning that the expansion of a well-behaved function $f(v) = o_{v\to+\infty}[\exp(\frac{v^2}{2})]$ satisfying the initial differential equation can be written:

$$f(v) = \frac{< v\varphi_g|f >}{< v\varphi_g|\varphi_0 >}\varphi_0 + \frac{< v\varphi_0|f >}{< v\varphi_0|\varphi_g >}\varphi_g + \sum_{\alpha^2\in\mathbb{N}^*} \frac{< v\varphi_\alpha|f >}{< v\varphi_\alpha|\varphi_\alpha >}\varphi_\alpha$$

$$\boxed{f(v) =< v^2|f > + < v|f > v + \sum_{\alpha^2\in\mathbb{N}^*} \frac{< v\varphi_\alpha|f >}{< v\varphi_\alpha|\varphi_\alpha >}\varphi_\alpha} \tag{118}$$

From the properties of the Hermite polynomials, we also obtain:

$$< 1|\varphi_{\alpha>0} >= \alpha^{\alpha^2} \exp(-\frac{\alpha^2}{2})$$

$$< 1|\varphi_{\alpha<0} >= (-1)^{\alpha^2}\alpha^{\alpha^2} \exp(-\frac{\alpha^2}{2})$$

B.3 Characteristics of the $\varphi_\alpha-$functions

The four lowest orders functions can be found by setting:

$$\alpha \in \pm\{1, \sqrt{2}\}$$

From definition 114, we find:

$$\varphi_{-\sqrt{2}}(v) = \{v^2 - 4\sqrt{2}v + 7\} \exp(\sqrt{2}v - 2)$$

$$\varphi_{-1}(v) = (v - 2) \exp(v - 1)$$

$$\varphi_1(v) = (v + 2) \exp(-v - 1)$$

$$\varphi_{\sqrt{2}}(v) = \{v^2 + 4\sqrt{2}v + 7\} \exp(-\sqrt{2}v - 2)$$

Note that we actually only need to study $\alpha > 0$ or $\alpha < 0$ as from the evenness/oddness of the Hermite polynomials:

$$\varphi_{-\alpha}(v) = (-1)^{\alpha^2}\varphi_\alpha(v)$$

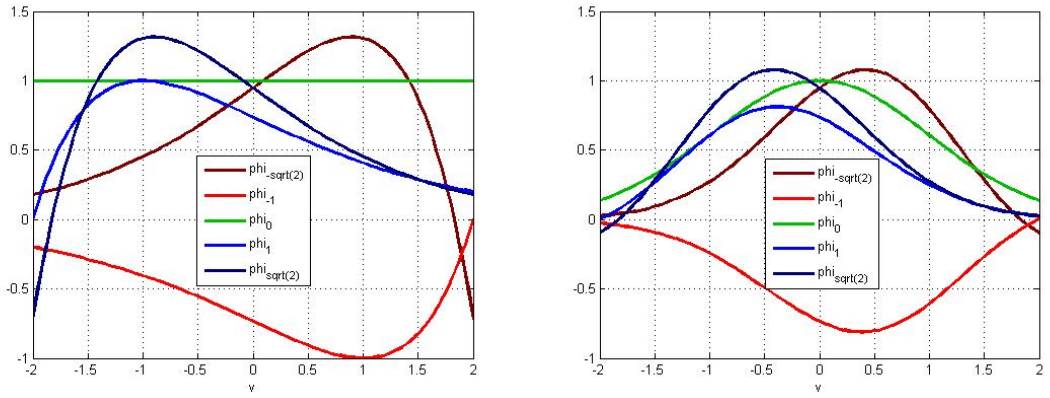

Figure 15: $\varphi_\alpha(\hat{v})$ (left) and $\exp(-\frac{\hat{v}^2}{2})\varphi_\alpha(\hat{v})$ (right) for $\alpha \in \{-\sqrt{2}, -1, 0, 1, \sqrt{2}\}$

C Dry and saturated spikes of the PDF

C.1 Definition

By construction of the n-stream (D) and OU models (4.1), the distribution comprises three "group" of parcels:

1. The parcels moving northwards in a zone where $y > y^*(q)$ are supersaturated and condensate as they move Northwards, constituting the saturated spike $\delta^-[q - q^*(y)]$.

2. The moisture of the parcels which have last hit the Northern boundary $y = L$ where $q = q^*(L) = q_{\min}$ cannot be changed until the parcels hit the Southern boundary $y = 0$. They constitute the dry spike $\delta^+(q - q_{\min})$.

3. The remaining part of the PDF is referred as the smooth part.

C.2 Saturated spike in the two-stream model

To understand how condensation occurs in the two-stream model, we relax the assumption of instant condensation; in dimensionless form:

$$\frac{\partial}{\partial y}\begin{pmatrix} N \\ -S \end{pmatrix} - \frac{\partial}{\partial q}[\mathcal{C}(q,y)\begin{pmatrix} N \\ S \end{pmatrix}] = \frac{1}{2}\begin{pmatrix} S-N \\ N-S \end{pmatrix}$$

Using the notations introduced in 4, we define a dimensionless condensation time that we assume small $\lambda^{-1}\tau \ll 1$, and use:

$$\mathcal{C}(q,y) = \lambda^{-1}\tau[q - q^*(y)]\mathcal{H}[q - q^*(y)]$$

In the supersaturated region $q \geq q^*(y)$, the parcels condensate fast enough for their velocities not to change sign, and we can assume $S \ll N$. The second equation gives $S \sim \lambda^{-1}\tau N$ making the approximation self-consistent, while integrating the first equation in the super-saturated region gives:

$$\frac{dW}{dy} + \frac{W(y)}{2} = -\frac{dq^*}{dy} N[q^*(y), y] \tag{119}$$

where we have defined the total density of supersaturated parcels:

$$W(y) \stackrel{\text{def}}{=} L \int_{q^*(y)}^{+\infty} N(q, y) dq \tag{120}$$

D The n-stream model

D.1 Definition

The n-stream model is a natural discretization of the OU model and thus a generalization of the two-stream model. It is easy to think about it as n bits which can take the value ± 1. If we consider one combination of bits, the sum of the bits gives the velocity of the corresponding parcel's ensemble, and allows us to build the advection matrix of the process$\boldsymbol{A}$. The exchange of parcels between ensembles happens when one bit's value is modified, and the probability of switching from one velocity to another allows us to build the exchange matrix of the process $\boldsymbol{E}$. Finally, the normalization condition is obtained by considering the probability of a combination of bits. Mathematically, defining the vectorial PDF $\underline{\rho}(q, y)$:

$$\frac{\partial}{\partial t}\underline{\rho} + \frac{V}{\sqrt{n-1}} \boldsymbol{A} \frac{\partial}{\partial y}\underline{\rho} = \frac{\beta}{2} \boldsymbol{E} \underline{\rho} \tag{121}$$

$$\boldsymbol{A}_{ij} \stackrel{\text{def}}{=} (n + 1 - 2i)\delta_{ij} \tag{122}$$

$$\boldsymbol{E}_{ij} \stackrel{\text{def}}{=} -n\delta_{ij} + j\delta_{i+1,j} + (n - j)\delta_{i-1,j} \tag{123}$$

where δ_{ij} is the Kronecker symbol. The normalization condition for $\underline{\rho}(q, y)$ yields:

$$\int_{q_{\min}}^{q^*(y)} \rho_i(q, y) dq = \frac{(n-1)!}{(i-1)!(n-i)!} \frac{1}{2^{n-1} L} \tag{124}$$

By construction of the n-stream model, the diffusivity of the parcels is given by:

$$\kappa \stackrel{\text{def}}{=} \frac{V^2}{\beta} \tag{125}$$

D.2 The three-stream model

The steady three-stream equations can be written:

$$\frac{V}{\sqrt{2}} \frac{\partial}{\partial y} \begin{pmatrix} 2N \\ 0 \\ -2S \end{pmatrix} = \frac{\beta}{2} \begin{pmatrix} -2 & 1 & 0 \\ 2 & -2 & 2 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} N \\ M \\ S \end{pmatrix}$$

where (N, M, S) respectively represent the northwards, motionless and southwards parcel's PDFs. We now work with the dimensionless variables defined in the two-stream model: 83 and 84. The normalization condition for the vectorial PDF is:

$$\int_{q_{\min}}^{q^*(y)} \begin{pmatrix} N \\ M \\ S \end{pmatrix} dq = \frac{1}{4L} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

Following the same steps as for the two-stream model, we find that the vectorial PDF is the sum of:

- A smooth part:
$$\frac{\Phi(q)}{4L} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + C_{\text{smooth}}(q) \begin{pmatrix} y \\ 2(y+\sqrt{2}) \\ y+2\sqrt{2} \end{pmatrix}$$
- A dry spike:
$$C_{\text{dry}} \begin{pmatrix} y \\ 2(y+\sqrt{2}) \\ y+2\sqrt{2} \end{pmatrix}$$
- A saturated spike:
$$\frac{W(y)}{4L} \delta^-[q - q^*(y)] \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

where from the boundary and normalization conditions:

$$C_{\text{dry}} = \frac{1}{4L(L+2\sqrt{2})}$$

$$C_{\text{smooth}}(q) = \frac{1}{4L} \frac{d}{dq} \frac{\Lambda(q)}{y^*(q)+2\sqrt{2}}$$

$$W(y) = \frac{\sqrt{2}\Lambda[q_s(y)]}{y+2\sqrt{2}}$$

Defining the average of a function $f(q)$ by:

$$\overline{f}(y) \stackrel{\text{def}}{=} L \int_{q_{\min}}^{q^*(y)} f(q)(N+M+S)(q)dq$$

the average sub-saturation is:

$$\overline{q_{\text{sub}}} = \int_{q_{\min}}^{q^*(y)} \Phi q_{\text{sub}} + (y+\sqrt{2}) \int_{q_{\min}}^{q^*(y)} \frac{\Lambda(q)dq}{y^*(q)+2\sqrt{2}}$$

Defining the moisture flux as:

$$\mathcal{F}_q \stackrel{\text{def}}{=} L_{\text{dim}} \int_{q_{\min}}^{q^*(y)} \sqrt{2} V q (N-S)(q)dq$$

we can once again relate it to the moisture gradient in dimensional form:

$$\mathcal{F}_q = -\kappa(\frac{d\overline{q}}{dy} - W|\frac{dq^*}{dy}|)$$

There is very little difference between the 2/3-stream models, except for the presence of motionless parcels, which explains why the moisture flux is smaller in the three-stream case.

D.3 The four-stream model

The steady four-stream equations can be written:

$$\frac{V}{\sqrt{3}}\frac{\partial}{\partial y}\begin{pmatrix} 3N_3 \\ N_1 \\ -S_1 \\ -3S_3 \end{pmatrix} = \frac{\beta}{2}\begin{pmatrix} -3 & 1 & 0 & 0 \\ 3 & -3 & 2 & 0 \\ 0 & 2 & -3 & 3 \\ 0 & 0 & 1 & -3 \end{pmatrix}\begin{pmatrix} N_3 \\ N_1 \\ S_1 \\ S_3 \end{pmatrix}$$

where (N_3, N_1, S_1, S_3) respectively represent the fast northwards, slow northwards, slow southwards and fast southwards parcel's PDFs. We now work with the dimensionless variables defined in 83 and 84. The normalization condition for the vectorial PDF is:

$$\int_{q_{\min}}^{q^*(y)} \begin{pmatrix} N_3 \\ N_1 \\ S_1 \\ S_3 \end{pmatrix} dq = \frac{1}{8L}\begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix}$$

Following the same steps as for the 2/3-stream models, we find that the vectorial PDF is the sum of:

- A smooth part:
$$\frac{\Phi(q)}{8L}\begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix} + 147\sqrt{3}C_{\text{smooth},\pm}(q)\underline{c}_{\pm}(y)$$
- A dry spike:
$$147\sqrt{3}C_{\text{dry},\pm}\underline{c}_{\pm}(y)$$
- A saturated spike:
$$\frac{\delta^-[q - q^*(y)]}{8L}\begin{pmatrix} W_3(y) \\ W_1(y) \\ 0 \\ 0 \end{pmatrix}$$

where a sum over $\pm$ is implied and $\underline{c}_{\pm}(y)$ is given by:

$$\begin{pmatrix} (1 \mp 6\sqrt{2}) \pm \sqrt{6}(y + 2\sqrt{3}) - \exp(\pm\sqrt{6}y) \\ 3(1 \mp 6\sqrt{2}) \pm \sqrt{6}(3y + 8\sqrt{3}) - 3(1 \pm \sqrt{2})\exp(\pm\sqrt{6}y) \\ 3(1 \mp 6\sqrt{2}) \pm \sqrt{6}(3y + 10\sqrt{3}) - 3(5 \pm 4\sqrt{2})\exp(\pm\sqrt{6}y) \\ (1 \mp 6\sqrt{2}) + \pm\sqrt{6}(y + 4\sqrt{3}) + (3 \pm 2\sqrt{2})\exp(\pm\sqrt{6}y) \end{pmatrix}$$

From the boundary and normalization conditions, we can solve for the six unknowns $(C_{\text{smooth},\pm}(q), C_{\text{dry},\pm}, W_1(y), W_3(y))$:

$$C_{\text{dry},\pm} = -\frac{1}{8L}\frac{\sqrt{3}}{1764}\Gamma(L)$$

$$C_{\text{smooth},\pm}(q) = -\frac{1}{8L}\frac{\sqrt{3}}{1764}\frac{d(\Lambda\Gamma_{\pm}[y^*])}{dq}$$

$$\begin{pmatrix} W_3(y) \\ W_1(y) \end{pmatrix} = \mathcal{W}(y) \begin{pmatrix} -\sqrt{2} + 5\sqrt{2}\cosh(\sqrt{6}y) + 8\sinh(\sqrt{6}y) \\ 3[\sqrt{2} + 3\sqrt{2}\cosh(\sqrt{6}y) + 4\sinh(\sqrt{6}y)] \end{pmatrix}$$

where:

$$\Gamma_{\pm}(y) = \frac{\sqrt{2} + (3\sqrt{2} \mp 4)\exp(\mp\sqrt{6}y)}{-\sqrt{2} + \sqrt{2}(9 + 2\sqrt{3}y)\cosh(\sqrt{6}y) + (14 + 3\sqrt{3}y)\sinh(\sqrt{6}y)}$$

$$\mathcal{W}(y) = \frac{2\Lambda[q^*(y)]}{-\sqrt{2} + \sqrt{2}(9 + 2\sqrt{3}y)\cosh(\sqrt{6}y) + (14 + 3\sqrt{3}y)\sinh(\sqrt{6}y)}$$

Defining the average of a function $f(q)$ by:

$$\overline{f}(y) \stackrel{\text{def}}{=} L\int_{q_{\min}}^{q^*(y)} f(q)(N_3 + N_1 + S_1 + S_3)(q)dq$$

the average sub-saturation is:

$$\overline{q_{\text{sub}}} = \int_{q_{\min}}^{q^*(y)} \Phi q_{\text{sub}} - \mathcal{Q}(y)$$

where:

$$\mathcal{Q}(y) = \frac{1 \pm \sqrt{6}(\sqrt{3} + y) - 2(1 \pm \sqrt{2})\exp(\pm\sqrt{6}y)}{4}\int_{q_{\min}}^{q^*(y)} \Lambda\Gamma_{\pm}[y^*]$$

where a sum over $\pm$ is implied. Defining the moisture flux as:

$$\mathcal{F}_q \stackrel{\text{def}}{=} L_{\text{dim}}\int_{q_{\min}}^{q^*(y)} \sqrt{3}Vq(N_3 + \frac{N_1 - S_1}{3} - S_3)(q)dq$$

we need to use the first and second meridional derivative of the average moisture to write the gradient/flux relation:

$$\mathcal{F}_q = -\kappa(\frac{d\overline{q}}{dy} - W_{tot}|\frac{dq^*}{dy}|)(1 - \frac{2}{D})$$

$$+\kappa\mathcal{L}(\frac{d^2\overline{q}}{dy^2} - W_{tot}\frac{d^2q^*}{dy^2})(1 - \frac{2 + (\sqrt{2} - 1)\exp(-\frac{\sqrt{6}\beta y}{V})}{D})$$

where $W_{tot} = W_1 + W_3$ is the total amount of parcels which have condensed, $\mathcal{L} = \frac{V}{\sqrt{6}\beta}$ is the boundary layer decay-length of the 4-stream model and the denominator in dimensionless form is:

$$D(y) = 2 + \cosh(\sqrt{6}y) + \sqrt{2}\sinh(\sqrt{6}y)$$

Qualitatively, the 4-stream model adds a boundary layer structure to the solution, which explains the better agreement of the analytical 4-stream moisture flux with the OU flux for small y (cf figure 14). Furthermore, the boundary layer provides a positive contribution depending on $\frac{d^2q}{dy^2}$ to the moisture flux, once again decreased by the amount of condensed parcels.

E Brownian motion with condensation a longitudinally averaged sphere

E.1 Preliminary assumptions

We are now interested in the advection-condensation problem on a isentropic sphere of radius a, parametrized by longitude λ and latitude θ. We make the following simplifying assumptions:

1. By taking a longitudinal average of the model, we reduce our problem to a 1D process on the surface of a sphere.

2. The domain we work on extends from the Equator ($\theta = 0$) to a given latitude $\theta_L \in]0, \frac{\pi}{2}[$.

3. We suppose that the saturation specific humidity monotonically decreases from the Southern boundary of the domain ($\theta = 0$) where $q^* = q_{max}$ to the Northern boundary of the domain ($\theta = \theta_L$) where $q^* = q_{min} = \min_\theta q(\theta)$.

4. We model the source term $\mathcal{S}$ following Sukhatme et al.. The model "resets" the specific humidity q to a random value chosen from a specified distribution $\Phi(q)$ when it encounters the Southern boundary of the domain ($\theta = 0$). Physically, we could think of the fictive domain $\theta \leq 0$ as the Tropics, that remoisten the dry parcels that are advected southward by the mid-latitude eddies. The normalization condition on Φ yields:
$$\int_{q_{min}}^{q_{max}} \Phi = 1$$

5. We assume that the moist parcels on this sphere have a constant diffusivity coefficient κ.

E.2 Normalization of the concentration of moist parcel

On the sphere of constant radius, we define $c(\lambda, \theta, t)$ [m^{-2}] as the physical concentration of moist parcels. It can be computed from the generalized concentration γ in (q, λ, θ, t) space by integrating over the specific humidity q:

$$c(\lambda, \theta, t) \overset{\text{def}}{=} \int_{q_{min}}^{q^*(\theta)} \gamma(q, \lambda, \theta, t) dq$$

By definition, the concentration of moist parcels c verifies the following normalization condition:

$$N_{tot} = a^2 \int_0^{2\pi} d\lambda \int_0^{\theta_L} d\theta \cos(\theta) c(\lambda, \theta, t) = a^2 \int_{q_{min}}^{q^*(\theta)} dq \int_0^{2\pi} d\lambda \int_0^{\theta_L} d\theta \cos(\theta) \gamma(q, \lambda, \theta, t) \tag{126}$$

Defining the zonally averaged generalized condensation as:

$$< \gamma > (q, \theta, t) \overset{\text{def}}{=} \frac{1}{2\pi} \int_0^{2\pi} d\lambda \gamma(q, \lambda, \theta, t)$$

and introducing the distribution:

$$\nu(q, \theta, t) \overset{\text{def}}{=} \frac{2\pi a^2}{N_{tot}} < \gamma > (q, \theta, t) = \frac{a^2}{N_{tot}} \int_0^{2\pi} d\lambda \gamma(q, \lambda, \theta, t)$$

the previous normalization condition 126 yields:

$$\boxed{\int_0^{\theta_L} \int_{q_{min}}^{q^*(\theta)} \nu(q, \theta, t) \cos\theta d\theta dq = 1}$$

Integrating $\nu(q, \theta)$ over (q) produces the marginal density:

$$p(\theta, t) \overset{\text{def}}{=} \int_{q_{min}}^{q^*(\theta)} \nu(q, \theta) dq$$

In the stationary limit $\lim_{t \to +\infty} p(t) = p_\infty$ and $\underline{\nabla p_\infty} = \underline{0}$, which allows us to compute p_∞ through its normalization condition:

$$\int_0^{\theta_L} p_\infty \cos\theta d\theta = 1 \;\Rightarrow\; p_\infty = \sin^{-1}(\theta_L)$$

As a consequence, if we seek a long-time equilibrium solution $\nu_\infty = \lim_{t \to +\infty} \nu$, we require:

$$\boxed{1 = \sin\theta_L \cdot \int_{q_{min}}^{q^*(\theta)} \nu_\infty(q, \theta) dq} \tag{127}$$

E.3 FPE and SDE of the PDF

The generalized concentration γ of moist parcels verifies the evolution-condensation-diffusion equation on this sphere:

$$\frac{\partial \gamma}{\partial t} + \frac{\partial}{\partial q}[(\mathcal{S} - \mathcal{C})\gamma] = \kappa \nabla^2 \gamma = \frac{\kappa}{a^2} \{ \frac{1}{\cos\theta} \frac{\partial}{\partial \theta} [\cos\theta \frac{\partial \gamma}{\partial \theta}] + \frac{1}{\cos^2\theta} \frac{\partial^2 \gamma}{\partial \lambda^2} \}$$

Integrating the previous relation from $\lambda = 0$ to $\lambda = 2\pi$ and using the fact that:

- $\frac{\partial \gamma}{\partial \lambda}(\lambda = 0) = \frac{\partial \gamma}{\partial \lambda}(\lambda = 2\pi)$ because $\frac{\partial \gamma}{\partial \lambda}$ is a continuous function of longitude.

- $\frac{\partial \gamma}{\partial q}(\lambda = 0) = \frac{\partial \gamma}{\partial q}(\lambda = 2\pi)$ because $\frac{\partial \gamma}{\partial q}$ is a continuous function of longitude.
- $\frac{\partial}{\partial \lambda}(\mathcal{S} - \mathcal{C}) = 0$, because the source and the sinks in the system are longitudinally invariant.

we obtain:

$$\frac{\partial < \gamma >}{\partial t} = \frac{\kappa}{a^2 \cos\theta} \frac{\partial}{\partial \theta}[\cos\theta \frac{\partial < \gamma >}{\partial \theta}]$$

Multiplying the previous equation by $\frac{2\pi a^2}{N_{tot}}$, we show that ν verifies the following evolution-diffusion equation:

$$\boxed{\frac{\partial \nu}{\partial t} = \frac{\kappa}{a^2 \cos\theta} \frac{\partial}{\partial \theta}[\cos\theta \frac{\partial \nu}{\partial \theta}]} \tag{128}$$

We are now interested in the PDF $n(q, \theta, t)$ of Θ, which is the random variable associated to the latitudinal motion of the moist parcel. From the normalization condition on $n(q, \theta, t)$:

$$\int_0^{\theta_L} \int_{q_{min}}^{q^*(\theta)} n(q, \theta, t) d\theta dq = 1$$

and the normalization condition on ν_∞ (127) we can identify n as:

$$\boxed{n(q, \theta, t) = \cos\theta \cdot \nu(q, \theta, t)}$$

It is then immediate to derive the FPE for n from the equation that ν satisfies:

$$\frac{\partial n}{\partial t} = \frac{\kappa}{a^2} \frac{\partial}{\partial \theta}[\tan\theta \cdot n + \frac{\partial n}{\partial \theta}]$$

By identifying the drift and the diffusion terms, we recognize that the previous FPE corresponds to the Ito's SDE for a random walk on a sphere with diffusivity κ:

$$\boxed{d\Theta(t) = -\frac{\kappa \tan[\Theta(t)]}{a^2} dt + \frac{\sqrt{2\kappa}}{a} dW(t)} \tag{129}$$

We can see that the random walk is biased towards small values of Θ as the surface of the sphere decreases with the latitude θ. The SDE for the specific humidity stays the same:

$$dQ(t) = \{\mathcal{S}[\Theta(t)] - \mathcal{C}[\Theta(t), Q(t)]\} dt$$

E.4 Steady PDF

From equation 128, we can prove that the steady Markov diffusive process $\nu_\infty = \lim_{t \to +\infty} \nu$ verifies Laplace's equation:

$$\frac{\partial}{\partial \theta}[\cos\theta \frac{\partial \nu_\infty}{\partial \theta}] = 0$$

Integrating this equation is immediate and gives:

$$\nu_\infty(q, \theta) = A(q)\mathcal{L}(\theta) + B(q)$$

where we have defined the function:

$$\mathcal{L}(\theta) \overset{\text{def}}{=} \int_0^\theta \frac{d\theta'}{\cos(\theta')} = \ln[\frac{1+\sin(\theta)}{\cos(\theta)}]$$

and (A, B) need to be determined:

- First, we consider the origin $\theta = 0$ where $\nu_\infty(q, 0) = B(q) = \beta\Phi(q)$ where the second equality comes from the resetting condition at $(\theta = 0)$. We combine the normalization condition on Φ and the integral condition on ν_∞ to find β:

$$\begin{cases} \int_{q_{min}}^{q_{max}} \beta\Phi(q)dq = \beta \\ \int_{q_{min}}^{q_{max}} \nu_\infty dq = \sin^{-1}\theta_L \end{cases} \Rightarrow \ \beta = \sin^{-1}\theta_L$$

- Because the only moistening mechanism at $(\theta = 0)$, the driest parcel $q = q_{min}$ will keep their dryness until they hit the Southern boundary. Consequently, ν_∞ will be peaked at q_{min}, and we need to take into account the "dry spike" $\delta^+(q - q_{min})$ in the distribution:

$$\nu_\infty(q, \theta) = [A_1 \cdot \delta(q - q_{min}) + A_2(q)] \cdot \mathcal{L}(\theta) + \frac{\Phi(q)}{\sin\theta_L}$$

  where (A_1, A_2) are to be determined. At $\theta = \theta_L$ where $q^* = q_{min}$:

$$\int_{q_{min}}^{q_{max}} \nu_\infty(q, \theta_L)dq = A_1\mathcal{L}_L = \sin^{-1}\theta_L \ \Rightarrow \ A_1 = \{\sin\theta_L \cdot \mathcal{L}_L\}^{-1}$$

  where we have defined :

$$\mathcal{L}_L = \mathcal{L}(\theta_L) = \ln[\frac{1+\sin(\theta_L)}{\cos(\theta_L)}]$$

  which verifies:

$$\begin{cases} \mathcal{L}(\theta_L = 0) = 0 \\ \mathcal{L}(\theta_L \to \frac{\pi}{2}) \to +\infty \end{cases}$$

- Finally, we substitute ν_∞ in the normalization condition 127:

$$1 = \sin\theta_L \int_{q_{min}}^{q^*(\theta)} \nu_\infty dq = \frac{\mathcal{L}(\theta)}{\mathcal{L}_L} + \sin\theta_L\mathcal{L}(\theta) \int_{q_{min}}^{q^*(\theta)} A_2(q)dq + \int_{q_{min}}^{q^*(\theta)} \Phi(q)dq$$

$$\frac{1}{\mathcal{L}(\theta)} \int_{q^*(\theta)}^{q_{max}} \Phi(q)dq = \frac{1}{\mathcal{L}(\theta)} - \frac{1}{\mathcal{L}(\theta)} \int_{q_{min}}^{q^*(\theta)} \Phi(q)dq = \frac{1}{\mathcal{L}_L} + \sin\theta_L \int_{q_{min}}^{q^*(\theta)} A_2(q)dq$$

  We solve for A_2 by using q rather than θ as the independent variable; defining θ^* as the reciprocal function of q^*, we can rewrite the previous equation as:

$$\frac{1}{\mathcal{L}[\theta^*(q)]} \int_q^{q_{max}} \Phi(q')dq' = \frac{1}{\mathcal{L}_L} + \sin\theta_L \int_{q_{min}}^{q} A_2(q')dq'$$

  We obtain A_2 by taking the derivative of the previous equation:

$$\sin\theta_L \cdot A_2(q) = \frac{d}{dq}\frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]}$$

  where once again Λ is the CDF of Φ which is defined in 90 so that $\Lambda(q_{min}) = 1$.

In conclusion, we can write ν_∞ as a function of (Λ, θ^*):

$$\boxed{\sin\theta_L \cdot \nu_\infty = [\frac{\delta^+(q-q_{min})}{\mathcal{L}_L} + \frac{d}{dq}\frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]}]\mathcal{L}(\theta) - \frac{d\Lambda}{dq}} \tag{130}$$

In the limit $\theta_L \to \frac{\pi}{2}$ (full Northern hemisphere), the distribution reduces to:

$$\nu_\infty = \mathcal{L}(\theta)\frac{d}{dq}\frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]} - \frac{d\Lambda}{dq}$$

In that limit, no parcels can reach the Northern pole ($\theta \to \frac{\pi}{2}$), which surface is 0, which suppresses the dry spike in the distribution. Note that the previous expressions are only valid for $q \leq q^*(\theta)$:

$$\boxed{\nu_\infty[q > q^*(\theta)] = 0}$$

which means that we have completely solved the FPE in this particular case. The steady PDF in spherical coordinates $n_\infty = \lim_{t\to+\infty} n$ can be directly computed from the previous solution:

$$n_\infty(q,\theta) = \cos(\theta) \cdot \nu_\infty(q,\theta)$$

E.5 Global steady PDF

By definition, the global PDF of specific humidity is:

$$P_Q(q) \overset{\text{def}}{=} \int_0^{\theta^*(q)} n_\infty(q,\theta)d\theta = \int_0^{\theta^*(q)} \cos\theta \cdot \nu_\infty(q,\theta)d\theta$$

Using the solution computed previously for ν_∞ (130), we can evaluate this integral:

$$\boxed{\sin\theta_L \cdot g(q) = [\frac{\delta^+(q-q_{min})}{\mathcal{L}_L} + \frac{d}{dq}\frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]}]\mathcal{K}[\theta^*(q)] - \frac{d\Lambda}{dq}\sin[\theta^*(q)]}$$

where we have defined the function:

$$\mathcal{K}(\theta) \overset{\text{def}}{=} \int_0^{\theta} \cos\theta' \cdot \mathcal{L}(\theta')d\theta' = \ln(\cos\theta) + \sin(\theta) \cdot \mathcal{L}(\theta)$$

E.6 Averages

With the knowledge of ν_∞, we can compute the average of any function $f(q)$, which we define by:

$$\overline{f}(\theta) = \sin\theta_L \int_{q_{min}}^{q^*(\theta)} f(q)\nu_\infty(q,\theta)dq$$

Using the expression we computed for ν_∞ (130), we can use integration by parts to obtain $\overline{f}$ as a function of $(\Lambda, q^*(\theta), f, \frac{df}{dq})$:

$$\boxed{\overline{f}(\theta) = \{\Lambda f\}[q^*(\theta)] - \mathcal{L}(\theta)\int_{q_{min}}^{q^*(\theta)}\{\frac{\Lambda}{\mathcal{L}[\theta^*]}\frac{df}{dq}\}(q)dq - \int_{q_{min}}^{q^*(\theta)}\{f\frac{d\Lambda}{dq}\}(q)dq} \tag{131}$$

Once again, we can compute the meridional gradient using Leibniz's formula to differentiate integrals:

$$\boxed{\frac{d\overline{f}}{d\theta} = -\frac{1}{\cos\theta}\int_{q_{min}}^{q^*(\theta)} \{\frac{\Lambda}{\mathcal{L}[\theta^*]}\frac{df}{dq}\}(q)dq} \tag{132}$$

We can use the previous expressions 131 and 132 to compute the average moisture content $\overline{q}(\theta)$ and its meridional gradient:

$$\overline{q}(\theta) = q^*\Lambda(q^*) - \mathcal{L}(\theta)\int_{q_{min}}^{q^*(\theta)} \frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]}dq - \int_{q_{min}}^{q^*(\theta)} q\frac{d\Lambda}{dq}dq$$

$$\frac{d\overline{q}}{d\theta} = -\frac{1}{\cos\theta}\int_{q_{min}}^{q^*(\theta)} \frac{\Lambda(q)}{\mathcal{L}[\theta^*(q)]}dq$$

E.7 Application to an exponentially decreasing $q^*(\theta)$ and a given remoistening $\Phi(q)$

We make the following assumptions:

1. We assume that the saturation specific humidity profile exponentially decreases with latitude:
$$q^*(\theta) \approx q_{max}\exp(-\alpha\theta) \;\Leftrightarrow\; \theta^*(q) \approx -\frac{1}{\alpha}\ln(\frac{q}{q_{max}})$$

2. We assume that the "Southern boundary resetting" produces complete saturation at $\theta = 0$ (case I of complete remoistening):
$$\Phi_I(q) = \delta^-(q - q_{max})$$
or that it remoistens the parcels following a uniform distribution (case II of uniform remoistening):
$$\Phi_{II}(q) = \frac{1}{q_{max} - q_{min}}$$

In case I, the steady distribution ν_∞ can be written:

- For $\theta = 0$ (Southern edge):
$$\nu_\infty(q, 0) = \frac{\delta^-(q - q_{max})}{\sin\theta_L}$$

- For $\theta \in]0, \theta_L]$ (Interior of the domain and Northern edge):
$$\frac{\sin\theta_L}{\mathcal{L}(\theta)}\nu_\infty(q,\theta) = \frac{\delta^+(q - q_{min})}{\mathcal{L}_L} + \frac{1}{(\alpha q)\cdot[\mathcal{L}^2\cos][\theta^*(q)]}$$

Similarly, the distribution g is given by:

- For $q = q_{max}$ (Southern edge):
$$g(q_{max}) = 0$$

- For $q \in [q_{min}, q_{max}[$ (Interior of the domain and Northern edge):

$$\frac{\sin\theta_L}{\mathcal{K}[\theta^*(q)]} g(q) = \frac{\delta^+(q - q_{min})}{\mathcal{L}_L} + \frac{1}{(\alpha q) \cdot [\mathcal{L}^2 \cos][\theta^*(q)]}$$

To check the validity of our general solution in this specific case, we numerically simulate the SDE on a sphere with:

$$\begin{cases} N_{tot} = 3.10^4 \\ q_{min} = 0.1 \\ q_{max} = 1 \\ \kappa = 1 \\ a = 1 \end{cases}$$

and compare the numerical PDF to the marginal distributions $f(\theta) = \cos\theta p_\infty$ and $g(q)$ for different values of the latitudinal extent of the domain θ_L.

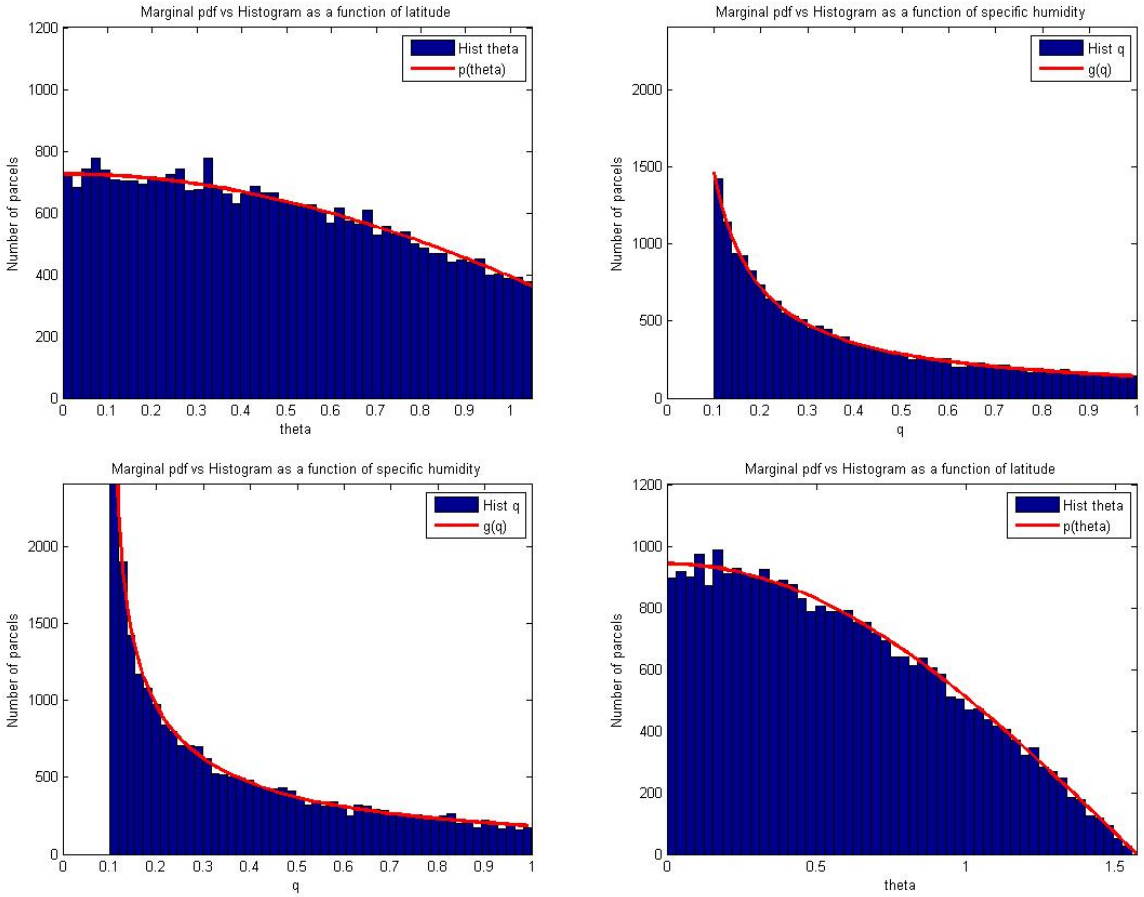

Figure 16: $\cos\theta \cdot p_\infty$ (left) and $g(q)$ (right) for $\theta_L = \frac{\pi}{3}$ (top) and $\theta_L = \frac{\pi}{2}$ (bottom) in case I of a complete remoistening

In case II, we obtain:

$$\sin\theta_L\cdot\nu_\infty = [\frac{\delta^+(q-q_{min})}{\mathcal{L}_L}+\frac{q_{max}-q}{q_{max}-q_{min}}\frac{1}{\alpha q[\mathcal{L}^2\cos][\theta^*(q)]}]\mathcal{L}(\theta)+\frac{1}{q_{max}-q_{min}}\{1-\frac{\mathcal{L}(\theta)}{\mathcal{L}[\theta^*(q)]}\}$$

$$\sin\theta_L\cdot g(q) = [\frac{\delta^+(q-q_{min})}{\mathcal{L}_L}+\frac{q_{max}-q}{q_{max}-q_{min}}\frac{1}{\alpha q[\mathcal{L}^2\cos][\theta^*(q)]}]\mathcal{K}[\theta^*(q)]+\frac{1}{q_{max}-q_{min}}\{\sin[\theta^*(q)]-\frac{\mathcal{K}[\theta^*(q)]}{\mathcal{L}[\theta^*(q)]}\}$$

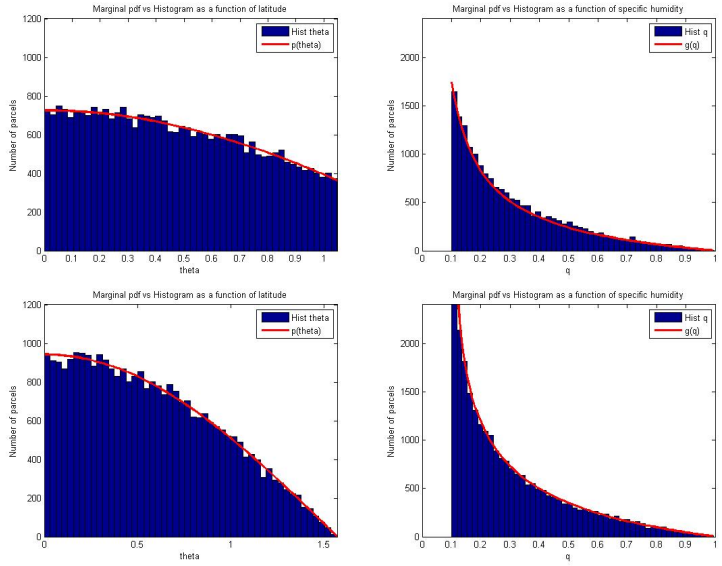

Figure 17: $\cos\theta\cdot p_\infty$ (left) and $g(q)$ (right) for $\theta_L=\frac{\pi}{3}$ (top) and $\theta_L=\frac{\pi}{2}$ (bottom) in case II of a uniform remoistening

F Incorporation of the saturated peak in the special case $\Phi_I(q)=\delta^-(q-q_{max})$

Here we will take into account the saturated peak in the initial conditions of the two-stream model 81, and prove that only the moist peak at $q=q_{max}$ is modified. The argument can be easily generalized to the Ornstein-Uhlenbeck process 4.1. Keeping the same notations, we add the dry peak and the saturated peak to the solution without applying any boundary condition:

$$\begin{cases}\mathcal{N}(q,y)=y[A_{\text{smooth}}(q)+A_{dry}\delta^+(q-q_{min})]+B(q)+\mathcal{W}(y)\delta^-(q-q*)\\ S(q,y)=(y+2)[A_{\text{smooth}}(q)+A_{dry}\delta^+(q-q_{min})]+B(q)\end{cases}$$

The normalized boundary condition at $y=0$ (87) gives:

$$\Phi(q)=2LN(q,0)=2LB(q)+2L\mathcal{W}(0)\Phi(q)\ \Rightarrow\ 2LB(q)=[1-2L\mathcal{W}(0)]\Phi(q)$$

where we have used the fact that $\Phi(q)=\delta^-(q-q_{max})$ and $q^*(0)=q_{max}$. The dry peak intensity doesn't change as it is determined by the normalization condition for the southwards

PDF S 82 at $y = L$ where $q = q^*(L) = q_{min}$:

$$2LA_{\text{dry}} = \frac{1}{2+L}$$

Again, the southwards PDF S satisfies the normalization condition 82 for all $y \in [0, L]$, and does not have any saturated peak:

$$1 = 2L\int_{q_{min}}^{q^*(y)} S(q,y)dq = [y+2][\frac{1}{2+L} + 2L\int_{q_{min}}^{q^*(y)} A_{\text{smooth}}(q)dq] + [1-2L\mathcal{W}(0)]\int_{q_{min}}^{q^*(y)} \Phi(q)dq$$

$$2L\int_{q_{min}}^{q^*(y)} A_{\text{smooth}}(q)dq = \frac{1}{y+2}[\int_{q^*(y)}^{q_{max}} \Phi(q)dq + 2L\mathcal{W}(0)\int_{q_{min}}^{q^*(y)} \Phi(q)dq] - \frac{1}{2+L}$$

Switching the independent variable from y to q , changing the bounds of the integrals carefully according to figure 2, and differentiating the previous equation with respect to q, we obtain:

$$2LA_{\text{smooth}} = \frac{d}{dq}\frac{1 + 2L\mathcal{W}(0)\delta_{qq_{max}}}{2 + y^*(q)}$$

where we have used the fact that:

$$\begin{cases} \int_q^{q_{max}} \Phi(q')dq' = \int_q^{q_{max}} \delta^-(q' - q_{max})dq' = 1 \\ \int_{q_{min}}^{q} \Phi(q')dq' = \int_{q_{min}}^{q} \delta^-(q' - q_{max})dq' = \delta_{qq_{max}} \end{cases}$$

where $\delta_{qq_{max}}$ is the Kronecker symbol for $q = q_{max}$. Finally, we find the weight of the saturated parcels $\mathcal{W}$ by applying the normalization condition 82 for $\mathcal{N}$:

$$1 = 2L\int_{q_{min}}^{q^*(y)} \mathcal{N}(q,y)dy$$

$$1 = \frac{y[1 + 2L\mathcal{W}(0)\delta_{q^*q_{max}}]}{2+y} + 2L\mathcal{W}(y)$$

$$2L\mathcal{W}(y) = \frac{2}{2+y}$$

leading to a PDF that only comprises a dry and a saturated peak in this special case:

$$\boxed{2LS(q,y) = (2+y)[\frac{\delta^+(q - q_{min})}{2+L} + \frac{d}{dq}\frac{1 + \delta_{qq_{max}}}{2 + y^*(q)}]}$$

$$2LS(q,y) = (2+y)[\frac{\delta^+(q - q_{min})}{2+L} + \frac{\delta^-(q - q_{max})}{2 + y^*(q)} - \frac{dy^*}{dq}\frac{1 + \delta_{qq_{max}}}{(2+y^*)^2}]$$

$$\boxed{2L\mathcal{N}(q,y) = y[\frac{\delta^+(q - q_{min})}{2+L} + \frac{d}{dq}\frac{1 + \delta_{qq_{max}}}{2 + y^*(q)}] + \frac{2\delta^-(q - q^*)}{2+y}}$$

$$2L\mathcal{N}(q,y) = y[\frac{\delta^+(q - q_{min})}{2+L} + + \frac{\delta^-(q - q_{max})}{2 + y^*(q)} - \frac{dy^*}{dq}\frac{1 + \delta_{qq_{max}}}{(2+y^*)^2}] + \frac{2\delta^-(q - q^*)}{2+y}$$

In conclusion, the distributions agree with the ones computed in 91 and 89 everywhere except at $q = q_{max}$, which means that:

- To compute the analytical solution, we can safely apply the boundary condition on N directly without having to worry about the saturated peak.
- The limit $\Phi(q) \to \delta^-(q - q_{max})$ is singular at $q = q_{max}$, where a moist peak $\delta^-(q - q_{max})$ linearly increasing with y appears.

References

[1] Boer, G. J., & Shepherd, T. G. (1983). *Large-Scale Two-Dimensional Turbulence in the Atmosphere.* Journal of the Atmospheric Sciences. doi:10.1175/1520-0469(1983)040<0164:LSTDTI>2.0.CO;2

[2] Brogniez, H., Roca, R., & Picon, L. (2009). *A study of the free tropospheric humidity interannual variability using meteosat data and an advection-condensation transport model.* Journal of Climate, 22(24), 6773–6787. doi:10.1175/2009JCLI2963.1

[3] Brown, R. G., & Zhang, C. (1997). *Variability of Midtropospheric Moisture and Its Effect on Cloud-Top Height Distribution during TOGA COARE*.* Journal of the Atmospheric Sciences, 54(23), 2760–2774. doi:10.1175/1520-0469(1997)054<2760:VOMMAI>2.0.CO;2

[4] Couhert, A., Schneider, T., Li, J., Waliser, D. E., & Tompkins, A. M. (2010). *The maintenance of the relative humidity of the subtropical free troposphere.* Journal of Climate, 23(2), 390–403. doi:10.1175/2009JCLI2952.1

[5] Dessler, A. E., & Sherwood, S. C. (2000). *Simulations of tropical upper tropospheric humidity.* Journal of Geophysical Research, 105(D15), 20155. doi:10.1029/2000JD900231

[6] K. A. Emanuel, (1994). Atmospheric Convection. *Oxford University Press.*

[7] Emanuel, K. A., & Pierrehumbert, R. T. (1995). *Microphysical and dynamical control of tropospheric water vapor.* NATO ASI Series I Global Enivironmental Change, 35, 17–28. doi:10.1007/978-3-642-61051-6_2

[8] Fick, A. (1851). *Uber diffusion.* Annalen Der Physik. doi:10.1002/andp.18551700105

[9] Flierl, G. R., & McGillicuddy, D. J. (2002). *Mesoscale and Submesoscale Physical – Biological Interactions.* The Sea (Vol. 12).

[10] Hurley, J. V., Galewsky, J., Worden, J., & Noone, D. (2012). *A test of the advection-condensation model for subtropical water vapor using stable isotopologue observations from Mauna Loa Observatory, Hawaii.* Journal of Geophysical Research: Atmospheres, 117(19),15. doi:10.1029/2012JD018029

[11] Mote, P. W., & Frey, R. (2006). *Variability of clouds and water vapor in low latitudes: View from Moderate Resolution Imaging Spectroradiometer (MODIS).* Journal of Geophysical Research: Atmospheres, 111(16), 1–14. doi:10.1029/2005JD006791

[12] O'Gorman, P. A., & Schneider, T. (2006). *Stochastic Models for the Kinematics of Moisture Transport and Condensation in Homogeneous Turbulent Flows.* Journal of the Atmospheric Sciences, 63(11), 2992–3005. doi:10.1175/JAS3794.1

[13] O'Gorman, P. A., Schneider, T., Lamquin, N., & Singh, M. (2011). *The Relative Humidity in an Isentropic Advection – Condensation Model : Limited Poleward Influence and Properties of Subtropical Minima.* Journal of the Atmospheric Sciences, 3079–3093. doi:10.1175/JAS-D-11-067.1

[14] Pasquero, C., Provenzale, A., & Babiano, A. (2001). *Parameterization of dispersion in two-dimensional turbulence.* Journal of Fluid Mechanics, 439, 279303. doi:10.1017/S0022112001004499

[15] Pierrehumbert, R. T., Brogniez, H., & Roca, R. (2007). *On the relative humidity of the atmosphere.* The Global Circulation of the Atmosphere, 143–185.

[16] Pierrehumbert, R. T., & Yang, H. (1993). *Global chaotic mixing on isentropic surfaces.* Journal of the Atmospheric Sciences. doi:10.1175/1520-0469(1993)050<2462:GCMOIS>2.0.CO;2

[17] Ryoo, J. M., Igusa, T., & Waugh, D. W. (2009). *PDFs of tropical tropospheric humidity: Measurements and theory.* Journal of Climate, 22(12), 3357–3373. doi:10.1175/2008JCLI2747.1

[18] Sawford, B. (2001). *Turbulent relative dispersion.* Annual Review of Fluid Mechanics, 33, 289–317. doi:10.1146/annurev.fluid.33.1.289

[19] Schneider, T., O'Gorman, P. A., & Levine, X. (2010). *Water vapor and the dynamics of climate changes.* Reviews of Geophysics, (48), 1–22. doi:10.1029/2009RG000302.1.INTRODUCTION

[20] Sherwood, S. C., Kursinski, E. R., & Read, W. G. (2006). *A distribution law for free-tropospheric relative humidity.* Journal of Climate, 19(24), 6267–6277. doi:10.1175/JCLI3978.1

[21] Sherwood, S. C., Roca, R., Weckwerth, T. M., & Andronova, N. G. (2010). *Tropospheric Water Vapor , Convection , and Climate.* Reviews of Geophysics, 48(2009), 1–29. doi:10.1029/2009RG000301.1.INTRODUCTION

[22] Sukhatme, J. (2004). *Lagrangian velocity correlations and absolute dispersion in the midlatitude troposphere.* Journal of the Atmospheric Sciences, 62, 3831–3836. doi:10.1175/JAS3560.1

[23] Sukhatme, J., & Young, W. R. (2011). *The advection-condensation model and water-vapour probability density functions.* Quarterly Journal of the Royal Meteorological Society, 137(659), 1561–1572. doi:10.1002/qj.869

[24] Taylor, G. I. (1921). *Diffusion by continuous movements.* Proc. Lond. Math. Soc. doi:10.1112/plms/s2-20.1.196

[25] Tompkins, A. M. (2001). *Organization of tropical convection in low vertical wind shears: The role of cold pools.* Journal of the Atmospheric Sciences, 58(13), doi:10.1175/1520-0469(2001)058<1650:OOTCIL>2.0.CO;2

[26] Thomson, D. (1987). *Criteria for the selection of stochastic models of particle trajectories in turbulent flows.* Journal of Fluid Mechanics, 180, 529=556. doi:10.1017/S0022112087001940

[27] Warhaft, Z. (2000). *Passive scalars in turbulent flows.* Annual Review of Fluid Mechanics, 32, 203–240. doi:10.1146/annurev.fluid.32.1.203

[28] Wing, A. A, & Emanuel, K. A. (2014). *Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations.* Journal of Advances in Modeling Earth Systems, 5(November), doi:10.1002/2013MS000269

[29] Uhlenbeck, G. E., & Ornstein, L. S. (1930). *On the theory of the Brownian motion.* Physical Review Letters, 823–841. doi:http://dx.doi.org/10.1103/PhysRev.36.823

Coupled Reduced Equations for Strongly Stratified Flows

Cesar B. Rocha

October 15, 2015

Abstract

We present a set of reduced equations in the limit of strong stratification. The asymptotics lead to the hydrostatic primitive equations for the slow/large scale flow and non-hydrostatic, quasi-linear equations for the perturbations. There is no closure problem and the system conserves energy. We explore the properties of this coupled system of equations by studying solutions of a two-dimensional toy problem. This simple problem displays interesting dynamics with $\mathcal{O}(1)$ feedbacks between mean and perturbations. Even in this toy problem, the Reynolds stresses and buoyancy fluxes are not sign-definite in most of the regions of the parameter space.

1 Introduction

Strongly stratified turbulence has been used as a paradigm to interpret observations of strongly stratified turbulent geophysical flows [7]. Turbulence dominated by strong stable stratification occurs at scales larger than the Ozmidov scale, l_O,

$$l_O \stackrel{\text{def}}{=} \left(\frac{\epsilon}{N^3}\right)^{1/2}, \tag{1}$$

where ϵ is the kinetic energy rate of dissipation, and N is the buoyancy frequency. The Ozmidov scale is the horizontal scale of the largest overturns in a stratified flow. At scales much smaller than l_O, the flow is essentially unaffected by stratification, whereas at scales larger than l_O the effects of stratification dominates.

In the ocean l_O is $\mathcal{O}(1)$ m and in the atmosphere it is $\mathcal{O}(100)$ km [1, 3, 7]. At $\mathcal{O}(l_O)$ scales, the Rossby number of the flow is large and rotation is unimportant. The most prominent coherent structures that emerge in strongly stratified turbulence are strongly anisotropic. Indeed, stratified turbulence is often termed "pancake" turbulence [2, 7].

In the ocean, stratified turbulence is important at scales between 100 and 1 m and in the atmosphere at scales between 10 km and 100 m [7]. Kinetic and potential energy horizontal wavenumber spectra calculated from observations in the atmosphere and in the ocean are consistent with predictions of stably stratified turbulence[1]. These scales are barely resolved in regional models and not resolved at all in general circulation models. Understanding of the dynamics of strongly stratified turbulence, and in particular how the small scale stratified turbulence interacts with large scale flows, is important for parameterizing these effects into numerical models.

[1] The spectra are also consistent with predictions of other theories, e.g. [4].

One approach towards understanding and modeling strongly stratified turbulence has been direct numerical simulations [1, 3, 7]. Figure 1 shows different regimes of stratified flows in the $\mathrm{Re} - 1/\mathrm{Fr}$ space, where (for details and definitions see section 2).

$$\mathrm{Fr} \stackrel{\mathrm{def}}{=} \frac{\mathcal{U}}{N\mathcal{L}}\,, \tag{2}$$

is the horizontal Froude number (and $1/\mathrm{Fr}^2$ is a measure of the strength of the stratification), and

$$\mathrm{Re} \stackrel{\mathrm{def}}{=} \frac{\mathcal{U}\,\mathcal{L}}{\nu}\,, \tag{3}$$

is the Reynolds number. Dimensional analysis suggests that strongly stratified turbulence regime develops when $\mathrm{ReFr}^2 > 1$ [3].

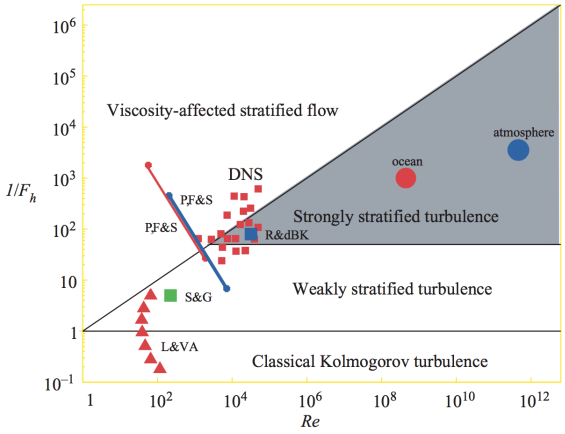

Figure 1: The different regimes of stratified flows as a function of the Reynolds number and the horizontal Froude number. Source: [3].

Direct numerical simulations with parameters of relevance to geophysical flows are currently unattainable through direct numerical simulations [1, 3]. Indeed, figure 1 shows that current direct numerical simulations are decades away in parameter space from the regimes of relevance to geophysical flows. Conclusions and extrapolations based on current low ReFr^2 simulations should be interpreted with care [1].

Here we take a different approach to study this problem. Inspired by the success of multi-scale asymptotics to study turbulent flows with strong constraints [6], e.g rotating convection [8], we dig deep into the scaling for strongly stratified flows and present a set of reduced equations for its modeling (Section 2). To explore the main properties of this system of equations and showcase its usefulness we study a toy initial value problem of stratified shear instability (Section 3). A summary and inquiries for future research are given in section 4.

2 Reduced Equations for Strongly Stratified Flows

Our starting point is the familiar Boussinesq equations. The momentum and mass conservation equations are reduced to

$$\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u} = -\boldsymbol{\nabla} p + \nu\triangle\boldsymbol{u} - b\,\hat{\mathbf{k}}\,, \tag{4}$$

and

$$\boldsymbol{\nabla}\cdot\boldsymbol{u} = 0\,, \tag{5}$$

where $b \stackrel{\text{def}}{=} -\frac{g}{\rho_0}\rho$ is the negative of the reduced gravity, commonly referred to buoyancy in the oceanographic community (e.g., [9]), and $\hat{k}$ is the unit vector in the vertical direction. The system is closed with the thermodynamic equation

$$\partial_t\, b + \boldsymbol{u}\cdot\boldsymbol{\nabla}\, b = \kappa\triangle\, b - N^2(z)\, w\,, \tag{6}$$

where $N^2(z) = \frac{\partial B}{\partial z}$ is the background buoyancy frequency.

We start by nondimensionalizing the Boussinesq equations (4)-(6). Let

$$(x,y) = (x^\star, y^\star)\,\mathcal{L}\,, \qquad z = z^\star\,\mathcal{H}\,, \qquad (u,v) = (u^\star, v^\star)\,\mathcal{U}\,, \qquad t = t^\star\,\frac{\mathcal{L}}{\mathcal{U}}\,, \qquad p = p^\star\,\mathcal{U}^2\,. \tag{7}$$

From the vertical momentum equation we find the scale for the buoyancy is

$$b = b^\star\,\frac{\mathcal{U}^2}{\mathcal{H}}\,. \tag{8}$$

To find a scale for the vertical velocity, we insist that the horizontal advection of buoyancy balances, to lowest order, the vertical advection of background buoyancy in (6). Hence

$$w = w^\star\,\frac{\mathrm{Fr}^2}{\alpha}\mathcal{U}\,, \tag{9}$$

where the horizontal Froude number is

$$\mathrm{Fr} \stackrel{\text{def}}{=} \frac{\mathcal{U}}{N\mathcal{L}}\,, \tag{10}$$

and the aspect ratio of the flow is

$$\alpha \stackrel{\text{def}}{=} \frac{\mathcal{H}}{\mathcal{L}}\,. \tag{11}$$

Using the scaling (7) through (9) in the Boussinesq equations (4) through (6), we obtain

$$\partial_{t^\star}\boldsymbol{u}_h^\star + \boldsymbol{u}_h^\star\cdot\boldsymbol{\nabla}^\star\boldsymbol{u}_h^\star + \frac{\mathrm{Fr}^2}{\alpha^2}w^\star\partial_{z^\star}\boldsymbol{u}_h^\star = -\boldsymbol{\nabla}_h^\star p^\star + \frac{1}{\mathrm{Re}}\left[\triangle_h^\star\boldsymbol{u}_h^\star + \frac{1}{\alpha^2}\partial_{z^\star}^2\boldsymbol{u}_h^\star\right]\,, \tag{12}$$

$$\mathrm{Fr}^2\left[\partial_{t^\star}w_h^\star + \boldsymbol{u}_h^\star\cdot\boldsymbol{\nabla}^\star w_h^\star + \frac{\mathrm{Fr}^2}{\alpha^2}\partial_{z^\star}w^\star\right] = -\partial_{z^\star}p^\star + b^\star + \frac{\mathrm{Fr}^2}{\mathrm{Re}}\left[\triangle_h^\star w^\star + \frac{1}{\alpha^2}\partial_{z^\star}^2 w^\star\right]\,, \tag{13}$$

$$\boldsymbol{\nabla}^\star\cdot\boldsymbol{u}_h^\star + \frac{\mathrm{Fr}^2}{\alpha^2}\partial_{z^\star}w^\star = 0\,, \tag{14}$$

and

$$\partial_{t^\star} b^\star + \boldsymbol{u}_h^\star \cdot \boldsymbol{\nabla}^\star b^\star + \frac{\mathrm{Fr}^2}{\alpha^2} w^\star \partial_{z^\star} b^\star = w + \frac{1}{\mathrm{Re}\ \mathrm{Pr}} \left[\triangle_h^\star b^\star + \frac{1}{\alpha^2} \partial_{z^\star}^2 b^\star \right] , \tag{15}$$

where the Reynolds number is

$$\mathrm{Re} \stackrel{\mathrm{def}}{=} \frac{\mathcal{U}\,\mathcal{L}}{\nu} , \tag{16}$$

and the Prandtl number is

$$\mathrm{Pr} \stackrel{\mathrm{def}}{=} \frac{\kappa}{\nu} . \tag{17}$$

From now on, we drop the star super scripts — all variables are nondimensional unless otherwise stated.

Note that a full quasi-two-dimensionalization of the system (to lowest order in Fr) depends on the ratio $\frac{\mathrm{Fr}^2}{\alpha^2}$ [2]. If that ratio goes to zero as $\mathrm{Fr} \to 0$, then the system consists of two-dimensional layers, with properties similar to two-dimensional turbulence. Alternatively, a more interesting case is when the aspect ratio α adjust so that $\frac{\mathrm{Fr}}{\alpha} = \mathcal{O}(1)$ (see appendix A for details). This implies that the vertical scale of the motion is

$$\mathcal{H} \sim \frac{\mathcal{U}}{N} . \tag{18}$$

Notice that with this scaling, the potential energy of the flow has the same order of the kinetic energy

$$\frac{b^2}{N^2} = \frac{\mathrm{Fr}^2}{\alpha^2} \mathcal{U}^2 \sim \mathcal{O}(\mathcal{U}^2) , \tag{19}$$

and therefore there is an approximate equipartition between kinetic energy and potential energy.

We assume that the system evolves in two different sets of scales. That is, all fields depend on fast/short scales (χ_h, τ) and slow/long $(\mathbf{x}_h, t)$. A single vertical scale is assumed. The fast time variable is

$$\tau \stackrel{\mathrm{def}}{=} \mathrm{Fr}^{-1}\, t , \qquad \text{so that} \qquad \partial_t \to \partial_t + \mathrm{Fr}^{-1} \partial_\tau . \tag{20}$$

Similarly, the horizontal short spatial variable is

$$\boldsymbol{\chi} \stackrel{\mathrm{def}}{=} \mathrm{Fr}^{-1}\, \boldsymbol{x} , \qquad \text{so that} \qquad \boldsymbol{\nabla}_h^{\boldsymbol{x}} \to \boldsymbol{\nabla}_h^{\boldsymbol{x}} + \mathrm{Fr}^{-1} \boldsymbol{\nabla}_h^{\boldsymbol{\chi}} . \tag{21}$$

Any field f depends on both fast/short and slow/long scales, and is decomposed into

$$f = \bar{f} + \mathrm{Fr}^n f' + o(\mathrm{Fr}^n) , \tag{22}$$

where the overbar denotes the average over the slow/short scales (e.g. [6]):

$$\bar{f}(\boldsymbol{x}, t, z) \stackrel{\mathrm{def}}{=} \lim_{S,T \to \infty} \frac{1}{S\,T} \iint_{\mathcal{A}} \int_T f(\boldsymbol{x}, t, \tau, \boldsymbol{\chi}, z) \mathrm{d}\tau \mathrm{d}\vec{\chi} . \tag{23}$$

Note that, by definition, $\bar{f}' = 0$. The correct power n depends on the dynamical field, and is chosen in order to obtain a nontrivial balance. We have

$$\begin{aligned} \boldsymbol{u}_h &= \bar{\boldsymbol{u}}_h + \mathrm{Fr}^{1/2}\, \boldsymbol{u}'_h + \mathcal{O}(\mathrm{Fr}^{3/2})\,, \\ p &= \bar{p} + \mathrm{Fr}^{1/2}\, p' + \mathcal{O}(\mathrm{Fr}^{3/2})\,, \\ b &= \bar{b} + \mathrm{Fr}^{1/2}\, b' + \mathcal{O}(\mathrm{Fr}^{3/2})\,, \\ w &= \bar{w} + \mathrm{Fr}^{-1/2}\, w' + \mathcal{O}(\mathrm{Fr}^{0})\,. \end{aligned} \tag{24}$$

Note that, in dimensional form, the fluctuations are isotropic, i. e., $(\boldsymbol{u}'_h, w') \sim \mathrm{Fr}^{1/2}\,\mathcal{U}$.

Reduced equations for $\mathrm{Fr} \ll 1$ and $\mathrm{Re} \gg 1$

We now introduce the rescaled operators (20) and (21) into the nondimensional equations (12) through (15), and average over the fast/short scales to obtain

$$\partial_t \bar{\boldsymbol{u}}_h + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{x}}_h \bar{\boldsymbol{u}}_h + \bar{w}\partial_z \bar{\boldsymbol{u}}_h = -\boldsymbol{\nabla}^{\boldsymbol{x}}_h \bar{p} + \frac{1}{\mathrm{Re}_b}\partial_z^2 \bar{\boldsymbol{u}}_h - \partial_z(\overline{w'\boldsymbol{u}'_h})\,, \tag{25}$$

$$0 = -\partial_z \overline{p} + \overline{b}\,, \tag{26}$$

$$\boldsymbol{\nabla}^{\boldsymbol{x}}_h \cdot \bar{\boldsymbol{u}}_h + \partial_z \overline{w} = 0\,, \tag{27}$$

and

$$\partial_t \bar{b} + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{x}}_h \bar{b} + \bar{w}\partial_z \bar{b} = -\bar{w} + \frac{1}{\mathrm{Re}_b\,\mathrm{Pr}}\partial_z^2 \bar{b} - \partial_z(\overline{w'b'})\,, \tag{28}$$

where the buoyancy Reynold number is $\mathrm{Re}_b \stackrel{\mathrm{def}}{=} \mathrm{Fr}^2\,\mathrm{Re}$.

The equations for the perturbations are obtained by subtracting the mean equations above from Boussinesq equations (with $\mathrm{Fr} \ll 1$ and $\mathrm{Re} \gg 1$). We obtain, to lowest order,

$$\partial_\tau \boldsymbol{u}'_h + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{\chi}}_h \boldsymbol{u}'_h + w'\partial_z \bar{\boldsymbol{u}}_h = -\boldsymbol{\nabla}^{\boldsymbol{\chi}}_h p' + \frac{\mathrm{Fr}}{\mathrm{Re}_b}\left(\triangle^{\boldsymbol{\chi}}_h + \partial_z^2\right)\boldsymbol{u}'_h\,, \tag{29}$$

$$\partial_\tau w' + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{\chi}}_h w' = -\partial_z p' + b' + \frac{\mathrm{Fr}}{\mathrm{Re}_b}\left(\triangle^{\boldsymbol{\chi}}_h + \partial_z^2\right) w'\,, \tag{30}$$

$$\partial_\tau b' + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{\chi}}_h b' + w'(\partial_z \bar{b} + 1) = \frac{\mathrm{Fr}}{\mathrm{Re}_b\,\mathrm{Pr}}\left(\triangle^{\boldsymbol{\chi}}_h + \partial_z^2\right) b'\,, \tag{31}$$

$$\boldsymbol{\nabla}^{\boldsymbol{\chi}}_h \cdot \boldsymbol{u}'_h + \partial_z w' = 0\,. \tag{32}$$

In some calculations, it is convenient to rewrite the equations for the perturbations (29) through (32) in terms of the slow/small scale independent variables. For future reference, these equations are:

$$\mathrm{Fr}\,\left(\partial_t \boldsymbol{u}'_h + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{x}}_h \boldsymbol{u}'_h\right) + w'\partial_z \bar{\boldsymbol{u}}_h = -\mathrm{Fr}\boldsymbol{\nabla}^{\boldsymbol{x}}_h p' + \frac{\mathrm{Fr}}{\mathrm{Re}_b}\left(\mathrm{Fr}^2 \triangle^x_h + \partial_z^2\right)\boldsymbol{u}'_h\,, \tag{33}$$

$$\mathrm{Fr}\,\left(\partial_t w' + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}^{\boldsymbol{x}}_h w'\right) = -\partial_z p' + b' + \frac{\mathrm{Fr}}{\mathrm{Re}_b}\left(\mathrm{Fr}^2\,\triangle^{\boldsymbol{x}}_h + \partial_z^2\right) w'\,, \tag{34}$$

$$\mathrm{Fr}\,\left(\partial_t b' + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}_h^{\boldsymbol{x}} b'\right) + w'(\partial_z \bar{b} + 1) = \frac{\mathrm{Fr}}{\mathrm{Re}_b\,\mathrm{Pr}}\left(\mathrm{Fr}^2\,\triangle_h^{\boldsymbol{x}} + \partial_z^2\right) b'\,, \tag{35}$$

$$\mathrm{Fr}\,\boldsymbol{\nabla}_h^{\boldsymbol{x}} \cdot \boldsymbol{u}_h' + \partial_z w' = 0\,. \tag{36}$$

Alternatively, the quasi-linear system above can be expressed in terms of a single variable w'

$$\left[\mathsf{L}^\chi - \frac{\mathrm{Fr}}{\mathrm{Re}_b\,\mathrm{Pr}}\left(\triangle_h^\chi + \partial_z^2\right)\right]\left[\mathsf{L}^\chi\left(\triangle_h^\chi + \partial_z^2\right) - \partial_z^2 \bar{\boldsymbol{u}}_h \cdot \nabla_h^\chi\right] w' + (\partial_z \bar{b} + 1)\triangle_h^\chi w'$$
$$= \frac{\mathrm{Fr}}{\mathrm{Re}_b}\left[\mathsf{L}^\chi - \frac{\mathrm{Fr}}{\mathrm{Re}_b\,\mathrm{Pr}}\left(\triangle_h^\chi + \partial_z^2\right)\right]\left(\triangle_h^\chi + \partial_z^2\right)^2 w'\,, \tag{37}$$

where the quasi-linear operator is

$$\mathsf{L}^\chi \stackrel{\text{def}}{=} \partial_\tau + \bar{\boldsymbol{u}}_h \cdot \boldsymbol{\nabla}_h^\chi\,. \tag{38}$$

The two-dimensional (x, z) version of (37) reduces to the Taylor-Goldstein equation in the inviscid limit (see appendix B).

Conservation of energy

The coupled system (25)-(28) and (29)-(31) conserves total energy in the inviscid limit ($\mathrm{Re} \to \infty$)

$$\frac{\mathrm{d}E}{\mathrm{d}t} = 0\,, \tag{39}$$

where the nondimensional total energy is

$$E = \frac{1}{2}\iiint\left[\left(|\bar{\boldsymbol{u}}_h|^2 + \bar{b}^2\right) + \mathrm{Fr}\left(\overline{|\boldsymbol{u}_h'|^2} + \overline{w'^2} + \overline{b'^2}\right) + \mathcal{O}(\mathrm{Fr}^2)\right]\mathrm{d}V\,, \tag{40}$$

and we assumed harmless boundary conditions such as no-flux or triple periodicity (see appendix C for details). The dimensional potential energy has the form b^2/N^2. To lowest order the kinetic energy is due to the horizontal flow. Consistent with our approximation, the system is isotropic at $\mathcal{O}(\mathrm{Fr})$.

3 Decaying Problems

To begin exploring the reduced system, we study solutions of a very simplified problem. We consider two-dimensional system $(x - z)$ with solutions of the slow field independent of the slow horizontal coordinate (x). In other words, the "pancakes" are infinitely long. A different interpretation is that we are looking for the zeroth mode of the slow field, or the x-average. Thus, the slow/long equations reduce to

$$\partial_t \bar{\boldsymbol{u}}_h = \frac{1}{\mathrm{Re}_b}\partial_z^2 \bar{\boldsymbol{u}}_h - \partial_z \overline{w'\boldsymbol{u}_h'}\,, \tag{41}$$

$$\partial_t \bar{b} = \frac{1}{\mathrm{Re}_b\,\mathrm{Pr}}\partial_z^2 \bar{b} - \partial_z \overline{w'b'}\,. \tag{42}$$

In the absence of perturbations, these equations simply reduce to the diffusion equation.

The initial conditions for this toy problem consists of two shear layers, with a strong stratification across the shear regions, and very low stratification between the layers. The two shear-layer set up was chosen for numerical convenience owing to its periodicity in z. The functional forms are

$$\bar{u}(z) = \tanh\left[m_0\left(z - \tfrac{\pi}{2}\right)\right] - \tanh\left[m_0\left(z - \tfrac{3\pi}{2}\right)\right] - 1\,, \tag{43}$$

and

$$\bar{b}(z) = A_0\left(\tanh[m_0(z - \tfrac{3\pi}{2})] - \tanh[m_0(z - \tfrac{\pi}{2}])\right(z - \pi). \tag{44}$$

Notice that the total background stratification is $B(z) = z + \bar{b}$. In the following experiments, we fix the mean shear m_0 and vary the distance between the two shear layers (Figure 2). We also fix $\mathrm{Fr} = 0.02$ and $\mathrm{Pr} = 1$.

The slow equations (41) and (42) together with the fast equations are solved (29) through (36) are sovled numerically with a standard Fourier spectral method. The whole system is solved in a single spatial scale and a single time scale. To compute the fluxes whose divergences force the slow equations, we only average the perturbations in space; solutions are not very sensitive to average both in space and time (see appendix D). The system is marched forward using a fourth order implicit-explicit Runge-Kutta time stepper.

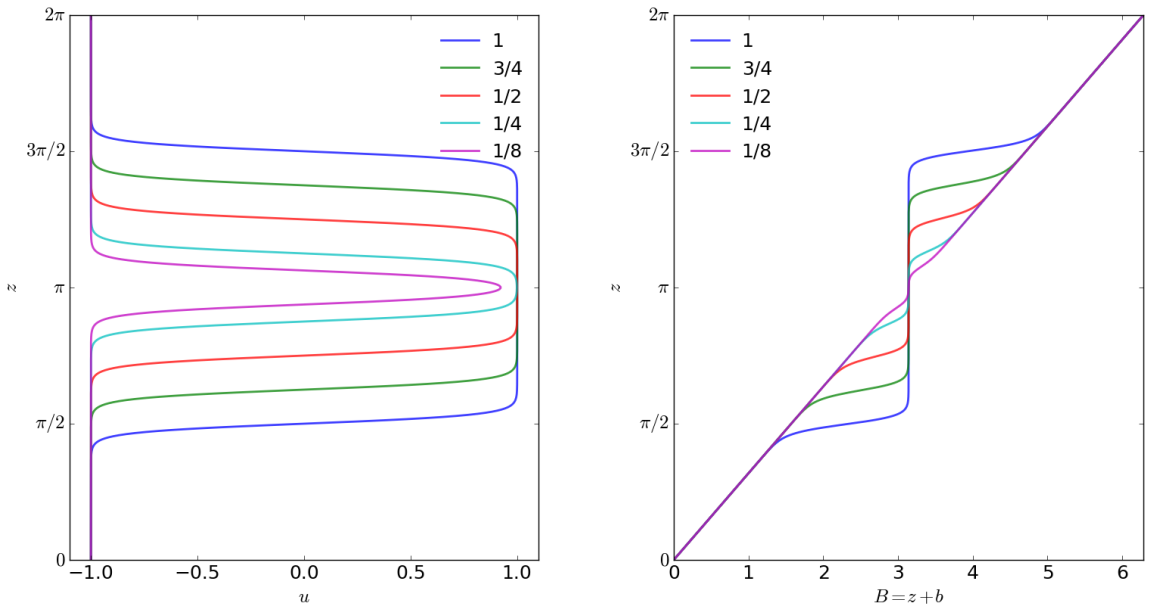

Figure 2: The initial condition for the toy decaying problem. We use various values of h, the distance between the two shear layers. (Left) horizontal velocity and (right) buoyancy.

3.1 Linear stability analysis

Using the stationary mean velocity (43) and stratification (44), we perform a linear stability analysis of the quasilinear equations for the perturbations. Figure 3 shows the growth rates for the base state profiles with $h = 1$ through $h = 1/8$ and various Re_b. The growth rates increase with Re_b. For large h, where the two shear layers are virtually independent the unstable modes span a wider region of the horizontal wavenumber space, whereas for small h the unstable modes are confined to wavenumbers $k < 10$. The growth rate of the most unstable mode increase with decreasing h. There is no dramatic transition of the stability properties as a function of Re_b, but the growth rates increase significantly for $\mathrm{Re}_b > 10$, particularly at small h.

These modes are stratified shear instablities. Figure 4 depicts the wavestructure for the most unstable modes in two cases: $h = 1$ and $h = 8$. The structure of the vertical velocity clearly show the classic "tilt against the shear". Also clearly depicted is the fact that with $h = 1$ the shear layers are independent whereas with $h = 1/8$ they are coupled. These modes are not classic Kelvin-Helmholtz instablities, e.g., Drazin & Reid [5]. In particular, the vorticity structure of these instabilities are composed of opposite sign votices with a phase shift.

3.2 Initial value problems

We now consider simulations of initial value problems. In particular, we discuss the structure of different solutions with varying h and Re_b. All simulations are initialized with the initial conditions (Figure 2). The perturbations are initialized with a small seed (10^{-6}) random field.

Figure 5 shows the time series for the evolution of the kinetic energy (KE) and potential energy (PE) of both slow/long fields and perturbations for an experiment with $\mathrm{Re}_b = 10$ and $h = 1$. With these parameters, the mean flow is signigicatlly damped by viscosity and diffusivity. Nevertheless, even at low Re_b, there is an interesting interaction between the mean (slow/long) fields and the perturbations. Once the pertubartions pick the right phase, they grow exponentially, with a rate consistent with the prediction of the linear stability analysis (see dashed green line in figure 5). The perturbations peak at about $t = 5$, and rapidly decay. This rapid decay is not accounted for by the viscous terms. Indeed, the perturbations accelerate the flow in the initial phase of their decay (see snapshots of slow velocity in the bottom panel of figure 5). For times larger than about 10, the perturbations have decayed significantly, so that the mean flow diffusively tends towards a state of rest with linear stratification.

Increasing Re_b changed qualitatively and quantitatively that picture. For instance, figure 11 shows the time series of the different components of energy for an experiment with $\mathrm{Re}_b = 200$ and $h = 1$. As in the $\mathrm{Re}_b = 10$ case, the perturbations growth eponentially after picking up the right phase, and the growth rate is consistent with the predictions for the most unstable mode. However, at about $t = 4$, the perturbations saturate. Because there are no non-linearities in the equatins for the perturbations, this saturation is due to a change in the slow horizontal velocity and buoyancy profiles. The perturbations then plateau for about 5 time units before they start decay. The initial decay is relatively fast, and it is not accounted for by viscosity. Indeed, as the perturbations decay, they accelerate the slow

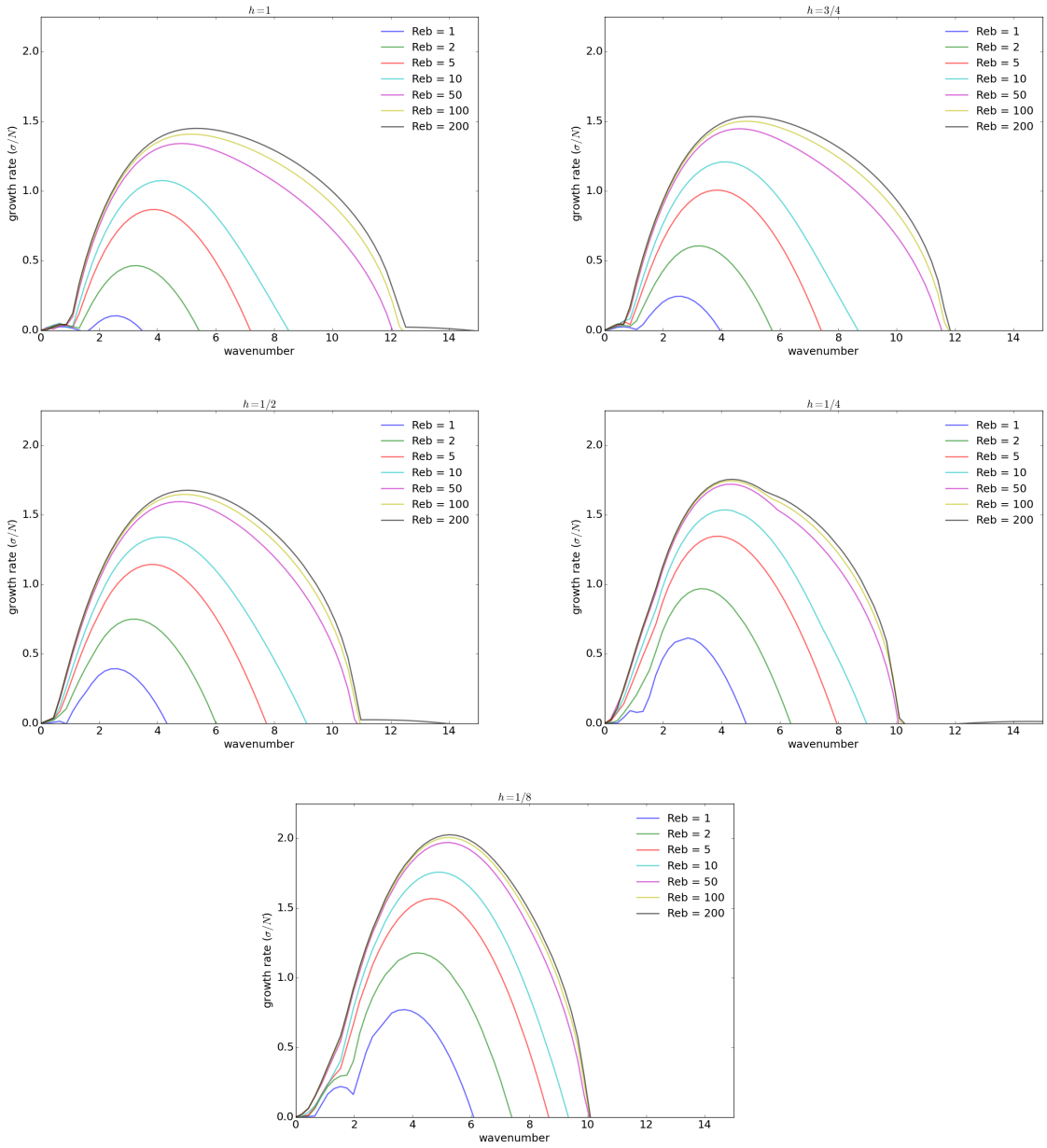

Figure 3: Growth rates for linear unstable modes for initial conditions with $h = 1$, $h = 3/4$, $h = 1/2$, $h = 1/4$, and $h = 1/8$.

flow (see lower panel of figure 11), significantly changing its shape. At about $t = 60$, the perturbations start decaying very slowly, consistent with the visous and diffusive rates. The slow buoyancy profile is also significantly changed. In particular, regions initially strongly stratified are mixed up, and part of the region initially nearly mixed are restratified.

Reducing h allows for the two shear layers to interact. In particular, with $h = 1/8$, the initial condition is essentially a jet localized in the middle of the domain. With relatively

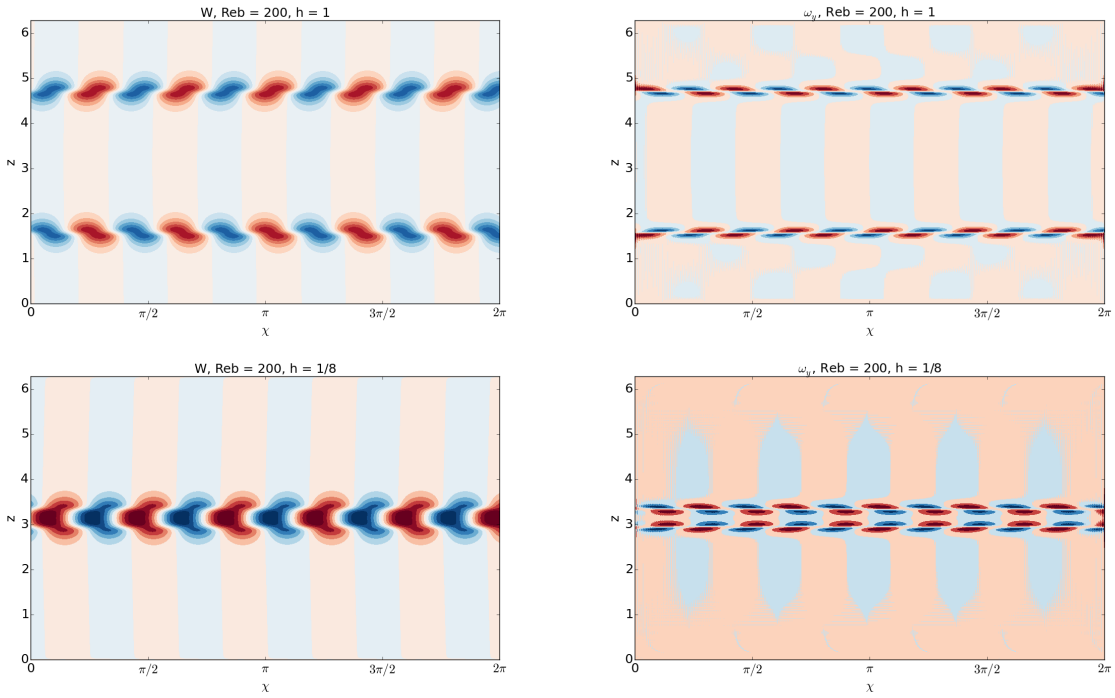

Figure 4: The wavestructure of the unstable modes for $h = 1$ (top) and $h = 1/8$ (bottom): (left) vertical velocity and (right) horizontal vorticity.

large Re_b, the evolution of the energy is similar to the cases . The perturbations grow exponentially, and then quickly saturate. The decay of the perturbations also occur in two stages, first a relatively fast decay, followed by a slow viscous decay. However, in this case the mean flow is not accelerated during the decay of the perturbations. That is, for this particular case, the buoyancy flux and the Reynolds stress are sign-definite.

3.3 Bulk properties

To characterize the solutions of this model problem in parameter space, we calculate some diagnostic bulk propoerties. In particular, the energy partition of the perturbations is

$$\gamma \stackrel{\text{def}}{=} \frac{PE_{fast}}{PE_{fast} + KE_{fast}} \,. \tag{45}$$

This particular bulk property is of interest because it is a proxy of mixing efficiency. Because we are dealing with unforced problem, this property is calculated when the perturbations peak or saturate. Figure 11 shows the distribution of γ in the $h - \mathrm{Re}_b$ space. Typically, γ increases with h, with maximum values of 0.7 for $h = 1$ (and $\mathrm{Re}_b \geq 50$); γ is about 0.35 for $h = 1/8$. More over, it is quite surprising that for $\mathrm{Re}_b \geq 50$, γ is nearly independent of Re_b. There is no simple scaling arguments that collapse this data into a single curve. That is, γ is not a simple function of h or the initial potential energy.

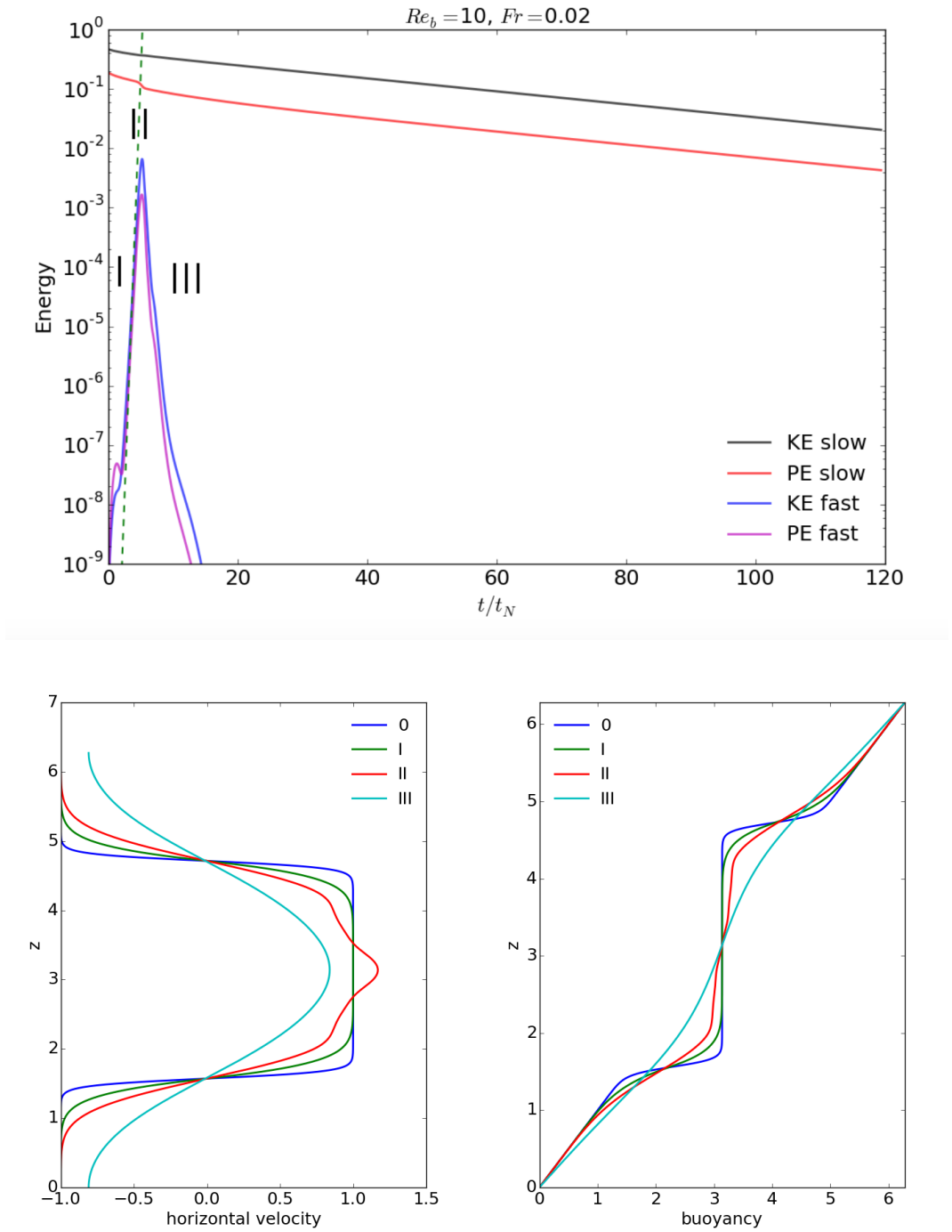

Figure 5: Solution to the IVP with $\mathrm{Re}_b = 10$ and $h = 1$. The upper panel show the time series for different components of energy. The bottom panels depict the evolution of the slow flow at different stages (marked in the energy plot). The green dashed line depicts the linear growth for the most unstable mode.

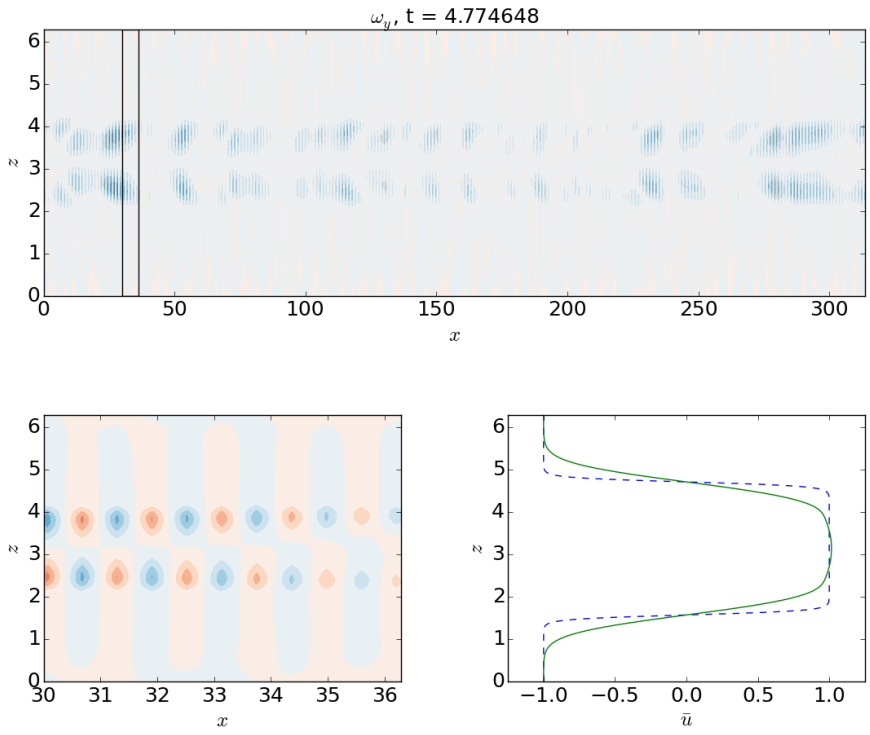

Figure 6: Snapshot of the structure of the solution to the IVP with $\mathrm{Re}_b = 10$ and $h = 1$. The upper panel shows the horizontal vorticity field. The lower left panel is a zoom-in of the upper panel in the region marked by the black square. The lower right panel shows the snapshot of the slow horizontal velocity (green) together with the initial condition.

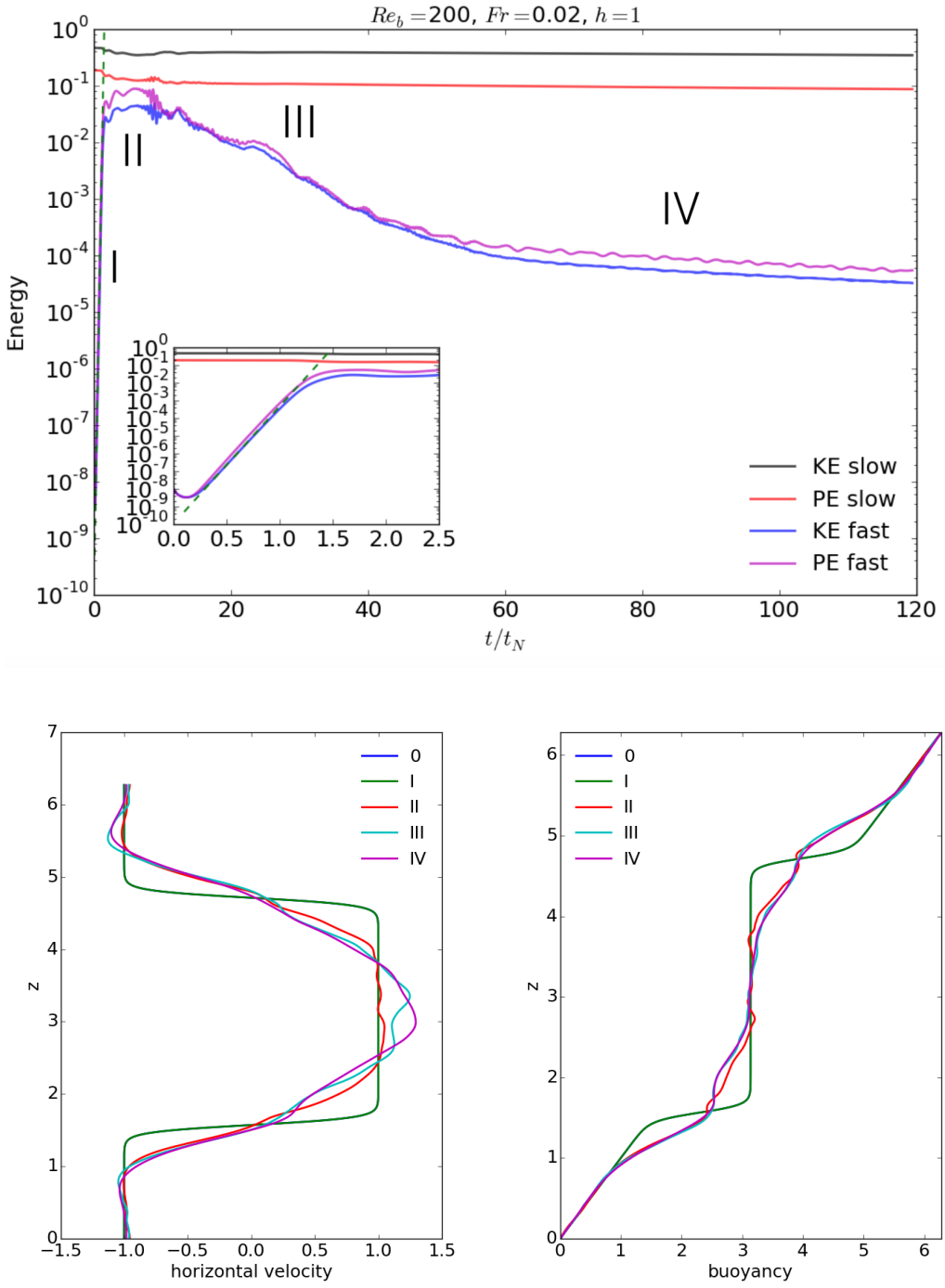

Figure 7: Solution to the IVP with $\mathrm{Re}_b = 200$ and $h = 1$. The upper panel show the time series for different components of energy. The bottom panels depict the evolution of the slow flow at different stages (marked in the energy plot). The green dashed line depicts the linear growth for the most unstable mode.

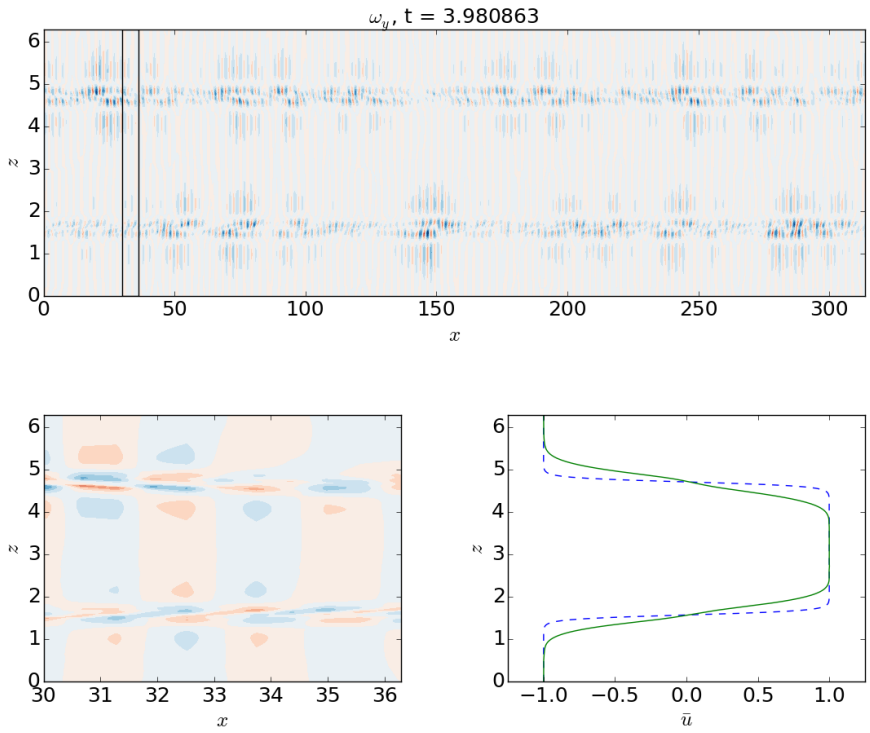

Figure 8: Snapshot of the structure of the solution to the IVP with $\mathrm{Re}_b = 200$ and $h = 1$. The upper panel shows the horizontal vorticity field. The lower left panel is a zoom-in of the upper panel in the region marked by the black square. The lower right panel shows the snapshot of the slow horizontal velocity (green) together with the initial condition.

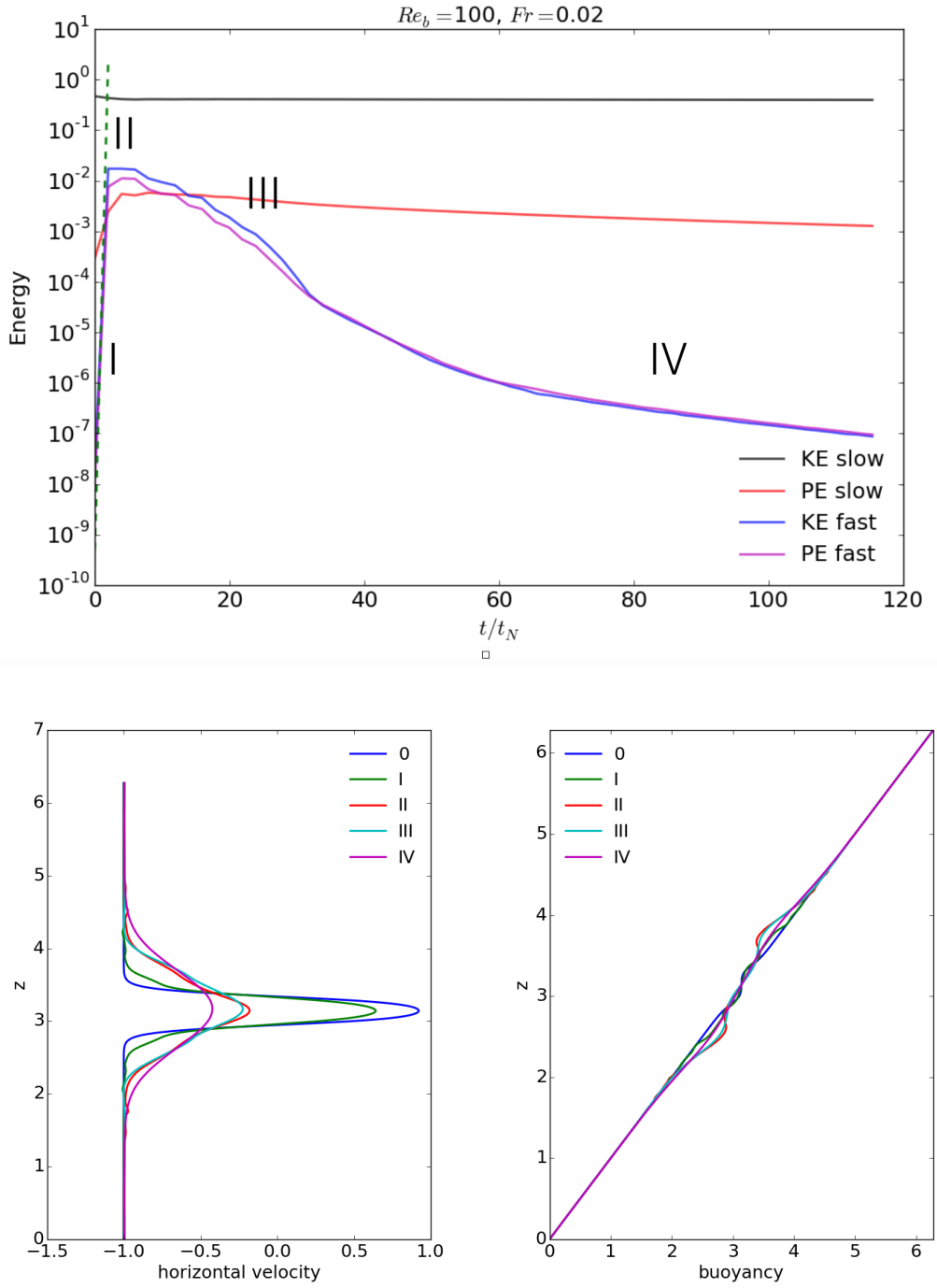

Figure 9: Solution to the IVP with $\mathrm{Re}_b = 100$ and $h = 1_8$. The upper panel show the time series for different components of energy. The bottom panels depict the evolution of the slow flow at different stages (marked in the energy plot). The green dashed line depicts the linear growth for the most unstable mode.

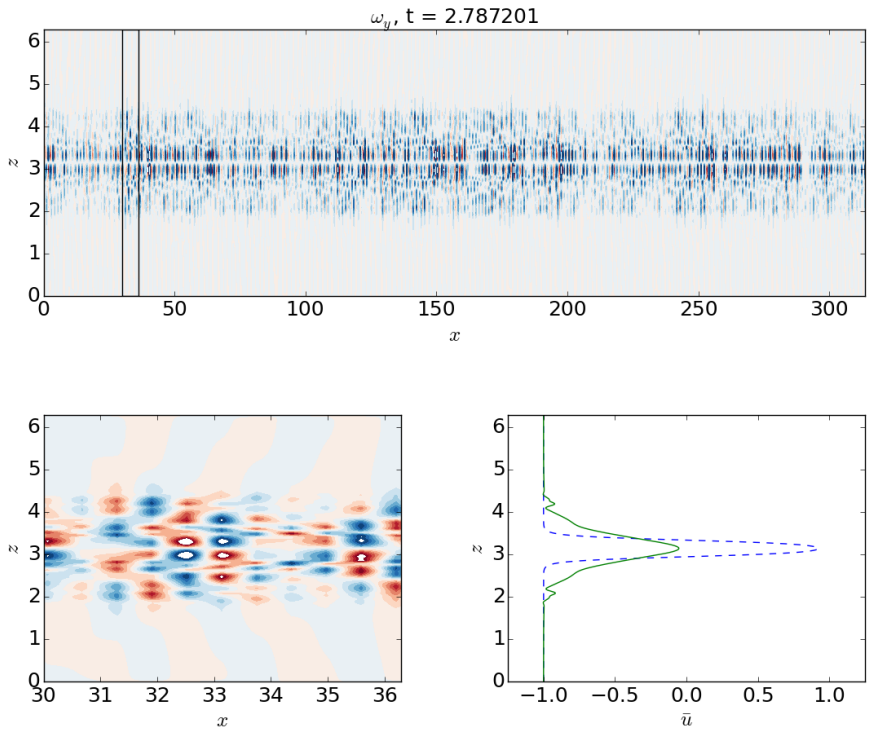

Figure 10: Snapshot of the structure of the solution to the IVP with $\mathrm{Re}_b = 100$ and $h = 1/8$. The upper panel shows the horizontal vorticity field. The lower left panel is a zoom-in of the upper panel in the region marked by the black square. The lower right panel shows the snapshot of the slow horizontal velocity (green) together with the initial condition.

We also calculate the gain of the perturbations

$$G \stackrel{\text{def}}{=} 2 \times 10^6 \times (KE_{fast} + PE_{fast}) \,, \tag{46}$$

where we remind the reader that 10^{-6} is the magnitude of the initial random seed. First, we note that G is relatively large ($> 10^7$) across the parameter space (Figure 11). The pattern of G, however, has much more structure than the pattern of γ. As expected, the gain typically increases with Re_b. Also G is minumum for $h = 1/8$ likely because the initial potential energy is much smaller than with larger h. G peaks for $1/2 \leq h \leq 3/4$.

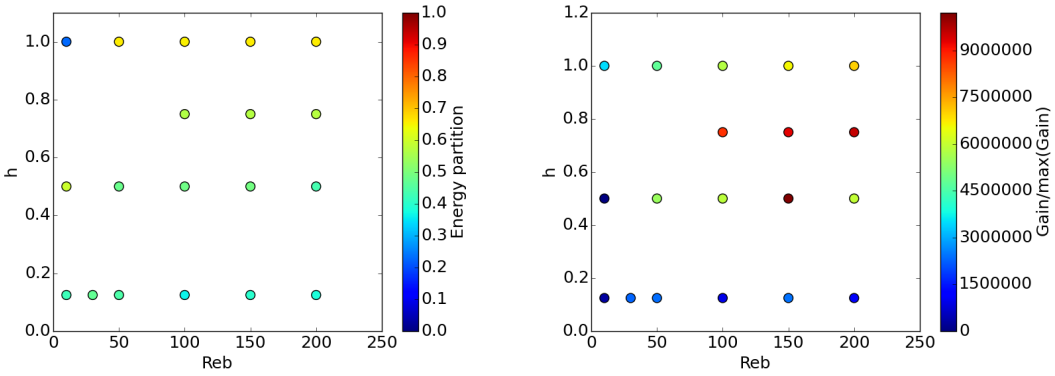

Figure 11: Bulk properties as a function of Re_b and h. (Left) Energy partition and (right) energy gain.

4 Final Remarks

In this study we have taken advantage of the strong anisotropy driven by strong stratification to simplify the modeling of stratified turbulent flows. In particular, the asymptotics lead to the hydrostratic primitive equations of oceanography for the slow fields coupled by the vertical divergence of the Reynolds stress and buoyancy flux to a quasi-linear system for the perturbations. Because the system is quasi-linear for the perturbations, there is no closure problem; the system is closed.

Using a simple toy problem, we have demonstred the propertities and utility of this reduced system. This example shows that the coupled system displays interesting dynamics with $\mathcal{O}(1)$ feedbacks between mean and perturbations. We emphasize that the Reynolds stresses and buoyancy fluxes associated with the perturbations are not sign-definite in many regions of the parameter space.

Future work include introducing back variations in slow spatial coordinate, and therefore assessing the importance of vertical vorticity in the slow flow; comparisons with direct numerical simulations; and three-dimensional solutions. It is our hope that these equations would be used to study strongly stratified turbulent flows in regions of the parameter of geophysical relevance, which are currently unattainable through direct numerical simulations.

Acknowledgements

Greg Chini and Keith Julien started working on this project in Fall 2014. I thank Greg Chini, Colm Caulfield, and Keith Julien for proposing this project, and for their generous advice over the Summer. I had useful discussions with Daniel Lecoanet, who also helped with code in the initial stages of this project.

A The self-similarity of strongly stratified flows

This appendix justifies the scaling choice $\mathrm{Fr} \sim \alpha$, which implies that the vertical scale is $\mathcal{H} \sim \mathcal{U}/N$. Following [2] we consider the dimensional inviscid Boussinesq equations in the limit $\mathrm{Fr} \to 0$

$$\partial_t \boldsymbol{u}_h + \boldsymbol{u}_h \cdot \boldsymbol{\nabla}_h \boldsymbol{u}_h + w \partial_z \boldsymbol{u}_h = -\boldsymbol{\nabla}_h \phi \,, \tag{47}$$

$$0 = -\partial_z \phi + b \,, \tag{48}$$

$$\boldsymbol{\nabla}_h \cdot \boldsymbol{u}_h + \partial_z w \,, \tag{49}$$

and

$$\partial_t b + \boldsymbol{u}_h \cdot \boldsymbol{\nabla}_h b + w \partial_z b = w \,, \tag{50}$$

Equations (47) through (50) are invariant under the transformations

$$N = N^*/A \,, \qquad z = A z* \,, \qquad w = A\, w^* \,, \qquad \text{and} \qquad b = b^*/A \,, \tag{51}$$

where A is a constant. Thus, for constant N, the solutions to (47) through (50) can be determined from the solutions from the same equations with $N^* = 1$. With $A = 1/N$, we have

$$\begin{aligned} \boldsymbol{u}_h &= \boldsymbol{u}_h^*(x, y, zN, t) \,, \\ w &= \frac{1}{N} w^*(x, y, zN, t) \,, \\ b &= N b^*(x, y, zN, t) \,. \end{aligned} \tag{52}$$

Note that the vertical length scale is inversely proportional to the buoyancy frequency; as the stratification increases the vertical scale decreases $\mathcal{H} \propto 1/N$. Thus, on dimensional grounds, we have $\mathcal{H} \sim \mathcal{U}/N$. This self similarity implies that $\mathrm{Fr} \sim \alpha$, and therefore as $\mathrm{Fr} \to 0$, the aspect ratio adjusts so that the ratio Fr/α remains $\mathcal{O}(1)$. This has profound consequences for the dynamics because the system does not become two-dimensional as $\mathrm{Fr} \to 0$.

B The Taylor-Goldstein equation

Consider the inviscid two-dimensional (x, z) version of equation (37).

$$(\partial_\tau + \bar{u}\partial_\chi) \left[(\partial_\tau + \bar{u}\partial_\chi) \left(\partial_\chi^2 + \partial_z^2 \right) - \partial_z^2 \bar{u} \partial_\chi \right] w' + (\partial_z \bar{b} + 1) \partial_\chi^2 w' = 0 \,. \tag{53}$$

Now, we assume wave-like solutions $w' = \hat{w}(z) \mathrm{e}^{\mathrm{i}k(x - c\tau)}$, to obtain

$$(\bar{u} - c)^2 \left(\hat{w}_{zz} - k^2 \hat{w} \right) + \big[\underbrace{(\bar{b}_z + 1)}_{\equiv N^2(z)} - (\bar{u} - c)\, \bar{u}_{zz} \big] \hat{w} = 0 \,, \tag{54}$$

which is the celebrated Taylor-Goldstein equation. In a bounded vertical domain of length 2π, we assume no-normal flow at the boundaries, which reduces to

$$\hat{w} = 0 \,, \qquad z = 0,\, 2\pi \,. \tag{55}$$

The Taylor-Goldstein equation has been significantly analyzed. Among the most important results, the sufficient condition for stability due to John Miles and Lou Howard is that the gradient Richardson's number be greater than a quarter:

$$\mathrm{Ri} \stackrel{\text{def}}{=} \frac{N^2(z)}{\bar{u}_z^2} > \frac{1}{4}\,. \tag{56}$$

Also, again due to Lou Howard, the unstable phase speed $c_r \stackrel{\text{def}}{=} \mathrm{Re}\{c\}$ is bounded by $\bar{u}_{min} < c_r < \bar{u}_{max}$, and $c_i \stackrel{\text{def}}{=} \mathrm{Im}\{c\}$ is bounded by

$$c_i^2 \leq \left[\tfrac{1}{2}\left(\bar{u}_{max} - \bar{u}_{min}\right)\right]^2 - \left[c_r - \tfrac{1}{2}\left(\bar{u}_{max} + \bar{u}_{min}\right)\right]^2 . \tag{57}$$

An upper bound on the growth rate is

$$k\, c_i \leq \frac{k}{2}\left(\bar{u}_{max} - \bar{u}_{min}\right) . \tag{58}$$

C Conservation of energy

We form an equation for the kinetic energy of the slow/large flow by dotting the (25) and adding to $\bar{w}$ times (28) to obtain

Slow flow

$$\partial_t \tfrac{1}{2}|\bar{\boldsymbol{u}}_h|^2 = \bar{w}\bar{b} + \overline{w'\boldsymbol{u}_h'}\cdot\partial_z\bar{\boldsymbol{u}}_h + \boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\boldsymbol{T}^{\boldsymbol{x}} + \partial_z T^z - \frac{1}{\mathrm{Re}_b}\left|\partial_z\bar{\boldsymbol{u}}_h\right|^2 , \tag{59}$$

where

$$\boldsymbol{T}^{\boldsymbol{x}} = -\bar{\boldsymbol{u}}_h\left(\bar{p} + \tfrac{1}{2}|\bar{\boldsymbol{u}}_h|^2\right) , \tag{60}$$

$$T^z = \frac{1}{\mathrm{Re}_b}\partial_z\tfrac{1}{2}|\bar{\boldsymbol{u}}_h|^2 + \bar{\boldsymbol{u}}_h\cdot\overline{w'\boldsymbol{u}_h'} - \bar{w}p\,. \tag{61}$$

and we used the continuity equation (27). Similarly, an equation for the potential energy density is formed by multiplying (28) by $\bar{b}$

$$\partial_t \tfrac{1}{2}\bar{b}^2 = -\bar{w}\bar{b} + \overline{w'b'}\partial_z\bar{b} + \boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\boldsymbol{B}^{\boldsymbol{x}} + \partial_z B^z - \frac{1}{\mathrm{Re}\,\mathrm{Pr}}\left(\partial_z b\right)^2 , \tag{62}$$

where

$$\boldsymbol{B}^{\boldsymbol{x}} = -\bar{\boldsymbol{u}}_h\tfrac{1}{2}\bar{b}^2\,, \tag{63}$$

and

$$B^z = \frac{1}{\mathrm{Re}_b\,\mathrm{Pr}}\partial_z\tfrac{1}{2}b^2 + -\bar{w}\tfrac{1}{2}\bar{b}^2 - \bar{b}\,\overline{w'b'}\,. \tag{64}$$

The equation for the slow/large scale total energy is then

$$\partial_t\tfrac{1}{2}\left(|\bar{\boldsymbol{u}}_h|^2 + \bar{b}^2\right) = \overline{w'\boldsymbol{u}_h'}\cdot\partial_z\bar{\boldsymbol{u}}_h + \overline{w'b'}\partial_z\bar{b} + \boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\left(\boldsymbol{T}^{\boldsymbol{x}} + \boldsymbol{B}^{\boldsymbol{x}}\right) + \partial_z\left(T^z + B^z\right) . \tag{65}$$

Fast flow

To obtain an equation for the energy density of the fast flow, we dot (33) with $\boldsymbol{u}_h'$, add to w' times (34), and average over the fast time to obtain

$$\mathrm{Fr}\,\partial_t \tfrac{1}{2}\left(\overline{\boldsymbol{u}_h'^2}+\overline{w'^2}\right)=\overline{w'b'}-\overline{\boldsymbol{u}_h'w'}\partial_z\bar{\boldsymbol{u}}_h'+\boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\boldsymbol{T}_2^{\boldsymbol{x}}+\partial_z T_2^z-\frac{\mathrm{Fr}}{\mathrm{Re}_b}\left(\mathrm{Fr}^2\overline{\left|\boldsymbol{\nabla}_h^{\boldsymbol{x}}\boldsymbol{u}_h'\right|^2}+\overline{\left|\partial_z\boldsymbol{u}_h'\right|^2}\right), \tag{66}$$

where

$$\boldsymbol{T}_2^{\boldsymbol{x}}=-\mathrm{Fr}\bar{\boldsymbol{u}}_h\left(p'+\tfrac{1}{2}\overline{|\boldsymbol{u}_h'|}^2+\tfrac{1}{2}\overline{w'^2}\right)+\frac{\mathrm{Fr}^3}{\mathrm{Re}_b}\boldsymbol{\nabla}_h^{\boldsymbol{x}}\tfrac{1}{2}\overline{|\boldsymbol{u}_h'|^2}, \tag{67}$$

$$T_2^z=\bar{\boldsymbol{u}}_h\cdot\overline{w'\boldsymbol{u}_h'}-\bar{w}p+\frac{\mathrm{Fr}}{\mathrm{Re}_b}\partial_z\tfrac{1}{2}\overline{|\boldsymbol{u}_h'|^2}. \tag{68}$$

Similarly, we obtain an equation for the potential energy density of the perturbation

$$\mathrm{Fr}\,\partial_t\tfrac{1}{2}\overline{b'^2}=-\overline{w'b'}-\overline{w'b'}\partial_z\bar{b}+\boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\boldsymbol{B}_2+\partial_z B_z-\frac{1}{\mathrm{Re}_b\mathrm{Fr}}\left(\mathrm{Fr}^2\overline{|\boldsymbol{\nabla}_h^{\boldsymbol{x}}b'|^2}+(\partial_z b')^2\right). \tag{69}$$

where

$$\boldsymbol{B}_2^{\boldsymbol{x}}=-\mathrm{Fr}\tfrac{1}{2}\boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot\bar{\boldsymbol{u}}_h\overline{b'^2}+\frac{\mathrm{Fr}^3}{\mathrm{Re}_b\,\mathrm{Pr}}\boldsymbol{\nabla}_h^{\boldsymbol{x}}\tfrac{1}{2}\overline{b'^2}, \tag{70}$$

$$B_2^z=\frac{\mathrm{Fr}}{\mathrm{Re}_b\,\mathrm{Pr}}\partial_z\overline{b'^2}. \tag{71}$$

The equation for the total energy for the leading order perturbations is

$$\mathrm{Fr}\,\partial_t\tfrac{1}{2}\left(\overline{|\boldsymbol{u}_h'|}^2+\overline{w'^2}+\overline{b'^2}\right)=-\overline{w'\boldsymbol{u}_h'}\cdot\partial_z\bar{\boldsymbol{u}}_h-\overline{w'b'}\partial_z\bar{b}+\boldsymbol{\nabla}_h^{\boldsymbol{x}}\cdot(\boldsymbol{T}_2^{\boldsymbol{x}}+\boldsymbol{B}_2^{\boldsymbol{x}})+\partial_z\left(T_2^z+B_2^z\right). \tag{72}$$

D Sensitivity of numerical solutions to time averaging

Formally, the averaging over the small scales is defined over time and space (23). For computational convenience, however, it is convenient to march coupled system of PDEs without with a single time-scale without time-averaging. To test the sensitivity of the evolution to the system to time-averaging, we perform two numerical simulations with the same parameters ($Fr=0.1$, $Re_b=80$, $k_0=4.5$, $m_0=10$). Time-averaging the small/fast flow over $1/Fr$ has insignificant effect on the initial evolution of the slow flow (figure 12). The secondary stage stage of evolution of the flow has quantitative differences, but overall no dramatic qualitative differences. We therefore conclude that, for this set of parameters, not averaging on time is not qualitatively misleading. It is not obvious whether such results hold at smaller Froude numbers, but we nevertheless make this assumption for computational feasibility.

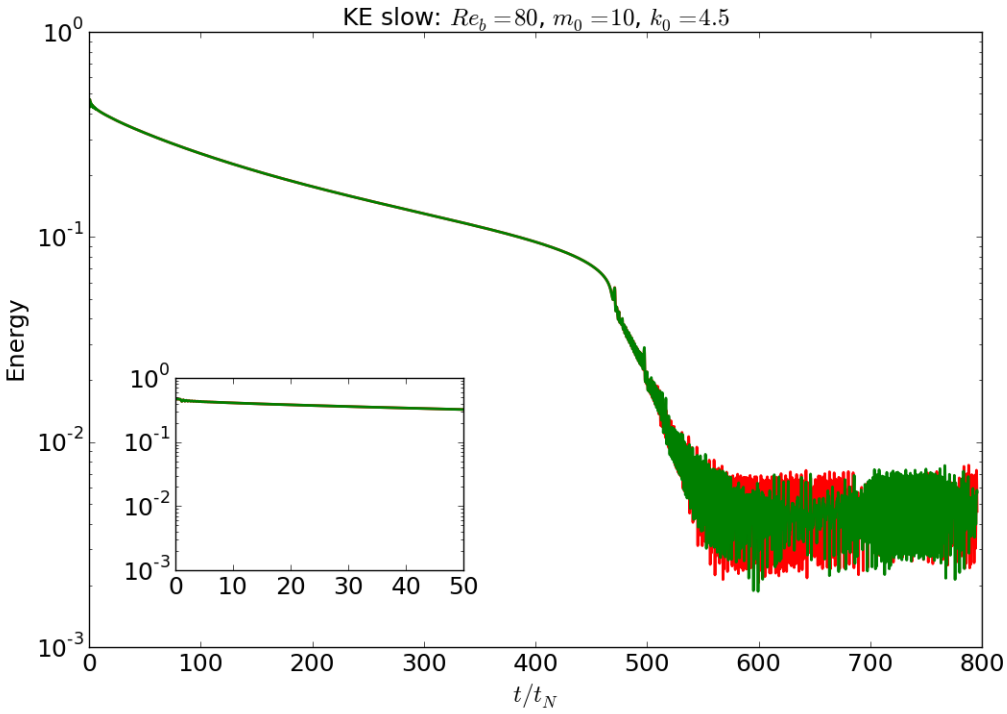

Figure 12: Kinetic energy of the slow flow for experiments with (red) and without (green) averaging. The results are very similar in the initial stages, but differ quantitatively in the secondary, oscillatory state.

References

[1] P. Bartello and S. Tobias, *Sensitivity of stratified turbulence to the buoyancy reynolds number*, Journal of Fluid Mechanics, 725 (2013), pp. 1–22.

[2] P. Billant and J.-M. Chomaz, *Self-similarity of strongly stratified inviscid flows*, Physics of Fluids, 13 (2001), pp. 1645–1651.

[3] G. Brethouwer, P. Billant, E. Lindborg, and J.-M. Chomaz, *Scaling analysis and simulation of strongly stratified turbulent flows*, Journal of Fluid Mechanics, 585 (2007), pp. 343–368.

[4] J. Callies, R. Ferrari, and O. Bühler, *Transition from geostrophic turbulence to inertia–gravity waves in the atmospheric energy spectrum*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 17033–17038.

[5] P. G. Drazin and W. H. Reid, *Hydrodynamic stability*, Cambridge University Press, 2004.

[6] K. Julien and E. Knobloch, *Reduced models for fluid flows with strong constraints*, Journal of Mathematical physics, 48 (2007), p. 065405.

[7] J. Riley and E. Lindborg, *Recent progress in stratified turbulence*, (2010).

[8] M. Sprague, K. Julien, E. Knobloch, and J. Werne, *Numerical simulation of an asymptotically reduced system for rotationally constrained convection*, Journal of Fluid Mechanics, 551 (2006), pp. 141–174.

[9] G. K. Vallis, *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*, Cambridge University Press, 2006.

Population Dynamics Forced by Stochastic Catastrophic Events

Christopher Spalding

October 15, 2015

1 Introduction

In the modern world, population extinctions are often seen in somewhat of a negative light. However, within an evolutionary context their action has been instrumental in the sculpting of the modern biosphere. Removing old forms clears the way for the emergence of new species with novel traits. Despite such importance, the mechanisms by which extinctions play out in nature are still mysterious. For example, the Earth offered up 5 mass extinction events within the past half-billion years, but it is unknown whether similar biodiversity would exist today were it not for these events. Perhaps the smaller-scale stochastic forcing of climate and competition would have sufficed.

Traditionally, environmental forcing has been considered in two broad ways. Commonly, a coloured noise term is added to the equations of population dynamics, which represents environmental forcing over a given autocorrelation time [1, 2, 3]. Though not a population model, it has been shown that cells may undergo transitions between two bistable states over a timescale that is minimised by an appropriate choice for the autocorrelation time of the noise [3]. Such a dependence on timescale is interesting when one considers that both the cell problem and the extinction problem reduce to a mean 'exit time' problem [4], where one asks what the mean time is for stochastic dynamics to take a system through a given point (zero individuals in the case of extinction).

Another way of looking at environmental stochasticity is to consider that most environmentally driven extinction occurs during particularly detrimental events, known as Catastrophes. The typical strategy for analysis of these is to suppose that catastrophes come along at a prescribed rate ν and instantaneously remove a given fraction of the population $1-p$ each time they do so [5]. This method may be related to the former method by noting that the autocorrelation time and $1/\nu$ are qualitatively similar. Accordingly, just as there is an autocorrelation timescale that minimizes transition times in cells, there may exist a frequency over which catastrophes are the most detrimental in terms of population extinctions - *a most catastrophic catastrophe*.

The importance of the correlation time and/or the time between catastrophic events is of great interest because it suggests that despite the forcing being intrinsically random, the randomness has more of an effect upon some timescales than others. To emphasise the significance of this fact, consider a deterministic environmental forcing such as the day night or seasonal cycles. Many species have adapted their life cycles to take advantage of these

predictable timescales, be it the breeding of mammals in spring or the growth-mortality daily cycles of phytoplankton [6]. It is less obvious whether a stochastic forcing, with no deterministic periodicity, can select for a specific timescale within a population.

To investigate the influence of time-scale upon population extinctions we model biological populations that experience stochastic catastrophic events with a typical frequency. To illustrate the problem, consider a favourable environment that is nevertheless struck randomly by deleterious events during which the death rate is enhanced. If these events are extremely rare, but last a long time, such as mass extinction events, the mean time to extinction of a given population might simply be the time until the next bad event. Consequently, if the events become more frequent, the extinction time goes down. Eventually, we reach a point where the events are not long-lived enough to make extinctions likely each time an event occurs. The extinction time then begins to lengthen as populations are usually able to survive through at least one event, having to wait for several before extinction occurs. This argument suggests that there is a frequency of stochastic events that minimises the extinction time.

2 Modelling Stochastic Populations

Not only are populations typically influenced by a stochastic environment, their dynamics are intrinsically stochastic. Specifically, births and deaths occur with a given probability, with the probability dependent upon the number of individuals. There always exists a possibility that no individual will reproduce over the individuals' lifetimes and thus extinction may occur by chance. Such intrinsic randomness is known as **demographic** stochasticity, distinct form **environmental** stochasticity. We investigate a population subject to both forms of stochastism.

We set up the problem as follows. Suppose there is a probability $P_n(\tau)$ that a population possesses n members at a time τ. Upon advancing time forward by one unit, the probability at each n will change in one of two ways. Probability may enter from some other n, which in the case of births means a probability flow from $n-1$ to n and deaths come from $n+1$. Alternatively, births and deaths at level n remove probability from P_n. We may write this process succinctly as

$$P_n(\tau+1) = \sum_m \pi_{nm} P_m(\tau), \tag{1}$$

where π_{nm} is the probability that the population moves from m to n. Population models usually consider only nearest neighbours to exchange probability, such as might be expected from single birth and death events. However, we retain generality at first.

One of the terms in the sum multiples $P_n(\tau)$ and so we pick it out

$$P_n(\tau+1) = \pi_{nn} P_n(\tau) + \sum_{m \neq n} \pi_{nm} P_m(\tau). \tag{2}$$

The probability of staying at the same n (π_{nn}) is just 1 minus the probability of transferring to anything else, or

$$\pi_{nn} = 1 - \sum_{m \neq n} \pi_{nm}. \tag{3}$$

Accordingly, we arrive at the discrete-time master equation

$$P_n(\tau+1) - P_n(\tau) = \sum_{m\neq n} \Big(\pi_{nm} P_m(\tau) - \pi_{mn} P_n(\tau) \Big). \tag{4}$$

It is often more convenient to take the continuous-time limit of equation (4), but we must then recast the probabilities π_{mn} as *rates* of probability flow R_{mn}. Upon doing so, we obtain the Master Equation

$$\frac{dP_n(t)}{dt} = \sum_{m\neq n} \Big(R_{nm} P_m(t) - R_{mn} P_n(t) \Big), \tag{5}$$

which may be written in matrix-vector form as

$$\dot{\mathbf{p}}(t) = \mathbf{Q}\,\mathbf{p}(t), \tag{6}$$

where we have defined the transition matrix $\mathbf{Q}$ and the vector $\mathbf{p}$ each of the elements of which corresponding to the probability at a given population level.

2.1 Birth-death model

Having set-up the Master Equation, we now prescribe forms for the elements of the transition matrix $\mathbf{Q}$. We require a death rate, δ_n and birth rate β_n as a function of individual number. Probability leaves step n if a birth or a death occurs, such that $R_{n+1,n}P_n = \delta_n$ and $R_{n-1,n}P_n = \beta_n$. Probability is gained by way of deaths from $n+1$ and births form $n-1$ such that the Master equation becomes

$$\frac{dP_n(t)}{dt} = -(\beta_n + \delta_n)P_n + \beta_{n-1}P_{n-1} + \delta_{n+1}P_{n+1}. \tag{7}$$

We will choose the exact forms for β_n and δ_n below. For now, suppose that there exists a stochastic, environmental variable $\mathcal{I}$ that modulates the death rate according to

$$\delta_n = \delta_n^{(0)} + \mathcal{I}(t)n, \tag{8}$$

where $\delta_n^{(0)}$ is the death rate when $\mathcal{I} = 0$.

Despite being probabilities, we have now introduced stochastic functions into β and δ themselves, removing the usefulness of the Master Equation. Strictly speaking, we must instead introduce a second stochastic dimension, described by a co-ordinate $\mathcal{I}$. However, for simplicity, we choose to allow $\mathcal{I}$ to take one of only two values, corresponding to a "good" state and a "bad" state. The good state has $\mathcal{I} = 0$ whereas in the bad state, $\mathcal{I} = aA$, where $A > 0$ is a constant and a is a constant with dimensions of inverse time that scales the organisms' generational overturning rate (see below).

The benefit of our two state system is that we can get around the requirement for a second dimension by defining two separate probability distributions, P_n^+ and P_n^-, where the former corresponds to the bad state and the latter the good state (the "plus" represents an enhanced death rate). We suppose that the system switches from the bad to the good state randomly, but with a typical frequency α. Conversely, the system switches from good to

bad at rate $\epsilon\alpha$ where $\epsilon < 1$ corresponds to an environment that is in the good state more often than the bad state on average. [1] Note that the mean value of $\mathcal{I}(t)$, which we refer to as $\bar{A}$ is given by

$$\bar{A} = \frac{\epsilon}{1+\epsilon} A. \tag{9}$$

We are now in a position to write separate master equations for P_n^+ and P_n^-. In particular, we add a probability flow of αP_n^+ from the plus state to the minus state and likewise a flow of $\epsilon\alpha P_n^-$ from the minus state to the plus state. Including these terms, we obtain the Master Equations:

$$\begin{aligned}
\frac{dP_n^+(t)}{dt} &= -(\beta_n + \delta_n^{(0)} + aAn)P_n^+ + \beta_{n-1}P_{n-1}^+ + \left(\delta_{n+1}^{(0)} + aA(n+1)\right)P_{n+1}^+ - \alpha\, P_n^+ + \epsilon\,\alpha\, P_n^- \\
\frac{dP_n^-(t)}{dt} &= -(\beta_n + \delta_n^{(0)})P_n^- + \beta_{n-1}P_{n-1}^- + \delta_{n+1}^{(0)}P_{n+1}^- + \alpha\, P_n^+ - \epsilon\,\alpha\, P_n^-.
\end{aligned} \tag{10}$$

These equations describe the probabilistic trajectory of a population subject to the stochastic forcing described above. An alternative view points to consider trajectories of the population size. We do not adopt such an approach, but a typical realisation is illustrated in figure 6, where we plot the number of individuals (scaled by the carrying capacity) as a function of time under the influence of stochastic catastrophic events.

2.2 Matrix approach

It is more convenient to analyse the equations above in matrix form. Accordingly, we recast the Master Equations into the form:

$$\frac{d}{dt}\begin{pmatrix} P_n^+ \\ P_n^- \end{pmatrix} = \begin{pmatrix} M_{mn}^{(0)} + aAO_{mn} - \alpha I_{mn} & \epsilon\alpha I_{mn} \\ \alpha I_{mn} & M_{mn}^{(0)} - \alpha\epsilon I_{mn} \end{pmatrix}\begin{pmatrix} P_n^+ \\ P_n^- \end{pmatrix}, \tag{11}$$

where $M_{mn}^{(0)}$ is a tridiagonal matrix that represents birth and death rates in the good state. To include the environmentally-enhanced death rate, we define the matrix

$$O_{mn}P_n \equiv (n+1)P_{n+1} - nP_n, \tag{12}$$

along with the identity matrix, given by I_{mn}.

In order to complete the specification of the problem, we prescribe functional forms for the birth and death rates. We choose to have the birth rate grow linearly with n while the death rate grows quadratically, forcing an "equilibrium" number $n = K$ (the carrying capacity) at which $\delta_n^{(0)} = \beta_n$. The carrying capacity here corresponds to the stationary solution of the good state in the limit where dynamics are deterministic. Specifically, we define

$$\begin{aligned}
\delta_n^{(0)} &= an\left(\frac{n}{K}\right) \\
\beta_n &= an
\end{aligned} \tag{13}$$

[1] Switching between states in this manner is known as a Telegraph Process [4].

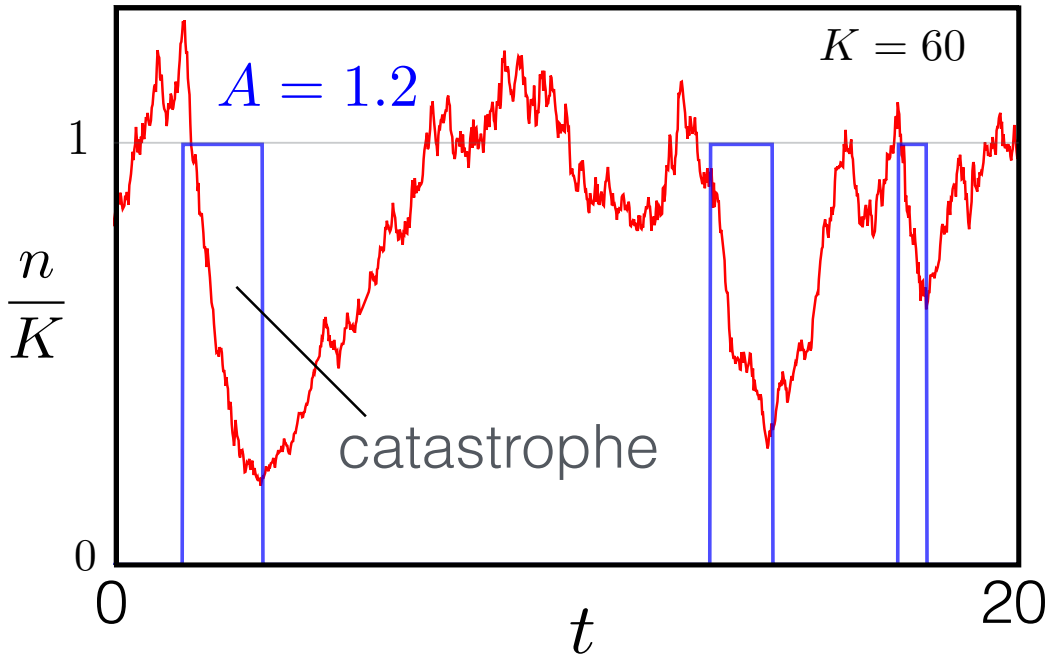

Figure 1: A typical realisation of the population size, n, normalised by the carrying capacity $K = 60$ under the action of stochastically distributed catastrophic events. The events typically last long enough to initiate a significant decline in the population numbers. The typical switching frequency $\alpha = 1$, with time measured in units of $1/a$. More rapid fluctuations would increase the extinction time because each catastrophe becomes significantly less detrimental. On the other hand, a slowing of the events would decrease extinction time, until the events' severity is offset by their scarcity.

and so as before, a measures the typical rate at which births and deaths proceed. Note that in this form, the linear environmental augmentation of death rate (An) naively appears commensurate with a decrease in birth rates. This statement is only true in the deterministic case; as we show below, adding both deaths and births generates more noise in the system than simply subtracting births, even if the resulting "deterministic" growth rate is the same in both cases.

2.3 Numerical solution

With all parts of the problem defined, we can now integrate equation (11) numerically to describe the time evolution of P_n under our specified stochastic forcing. The most important quantity to be extracted from the model is the extinction rate $\mathcal{R}_\mathrm{e}$ which for now we simply define as

$$\mathcal{R}_\mathrm{e} \equiv \frac{1}{P_0}\frac{dP_0}{dt}. \tag{14}$$

We present the time evolution of P_0 in Figure (3) along with its rate of change, using the initial condition that $P_n^+ = 0$ and P_n^- is drawn from the quasi-stationary distribution that would describe P_n in the absence of environmental forcing. This choice leads to a transient period, during which, probability flows from the minus state to the plus state until a quasi-steady state is reached for both. The timescale of the transient dynamics is related to how quickly equilibrium is established in the bad state.

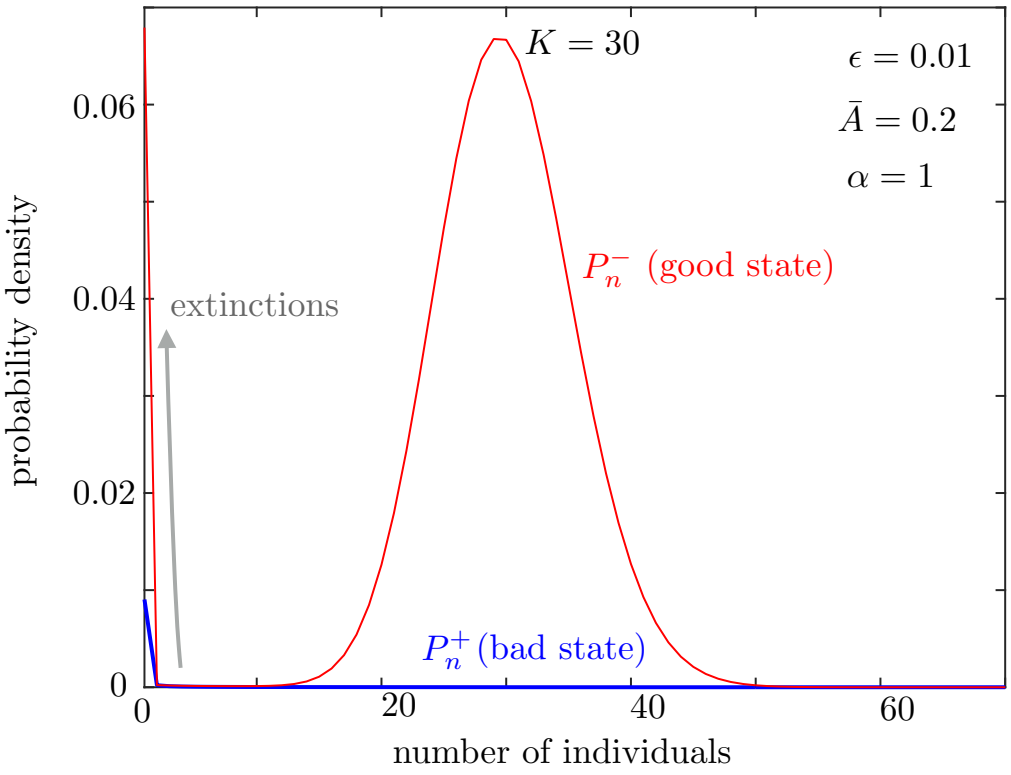

Figure 2: The quasi-steady distribution for the plus state (P_n^+, blue) and the minus state (P_n^-, red) resulting from an environmental perturbation with frequency α. Notice the pile-up of probability in the extinct state and the peak at $n = K$, but only for the minus state. The plus state only has a maximum at the origin.

In many problems, the transient period, which depends upon initial conditions, is of great importance. For example, initial conditions are crucial in computing the probability that any given event will lead to extinction. Additionally, if a system possesses multiple (quasi-) steady states, the initial configuration may determine where the long-time state will decay to. These complications do not apply to our considerations here, where we are interested in the long-term mean extinction time appropriate to an ensemble of populations under the influence of the environmental forcing prescribed above. Essentially, we neglect the probability that extinction occurs before the transient evolution decays.

Repeating the numerical integrations for a wide parameter-space would be time consuming. Instead, we take advantage of the quasi-steady evolution by seeking the lowest (in

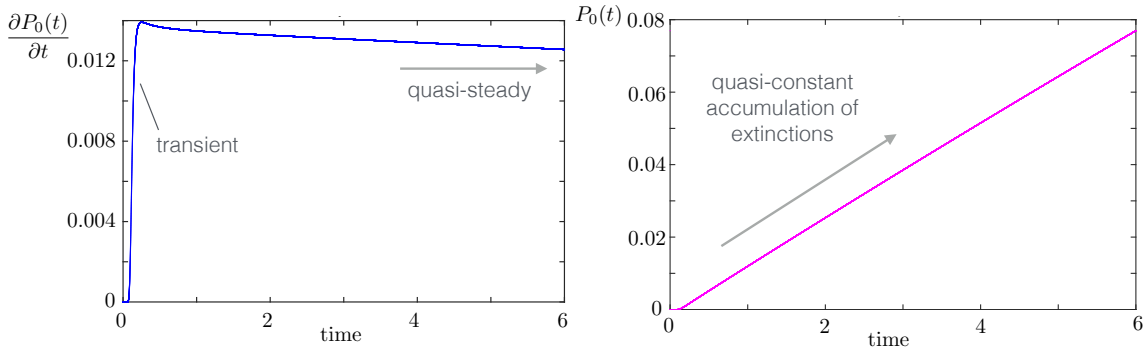

Figure 3: The flow of probability into the extinct state for the parameters of Figure 2. On the left is the rate of extinctions where the right indicates the probability of being extinct at a given time. After a brief transient period, extinctions begin to occur at a fairly constant rate. It is this quasi-steady rate we seek.

magnitude) eigenvalue λ_0 of the matrix on the RHS of equation (11). The inverse of λ_0 gives us approximately the mean time to extinction (MTE) of the population as we show next.

2.3.1 Mean time to extinction

After a time t, the probability a population is not extinct $\mathcal{P}(T_e > t)$ is simply the sum over all $n > 0$ of P_n,

$$\mathcal{P}(T_e > t) = \sum_{n>0} P_n(t), \tag{15}$$

and the probability distribution for extinction times is simply the (negative) time derivative of $\mathcal{P}(T_e > t)$. Suppose now that each P_n may be written as a sum of eigenmodes

$$P_n = \sum_m P_{nm} e^{-\lambda_m t}, \tag{16}$$

such that the mean extinction time can be written as

$$\begin{aligned} \bar{T}_e &= -\sum_{n>0}\sum_m P_{mn} \int_0^\infty T\left[\frac{\partial e^{-\lambda_m T}}{\partial T}\right] dT \\ &= -\sum_{n>0}\sum_m P_{mn} \frac{1}{\lambda_m}, \end{aligned} \tag{17}$$

where the second equality only holds when all $\lambda_m > 0$, i.e., there is no truly stationary solution except certain extinction ($P_0|_{t=\infty} = 1$). The transient state is rapid compared to the long-term quasi-steady decay of probability and so we may say that $\lambda_0 \ll \lambda_1 < \lambda_2$, etc. In other words, the sum over m above collapses to a single term, that of $m = 0$. Furthermore, the eigenvectors P_{mn} are unit-normalized, such that the double summation

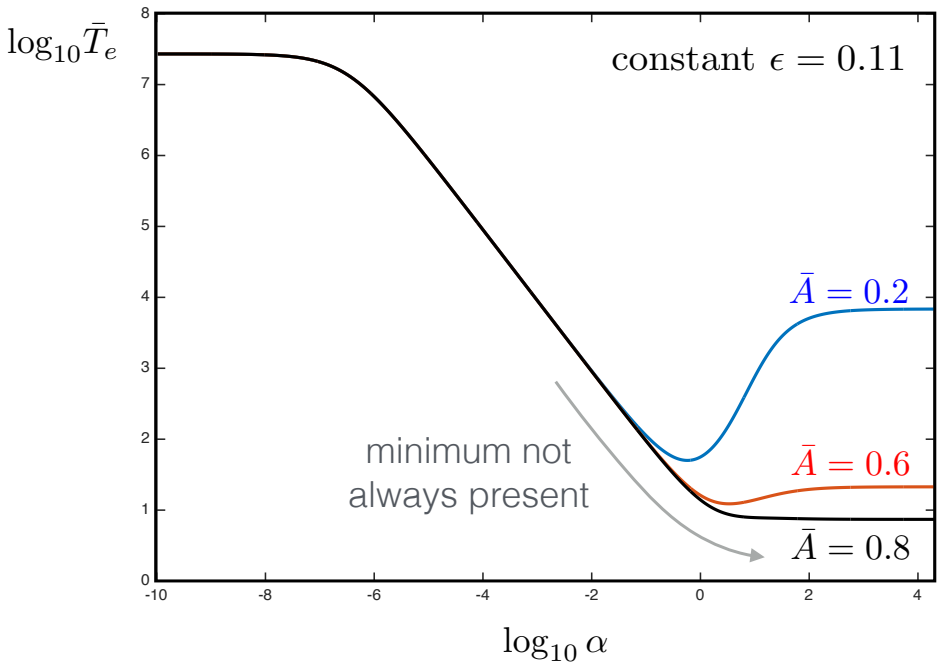

Figure 4: Similar to Figure 4, except we demonstrate that for large enough $\bar{A}$, the minimum ceases to exist.

reduces to a single inverse eigenvalue and the mean extinction is given by

$$\boxed{\bar{T}_e \approx \lambda_0^{-1}} \tag{18}$$

which is equivalent to our ascertain above.

2.4 Eigenvalue

We now compute the lowest eigenvalue numerically for a range of parameters. If $\bar{A} \gtrsim 1$, we find that the mean extinction time decreases monotonically with α (Figure 4). However, this is not the case we are interested in because such detrimental mean values make it unlikely a population would exist in the first place and, furthermore, the transient solution would become of significance in that case. Below, we consider only $\bar{A} < 1$, where the mean growth rate is not always negative, leaving room for stochastic periods of recovery. We present the results in Figure 5.

We immediately point out that there exists an α corresponding to a minimum in the mean extinction time. Such a minimum is reminiscent of several important results from the literature. For example, from the field of cell biology, cell-differentiation is sometimes thought to proceed by way of a transition from one stable state to another [7]. Theoretical models have shown that noise-induced transitions between the states can be minimised by autocorrelating the noise over some critical timescale [3]. The mean extinction time of a metapopulation can be maximised in an analogous fashion by allowing some critical degree of migration between populations [8].

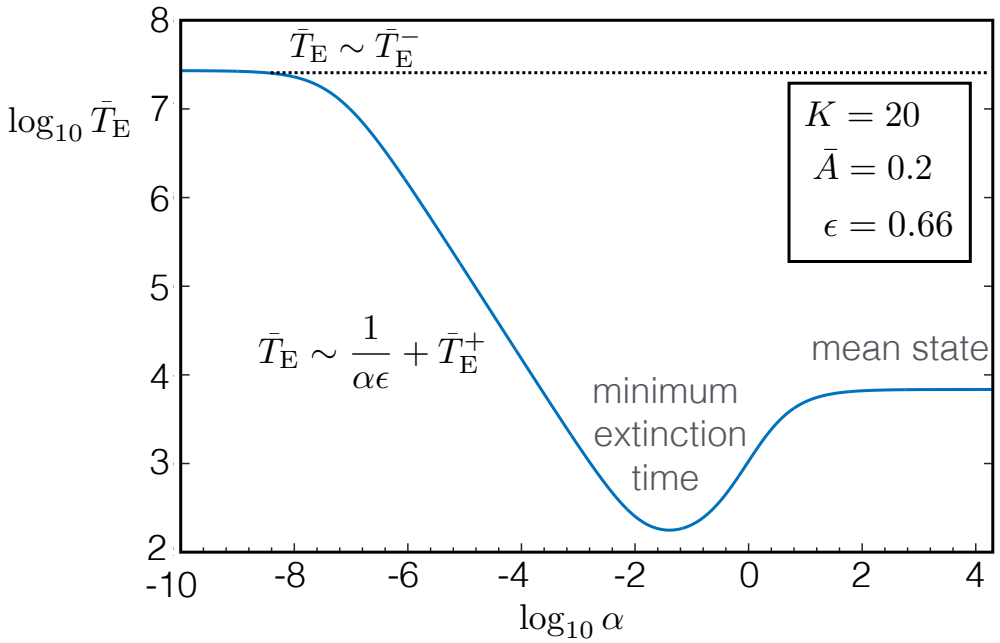

Figure 5: The mean extinction time as a function of the stochastic rate parameter α. Notice the minimum. We explain the origin of the minimum in the text.

The specific cases above are related to a more general problem - the mean first passage time of a particle over a potential barrier whose height fluctuates between a high and a low state [9]. Referred to as "Resonant Activation" (RA), the optimum stochastic frequency was found to be similar to the time it takes to escape when the barrier is at its lowest. Returning to our problem, the analogous conclusion would be that the worst α would correspond with the extinction time in the bad state. We find numerically that this statement is approximately correct at large A (with the agreement being better at larger K). However, for $A > 1$ but not significantly so, a better approximation is to solve for the α that makes the variance equal to the turnover time, or

$$\begin{aligned} \sigma^2 &= \frac{2A\bar{A}a^2}{\alpha(1+\epsilon)^2} \\ &= a \\ \rightarrow \quad \alpha_{\rm res} &\approx 2A\bar{A}a\left(1-\frac{\bar{A}}{A}\right)^{-2}. \end{aligned} \tag{19}$$

We do not yet have a theory to explain this correspondence.

2.5 Origin of "resonance"

In Figure 5, we highlight the dominant influence upon extinction time as a function of α. As $\alpha \to 0$, the mean extinction time is the mean of the extinction time in the bad state and good state [10]. Owing to the greatly reduced extinction time in the bad compared to the

good state, along with $\epsilon < 1$, the mean extinction time in the small α limit simply becomes approximately the extinction time in the good state, $\bar{T}_e^-$.

Increase α such that its inverse $\alpha^{-1} \lesssim \bar{T}_e^-$ and we enter the regime where a population will not typically go extinct before it encounters a bad state. However, once the bad state hits, with high probability, the population goes extinct before another good state is encountered. Accordingly, the mean extinction time in this regime is roughly the mean time to enter a bad state ($1/\epsilon\alpha$), plus the extinction time in the bad state:

$$\bar{T}_e \sim 1/\epsilon\alpha + \bar{T}_e^+. \tag{20}$$

Eventually, α becomes sufficiently large that the most probable trajectory makes it through at least one bad event, recovering during the subsequent good state. Accordingly, the mean extinction time will begin to rise above the extrapolation of $1/\epsilon\alpha + \bar{T}_e^+$. The α at which such a transition occurs corresponds approximately to the case when $\bar{T}_e^+ \approx 1/\alpha$. Substituting this condition into the small-α expression above, we arrive at an estimate for the minimum extinction time

$$\bar{T}_e^{(\min)} \approx \bar{T}_e^+ \left(1 + \frac{1}{\epsilon}\right). \tag{21}$$

Unfortunately, this approximation is only good to within an order of magnitude or so.

One more regime may be described analytically: the limit $\alpha \to \infty$. In this regime, the environment switches so rapidly that the population only "sees" the mean value of environmental forcing $\bar{A}$. Accordingly, the extinction time $\bar{T}_e^{(\infty)}$ corresponds with that which would be calculated in a stationary environment where δ_n is modulated by $\bar{A}n$. Notice that there is no guarantee in general that $\bar{T}_e^{(\infty)} > \bar{T}_e^{(\min)}$, as is apparent from Figure 4. Rather, a minimum exists only if the mean state is sufficiently favourable, or, that increasing α above α_{res} allows the population sufficient time to recover in the good periods, otherwise, increasing α is monotonically more detrimental.

2.6 Implications/Applications

The existence of a minimum extinction time is of potential significance in a variety of ways. First, our results suggest that even stochastic forcing can lead to selection upon a population's reproductive timescale. Recall that all times here are essentially scaled by a, the organism's life cycle turnover rate. If the population is forced at α_{res}, it can increase its mean extinction time by changing a, with an increase in a corresponding with movement to the left in Figure 4. This result is interesting because, naively, one would not expect a population to be able to adapt in this way because its biology cannot 'know' when the next bad event will occur.

On a longer timescale, the existence of a 'most catastrophic catastrophe' at α_{res} implies an important realisation. Of all extinctions throughout evolution history, it appears that to a first approximation, more populations have been removed by fluctuations over roughly the timescale of a life-cycle than any other timescale, though exactly what that timescale is depends upon the species. Considering annual life cycles, fluctuations on a year to year timescale, such as droughts or ice extent, have removed more populations than extremely rare events such as the bollide impact that sealed the dinosaurs' fate. However, as far as

recovery goes, the aftermath of a mass extinction probably leads to qualitatively different evolutionary trajectories than occur subsequent to the various background extinctions trickling in as a result of stochastic forcing at α_{res}.

Also of significance, global warming is thought increase the frequency of extreme events. An important issue is knowing how these changes may impact different species. This work suggests that each species' life-cycle timescale is crucial to understanding their extinction risks. For example, as the frequency of events changes, it may enter the resonance of some species, whilst leaving the resonance of others, actually reducing their extinction risks.

The final implication we mention here is with regard to treating infectious diseases. An extinction is good when it means wiping out an illness. Suppose multiple populations are infected by a pathogen, for which treatment is available, but only at a limited supply rate. An important question is how to best distribute the medicine to minimise the extinction time - the time it takes to wipe out the disease. The minimum here may inform how to best go about such treatments.

3 Continuum Limit and Fokker-Planck Approximation

In the form thus far adopted, i.e., discrete population numbers, we are unable to write down any simple, closed-form expressions for extinction times. However, we highlighted several regimes where the extinction time of the full system was dominated by the plus state, the minus state, or some well-defined combination thereof. Analytic approximations for these extinction times may be obtained, but only in large-K limit. In such a limit, we may approximate the population size as a continuum, thereby re-casting the master equation in terms of a Fokker-Planck equation, from which, the mean extinction time may be drawn. It must be cautioned that the continuum results will not be quantitatively the same as the discrete problem [10], but in most cases the general qualitative nature is preserved.

Above, the population may be thought of as occupying one of a semi-infinite number of "steps" n at any one time. Now suppose we define a new variable

$$x \equiv \frac{n}{K}, \tag{22}$$

such that the space between steps is reduced by a factor K. If we make K large, meaning that the equilibrium population size is large, the space of x becomes closer to a continuum. The continuum version of the master equation is derived by considering that probability may flow into position x from other x' at a rate $W(x-x')R(x')\rho(x')$, but flow out at a rate $R(x)\rho(x)$, with $\rho(x)$ taking the place of P_n as the probability density function. It evolves according to the equation

$$\frac{d\rho(x)}{dt} = -R(x)\rho(x) + \int_0^\infty R(x')W(x-x')P(x')dx'. \tag{23}$$

Upon comparison with the birth-death process above, we see that the function $W(x-x')$ forces probability to only flow between nearest neighbours, corresponding to an interval $\Delta x = 1/K$. Accordingly, W takes the form of two Dirac delta functions, one at $x + 1/K$ and one at $x - 1/K$. We may now integrate equation 23, then perform a Taylor expansion

to order $\mathcal{O}(1/K)^2$ about x such that

$$\frac{d\rho(x)}{dt} \approx \frac{\partial}{\partial x}\bigg[- f(x) + \frac{1}{2K}\frac{\partial}{\partial x}\Big(g^2(x)\Big)\bigg]\rho(x) \tag{24}$$

where we have defined the "drift" $f(x)$ and "diffusion" $g(x)$ using the continuous analogues of δ_n and β_n above:

$$\begin{aligned} f(x) &= \beta(x) - \delta(x) & g(x)^2 &= \beta(x) + \delta(x) \\ &= ax(1 - \mathcal{A} - x) & &= ax(1 + \mathcal{A} + x). \end{aligned} \tag{25}$$

The Fokker-Planck equation and functions $f(x)$ and $g(x)$ make be applied to the good state by setting $\mathcal{A} = 0$, the bad state using $\mathcal{A} = A$ and the mean, large-α limit using $\mathcal{A} = \bar{A}$. Within each of these regimes, we may calculate an approximate expression for the MTE. However, the functional form of the solution depends upon whether $\mathcal{A} > 1$ or $\mathcal{A} < 1$. We do not consider the case where $\mathcal{A} \sim 1$.

3.1 Asymptotic solutions

In what follows, we solve for the MTE by supposing that the PDF is stationary, but that there exists a current at $x \to \infty$ that is balanced by extinctions at the origin. The equation to to solve is

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}\Big[a\,x\,(1 - A - x)\,\rho\Big] + \frac{1}{2\,K}\frac{\partial^2}{\partial x^2}\Big[a\,x\,(1 + A + x)\,\rho\Big] \tag{26}$$

which we rewrite as

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial x}\Big[u(x)\,\rho\Big] + \frac{1}{2\,K}\frac{\partial^2}{\partial x^2}\Big[v(x)\,\rho\Big] \\ &= -\frac{\partial J}{\partial x} \end{aligned} \tag{27}$$

We now solve the steady-state equation to extract the conserved current

$$J = -\frac{1}{2K}\frac{\partial}{\partial x}\Big[v(x)\rho(x)\Big] + u(x)\rho(x), \tag{28}$$

which has the general solution

$$\rho = -\frac{2KJ}{a}\frac{1}{x(x + 1 + \mathcal{A})}e^{2K\mathcal{F}(x)}\int_0^x e^{-2K\mathcal{F}(x')}dx', \tag{29}$$

where

$$\begin{aligned} \mathcal{F}(x) \equiv \int \frac{u(x)}{v(x)}dx &= \int \frac{1 - \mathcal{A} - x}{1 + \mathcal{A} + x}dx \\ &= \int \bigg[- 1 + \frac{2}{1 + \mathcal{A} + x}\bigg]dx \\ &= -x + 2\ln(1 + \mathcal{A} + x) \end{aligned} \tag{30}$$

and so the solution for ρ reads

$$\rho = \frac{2KJ}{a} e^{2K(2\ln(1+\mathcal{A}+x)-x)} \frac{1}{x(x+1+\mathcal{A})} \int_0^x e^{-2K(2\ln(1+\mathcal{A}+x')-x')} dx'. \tag{31}$$

The current J is obtained by requiring that the integral over all space of the PDF is unity:

$$\begin{aligned} \int_0^\infty \rho \, dx &= -\frac{2KJ}{a} \int_0^\infty e^{2K(2\ln(1+A+x)-x)} \frac{1}{x(x+1+A)} \int_0^x e^{-2K(2\ln(1+A+x')-x')} dx' dx \\ &= 1 \end{aligned} \tag{32}$$

We cannot solve this equation exactly, however, we may make progress by analysing the large K case. Specifically, where $\mathcal{F}(x)$ is positive, $\exp(2K\mathcal{F}(x))$ is very large and vice versa, such that we can approximate which regions of the integral make the largest contribution. The expansions must be carried out separately for $\mathcal{A} > 1$ and $\mathcal{A} < 1$. The reason is that when $\mathcal{A}$ is smaller, ρ will be Gaussian-like about $x = 1 - \mathcal{A}$, where the deterministic dynamics would reach a stead state. However, if $\mathcal{A} > 1$, this argument breaks down because the deterministic dynamics do not possess a steady solution and the PDF becomes pressed up against the origin. The integral above behaves very differently within these two regimes

3.1.1 Calculation for $\mathcal{A} < 1$

We begin with $A < 1$, such that the PDF has a Gaussian-like peak around $x = 1 - A$. The second integrand (the one over x' in 31) is approximately constant within the region of interest (under the Gaussian). More precisely,

$$\begin{aligned} \int_0^x e^{-2K(2\ln(1+\mathcal{A}+x')-x')} dx' &\approx \int_0^x e^{-4K\ln(1+\mathcal{A})} \int_0^\infty e^{-2K\frac{1-\mathcal{A}}{1+\mathcal{A}}x'} dx' \\ &= e^{-4K\ln(1+\mathcal{A})} \frac{1}{2K} \frac{1-\mathcal{A}}{1+\mathcal{A}}. \end{aligned} \tag{33}$$

Therefore, the integral becomes

$$\begin{aligned} 1 &\approx -\frac{1}{2K} \frac{1+\mathcal{A}}{1-\mathcal{A}} e^{-4K\ln(1+\mathcal{A})} \frac{2KJ}{a} \int_{-\infty}^\infty \frac{1}{2(1-\mathcal{A})} e^{2K(2\ln(2)-1+\mathcal{A})} e^{-\frac{1}{2}Ky^2} dy \\ &= -\frac{J}{2a} \frac{1+\mathcal{A}}{(1-\mathcal{A})^2} e^{2K(2\ln 2-1+\mathcal{A}-2\ln(1+\mathcal{A})} \sqrt{\frac{2\pi}{K}} \\ \rightarrow \quad J &= -\sqrt{K} a \frac{(1-\mathcal{A})^2}{1+\mathcal{A}} \sqrt{\frac{2}{\pi}} e^{-2Kc_1} \\ c_1 &\equiv 2\ln 2 - 1 + \mathcal{A} - 2\ln(1+\mathcal{A}), \end{aligned} \tag{34}$$

and so the mean extinction time

$$\boxed{-J^{-1} \approx \frac{1+\mathcal{A}}{a(1-\mathcal{A})^2} \sqrt{\frac{\pi}{2K}} \exp\left(2K[2\ln 2 - 1 + \mathcal{A} - 2\ln(1+\mathcal{A})]\right)} \tag{35}$$

with the most crucial result being the square root combined with an exponential. This expression, for $\mathcal{A} < 1$, is most applicable to the good state and the mean states.

3.1.2 Calculation for $\mathcal{A} > 1$

We return to the original equation to solve:

$$\int_0^\infty \rho dx = -\frac{2KJ}{a}\int_0^\infty e^{2K(2\ln(1+A+x)-x)}\frac{1}{x(x+1+A)}\int_0^{x'} e^{-2K(2\ln(1+A+x')-x')}dx'dx, \tag{36}$$

where, this time, $A > 1$ and so the PDF is pressed up against $x = 0$ (see numerics for P^+). Accordingly, the second integral will receive most of its contribution from $x \approx 0$. The first integral must again be expanded about small x' but **is no longer approximately constant** within the region of interest. Accordingly, we expand the exponent but do not approximate it as linear in x:

$$\begin{aligned}\int_0^x e^{-2K(2\ln(1+A+x')-x')}dx' &\approx \int_0^x e^{-4K\ln(1+A)}e^{-2K\frac{1-A}{1+A}x'}dx' \\ &= e^{-4K\ln(1+A)}\frac{1+A}{1-A}\frac{1}{2K}\left(1-e^{-2K\frac{1-A}{1+A}x}\right).\end{aligned} \tag{37}$$

We deviate once again from the $\mathcal{A} < 1$ case by expanding the exponential term in small x rather than about the maximum in x (because the maximum is now at $x < 0$). Specifically, we approximate

$$e^{2K(2\ln(1+\mathcal{A}+x)-x)} \approx e^{4K\ln(1+\mathcal{A})}e^{2K\frac{1-\mathcal{A}}{1+\mathcal{A}}x} \tag{38}$$

which after substitution yields

$$1 \approx -\frac{J}{a}\frac{1+\mathcal{A}}{1-\mathcal{A}}\int_0^\infty \frac{1}{x(1+\mathcal{A}+x)}\left(e^{2K\frac{1-\mathcal{A}}{1+\mathcal{A}}x}-1\right)dx, \tag{39}$$

which is a relatively simple form as some exponential terms have cancelled out.

Next, we define the positive quantities

$$\begin{aligned}B &\equiv \frac{\mathcal{A}-1}{\mathcal{A}+1} \\ \lambda &\equiv 2K(\mathcal{A}-1)\end{aligned} \tag{40}$$

and a positive variable

$$\xi \equiv 2KBx \tag{41}$$

such that the integral becomes

$$1 \approx \frac{J}{a}2K\int_0^\infty \frac{1}{\xi(\xi+\lambda)}\left(e^{-\xi}-1\right)d\xi. \tag{42}$$

Integrating by parts:

$$\frac{a}{2KJ} \approx \frac{1}{\lambda}{}_0^\infty\left[\ln\left(\frac{\xi}{\xi+\lambda}\right)(e^{-\xi}-1)\right] + \frac{1}{\lambda}\int_0^\infty \ln\left(\frac{\xi}{\xi+\lambda}\right)e^{-\xi}d\xi. \tag{43}$$

whence the boundary term vanishes, leading to the compact integral

$$\frac{a}{2KJ} \approx \frac{1}{\lambda}\int_0^\infty \ln\left(\frac{\xi}{\xi+\lambda}\right)e^{-\xi}d\xi. \tag{44}$$

Adding and subtracting $\ln(\lambda)$ yields

$$\begin{aligned}\frac{a\lambda}{2KJ} &= \int_0^\infty \ln\left(\frac{\xi}{\lambda}\right)e^{-\xi}d\xi - \int_0^\infty \ln\left(1+\frac{\xi}{\lambda}\right)e^{-\xi}d\xi \\ &= -\gamma - \int_0^\infty \ln(\lambda)e^{-\xi}d\xi - \int_0^\infty \ln\left(1+\frac{\xi}{\lambda}\right)e^{-\xi}d\xi \\ &\approx -\gamma - \ln\lambda - \int_0^\infty \left(\frac{\xi}{\lambda}+\frac{\xi^2}{2\lambda^2}\right)e^{-\xi}d\xi \\ &= -\gamma - \ln\lambda - \frac{1}{\lambda} - \frac{1}{\lambda^2}. \end{aligned} \tag{45}$$

Finally, the mean extinction time, neglecting $1/K$ terms, becomes

$$\boxed{-J^{-1} \approx \frac{1}{a(\mathcal{A}-1)}\Big(\gamma + \ln(2K(\mathcal{A}-1))\Big)} \tag{46}$$

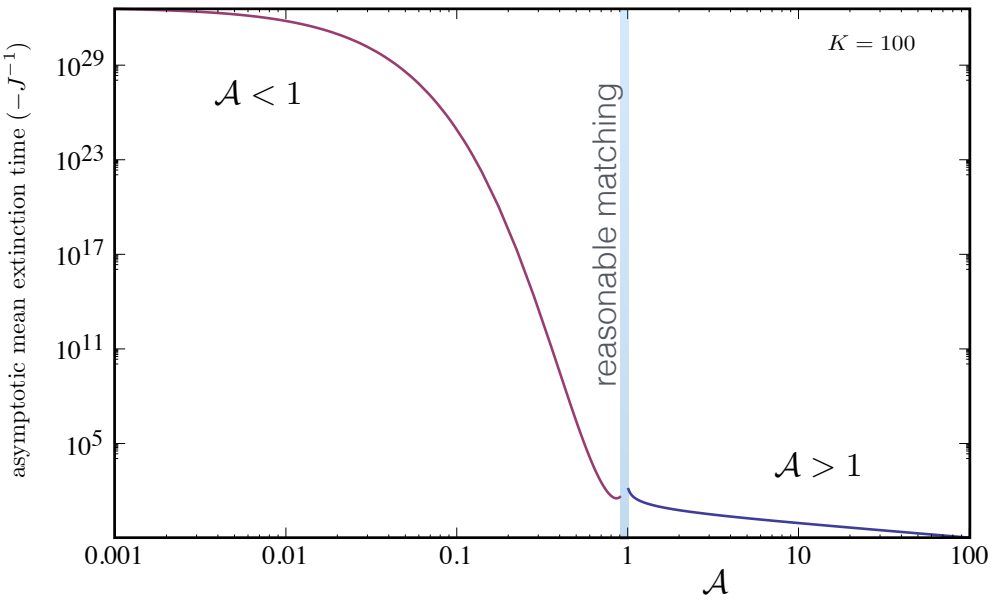

Figure 6: Asymptotic solutions for extinction time in both cases $\mathcal{A} < 1$ and $\mathcal{A} > 1$. We exclude the region near $\mathcal{A}$ as our expansions were not valid there. There is a reasonable degree of matching between the two regimes, which could be improved by utilised an expansion of $\mathcal{A}$ about unity.

Using the above expressions from extinction time, we may approximate the dynamics of the full, environmentally-forced system in the $\alpha \to 0$, small α and $\alpha \to \infty$ limits. Determining whether a minimum extinction time exists or not requires a comparison between

the extinction time at α_{res} and the extinction time at $\alpha \to \infty$, which requires an expression for the extinction time near $\mathcal{A} = 1$, as this is where the numerics suggest the minimum disappears. We do not yet have an expression in this regime. However, such a calculation is fairly straight forward and becomes applicable in the case where low extinction times characterise even the mean state.

4 Stochastic Calculus Form

The goal of this section is to obtain a stochastic differential equation that approximates the evolution of the population size, which now takes the form of a stochastic variable ${X_t}^2$. A well-known result is that the probability density function satisfying a Fokker-Planck equation of the form 24, describes a trajectory given by the so-called Itô Stochastic Differential equation

$$dX_t = f(X_t)dt + \left[g(X_t)/\sqrt{K}\,\right] \cdot dW_t, \tag{47}$$

where W_t describes a Weiner Process, whose time derivative produces delta-correlated Gaussian White Noise. Notice that if we suppose $K \to \infty$, the equation aproaches a deterministic differential equation with the population increasing simply as the different between births and deaths ($f(X_t)$).

The form above suggests that the noise associated with random births and deaths is intrinsically white, or at least can be modelled as such. More specifically, there exist two different interpretations of stochastically-forced systems - Itô and Stratonovich. In the Itô interpretation, one considers the system to be forced by noise that exactly satisfies both the Martingale and Markovian conditions [11]. The former implies that the expectation value at some future time equals the current state, whereas the latter suggests that the future state depends only upon the current conditions. The Stratonovich interpretation supposes that no real noise precisely satisfies these criteria and so deriving the noise term requires considering some real noise before taking the white-noise limit. These two interpretations yield different functional forms for the SDE.

Now suppose we wanted to add environmental forcing to the Itô SDE above. Owing to the 'realness' of environmental forcing, it has previously been suggested that one adds a noise term using Stratonovich calculus [12] thereby mixing the interpretations. In what follows, we do not explicitly make this assumption, but rather, we derive an SDE describing the environmentally forced situation, taking particular limits upon α. What we find is an expression that appears to mix the Itô and Stratonovich interpretations, but only within the appropriate limits, suggesting that merely postulating an extra, Stratonovich term is not adequate in most scenarios.

4.1 Mixed interpretations

In order to proceed we must obtain an approximate solution for the full probability

$$P_n = P_n^+ + P_n^-, \tag{48}$$

[2] As a convention, the subscript 't' in the field of stochastic calculus refers to 'as a function of time', rather than 'time derivative'.

which is easier at first for the discrete case. To do so, we add the Master equations (10) to obtain

$$\begin{aligned}\frac{dP_n^+}{dt} &= \big(M_{mn} + aAO_{mn} - \alpha I_{mn}\big)P_n^+ + \epsilon\alpha I_{mn}(P_n - P_n^+)\\ \frac{dP_n}{dt} &= M_{nm}P_n + aAO_{nm}P_n^+.\end{aligned} \tag{49}$$

First, we define a matrix

$$C_{mn} \equiv -(M_{mn} + aAO_{mn} - \alpha(1+\epsilon)I_{mn}), \tag{50}$$

such that the master equation for the bad state becomes

$$\frac{dP_n^+}{dt} + C_{mn}P_n^+ = \epsilon\,\alpha\,I_{mn}P_n. \tag{51}$$

For convenience, we now make somewhat of an abuse of notation, using C in favour of C_{mn}, in terms of which, the general solution to the equation above is

$$P_n^+(t) - e^{-C(t-t_0)}P_n^+(t_0) = \epsilon\,\alpha\int_{t_0}^{t} e^{-C(t-s)}P_n(s)ds, \tag{52}$$

where it must be remembered that C is subject to the rules of matrix algebra.

As noted above in the eigenvalue problem, we are interested in the quasi-steady behaviour. Accordingly, it is appropriate to take as our lower bound $t_0 \to \infty$, thereby eliminating the boundary term on the LHS. Furthermore, we may solve the integral on the RHS by way of the following arguments. The matrix C possess a full spectrum of positive eigenvalues, however, by taking the limit $t_0 \to -\infty$, we are essentially stating that the time-evolution of P_n is dominated by the lowest (in magnitude) eigenvalues. Consequently, we may suppose that the integral over all past times s on the RHS obtains the majority of its contribution from the recent past, i.e., small $t-s$. Accordingly, we Taylor expand about $s = t$:

$$P(s) \approx P(t) + (s-t)\frac{dP}{ds}\bigg|_t + \mathcal{O}(s-t)^2. \tag{53}$$

Upon substitution into the general solution (52) we integrate to arrive at the result

$$P_n^+ \approx \epsilon\alpha\bigg(C^{-1}P_n + C^{-2}\frac{dP_n}{dt}\bigg). \tag{54}$$

Finally, we may insert this expression into the master equation for P_n to obtain one single equation for the dynamics, which after some rearranging, takes the form

$$\frac{dP_n}{dt} = \Big(1 - aA\epsilon\alpha O_{mn}C_{mn}^{-2}\Big)^{-1}\Big(M_{mn} + aA\epsilon\alpha OC_{mn}^{-1}\Big)P_n. \tag{55}$$

We now carry out similar Taylor expansions as before to obtain the continuum approximation to the above equation. However, in addition, we suppose that α and A are large, a

condition satisfied by the white-noise limit of our chosen environmental forcing. The final expression, up to terms with second derivatives, is

$$\frac{\partial \rho(x,t)}{\partial t} \approx \frac{\partial}{\partial x}\Bigg[-ax(1-\bar{A}-x) + \frac{1}{2K}\frac{\partial}{\partial x}\Big(ax(1+\bar{A}+x)\Big)\Bigg]\rho(x,t) + \frac{1}{2}\frac{\partial}{\partial x}\Big[\sigma\frac{\partial}{\partial x}\sigma\Big]\rho(x,t), \tag{56}$$

where the notation is such that all terms to the left of $\rho(x,t)$ operate upon it. Here, $\sigma^2 = a^2A^2\epsilon/(\alpha(1+\epsilon)^2)$ is the variance of the noise.

There are two crucial things to note about the form above. First, if one truly considers $\alpha = \infty$, then the equation becomes that of the mean state. Importantly, the mean environmental state *acts upon the demographic stochasticity*. However, with α large but finite, an extra term arises on the right that has the form one would infer from a Stratonovich interpretation of white noise with variance σ^2. Accordingly, by prescribing a real source of noise, we could extract the Stratonovich form, but only by making approximations that exclude some of the more interesting aspects of the dynamics, i.e., the the minimum at α_{res}.

Acknowledgements

I would like to thank my two advisors, Charlie Doering and Glenn Flierl for their invaluable mentorship during the Woods Hole GFD Summer School. Charlie's ability to convey detailed concepts was rivalled only by his dedication as a first baseman. Glenn's rational insight always helped simplify problems, even those we hadn't come up with yet. Their combination provided me a very privileged research environment for which I am immensely grateful. Also worth a mention is John Wettlaufer. Throughout the summer, he showed great restraint in keeping a particularly eminent group of professors in line as they vied for attention during seminars. Last but not least I am thankful to Andy Thompson, not just for informing me of the summer school and aiding my application, but also for helping to inspire my interest in geophysical fluid dynamics through two years of excellent teaching and exciting discussion.

References

[1] Shahrezaei, V., Ollivier, J. F., & Swain, P. S. (2008). Colored extrinsic fluctuations and stochastic gene expression. Molecular systems biology, 4(1), 196.

[2] Kamenev, A., Meerson, B., & Shklovskii, B. (2008). How colored environmental noise affects population extinction. Physical review letters, 101(26), 268103.

[3] Assaf, M., Roberts, E., Luthey-Schulten, Z., & Goldenfeld, N. (2013). Extrinsic noise driven phenotype switching in a self-regulating gene. Physical review letters, 111(5), 058102.

[4] Gardiner, C. W. (1985). Handbook of stochastic methods (Vol. 4). Berlin: Springer.

[5] Cairns, B. J., Ross, J. V., & Taimre, T. (2007). A comparison of models for predicting population persistence. ecological modelling, 201(1), 19-26.

[6] Ribalet, F., Swalwell, J., Clayton, S., Jimnez, V., Sudek, S., Lin, Y., ... & Armbrust, E. V. (2015). Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. Proceedings of the National Academy of Sciences, 201424279.

[7] Chang, H. H., Oh, P. Y., Ingber, D. E., & Huang, S. (2006). Multistable and multistep dynamics in neutrophil differentiation. BMC cell biology, 7(1), 11.

[8] Khasin, M., Meerson, B., Khain, E., & Sander, L. M. (2012). Minimizing the population extinction risk by migration. Physical review letters, 109(13), 138104.

[9] Doering, C. R., & Gadoua, J. C. (1992). Resonant activation over a fluctuating barrier. Physical review letters, 69(16), 2318.

[10] Doering, C. R., Sargsyan, K. V., & Sander, L. M. (2005). Extinction Times for Birth-Death Processes: Exact Results, Continuum Asymptotics, and the Failure of the Fokker–Planck Approximation. Multiscale Modeling & Simulation, 3(2), 283-299.

[11] Moon, W., & Wettlaufer, J. S. (2014). On the interpretation of Stratonovich calculus. New Journal of Physics, 16(5), 055017.

[12] Hakoyama, H., & Iwasa, Y. (2000). Extinction risk of a density-dependent population estimated from a time series of population size. Journal of theoretical biology, 204(3), 337-359.

Punctuated Plume Penetration: Entrainment Dynamics of the Layered Filling Box

Yana Bebieva

October 15, 2015

1 Introduction

Many field observations indicate a layered stratification in the Global Ocean. For example, double-diffusive staircase structures are prevalent across the Arctic Ocean inhibiting different depth regions: above the Atlantic Water layer at around 400 m and in the deep Arctic above the homogeneous bottom layer. There are proposed scenarios on how brine-enriched shelf water plumes can penetrate towards the deep Arctic (see e.g., [1, 2]) causing slow deep water ventilation. Presumably, while descending such dense plumes pass through the regions of a layered stratification that can modify the plume dynamics. The influence of the layered stratification on plume dynamics and a feedback mechanisms have not been investigated yet.

Another example where plume interacts with the layered stratification is melting of marine-terminating glaciers. Recently it has been shown that glacial melting rate can be constrained by considering interaction of turbulent plume with a linearly stratification in a presence of repeating layered intrusions [11].

In this study by conducting a series of experiments and using the numerical model, we examine how layered stratification can change the plume dynamics and how plume can modify the initial stratification.

This report is organized as follows. In the next section, we describe the experimental technique. We describe the observed dynamics of plume penetration through the stratified environment and interaction with the layers in section 2.2. In section 3, we present the developed numerical model and discuss model outputs. Findings are summarized and discussed in section 4.

2 Experiments

2.1 Experimental procedure

We have conducted a series of experiments to understand the interaction between layered stratification and a turbulent plume. In these experiments we used: a square tank (42 cm $\times$ 42 cm $\times$ 70 cm), a plume source, a conductivity–temperature (CT) probe and a light projector and a camera to create shadowgraph images (Figure 1).

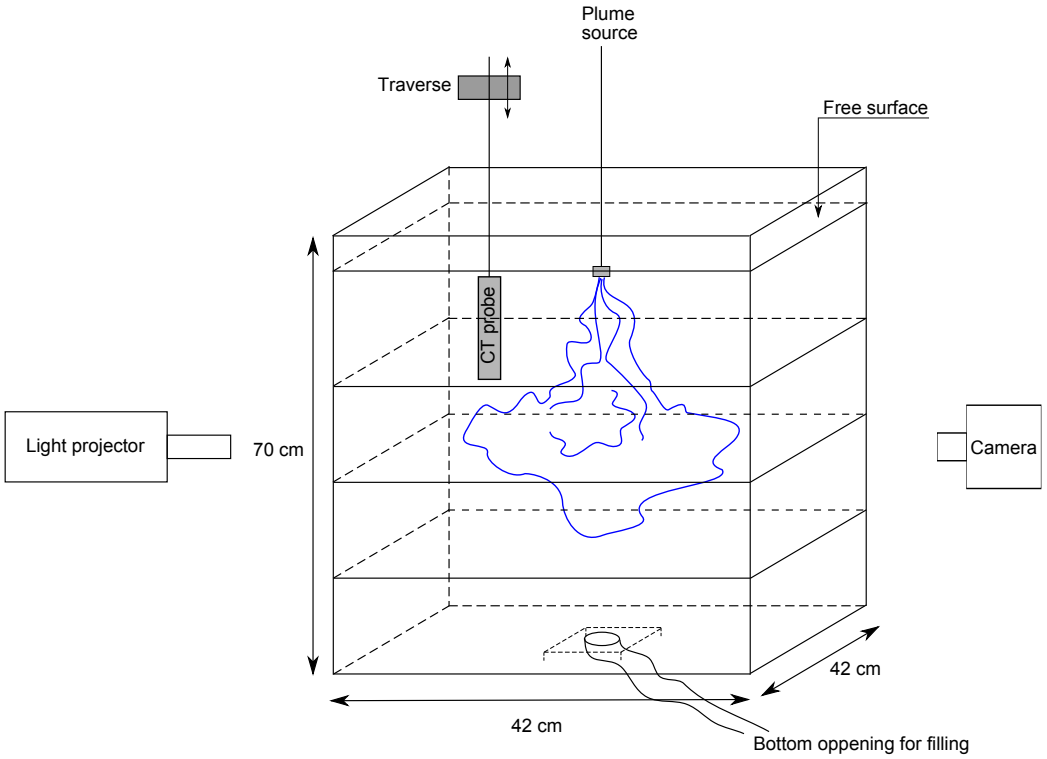

Figure 1: A schematic of the experimental set-up showing a square tank with layered stratification. Blue line shows a salt solution plume, released in the middle of the tank. Evolution of stratification is measured by a traversing conductivity–temperature (CT) probe from the depth where a plume released till maximum depth CT probe can reach, covering the total depth of about 30 cm. This schematic shows the initial stage of the experiment when the plume has been just released and depth of the free surface corresponds to the location of the plume source.

To establish a layered stratification we prepared solutions with different density (i.e., salinity) and fill the tank through a bottom opening layer by layer starting with fresh water and using a very low flow rate to reduce turbulent mixing and create sharp interfaces. The bottom opening was covered by an elevated metallic plate to spread incoming flow horizontally. With such technique an average height of interfaces is about 1–1.5 cm. We used one value of bulk stratification $N_0 \approx 1.13$ s^{-1}, apply two different flow rates ($Q_1 \approx 0.95$ cm^3/s and $Q_2 \approx 1.9$ cm^3/s) and choose three layer depths of 10 cm, 5 cm and 2.5 cm (that becomes effectively a linear stratification after some time due to diffusive mixing). In total we had 6 experiments.

Salt solution plumes (with 20% salinity) were released in the middle of the tank and the CT probe was located at a diagonal position about 1/4 from one of the corners in order to sample density evolution in time and not to have influence from turbulent mixing due to plume propagation. CT measurements have a vertical resolution of ~ 0.6 mm and only down-going profiles were used in the analysis because the CT sensors are located at the base of the profiler so that they are affected by turbulence in the wake of the instrument

in the up-cast profiles. The total profiled depth is about 30 cm, spanning from the depth where a plume is released till ∼20 cm from the bottom of the tank. It takes ∼90 s to finish one cycle of sampling.

In addition to the data derived from the CT measurements we used shadowgraph technique of the flow pattern visualization. Pictures were taken every 30 seconds, allowing to sample the first front propagation with 3 times higher frequency compared to the CT measurements. We used the Hough transform [8] to identify locations of interfaces in each image and calculate rate of first front descent. This method allows to distinguish only well-separated interfaces. At the moments when an interface reaches the one bellow its location is no longer recognizable and there are some missing data. Signal reappears again when this bottom interface starts to move downwards from its original position (see Figure 2).

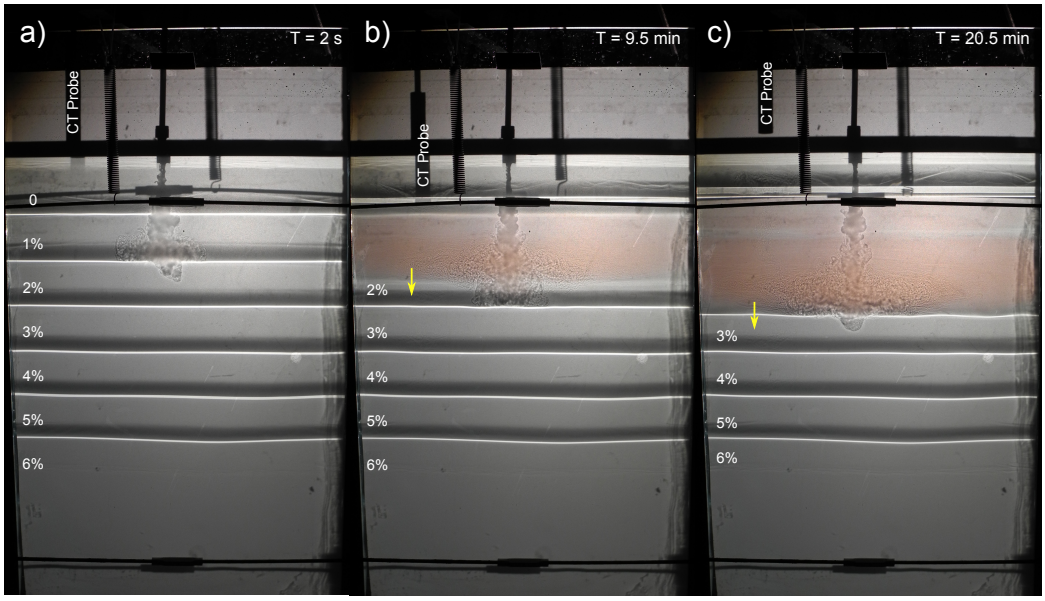

Figure 2: Shadowgraph (corresponding to the experiment $N_0 \approx 1.13$ s^{-1}, $Q_0 \approx 1.9$ cm^3/s, $h = 5$ cm) showing evolution of the background stratification due to interaction with the dense (salinity of 20%) plume at a) 2 s, b) 9.6 min and c) 20.5 min after the beginning of the experiment. Salinity is marked in percentage for each layer. Yellow arrows represent a direction of the first front propagation. Reddish color shows contaminated fluid. An uncontaminated layer (see text) is visible in c).

2.2 Experimental results

2.2.1 Qualitative observations

In order to examine the impact of plume on layered stratification, we begin by characterizing the qualitative observations of mixing within the layers and at the interface.

After a plume has been released, it entrains ambient fluid as it descends. Depending on the initial buoyancy flux the plume can pass through a few layers (interfaces) before the buoyancy flux becomes zero (at the depth z_F, note that at the very first moment of

the experiment this level always coincides with one of the artificially created interfaces) and then it spreads out laterally. Some portion of the plume, however, penetrate further till the depth where momentum flux is zero (z_M, this level does not necessarily coincide with an interface as shown in Figure 3, see e.g., Figure 2a, 2c). In the overshooting region (depth range between z_F and z_M) penetrative entrainment process is taking place and the plume fluid is mixes with the more dense ambient fluid beneath z_F. This mixture eventually ascends towards its neutral density level that is close to z_F but bellow this depth. Thus, the penetrative entrainment modifies the ambient stratification by introducing lighter fluid in place of the original dense fluid within the mixed layer (Figure 4a). Depth where the ambient stratification (after some time from the beginning of the experiment) begins to differ from the original is defined as a penetrative depth (z_p) of the descending front (Figure 3,4). The other process that happens as the plume fluid spreads out at z_F is a formation of a stable stratification above this level according to a "filling box" model described by [3].

As time evolves, a plume head in an overshooting region reaches the next artificial interface and entrains even more dense fluid from the layer below that interface (as shown in Figure 3). This accelerates the propagation rate of the descending front, since each time portion of the dense fluid in the final mixture due to penetrative entrainment increases and neutral density level is moving downward faster compared to the previous stage when penetrative entrainment operates only within the mixed layer. Meanwhile, due to the filling box process density of the plume fluid at z_F increases and z_F approaches z_P. At the moment when $z_F = z_P$, the density of the plume fluid becomes equal to the density of a remainder portion of the initial mixed layer and the plume punches through this remnant towards the next artificial interface leaving behind an uncontaminated background fluid (Figure 3, 4a).

This process continues, however, only first ~ 2 uncontaminated layers are formed as the remainders from the initial mixed layers. When plume propagates deeper z_F can not catch up with z_P any more, because it takes a longer time to accumulate density through a filling box mechanism and penetrative entrainment takes over and erodes the background stratification in a way, that every time when z_P approaches an artificial interface an actual jump in density is even larger than it was in the original stratification (Figure 4b).

2.2.2 Quantitative observations

It is of interest to examine the variations in center of mass of the contaminated fluid and the rate of the descending front across the set of experiments. The center of mass can be expressed as

$$Z_{CM} = \frac{\int(\rho(z,t) - \rho(z,0))z dz}{\int(\rho(z,t) - \rho(z,0))dz}, \tag{1}$$

where $\rho(z,0)$ is density at the beginning of the experiment ($t = 0$) and $\rho(z,t)$ is density after some time, z is the vertical coordinate. Figure 5 indicates that data from all experiments fall around the same curve showing that during the first instants the center of mass moves towards the plume source because all added plume fluid is accumulated within the top layers. Later in time, when the plume fluid is transported down into deeper regions, the center of mass begins to move away from the plume source. The descend rate of center of

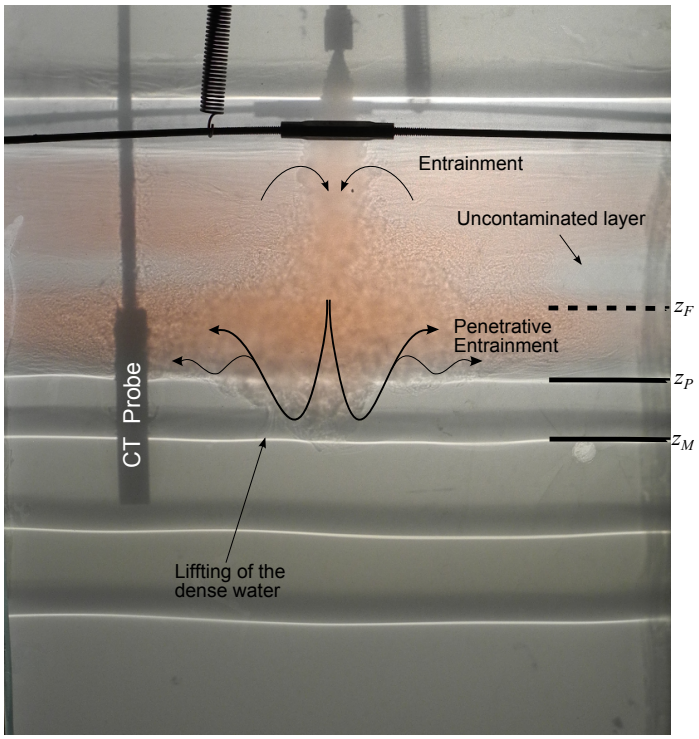

Figure 3: Example of a shadowgraph image showing two different types of mixing of plume fluid with the ambient water: entertainment from the sides of the plume and penetrative entertainment from the base of the plume). Uncontaminated layer is moving upward (see text for details). Location of the penetration depth (z_P), the depth where momentum flux is zero (z_M) are shown by solid black lines and approximate location of the depth where buoyancy flux is zero (z_F) is shown by dashed line.

mass is about the same for all experiments, meaning that layered stratification does not influence the the bulk property of the propagation of the contaminated fluid within the basin.

To compute the second important measure of the distribution of contaminated fluid in the basin, the descend rate of the penetration level, we used data derived from shadowgraph images (as described in section 2.1). The rate of penetration of the descending front in the experiments with a linear stratification (Figure 6) follows $t^{1/2}$ that is in agreement with "stratified filling box" experiments performed by [5], where continuous linear stratification was used. In the experiments with the layered stratification, z_p descends with a changing rate depending on the region of the profile where penetrative entrainment takes place. When ambient fluid entrains only from one mixed layer, z_P propagates very slowly, however, once the penetrative entrainment encompasses some portion of the layer below (in other words, when z_M crosses next artificial interface), the descend rate increases. At the early stages of the experiment (Figure 6, at time around 5 and 9) the plume punches through the remnant of the original layer as described in section 2.2.1 and a jump in propagation of z_P is observed. Note that the depth of this jump may contain some artificial signal due to the method we

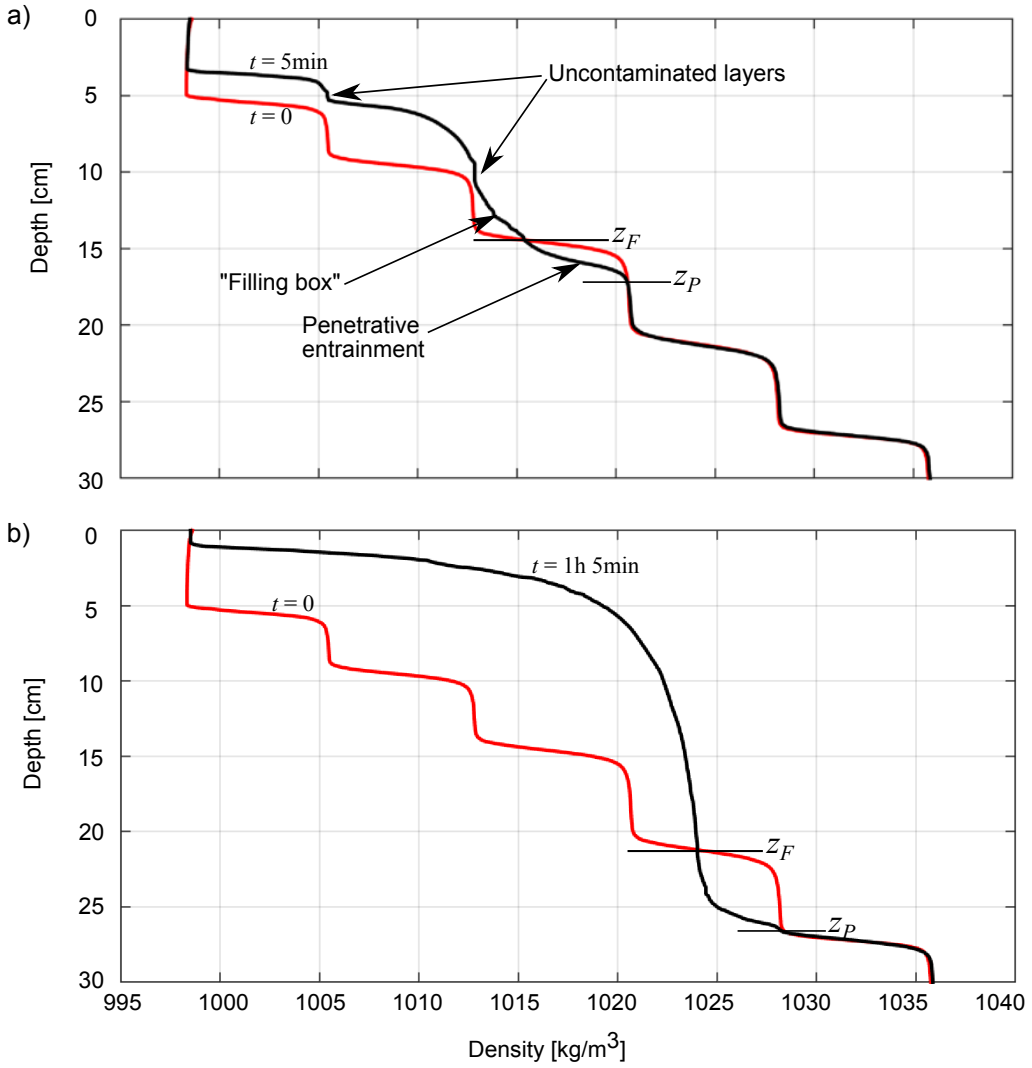

Figure 4: Density profiles measured by a CT probe in the experiment with parameters $N_0 \approx 1.13$ s^{-1}, $Q_0 \approx 1.9$ cm^3/s and $H = 5$ cm at the beginning $t = 0$ (red line) and after a) $t \sim$5 min (red line) and b) $t \sim$1 h 5 min (red line). z_F is the depth where the buoyancy flux is zero, z_P is the penetrative depth of the descending front. Two regions of uncontaminated fluid and portions of the profiles that have been modified from the original due to "filling box" and penetrative entrainment processes are marked in a).

used to detect level z_P (see section 2.1).

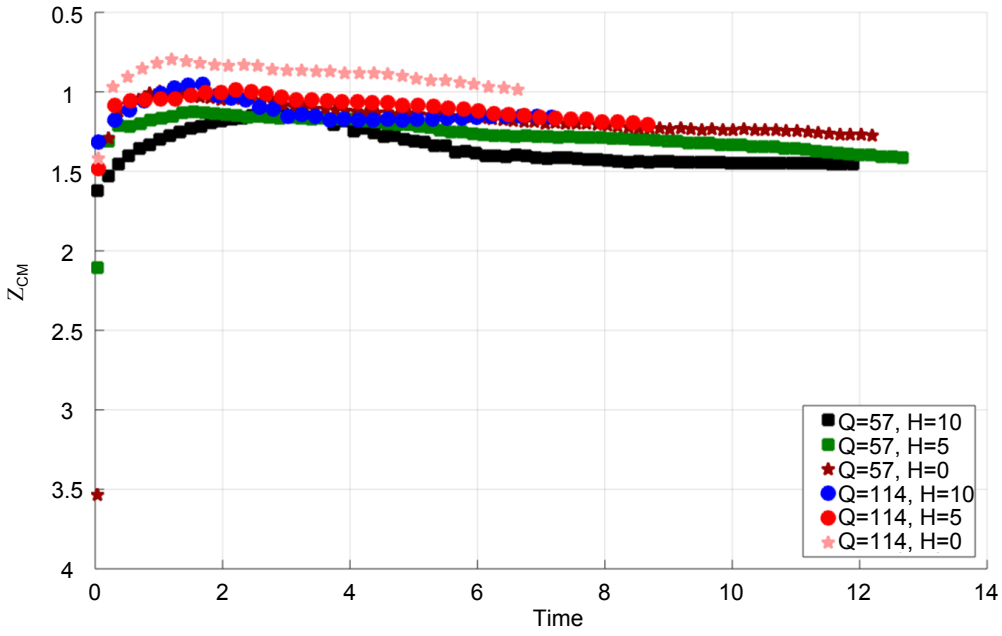

Figure 5: Non-dimensional center of mass versus non-dimensional time across the set of experiments. The nondimensionalization is discussed in section 3.1. For each experiment the flow rate (Q, in cm^3/s) and height of the mixed layers (H, in cm) are shown in the legend. $H = 0$ means linear stratification.

3 Numerical Model

3.1 Formulation of the model

To model the dynamics of the plume in a layered stratified environment we use the theory of turbulent buoyant plumes [10]. Consider conservation equations of mass, momentum and buoyancy averaged over a horizontal cross-section for a steady plume

$$\frac{dQ}{dz} = 2\epsilon_p M^{1/2} \tag{2a}$$

$$\frac{dM}{dz} = \frac{FQ}{M} \tag{2b}$$

$$\frac{dF}{dz} = -QN^2, \tag{2c}$$

where πQ is volume flux, πM is specific momentum flux and πF is buoyancy flux. Buoyancy frequency is defined as

$$N^2 = -\frac{g}{\rho_{ref}}\frac{d\overline{\rho}}{dz}.$$

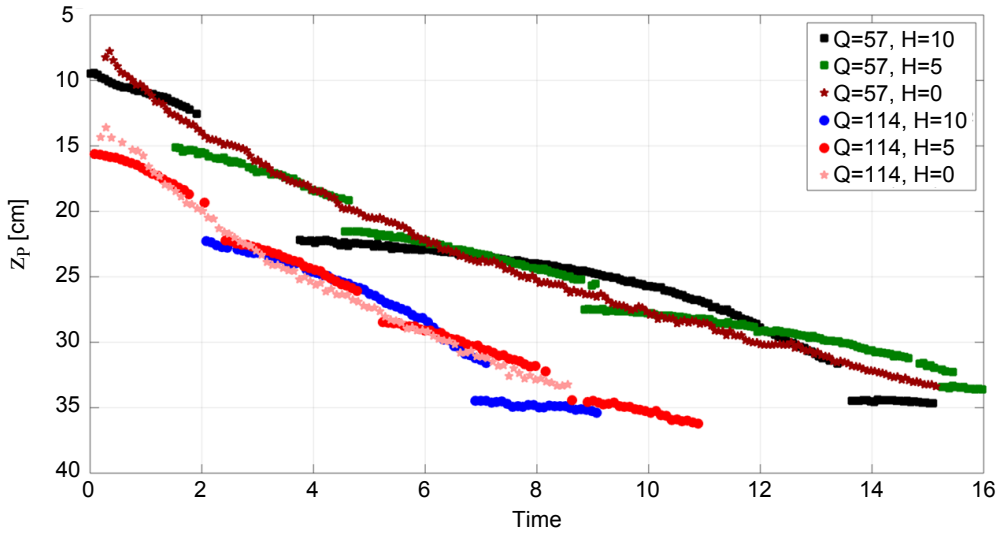

Figure 6: Penetration level (z_P) versus non-dimensional time across the set of experiments. The nondimensionalization is discussed in section 3.1. For each experiment the flow rate (Q, in cm^3/s) and height of the mixed layers (H, in cm) are shown in the legend. $H = 0$ means linear stratification.

where $\frac{d\bar{\rho}}{dz}$ is background bulk density gradient over the staircase.

Following [6] we use a characteristic length scale of the plume height $H_p = (2\epsilon_p)^{-1/2} F_s^{1/4} N_0^{-3/4}$, where ϵ_p is the entrainment constant, F_s is the source specific buoyancy flux, and N_0 is the Brunt–Väisälä frequency determined using the bulk density gradient over a staircase, to nondimensionalize variables as

$$\hat{z} = \frac{z}{H_p} \tag{3a}$$

$$\hat{N} = \frac{N}{N_0} \tag{3b}$$

$$\hat{F} = \frac{F}{F_s} \tag{3c}$$

$$\hat{Q} = \frac{Q}{(2\epsilon_p)^{4/3} F_s^{1/3} H_p^{5/3}} \tag{3d}$$

$$\hat{M} = \frac{M}{(2\epsilon_p)^{2/3} F_s^{2/3} H_p^{4/3}}. \tag{3e}$$

$$\hat{t} = \frac{t}{(2\epsilon_p)^{4/3} F_s^{1/3} H_p^{2/3} A^{-1}}, \tag{3f}$$

where A is a cross section of the basin. The buoyancy flux at the source can be computed using the $F_s = g'_p Q_s$, where πQ_s is the source flow rate and $g'_p = g(\rho_p - \rho)/\rho$ is the reduced

gravity with ρ_p being a density of the plume source fluid. The non-dimensional governing equations are then

$$\frac{d}{d\hat{z}}\hat{Q} = \hat{M}^{1/2} \tag{4a}$$

$$\hat{M}\frac{d}{d\hat{z}}\hat{M} = \hat{F}\hat{Q} \tag{4b}$$

$$\frac{d}{d\hat{z}}\hat{F} = -\hat{Q}\hat{N}^2. \tag{4c}$$

In case of a staircase, it is simplest to construct the model such that equations (4) are evaluated in the mixed layers (letting $N = 0$) and at each interface the buoyancy flux is adjusted as $\hat{F}_{i+1} = \hat{F}_i - \hat{g'}_{i,i+1}\hat{Q}_i$, where $\hat{g'}_{i,i+1}$ is the reduced gravity between two adjacent layers i and $i+1$, and $\hat{Q}_{i,i+1}$ is the volume flux at the interface (all notations are defined in Figure 10). Therefore, we solve the system of equations (4) with the number of the equations equal to the number of a mixed layers in a staircase.

3.2 Model output

Solutions of the equations (4) with initial conditions $Q_s \approx 0$, $M_s \approx 0$ and $F_s = 1$ show that z_M is almost insensitive to the height of a mixed layer (H) when $H < 1.8$ and proportional to the layer height for $H > 1.8$ (Figure 8, 9). The regime shift occurs when the buoyancy flux becomes negative at the very first interface (e.g., see Figure 8d), because in this case regardless of the height of the layer, the plume always reaches the base of the first layer and then overshoots the interface to decrease its momentum to zero. Since the buoyancy flux is constant within the mixed layer and reduces by a finite amount at the interface, the buoyancy flux profile also has a staircase structure. Given this, z_F always equals to the height of one of the interfaces and in the occasions when a plume looses its buoyancy flux within the very first layer, z_F is equal to the layer height. Thus, in Figure 9b) we observe a linear trends with the coefficient of proportionality 1, 2, 3, etc. starting from the most right line.

In our experiment, however, the plume is time-dependent and in a confined basin. Therefore, over the course of the experiment, the background stratification continuously changes due to the "filling box" and penetrative entrainment processes (as discussed in section 2.2.1). To account for the "filling box" process we use a numerical scheme proposed by [7]. In this case, from a conservation of volume in the environment we can compute the location of the top interface of the newly formed layer as $h_{i,j} = h_{i,i+1} - \hat{Q}_{i,i+1}\Delta t$, where h_i is the position of the interface that corresponds to z_F, $\hat{Q}_i$ is the volume flux at this interface and Δt is a time increment in integration. Penetrative entrainment can be also described in terms of formation of a new layer beneath z_F. In this case we assume that the entrained volume flux is linearly proportional to the volume flux of the plume at z_F [9, 4] and the height of the interface beneath z_F can be found from $h_{j,i+1} = h_{i,i+1} + E\hat{Q}_{i,i+1}\Delta t$ (indecies here are counted from the plume source, i.e., $i+1$ is located further from the source compared to j as shown in Figure 10), where E is an empirically determined constant. The density of this layer depends on how many interfaces (layers) are swept be the plume during the overshooting, and it can be computed as a weighted mean. Subsequently, these two constructed

layers are mixed to generate one mixed layer that is added to the original staircase. During this process, one "old" interface $(i, i+1)$ is being replaced by two new interfaces (i, j) and $(j, i+1)$. This leads to an additional equation in the system (4). When the depth of a generated layer is smaller than a chosen threshold l, then this layer collapses; in this case the background density is redistributed whereas the number of equations remains the same. Note also that if a generated layer via the penetrative entrainment process is more dense than a layer below, this causes merging.

In this model we have three turning parameters: side entrainment coefficient ϵ_p, penetrative entrainment coefficient E and a threshold value for layer collapsing l. To find the best combination of these parameters we choose to simulate the experiment with $N_0 \approx 1.13$ s^{-1}, $Q_0 \approx 1.9$ cm^3/s and $h = 5$ cm that clearly shows uncontaminated layers. As described in section 2.1, during the setup processes it was hard to obtain very thin interfaces, and the generated staircases, in fact, are a combination of mixed layers and regions with linear stratification. Thus, to obtain the most realistic model output and compare it with the observations we initialize the model with an experimental profile at time $t = 0$, that has been divided into a number of mixed layers with various heights (i.e., many thin layers in place of an interface, see Figure 11).

Having explored the range of parameters, we concluded that the model with $\epsilon_p = 0.08$, $E = 0.09$ and $l = 0.1$ cm (this value was non-dimensionalized correspondingly to use in a model) has the best representation of the observed physical processes. Figure 11 indicates that the model generates uncontaminated layers and modifies the background density profile due to the penetrative entrainment and the shape of the profile from the model output resembles that from the experiment.

4 Results and Discussion

From the experiments we concluded that the rate of penetration in the layered stratification depends on the location within the staircase, however, overall it follows the penetration rate that is characteristic for the linear stratification. Position of center of mass does not depend whether density profile is linear or consists of layers. The major difference between linear and staircase stratification is that in the second case uncontaminated layers are generated during the first stages of the experiment. CT measurements also show that structure of the density profiles for three examined stratification (layered with 5 and 10 cm layer depth and linear but with the same bulk density gradient) at any particular instant is very different (Figure 7).

Despite this similarity between the model output and observed profiles, further analysis of the numerical simulations has revealed that the model and experimental results differ significantly when penetration depth and location of center of mass are considered. The possible explanation can be in oversimplified parametrization of the penetrative entrainment process. To improve it, one can construct more realistic model where generated layers do not necessary ascend all the way till z_F, causing further merging, instead, these layers spread out at the level of neutral buoyancy. Further, relative contribution of the layers (covered during the overshooting) to the density can be accounted for if the total volume flux ($E\hat{Q}_{i,i+1}$) is distributed over the layers proportionally to their depths and distance from

z_F. Another possible source of error is the effect of internal waves on momentum reduction at the interfaces that is completely neglected in the current version of the model.

This study has demonstrated that although layered stratification in comparison to the linear stratification does not influence the bulk properties (penetration depth, location of center of mass), the profile structure is very different. The distinctive difference between the layered and linear stratification is in the presence of the uncontaminated layers in case of the initial layered stratification.

5 Acknowledgments

I would like to thank Colm-cille Caulfield for giving me the chance to work on this problem and providing continuous supervision throughout the summer. The GFD laboratory at WHOI provided invaluable technical support and I am very grateful to Anders Jensen, who has helped me a lot with laboratory set-up. I would like to thank John Wettlaufer for his technical advises and Karl Helfrich for helping me with data acquisition. I also would like to thank Mary-Louise Timmermans for a lot of insightful discussions. Paolo Luzzatto-Fegiz is thanked for providing the traverse operation codes and I appreciate many helpful suggestions that came from former GFD fellows Georgy Manucharyan, Srikanth Toppaladoddi and Andrew Wells.

References

[1] Knut Aagaard, JH Swift, and EC Carmack, *Thermohaline circulation in the arctic mediterranean seas*, Journal of Geophysical Research: Oceans (1978–2012), 90 (1985), pp. 4833–4846.

[2] Anna Akimova, Ursula Schauer, Sergey Danilov, and Ismael Núñez-Riboni, *The role of the deep mixing in the storfjorden shelf water plume*, Deep Sea Research Part I: Oceanographic Research Papers, 58 (2011), pp. 403–414.

[3] WD Baines and JS Turner, *Turbulent buoyant convection from a source in a confined region*, Journal of Fluid mechanics, 37 (1969), pp. 51–80.

[4] DJ Bower, CP Caulfield, SD Fitzgerald, and AW Woods, *Transient ventilation dynamics following a change in strength of a point source of heat*, Journal of Fluid Mechanics, 614 (2008), pp. 15–37.

[5] Silvana S. S. Cardoso and Andrew W. Woods, *Mixing by a turbulent plume in a confined stratified region*, Journal of Fluid Mechanics, 250 (1993), pp. 277–305.

[6] CP Caulfield and Andrew W Woods, *Turbulent gravitational convection from a point source in a non-uniformly stratified environment*, Journal of Fluid Mechanics, 360 (1998), pp. 229–248.

[7] AE Germeles, *Forced plumes and mixing of liquids in tanks*, Journal of Fluid Mechanics, 71 (1975), pp. 601–623.

[8] John Illingworth and Josef Kittler, *A survey of the hough transform*, Computer vision, graphics, and image processing, 44 (1988), pp. 87–116.

[9] YJP Lin and PF Linden, *The entrainment due to a turbulent fountain at a density interface*, Journal of Fluid Mechanics, 542 (2005), pp. 25–52.

[10] BR Morton, Geoffrey Taylor, and JS Turner, *Turbulent gravitational convection from maintained and instantaneous sources*, in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 234, The Royal Society, 1956, pp. 1–23.

[11] Andrew Wells and Samuel Magorrian, *Turbulent plumes from ice melting into a linearly stratified ocean*, Bulletin of the American Physical Society, 60 (2015).

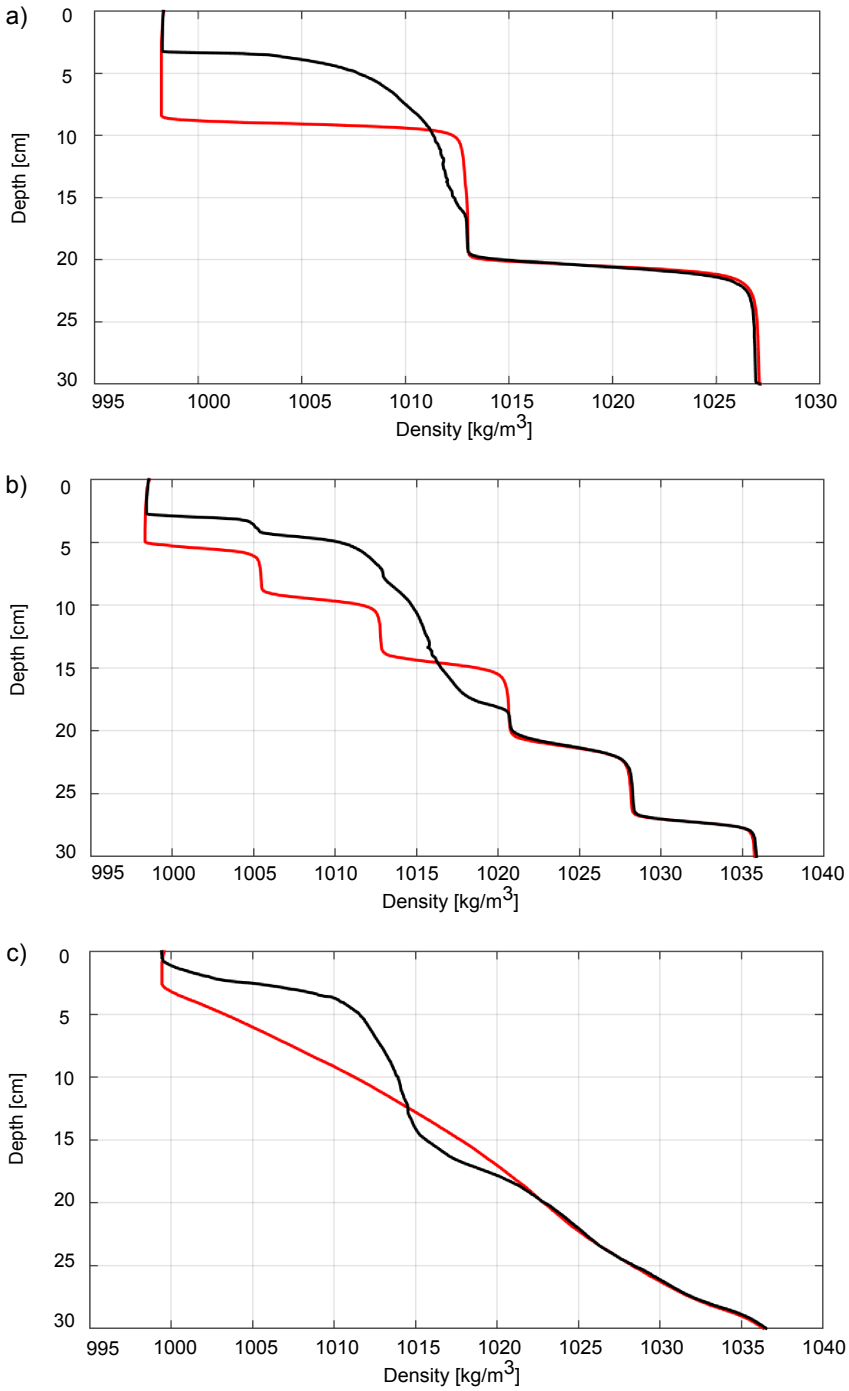

Figure 7: Density profiles measured by a CT probe in the experiment with parameters $N_0 \approx 1.13$ s^{-1}, $Q_0 \approx 1.9$ cm^3/s and various layer height a) $H = 10$ cm, b) $H = 5$ cm and c) linear stratification. Red line shows the initial profiles, black line shows the profiles after 8 min from the beginning of the experiment.

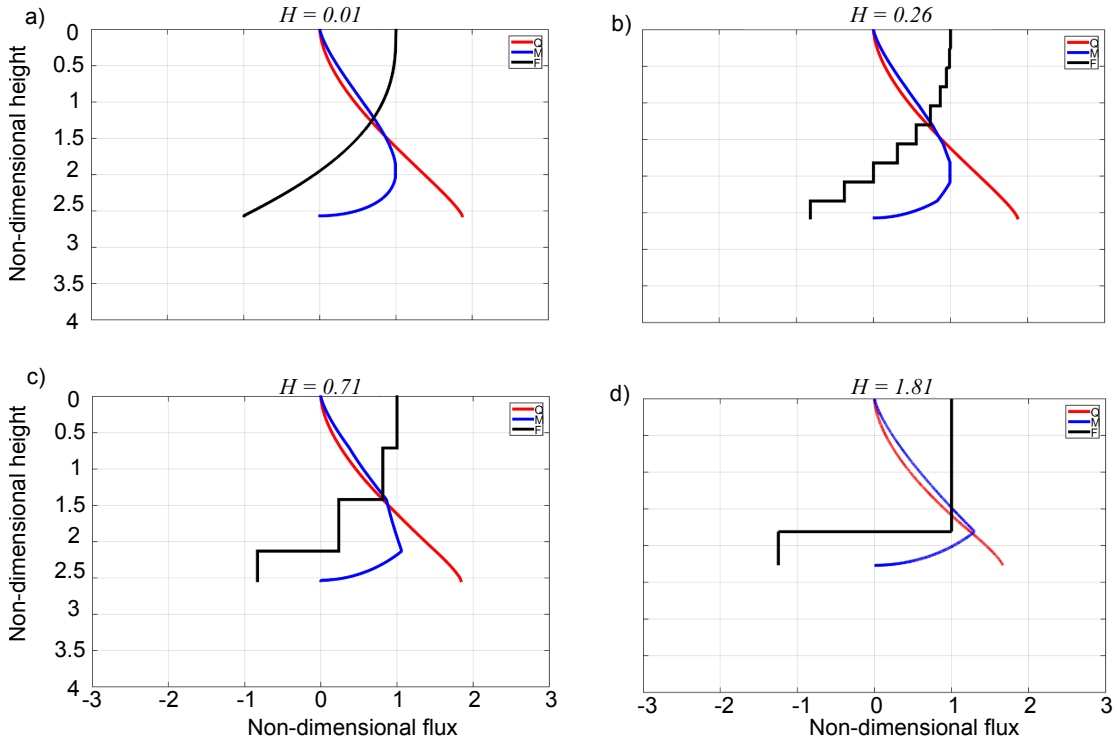

Figure 8: Solution of the equations (4) for volume flux (red), momentum flux (blue) and buoyancy flux (black) applying different non-dimensional height of the layers in a staircase a) $H = 0.01$, b) $H = 0.26$, c) $H = 0.71$ and d) $H = 1.81$. All fluxes are non-dimensional according to (3). Plume source is located at the top of the domain. Initial conditions are $Q_s \approx 0$, $M_s \approx 0$ and $F_s = 1$. All simulations were performed applying the same initial bulk density gradient.

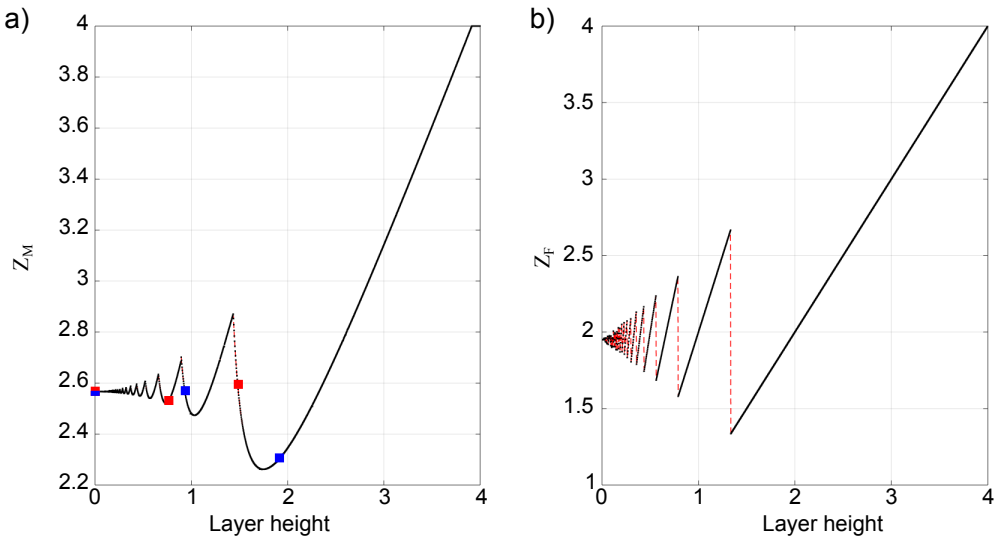

Figure 9: Dependence of a) z_M and b) z_F on nondenominational layer height. In a) red squares show the experiments with $Q_0 \approx 1.9$ cm^3/s and blue squares show the experiments with $Q_0 \approx 0.95$ cm^3/s. The red-blue square represents two experiments with the linear stratification (layer height is zero) and two different flow rates (specified above). Red dashed line serves as a connection line between data points.

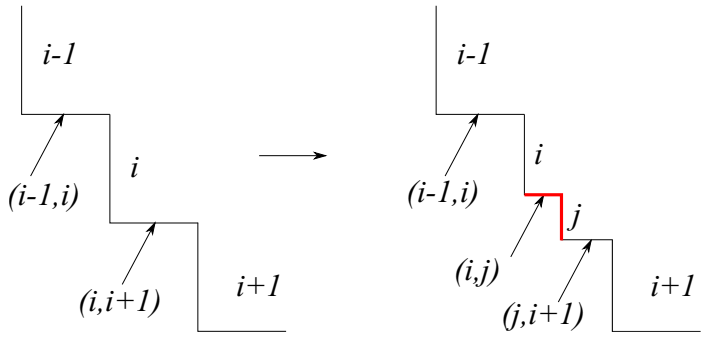

Figure 10: Schematic showing all notations using in description of a new mixed layer (red) generation due to "filling box" and penetrative entrainment processes. All layers are defined by a single index, all interfaces are defined by two indecies of the adjacent layers.

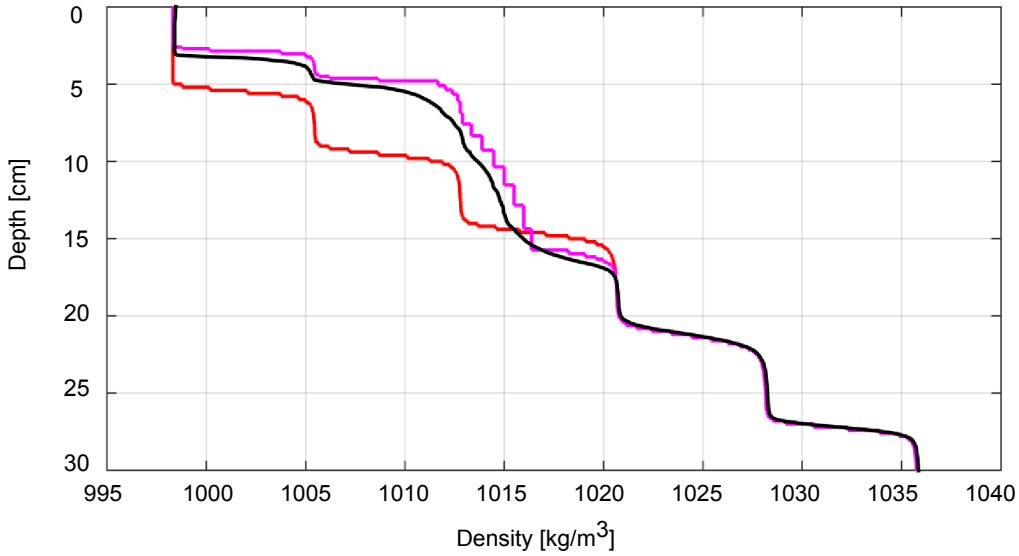

Figure 11: Density profiles measured by a CT probe in the experiment with parameters $N_0 \approx 1.13$ s^{-1}, $Q_0 \approx 1.9$ cm^3/s and $H = 5$ cm. Red line shows the initial profile that has been split into a number of mixed layers to serve as a model input, black line shows the measured profile after ∼7 min from the beginning of the experiment, and magenta line shows the model output after ∼7 min of model time.

An Optimal Exit Time Problem

Florence Marcotte

October 15, 2015

Introduction

Optimisation of mixing for a passive scalar subject to advection-diffusion by an incompressible flow has a long and celebrated story. The question of its measure is a major issue and various metrics have been adopted in order to quantify the mixing efficiency in terms of suppression of the scalar variances by a stirring flow ([13], [2] for a quantification of mixing efficiency on multiple scales, [7], [5], and the review [15]). Using these different metrics, various related problems have been addressed - for instance, optimising either the flow ([8], for an unsteady problem) or a source-sink distribution ([14], steady case) given a fixed energy budget, and deriving sharp asymptotic bounds on the mixing enhancement due to advection ([13], [9], [12]). In the present report we will address the following question: given a fixed kinetic energy constraint and an initial, uniform distribution of particles, what steady, incompressible flow achieves minimum exit time (in L^1-norm) from a cavity with absorbing boundary conditions?

The equation governing the advection-diffusion of a passive scalar c by an autonomous flow with velocity field u and constant, homogeneous molecular diffusivity κ is:

$$\frac{\partial c^\dagger}{\partial t^\dagger} = -u^\dagger \cdot \nabla^\dagger c^\dagger + \kappa \triangle^\dagger c^\dagger. \tag{1}$$

In what follows, the flow is incompressible ($\nabla^\dagger \cdot u^\dagger = 0$) and the magnitude of the velocity field is tuned through a fixed kinetic energy constraint:

$$\left\langle |u^\dagger|^2 \right\rangle = 2E^\dagger, \tag{2}$$

where $\langle \cdot \rangle = \int_\Omega \cdot d\Omega$, so that $\langle | \cdot | \rangle$ denotes the L^1-norm. Using the typical size of the domain as the unit length and a diffusive timescale, (1) then becomes in dimensionless variables:

$$\frac{\partial c}{\partial t} = -u \cdot \nabla c + \triangle c. \tag{3}$$

Note that with this choice of non-dimensionalisation, the magnitude of the dimensionless velocity u can be regarded as a Péclet number $Pe = \frac{UL}{\kappa}$, which quantifies the ratio between diffusive and advective timescales. All the scalings in terms of dimensionless, kinetic energy E can thus be regarded as equivalent scalings in terms of squared Péclet number Pe^2.

1 The Exit Time Problem

A flow with velocity u is prescribed in a cavity where the concentration of released particles satisfies the advection diffusion equation (3) with absorbing (homogeneous Dirichlet) boundary conditions, meaning that particles are removed from the system as soon as they first hit the boundary. Suppose we initially release particles at a location x_0:

$$c(x,0) = \delta(x - x_0), \tag{4}$$

and let $f(t, x_0)$ the probability density for a particle released in x_0 at $t = 0$ to hit the wall (and therefore exit the domain) at time t. The fraction $S(t, x_0)$ of surviving particles after a time t is given by integrating the probability density for particles to exit the domain at later times, and we have:

$$S(t,x_0) = \langle c(x,t,x_0)\rangle = \int_t^\infty f(s,x_0)ds, \tag{5}$$

implying

$$\frac{dS(t,x_0)}{dt} = -f(t,x_0). \tag{6}$$

The expected exit time for particles issued from x_0 at $t = 0$ is then, using the fact that no particle is expected to survive at infinite times ($S(t,x_0) \underset{t\to\infty}{\to} 0$):

$$\begin{aligned} T(x_0) &= \int_0^\infty t f(t,x_0)dt && (7)\\ &= -\int_0^\infty t\frac{dS(t,x_0)}{dt}dt && (8)\\ &= \int_0^\infty S(t,x_0)dt && (9)\\ &= \left\langle \int_0^\infty c(x,t,x_0)dt \right\rangle. && (10) \end{aligned}$$

Let us introduce the linear operator $\mathscr{L} = u\cdot\nabla - \triangle$ and prove that

$$\mathscr{L}^* T = 1, \tag{11}$$

where $\mathscr{L}^*$ is the adjoint of $\mathscr{L}$, which we shall determine to be $\mathscr{L}^* = -u\cdot\nabla - \triangle$. The particles which are released on the boundary are expected to instantly leave the domain, therefore T satisfies the homogeneous Dirichlet boundary condition:

$$T = 0 \qquad \text{on} \quad \partial\Omega. \tag{12}$$

Let us apply $\mathscr{L}$ to $\int_0^\infty c(x,t,x_0)dt$. Using (3) we get:

$$\begin{aligned} \mathscr{L}\int_0^\infty c(x,t,x_0)dt &= \int_0^\infty \mathscr{L}c(x,t,x_0)dt && (13)\\ &= \int_0^\infty -\frac{\partial c}{\partial t}dt && (14)\\ &= c(x,0) && (15)\\ &= \delta(x-x_0). && (16) \end{aligned}$$

From now on we write $\mathscr{C} = \int_0^\infty c(x,t,x_0)dt$, so that (16) reduces to $\mathscr{L}\mathscr{C} = \delta(x-x_0)$, and we assume that the initial distribution of particles is spatially homogeneous. Let U be the unique solution of the equation $\mathscr{L}^*X = 1$ with $X = 0$ on $\partial\Omega$. Then for every x_0,

$$U(x_0) = \langle \delta(x-x_0)U \rangle = \langle \mathscr{L}\mathscr{C}U \rangle = \langle \mathscr{C}\mathscr{L}^*U \rangle = \langle \mathscr{C} \rangle = T(x_0), \tag{17}$$

therefore $U = T$.

The boundary terms in the calculation of the adjoint vanish due to impermeable boundary condition ($u \cdot n = 0$ on $\partial\Omega$), (12) and the fact that $\mathscr{C} = 0$ on the boundary:

$$\langle T\mathscr{L}\mathscr{C} \rangle = \langle T(u\cdot\nabla - \triangle)\mathscr{C} \rangle \tag{18}$$

$$= \langle \mathscr{C}(-u\cdot\nabla - \triangle)T \rangle + \int_{\partial\Omega} (uT\mathscr{C} - T\nabla\mathscr{C} + \mathscr{C}\nabla T)\cdot n d(\partial\Omega) \tag{19}$$

$$= \langle \mathscr{C}\mathscr{L}^*T \rangle . \tag{20}$$

2 Formulation of the Variational Problem

2.1 ... for the minimum exit time

The problem considered here is the minimisation of the particles expected exit time from a cavity with absorbing boundary conditions on the periphery and uniform initial distribution of particles, where we will use the L^1-norm as a metric. In the absence of stirring ($u = 0$), the transport is purely conductive and the mean expected exit time solely depends on the fluid molecular diffusivity κ. As will be seen in the next section, stirring always results in lowering the mean exit time for the particles with our choice of a metric. Note that this result is true for the L^1-norm but not, for example, for the L^∞-norm as demonstrated by [5], who proved that for any 2D, simply connected domain different from a disc, there always exists a flow that enhances the largest exist time compared to the pure conduction case (see Theorem 1.1 in [5]). In the following we will also consider a 2D domain with incompressible flow ($\nabla\cdot u = 0$), and therefore we introduce the stream-function Ψ such that:

$$u_x = -\frac{\partial\Psi}{\partial y} \qquad \text{and} \qquad u_y = \frac{\partial\Psi}{\partial x}. \tag{21}$$

We aim at determining the structure of the flow that realizes optimally efficient stirring, under a given energy constraint. The expected exit time and stream-function have to satisfy:

$$\mathscr{L}^*T = 1, \tag{22}$$

$$\left\langle |\nabla\Psi|^2 \right\rangle = 2E, \tag{23}$$

where we write $\mathscr{L}^*T = -J(\Psi,T) - \triangle T$, with the Jacobian $J(a,b) = \partial_x a\partial_y b - \partial_y a\partial_x b$. We thus define the following functional, to be minimised in order to achieve minimal mean exit time under the above constraints:

$$\mathscr{F}(T,\Psi,\Theta,\mu) = \langle T \rangle - \langle \Theta\left(\mathscr{L}^*T - 1\right) \rangle + \frac{\mu}{2}\left(\left\langle |\nabla\Psi|^2 \right\rangle - 2E \right), \tag{24}$$

where Θ, μ are Lagrange multipliers. With impermeability boundary conditions ($\Psi = 0$ on the wall), we get the first variations of $\mathscr{F}$:

$$\delta_T \mathscr{F} = \langle \delta T \rangle - \langle \Theta \left(\mathscr{L}^* \delta T \right) \rangle = \langle \delta T \rangle - \langle \left(\mathscr{L} \Theta \right) \delta T \rangle , \tag{25}$$

$$\delta_\Psi \mathscr{F} = \langle \Theta J(\delta\Psi, T) \rangle + \mu \langle \nabla(\delta\Psi) \cdot \nabla\Psi \rangle = \langle \delta\Psi J(T, \Theta) \rangle - \mu \langle \delta\Psi \triangle \Psi \rangle , \tag{26}$$

$$\delta_\Theta \mathscr{F} = \langle \delta\Theta \left(\mathscr{L}^* T - 1 \right) \rangle . \tag{27}$$

We look for an extremum by setting the functional derivative of $\mathscr{F}$ to 0, which gives the Euler-Lagrange equations:

$$\mathscr{L}\Theta = 1, \tag{28}$$

$$J(T, \Theta) - \mu \triangle \Psi = 0, \tag{29}$$

$$\mathscr{L}^* T = 1. \tag{30}$$

2.2 ... for an internal heating problem

Let us temporarily drop our previous notations and consider now the steady-state of temperature $\Theta^\dagger$ in an internally heated flow with heat source $Q^\dagger(x)$, divergence-free velocity field $u^\dagger$ and constant, homogeneous molecular diffusivity κ:

$$0 = -u^\dagger \cdot \nabla^\dagger \Theta^\dagger + \kappa \triangle^\dagger \Theta^\dagger + Q^\dagger \tag{31}$$

or equivalently, in dimensionless variables and diffusive timescale:

$$\mathscr{L}\Theta = Q. \tag{32}$$

Let us consider a uniform heating source and fix $Q = 1$. We look for optimum cooling by a 2D flow using the L^1-norm of the temperature as a metric, under incompressibility and fixed energy constraints. Let the functional:

$$\mathscr{F}(\Theta, \Psi, T, \mu) = \langle \theta \rangle - \langle T \left(\mathscr{L}\Theta - 1 \right) \rangle + \frac{\mu}{2} \left(\left\langle |\nabla\Psi|^2 \right\rangle - 2E \right) , \tag{33}$$

where T, μ are Lagrange multipliers (again, let us forget about last paragraph notations!). With impermeability boundary conditions ($\Psi = 0$ on the wall), the first variations of $\mathscr{F}$ are:

$$\delta_\Theta \mathscr{F} = \langle \delta\Theta \rangle - \langle T \left(\mathscr{L} \delta\Theta \right) \rangle = \langle \delta\Theta \rangle - \langle \left(\mathscr{L}^* T \right) \delta\Theta \rangle , \tag{34}$$

$$\delta_\Psi \mathscr{F} = - \langle T J(\delta\Psi, \Theta) \rangle + \mu \langle \nabla(\delta\Psi) \cdot \nabla\Psi \rangle = \langle \delta\Psi J(T, \Theta) \rangle - \mu \langle \delta\Psi \triangle \Psi \rangle , \tag{35}$$

$$\delta_T \mathscr{F} = \langle \delta T \left(\mathscr{L}\Theta - 1 \right) \rangle . \tag{36}$$

Looking for an extremum implies setting the functional derivative of $\mathscr{F}$ to 0, which now yields:

$$\mathscr{L}^* T = 1, \tag{37}$$

$$J(T, \Theta) - \mu \triangle \Psi = 0, \tag{38}$$

$$\mathscr{L}\Theta = 1. \tag{39}$$

Remarkable here is the conjugacy between both systems: under fixed energy constraint, optimization of cooling in the internal heating problem and minimisation of expected time in the exit time problem require solving the same set of Euler-Lagrange equations, the Lagrange-multiplier in the first problem (28)-(30) satisfying the same equation as the passive scalar in the second one (37)-(39) and vice versa. Solving for the exit time problem therefore provides a solution for the internal heating problem as well. We will focus on the former in the next sections, although the later remains an underlying motivation as it may be relevant for many engineering purposes.

3 The System

This section is dedicated to the introduction of the various notations and relationships that will be used in undertaking an analytical solution of the Euler-Lagrange equations. The master equations are expressed in terms of the new variables and the geometry of the domain is then restricted to the unit disc. Finally, the introduction of a particular *Ansatz* results in recasting the variational problem in a 1D, non-linear eigenvalue problem to be solved under fixed energy constraint.

Re-writing (28)-(30) in terms of Jacobians, we now have:

$$-J(\Psi, T) - \triangle T = 1, \tag{40}$$

$$J(\Psi, \Theta) - \triangle \Theta = 1, \tag{41}$$

$$-J(T, \Theta) + \mu \triangle \Psi = 0, \tag{42}$$

to be solved under the energy constraint:

$$\left\langle |\nabla \Psi|^2 \right\rangle = 2E, \tag{43}$$

and homogeneous Dirichlet boundary conditions for T,Θ and Ψ on $\partial\Omega$. Let

$$T = C + \tau \qquad \text{and} \qquad \Theta = C + \sigma, \tag{44}$$

where C is the pure conduction solution in the domain:

$$-\triangle C = 1. \tag{45}$$

Since C verifies (45), the system (40)-(42) becomes after subtraction:

$$-J(\Psi, C) - J(\Psi, \tau) - \triangle \tau = 0, \tag{46}$$

$$J(\Psi, C) + J(\Psi, \sigma) - \triangle \sigma = 0, \tag{47}$$

$$J(\sigma, \tau) + \mu \triangle \Psi = J(\tau - \sigma, C). \tag{48}$$

Introducing the new variables

$$\xi = \tau - \sigma \qquad \text{and} \qquad \eta = \tau + \sigma, \tag{49}$$

in order to simplify the calculations and noting that $J(\tau, \sigma) = \frac{1}{2} J(\xi, \eta)$, we obtain the following system, whose solutions correspond to the extrema of the cost function $\mathscr{F}$:

$$-J(\Psi, \xi) - \triangle \eta = 0, \tag{50}$$

$$-J(\Psi, \eta) - \triangle \xi = 2J(\Psi, C), \tag{51}$$

$$\frac{1}{2} J(\eta, \xi) + \mu \triangle \Psi = J(\xi, C). \tag{52}$$

3.1 A few useful identities

Before we attempt to solve for the system (50)-(52) in terms of ξ, η and Ψ, it will prove useful to write down a few integral relationships obtained from taking the scalar product of the master equations (under their various forms) against the different variables. In this paragraph, we justify that stirring effectively improves the mixing compared to the case where the flow is at rest ($u = 0$) and derive two different expressions for the mean expected exit time which we will need later on.

Let us prove first that stirring enhances mixing with our choice of metric. Starting back from the original equation (40) and multiplying it by T, integration over the domain leads to:

$$\boxed{\left\langle |\nabla T|^2 \right\rangle = \langle T \rangle}. \tag{53}$$

A similar operation on (45) yields

$$\boxed{\left\langle |\nabla C^2 \right\rangle = \langle C \rangle}. \tag{54}$$

As the expected exit time can be decomposed as $T = C + \tau$, we can re-write:

$$\langle T \rangle = \left\langle |\nabla C|^2 \right\rangle + 2 \langle \nabla \tau \cdot \nabla C \rangle + \left\langle |\nabla \tau|^2 \right\rangle. \tag{55}$$

Multiplying (46) by respectively C and τ and integrating over the domain leads to:

$$0 = \quad - \langle CJ(\Psi, C) \rangle - \langle CJ(\Psi, \tau) \rangle - \langle C \triangle \tau \rangle = \quad - \langle CJ(\Psi, \tau) \rangle + \langle \nabla C \nabla \tau \rangle, \tag{56}$$

$$0 = \quad - \langle \tau J(\Psi, C) \rangle - \langle \tau J(\Psi, \tau) \rangle - \langle \tau \triangle \tau \rangle = \quad + \langle CJ(\Psi, \tau) \rangle + \left\langle (\nabla \tau)^2 \right\rangle. \tag{57}$$

Thus

$$\langle \nabla C \cdot \nabla \tau \rangle = - \left\langle |\nabla \tau|^2 \right\rangle \tag{58}$$

and

$$\left\langle |\nabla T|^2 \right\rangle = \left\langle |\nabla C|^2 \right\rangle - \left\langle |\nabla \tau|^2 \right\rangle \qquad \text{i.e.} \qquad \boxed{\langle T \rangle = \langle C \rangle - \left\langle |\nabla \tau|^2 \right\rangle}, \tag{59}$$

which shows that stirring can only lower the L^1-norm for the expected exit time (compared to the purely conductive case).

Multiplying now (50), (51), (52) by, respectively, η, ξ and Ψ and integrating over the domain yields:

$$\left\langle |\nabla \eta|^2 \right\rangle = \langle \eta J(\Psi, \xi) \rangle = \langle \Psi J(\xi, \eta) \rangle, \tag{60}$$

$$\langle \eta J(\Psi, \xi) \rangle + \left\langle |\nabla \xi|^2 \right\rangle = 2 \langle \xi J(\Psi, C) \rangle, \tag{61}$$

$$-\mu \left\langle |\nabla \Psi|^2 \right\rangle = \frac{1}{2} \langle \Psi J(\xi, \eta) \rangle + \langle \Psi J(\xi, C) \rangle. \tag{62}$$

It follows that

$$\left\langle |\nabla \eta|^2 \right\rangle + \left\langle |\nabla \xi|^2 \right\rangle = -2 \langle \Psi J(\xi, C) \rangle, \tag{63}$$

and

$$2E\mu = -\frac{1}{2}\left\langle |\nabla\eta|^2 \right\rangle + \frac{1}{2}\left\langle |\nabla\eta|^2 \right\rangle + \frac{1}{2}\left\langle |\nabla\xi|^2 \right\rangle. \tag{64}$$

We also have

$$4|\nabla\tau|^2 = |\nabla\xi|^2 + |\nabla\eta|^2 + 2\nabla\xi\cdot\nabla\eta, \tag{65}$$

$$4\left\langle |\nabla\tau|^2 \right\rangle = \left\langle |\nabla\xi|^2 \right\rangle + \left\langle |\nabla\eta|^2 \right\rangle, \tag{66}$$

as multiplying (50) by ξ and integrating over the domain yields:

$$\langle \nabla\xi\cdot\nabla\eta \rangle = 0. \tag{67}$$

Therefore

$$\left\langle |\nabla\tau|^2 \right\rangle = E\mu + \frac{1}{4}\left\langle |\nabla\eta|^2 \right\rangle \qquad \text{i.e.} \qquad \boxed{\langle T\rangle = \langle C\rangle - E\mu - \frac{1}{4}\left\langle |\nabla\eta|^2 \right\rangle}. \tag{68}$$

Moreover, we can also write:

$$\langle T\rangle = \langle C\rangle + \frac{1}{2}\langle \eta\rangle + \frac{1}{2}\langle \xi\rangle. \tag{69}$$

Multiplication of (40) and (41) by ξ and integration over the domain yields:

$$-\langle \xi J(\Psi,T)\rangle - \langle \xi\triangle T\rangle = \langle \xi\rangle \qquad \text{and} \quad \langle \xi J(\Psi,\Theta)\rangle - \langle \xi\triangle\Theta\rangle = \langle \xi\rangle. \tag{70}$$

Hence

$$2\langle \xi\rangle = -\langle \xi\triangle\eta\rangle - \langle \xi J(\Psi,\xi)\rangle = \langle \nabla\xi\cdot\nabla\eta\rangle = 0, \tag{71}$$

which, once combined with (69), provides an alternative expression for $\langle T\rangle$:

$$\boxed{\langle T\rangle = \langle C\rangle + \frac{1}{2}\langle \eta\rangle}. \tag{72}$$

3.2 The disc case

From now on the domain is assumed to be a disc of radius 1 and cylindrical coordinates are adopted. The conduction solution in the unit disc is given by $C = \frac{1-r^2}{4}$ and the system above simplifies to:

$$-J(\Psi,\xi) - \triangle\eta = 0, \tag{73}$$

$$-J(\Psi,\eta) - \triangle\xi = \frac{\partial\Psi}{\partial\theta}, \tag{74}$$

$$\mu\triangle\Psi = \frac{1}{2}J(\xi,\eta) + \frac{1}{2}\frac{\partial\xi}{\partial\theta}, \tag{75}$$

with the kinetic energy constraint

$$2E = \left\langle |\nabla\Psi|^2 \right\rangle. \tag{76}$$

3.2.1 Preliminary remark

If we consider small-amplitude perturbation from the pure conduction problem (say $E = \epsilon$), with $\Psi = O(\epsilon)$ and, accordingly, $\tau = O(\epsilon)$, $\sigma = O(\epsilon)$, then the system (46)-(47) becomes at order $O(\epsilon)$:

$$-J(\Psi, C) - \triangle\tau = 0, \tag{77}$$

$$J(\Psi, C) - \triangle\sigma = 0, \tag{78}$$

which implies $\tau + \sigma = o(\epsilon)$ since τ and σ verify the same boundary conditions.
Therefore $\eta = o(\epsilon)$ and from (51)-(52) we get at order $O(\epsilon)$:

$$-\triangle\xi = \frac{\partial\Psi}{\partial\theta}, \qquad \text{and} \qquad \mu\triangle\Psi = \frac{1}{2}\frac{\partial\xi}{\partial\theta}. \tag{79}$$

Combining both equation in order to eliminate one variable leads to the same equation for ξ and Ψ: [1]

$$\mu\triangle^2\xi = -\frac{1}{2}\frac{\partial^2\xi}{\partial\theta^2}, \qquad \mu\triangle^2\Psi = -\frac{1}{2}\frac{\partial^2\Psi}{\partial\theta^2}. \tag{80}$$

As ξ and Ψ also satisfy the same homogeneous Dirichlet BCs, we can look for $\Psi = \lambda\xi$ with λ a proportionality constant, to be determined. For the two equations of (79) to be consistent we need $\boxed{\lambda = (2\mu)^{-\frac{1}{2}}}$. Assuming continuation of the linear solution into the non-linear regime, this observation provides a guess for a class of solutions satisfying the following *Ansatz*.

3.2.2 The *Ansatz*

Let's look for a variable-separated solution of the form:

$$\xi = A(r)\cos(m\theta), \qquad \text{and} \qquad \Psi = B(r)\sin(m\theta), \tag{81}$$

with m an integer. Consistency in (74) requires retaining only $\cos(m\theta)$ terms, hence $\frac{\partial\eta}{\partial\theta} = 0$. Plugging these forms into (73)-(75), we get

$$\frac{m}{r}B'A\sin^2(m\theta) + \frac{m}{r}BA'\cos^2(m\theta) - \frac{1}{r}\frac{d}{dr}(r\eta') = 0, \tag{82}$$

$$\frac{m}{r}B\eta' - \frac{1}{r}(rA'' + A') + \frac{m^2}{r^2}A - mB = 0, \tag{83}$$

$$\frac{\mu}{r}(rB'' + B') - \frac{\mu m^2}{r^2}B - \frac{m}{2r}A\eta' + \frac{m}{2}A = 0, \tag{84}$$

[1] Assuming now separation of variables with $\Psi = \hat{\Psi}(r)\sin(m\theta)$ (or $\Psi = \hat{\Psi}(r)\cos(m\theta)$), we notice that Ψ necessarily verifies (80) if it verifies Helmholtz' equation

$$\triangle\Psi = -\frac{m}{\sqrt{2\mu}}\Psi,$$

as then

$$\triangle^2\Psi = \frac{m^2}{2\mu}\Psi = -\frac{1}{2\mu}\frac{\partial^2\Psi}{\partial\theta^2}.$$

Based on the previous analysis for small-energy stirring, we require A and B be proportional (let $B = \lambda A$) such that

$$B'A = A'B. \tag{85}$$

Then we can re-write (82) as:

$$B'B\lambda^{-1}m = \frac{d}{dr}(r\eta'), \tag{86}$$

hence

$$r\eta' = \frac{m}{2\lambda}B^2, \tag{87}$$

where the integration constant has been set to 0 so as to ensure regularity of η at $r = 0$ (regularity of the velocity field across the disc center requires $B(0) = 0$ for $m \neq 0$). Using $A = B\lambda^{-1}$, (83) and (84) become

$$(r^2B'' + rB') - m^2B - \lambda m r\eta' B + \lambda m r^2 B = 0, \tag{88}$$

$$(r^2B'' + rB') - m^2B - \lambda\frac{m}{2\mu}r\eta' B + \lambda\frac{m}{2\mu}r^2 B = 0, \tag{89}$$

with $\lambda = (2\mu)^{-\frac{1}{2}}$ for (88) and (89) to be equivalent. Then using (87) the ODE on B becomes:

$$\boxed{r^2B'' + rB' + \left(r^2\lambda m - m^2\right)B = \frac{m^2}{2}B^3}. \tag{90}$$

This eigenproblem, with eigenvalue $m\lambda$, has to be solved under homogeneous Dirichlet boundary conditions $B(0) = 0$ and $B(1) = 0$ and the energy constraint:

$$2E = \left\langle |\nabla\Psi|^2 \right\rangle = \iint \left(B'^2 \sin^2(m\theta) + \frac{B^2m^2}{r^2}\cos^2(m\theta) \right) r dr d\theta \tag{91}$$

$$\boxed{2E = \pi \int_0^1 \left(B'^2 + \frac{B^2m^2}{r^2} \right) r dr}. \tag{92}$$

An analytical solution of (90) under the energy constraint and boundary conditions mentioned above will be undertaken asymptotically in the two limiting cases of small ($E \to 0$, or equivalently $Pe \to 0$) and large-energy flows ($E \to \infty$, $Pe \to \infty$).

4 Small E Case: The Linear Problem

In the limit of small-energy flow, the cubic term is omitted from (90) and we recover a Bessel equation:

$$r^2B'' + rB' + \left(r^2\lambda m - m^2\right)B = 0. \tag{93}$$

Rescaling $\tilde{r} = \alpha r$ with $\tilde{B}(\tilde{r}) = B(r)$, and choosing $\alpha = \sqrt{\lambda m}$ yields

$$\tilde{r}^2\tilde{B}'' + \tilde{r}\tilde{B}' + \left(\tilde{r}^2 - m^2\right)\tilde{B} = 0. \tag{94}$$

We look for non-axisymmetric solutions as the $m = 0$ mode evidently has poor mixing properties, the flow only sweeping particles along the iso-contours of C in the purely diffusive case. We expect the optimum flow pattern to favor radial transport so as to quickly expel particles toward the boundary.

For a given mode m, non-singular solutions of (94) are proportional to the Bessel function $J_m(\tilde{r})$ whose (stricly) positive roots determine the constant λ (and therefore μ) so as to meet the homogeneous Dirichlet boundary condition $B(1) = 0$. This requires, for a given m:

$$\sqrt{m\lambda} = j_{m,n} = \sqrt{\frac{m}{\sqrt{2\mu}}} \tag{95}$$

where $j_{m,n}$ is the n-th positive root of the Bessel function $J_m(\tilde{r})$.

Coming back to the linearized ODE, we divide (93) by r and rewrite it as:

$$(rB')' - \frac{m^2}{r}B = -\lambda m r B \tag{96}$$

Multiply by B and integrate in r leads in

$$\begin{aligned}\int_0^1 \lambda m r B^2 dr &= \int_0^1 \left(-B(rB')' + \frac{m^2}{r}B^2\right) dr \\ &= \int_0^1 rB'^2 + \frac{m^2}{r}B^2 dr\end{aligned}$$

since $B = 0$ on $r = 0$ and $r = 1$. Thus with $\tilde{B} = \beta J_m$ we have

$$\frac{2E}{\pi} = \lambda m \int_0^1 rB^2 dr = j_{m,n}^2 \int_0^1 r\tilde{B}(j_{m,n}r)^2 dr = j_{m,n}^2 \beta^2 \int_0^1 r J_m(j_{m,n}r)^2 dr \tag{97}$$

and using the result (see [6]) that $\int_0^1 r J_m(j_{m,n}r)^2 dr = \frac{1}{2}J'_m(j_{m,n})^2$, we finally obtain the amplitude of the solution:

$$\beta^2 = \frac{4E}{\pi j_{m,n}^2} \frac{1}{J'_m(j_{m,n})^2}. \tag{98}$$

The mean expected exit time is given by (68), where in the unit disc case $\langle C \rangle = \frac{\pi}{8}$. Since $B = O(E)$ and $\left\langle |\nabla \eta|^2 \right\rangle = O(E^4)$ (following (87)), this last contribution can be neglected in the calculation of $\langle T \rangle$ and we find:

$$\frac{\langle T \rangle}{\langle C \rangle} = 1 - \mu \frac{8E}{\pi}. \tag{99}$$

Optimal stirring efficiency therefore requires μ be as large as possible. Since

$$\mu = \frac{1}{2}\frac{m^2}{j_{m,n}^4}, \tag{100}$$

this yields

$$\boxed{\langle T \rangle = \frac{\pi}{8} - \frac{m^2}{2{j_{m,n}}^4}E,} \tag{101}$$

and the optimal streamlines pattern necessarily displays a single cell in the radial direction ($n = 1$). For $m = 1$, $n = 1$, we find $\mu = 0.00232$; for $m = 2$, $n = 1$: $\mu = 0.00287$; for $m = 3$, $n = 1$: $\mu = 0.00271$. Thus optimal stirring at low energy budget (or low Péclet number) is realised for $\boxed{(m, n) = (2, 1)}$ (see pattern on figure 1).

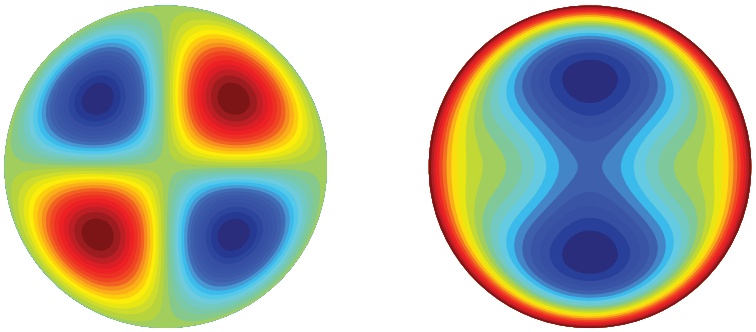

Figure 1: *Left:* Streamlines pattern and *Right:* τ-isocontours for the optimal stirring flow at small energy E, where $\tau = T - C$ is the correction to the purely conductive exit time (dark blue denotes "large", negative values, with $\tau = 0$ on the boundary).

5 The Large Energy Limit

5.1 Fixed m, large E case

For a given mode m, let us first consider the asymptotic behavior of (90) at large energy (the details of the calculations are left to the appendix).

Dominant balance in the bulk of the flow occurs between the eigenvalue term and the cubic one provided $\lambda \sim B^2$, resulting in a linear leading term for the outer solution in the limit of infinitely large E. A peripheral boundary layer, of thickness ϵ, accommodates for the homogeneous Dirichlet boundary condition in $r = 1$, while local analysis reveals a behavior of $B \sim r^m$ near the origin, in a region whose typical thickness goes to 0 as E tends to infinity (this internal layer existing for $m > 1$ only). Neglecting this region in a first approximation, we thus form the following composite solution:

$$B \approx \sqrt{\frac{2\lambda}{m}} r \tanh\left(\sqrt{\frac{m\lambda}{2}}(r-1)\right). \tag{102}$$

The energy constraint on (102) requires $B = O(E^{\frac{1}{3}}) = \epsilon^{-1}$ and determines the value of λ at leading order:

$$\lambda = \left(\frac{9m}{2\pi^2}\right)^{\frac{1}{3}} E^{\frac{2}{3}} = \frac{1}{\sqrt{2\mu}}. \tag{103}$$

We then compute the L^1-norm of the expected exit time at infinitely large E:

$$\langle T \rangle = \langle C \rangle + \frac{1}{2}\langle \eta \rangle = \frac{\pi}{8} - \frac{\pi}{2}\int_0^1 \eta' r^2 dr = \frac{\pi}{8} - \frac{\pi}{4}\int_0^1 \lambda^{-1} m B^2 r dr. \tag{104}$$

Replacing B by the composite solution (102) and splitting the integral at $1-\delta$ with $\epsilon \ll$

$\delta \ll 1$, we roughly get after introducing $x = \sqrt{\frac{m\lambda}{2}}(r-1)$:

$$\langle T \rangle \approx \frac{\pi}{8} - \frac{\pi}{2}\int_0^{1-\delta} r^3 dr - \frac{\pi}{2}\sqrt{\frac{2}{m\lambda}}\int_0^{\sqrt{\frac{m\lambda}{2}}\delta} \tanh^2(x)dx \tag{105}$$

$$\approx \frac{\pi}{8} - \frac{\pi}{8}(1-\delta)^4 - \frac{\pi}{2}\delta + \frac{\pi}{2}\sqrt{\frac{2}{m\lambda}}. \tag{106}$$

Finally

$$\langle T \rangle = \frac{\pi}{2}\sqrt{\frac{2}{m\lambda}} + o(\delta) + o(\epsilon), \tag{107}$$

and, replacing λ by (103), we find at leading order:

$$\boxed{\langle T \rangle = \left(\frac{\pi^4}{6}\right)^{\frac{1}{3}} m^{-\frac{2}{3}} E^{-\frac{1}{3}}}. \tag{108}$$

The calculation of the mean exit time in (105) shows that the contribution of the conductive exit time $\langle C \rangle = \frac{\pi}{8} = O(1)$ is exactly canceled by the flow in the bulk (outer region) provided we assume linear behavior for B everywhere except in the wall boundary layer. On the other hand, the remaining, leading-order mean exit time $O(\epsilon)$ is solely determined by this peripheral boundary layer profile.

The large E, fixed m expected exit time (108) appears to correctly describe the asymptotic behavior of the solution for a given mode as will be seen in the next paragraph. However, asymptotics at fixed m do not provide any evidence for the existence of an optimum flow pattern as the mean exit time goes to 0 if m is chosen arbitrarily large. The existence of an optimum, if any, must result from a "penalty" on large wavenumbers m, which will arise from taking the distinguished limit for large m and large E.

5.2 Some numerics

For different values of the wavenumber m, we solve for the non-linear eigenvalue problem (90) by means of a continuation method, using the dedicated **bvp5c** function ([11]) in MATLAB with λ as a parameter: starting from the Bessel function solution as an initial in the quasi-linear regime (typically $E = 0.001 - 0.1$ depending on m), E is gradually increased up to $E = 10^8$, the output of each computation providing a "first guess" for the next one. [2]

The numerical results are in excellent agreement with the asymptotics and reproduce the scaling (103) of $\lambda = O(E^{\frac{2}{3}})$. As can be seen on figure 2 *(top)*, the numerical mean exit time for a given m perfectly superimposes with the large-energy asymptotics provided E is large enough, with a decay $\sim E^{-\frac{1}{3}}$.

[2] The function bvp5c (as its older version bvp4c) relies on a collocation method where the non-linear eigenvalue problem is solved iteratively by linearizing around the (user-provided) initial solution. This unfortunately prevents one from capturing all the possibly existing solutions (and in particular, flows with more cellular structures in the radial direction) for numerical convergence requires a relatively accurate first guess.

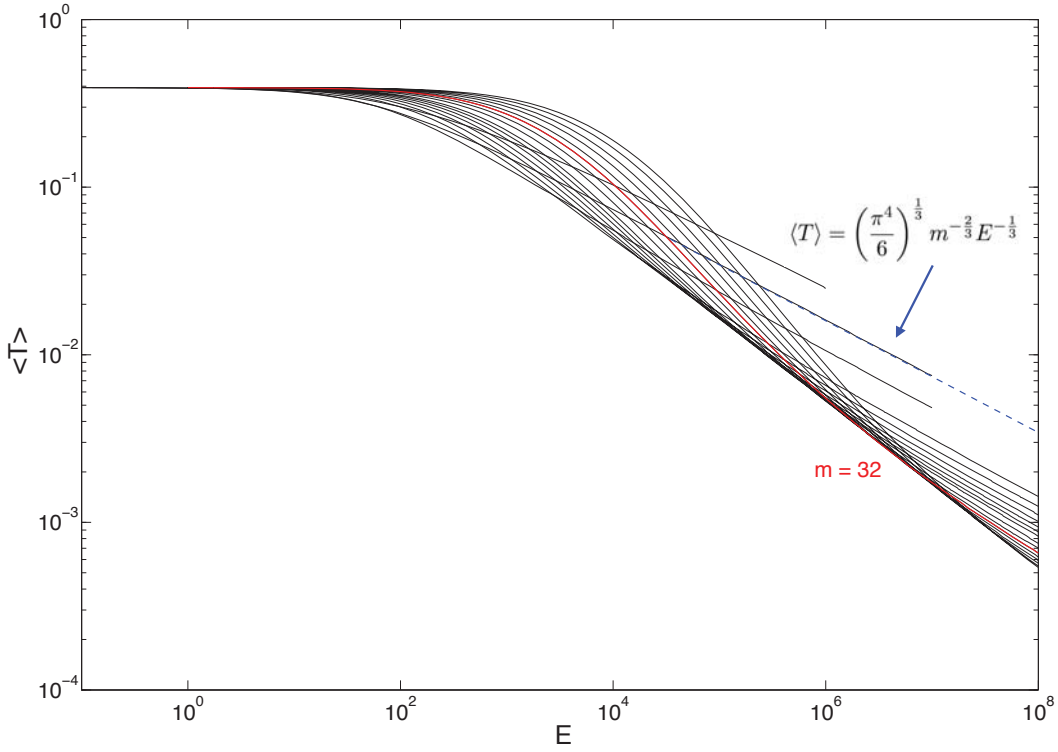

Figure 2: Solid lines: mean expected exit time versus energy (numerical result from bvp5c), for the wavenumber: $m = 1, 2, 4, 8, 10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 42, 48, 56, 64$. Dashed, blue line: large E, fixed m asymptotics, for $m = 2$. As an example, the result obtained for mode $m = 32$ is highlighted in red.

The superposition of mean exit times for various wavenumbers on figure 2 clearly indicates the existence of an optimal m at a given energy E. The minimal mean exit time corresponds to the lower envelope of the various m graphs, whose equation seems to satisfy a power law in E (namely $\sim E^{-\frac{1}{2}}$, as will be shown in the following).

5.3 Large m, large E case

The "penalty" on large wavenumbers actually results from the presence of an internal layer (for $m > 1$): indeed for large m, the streamlines of the divergence-free flow strongly tighten near the center of the disc where diffusion is likely to overpower radial transport. This creates a very-low velocity region (a "stagnant" zone) which widen at fixed E with increasing m - and shrinks at increasing E for fixed m. In this region, the flow is nearly ineffective and the exit time corresponds to the purely conductive one. Hence for a given energy budget, an optimum flow has to be found in combining a large number of cells that efficiently expel particles toward the wall with a stagnation area of limited extent at the center of the disc.

5.3.1 A composite solution for the large-energy flow

The outer solution (bulk): If we consider large wavenumbers and assume that m scales as a power of E, (90) at leading order degenerates into:

$$\left(m\lambda r^2 - m^2\right) B = \frac{m^2}{2} B^3. \tag{109}$$

As in the fixed m case, balancing the eigenvalue and the cubic terms yields $\lambda \sim mB^2$. The difference in the solution arises from the second term in the LHS of (109), whose weight becomes comparable to the one of the first term for a typical radius $r_\times$ such that:

$$r_\times = \sqrt{\frac{m}{\lambda}}. \tag{110}$$

The solution of (109) (outer solution) is then

$$B = \sqrt{\frac{2\lambda}{m}(r^2 - r_\times^2)}. \tag{111}$$

This solution breaks down for a radius of $r \approx r_\times$, which is the typical thickness of the stagnation zone. (Note that we have then $r_\times \sim B^{-1}$, implying that the stagnation zone shrinks as the energy budget increases.) As r approaches 1, we have

$$B \approx \sqrt{\frac{2\lambda}{m}} r \left(1 - \frac{1}{2}\frac{r_\times^2}{r^2}\right) \tag{112}$$

$$\approx \sqrt{\frac{2\lambda}{m}} r \qquad \text{at leading order since } r_\times^2 \sim B^{-2}. \tag{113}$$

The leading terms in (109) as the wall is approached are therefore the first and the third ones, which have to be retained while solving for B near $r = 1$.

The inner solution ($r \to 1$): A boundary layer develops on the wall so as to accommodate for the Dirichlet homogeneous condition on B. Writing ε the typical thickness of this peripheral layer, we rescale the radial coordinate as $r = 1 - \varepsilon\rho$. Expressing (90) in the fast variable ρ yields:

$$(1-\varepsilon\rho)^2\varepsilon^{-2}B'' + (1-\varepsilon\rho)\varepsilon^{-1}B' + m\lambda(1-\varepsilon\rho)^2 B - m^2 B = \frac{m^2}{2}B^3, \tag{114}$$

and retaining the higher order derivative in the dominant balance implies $\lambda m \sim \varepsilon^{-2}$. At leading order, since $\lambda \sim mB^2$ we have

$$\varepsilon^{-2}B'' + (m\lambda - m^2)B = \frac{m^2}{2}B^3. \tag{115}$$

The solution of (115) (inner solution) satisfying the boundary condition in $r = 1$ (or $\rho = 0$) is

$$B = \sqrt{\frac{2\lambda}{m}\left(1 - \frac{m}{\lambda}\right)} \tanh\left(k\rho\right), \tag{116}$$

where $k = \varepsilon\sqrt{\frac{m\lambda - m^2}{2}}$. This inner solution clearly satisfies asymptotic matching with the outer solution (using (111)).

The stagnation zone (internal layer): Local analysis in the vicinity of the center reveals that $B \sim r^m$ as r goes to zero - there the stirring is largely ineffective. For this reason the composite solution for B which is proposed in the next paragraph, where $B = 0$ is assumed everywhere in the stagnation zone, turns out to provide sufficient accuracy for the calculations to come. However, a complete description of the asymptotic solution is much more satisfactory and we aim at capturing the region where the outer solution (111) breaks down as well as its reconnection with the stagnant zone, which seems hard to reconcile with the infinite slope of the bulk solution in $r = r_\times$. Introducing the change of variables $t = \ln\left(\frac{r}{r_\times}\right)$, (90) becomes:

$$B_{tt} + \left(\lambda m r_\times^2 e^{2t} - m^2\right) B = \frac{m^2}{2} B^3. \tag{117}$$

Linearizing around $t = 0$ (or equivalently $r = r_\times$) yields

$$B_{tt} + \left(\lambda m r_\times^2 (1 + 2t) - m^2\right) B = \frac{m^2}{2} B^3 \qquad \text{at leading order,} \tag{118}$$

$$B_{tt} + 2m^2 t B = \frac{m^2}{2} B^3 \qquad \text{as } r_\times = \sqrt{\frac{m}{\lambda}}. \tag{119}$$

Let us rescale the variable $s = \alpha t$ and the unknown function $\beta b(s) = B(t)$. The choice of $\alpha = (2m^2)^{\frac{1}{3}}$ and $\beta = \left(\frac{16}{m}\right)^{\frac{1}{3}}$ results in recovering a second Painlevé's transcendent equation with constant zero:

$$b'' = 2b^3 - sb. \tag{120}$$

Fortunately, this equation does admit a particular solution - namely the Hastings-McLeod solution (see [4], up to a change of sign $x \leftarrow -x$) - which asymptotically satisfies:

$$b_{HM}(s) \underset{s \to -\infty}{\to} 0 \qquad \text{and} \qquad b_{HM}(s) \underset{s \to +\infty}{\sim} \sqrt{\frac{s}{2}}. \tag{121}$$

Moreover, the Hastings-McLeod solution decays toward 0 as a Airy function as $s \to -\infty$, thus displaying the correct behavior for $r \to 0$, and its asymptotic matching with the bulk solution (111) takes care of itself as $r > r_\times$: there

$$B_{HM}(r) \sim \left(\frac{16}{m}\right)^{\frac{1}{3}} \sqrt{\frac{(2m^2)^{\frac{1}{3}}}{2} \ln\left(1 + \frac{r - r_\times}{r_\times}\right)}, \tag{122}$$

$$\sim 2\sqrt{\frac{(r - r_\times)}{r_\times}}, \tag{123}$$

while the bulk solution (111) in the vicinity of $r_\times$ is equivalent to:

$$B(r) = \sqrt{\frac{2}{r_\times^2}(r^2 - r_\times^2)} \underset{r \to r_\times^+}{\sim} 2\sqrt{\frac{(r - r_\times)}{r_\times}}. \tag{124}$$

Nevertheless, the calculation of the asymptotic mean exit time in the next paragraphs will be made considerably simpler by ignoring this last refinement and adopting the expression (125)!

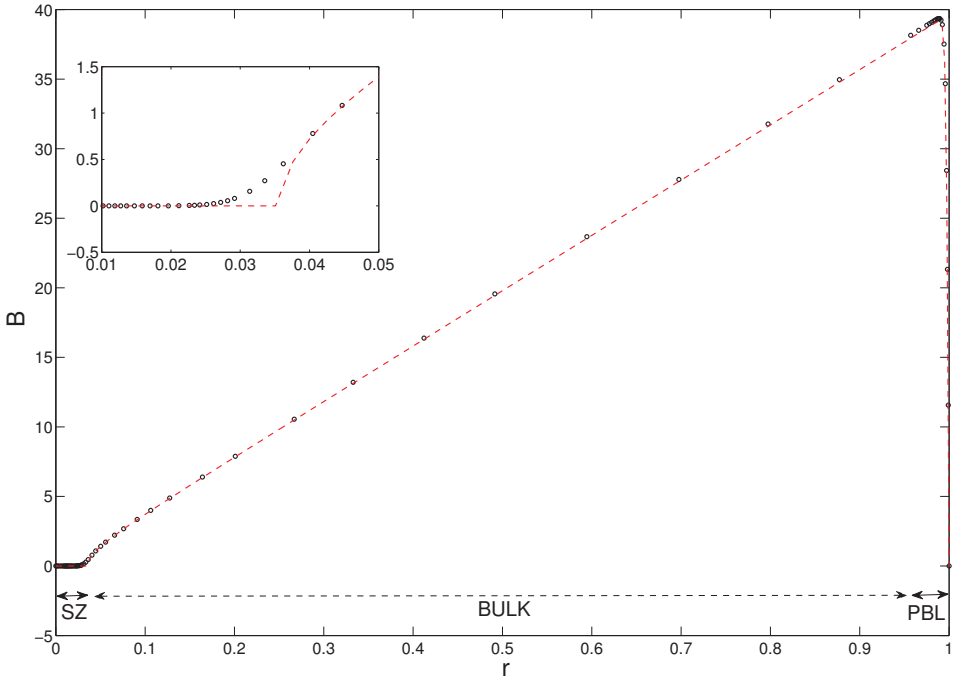

Figure 3: Black circles: numerical solution for $m = 16$, $E = 8.4e5$, computed with bvp5c by a continuation method. Red, dashed line: approximated composite (following (125)) for the corresponding parameters, with the eigenvalue $\lambda(m, E)$ provided by the dispersion relation (142) (see next section). *Inset:* Blow-up on the vicinity of $r_\times$. The different flow regions are indicated on the graph: SZ = internal boundary layer or "stagnation zone", PBL = peripheral boundary layer, BULK = main flow.

A composite solution In the following, let us approximate the full solution by the composite:

$$\boxed{B \approx \begin{cases} \sqrt{\frac{2\lambda}{m}(r^2 - r_\times^2)}\,\tanh\left[k\varepsilon^{-1}(1-r)\right], & \text{if } r_\times < r; \\ 0, & \text{if } 0 < r < r_\times. \end{cases}} \tag{125}$$

Indeed, superposition with the numerical solution of (90) on figure 3 (obtained with the function bvp5c using our continuation method) shows excellent agreement except for the small region restricted to the vicinity of $r_\times$, which we should neglect in the next calculations.

5.3.2 Dispersion relation and optimal exit time

So far the eigenvalue λ is still unknown and its relationship with m and E will be determined by the fixed energy budget, and will in turn be used in the minimisation of the mean expected exit time. Multiplying (90) by B and integrating yields a useful, alternative expression for the energy constraint:

$$rB'' + B' + (m\lambda r - \frac{m^2}{r})B = \frac{m^2}{2}B^3 \tag{126}$$

$$\implies \int_0^1 \left(B(rB')' + (m\lambda r - \frac{m^2}{r})B^2\right) dr = \int_0^1 \frac{m^2}{2}B^4 dr \tag{127}$$

$$\int_0^1 \left(rB'^2 + \frac{m^2}{r}B^2\right) dr = \int_0^1 \left(m\lambda r B^2 - \frac{m^2}{2}B^4\right) dr \tag{128}$$

$$\frac{2E}{\pi} = \int_0^1 \left(m\lambda r B^2 - \frac{m^2}{2}B^4\right) dr. \tag{129}$$

Injecting (125) into the last expression, we find:

$$\frac{2E}{\pi} = m^2 \int_0^1 \left(\frac{\lambda}{m} r B^2 - \frac{1}{2}\frac{B^4}{r}\right) dr = m^2 \int_0^1 r B^2 \left(\frac{1}{r_\times^2} - \frac{1}{2}\frac{B^2}{r^2}\right) dr. \tag{130}$$

Noting that the term inside the brackets can be rewritten as

$$\frac{1}{r_\times^2} - \frac{1}{2}\frac{B^2}{r^2} = \frac{1}{r_\times^2} - \frac{1}{r^2}\left(\frac{r^2}{r_\times^2} - 1\right)\tanh^2(k\rho) \tag{131}$$

$$= \frac{1}{r_\times^2} - \frac{1}{r_\times^2}\tanh^2(k\rho) + \frac{1}{r^2}\tanh^2(k\rho) \tag{132}$$

$$= \frac{1}{r_\times^2}\cosh^{-2}(k\rho) + \frac{1}{r^2}\tanh^2(k\rho), \tag{133}$$

we obtain:

$$\frac{2E}{\pi} = m^2 \int_{r_\times}^1 2r \left(\frac{r^2}{r_\times^2} - 1\right) \tanh^2(k\rho) \left(\frac{\cosh^{-2}(k\rho)}{r_\times^2} + \frac{\tanh^2(k\rho)}{r^2}\right) dr \tag{134}$$

$$= \frac{2m^2}{r_\times^2} \int_{r_\times}^1 \left(r^3 - r r_\times^2\right) \left(1 - \cosh^{-2}(k\rho)\right) \left(\frac{\cosh^{-2}(k\rho)}{r_\times^2} + \frac{\tanh^2(k\rho)}{r^2}\right) dr \tag{135}$$

$$= \frac{2m^2}{r_\times^2} \int_{r_\times}^1 \left(r^3 - r r_\times^2\right) \left(\frac{\cosh^{-2}(k\rho)}{r_\times^2} - \frac{\cosh^{-4}(k\rho)}{r_\times^2} + \frac{1}{r^2} - \frac{2\cosh^{-2}(k\rho)}{r^2} + \frac{\cosh^{-4}(k\rho)}{r^2}\right) dr \tag{136}$$

$$\frac{E}{\pi}\frac{r_\times^2}{m^2} = \int_{r_\times}^1 \left(\underbrace{(r - \frac{r_\times^2}{r})}_{(a)} + \underbrace{\frac{1}{\cosh^2}(k\rho)\left(\frac{r^3}{r_\times^2} - 3r + \frac{2r_\times^2}{r}\right)}_{(b)} + \underbrace{\frac{1}{\cosh^4}(k\rho)(k\rho)\left(-\frac{r^3}{r_\times^2} + 2r - \frac{r_\times^2}{r}\right)}_{(c)} \right) dr. \tag{137}$$

Retaining only the leading term in each contribution (a),(b),(c) yields:

$$\frac{E}{\pi}\frac{r_\times^2}{m^2} = \int_{r_\times}^1 r dr + \frac{\varepsilon}{k}\int_0^{\frac{k}{\varepsilon}r_\times} \frac{(1-\frac{\varepsilon}{k}x)^3}{r_\times^2\cosh^2(x)}dx - \frac{\varepsilon}{k}\int_0^{\frac{k}{\varepsilon}r_\times} \frac{(1-\frac{\varepsilon}{k}x)^3}{r_\times^2\cosh^4(x)}dx \tag{138}$$

$$\approx \frac{1}{2} + \frac{\varepsilon}{k}\int_0^{\frac{k}{\varepsilon}r_\times} \frac{1}{r_\times^2\cosh^2(x)}dx - \frac{\varepsilon}{k}\int_0^{\frac{k}{\varepsilon}r_\times} \frac{1}{r_\times^2\cosh^4(x)}dx \tag{139}$$

$$\approx \frac{1}{2} + \frac{1}{r_\times^2}\frac{\varepsilon}{k} - \frac{2}{3r_\times^2}\frac{\varepsilon}{k} \tag{140}$$

$$\approx \frac{1}{2} + \frac{1}{3r_\times^2}\frac{\varepsilon}{k}. \tag{141}$$

Finally, recalling that at leading order $k = \varepsilon\sqrt{\frac{m\lambda}{2}}$ and $r_\times = \sqrt{\frac{m}{\lambda}}$, we obtain a compact expression for the energy constraint as a function of m and λ:

$$\boxed{\frac{2E}{\pi} = m\lambda + \frac{2\sqrt{2}}{3}\sqrt{\frac{\lambda^3}{m}}}. \tag{142}$$

Similarly, we compute the mean exit time at leading order:

$$\langle T\rangle = \frac{\pi}{8} - \frac{\pi}{4}\int_0^1 \gamma m B^2 r dr \tag{143}$$

$$= \frac{\pi}{8} - \frac{\pi}{2}\int_{r_\times}^1 r^3(1 - \frac{r_\times^2}{r^2})\tanh^2(k\rho)dr \tag{144}$$

$$= \frac{\pi}{8} - \frac{\pi}{2}\frac{\varepsilon}{k}\int_0^{\frac{k}{\varepsilon}r_\times} (1-\frac{\varepsilon}{k}x)^3\left(1 - \frac{r_\times^2}{(1-\frac{\varepsilon}{k}x)^2}\right)\tanh^2(k\rho)dr \tag{145}$$

$$\approx \frac{\pi}{8} - \frac{\pi}{2}(\frac{1}{4} - \frac{r_\times^2}{2} + \frac{\delta_{SZ}^4}{4} + \frac{\varepsilon}{k}) \tag{146}$$

$$\approx \frac{\pi r_\times^2}{4} + \frac{\pi}{2}\frac{\varepsilon}{k}, \tag{147}$$

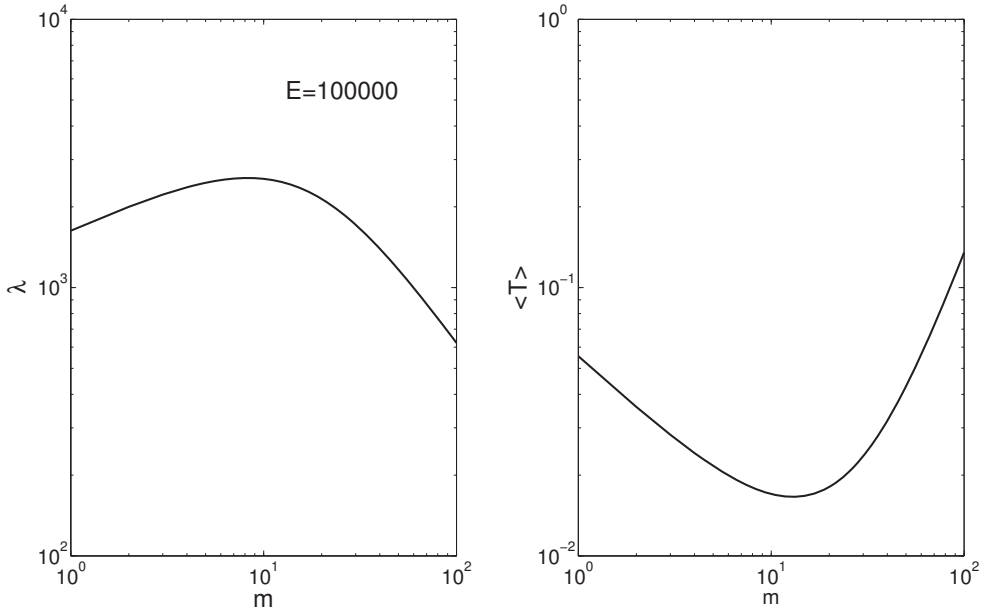

Figure 4: *Left:* eigenvalue λ against azimuthal mode m, for a given kinetic energy budget $E = 10^6$. *Right:* expected exit time (in L^1-norm) against azimuthal mode m.

which in terms of m and λ gives:

$$\boxed{\langle T\rangle = \frac{\pi}{4}\frac{m}{\lambda} + \frac{\pi}{\sqrt{2}}(m\lambda)^{-\frac{1}{2}}.} \tag{148}$$

For a given E, we can already draw - numerically - the graph of λ and $\langle T\rangle$ as a function of m as illustrated on figure 5.3.2.

Considering the energy constraint, we can now guess the scaling for m and λ requiring all the terms to be of the same order. If $m = O(E^\nu)$ and $\lambda = O(E^\beta)$, we get from (142) that $1 = \nu + \beta = \frac{3}{2}\beta - \frac{1}{2}\nu$ i.e.

$$\boxed{\nu = \frac{1}{4}} \qquad \text{and} \qquad \boxed{\beta = \frac{3}{4}}. \tag{149}$$

As a result, we also know the stagnation zone and the peripheral typical thicknesses:

$$\boxed{r_\times = O(E^{-\frac{1}{4}})} \qquad \text{and} \qquad \boxed{\varepsilon = O(E^{-\frac{1}{2}})}. \tag{150}$$

Accordingly, let us write $m = \left(\frac{2E}{\pi}\right)^{\frac{1}{4}}\tilde{m}$, $\lambda = \left(\frac{2E}{\pi}\right)^{\frac{3}{4}}\tilde{\lambda}$ and the conduction-normalised mean exit time $\frac{8}{\pi}\langle T\rangle = \left(\frac{2E}{\pi}\right)^{-\frac{1}{2}}\tilde{T}$. The energy constraint then becomes a dispersion relation between m and λ:

$$1 = \tilde{m}\tilde{\lambda} + \frac{2\sqrt{2}}{3}\frac{\tilde{\lambda}^{\frac{3}{2}}}{\tilde{m}^{\frac{1}{2}}}, \tag{151}$$

and the exit time estimate is now:

$$\tilde{T} = 2\tilde{m}\tilde{\lambda}^{-1} + \frac{8}{\sqrt{2}}(\tilde{m}\tilde{\lambda})^{-\frac{1}{2}}. \tag{152}$$

For the sake of simplification we introduce a new variable $Z = \tilde{\lambda}^{-\frac{1}{2}}\tilde{m}^{\frac{3}{2}}$ and re-write both equations in terms of Z, thus eliminating $\tilde{\lambda}$. This yields respectively

$$1 = \tilde{m}^4\left(\frac{1}{Z^2} + \frac{2\sqrt{2}}{3Z^3}\right) \qquad \text{and} \qquad \tilde{T} = \left(2Z + 4\sqrt{2}\right)\left(1 + \frac{2\sqrt{2}}{3Z}\right)^{\frac{1}{2}}. \tag{153}$$

The energy constraint provides an expression for $\tilde{m}$ as a function of Z and, plugging this into the second expression in (153), we derive the conduction-normalized mean exit time as a function of Z solely. The asymptotic behavior of $\tilde{T}(Z)$ as $Z \to 0$ and $Z \to \infty$ clearly indicates the existence of a global minimum on R_+, which is obtained for

$$\frac{\partial \tilde{T}}{\partial Z} = 0 \qquad \Longrightarrow \qquad 0 = Z^2 + \frac{\sqrt{2}}{3}Z - \frac{4}{3} \qquad \Longrightarrow \qquad Z_{>0} = \frac{2\sqrt{2}}{3}. \tag{154}$$

Therefore the value of $\tilde{m}^*$ and $\tilde{\lambda}^*$ that correspond to optimal flow efficiency are

$$\tilde{m}^* = \sqrt{\frac{2}{3}} \quad \text{and} \quad \tilde{\lambda}^* = \sqrt{\frac{3}{8}} \quad \text{i.e.} \quad m^* = \sqrt{\frac{2}{3}}\left(\frac{2E}{\pi}\right)^{\frac{1}{4}} \quad \text{and} \quad \lambda^* = \sqrt{\frac{3}{8}}\left(\frac{2E}{\pi}\right)^{\frac{3}{4}}. \tag{155}$$

For this optimal mode m^*, the achieved, conduction-normalized mean exit time is then:

$$\boxed{\frac{\langle T\rangle}{\langle C\rangle} = \frac{32}{3}\left(\frac{2E}{\pi}\right)^{-\frac{1}{2}}}. \tag{156}$$

As can be seen on figure 5, (156) is in excellent agreement with the numerical solutions of the eigenvalue problem (90) found with bvp5c: the straight (dashed) line corresponding to the asymptotic optimal exit time perfectly matches the lower envelope of the different modes m for E larger than 10^4 approximately.

5.4 A fast mixer

The result in (156) is expressed in terms of dimensionless energy. Using the fact that the master equations were made dimensionless in diffusive timescale, we have in dimensional variables:

$$E^\dagger = \left\langle |u^\dagger|^2 \right\rangle \propto \kappa^2 E \qquad \text{and} \qquad \left\langle C^\dagger \right\rangle \propto \frac{L^4}{\kappa} \qquad \Longrightarrow \left\langle T^\dagger \right\rangle \propto \frac{L^4}{\kappa}\frac{\kappa}{E^{\dagger\frac{1}{2}}} \tag{157}$$

i.e.

$$\boxed{\left\langle T^\dagger \right\rangle \propto \frac{L^3}{U}}. \tag{158}$$

It is important to emphasize that due to the very particular scaling of (156), the dimensional mean exit time (158) is completely independent of the molecular diffusivity: κ may be

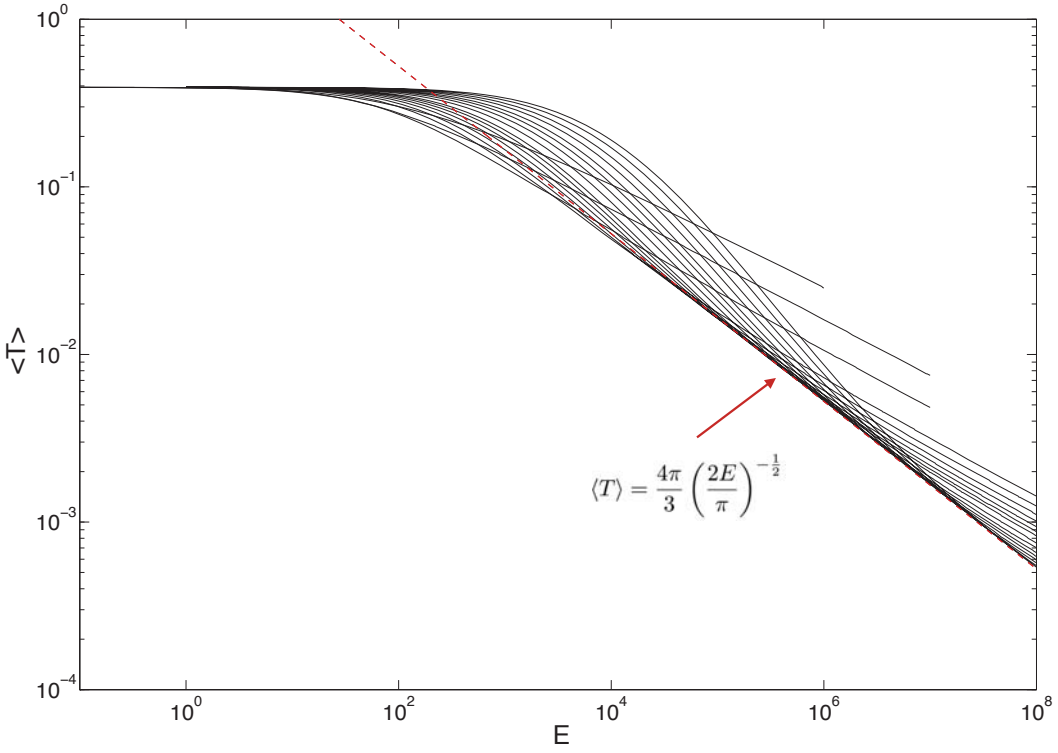

Figure 5: Blue, solid lines: mean expected exit time (from bvp5c) as a function of the kinetic energy, for different wavenumbers (as in figure 2). The black, dashed line represents the asymptotic, optimal exit time as computed from (156).

chosen arbitrarily small (but non-zero) and the particles will be expelled from the domain in purely mechanical time (once they have reached the external boundary layer, diffusivity is still needed for the particles to exit the domain despite impermeability boundary condition). This result is consistent with the bound on mixing efficiency derived in [13] at large Péclet number, which turns out to be independent of the molecular diffusivity (and is expected to hold under turbulent or chaotic mixing). A comparison can be drawn between this physical result and the existence of the so-called fast dynamos (as, for instance, the Ponomarenko ([10]) or the ABC flows dynamo ([1], [3]), whose growth-rate becomes independent of the magnetic diffusivity in the limit of large magnetic Reynolds number.

Although the problem we consider here is quite different from the one addressed for instance by [14] (source optimization), and even though we use a different metric for quantifying the mixing efficiency (they consider the ratio of the L^2-norms for the scalar concentration without and with stirring, see the case $p = 0$ in [14]), we could recast (156) in terms of a "mixing enhancement factor"

$$\mathscr{E} = \frac{\langle C\rangle}{\langle T\rangle} = \frac{3}{32}\left(\frac{2E}{\pi}\right)^{\frac{1}{2}}, \tag{159}$$

and recover a similar linear dependency of the enhancement factor with Péclet number in the asymptotic $Pe \to \infty$ regime.

Conclusion

Although neither our analytical approach (due to the choice of the *Ansatz* (85)), nor the numerical one (due to the use of a continuation method, the routine bvp5c being highly sensitive to the initial) enables us to capture any existing bifurcation in the large-E regime, we exhibit a solution which at least achieves mechanical exit time. Direct numerical simulation would be necessary to ensure that the solution we find with our choice of an *Ansatz*, which is suggested by the small-energy case, truly corresponds to the general optimum over all possible flows, including small-scale, cellular patterns. However, assuming a stronger exponent for E in (156) - say $E^{-\alpha}$ with $\alpha > \frac{1}{2}$ - would yield

$$\left\langle T^\dagger \right\rangle \propto \left(\frac{L^4}{E^{\dagger\alpha}} \right) \kappa^{2\alpha-1}, \tag{160}$$

meaning that the L^1-norm of the expected exit time corresponding to the optimal flow would be enhanced by increasing molecular diffusivity, which seems unconvincing as a higher diffusivity tends to improve mixing. We therefore expect any improvement compared to the solution we find with (85) to deal only with the prefactor in (156). An important source for improvement would presumably reside in considering the mixing enhancement achieved by a time-dependent flow, which we have not attempted, but would be of considerable interest also from an engineering point of view.

Acknowledgements

I would foremost like to thank Bill Young, Jean-Luc Thiffeault and Charlie Doering for their patient guidance and contagious enthusiasm throughout the program. A special thanks goes to Stefan Llewellyn Smith, Basile Gallet and Greg Chini for all the helpful discussions on the blackboard, and to Oliver Bühler and John Wettlaufer for organising such an enjoyable and fruitful summer. Finally, a huge thanks to all the GFD staff members and fellows for the lively debates and the friendly atmosphere of Walsh cottage!

References

[1] Childress, S. & Gilbert, A.D. 1995 *Stretch, Twist, Fold: The Fast Dynamo* Springer, Berlin

[2] Doering, C.R. & Thiffeault, J.-L. 2006 Multiscale mixing efficiencies for steady sources, *Phys. Rev. E*, **74 (2)** 025301(R).

[3] Galloway, D. & Frisch, U. 1986 Dynamo action in a family of flows with chaotic streamlines, *Geophysical & Astrophysical Fluid Dynamics* **36**, 53–83.

[4] Huang, M., Xu, S.-X. & Zhang, L. 2015 Location of Poles for the Hastings-McLeod Solution to the Second Painlev Equation, *arXiv:1410.3338v2 [math.CA]*

[5] Iyer, G., Novikov, A., Ryzhik, L. & Zlatos, A. 2010 Exit times of diffusions with incompressible drift, *SIAM J. Math. Anal.* **42** 2484–98.

[6] Lebedev, N. N. 1972 *Special functions and their applications* Dover Publications, Inc., New-York

[7] Lin, Z., Bodova, K. & Doering, C.R. 2010 Models and measures of mixing and effective diffusion, *Discrete Cont. Dyn. Syst.* **28** 259–74.

[8] Lin, Z., Thiffeault, J.-L. & Doering, C.R. 2011 Optimal stirring strategies for passive scalar mixing, *J. Fluid Mech.* **675** 465–76.

[9] Plasting, S., & Young, W.R. 2006 A bound on scalar variance for the advection-diffusion equation, *J. Fluid Mech.* **552** 289–298.

[10] Ponomarenko, Y. B. 1973 On the theory of the hydromagnetic dynamo, *J. App. Mech. & Techn. Physics* **14** 775

[11] Shampine, L.F., Reichelt, M.W. & Kierzenka, J. 2000 Solving Boundary Value Problems for Ordinary Differential Equations in MATLAB with bvp4c, available at ftp://ftp.mathworks.com/pub/doc/papers/bvp/

[12] Shaw, T. A., Thiffeault, J.-L. & Doering, C.R. 2007 Stirring up trouble: Multi-scale mixing measures for steady scalar sources, *Physica D*, **231 (2)** 143–164.

[13] Thiffeault, J.-L., Doering, C.R. & Gibbon, J.D. 2004 A bound on mixing efficiency for the advection-diffusion equation, *J. Fluid Mech.* **521**, 05–114.

[14] Thiffeault, J.-L. & Pavliotis, G.A. 2008 Optimizing the source distribution in fluid mixing, *Physica D* **237** 918–29.

[15] Thiffeault, J.-L. 2012 Using multiscale norms to quantify mixing and transport, *Nonlinearity* **25** 1–44.

A Large E, fixed m case

In the case of large E, we look at the dominant balance taking place in the bulk of the flow. Introducing the rescaled variables $B = E^\alpha \tilde{B}$ and $\lambda = E^\beta \tilde{\lambda}$, (90) becomes:

$$r^2 \tilde{B}'' E^\alpha + r\tilde{B}' E^\alpha + r^2 \tilde{\lambda} m \tilde{B} E^{\alpha+\beta} - m^2 \tilde{B} E^\alpha = \frac{m^2}{2} \tilde{B}^3 E^{3\alpha}. \tag{161}$$

The only contribution likely to balance the cubic term is the one containing the eigenvalue, requiring $\beta = 2\alpha$. Then at leading order (90) degenerates into:

$$r^2 \tilde{\lambda} m \tilde{B} = \frac{m^2}{2} \tilde{B}^3, \tag{162}$$

which directly gives the outer solution

$$B_o = \tilde{B} E^\alpha = \pm \sqrt{\frac{2}{m}} \tilde{\lambda} E^\alpha r \tag{163}$$

This solution does not satisfy $B(1) = 0$. Proper rescaling of the radial coordinate is required so as to accommodate for the boundary condition: we introduce a new scale $r = 1 - \epsilon\rho$. The inner solution has to satisfy asymptotic matching with the outer solution, which suggests using $B = E^\alpha \bar{B}$ to get

$$\frac{(1-\epsilon\rho)^2}{\epsilon^2} \bar{B}'' E^\alpha + \frac{(1-\epsilon\rho)}{\epsilon} \bar{B}' E^\alpha + (1-\epsilon\rho)^2 \tilde{\lambda} m \bar{B} E^{3\alpha} - m^2 \bar{B} E^\alpha = \frac{m^2}{2} \bar{B}^3 E^{3\alpha}. \tag{164}$$

Dominant balance and the requirement that the higher-order derivative be retained yield $\epsilon = O(E^{-\alpha})$. At leading order, (164) then becomes:

$$\bar{B}'' + \tilde{\lambda} m \bar{B} = \frac{m^2}{2} \bar{B}^3, \tag{165}$$

a solution of which can be found under the form $\bar{B} = \bar{B}_\infty \tanh(k\rho)$ where asymptotic matching with the outer solution as $\rho \to \infty$ leads to $\bar{B}_\infty = \pm\sqrt{\frac{2\tilde{\lambda}}{m}}$ and $k = \pm\sqrt{\frac{m\tilde{\lambda}}{2}}$. At leading order the inner solution is then, up to a change of sign:

$$B_i = \sqrt{\frac{2\tilde{\lambda}}{m}} E^\alpha \tanh\left(\sqrt{\frac{m\tilde{\lambda}}{2}}\rho\right). \tag{166}$$

The energy constraint determines α:

$$\frac{2E}{\pi} = \int_0^1 rB'^2 + \frac{m^2}{r} B^2 dr = \int_0^{1-\delta} (rB_o'^2 + \frac{m^2}{r} B_o^2) dr + \int_{1-\delta}^1 (B_i'^2 + m^2 B_i^2) dr \tag{167}$$

$$\frac{2E}{\pi} = \int_0^{1-\delta} \underbrace{B_\infty^2 r(1+m^2) dr}_{=O(E^{2\alpha})} + \int_{''-\infty''}^0 \left(\underbrace{\frac{B_\infty^2 k^2}{\epsilon} \left(1 - \tanh^2(k\rho)\right)^2}_{=O(E^{2\alpha\epsilon^{-1}})} + \underbrace{m^2 B_\infty^2 \epsilon \tanh^2(k\rho))}_{=O(E^{2\alpha\epsilon})} \right) d\rho \tag{168}$$

Dominant balance requires $\epsilon = E^{-\alpha}$ and $\alpha = \frac{1}{3}$, hence $\beta = \frac{2}{3}$. The leading terms in the energy constraint then determine the factor $\tilde{\lambda}$:

$$\frac{2}{\pi} = \int_{''-\infty''}^{0} \frac{\tilde{B}_\infty^2 k^2}{\epsilon} \left(1 - \tanh^2(k\rho)\right)^2 d\rho = \frac{\tilde{B}_\infty^2 k^2}{\epsilon} \int_{''-\infty''}^{0} \left(1 - 2\tanh^2(k\rho) + \tanh^4(k\rho)\right) d\rho \tag{169}$$

and since $\tanh^2 - \tanh^4 = \tanh^2 \tanh'$, we finally get:

$$\frac{2}{\pi} = \frac{\tilde{B}_\infty^2 k^2}{\epsilon} \left[\frac{\tanh(k\rho)}{k} - \frac{\tanh^3(k\rho)}{3k}\right]_{-\infty}^{0} = \frac{2\tilde{B}_\infty^2 k}{3\epsilon} = \frac{2\sqrt{2}\tilde{\lambda}^{\frac{3}{2}}}{3\sqrt{m}} \tag{170}$$

A Stochastic Approach to Examining the Seasonal Predictability of Arctic Sea Ice

Anna FitzMaurice

October 15, 2015

1 Introduction

1.1 Motivation

In the context of a warming climate, a key question of interest is the fate of the Arctic sea ice. Reductions in the sea ice extent have the potential to greatly affect Arctic communities and ecosystems, and an ice-free Arctic would have a range of commercial and geopolitical implications related to the opening of previously unnavigable shipping routes. More fundamentally, the Arctic sea ice is often viewed as a diagnostic for the state of the warming Earth system, and its decline is one of the most striking indicators of modern climate change. While it is apparent that with sufficient warming the sea ice will eventually all be lost in the Arctic, the specifics of this decline are less well understood. This is related to the fact that superposed on the clear seasonal cycle and general decreasing trend displayed in the satellite record of the Arctic sea ice extent (shown in Figure 1) is significant variability on the annual timescale.

The seasonal growth and decline of Arctic sea ice is tied to the large variations in insolation at high latitudes over the course of the year. During the winter when the daylight hours are short or non-existent, the Arctic ocean cools sufficiently to allow ice to form, attaining its maximum extent in March. As the daylight hours increase through spring and summer, the Arctic warms and the ice melts back to reach its minimum extent in early September. The variability observed on top of this seasonal cycle is in part due to the interannual variability of the underlying climate, but also central to this variability is the existence of strong feedbacks in the sea ice system. Positive feedbacks act to magnify the effect of any small perturbation to the sea ice extent, while negative feedbacks dampen perturbations. It is consequently not just the underlying climate variability that complicates seasonal prediction, but also fluctuations in the strength of these feedbacks on different timescales.

Due to the intrinsic sensitivity of the sea ice system to perturbations, the ability of modern Global Climate Models (GCMs) to predict the Arctic sea ice extent seasonally is limited. In this project, it is hoped that insight might be gained by application of stochastic theory to an idealized sea ice model, which captures the essential physics of the sea ice system while allowing for unmodelled variability by inclusion of a noise term. The physical understanding derived from this study has implications for the interpretation of

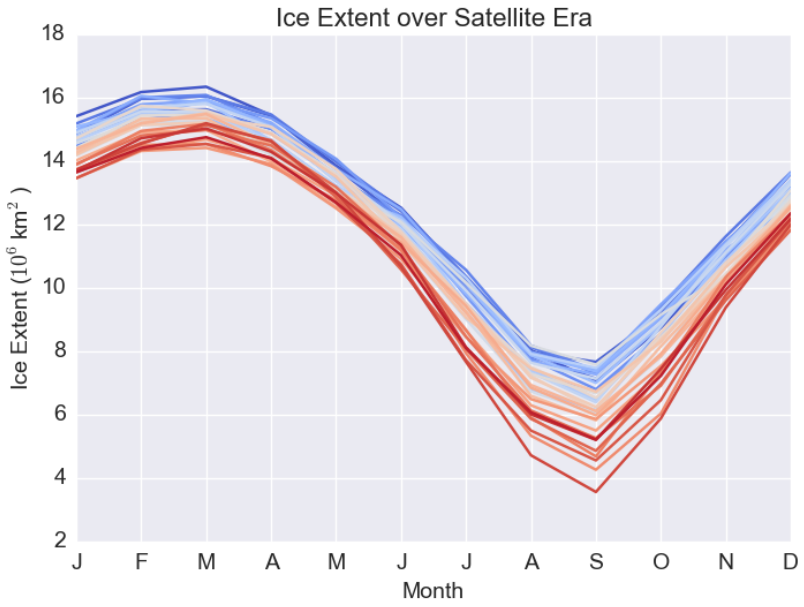

Figure 1: Area extent of Arctic sea ice since the beginning of the satellite era. Earlier years are shown in cold colors, with later years in warmer colors. In addition to the general decreasing trend over the satellite era, there is interannual noise in the ice extent, with a maximum variability displayed in September. (Data Source: NSIDC)

GCM results concerning short term Arctic sea ice prediction, and helps identify the key processes GCMs must capture if they are to be successful at seasonal forecasting.

1.2 Defining predictability

Predictability describes our ability to say something about the state of a system at some future time, given a specific initial condition. Mathematically, there are several ways to quantify what is intuitively understood by predictability. A measure frequently used to make statements about the predictability of Arctic sea ice is the correlation coefficient between sea ice anomalies at a given start month and subsequent months of increasing lag time, as shown for the sea ice extent data from the satellite era in Figure 2A. The decrease in the correlation coefficient with increasing lag is associated with a decline in predictability. However, how the time variation of the correlation coefficient relates to a stochastic interpretation of the system is not immediately obvious. Instead, a more transparent measure of predictability is the standard deviation of anomaly trajectories for a given point in time: if the standard deviation increases from a given start month to a later month, we might expect predictability to decrease due to the divergence of trajectories from nearby initial conditions. On the other hand, if the standard deviation decreases over the forecast period, the convergence of trajectories might be associated with an increase in predictability.

The standard deviation of the ice anomaly (once the seasonal cycle is removed) for the

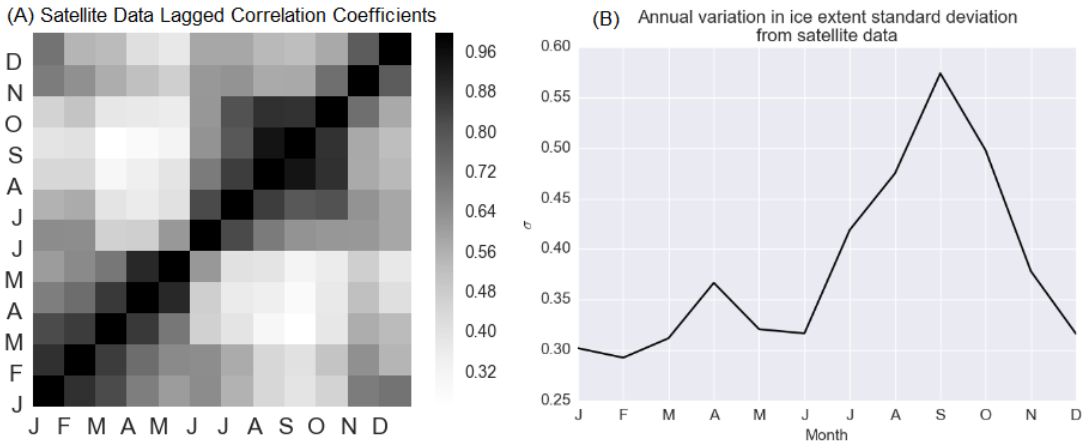

Figure 2: Two measures of predictability for the detrended Arctic sea ice extent satellite data: (A) lagged correlation coefficients, and (B) path standard deviation. (Data Source: NSIDC)

detrended sea ice extent satellite data is plotted in Figure 2B. The standard deviation is seen to increase throughout the year from a winter minimum, to a maximum that coincides with the sea ice minimum in September. This suggests poor predictability of the summer minimum of Arctic sea ice in winter due to the divergence of nearby trajectories going from March to September associated with the increase in standard deviation over this time. In turn, this begs the question of whether a spring "predictability barrier" might exist for the Arctic sea ice, similar to that discussed in the literature for El Niño prediction.

1.3 Model skill

The stochastic interpretation of predictability is closely related to the concept of skill, defined in the modeling community as

$$\text{skill} \equiv 1 - \frac{\sigma_{\text{sim}}^2}{\sigma_{\text{ref}}^2} \ , \tag{1}$$

where σ_{ref} is the reference standard deviation of anomalies for a given period, calculated from a long control run, and σ_{sim} is the standard deviation of runs from a given initial condition. If the initial value problem yields the same variance as the reference, the model has no skill. As the variance of the initialized model ensemble decreases relative to the reference standard deviation, the model's skill is said to increase.

In what follows, we use stochastic theory applied to a simple model of Arctic sea ice to explain the physical origins of variations in the predictability of Arctic sea ice, and to better understand the concept of skill. In Section 2 the simple model is introduced, and the key features of its deterministic solutions explained. Section 3 treats the application of stochastic theory to this model, and finally Section 4 considers the use of GCMs for seasonal prediction in light of the findings of the previous section.

2 A Simple Model for Arctic Sea Ice

2.1 Model description

By working with a simple model of the Arctic sea ice, it is hoped that the fundamental features of the system that act to control the annual variations in predictability might be deduced. The energy balance model of Eisenman & Wettlaufer (2009) [1] (hereon E&W09) is appropriate for this purpose in that it captures the essential physics of the Arctic sea ice, without additional complications. The model is based on a simplified version of the thermodynamic model of [3] coupled to a two-stream model of the atmosphere, which feels a heat flux convergence that depends on the meridional temperature gradient.

The model evolves the single variable E, which describes the latent heat of the sea ice or the heat content of the ocean mixed layer, depending on whether the water is frozen or not. In the case that ice is present, E is determined from the latent heat of freezing L_i times the ice thickness h_i, while in the ice-free case E is the product of the ocean heat capacity c_{ml}, the mixed layer depth H_{ml} (assumed constant), and the ocean temperature T_{ml}. Hence,

$$E \equiv \begin{cases} -L_i h_i & E < 0 \quad \text{(sea ice)} \\ c_{ml} H_{ml} T_{ml} & E \geq 0 \quad \text{(ocean).} \end{cases} \tag{2}$$

The energy variable evolves according to equation

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \left[1 - \alpha(E)\right] F_s(t) - F_0(t) + \Delta F_0 - F_T(t) T(t, E) + F_B + \nu_0 \mathcal{R}(-E), \tag{3}$$

where α is the state-dependent albedo, F_B is the oceanic heat flux to the base of the ice, and $\mathcal{R}$ is the ramp function defined such that $\nu_0 \mathcal{R}(-E)$ represents a constant export of 10% year^{-1} of the sea ice in the case with ice, and is zero otherwise. The terms $F_0 + F_T T$ are the linearization of the Stephan-Boltzmann equation for the emission of longwave radiation about freezing temperature T_{Fr} (adapted to include the dependence of the atmospheric heat flux on the meridional temperature gradient, and to allow the atmosphere to be partially opaque), and ΔF_0 is an adjustable parameter used to apply radiative forcing perturbations to the model (to mimic increased atmospheric CO_2 levels, for instance). For given $F_s, F_0, \Delta F_0, F_T$, and F_B, to close the equations it remains to specify functional forms for α and T.

The albedo of solid ice α_i is very high compared to the substantially lower albedo of open water α_{ml}. As sea ice melts, pools of meltwater appear on the surface, leads open up, and the ice thins and eventually disintegrates. This is modelled in E&W09 by allowing the albedo to decrease smoothly from α_i to α_{ml} via

$$\alpha(E) = \frac{1}{2}(\alpha_{ml} + \alpha_i) + \frac{1}{2}(\alpha_{ml} - \alpha_i) \tanh\left(\frac{E}{L_i h_\alpha}\right), \tag{4}$$

for thickness parameter h_α.

The surface temperature is given by

$$T(t, E) = \begin{cases} -\mathcal{R}\left[\frac{(1-\alpha_i)F_s(t) - F_0(t) + \Delta F_0}{k_i L_i / E - F_T(t)}\right] & E < 0 \\ \frac{E}{c_{ml} H_{ml}} & E \geq 0, \end{cases} \tag{5}$$

which expresses the energy balance in the three possible cases of an ice layer that is below freezing, melting ice, and an ice-free ocean. If the ice is below the freezing point, E and T are negative and ice growth occurs until a balance is reached between the surface radiative fluxes and the upward heat flux in the ice. For ice at the freezing point, ablation may occur and release latent heat, thinning the ice. These two regimes are expressed by the ramp function in equation 5. When all the ice is melted, so $E \geq 0$, the mixed layer is able to absorb energy and raise its temperature (dependent on the mixed layer heat capacity c_{ml} and depth H_{ml}) to maintain radiative equilibrium.

2.2 Behavior of the deterministic model

For a sufficiently low forcing ΔF_0, integration of the deterministic model yields a solution with perennial sea ice, while for high ΔF_0, an ice-free state results. In between, E&W09 find a stable seasonally ice-free state in the full nonlinear model, but show that this solution is unstable in a partially linearized version of the model. Historically it has been proposed that seasonal ice states must be unstable due to the ice-albedo feedback, which is always positive. This has lead to concerns regarding a "tipping point" in the stability of the Arctic sea ice, due to the possibility of hysteresis in the bifurcation diagram of such a system. However, in the nonlinear model E&W09 find that competing longwave effects, which allow thin ice to grow more quickly than thick ice, act to stabilize seasonal sea ice solutions. The underlying structure of the system is still the same though, and additional heating once this seasonally stable ice state has been reached could result in hysteresis behavior as the system jumps to the permanently ice-free state.

A minimal model for a seasonally stable ice cover is analyzed in [4], in which it is deduced that the minimal condition for stability of a seasonal ice state is that the seasonal cycle be broken into more than two periods such that there may be heat loss by the ocean during the period when the summer is ice-free. This follows intuitively from the understanding that the ocean mixed layer must lose its accumulated heat before ice can form.

3 Stochastic Theory & its Implications for Predictability

While the deterministic deterministic E&W09 model captures the essence of the seasonal cycle of Arctic sea ice, it was seen in the satellite data in Figure 1 that the real ice extent is subject to small amplitude variability superposed on this cycle. To simulate this variability, a noise term may be added to the governing ODE of the deterministic model, turning it into the Stochastic Differential Equation (SDE)

$$\frac{\mathrm{d}E}{\mathrm{d}t} = f(E,t) + \sigma\xi, \tag{6}$$

where the term $f(E,t)$ encapsulates all the right-hand side forcings of Equation 3.

3.1 Stochastic perturbation theory

The small magnitude of the noise in the sea ice extent in relation to the amplitude of its seasonal cycle motivates the application of the stochastic perturbation theory developed by

Moon & Wettlaufer (2013) [5], in which the solution to the SDE is sought as a perturbation about the time-dependent ODE solution in terms of small parameter σ, which is the noise magnitude.

Starting from the deterministic equation

$$\frac{\mathrm{d}E_s}{\mathrm{d}t} = f(E_s, t), \tag{7}$$

we suppose that the SDE

$$\frac{\mathrm{d}E}{\mathrm{d}t} = f(E, t) + \sigma\xi \tag{8}$$

is solved by $E = E_s + \eta(t)$, for η small, so that the right-hand side of Equation 8 may be expanded about ODE solution E_s as

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} \approx \left[\frac{\partial f}{\partial E}\Big|_{E_s}\eta + \frac{1}{2}\frac{\partial^2 f}{\partial E^2}\Big|_{E_s}\eta^2 + ...\right] + \sigma\xi \tag{9}$$

$$= c(t)\eta + d(t)\eta^2 + \sigma\xi, \tag{10}$$

where time-dependent parameters c and d are defined such that $c(t) \equiv \partial_E f|_{E_s}$ and $d(t) \equiv \partial_E^2 f|_{E_s}/2$.

It follows from the above that the time evolution of perturbations to the underlying seasonal cycle of the zero-dimensional sea ice model are entirely governed at leading order by the parameters c and d. These parameters vary over the course of the year, and display a seasonal cycle just like the sea ice. Over short time periods for which perturbations remain small and the above theory is valid, c and d may be considered constant to simplify analysis. In what follows, a consideration of the system's behaviour for constant c and d of different signs will allow the roles played by c and d in determining the evolution of perturbations to the sea ice, and their resultant impact on the predictability of the system, to be understood.

3.2 Numerical solution

To integrate Equation 10, the Euler-Maruyama method may be employed. The Euler-Maruyama method is an adaptation of the Euler method for SDEs. At each timestep in the standard Euler integration, a Guassian random variable is drawn from normal distribution of mean zero and standard deviation 1. This value is scaled by the size of the timestep and the noise amplitude, and added to account for the noise term in the SDE. Explicitly

$$\eta_{t+1} = \eta_t + f(\eta_t)\mathrm{d}t + \sigma\mathrm{d}W, \tag{11}$$

where $\mathrm{d}W \sim N(0, \sqrt{\mathrm{d}t})$. If an ensemble of such integrations are run for a given initial condition, each with different realisations of the Guassian noise term at each timestep, a pdf of the solution may be obtained from the spread of the ensemble at a given time. This method is explained in full in [2].

Ensembles of integrations from a given small initial perturbation for constant c and d are shown in Figure 3. For different signs of c, very different behaviour is observed, with the divergence of trajectories for $c > 0$, and convergence for $c < 0$. To understand this behaviour, and the importance of the values of c and d, an analytical solution is sought.

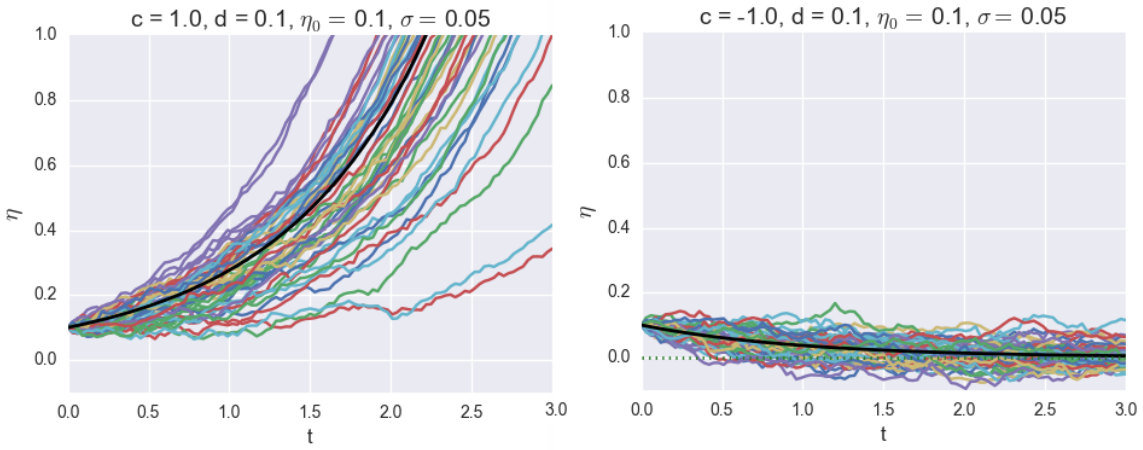

Figure 3: Realisations of integrations of Equation 10 using the Euler-Maruyama method for $c > 0$ (left) and $c < 0$ (right). The ensemble mean is shown in black.

3.3 Analytical solution

Expanding η in powers of the (small) noise magnitude σ as $\eta = \eta_0 + \sigma\eta_1 + \sigma^2\eta_2 + ...$, and considering the equation at subsequent orders of σ,

$$\begin{aligned} O(1): \quad \frac{\mathrm{d}\eta_0}{\mathrm{d}t} &= c(t)\eta_0 + d(t)\eta_0^2 \\ O(\sigma): \quad \frac{\mathrm{d}\eta_1}{\mathrm{d}t} &= \left[c(t) + 2d(t)\eta_0(t)\right]\eta_1 + \xi \\ O(\sigma^2): \quad \frac{\mathrm{d}\eta_2}{\mathrm{d}t} &= \left[c(t) + 2d(t)\eta_0(t)\right]\eta_2 + d(t)\eta_1(t)^2, \\ \vdots \end{aligned} \tag{12}$$

It is immediate that the equation is deterministic at leading order, with solution

$$\eta_0 = \frac{c\eta_0(0)e^{ct}}{c + d\eta_0(0) - d\eta_0(0)e^{ct}}, \tag{13}$$

whose evolution is described by potential $V(t) \equiv -\frac{1}{2}c\eta_0^2 - \frac{1}{3}d\eta_0^3$. The shape of this potential for the different sign cases is shown in Figure 4. As we supposed η is a small perturbation to the deterministic solution and so close to zero, the salient point to take from Figure 4 is that zero is a steady state for $c < 0$, and and unsteady state for $c > 0$. In this respect c determines the stability of the system, explaining the divergence seen in the numerical solution for $c > 0$, and the convergence for $c < 0$. The parameter d, meanwhile, controls the concavity of the potential, and influences the asymmetry of the response to positive versus negative perturbations.

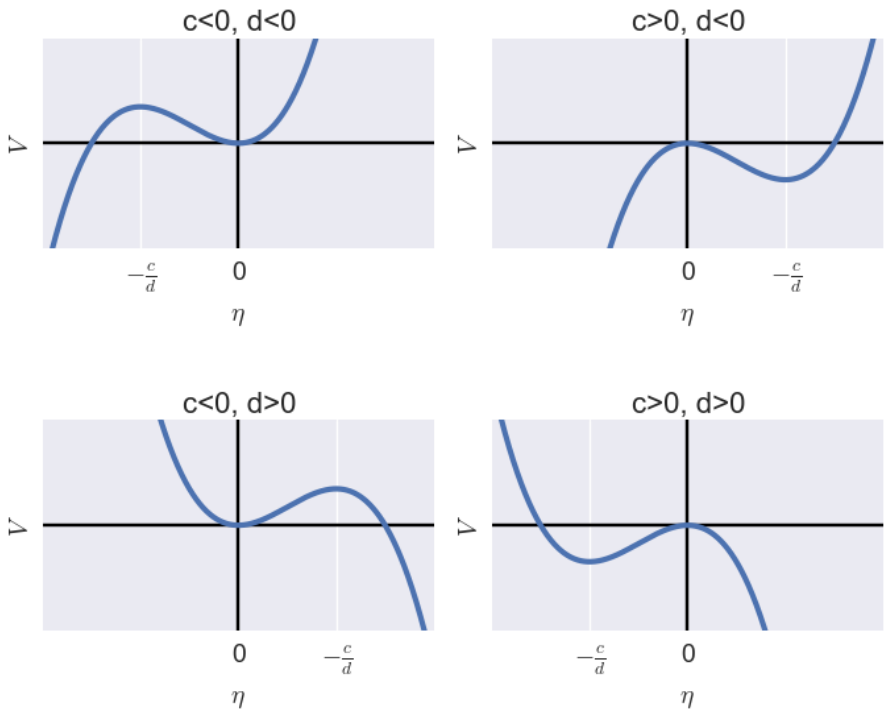

Figure 4: A schematic of the potential governing the evolution of the deterministic first order solution to the SDE.

3.3.1 Order σ

The stochastic term ξ enters the equation at first order in σ, so it is possible to write down the Fokker-Planck equation for the evolution of the pdf ρ of η_1 for a given initial condition:

$$\frac{\partial \rho}{\partial t} = -\big[c(t) + 2d(t)\eta_0(t)\big]\frac{\partial}{\partial \eta_1}(\eta_1 \rho) + \frac{1}{2}\frac{\partial^2 \rho}{\partial \eta_1^2}. \tag{14}$$

This equation is complicated by the fact that the $O(\sigma)$ equation is non-autonomous, due to c, d, and η_0 being time-dependent. To solve, the Fourier transform in η_1 is taken to give

$$\frac{\partial \hat{\rho}}{\partial t} = \tilde{c}k\frac{\partial \hat{\rho}}{\partial k} - \frac{k^2}{2}\hat{\rho}, \tag{15}$$

where $\tilde{c(t)} \equiv c(t) + 2d(t)\eta_0(t)$. The characteristic equations for this PDE are

$$\frac{\mathrm{d}k}{\mathrm{d}t} = -\tilde{c}k \tag{16}$$

$$\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}t} = -\frac{k^2}{2}\hat{\rho}, \tag{17}$$

which integrate to give

$$k = k_0 \exp\left(-\int_0^t \tilde{c}(s)\ \mathrm{d}s\right) \tag{18}$$

$$\hat{\rho} = \exp\left(-\int_0^t \frac{k^2}{2}\ \mathrm{d}s\right). \tag{19}$$

Substituting for k in the equation for $\hat{\rho}$, taking k_0 outside of the integral and then substituting back using $k_0 = k\exp\left(\int_0^t \tilde{c}(s)\ \mathrm{d}s\right)$, $\hat{\rho}$ is found to be

$$\hat{\rho} = \exp\left[-\frac{k^2}{2}\left(e^{2\int_0^t \tilde{c}(s)\ \mathrm{d}s}\int_0^t e^{-2\int_0^s \tilde{c}(r)\ \mathrm{d}r}\ \mathrm{d}s\right)\right]. \tag{20}$$

Finally, on inverting the Fourier transform we have

$$\rho = \frac{1}{\sqrt{2\pi}\sigma_T}\exp\left(-\frac{\eta_1^2}{2\sigma_T^2}\right), \tag{21}$$

where the time-evolving standard deviation σ_T^2 is described by

$$\sigma_T^2 = e^{2I(t)}\int_0^t e^{-2I(s)}\ \mathrm{d}s, \tag{22}$$

$$I(t) = \int_0^t \big[c(s) + 2d(s)\eta_0(s)\big]\ \mathrm{d}s. \tag{23}$$

From this it can be deduced that the standard deviation, and hence the predictability, of perturbation trajectory ensembles is dependent on the parameters c and d. For negative c there is the possibility that the standard deviation will converge with time, and hence a

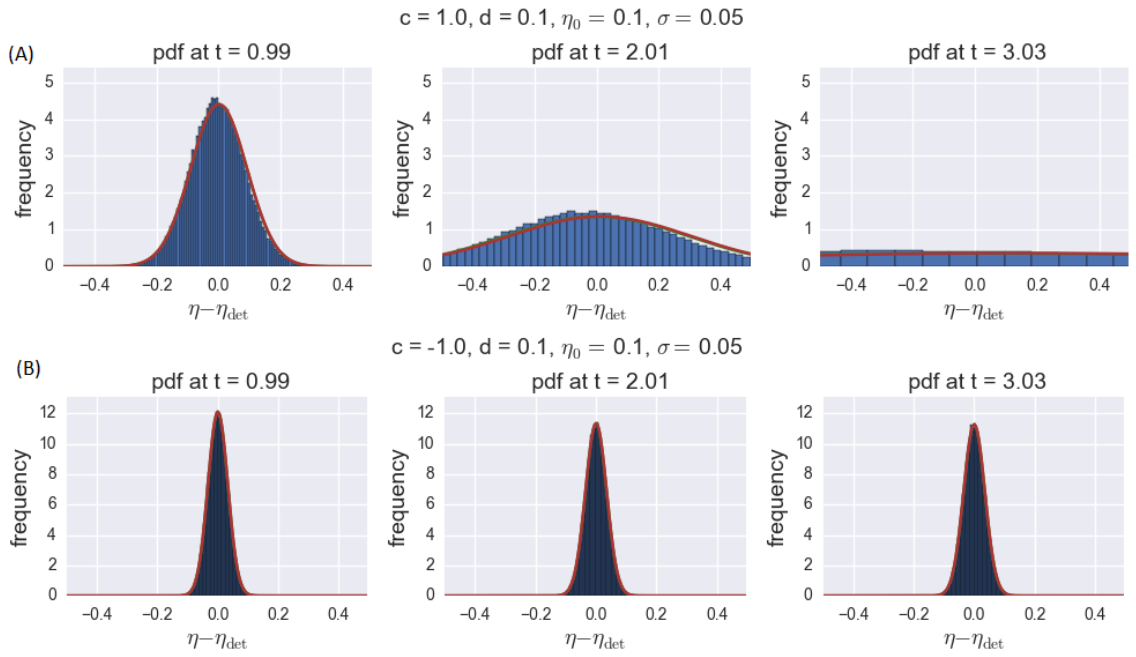

Figure 5: Left to right: Time evolution of the pdf of a perturbation η once its deterministic $O(1)$ part has been subtracted for (A) $c > 0$ and (B) $c < 0$ (c, d constant). Numerical integrations are shown in the blue histograms, and the $O(\sigma)$ analytical solution is overlaid in red.

degree of predictability will be maintained, while for positive c it can be seen that σ_T will diverge, and information will be lost. This agrees with the interpretation of c as defining the stability of the system at first order, as noise is magnified when the system is in an unstable state, and damped when the system is stabilizing.

This analytical solution is compared to the pdfs computed numerically from ensemble integrations of the SDE in Figure 5. In Figure 5A, for case $c > 0$ when the deterministic solution is close to an unstable steady state, the analytical solution agrees with the numerically observed rapid divergence of trajectories. In the stable case $c < 0$ shown in Figure 5B, both the analytical and numerical solution display the maintenance of a tight Guassian centred about the deterministic solution. There is some divergence in the tails between the analytical and numerical solutions for the case $c > 0$, but overall the agreement is fair.

The role of parameter d is explored in Figure 6, in which the time evolution of standard deviation σ_T^2 is plotted for various sign choices of c and d. There is either exponential growth or convergence of the standard deviation, depending on the sign of c, as seen in Figure 5. The parameter d has a lower order effect on the rate of this growth or convergence.

3.3.2 Approximations for small perturbations at small time

To understand the influence of c and d more clearly in Equation 22, consider the evolution of a small perturbation over a short time period, so $\epsilon \equiv d\eta_0(0) \ll 1$ and $ct \ll 1$. Such solutions

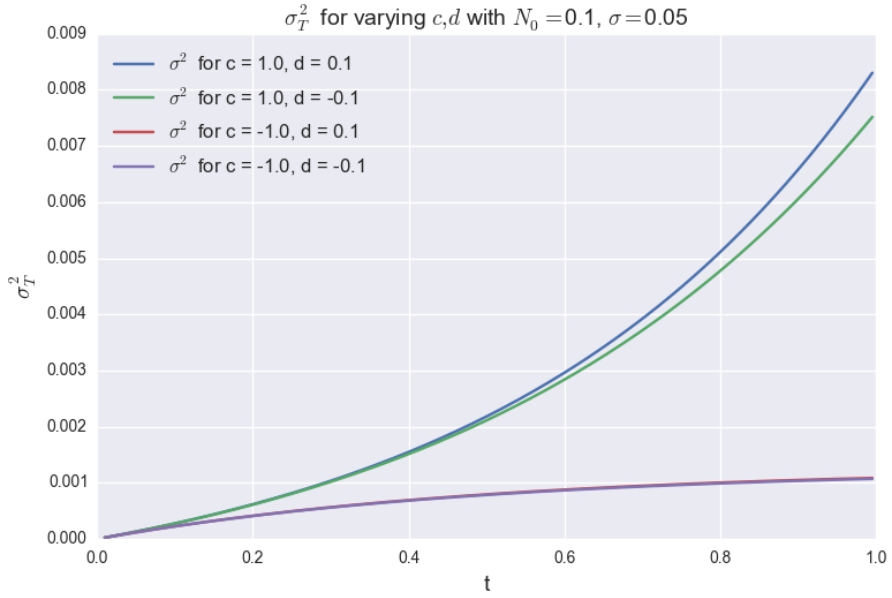

Figure 6: Time evolution of the $O(\sigma)$ standard deviation for various constant values of c and d.

are relevant as perturbations to the Arctic sea ice are generally small in comparison to the climatological value, and the initial evolution of the stochastic solution is of most interest in determining the roles of c and d.

On introducing small parameter ϵ, we have that by definition,

$$\tilde{c} = c + \frac{2c\epsilon e^{ct}}{c + \epsilon(1 - e^{ct})}. \tag{24}$$

So substituting into the expression for $I(t)$,

$$\begin{aligned} \mathrm{I}(t) &= ct + 2\int_0^t \frac{c\epsilon e^{cs}}{c + \epsilon(1 - e^{cs})} ds \\ &= ct + 2\log\left(\frac{c}{c + \epsilon(1 - e^{ct})}\right), \end{aligned} \tag{25}$$

assuming constant c and d. It follows that

$$\begin{aligned} \sigma_T^2 &= e^{2I(t)} \int_0^t e^{-2I(s)} ds \\ &= e^{2ct}\left(\frac{c}{c + \epsilon(1 - e^{ct})}\right)^4 \int_0^t e^{-2cs}\left(1 + \frac{\epsilon}{c}(1 - e^{cs})\right)^4 ds. \end{aligned} \tag{26}$$

Thus far, no assumptions have been made other than that c and d are constants. If it is now supposed that ϵ is small, the binomial parts of the above expression for the standard

deviation may be expanded to give

$$\begin{aligned}\sigma_T^2 &\simeq e^{2ct}\left(1-\frac{4\epsilon}{c}(1-e^{ct})\right)\int_0^t e^{-2cs}\left(1+\frac{4\epsilon}{c}(1-e^{cs})\right)ds \\ &= e^{2ct}\left(1-\frac{4\epsilon}{c}(1-e^{ct})\right)\left(\frac{1}{2c}(1-e^{-2ct})-\frac{2\epsilon}{c^2}(1-e^{-ct})^2\right) \\ &= \left(1+\frac{4\epsilon}{c}(e^{ct}-1)\right)\left(\frac{1}{2c}(e^{2ct}-1)-\frac{2\epsilon}{c^2}(e^{ct}-1)^2\right). \end{aligned} \tag{27}$$

Then to $O(\epsilon)$, the standard deviation becomes

$$\sigma_T^2 \simeq \frac{1}{2c}(e^{2ct}-1)+\epsilon\frac{2}{c^2}e^{ct}(e^{ct}-1)^2. \tag{28}$$

From this approximate expression, the role of c and d deduced from numerically integrating the full expression for the standard deviation can be seen directly. If c is negative, then the limit as $t\to\infty$ can be found as $-1/2c$, whereas if c is positive, the variance grows without bound with time. The parameter d only enters in combination with the initial perturbation $\eta_0(0)$ through $\epsilon \equiv d\eta_0(0)$. A positive value of ϵ increases the growth in the variance at second order, so positive perturbations to the sea ice extent lose predictability more rapidly when $d>0$ than when $d<0$, whereas negative perturbations lose predictability faster for $d<0$.

Applying the small time limit $ct \ll 1$, the exponentials may be expanded to give

$$\sigma_T^2 \simeq t(1+(2\epsilon+c)t), \tag{29}$$

demonstrating the role of both c and d (through the parameter ϵ) in the divergence of trajectories at small time. As discussed previously, both positive c and positive ϵ contribute to the loss of predictability due to the increase of σ_T^2 with time.

3.3.3 Order σ^2

At $O(\sigma^2)$, the stochasticity is no longer explicit in the η_2 equation, but enters through the appearance of the stochastic variable η_1. To obtain a Fokker-Planck equation at second order then, we return to equation 10, and write down Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial\eta}(\tilde{c}\eta+d\eta^2)\rho+\frac{\sigma^2}{2}\frac{\partial^2\rho}{\partial\eta^2}. \tag{30}$$

Now rescaling by defining y such that $\eta=\sigma y$,

$$\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial y}(\tilde{c}y+d\sigma y^2)\rho+\frac{1}{2}\frac{\partial^2\rho}{\partial\eta^2}, \tag{31}$$

the Fokker-Planck equation may be expanded in powers of σ as

$$\begin{aligned} O(1): &\quad \tfrac{\partial\rho_0}{\partial t} = -\tfrac{\partial}{\partial y}(\tilde{c}y\rho_0)+\tfrac{1}{2}\tfrac{\partial^2\rho_0}{\partial y^2} \\ O(\sigma): &\quad \tfrac{\partial\rho_1}{\partial t} = -\tfrac{\partial}{\partial y}(\tilde{c}y\rho_1)+\tfrac{1}{2}\tfrac{\partial^2\rho_1}{\partial y^2}-\tfrac{\partial}{\partial y}(dy^2\rho_0) \\ &\quad\vdots \end{aligned} \tag{32}$$

Due to the rescaling employed, the $O(1)$ equation is identical to that seen previously in the $O(\sigma)$ problem, and the $O(\sigma^2)$ solution comes from the $O(\sigma)$ Fokker-Planck equation above. Note that the form of the higher order equation is the same as that of the lower order equation, but with an additional forcing term determined by the lower order solution.

To solve, a Fourier transform is taken in y to give

$$\frac{\partial \hat{\rho_1}}{\partial t} = \tilde{c}k\frac{\partial \hat{\rho_1}}{\partial k} - \frac{k^2}{2}\hat{\rho_1} + idk\frac{\partial^2 \hat{\rho_0}}{\partial k^2}. \tag{33}$$

Substituting in the known Guassian form of $\hat{\rho_0}$, the method of characteristics may be applied similarly to before to solve for $\hat{\rho_1}$. It is found that

$$\hat{\rho_1} = id\left[S(t)k^3 - M(t)k\right]\exp\left(-\frac{\sigma_T^2 k^2}{2}\right), \tag{34}$$

$$\text{for:} \quad S(t) = e^{3I(t)}\int_0^t \sigma_T^4 e^{-3I(s)}\ \mathrm{d}s, \tag{35}$$

$$M(t) = e^{I(t)}\int_0^t \sigma_T^2 e^{-I(s)}\ \mathrm{d}s. \tag{36}$$

Inverting for ρ_1,

$$\rho_1 = dM(t)\frac{\partial \rho_0}{\partial y} + dS(t)\frac{\partial \rho_0}{\partial y^3}. \tag{37}$$

So recalling that $y = \eta/\sigma$,

$$\rho \approx \rho_0 + \sigma\rho_1 \tag{38}$$

$$= \rho_0 + \sigma^2 dM(t)\frac{\partial \rho_0}{\partial y} + \sigma^4 dS(t)\frac{\partial^3 \rho_0}{\partial y^3} \tag{39}$$

$$= \Big[1 - \frac{dM}{\sigma_T^2}\eta + \frac{dS}{\sigma_T^4}(3\eta - \eta^3)\Big]\rho_0. \tag{40}$$

3.3.4 Evolution of moments

The analytical form of the probability distribution calculated to second order in σ has the following moments

$$< \eta\rho > = -\sigma^2 dM \tag{41}$$

$$< \eta^2\rho > = (\sigma\sigma_T)^2 \tag{42}$$

$$< \eta^3\rho > = \sigma^4(3dM\sigma_T^2 - 6dS). \tag{43}$$

The distribution skewness, defined as

$$Sk \equiv \frac{< \eta^3\rho > -3 < \eta\rho >< \eta^2\rho > - < \eta\rho >^3}{< \eta^2\rho >^{3/2}} \tag{44}$$

can be calculated from these moments as

$$Sk = \frac{\sigma}{\sigma_T^3}\Big[- 6dS + 6d\sigma_T^2 M + \sigma^2 d^3 M^3\Big]. \tag{45}$$

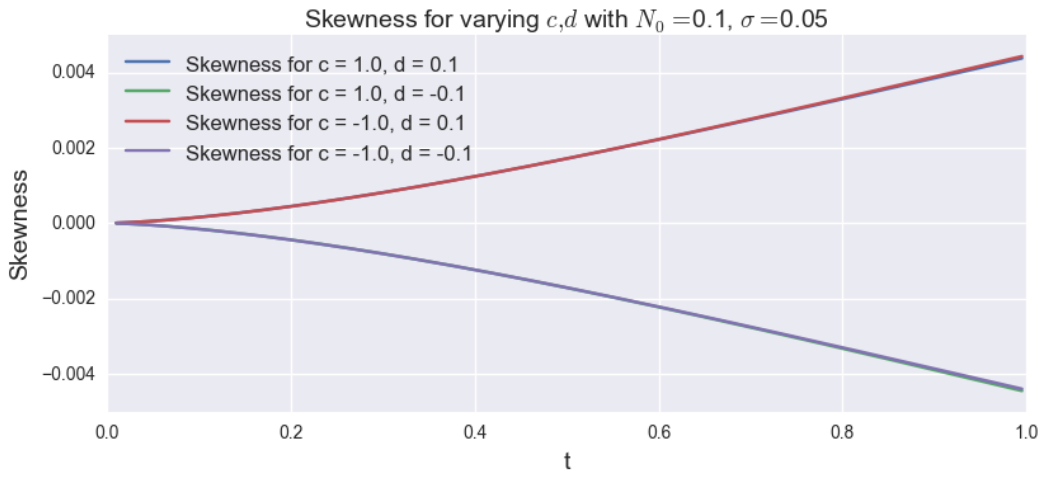

Figure 7: Evolution of the perturbation pdf skewness with time, for varyious constant values of parameters c and d.

This is plotted in Figure 7. Inclusion of the second order correction yields little improvement of the analytic solution, but it allows the interpretation of the parameter d as the asymmetry present in the system, when coupled with initial condition $\eta(0)$. For $\eta(0)d$ positive, the skewness evolves to positive values, while for $\eta(0)d$ negative, the skewness becomes more and more negative with time. This may be intuitively understood by referring to the potentials in Figure 4, in which it can be seen that if $\eta(0)d$ is positive, the initial condition is on the same side as the origin as the unstable steady state, while if $\eta(0)d$ is negative, it will be on the same side of the origin as the stable steady state. Depending on the stability of this nearby secondary fixed point, the tails of the distribution evolve differently, resulting in a positive or negative skewness.

3.4 Physical interpretation

From the simple model of Arctic sea ice, it has been seen that increases in the standard deviation of trajectories are caused by c positive, while convergence of trajectories is achieved for c negative. Returning to the definition of c as $\partial_E f|_{E_S}$, c may be interpreted as representing feedbacks in the sea ice system. For c positive, the forcing f increases with the ice thickness variable E implying that positive feedbacks dominate, while for c negative, an increase in E causes a reduction in the ODE forcing so that the system is controlled by negative feedbacks. This interpretation allows a physical explanation for the seasonal variation of the sea ice extent standard deviation in terms of the annual variation of feedback strengths in the Arctic.

The two main relevant feedbacks in the Arctic are the ice-albedo feedback and the longwave stabilization feedback. The ice-albedo feedback refers to the effect of changes in surface albedo with ice thickness and extent. Any initial ice melt reduces the surface albedo, causing more solar radiation to be absorbed, resulting in further melting. Conversely, a positive perturbation to the ice thickness or extent increases the surface albedo so that

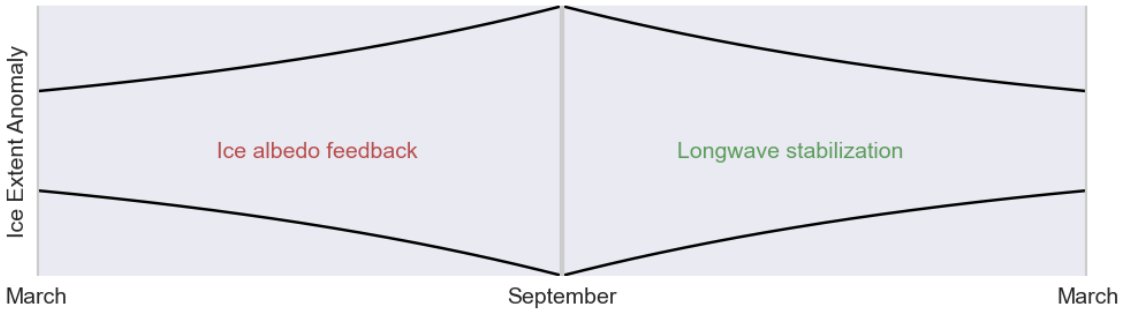

Figure 8: A schematic of the seasonal variation in trajectory standard deviation, and the physical mechanisms that control this. The spreading of trajectories in spring and summer is associated with the dominance of the positive ice-albedo feedback during this time of year, while the dominance of the negative longwave stabilization feedback in fall causes the re-focussing of trajectories going into the winter months.

more radiation is reflected, the system cools, and more ice is able to form. The ice-albedo feedback is therefore a positive feedback. The longwave stabilization feedback, on the other hand, describes the faster growth of thin ice than thick ice due to the nonlinearity of the system, and is a negative feedback. If there is a negative perturbation to the sea ice thickness, the speed at which is grows increases, pushing the system back to equilibrium. Similarly, the growth of positive perturbations is arrested, again stabilizing the system.

On a seasonal basis, the strength of the ice-albedo feedback varies in step with the solar radiation. In the winter when there is no sunlight to produce a positive ice-albedo feedback, the negative longwave stabilization feedback dominates, and we would anticipate a negative value for c. As the insolation increases through spring, the ice-albedo feedback increases in strength, raising the value of c to its summer maximum. As the value of c controls the growth of the standard deviation of trajectories, this explains the observed seasonal variation in sea ice extent standard deviation, which is low during winter and increases to a September maximum in step with the Arctic insolation levels. This situation is illustrated in the schematic in Figure 8.

The second parameter of relevance that came out of the preceding analysis was ϵ; the combination of $d \equiv \partial_E^2 f|_{E_S}/2$ and $\eta_0(0)$. It was seen that predictability was lost more quickly in the case of positive ϵ. The simple model shows the seasonal cycle of d as following a similar pattern to that of c; positive in the spring and summer, and negative over winter. As ϵ comprises both d and the sign of the initial perturbation, this suggests faster loss of predictability following anomalously high ice extents ($\eta_0(0) > 0$) than for anomalously low spring ice extents ($\eta_0(0) > 0$) in spring when $d > 0$, but the converse in fall when $d > 0$.

3.4.1 Changing predictability in a warming climate

With the aid of the physical framework outlined above, it is possible to attempt to answer the question of how the predictability of the Arctic sea ice extent might change in the context of a warming climate. Running the simple model of Eisenmann & Wettlaufer with

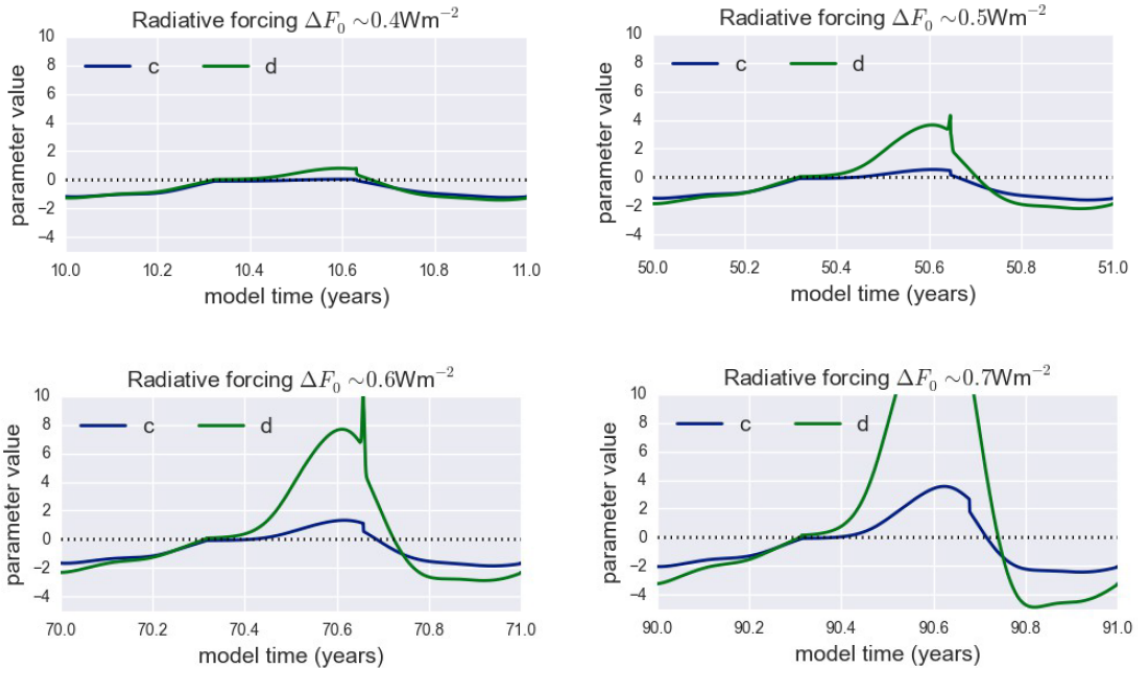

Figure 9: Steady state seasonal cycle of parameters c and d under increasing radiatve forcing F_0 from the E&W09 model. The seasonal cycles of both c and d are amplified as the forcing is increased from 0.4Wm^{-2} to 0.7Wm^{-2}.

increased values of radiative forcing F_0, the annual cycle of c is seen to amplify under global warming as shown in Figure 9. As c becomes more positive in summer months in a warmer climate, it might be anticipated that the predictability of the September minimum of Arctic sea ice will decline under global warming. Conversely, as c is seen to become more negative during winter as F_0 increases, the predictability of the winter maximum might increase.

Although the influence of $d\epsilon$ on predictability is secondary to that of c, the amplification of the seasonal cycle of d could likewise contribute to a worsening predictability of summer sea ice minima following positive perturbations to the spring ice extent, but might act to improve predictability following anomalously low spring extents. Again, the converse would be true for the prediction of the winter sea ice maximum, when the sign of d is reversed.

3.5 The backward equation

From analysis of the Fokker-Planck equation, it has been seen that predictability in the sense of trajectory standard deviation is limited by the dominance of positive feedbacks in the sea-ice system in spring. However, given an observed sea ice extent in September, what does this feature of the system mean for our ability to pinpoint the origin of such an anomaly? This problem calls for application of the Kolmogorov Backward Equation (KBE), which gives the pdf at some initial time $s < t_{\text{end}} \equiv 0$ of conditions that could have led to solution η_{end} at $t = t_{\text{end}}$ (where stochastic perturbation is applied as before so that η is a small anomaly superposed on the seasonal sea ice solution).

From the $O(\sigma)$ expansion fo Equation 10, the KBE may be written down as

$$-\frac{\partial \rho}{\partial s} = \big[c(s) + 2d(s)\eta_0(s)\big]\eta_1 \frac{\partial \rho}{\partial \eta_1} + \frac{1}{2}\frac{\partial^2 \rho}{\partial \eta_1^2}. \tag{46}$$

Fourier transforming and applying the method of characteristics as for the forward equation, the standard deviation of the leading order Guassian solution is found to be

$$\sigma_{TB}^2 = e^{-2I_B(s)} \int_s^0 e^{2I_B(r)} \, \mathrm{d}r \tag{47}$$

$$I_B(s) = \int_s^0 \big[c(r) + 2d(r)\eta_0(r)\big] \, \mathrm{d}r \ . \tag{48}$$

This is plotted backwards in time in Figure 10.

Figure 10 demonstrates that the evolution of the standard deviation of trajectories going back in time is controlled by the stability parameter c just as for forward trajectories. However, unlike for the forward case, a positive value of c leads to trajectories converging backwards in time, while a negative value of c causes them to diverge. Returning to the interpretation of c as the stability of the system, this can be understood in terms of the possible initial conditions that could have produced a given anomaly. If the system is in a stable regime ($c < 0$), trajectories may have originated from a large range of ICs within the potential well, so looking back in time the pdf of ICs that could have resulted in a given anomaly is broad. On the other hand, if the system is unstable ($c > 0$), there is a very low probability of the IC being far from the origin, as this is a potential hill when the system is unstable.

If the KBE standard deviation is plotted further back in time for the stable case, then the direction of the perturbation (controlled by asymmetry parameter $\eta_0(0)d$) is found to matter, with some divergence in trajectories occurring if this places the perturbation on the side of the origin with a potential well, as opposed to further convergence if the perturbation was on the opposite side of the origin (see Figure 4). While this agrees with the interpretation above, its physical relevance may be limited by the fact that stochastic perturbation theory is only valid for small perturbations, as so may fail far back in time when trajectories are far from the origin.

As for the forward equations, approximations for small perturbations at small time can be made for the backward equation to see directly the role of c, d, and $\eta_0(0)$ in the equation. For small parameter ϵ defined as before, for constant c and d the expression for $I_B(s)$ is

$$\begin{aligned} I_B(t) &= ct + 2\int_0^t \frac{c\epsilon e^{-cs}}{c + \epsilon(1 - e^{-cs})} ds \\ &= ct - 2\log\left(\frac{c}{c + \epsilon(1 - e^{-ct})}\right), \end{aligned} \tag{49}$$

where the change of variables $t \equiv -s$ has been made in the definition of I_B for ease of comparison with the forward equation. Then

$$\begin{aligned} \sigma_{TB}^2 &= e^{-2I_B(t)} \int_0^t e^{2I_B(s)} ds \\ &= e^{-2ct}\left(\frac{c}{c + \epsilon(1 - e^{-ct})}\right)^4 \int_0^t e^{2cs}\left(1 + \frac{\epsilon}{c}(1 - e^{-cs})\right)^4 ds. \end{aligned} \tag{50}$$

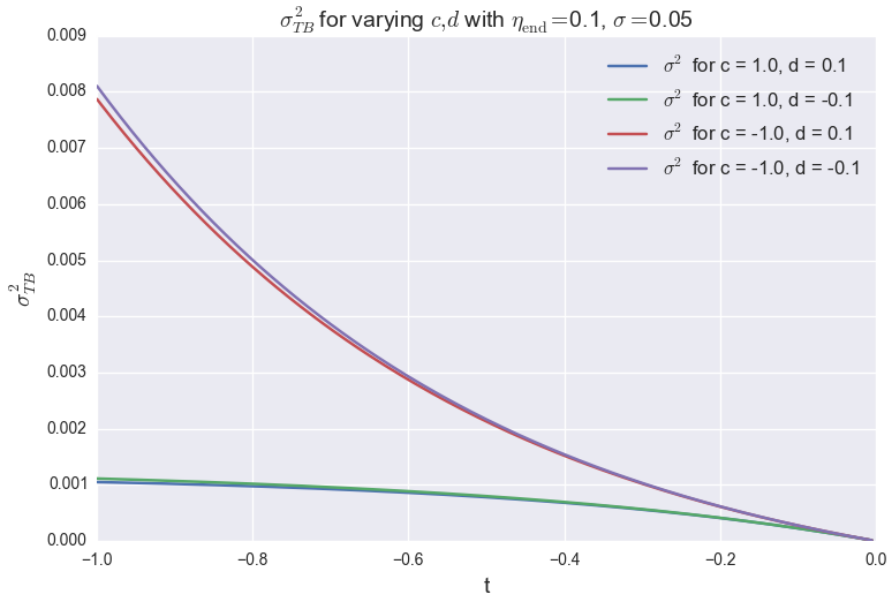

Figure 10: Trajectory standard deviation evolution backward in time, as calculated from applying the Kolmogorov Backward Equation to the $O(\sigma)$ problem for various constant values of parameters c and d.

Now assuming that ϵ is small, the binomials may be expanded as before, allowing integration for the approximate expression

$$\sigma_B^2 \simeq \left(1 - \frac{4\epsilon}{c}(1 - e^{-ct})\right)\left(\frac{1}{2c}(1 - e^{-2ct}) + \frac{2\epsilon}{c^2}(1 - e^{-ct})^2\right). \tag{51}$$

Then to $O(\epsilon)$,

$$\sigma_{TB}^2 \simeq \frac{1}{2c}(1 - e^{-2ct}) - \frac{2\epsilon}{c^2}e^{-ct}(1 - e^{-ct})^2. \tag{52}$$

It can immediately be seen from this expression that the influence of c on the standard deviation has changed signs; now a positive value of c allows the variance to converge, while negative c leads to exponential growth of the variance with time. Similarly, the sign of ϵ is flipped, with positive $d\eta_0(0)$ now reducing the divergence of trajectories.

These findings bode well for the attribution of a given year's September sea ice minimum to events earlier in the season, even if the forward equation suggests that the prediction of this minimum is inherently limited. There have been attempts in the literature to trace anomalous sea ice extents in specific years to the spring melt pond fraction [6], or the occurrence of given winter storms [7]. In particular, our understanding of the role of ϵ suggests attribution of extreme summer minima (i.e. negative $\eta_0(0)$) should be easier than the attribution of maxima (positive $\eta_0(0)$) when d has a positive value as in summer.

3.6 Understanding skill

Returning to the concept of skill defined in the Section 1.3, we may now relate this to the system's stability in a similar manner to the above treatment of predictability. Recall that

$$\text{skill} \equiv 1 - \frac{\sigma^2_{\text{sim}}}{\sigma^2_{\text{ref}}} \ , \tag{53}$$

for reference standard deviation σ^2_{ref}, and simulation standard deviation σ^2_{sim}. In the context of the simple model considered previously, σ^2_{sim} is the quantity σ^2_T defined in Equation **??**. To calculate σ^2_{ref}, the long time limit of the $O(\sigma)$ SDE is taken following [5], which gives

$$\sigma^2_{\text{ref}} = \exp\left(2\int_0^{\tilde{t}} c(s)\ \mathrm{d}s\right)\left[I_{\tilde{t}} + \frac{1}{e^{2\gamma}-1} I_T\right], \tag{54}$$

where

$$I_{\tilde{t}} \equiv \int_0^{\tilde{t}} \exp\left(-2\int_0^r c(s)\ \mathrm{d}s\right)\mathrm{d}r, \tag{55}$$

$$I_T \equiv \int_0^{T} \exp\left(-2\int_0^r c(s)\ \mathrm{d}s\right)\mathrm{d}r \tag{56}$$

$$\gamma \equiv -\int_0^T c(s)\ \mathrm{d}s, \tag{57}$$

for $t = nT + \tilde{t}$, where T is the period length of c, and $\tilde{t} = t \mod T$.

To make the skill calculation numerically tractable, consider a simplified two-season model in which $c = c_P > 0$ for six months ('spring'), and $c = -c_N < 0$ for six months ('fall');

$$c(t) = \begin{cases} c_P & 0 \le t < 1/2 \\ -c_N & 1/2 \le t < 1, \end{cases} \tag{58}$$

where c_P and c_N are positive and $c_N > c_P$ for stability. In this system,

$$\begin{aligned} I_T &= \int_0^{1/2} \exp\left(-2\int_0^r c(s)ds\right)dr + \int_{1/2}^1 \exp\left(-2\int_0^r c(s)\right)dr \\ &= \int_0^{1/2} e^{-2c_P r}dr + \int_{1/2}^1 \exp\left(-c_P + 2c_N(r-\frac{1}{2})\right)dr \\ &= \frac{1}{2c_P}(1-e^{-c_P}) + \frac{1}{2c_N}(e^{c_N-c_P} - e^{-c_P}). \end{aligned} \tag{59}$$

Also,

$$\begin{aligned} \gamma &= -\int_0^{1/2} c_P ds + \int_{1/2}^1 c_N ds \\ &= \frac{1}{2}(c_N - c_P). \end{aligned} \tag{60}$$

The value of $I_{\tilde{t}}$ depends on which half of the year the model is in. When $0 \leq \tilde{t} < 1/2$,

$$I_{\tilde{t}} = \int_0^{\tilde{t}} \exp(-2c_P r)dr = \frac{1}{2c_P}(1 - e^{-2c_P\tilde{t}}), \tag{61}$$

while if $1/2 \leq \tilde{t} < 1$,

$$\begin{aligned} I_{\tilde{t}} &= \int_0^{1/2} \exp(-2c_P r)dr + \int_{1/2}^{\tilde{t}} \exp\left(-c_P + 2c_N(r - \frac{1}{2})\right) dr \\ &= \frac{1}{2c_P}(1 - e^{-c_P}) + \frac{1}{2c_N}e^{-(c_P+c_N)}(e^{2c_N\tilde{t}} - e^{c_N}). \end{aligned} \tag{62}$$

The above expressions give reference standard deviation

$$\begin{aligned} \sigma^2_{\text{ref}} = e^{2c_P\tilde{t}} \Bigg[&\frac{1}{2c_P}(1 - e^{-2c_P\tilde{t}}) + \\ &\frac{1}{e^{(c_N-c_P)} - 1}\left(\frac{1}{2c_P}(1 - e^{-c_P}) + \frac{1}{2c_N}(e^{(c_N-c_P)} - e^{-c_P})\right)\Bigg] \end{aligned} \tag{63}$$

in the first part of the year (i.e. $0 \leq \tilde{t} < 1/2$), and

$$\begin{aligned} \sigma^2_{\text{ref}} = e^{c_P+c_N-2c_N\tilde{t}} \Bigg[&\frac{1}{2c_P}(1 - e^{-c_P}) + \frac{1}{2c_N}e^{-(c_P+c_N)}(e^{2c_N\tilde{t}} - e^{c_N}) \\ &+\frac{1}{e^{(c_N-c_P)} - 1}\left(\frac{1}{2c_P}(1 - e^{-c_P}) + \frac{1}{2c_N}(e^{(c_N-c_P)} - e^{-c_P}\right)\Bigg] \end{aligned} \tag{64}$$

in the latter part of the year ($1/2 \leq \tilde{t} < 1$).

Figure 11 shows the time evolution of the skill for such a model, in the case of a spring start of for the case of a fall start. It is apparent that skill is lost more rapidly when the system is in the unstable spring state than when the model is in the stable fall state. This agrees with the role of the stability parameter c in controlling whether nearby trajectories converge or diverge, as explained previously.

A concept often noted in the literature on model skill is that of *reemergence*. In general, model skill is expected to decrease monotonically away from a given initial condition. However, some models exhibit a slight increase in skill following their initial decline; a feature referred to as reemergence. One hypothesis put forward to explain this phenomenon is that of natural variability. Initial work has suggested that the simple model may demonstrate a reemergence of skill at long time if forced periodically, which motivates further consideration of this concept in future work.

4 Application of Results to Global Climate Models

In practice, the seasonal prediction of Arctic sea ice is not done using simple one-dimensional models, but using high resolution Global Climate Models (GCMs). From a consideration of the origins of predictability in a simple model, however, it is evident that GCMs must be capable of capturing the important feedbacks at play in the Arctic (principally, the

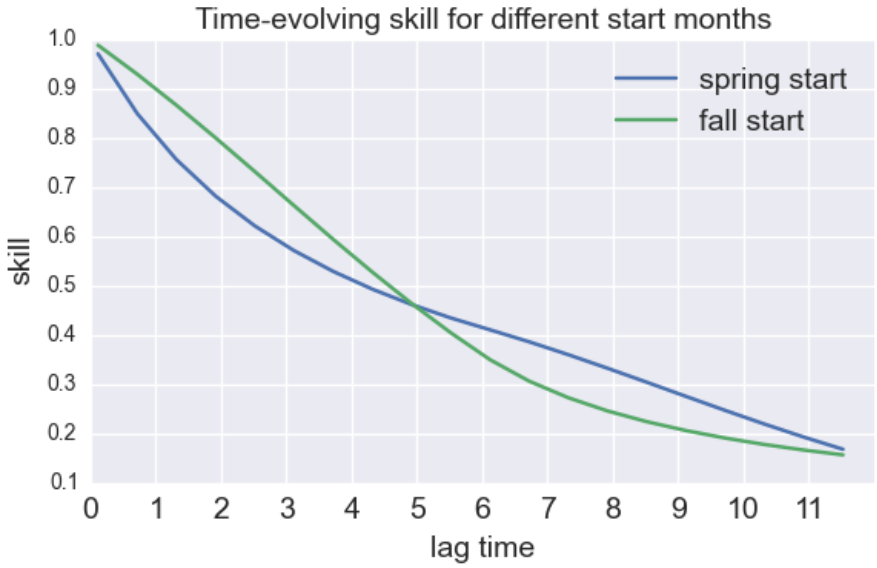

Figure 11: The decrease in skill over time for different start months; one during the unstable season (spring; $c = 0.1$) and one going into the stable season (fall; $c = -2.0$).

positive ice-albedo feedback and the negative longwave stabilization feedback) if they hope to predict Arctic sea ice on a seasonal basis. It might be expected that if such feedbacks are incorporated in the GCMs, they will show the same seasonal variation in standard deviation seen in the satellite data and explained by the simple model considered above.

4.1 CMIP5 data

To assess the success of GCMs at reproducing the seasonal cycle in the standard deviation sea ice anomaly trajectories that forms the basis of our understanding of predictability, data from the CMIP5 model ensemble is analysed. CMIP5 is the fifth Climate Model Inter-comparison Project, and gathers state of the art GCMs for a series of standardised runs with common forcings to enable comparison between the models. Available statistics include the ice thickness, the sea ice extent, and the ice volume. From the ice extent, a secondary metric known as the Equivalent Ice Extent (EIE) may be calculated, which is the area within the sea ice margin. This may be used to avoid biases resulting from the location of the ice in relation to land.

To analyse the GCM sea ice predictability, timeseries of the statistics listed above were taken from the CMIP5 historical runs. These runs start between 50 and 150 years before present, depending on the model, and are run through to the modern day with conditions mimicking those of the last century. For each model, the data was detrended, and its seasonal cycle was removed. The remaining timeseries anomalies were then split into years, with each year considered an independent anomaly trajectory. This enabled calculation of the annual cycle of the trajectory standard deviation, both for individual models and for the CMIP5 data as a whole, as shown in Figure 12 for the EIE.

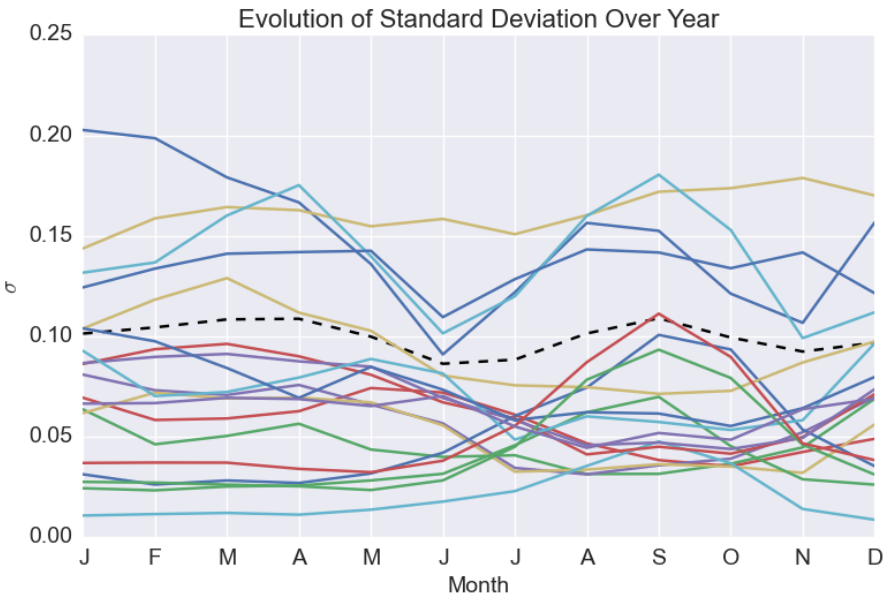

Figure 12: Annual evolution of the standard deviation of the EIE anomaly for the different models included in CMIP5, with the multimodel standard deviation shown as the black dashed line.

It is evident from Figure 12 that the models, in general, do not capture the seasonal cycle in the sea ice anomaly standard deviation seen in the satellite data. This is problematic given the relation of the seasonal variation in standard deviation to the annual cycle in feedback strengths, which is fundamental to the problem of seasonal prediction. Some insight can be gained into why the models might be failing by plotting the EIE before the seasonal cycle is removed (Figure 13). In Figure 13 it is seen that the models may be split into two rough categories; those that capture the naturally observed sinusoidal seasonal cycle in the Arctic sea ice, and those that have a somewhat unnatural cycloidal-shaped seasonal cycle. If the ice extent is considered instead of the EIE, however, this cycloidal behaviour is eliminated and all the models display a sinusoidal seasonal cycle. This suggests that the models may have been tuned to fit the ice extent rather than the EIE, resulting in the strange seasonal cycles seen in Figure 13, and a loss of the statistical behaviour pertinent to seasonal prediction.

It is hoped that the models might succeed better at reproducing the statistics of the ice thickness and ice volume, as these metrics are dominated by the less noisy multi-year ice, and consequently not as prone to the large interannual fluctuations experience by the sea ice extent.

4.2 Ensemble size & statistic convergence

When using GCM runs to make statements about the Arctic sea ice, a further issue is that of ensemble size. Under the assumption of independence, the CMIP5 data might be thought of as providing a collection of stochastic trajectories of Arctic sea ice extent. How many

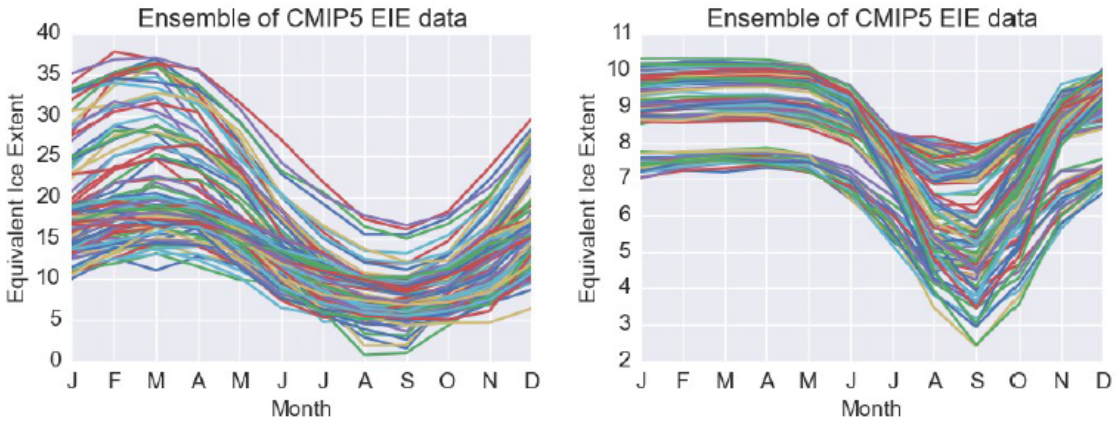

Figure 13: The seasonal cycle of Equivalent Ice Extent (EIE) for a selection trajectories from each of the CMIP5 models. Left: those models that capture a sinusoidal seasonal cycle, right: those models with a cycloidal seasonal cycle.

models then are needed to make robust statistical statements about the behaviour of ice, say under a forcing like that associated global warming?

In Figure 14, the simple model SDE is integrated for increasing ensemble sizes ranging from 10 to 1000, and the time evolution of the resultant standard deviation is compared to that computed from the analytic solution. Convergence occurs relatively quickly, with ensemble sizes of around 50 producing standard deviations that are in reasonable agreement with the analytic result. Even such a modest ensemble size is somewhat larger than the number of models included in CMIP5, though. This should be taken into consideration when drawing conclusions from GCM data derived from a limited number of runs.

5 Conclusions & Future Work

The application of stochastic perturbation theory to a simple model has allowed the annual variations in the predictability of Arctic sea ice to be traced back to the stability of the system, which arises from the presence of feedbacks. The dominance of the positive ice-albedo feedback in spring makes the system unstable, resulting in low predictability of the sea ice minimum going into summer. The negative longwave stabilization feedback in fall, however, focusses sea ice anomaly trajectories going into winter, increasing predictability during this period. It was seen that the problem of attribution works in reverse, with an unstable system making anomalies easy to attribute to events earlier in the season, and a stable system complicating attribution.

It was further shown that the system stability may be directly related to the concept on model skill. In the unstable regime (i.e. spring), model skill drops off faster than in the stable regime (i.e fall). This is likewise a consequence of trajectories diverging away from an initial condition when positive feedbacks dominate, and converging when negative feedbacks dominate. Change in the strength of the ice-albedo and longwave stabilization feedbacks under global warming would alter the stability properties of the system, affecting

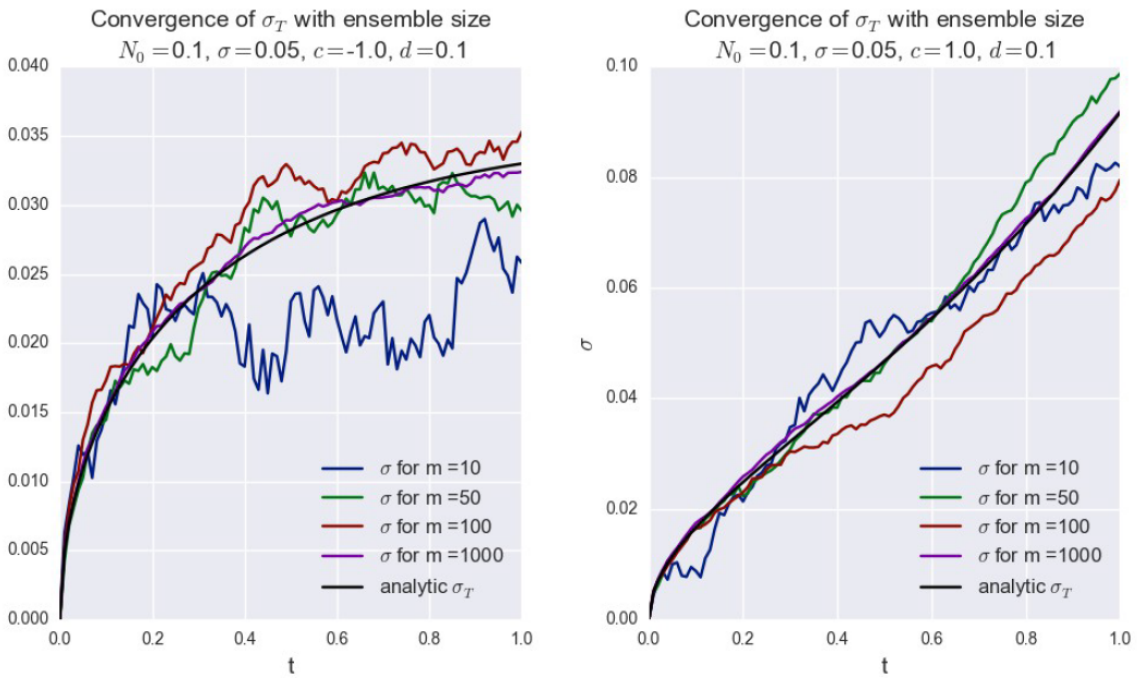

Figure 14: Convergence of the standard deviation from the numerical integration of ensembles to the analytically computed standard deviation when the system is in a stable regime (left) and an unstable regime (right).

our ability to predict the summer and winter sea ice extents.

A brief analysis of GCM data suggests that current generation models may experience difficulty in the seasonal prediction of Arctic sea ice as their failure to capture the annual cycle of sea ice variance suggests difficulty in reproducing principal physical feedbacks that underlie the evolution of sea ice anomalies.

It should be noted that the methodology presented in this project is by no means constrained to the Arctic sea ice system, but may in fact be applied to any problem for which there is relatively small magnitude noise superposed on a strong periodic cycle. As such, future work might focus on using this framework to consider other climate signals for which seasonal or interannual prediction is of interest.

References

[1] I. Eisenman and J. S. Wettlaufer, *Nonlinear threshold behavior during the loss of Arctic sea ice*, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009), pp. 28–32.

[2] P. K. Kloeden and E. Platen, *Numerical Solutions of Stochastic Differential Equations*, Springer-Verlag, Berlin, 2008.

[3] G. A. Maykut and N. Untersteiner, *Some results from a time-dependent thermodynamic model of sea ice*, Journal of Geophysical Research, 76 (1971), pp. 1550–1575.

[4] W. Moon and J. S. Wettlaufer, *On the existence of stable seasonally varying Arctic sea ice in simple models*, Journal of Geophysical Research, 117 (2012), p. C07007.

[5] ——, *A stochastic perturbation theory for non-autonomous systems*, Journal of Mathematical Physics, 54 (2013), p. 123303.

[6] D. Schröder, D. L. Feltham, D. Flocco, and M. Tsamados, *September Arctic sea-ice minimum predicted by spring melt-pond fraction*, Nature Climate Change, 4 (2014), pp. 353–357.

[7] J. Zhang, R. Lindsay, M. Steele, and A. Schweiger, *What drove the dramatic retreat of arctic sea ice during summer 2007 ?*, 35 (2008), pp. 1–5.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-2016-05	2.	3. Recipient's Accession No.
4. Title and Subtitle 2015 Program of Study: Stochastic Processes in Atmospheric & Oceanic Dynamics			5. Report Date November 2016 6.
7. Author(s) John Wettlaufer and Oliver Bühler			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C)OCE-1332750 (G)
12. Sponsoring Organization Name and Address National Science Foundation			13. Type of Report & Period Covered Technical Report 14.

15. Supplementary Notes
This report should be cited as: Woods Hole Oceanographic Institution Technical Report, WHOI-2016-05.

16. Abstract (Limit: 200 words)
Stochastic Processes in Atmospheric and Oceanic Dynamics was the theme at the 2015 GFD Program. John Wettlaufer and Oliver Bühler were the stochastic co-directors. Professors Charlie Doering (University of Michigan) and Henk Dijkstra (University of Utrecht) were the principal lecturers. Their lectures were collectively two pronged. The first prong was launched by Charlie Doering, who laid down the mathematical foundations of random variables, stochastic processes and the nature and analysis of stochastic differential equations. In the second, Henk Dijkstra took us through the many places in the Atmosphere, Ocean and Climate system where the infrastructure from the first prong plays out.

17. Document Analysis a. Descriptors

Geophysical Fluid Dynamics
Stochastic Processes
Atmospheric and Oceanic Dynamics

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Approved for public release; distribution unlimited.	UNCLASSIFIED	402
	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18) *See Instructions on Reverse*

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce