Signals and Hidden Information

by

Kevin Royce Vixie

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in

Systems Science

Portland State University
2001

To my parents –
who loved me
and inspired me
… and left before their time

To My Family, Beata and Levi –
for love and support

To My Brother Curtis –
for understanding
for insight
for encouragement

To my mentors Justin MacCarthy, Thomas M. Thompson,
Kenneth Wiggins, Chris Bretherton, John Erdman,
Serge Preston, David Sigeti,
and Andrew M. Fraser –
for their patience and inspiration.

To my friends –
Don, Larry, Jim, Jeff,
Erik, Ben, Tom, Jonathan, John,
Melanie, John, Richard, Andreas, Tad, and Steve,
David, Marta, Aunt Ida Mae, Uncle John and Steve Butterfield

Acknowledgments

In acknowledging those influential in aiding me in my progress towards this dissertation I find myself realizing that many people, from the earliest point in my life until the present time, have had a significant impact on this progress – impact that can be seen in who I am and what I have done. *If you do not like stories, don't read the rest of these acknowledgments ... it is a very personal story for which I give no apology.*

The road to this dissertation has been longer than the usual such road, with many more twists, delays, and detours. It began of course with my parents whose own brilliance, tenacity, and love had enormous impact on the establishment of the drive to create, explore and connect with others. My first mentor in mathematics, Justin MacCarthy, was living in Deming, NM when we first met – I, a naive eager learner, he an erudite linguist-mathematician. He inspired me by showing me all sorts of wonderful things in mathematics about which I had no idea. After finishing my "high school" years of self study in the high desert of Deming, I moved on to Walla Walla College where the mentoring (and patience!) of Thomas M. Thompson and Kenneth Wiggins had a deep impact – not only on my mathematical development, but also on my physical survival! Next came some years at the University of Washington whose high point was the time I spent as a student of Chris Bretherton's. By all rational paths, I should have finished a dissertation with him many years ago. At this point, the largest detour began – oceanography research, construction work, engineering consulting, work on a seed farm, a Masters Degree, elementary school teaching, setting up and running a physiology research Lab, being unemployed – and then a miracle occurred – I met my wife Beata. Shortly after marrying, I was visiting my cousin Jim who suggested I take up my previous habit of "Walkabout"'s in the woods and forests. This was the beginning of the end of the long detour. It was on those walks that I began to gain a new confidence, quietness, clarity, and insight that I had known before but not with such depth. I began to think about mathematics again. This progressed until I was thinking about mathematics most of the time (when I woke up in the morning, during my walks in the woods, while I was setting up experiments in the Lab, in the evening), and I decided to go back to graduate school. I thought of Andrew M. Fraser, who had previously impressed me with his abilities and his patience. I began to work with him – first part time, and then – three months after Levi was born (!) – full time. I found Andy's viewpoints and courses to be very inspiring. Going back to school after a long break for the purpose of working with Andy had the added, surprise benefit of the semi-simultaneous arrival in the systems science and mathematics programs of several other highly talented students, also on delayed tracks such as mine. This peer group was of great importance to a deeper development of my intuitions and technical abilities. Another unplanned benefit was the handful of very gifted mathematicians from whom I learned analysis and geometry. John Erdman inspired me in a way that is hard to overestimate. His modified Moore-method classes had a profound influence on me. Serge Preston reinforced my realization that mathematics had a flow – a life, that could be tapped by the intuition and made to work for you in the creation of new mathematics. I still remember his joke about the soviet policemen and the book. Through Andy, I came to Los Alamos where I benefitted greatly from the generous encouragement of David Sigeti and others.

The work contained in this dissertation has benefitted in a direct manner from conversations and interactions with (ordering not significant) David Sigeti, Murray Wolinsky, James Howse, Tim Sauer, Lai-Sang Young, Andrew M. Fraser, Pieter Swart, David Caraballo, Gary Sandine, John Pearson, Kevin L. Buescher, Andreas Rechtsteiner, Clint Scovel and Brendt Wohlberg.

Of these I would like to single out Andy Fraser and David Sigeti for their generous contribution of time for discussions.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

This dissertation is the sum of 4 papers representing work carried out at Los Alamos National Laboratory over the last 3 years. In this introduction, I will give a very brief and high level view of the results in those papers as well as the publication status of the same papers.

Science typically concerns itself with drawing conclusions based on measurements. These measurements very rarely capture the state being measured. This leads us to various *inverse problems*. Some problems, such as x-ray tomography or seismic exploration, are the typical examples, but in actuality, many other scientific questions find answers through the solution of some inverse problem. Let us formalize the situation which covers the type of problems studied in this dissertation. One begins with a space of states X and an *operator* or *mapping*, $F : X \to X$, by which we get an evolution of states $x \in X$ in time. For simplicity and without much loss of generality, we discretize time and let our time coordinate τ take on integral values: $\tau \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. We will denote the state at time $\tau = i$ as x_i. Our time evolution is then given by $x_{i+1} = F(x_i)$. To be more general one would permit the operator F to depend on time. For the purposes of this introduction, we assume that F does not depend on time. As the states evolve in time, we perform measurements via another operator $P : X \to Y$, where Y is our measurement space. The sequence of *measurements*, *observations* or *data* $\{..., y_{-2}, y_{-1}, y_0, y_1, y_2, ...\}$ is obtained by the operation of P on the sequence of states $\{..., x_{-2}, x_{-1}, x_0, x_1, x_2, ...\}$.

In addition to loss of information in the measurement process due to the fact that the level sets of P are entire hyper-surfaces in X, we typically have "noise" to deal with which means that for a point x_i in the state-space X, the corresponding measurement $P(x_i)$ will only be known to within some error. One might have a model for the "noise" or error: $\tilde{p}(y_i|x_i) = \tilde{p}(y_i - P(x_i))$ where $\tilde{p}()$ might, for example, be a multivariate Gaussian probability density.

Example 1.1.1. *If X is the instantaneous state of a fluid in some experimental setup (in which case X is infinite dimensional – from the point of view of the Navier-Stokes equations), then the measurement might be the velocities at n points in the fluid. In this case Y is $\mathbb{R}^n$ and P maps an infinite dimensional space into $\mathbb{R}^n$.*

The broad question we now have is, "How can we get back to the sequence of states and/or the operator F given some or all of the data $\{..., y_{-2}, y_{-1}, y_0, y_1, y_2, ...\}$?" Getting "back to the sequence of states and/or the operator F ", obviously includes obtaining information *clearly contained* in state and operator knowledge. But I would also include recovery of information which is in principle obtainable, but in fact difficult to obtain, from state and operator knowledge.

The four papers in this dissertation look at a few aspects of this question. Papers 1-4 are contained in chapters 2-5 respectively. In chapter 2, I settled a controversy concerning the test proposed by Hinich and Wolinsky which – they claimed – detected aliasing in some sampled stationary processes. They were right and I explained why. In chapter 3, I showed that one could drop the assumption of stationarity, look instead at samples drawn from a single waveform and still get detection of aliasing. The key concept, which was completely new, was that of *sampling stationarity.* In chapter 4, I examined the question of tomographic reconstructions of dynamically evolving 3-dimensional objects from a series of 2-dimensional projections or radiographs, *all from one viewing angle.* While the work in chapter 4 was original for me, some of the work turned out to have similar or identical precedents. What was new was the concept of reconstructing a single object from a minimal number of radiographs. All of the previous work except Aeyels paper [1] kept the dynamics governing time evolution fixed while varying the measurement function, while I looked at the case of fixed measurements and variable dynamics. The key proofs were original (since the theorems were new to me) and I believe that the approach to reconstruction through the **Extended Linear Transverse Intersection Theorem** (Theorem 4.6.2) is also new. In chapter

5 (paper 4) I have carefully dissected a concept defined by Fraser for the purpose of measuring model fidelity. In this case our task is not so much the solution of an inverse problem, but rather the evaluation of a proposed solution. What is new here is the careful explanations of the different types of Sinai-Ruelle-Bowen (SRB) measures and their relationships as well as the role finite precision plays in the computation of Lyapunov exponents. Theorem 5.5.1 and Conjecture 5.5.1 are new. Theorem 5.5.5 is also new.

The publication status of the papers are as follows. Chapter 2 has been published in the proceeding of ISSPA '99 [102]. Chapters 2,3, and 4 are on the LANL e-print archive [102, 101, 100]. I have been invited to submit chapter 4 for the proceedings of a special session of the 2001 New Orleans AMS meeting in which I gave an invited talk.

1.2 Tutorial on Aliasing

The sampling of continuous time signals or the subsampling of discrete signals almost always involves a loss of information. This is certainly not surprising. What is often surprising to the uninitiated is that there are important instances in which *nothing* is lost in the process of sampling! In this section we explain what *aliasing* is and why it can be avoided by *proper sampling.* In the following section we will denote the convolution of two functions $f_1(u)$ and $f_2(u)$ by $f_1(u) * f_2(u)$. We will denote the pointwise product of the same two functions by $f_1(u) \cdot f_2(u)$. We shall also be careless about the fact that while f might be a function, $f(u)$ is a particular value of that function, by using $f(u)$ to denote both the function and the values the function might attain.

Question 1. *Suppose that a continuous time signal $f(t)$ is sampled or measured at the times $i\Delta\tau$ where $i = ..., -2, -1, 0, 1, 2, ...$ and $\Delta\tau$ is some sampling interval. If one throws away the rest of the signal and keeps only the samples, how much has been lost?*[1]

Answer: *Almost everything – (see figure 1.1)*

But as mentioned above, there are important cases in which *nothing* is lost!

[1] To answer this in detail requires the consideration of generalized Fourier transforms and tempered distributions. We will present only an intuitive explanation, but include references at the end of the subsection for the details.

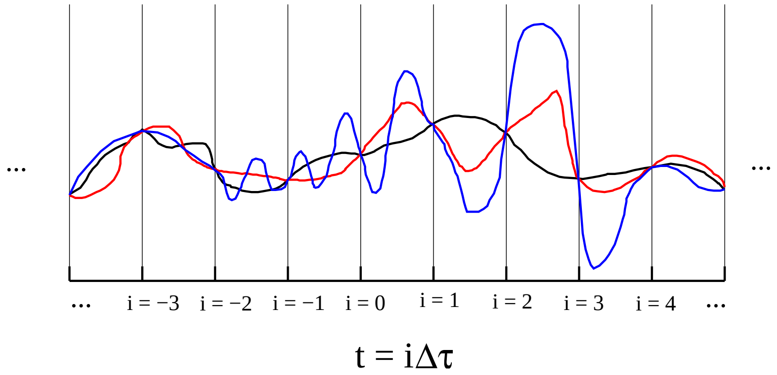


Figure 1.1: Each signal has the same set of samples at the sampling times and therefore cannot be distinguished on the basis of the samples

We introduce the notation and terminology now. Consider a signal $f(t)$ and let $F(\omega)$ be it's Fourier transform (where ω is in cycles per time-unit, not radians per time-unit). Suppose that the support of $F(\omega)$ is contained in the interval $[-\omega_h, \omega_h]$ (i.e. $F(\omega) = 0$ for all $\omega \notin [-\omega_h, \omega_h]$). Suppose also that we have sampled the signal $f(t)$ every $\Delta\tau$ time units. This will give us a sampling rate of $R_s = 1/\Delta\tau$.

Definition 1.2.1 (Proper Sampling). *If $R_s > 2\omega_h$ then the signal $f(t)$ is said to have been* **properly sampled**.

We now present a result which many find amazing at first glance.

Theorem 1.2.1 (Shannon Sampling Theorem). *If a signal is properly sampled, the samples alone are enough to reconstruct the original signal! (see figure 1.2)*

Explanation: We begin with some **Suggestive reasoning:** If two signals f_1 and f_2 have exactly the same samples every $\Delta\tau$ then, it must be the case that $g \equiv f_1 - f_2$ is zero at each of the sampling points. Now the only sinusoids that are zero at these points are sinusoids with frequencies $1/2\Delta t$, $2/2\Delta t$, $3/2\Delta t$, etc. This *suggests* that the Fourier expansion of the difference is made up of components with frequencies that are greater or equal to $1/2\Delta t$. **Graphical "proof":** reconstructibility of a properly sampled waveform can be seen to be possible by simply looking at the reconstruction convolution in both the time and frequency domains. We do this in figure 1.2. In line (A) we see both the signal $s(t)$ and it's Fourier transform $S(\omega)$. Line (B) shows the sampling function (also called the **Sampling Comb**) and it's transform. In the next line we see result of multiplying the signal and the sampling function (also called the *sampling comb*). Line (D) shows the appropriate sinc function and it's transform which permit the convolution shown in Line (E) to perfectly reconstruct the signal. By the same argument, we can see that if the support of the Fourier transform extends outside of the interval $(-1/2\Delta\tau < \omega < 1/2\Delta\tau)$ then the reconstruction $h(t) * (s(t) \cdot d_{\Delta t}(t))$ will not in general be equal to the the original signal $s(t)$ (see figure 1.3). The error or difference within $(-1/2\Delta\tau < \omega < 1/2\Delta\tau)$ between the transform of the reconstructed signal and the original signal's transform is termed *aliasing.* (see figure 1.3) **Actual Details:** See remark 1.2.1. **End of Explanation.**

Remark 1.2.1. *When does the Shannon sampling theorem hold? That is, in our intuitive explanation above, we said very little about assumptions of the theorem. In fact, we are interested in the case of persistent signals which have only distributional Fourier transforms which are in fact* tempered distributions. *One can show that in fact the sampling theorem holds for any finite sum of sinusoids and in fact does NOT hold for arbitrary functions which have generalized transforms (tempered distributions), but we believe that assuming the Shannon reconstruction works probably doesn't limit the type of persistent,* bounded *signal one is working with. (See lesson 38 of [34]as well as [60] and the references found there for more discussion of the reconstruction of signals from their samples in the case the signal DOES NOT lie in L^2)*

Finally, we define a few related terms which some readers may be unfamiliar with.

Replication Phenomenon For this we refer the reader to line C of figure 1.2 in which the transform of the sampling function or comb is convolved with the transform of the signal to get another transform which is the sum of a bunch of shifted replications of the original transform. This is the replication phenomena to which we will later refer. Notice that aliasing is nothing more than the overlap caused by having the "tines" of the comb transform too close together which in turn is caused by having the "tines" of the sampling comb itself to far apart (undersampling!).

Nyquist Frequency twice the highest frequency in the Fourier transform of the signal. In other words the Nyquist frequency is defined for any signal to be twice the highest frequency in the Fourier spectrum of that signal. If a signal is not bandlimited, then the Nyquist frequency is ∞.

Nyquist $\Delta\tau$ The time increment equivalent to $1/f_N$, where f_N is the Nyquist frequency in cycles per time unit.

Nyquist Limited Data Samples taken from a signal at a rate high enough to ensure proper sampling.

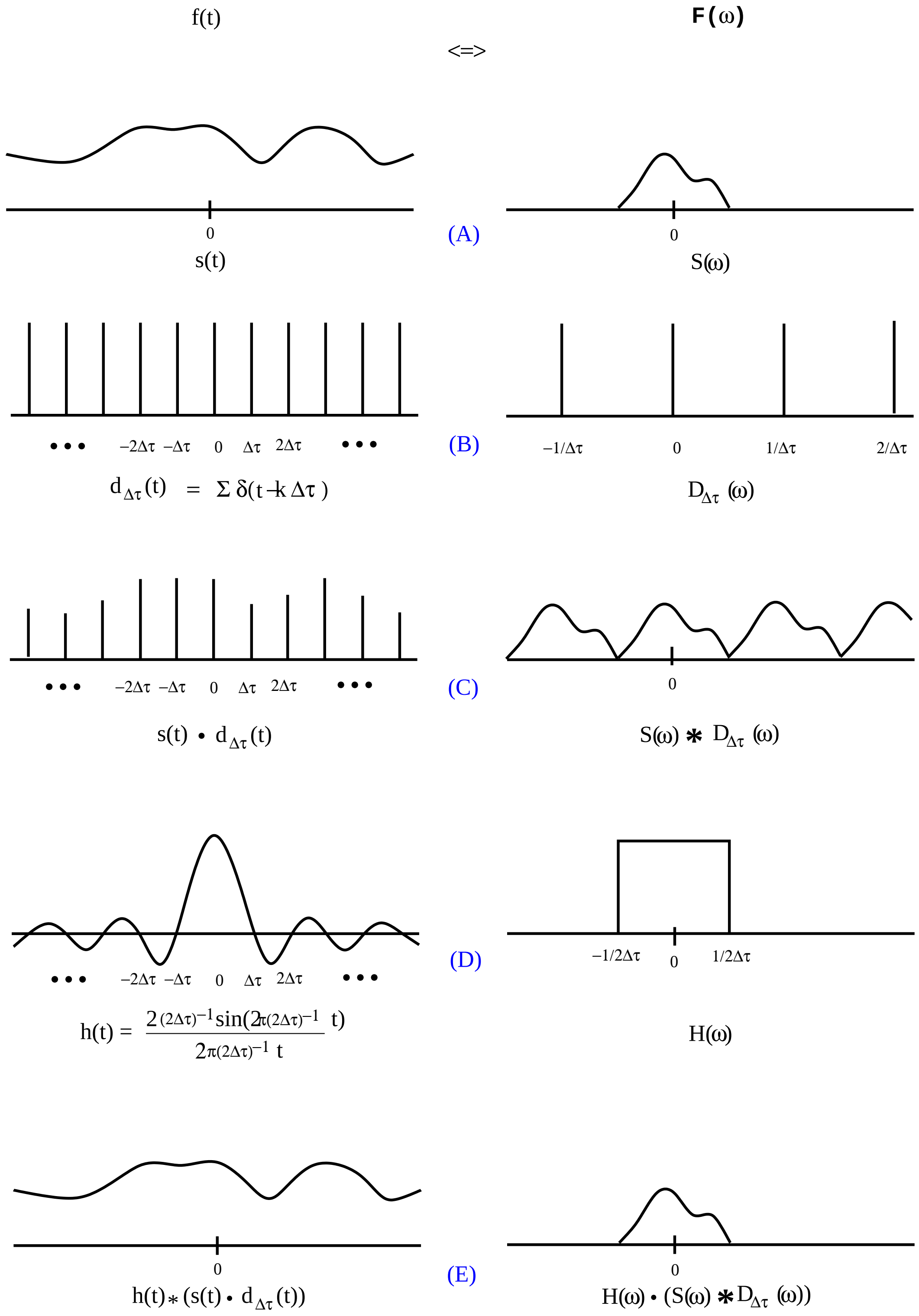


Figure 1.2: The reconstruction of a signal from it's samples via a convolution.

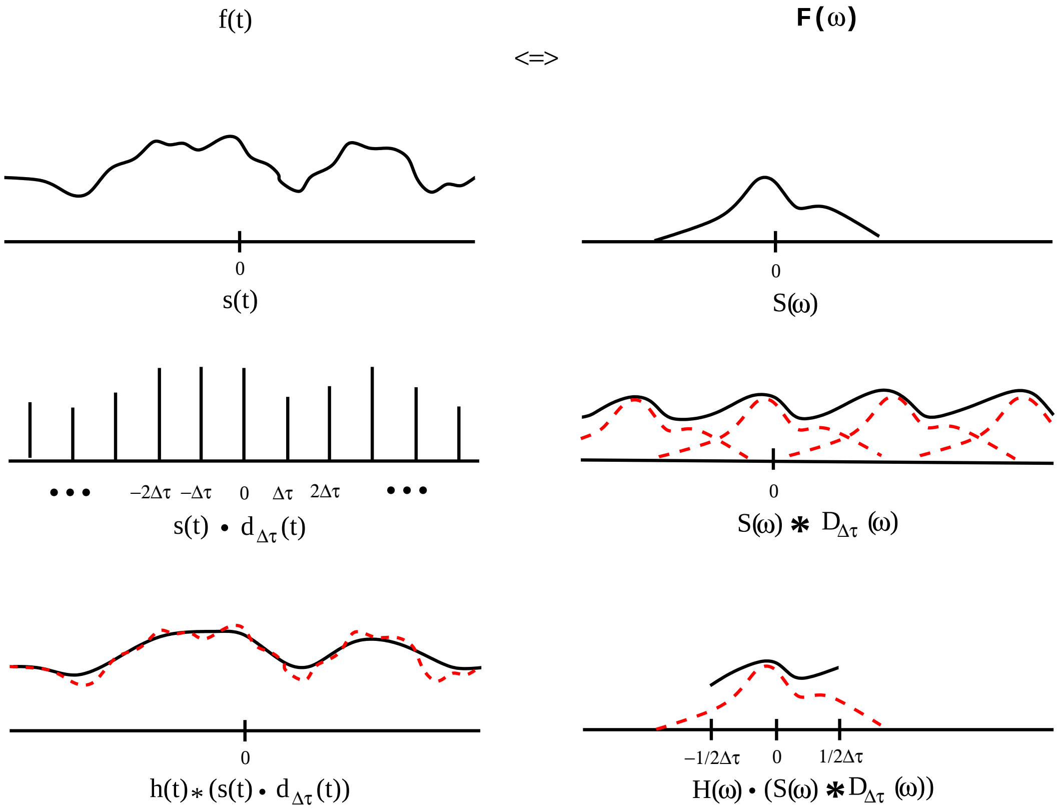


Figure 1.3: The reconstruction of an *undersampled* signal from it's samples via a convolution. **Second line, right:** The dashed lines are the summands which sum to the transform, $S(\omega) * D_{\Delta\tau}(\omega)$, shown by the solid line. **Bottom Line, left:** The solid line is the result of the convolution $h(t) * (s(t) \cdot d_{\Delta\tau}(t))$ – i.e. the Shannon reconstruction from the samples. The dashed line shows the original signal for comparison. **Bottom Line, right:** The solid line shows the transform of the reconstructed signal $(H(\omega) \cdot (S(\omega) * D_{\Delta\tau}(\omega)))$. The dashed line shows the transform of the original signal for comparison. The difference between the dashed and solid lines within the interval $(-1/(2\Delta\tau), 1/(2\Delta\tau))$ indicates the degree to which the high frequency information is "polluting" the low frequency reconstruction, i.e. the degree of *aliasing* present.

Chapter 2

The Bispectral Aliasing Test: A Clarification and Key Examples

Kevin R. Vixie [1]

2.1 Preamble

Controversy regarding the correctness of the bispectral aliasing test proposed by Hinich and Wolinsky [39] has been surprisingly long-lived. Two factors have prolonged this controversy. One factor is the presence of deep-seated intuitions that such a test is fundamentally impossible. Perhaps the most compelling objection is that, given a set of discrete-time samples, one can construct an unaliased continuous-time series which exactly fits those samples. Therefore, the samples alone can not show that the original time series was aliased. The second factor prolonging the debate has been an inability of its proponents to unseat those objections. In fact, as is shown here, all objections can be met and the test as stated is correct. In particular, the role of stationarity as prior knowledge in addition to knowledge of the sample values turns out to be crucial. Under certain conditions, including those addressed by the bispectral aliasing test, the continuous-time signals reconstructed from aliased samples are non-stationary. Therefore detecting aliasing in (at least some) stationary continuous-time processes both makes sense and can be done. The merits of the bispectral test for practical use are briefly

[1] The research for this chapter was done in collaboration with Murray Wolinsky and David E. Sigeti. The chapter has been published in the ISSPA '99 proceedings [102]

addressed, but our primary concern here is its theoretical soundness.

2.2 The Bispectral Aliasing Test

The domain of the discrete-time bispectrum is the two dimensional bifrequency $\{\omega_1, \omega_2\}$ plane. Assuming a real-valued discrete time series (sampled at integral time values), the usual replication phenomenon dictates that all non-redundant information is confined to the square $0 \leq \omega_1, \omega_2 \leq \pi$. When one fully accounts for symmetries, the non-redundant information in the bispectrum is confined to a particular triangle inside this square [18, 107].

This triangle naturally divides into two pieces. One piece is an isosceles triangle and is unproblematic. The other piece, somewhat unusual in shape, is the source of the controversy under discussion. Naive consideration of this triangle shows that it involves frequencies higher than the Nyquist frequency and therefore must have something to do with aliasing. Hinich and Wolinsky considered this more carefully and showed that the naive intuition is correct: if the discrete time series arises from sampling a stationary, band-limited, continuous-time process, and if the sampling rate is sufficiently rapid to avoid aliasing, then the discrete bispectrum is non-zero only in the isosceles triangular subset of the fundamental domain. Conversely, if the bispectrum of a sampled stationary continuous-time process is non-zero in the outer triangle, then the sampling rate was too slow to avoid aliasing.

It should be clearly understood that there is no assertion that aliasing in general can be detected. The statement is not "if a signal is aliased, then the outer triangle will have a non-zero bispectrum." Rather, the assertion is the converse, "if the outer-triangle shows a non-zero bispectrum, the (underlying) continuous-time signal must have been aliased."

At one level, this result is obvious and, in fact, the result was initially so-regarded [83]. However, doubt soon arose. Perhaps the most important source for suspicion is the argument based on reconstruction alluded to above.

In light of this objection, one is led to reconsider the association of the outer triangle with aliasing. One can take the position that there is no relation, as in [32]. One can decide that something is aliased, but that it is the bispectral estimator rather than the signal. (That is, the method of estimation of the bispectrum is introducing spurious data in the outer triangle). There is some plausibility to this claim, for the frequencies that are involved in the outer triangle are ω_1, ω_2, and $\omega_1 + \omega_2 - 2\pi$. This seems to be the position of Pflug et al. [53].

Or, one can try to delineate the conditions, if any, under which the test makes sense. This was done by Hinich and Messer in 1995[38]. They confirmed the validity of the original argument and stated its conclusions more carefully. In particular they conclude that a non-zero bispectrum in the outer triangle indicates a non-random signal or one of the following:

- a random, but non-stationary signal ;
- a random, stationary, but aliased signal, or;
- a random, stationary, properly-sampled signal which violates the mixing condition.

We believe that the analysis of Hinich and Messer, while entirely correct, did little to persuade the detractors of the test. In particular their analysis did not address the reconstruction objection and may have left the impression that the circumstances for which the test applies are unlikely to be met in practice.

In this paper, we show that the reconstruction objection is far from fatal. We further establish that stationarity is the only property which is crucial to the test. Since this property is required in order to define the bispectrum, one can legitimately apply the aliasing test whenever one is entitled to compute a bispectrum. Therefore the bispectral aliasing test is as theoretically sound as the bispectrum itself.

2.3 The Selection Rule and Brillinger's Formula

The bispectrum, defined to be the triple Fourier transform of the third-order autocorrelation, reduces to a function of two frequencies since stationarity confines the spectrum to the plane through the origin of the frequency domain perpendicular to the vector (1,1,1). The defining equation[2] is given by

[2] In this equation and the next we introduce notation that may be unfamiliar to some. $\mathfrak{F}_{123}$ is the 3-dimensional or triple Fourier Transform, $c_3(t_1, t_2, t_3) = \langle x_1(t_1) x_2(t_2) x_3(t_3) \rangle$ is the third order auto correlation of the stochastic signal $x(t)$, $b(\omega_1, \omega_2)$ is the bispectrum

$$\mathfrak{F}_{123}(c_3(t_1, t_2, t_3)) = b(\omega_1, \omega_2)\delta(\omega_1 + \omega_2 + \omega_3). \tag{2.1}$$

Another way of computing the bispectrum is to switch the order in which one does the Fourier transforming and the ensemble averaging. This leads to the following result.

$$b(\omega_1, \omega_2) = \langle X(\omega_1)X(\omega_2)X(\omega_3 = -\omega_1 - \omega_2)\rangle \tag{2.2}$$

If the process is band-limited and $X(\omega) = 0$ for $|\omega| > \pi$, then the bispectrum is confined to the intersection of the $(1, 1, 1)$ plane and the π-cube (i.e. $(\omega_1, \omega_2, \omega_3) \in [-\pi, \pi) \otimes [-\pi, \pi) \otimes [-\pi, \pi)$). The plane and its projection onto the (ω_1, ω_2) plane is shown in Figure 1. Upon sampling with unit time step, one obtains the usual replication in three dimensions. (Doing everything in 3-dimensions and projecting at the end keeps things simpler and makes it easier to avoid errors.) In particular, one gets that if the process is sampled at a frequency greater than twice the highest frequency component, then the bispectrum is confined to the replications of the tilted hexagon shown.

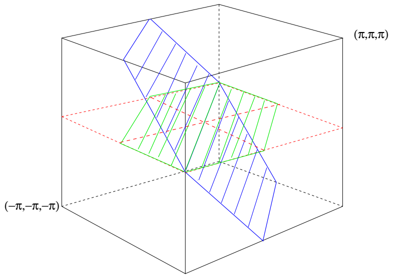


Figure 2.1: The origin of the bispectral fundamental domain

The replication gives the discrete-time bispectrum b_d:

$$b_d(\lambda_1, \lambda_2, \lambda_3) = \sum_{\omega_1+\omega_2+\omega_3=0} b(\omega_1, \omega_2, \omega_3). \tag{2.3}$$

(which this equation is defining) of the stochastic signal $x(t)$ and $\delta(\omega_1+\omega_2+\omega_3)$ is simply the Dirac delta of $\omega_1 + \omega_2 + \omega_3$. Note that the frequency that corresponds to t_i is ω_i for $i = 1, 2,$ and 3. In the next equation we use $X(\omega)$ to denote the Fourier transform of $x(t)$.

where $\omega_i = \lambda_i + 2\pi k$ for integer k.

Since the replication does not cause any overlaps, the outer triangle remains empty. This is the Hinich and Wolinsky aliasing theorem. (Note that the outer triangle is equivalent to the bigger triangle with vertices $(0,\pi,0)$, $(0,0,\pi)$ and $(0,\pi,\pi)$ by symmetries. See [18, 107] for details.)

2.4 The Reconstruction Objection

Suppose we have a stationary process $x(t)$ and we under-sample it by sampling at $t \in Z$. Then by convolving $x(t)$ with the appropriate *sinc* function we get a reconstructed process $x_r(t)$. This new process will have exactly the same samples as the original process and therefore exactly the same sampled bispectrum: yet it is not aliased. Therefore for any process that is under-sampled, we have another process producing an identical sampled process which is not under-sampled, showing that that one could not possibly detect aliasing via the bispectrum computed from samples!

The rub here is the fact that the reconstructed signal will not necessarily be stationary. Processes reconstructed from aliased samples of continuous-time signals are generally cyclostationary but not stationary. Some aliased processes do, in fact, reconstruct into stationary processes. But in the class of stationary signals for which the bispectral aliasing test gives positive results, reconstruction from aliased samples produces non-stationary processes.

To carefully illustrate this we will consider several stationary processes generated by taking a periodic signal with period T and giving it a random shift $\theta \in [0,T)$. *In this section, we will use units of cycles per second (cps) for frequencies, often giving the corresponding radians per second (rad/s) figure.* First consider a simple cosine process,

$$x(t) = \cos(2\pi\alpha t + 2\pi\alpha\theta), \tag{2.4}$$

where $\alpha = .75$, and θ is randomly chosen from [0,4/3). Upon sampling and reconstruction we get the cosine process given by

$$x_r(t) = \cos(2\pi\hat{\alpha} t + 2\pi\alpha\theta), \tag{2.5}$$

where $\hat{\alpha} = -0.25.$[3]

The key idea is that the signal appears at the lower frequency as dictated by its replication into the fundamental region of the frequency space – [-.5,.5) in cycles per second or [-π, π) in radians per second – but its reconstructed phase term is the same as the "source" component phase term.

Now consider

$$x(t) = \cos(2\pi\alpha t + 2\pi\alpha\theta) + \cos(2\pi\beta t + 2\pi\beta\theta), \tag{2.6}$$

where $\alpha = 1.0$, $\beta = 3.0$, and θ is chosen randomly from the interval $[0, 1)$. The reconstructed process one gets is

$$x(t) = \cos(2\pi\hat{\alpha}\pi t + 2\pi\alpha\theta) + \cos(\hat{\beta}\pi t + 2\pi\beta\theta), \tag{2.7}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the aliased frequencies. Although the phase terms $2\pi\alpha\theta$ and $2\pi\beta\theta$ remain unchanged, they now correspond to different time shifts so that we no longer have a single shifted waveform. In the case that we sample every time unit $\hat{\alpha} = \hat{\beta} = 0$ so that, by an unfortunate accident we have chosen to sample commensuratly with the signal period. If we choose another sampling time, and with probability 1 a random choice of sampling time will not lead to this problem, we get a reconstruction with two non-degenerate sinusoids. For example, if we choose Δt (our sampling interval) to be $e = 2.71...$, then we get a non-trivial reconstructed process that is shown in figure 2.2. The key idea is that the stochastic shift θ will now NOT correspond to a time shift since the factors multiplying the shift no longer correspond to the frequencies of the components in which they appear. This might be difficult to see. First one should be convinced of the fact that if the factor multiplying time divided by the factor multiplying θ **is the same for each component**, then the effect of different θ is simply a **rigid translation of the entire waveform**. Next, one should verify that as long as θ is uniformly

[3] If we sample the signal $cos(2\pi\alpha t + 2\pi\alpha\theta)$ to get $cos(2\pi\alpha i\Delta t + 2\pi\alpha\theta)$ we see that there are a countable number of possible sources for the samples since $cos(2\pi\alpha i\Delta t + 2\pi\alpha\theta) = cos(2\pi\alpha i\Delta t + 2\pi ki + 2\pi\alpha\theta)$ $k \in \mathbb{Z}$, and we can manipulate this to get $cos(2\pi(\alpha + k/\Delta t)i\Delta t + 2\pi\alpha\theta)$ so that choosing the single sinusoid with frequency (in cps) in $[-1/2\Delta t, 1/2\Delta t)$ means that the reconstructed waveform will be exactly $cos(2\pi(\alpha + k_a/\Delta t)i\Delta t + 2\pi\alpha\theta)$ where k_f is the unique integer such that $\alpha + k_f/\Delta t \in [-1/2\Delta t, 1/2\Delta t)$. This special frequency – $\alpha + k_f/\Delta t$ – is called the *aliased frequency*.

chosen from $[0, a)$ and changing θ from 0 to a moves the entire waveform exactly one period, then the process is stationary – i.e. statistics computed at any point in time will be the same as those from any other point in time.

Since $2\pi\alpha/2\pi\alpha = 2\pi\beta/2\pi\beta$ our original signal certainly has the rigid translation property. And since changing θ from 0 to 1 indeed translates the waveform by one period, we have that the original process is stationary. On the other hand it will generically be the case that $2\pi\hat{\alpha}/2\pi\alpha \neq 2\pi\hat{\beta}/2\pi\beta$ and the reconstructed waveform **will not** have the rigid translation property. Therefore we expect that the reconstructed signal will have periodic statistics.

Computation of the required expectations requires that one can "average over the ensemble." Since this stationary process is not ergodic, one can not get the result from a single realization of the process. It is at this point that some differences in perspective arise. Strictly speaking, in order to compute a bispectrum one must perform the ensemble average. A single realization does not suffice unless the process is ergodic.

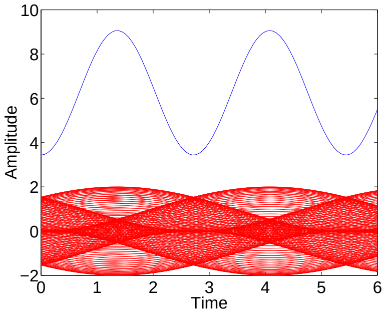


Figure 2.2: Cyclostationarity of a stochastic signal reconstructed from aliased samples. The lower part of the figure is the superposition of a large number of waveforms from the process. The outline of this lower part is the process envelope. (Remember that each point in the underlying probability space corresponds to an entire waveform.) The sampling interval is e. The upper curve is the sixth moment, chosen for ease of display. The original process is given by $x(t) = \sin(2\pi\alpha t + 2\pi\alpha\theta) + \sin(2\pi\beta t + 2\pi\beta\theta)$ with $\alpha = 1.0$, $\beta = 3.0$, $\theta \in [0, 1)$, and $\Delta t = e = 2.71\ldots$.

Finally, consider the process defined by

$$\begin{aligned} x(t) = &\cos(2\pi(5/20)t + 5 \cdot 2\pi\theta) + \\ &\cos(2\pi(6/20)t + 6 \cdot 2\pi\theta) + \\ &\cos(2\pi(-11/20)t - 11 \cdot 2\pi\theta) \end{aligned} \tag{2.8}$$

where θ is chosen randomly from [0,1).

Because, as explained above, changing θ rigidly translates the entire waveform over one period, the process is stationary and the bispectrum may be computed. This process has a spike in the bispectrum at (5/20 cps, 6/20 cps) or in radians/sec, a spike at $\omega_1 = 10/20\pi$, $\omega_2 = 12/20\pi$, which is in the outer triangle (again, see [107] for details concerning the outer triangle). The reconstructed signal is given by

$$\begin{aligned} x(t) = &\cos(2\pi(5/20)t + 5 \cdot 2\pi\theta) + \\ &\cos(2\pi(6/20)t + 6 \cdot 2\pi\theta) + \\ &\cos(2\pi(9/20)t - 11 \cdot 2\pi\theta) \end{aligned} \tag{2.9}$$

which is not stationary. Therefore we have a signal with nonempty outer triangle whose reconstruction is not stationary. This situation is exactly what the bispectral test implies happens whenever the outer triangle is nonempty. The loss of stationarity causes the Fourier transform of the triple autocorrelation to "move off" of the (1,1,1) plane.

Therefore, if one knows (or is willing to assume) that the process which generated the observed samples was stationary, one can rule out the unaliased reconstruction as the source of the samples. In a sense, the continuous time signal reconstructed from aliased samples of an original time series is a "measure zero" object. This result is very surprising to most people's intuitions. In chapter 3 we define a concept of stationarity appropriate for single waveforms allowing us to study the detection of aliasing outside of the framework of stochastic processes.

2.5 The Replication objection

Upon looking at Equation 2.2 one may observe that even if $X(\omega) = 0$ for $|\omega| > \pi$, sampling effectively fills in the spectrum at higher frequencies. This is the basis for the objection appearing in Swami [95]. This concern is addressed as follows. While the spectrum does indeed fill out upon sampling, the undesired expectations remain zero. Consider a (statistically stationary) ensemble constructed by uniformly translating a periodic or finite-duration waveform $x(t)$. Two operations are necessary to produce the discrete-time ensemble; a uniform shift in time over a period T, which introduces linear phase factors, and sampling, which produces spectrum replication. These operations do not commute: i.e., one wants to time-shift the waveform first and then sample, rather than to shift its samples. For the shifted samples $x_s(t+\theta)$ one finds

$$\mathfrak{F}(x_s(t+\theta)) = e^{i\theta\omega}\mathfrak{F}(x_s(t)) \tag{2.10}$$

But for the sampled shifted waveforms $x(t+\theta)|_s$, i.e., the waveforms needed to construct a stationary ensemble, the phase of the original signal is propagated to higher frequencies periodically rather than linearly. This difference leads to the vanishing of unwanted expectations.

For example, consider the process given by the randomly shifted sum of unit amplitude cosine waves with frequencies at $n/20$ (rad/s) where n takes integer values from 1 to 19. The sampled spectrum has components at $\omega_1 = 10\pi/20$, $\omega_2 = 11\pi/20$ and $\omega_3 = -21\pi/20$ but the average

$$\langle X(\omega_1)X(\omega_2)X(\omega_3)\rangle \tag{2.11}$$

reduces to

$$\langle e^{i10\theta\pi}e^{i11\theta\pi}e^{i19\theta\pi}\rangle_\theta \tag{2.12}$$

where θ is chosen with uniform probability from [0,1). This average vanishes. Therefore, the potential contribution in the outer triangle is zero because averaging kills it. This is in contrast to the case where the average is zero because the spectral amplitudes are themselves zero (as in the proof of the aliasing test).

2.6 Empirical counter-examples

Other objections to the test have been made. Frequently these objections involve a (purported) counter-example to the bispectral aliasing test. A particularly clear example is provided by Frazer, Reilly and Boashash [32]. Here the authors do two things. They present an example of an aliased signal which the aliasing test fails to mark as aliased. The example is unproblematic: neither the aliasing test nor any aliasing test we are aware of will detect all aliased signals. It is not, however, a counter-example to the test. Since there is nothing in the outer triangle, the bispectral aliasing test makes no assertion regarding the presence of aliasing.

The other example the authors provide is more interesting. It consists of a signal involving coupled sinusoids at $\omega_1 = 0.3125Hz$, $\omega_2 = 0.25Hz$ and $\omega_3 = .4375Hz$ and the authors show that there is a peak in the outer triangle under conditions which rule out aliasing. As the authors note these frequencies sum to 1 Hz ((the sampling rate). Under these conditions the authors are correct in asserting that the aliasing test gives a positive result, which they believe to be incorrect. However, what the aliasing test actually indicates is that this signal is non-stationary. The particular interaction which the authors have constructed is not one for which the continuous-time selection criteria is met, i.e., the frequencies involved do not sum to zero. Therefore, even though the *samples* of this signal do meet the *discrete-time* stationarity condition, the underlying continuous time signal does not meet the stationarity condition and consequently this signal does not provide us with a counterexample to the aliasing test. Since details of the signal generation are not given (for example, exactly what filter was used to low-pass filter the data?), it is not clear where, in the processing, the signal loses stationarity – if it ever was stationary.

One can look at these results in various ways. Our position is that neither example constitutes a counter-example to the validity of the aliasing test in theory, though they both show that the test is limited in practice. The first example shows that there are aliased signals which the test does not see. This is obvious anyway since there are signals with zero bispectrum whose samples can be aliased. The second example shows that the term “aliasing test” must be restricted to stationary signals. As stated earlier, this restriction is inherent in the definition of the bispectrum.

2.7 Conclusions

So, is this something for nothing? How can one get information about higher frequency amplitudes from what is usually thought of as Nyquist-limited data? The answer is of course that the assumption of stationarity is far from nothing. But, to exploit stationarity one must be able to perform the ensemble averaging indicated in the definition of the bispectrum. This implies that one must either have an ergodic process or have access to sufficiently many sample paths.

It is certainly possible that, in practice, the bispectrum can be usefully applied to signals for which there is no theoretical justification. For such uses the aliasing test is silent. However, it is essential that a clear understanding of the fundamental properties of higher-order spectra be available. And the present authors believe that correct understanding of the outer triangle leads to deeper insight of the meaning of the bispectrum in general.

Chapter 3

Detection of Aliasing in Persistent Signals

Kevin R. Vixie [1]

3.1 Introduction

Detection of aliasing from temporal samples alone, with no restrictions on the original continuous-time source, is impossible because any set of samples may be reconstructed (using convolution with the sinc function) to a properly sampled signal having the same samples. However, quite often additional information about the source is available. It is, of course, obvious that tight constraints on the source would permit perfect reconstructions of vastly under-sampled signals. For example, the constraint that the data comes from a linear function of time makes any two samples sufficient. A less extreme example is a signal with a lower as well as an upper frequency cutoff (a *bandpass signal*). For bandpass signals, it is well-known that one can sample at a rate below twice the highest frequency while still achieving perfect signal recovery (see [48, p.138, theorem 13.3]).

What are the weakest constraints that one can put on the signal and still get something—detection of aliasing, for example? Here, we examine constraints of stationarity. In 1988 Hinich and Wolinsky [39] suggested a bispectral test for

[1] The research for this chapter was done in collaboration with David E. Sigeti and Murray Wolinsky.

detecting aliasing in temporally sampled stationary stochastic processes[2]. The test aroused some controversy [95, 32] which is examined in [38] and chapter 2. In chapter 2 we show, in detail, that the test *does* detect aliasing in some signal processes and that it is the constraint of stationarity that makes the detection of aliasing possible. Briefly, if we under-sample a stationary process and then reconstruct a continuous-time signal from the samples using the Shannon sinc filter, the reconstructed process will not, in general, be stationary. In contrast, a *proper* sampling followed by reconstruction will not destroy stationarity because this procedure just reconstructs the original signal. Detecting non-stationarity in the reconstructed process thus suffices to establish the existence of aliasing in the time series, provided it can be assumed that the original signal was stationary. These results are reviewed in Section 3.3.

Applying these concepts requires either a random sample of the paths of the process or an assumption of ergodicity which makes it possible to extract statistics from a single sample path. In this paper we attempt to generalize the results for stationary processes to the more common case where we have only a single sample path and can make no assumption of ergodicity. In other words, we look for ways to discover under-sampling in a time series drawn from a single waveform, which may or may not be a sample path of some underlying stochastic process. We define *sampling stationarity*, a form of stationarity that makes sense for single waveforms, and show that it can be used to detect aliasing in complex, continuous-spectrum signals. We present reasons to believe that sampling stationarity should be a generic[3] property of signals and that the destruction of sampling stationarity by the process of under-sampling and reconstruction should occur quite generally. Finally, we explain how it might be possible to use the reconstructed sample statistics plots (RSS plots) that we use to detect aliasing to obtain additional information about individual Fourier components beyond the Nyquist frequency.

The remainder of this paper proceeds as follows. After illustrating the key idea of this paper with an example in Section 3.2, we proceed, in Section 3.3, to demonstrate how a constraint of stationarity permits the detection of under-sampling in

[2] In the following, we will use the terms *signal process* or just *process* for stochastic signal processes. Except when we use the terms *sample path* for a realization of a stochastic process or *random sample*, the word "sample" will refer to temporal sampling.

[3] We use the term *generic* in a nontechnical sense. The term usually occurs in a situation where one would like to say "with probability 1" but where no obvious probability measure exists.

some signal processes. Then in Section 3.4 we define sampling stationarity. In Section 3.4.1 we use examples to show that the concept of sampling stationarity does, indeed, enable detection of aliasing for nontrivial signals. In Section 3.4.2 we consider the case of periodic signals. For this class of signals, we provide a complete explanation of how (and when) the method of high-frequency detection works. The possible extension of this explanation to non-periodic signals is then discussed in Section 3.4.3. In Section 3.4.4, we present some reasons to believe that the plots that we have used to detect aliasing may also be used to recover some portion of the original signal's high-frequency content. This is followed by suggestions for further work (Section 3.5) and a conclusion that summarizes the work in this paper (Section 3.6). Two after-notes contain computational and mathematical details.

3.2 Example

The key idea of our approach is captured by a very simple example. Suppose that we sample a square wave that takes the values -1 and 1. There is a unique properly band-limited signal that has this time series as its samples. We can compute this signal by applying the Shannon sinc filter to our time series. We may regard this computation as an attempt to reconstruct the original continuous-time signal. If we can reject this reconstructed signal as the source of our samples, then we must conclude that the time series contains aliased components.

Note that our given time series consists only of -1's and 1's. The reconstructed signal, on the other hand, is necessarily a continuous function of time, taking on all values in the interval $[-1, 1]$ (and, in fact, beyond). The only way that we could have obtained a sequence of -1's and 1's by sampling the reconstructed signal is if we had chosen a particular sampling rate (or one of its sub-harmonics) and a unique shift of the sampling comb. Any other combination of sampling rate and shift would have produced a series that takes a continuum of values. The probability of having chosen the special sample rate and shift that give a sequence of -1's and 1's is clearly zero, provided that our sampling rate was chosen independently of the source. With this proviso, then, we can reject (with confidence level 1) the hypothesis that our time series consisting of -1's and 1's came from sampling the reconstructed signal.

The assumption that the sampling rate was chosen independently of the source is justified in most (but not all) cases of practical importance because we can rule out any interdependence between the source and the sampling rate on physical grounds. For example, if a signal produced by a distant source is sampled at a predetermined rate, such a coupling is clearly out of the question—it would amount to believing that the process that produced the signal "knew" when we were going to sample at a distant location.

We can draw valid conclusions from a sampled signal about Fourier components beyond the Nyquist frequency only if we can put constraints on the original continuous-time source. How can we characterize the constraints that we are imposing in this case? Effectively, we are assuming that the sample times (which are determined by the sample rate and shift) do not play a distinguished role in the source. Showing that the sampling times are distinguished in the reconstructed signal then suffices to reject the reconstructed signal as the original source of the samples.

How, then, can we extend this analysis to more general classes of signals? In the case of a square wave (or any signal that takes on a finite number of values), the appearance of the time series produced by sampling the reconstructed signal at the given sampling times could not be more different from the appearance of a time series produced by sampling at any other shift of the sampling comb. Thus it is clear what we mean when we say that the sample times are distinguished in the reconstructed signal. For more general signals, however, it is not so clear exactly what it means for the sample times to be distinguished.

There is one obvious case in which we can be assured that the sample times are not distinguished in the original signal and in which we can detect the distinguished character of the sample times in the reconstructed signal. If the original signal is, in fact, a stationary signal process, then, by definition, *no* time is distinguished. The appearance of non-stationarity in the reconstructed signal would then indicate the presence of aliasing in the time series. The detection of aliasing in time series from stationary signal processes is the subject of the next section. Following that, we use our example of sampling from a square wave and insights from the case of stationary processes to develop a method for detection of aliasing in single waveforms.

3.3 Detection of Aliasing in Stationary Processes

Consider the case of detection of aliasing in stationary signal processes. We start with the simplest stationary processes imaginable—randomly shifted periodic signals. If we have a waveform, $x(t)$, with period T, then we can produce a stationary process by adding to t a random time shift, θ, that is evenly distributed on $[0, T)$. A sample path of our process then has the form $x(t + \theta)$ for a particular choice of θ.

Consider then the effect of under-sampling and reconstruction on a simple sine process,

$$x(t) = \sin(2\pi f t + 2\pi f \theta), \tag{3.1}$$

where θ is evenly distributed on $[0, f^{-1})$. If we under-sample with a sampling interval Δt, corresponding to the Nyquist band $[-(2\Delta t)^{-1}, (2\Delta t)^{-1})$, and then reconstruct via convolution with the sinc filter, we get the sine process given by

$$x_r(t) = \sin(2\pi \hat{f} t + 2\pi f \theta). \tag{3.2}$$

Here, $\hat{f}$ is the aliased frequency, given by $\hat{f} = f + k_f/\Delta t$ where k_f is the unique integer that places $\hat{f}$ in the Nyquist band. The key point is that the phase of the reconstructed signal is the same as the phase of the source even though the frequency has changed to the aliased value $\hat{f}$. For a process with a single harmonic, the reconstructed signal remains stationary because the phase term, $2\pi f \theta$, is evenly distributed on 2π.

Consider then a second signal process,

$$x(t) = \sin(2\pi \alpha t + 2\pi \alpha \theta) + \sin(2\pi \beta t + 2\pi \beta \theta), \tag{3.3}$$

where β is an integer multiple of α and θ is chosen randomly from the interval $[0, \alpha^{-1})$. Since the time shift, θ, is the same for both components, this is, for the various values of θ, just a shifted waveform of a given shape. Since θ is evenly distributed over the period, α^{-1}, the process is stationary.

If we sample this process at a rate low enough for both components to be aliased and then reconstruct using the sinc filter, we get

$$x_r(t) = \sin(2\pi \hat{\alpha} t + 2\pi \alpha \theta) + \sin(2\pi \hat{\beta} t + 2\pi \beta \theta), \tag{3.4}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the aliased frequencies. Although the phase terms, $2\pi\alpha\theta$ and $2\pi\beta\theta$, are still evenly distributed over 2π, they now correspond to *different* time

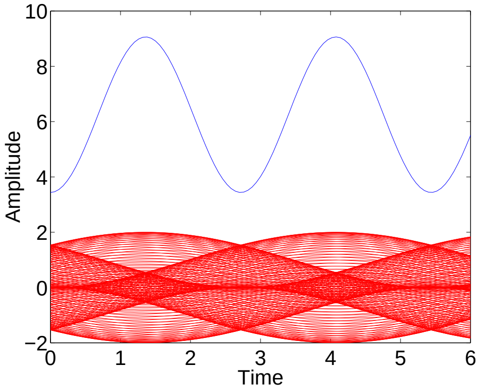


Figure 3.1: Plot illustrating the non-stationarity of a randomly shifted waveform that has been under-sampled and then reconstructed. The upper curve is the sixth moment[5]. The lower curve shows the process envelope. The original process is given by Equation 3.3 with $\alpha = 1.0$, $\beta = 3.0$, $\theta \in [0, 1)$, and $\Delta t = e = 2.71...$. (Another way of looking at the process envelope is that it is the area covered by graphs of *all* signals in the process – stationarity would imply no structure along the time axis).

shifts for the two components. Thus, we no longer have a single shifted waveform and we can expect, in general, that stationarity will have been lost.

We illustrate this loss of stationarity on an example by setting $\alpha = 1.0$ and $\beta = 3.0$ in Equation 3.3 and choosing a sample time, Δt, equal to $e = 2.71...$. We may detect the loss of stationarity by examining the envelope of the reconstructed process. We define the envelope of a process, X_t, as the support of the probability density of X_t as a function of t. Another definition that would often coincide is that the process envelope at time t is the smallest interval containing the support of the process at time t. For a process produced by randomly shifting a periodic waveform, the envelope may be conveniently displayed by plotting the sample paths corresponding to a representative collection of time shifts as in Figure 3.1. Clearly, a stationary process must have a constant envelope. If we compute the envelope for the process defined in Equation 3.4 with the parameter values that we have specified, we get an oscillating figure (see Figure 3.1). This implies that the signal is non-stationary. In fact, it is cyclostationary with period equal to the sampling interval.

[5] The sixth moment was chosen for clarity of presentation. The second moment remains constant in this case. The fourth moment does oscillate but the scale of its

As explained above, we do not lose stationarity when our original signal is a single sine wave. Nor do we lose stationarity when ratios between frequencies are preserved under the aliasing. For example, if $\Delta t = 1.0$ and the original frequencies are $(10/9, 20/9, 30/9)$, they would alias to $(1/9, 2/9, 3/9)$ and we would obtain another stationary process. But this situation is very special (non-generic). In general, more than one Fourier component is present and we do not have the special relationships between the sampling rate and the component frequencies that preserve ratios between frequencies when under-sampling. Thus, we expect that, generically, a stationary process formed by randomly shifting a periodic waveform will lose stationarity upon under-sampling and reconstruction.

Within the context of single shifted waveforms, the destruction of stationarity can occur in some remarkable situations. Consider that under-sampling and reconstruction can break stationarity even when only one of the components (β, say) is aliased and when β aliases to α or $-\alpha$. In other words, stationarity can be broken even when the two components, after under-sampling, lie right on top of each other. We can see this by choosing $\alpha = 0.25$, $\beta = 0.75$, and $\Delta t = 1.0$ in Equation 3.3. The envelope for the reconstructed process is shown in Figure 3.2 where the non-stationarity is apparent. (Of course, we cannot possibly detect this loss of stationarity by examining only a single sample path, since the sample path will never be more than a single sine wave of some amplitude and phase.)

Not all stationary processes are randomly shifted periodic waveforms. What can we say about more general stationary processes? It is clear that, if there exist *no* phase relationships between any of the Fourier components of the process, then under-sampling and reconstruction will not destroy stationarity. For a generic stationary process, though, we would expect at least some sets of components to exhibit phase relations. In that case, we would expect stationarity to be destroyed because it is difficult to imagine how the destruction of stationarity associated with one set of components could somehow be canceled out by the presence of other incommensurate components.

This argument, together with the observation that there is simply no reason to believe that stationarity should be preserved under under-sampling and reconstruction, suggests that the loss of stationarity should be a general feature of stationary processes.

oscillation is too small to allow meaningful display of the moment and the envelope on the same scale.

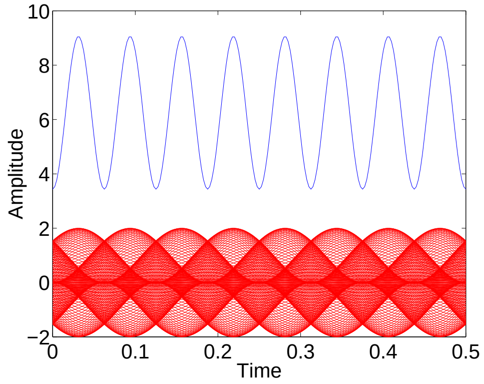


Figure 3.2: The sixth moment (upper curve) and the process envelope (lower curve) of the process given by Equation 3.3 with $\alpha = 0.25$, $\beta = 0.75$, $\theta \in [0, 4)$, and $\Delta t = 1.0$.

3.4 Detection of Aliasing in Single Sample Paths

The method that we have just used to detect aliasing in a sampled stationary process requires complete knowledge of the discrete-time process obtained by sampling the original continuous-time source. Usually, however, we have available to us only a single sample path. Therefore, we require a method for detecting aliasing in a single waveform which may or may not be a sample path of a stochastic signal process.

We may develop such a method by reconsidering the example of sampling from a square wave discussed in Section 3.2 in light of our discussion of the effect of under-sampling on stationary signal processes. Recall that the sampled time series from the square wave takes on the values -1 and 1. We may state this in statistical language by saying that the one-time probability density of the time series consists of two Dirac delta functions centered at -1 and 1, respectively. Now, we would have obtained the same one-time statistics if we had sampled the original square wave with any shift of the sampling comb. We will say that a waveform has *sampling stationarity* for a given sampling interval if the one-time sample statistics do not change as the position of the sample comb is shifted along the waveform. *Observing that the original square wave had sampling stationarity for the given sampling interval is essentially equivalent to saying that the sample times were* *not distinguished in the source.*[6] Note that the signal obtained by applying the Shannon sinc filter to the time series does *not* have sampling stationarity—the one-time statistics of the reconstructed signal vary dramatically with shifts of the sampling comb (see Figure 3.3). This lack of sampling stationarity corresponds to the distinguished role of the sample times in the reconstructed signal. Of course, this distinguished role for the sample times is what allowed us to reject the reconstructed signal as a candidate for the original source of the samples and, thus, to conclude that the sampled series contained aliased components.

This discussion suggests the following test for aliasing in signals (no underlying stochastic process assumed). Collect the statistics on the recorded samples. Reconstruct the signal at various shifts of the sampling comb and collect the statistics at these reconstructed samples. Compare with the original statistics. (We use the term "statistics" loosely, without the assumption that the samples are independent samples of some underlying probability distribution.) If we find that the reconstruction has different statistics at some shift of the sampling comb, an assumption of sampling stationarity for the original signal implies that the reconstruction is not the original signal and therefore that the signal was under-sampled.

For this test to be at all useful, two questions must be answered:

1. Are typical signals characterized by sampling stationarity?
2. Do typical under-samplings reconstruct to signals for which sampling stationarity is violated?

The answer to question 1 is clearly "yes" for sample paths of ergodic stationary processes and for signals from ergodic dynamical systems. It is also clear that there are other classes of signals which possess sampling stationarity. For example, general periodic signals (not just square waves) possess sampling stationarity if the sampling interval is incommensurate with the signal period (a generic condition). Below, we conjecture that sampling stationarity is a generic property of signals.

[6] Note that the original square wave does not have sampling stationarity for a sampling interval equal to its period. In general, a periodic signal will not have sampling stationarity with respect to sampling intervals commensurate with its period. However, the set of sampling intervals that are commensurate with a given period has Lebesgue measure zero. Clearly, the probability of choosing such a special sampling interval is 0 under the assumption that the sample times are chosen independently of the source.

The examples that we consider next suggest that the answer to the second question is also "yes".

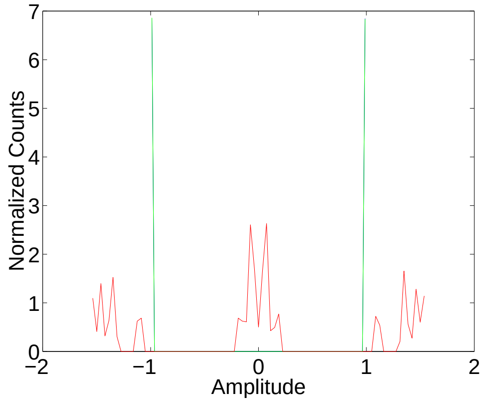


Figure 3.3: Sample statistics of data and reconstruction from a square wave. The plot shows sample statistics of the data in blue and green (which are indistinguishable) and the reconstruction in red. (See the beginning of Section 3.4.1 for an explanation of the blue and green histograms.) The red histogram is obviously very different from the blue and green with which it would coincide if the reconstruction had sampling stationarity. The blue and green histograms have been rescaled so as to make the three histograms of comparable height.

3.4.1 Examples

In each of the examples listed below, the time series to which we apply our test for aliasing was split into two interleaving series, D_1 from the samples taken at [0,2Δt,4Δt, ...] and D_2 from the samples taken at $[\Delta t, 3\Delta t, 5\Delta t, ...]$. The sample statistics corresponding to D_1 and D_2 are plotted in blue and green, respectively. For original signals with sampling stationarity, these two histograms will coincide. We then produce a reconstruction from D_1, computed at the times corresponding to D_2. The sample statistics corresponding to this reconstructed series are shown in red. If the red histogram is significantly different from the blue, then the reconstructed signal does not have sampling stationarity.

Example: For a periodic signal, the generic condition of incommensurability of the sampling interval and the signal period implies that the signal has the property of sampling stationarity. But we also find that under-sampling and reconstructing

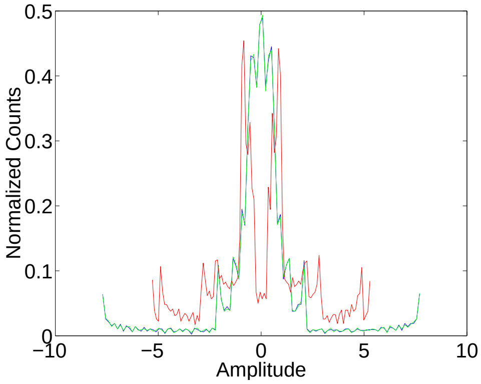


Figure 3.4: Sample statistics of data and reconstruction from a periodic signal. The blue and green histograms coincide, indicating that the original signal had sampling stationarity. The red histogram, showing the statistics of the reconstructed signal, is obviously very different from the blue, showing that the reconstruction does not have sampling stationarity.

produces a signal that does NOT have sampling stationarity, as illustrated in Figure 3.4. The data for the plot were generated by sampling ($\Delta t = e/16$) a sum of sines with frequencies $(0, 1, 2, ..., 10)$ and random amplitudes that ranged between .78 and 1.22.

Example: If our signal consists of a sum of sine waves with incommensurate frequencies, then we cannot detect aliasing by this method (see Figure 3.5). Although such a signal will have sampling stationarity, the sampling stationarity will not be broken by under-sampling and reconstruction because it is impossible to have relationships between the phases of different Fourier components (see Section 3.4.2).

Example: The previous example might lead to the suspicion that this method works only for periodic signals (or step signals such as the square wave). However, the presence of incommensurate Fourier components does not necessarily destroy the ability to detect aliasing in a periodic waveform with more than one harmonic component. Figure 3.6 shows the result of combining a periodic waveform with incommensurate harmonics. The total power in the incommensurate harmonics is about 21% of the power in the periodic waveform. The sampling stationarity of the original signal and the breakdown of sampling stationarity with under-sampling and reconstruction are apparent. This shows that, as long as *some* of our aliased Fourier components are commensurate with other components, the method can

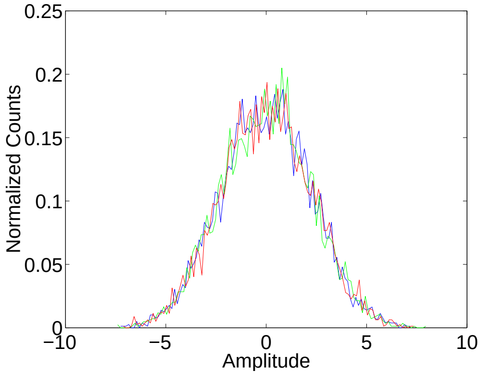


Figure 3.5: Sample statistics of data and reconstruction from a sum of sine waves with incommensurate frequencies. The blue and green histograms coincide. The red, showing the statistics of the reconstructed signal, is NOT obviously different.

work.

Example: So far, we have demonstrated that the method works for pure periodic signals and for periodic signals mixed with incommensurate harmonics. Figures 3.7, 3.8, and 3.9 show that the method works for much more complex signals with continuous spectra. The signals are taken from the Lorenz and Rössler systems (see section 3.7).

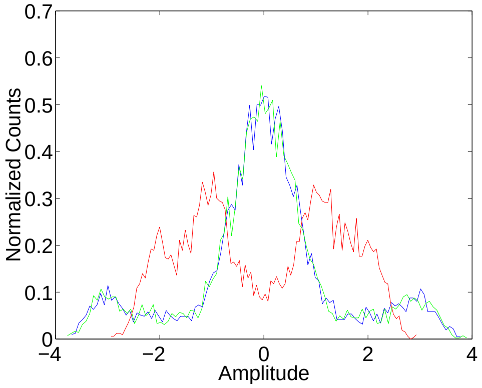


Figure 3.6: Sample statistics of data and reconstruction from a mixture of a periodic waveform and incommensurate harmonics. The blue and green histograms coincide. The red, showing the statistics of the reconstructed signal, is obviously different.

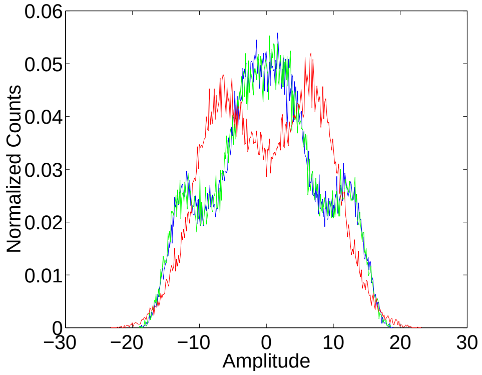


Figure 3.7: Sample statistics of data and reconstruction from the x-coordinate of the Lorenz model. The blue and green histograms coincide and the red is clearly different.

The success in detecting aliasing in time series from the Lorenz and Rössler systems suggests that the method may work for a very broad class of signals. Before attempting to determine how wide this class might actually be, we will look at the periodic case in order to begin to understand the precise mechanism of the method.

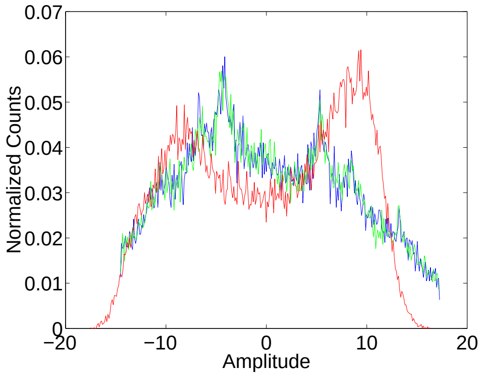


Figure 3.8: Sample statistics of data and reconstruction from the x-coordinate of the Rössler model. The blue and green histograms coincide and the red is clearly different.

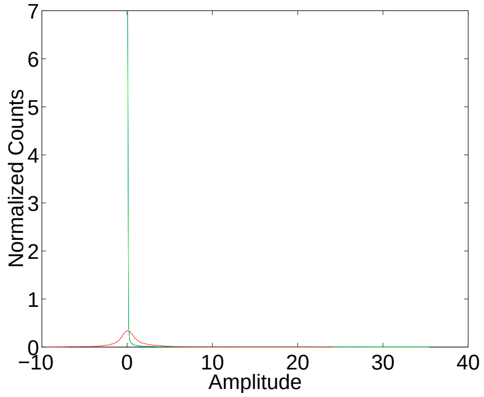


Figure 3.9: Sample statistics of data and reconstruction from the z-coordinate of the Rössler model. The blue and green histograms coincide and the red is clearly different.

3.4.2 A Closer Look at the Periodic Case

If one samples a periodic signal incommensurately with the signal period, the samples end up mixing evenly around the waveform (see section 3.8). Thus, all shifts of a sampling comb with a sampling interval that is incommensurate with the period will produce the same statistics. This implies that:

Theorem 3.4.1. *A periodic signal will have sampling stationarity with respect to any sampling interval that is incommensurate with the period of the signal.*

Conversely, sampling with an interval that *is* commensurate with the period will, in general, produce statistics that depend on the sampling shift. Theorem 3.4.1 implies that:

Theorem 3.4.2. *Every periodic signal has sampling stationarity with respect to all sampling intervals except for a set of intervals with (Lebesgue) measure zero.*

Thus, the probability of choosing a sampling interval for which a given periodic signal does not have sampling stationarity is zero, provided the interval is chosen independently of the signal.

What, then, is the effect of under-sampling and reconstruction on this sampling stationarity? The sample statistics are determined by the shape of the waveform (see section 3.8 for the exact formula). It can be shown that the process of undersampling and reconstruction is equivalent to sampling the original waveform at the same rate with the individual Fourier components shifted with respect to each other. When different components experience different time shifts, the shape of the effective waveform changes. Consequently, the statistics change. We now explain this in detail.

When we sample a single harmonic with frequency f and phase φ every Δt time units, we get the values

$$y_n = \sin\left(2\pi f n \Delta t + \varphi\right) \qquad n \in Z. \tag{3.5}$$

We will temporarily suppress the phase and rewrite this expression as

$$\begin{aligned} \sin\left(2\pi f n \Delta t\right) &= \sin\left(2\pi f n \Delta t + 2\pi k n\right) \\ &= \sin\left(2\pi\left(f + \frac{k}{\Delta t}\right) n \Delta t\right) \end{aligned} \tag{3.6}$$

for any integer k, so that the reconstruction of this component at points $1+s, 2+s, 3+s, \ldots$ is given by

$$\hat{y}_{n,s} = \sin\left(2\pi\left(f + \frac{k_f}{\Delta t}\right)\left(n+s\right)\Delta t\right), \tag{3.7}$$

where the reconstruction chooses precisely one of the integral k's, which we will call k_f, such that $f + k_f/\Delta t$ is in the interval $[-1/2\Delta t, 1/2\Delta t)$. We can now rewrite the reconstructed harmonic as

$$\begin{aligned} \hat{y}_{n,s} &= \sin\left(2\pi\left(f + \frac{k_f}{\Delta t}\right) n \Delta t + 2\pi\left(f + \frac{k_f}{\Delta t}\right) s \Delta t\right) \\ &= \sin\left(2\pi f n \Delta t + 2\pi k_f n + 2\pi f s \Delta t + 2\pi k_f s\right) \\ &= \sin\left(2\pi f n \Delta t + 2\pi f s \Delta t + 2\pi k_f s\right) \end{aligned} \tag{3.8}$$

where we drop $2\pi k_f n$ since k_f is an integer. *Thus, the reconstructed signal has samples at a shift, s, as though we were sampling the original waveform, but with the phase of the individual Fourier component shifted by the amount $2\pi f s \Delta t + 2\pi k_f s$.* The first term amounts to a time shift which is the same for all the components in the waveform. This implies that these first terms do not change the shape of the waveform and can be ignored. So we may consider the effective waveform (at a shift s) to be

$$\sum_i A_i \sin(2\pi f_i n \Delta t + 2\pi k_{f_i} s + \varphi_i) \tag{3.9}$$

where we have reinserted the phase. The term $2\pi k_{f_i} s$ amounts to a time shift that is different for different f_i. This difference in time shifts leads to a change in the shape of the effective waveform as s changes which in turn changes the sample statistics.

For a given waveform, it is clear that almost any change in the shape of the waveform will change the sample statistics. (For example, a generic choice of s will change the heights of the extrema, changing the locations of the singularities in the histogram. See section 3.8.) Thus, we conclude that *generically, periodic signals have sampling stationarity which is destroyed by under-sampling and reconstruction.*

3.4.3 Non-periodic Signals

Now, we want to use the insight that we have gained for the case of periodic signals to get a better understanding of the answers to the two questions at the end of Section 3.4. We begin with some general questions about the kinds of signals to which our method might possibly apply.

Consider first the case of transient signals. In order to be able to talk about sampling stationarity at all, we have to be able to take as many samples as we want (at the given sampling rate) in order to be able to estimate the one-time probability distribution to arbitrary accuracy. This implies that we must think of our signals as functions of infinite time. In this context, any transient signal has trivial sampling stationarity—the probability distribution is a delta function at zero. By the same token, under-sampling and reconstruction will not destroy this sampling stationarity. Thus, we need to restrict our attention to persistent (non-transient) signals.

Within the class of persistent signals, it is clear that we need the signals that we consider to have well-defined sample statistics for arbitrary sampling intervals and shifts. Given that we are discussing aliasing, our signals also need to have a Fourier transform (in some sense). The set of signals with well-defined power spectra (which will have, in general, singular components) will clearly meet these criteria, although the actual class to which our method applies may be larger. In the following, then, we may take the term *persistent signal* to refer to a signal with a well-defined, nonzero power spectrum.

Consider then the question of which signals have sampling stationarity for which sampling intervals. In the case of periodic signals, Theorems 3.4.1 and 3.4.2 provide what is essentially a complete answer—sampling stationarity holds for generic choices of signals and sampling intervals. At first glance one might try to generalize Theorem 3.4.1 to the following:

Conjecture 3.4.1. *Every signal has the property of sampling stationarity for every sampling interval Δt that is not commensurate with the period of any* singular *component of its spectrum.* (FALSE)

Unfortunately this conjecture is false as may be seen from the following counterexample. If we under-sample and reconstruct a signal with a purely continuous spectrum (such as our signal from the Lorenz system), we will introduce no new singular components. Thus, the reconstructed signal will have a purely continuous spectrum. If the conjecture were true, then, such a reconstructed signal would have sampling stationarity for all sampling intervals by virtue of having no singular components. Yet it is just the *lack* of sampling stationarity of this signal with respect to the given sampling interval that allows us to detect aliasing in this case. Thus, we know that there exist signals that lack sampling stationarity with respect to sampling intervals that are not commensurate with any singular component of their spectra and the conjecture is false. However, the reconstruction of an under-sampled signal has a very special relationship to the interval with which the sampling was done. Thus, one expects that re-sampling the reconstruction with a new sampling interval *not related to the original interval* will yield statistics that are again stationary with respect to shifts in the sampling comb. Therefore, we arrive at the following conjecture:

Conjecture 3.4.2. *Every signal has the property of sampling stationarity for* every *sampling interval Δt,* except *a set of Δt's with (Lebesgue) measure zero.*

This conjecture implies that, if one were to observe the reconstructed statistics varying with changes in the shift, this observation would be enough to conclude (with probability 1) that the samples came from an under-sampled waveform. In other words, *the truth of the conjecture would imply that the detection of under-sampling by the proposed method is generically free of false positives*

Next, we want to know when under-sampling and reconstruction of persistent non-periodic signals will yield new signals which *have* the property of sampling stationarity. (In other words, we also want to know when we can get false negatives.) The analysis that we have presented for periodic signals *suggests* that under-sampling and reconstruction should destroy sampling stationarity for general persistent signals in which at least some of the aliased Fourier components are commensurate with other components of the signal[7]. The reasoning is that each individual component can be regarded as a part of a family of harmonics and that the effective shape of the waveform associated with this family is changing with shifts of the sampling comb. There does not appear to be any reason to believe that combining different periodic waveforms, each of which is changing its sample statistics with shifts of the sampling comb, would result in sample statistics that do not change. Therefore, *it seems likely that, generically, persistent signals have sampling stationarity that is destroyed by under-sampling and reconstruction.* In order to turn this last statement into a well defined conjecture, it will be necessary to define precisely what is meant by "generically" in the case of persistent non-periodic signals. The question of exactly how to define "persistent" must also be answered. Since the transformation that takes us from a waveform to sample statistics is extremely nonlinear, a proof is likely to be difficult.

3.4.4 Recovery of High-Frequency Information

The next question that presents itself is whether or not we can recover information about individual aliased Fourier components using the sampling-shift dependence of the reconstructed statistics. Ideally, we would like to know how much of the signal at an individual frequency, f, in the Nyquist band comes from each frequency that aliases to f.

Consider the one-time probability density of the reconstructed signal, $p(x)$, as a function of both x and shift s. We call this two-dimensional surface a Reconstructed Sample Statistics (RSS) plot (see figures 3.10 and 3.11 for example RSS plots).

The RSS plot has dependencies on s tied directly to the quantities k_{f_i}. Each k_{f_i}, in turn, determines the particular copy of the Nyquist band in which its corresponding f_i is located. This chain of dependencies suggests that the RSS plot contains the information necessary to determine the contribution of each band to the signal at a given frequency in the Nyquist band. The inverse problem is greatly complicated by the interaction of the Fourier components and the nonlinear "pro-

[7] Note that the condition that the original signal must have harmonically related components (i.e. commensurate components) will be satisfied by any signal with a nonzero continuous part to its spectrum as well as by periodic signals.

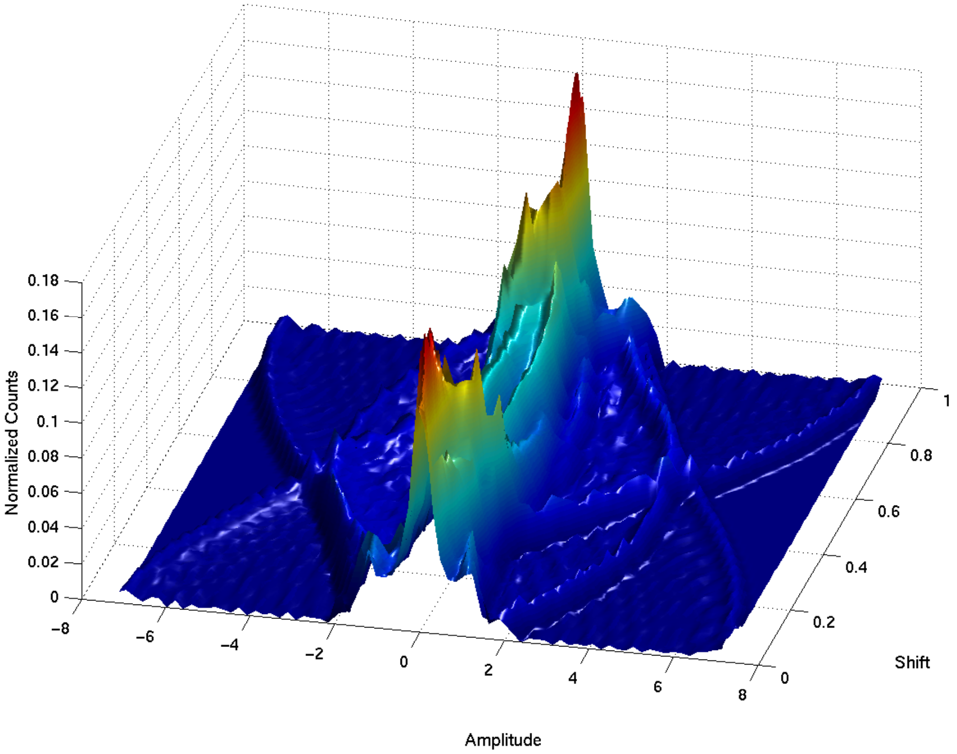


Figure 3.10: Example RSS Plot: A signal $f(t) = \sum_{i=0}^{10} c_i \sin(2\pi i t)$ with the c_i's chosen randomly in $[.75, 1.25]$, was sampled every $.0625 * \exp(1)$ time units ($=$ $.1698...$ time units). Since the Nyquist $\Delta\tau$ is is $.05$ time units, the signal is *undersampled* and the as the plot shows, the sample statistics change with the reconstruction shift.

jection" that turns the waveform into statistics. This nonlinear inverse problem will be a major focus of future work on the detection (and possibly correction) of aliasing.

3.5 Directions for Further Investigations

In addition to the work already alluded to on the inverse problem formed by the RSS plots, there are other issues to explore. Included among them are:

- What are the effects of noise on this method for detection of aliasing?

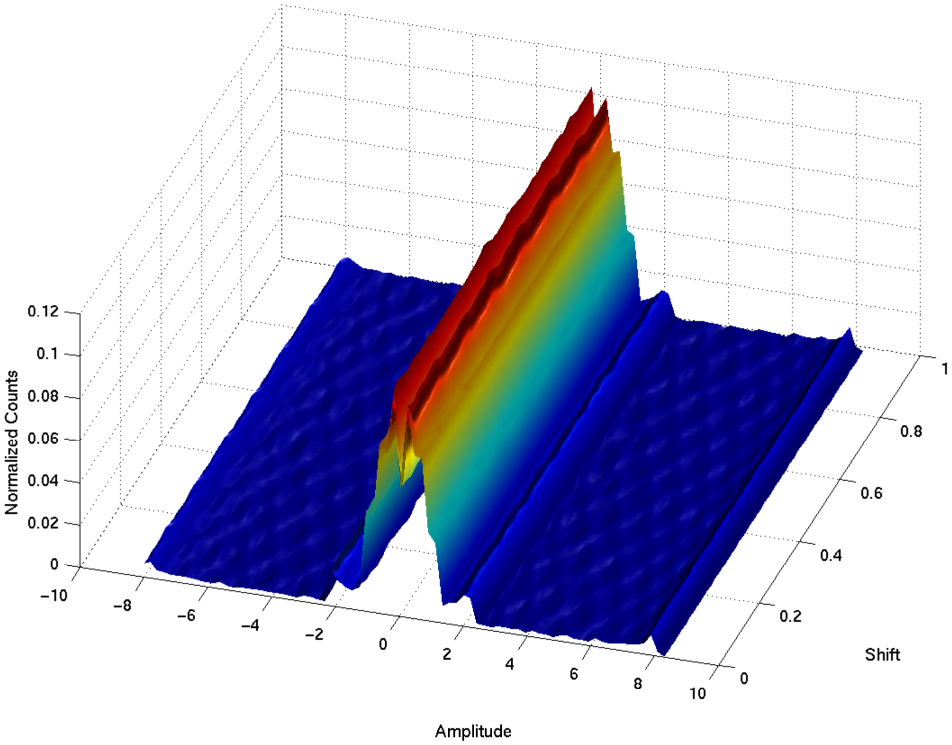


Figure 3.11: Another Example RSS Plot: A signal $f(t) = \sum_{i=0}^{10} c_i \sin(2\pi i t)$ with the c_i's chosen randomly in $[.75, 1.25]$, was sampled every $.0125 * \exp(1)$ time units ($= .03398...$ time units). Since the Nyquist $\Delta\tau$ is is $.05$ time units, the signal is *properly sampled* and as the plot shows, the sample statistics **do not** change with the reconstruction shift.

- What is the effect of near commensurability of sample interval and signal period?
- What is the effect of finite time-series length?
- How does the departure of the statistics of the reconstructed signal from stationarity depend on the fraction of the total power that lies outside the Nyquist band?

These questions are important to the practical usefulness of the method of high-frequency detection/recovery.

3.6 Conclusion

Although the idea of detection of aliasing is typically dismissed with references to the Nyquist criterion and the Shannon reconstruction theorem, we have demonstrated that detection of aliasing is possible with what appear, at first glance, to be very weak prior assumptions. The key concept is that of sampling stationarity. We emphasize that *this concept makes sense for single waveforms.* Although this concept arose in the consideration of step signals like square waves, its usefulness extends far beyond these signals. In particular, our method enables the detection of aliasing in samples from nontrivial waveforms such as measurements from motion on the Lorenz or Rössler attractors. As indicated above, many questions remain. Some of these are important for the practical utility of the concept of sampling stationarity and the associated RSS plots.

3.7 After-notes: Computational Details

The calculations represented in the paper were done with Matlab. The Lorenz equations,

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(R - z) - y \\ \dot{z} &= xy - bz, \end{aligned} \tag{3.10}$$

were integrated with parameter values of $\sigma = 10$, $R = 28$, and $b = 8/3$ using Matlab's "ODE45" which is an adaptive step size routine. Relative tolerance was set to the default value of 1.0×10^{-3} and absolute tolerance was set to the default 1.0×10^{-6}. Initial conditions were set at $x = y = z = 1$. Values for the x, y, and z coordinates were saved every 0.5 time units. 200,001 samples were taken and split into two interleaving time series each 100,000 samples long. The first series was used to reconstruct a signal via convolution with a sinc filter of length 200,001. These very long series and filters were used to minimize the effects of truncating the convolution at the ends of the series. The histograms for the reconstructed signal were computed from the middle 50% of the reconstructed series.

The Rössler equations,

$$\begin{aligned}\dot{x} =& - z - y \\ \dot{y} =& x + ay \\ \dot{z} =& b + z(x - c),\end{aligned} \tag{3.11}$$

were integrated in the same way with a sampling interval of 10 time units, and parameter values of $a = 0.15$, $b = 0.2$, and $c = 10$. In this way 200,001 samples were obtained and the splitting and reconstruction were done as described for the Lorenz equations.

All the histograms presented here were originally calculated from much shorter time series ($10,000$ samples). The features that allow us to conclude that aliased components are present were all clearly visible in the histograms made from shorter time series although, of course, the histograms were considerably rougher. We conclude that the results that we have presented are certainly not an artifact of finite-length time series.

3.8 After-notes: Sample Statistics for a Periodic Signal

If one samples a periodic signal, $h(t)$, incommensurately with the signal period T, the samples end up mixing evenly around the waveform[8]. The resulting histogram is proportional to the reciprocal of the derivative of the waveform. This follows from the fact that the probability of getting any particular t (position along the waveform) is uniformly distributed over $[0, T)$ which in turn implies that the probability of the interval $[y, y + dy)$ is the probability of the corresponding dt or $(1/T)(dy/h^{'}(t))$. More precisely, the probability density for y is

$$p(y) = \frac{1}{T} \sum_{t_\alpha \in \mathcal{T}(y)} (h^{'}(t_\alpha))^{-1} \tag{3.12}$$

[8] Proving this is surprisingly difficult. An equivalent problem is the determination of the density of points obtained by repeated iterations of an irrational rotation on a circle of unit circumference. The resulting distribution of points satisfies $n_{[a,b)}/N \approx (b - a)$ where $n_{[a,b)}$ is the number of iterates in $[a, b)$, N is the total number of iterates, and $[a, b) \subset [0, 1)$. See [89, p.39-40,29] for details.

where

$$\mathcal{T}(y) = \{t \mid t \in [0, T), h(t) = y\} \tag{3.13}$$

Note that the density will have $1/\sqrt(y)$ singularities at the local maxima and minima of $h(t)$. The form of the singularities follows from the fact that a generic waveform has maxima and minima with nonzero second derivative.

Chapter 4

Reconstruction from projections using dynamics: Noiseless Case

Kevin R. Vixie [1]

4.1 Introduction

Reconstructing a series of 3 dimensional density distributions from a finite number of 2 dimensional measurements is impossible unless prior assumptions of some sort are used [90]. The difficulty comes from the fact that, without fairly strict assumptions, many different density fields project to the same radiograph. To state this another way, a radiographic measurement device, thought of as a projection operator, has a nontrivial null space. Our approach in this paper is to discretize the object space and the radiograph (measurement) space. We then combine a sequence of radiographic measurements into one super-measurement. Combined with the operator which determines dynamics, the single time projection operator can be turned into an extended projection operator that maps a sequence of objects into a super-measurement. Due to the dynamical constraints, the dimension of the object sequence space does not grow as the length of that sequence increases. On the other hand, the size of the data space (the space of super-measurements) does, implying that eventually, the extended projection operator has a trivial null space.

Now a look ahead. In section 4.2, we briefly outline the problem. Section 4.3 outlines the notation used for the rest of the paper. In section 4.4 we introduce

[1] The research for this chapter was done in collaboration with Gary Sandine.

the problem in it's linear setting and give a technical overview of our results. The key notion turns out to be that of *transversality*. An intuitive footing for the entire paper is given. In the following section (sec.4.5), we illustrate the ideas with simple numerical examples. Section 4.6 contains the technical results of the paper. In subsection 4.6.1, we introduce and prove a theorem which bounds how slowly the dimension of the projection operator null space decreases as the number of measurements incorporated in the super-measurement goes up. There is a lower bound on the number of measurements that are needed to get a unique inversion. If $n/d =$ (dimension of the object space) $\div$ (dimension of the measurement space), then the lower bound is simply $\lceil n/d \rceil$ – the smallest integer greater or equal to n/d. We will say that a particular combination of linear system, L and measurement projection, P has the optimal reduction property if the number of measurements needed to get an unique inversion equals this lower bound. Subsection 4.6.2 looks at the prevalence of linear systems which (w.r.t. a fixed P) having the optimal reduction property. Next, in subsection 4.6.3, we outline a proof of the extension of one of the results to the case of nonlinear dynamics. The relation to known results is discussed in section 4.7. We close with a summary and discussion.

4.2 The Problem

Radiographic experiments measuring very fast events typically produce data consisting of a sequence of 2-d projections. These 2-d projections are created by bombarding some 3-d distribution of density – the object – with penetrating radiation of some sort such as high-energy x-rays or protons. The number of angles at which the data is taken is typically 1. We idealize this to get the following model. The object will be a point x in an object space X which changes from measurement to measurement as dictated by a linear operator L, acting on X. The measurements d which lie in the measurement space D will be generated by the action of a measurement or projection operator P. Thus, if the object and measurement at time $t \in \mathbb{N}$ are denoted x_t and d_t respectively, we can express the actions of the operators L and P in the following way: $x_{t+1} = L(x_t)$ and $d_t = P(x_t)$. See figure 4.1. We define the extended (or experimental) spaces to be the product spaces

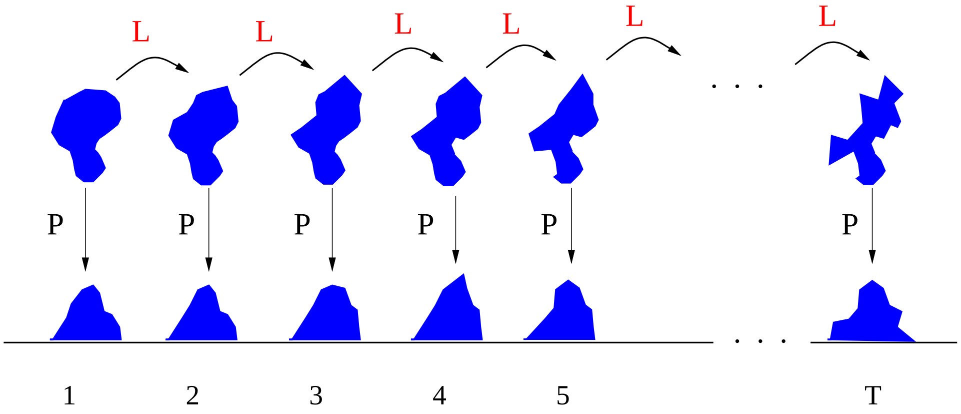


Figure 4.1: The Problem

$\tilde{X} \equiv X^T$ and $\tilde{D} \equiv D^T$ where T is the number of observation times in a particular experiment. If we have a particular sequence of points in the object space, then this sequence is a single point $\tilde{x} = (x_1, x_2, ..., x_T)$ in $\tilde{X}$. The measurement process produces a point $\tilde{d} = (d_1, d_2, ..., d_T)$ in the extended (or super-)measurement space $\tilde{D}$. This can be succinctly expressed using the extended projection operator $\tilde{P} \equiv P^T$ since then $\tilde{d}_t = \tilde{P}(\tilde{x}_t)$.

If A is defined to be the T-1 by T block matrix,

$$\begin{bmatrix} L & -I & 0 & 0 & \dots & 0 \\ 0 & L & -I & 0 & \dots & 0 \\ 0 & 0 & L & -I & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & & L & -I \end{bmatrix} \tag{4.1}$$

then the null space of this operator, denoted N_A, is exactly the set of elements of $\tilde{X}$ which satisfy the dynamics. That is, $N_A = \{\tilde{x} \in \tilde{X} | x_{t+1} = Lx_t \text{ where } \tilde{x} = (x_1, x_2, ..., x_T)\}$ = the set of sequences in $\tilde{X}$ which satisfy $x_{t+1} = Lx_t$.

Let $N_{\tilde{P}}$ be denote the null space of $\tilde{P}$. The inverse problem is now solvable when $N_A \cap N_{\tilde{P}} = \{0\}$.

4.3 A Pause for Notation

We now establish the notation we will use throughout, except in section 4.6.3 where we find it more convenient to modify the notation. This section should be used as a reference.

$T \equiv$ The number of radiographs.

$X \equiv$ The Space of objects - we will use $\mathbb{R}^n$.

$x \equiv$ An element of X.

$\mathcal{B} \equiv$ A basis for X. It has precisely n elements.

$b_i \equiv$ the ith element of $\mathcal{B}$.

$\tilde{X} \equiv X^T = X \times X \times ... \times X$.

$\tilde{x} \equiv$ An element $(x_1, ..., x_T)$ of $\tilde{X}$.

$\tilde{b}_i \equiv$ A basis element of $\tilde{X}$ given by $(b_i, L(b_i), ..., L^{T-1}(b_i))$. ($L$ assumed invertible.)

$\tilde{\mathcal{B}} \equiv$ A basis of $\tilde{X}$ given by $\tilde{b}_i$ for $i = 1, ..., n$.

$D \equiv$ The space of radiographs - we will use $\mathbb{R}^m$.

$d \equiv$ An element of D.

$\tilde{D} \equiv D^T = D \times D \times ... \times D$.

$\tilde{d} \equiv$ An element $(d_1, ..., d_T)$ of $\tilde{D}$.

$L \equiv$ The linear operator on X that gives the dynamics: $x_{i+1} = L(x_i)$.

$P \equiv$ The projection (measurement) operator $P : X \to D$. We assume that P is full rank since otherwise we may choose a smaller D and consider P with this restricted range to get a full rank P.

$N \equiv$ The null space of P.

$p \equiv$ The dimension of the null space of P.

$\tilde{P} \equiv$ The extended or product projection operator $\tilde{P} : \tilde{X} \to \tilde{D}$.

$$\tilde{P} = \begin{bmatrix} P & 0 & 0 & 0 & ... & 0 \\ 0 & P & 0 & 0 & ... & 0 \\ 0 & 0 & P & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & & 0 & P \end{bmatrix} \tag{4.2}$$

$A \equiv$ An operator from $\tilde{X}$ to X^{T-1}

$$\begin{bmatrix} L & -I & 0 & 0 & ... & 0 \\ 0 & L & -I & 0 & ... & 0 \\ 0 & 0 & L & -I & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & & L & -I \end{bmatrix} \tag{4.3}$$

$N_A \equiv$ null space of $A =$ set of $\tilde{x}$ such that $x_{i+1} = L(x_i)$ for $i = 1, ..., T-1$.

$\tilde{P}_{N_A} \equiv \tilde{P}(N_A)$.

$[A, ..., C] \equiv$ The Cartesian product $A \times ... \times C$.

And we use the following standard notation.

$\mathbf{dim}(H) \equiv$ The dimension of the space or subspace H

$M^{\perp} \equiv$ The orthogonal complement of M.

And the almost standard notation ...

$S \pitchfork Q \equiv S$ intersects Q transversely,

with the slight twist that when S and Q are subspaces, the intersection always includes the zero vector. So for example, a transverse intersection of two 1-dimensional subspaces of R^3 is just $\{0\}$, instead of empty intersection as would be the case for two 1-dimensional curves in R^3.

4.4 The Solution: An Overview

Let $X = \mathbb{R}^n$. Pick a basis, $\{b_i\}$, $i = 1...n$, for X. Then N_A is spanned by $\tilde{b}_i \in \tilde{X}$ where $\tilde{b}_i = (b_i, L(b_i), ..., L^{T-1}(b_i))$.

Example 4.4.1. *Suppose that the linear operator L has a complete set of eigenvectors $\{\omega_i\}$, where i goes from 1 to n. Then N_A is spanned by $\tilde{\omega}_i \in \tilde{X}$ where $\tilde{\omega}_i = (\omega_i, L(\omega_i), ..., L^{T-1}(\omega_i)) = (\omega_i, \lambda_i\omega_i, ..., \lambda_i^{T-1}\omega_i)$.*

Let $D = \mathbb{R}^m$. Form $E_{\tilde{P}(\tilde{b})}$, the n by mT matrix where row i is $\tilde{P}(\tilde{b}_i) = (P(b_i), PL(b_i), ..., PL^{T-1}(b_i))$. Then there is a unique $\tilde{x}$ for every $\tilde{P}(\tilde{x})$ iff the rank of $E_{\tilde{P}(\tilde{b})}$ is n. (This is virtually identical to the usual test for observability from control theory, except that I am not assuming that $T = n$. See for example [68] p. 178 or [91] p. 271). We would like to know how big T needs to be: how many measurements do we need? Before diving into the theorems and their proofs answering this question, we pause to give an intuitive overview of the ideas behind the theorems as well as a peek at the results themselves. We also insert a section containing a simple numerical example just before the section with all the main results.

Let us look at $N_A \cap N_{\tilde{P}}$ a little more carefully. If the "true" sequence in $\tilde{X}$ is $\tilde{x}^*$ and we have measured $\tilde{d}^* = \tilde{P}(\tilde{x}^*)$, then any $\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in N^T$ can be added to $\tilde{x}^*$ without changing the observed data $\tilde{P}(\tilde{x}^*)$. In other words, every point in $[x_1^* + N, x_2^* + N, ..., x_T^* + N]$ generates identical measurements – for every $\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in N^T$, $\tilde{d}^* = \tilde{P}(\tilde{x}^* + \mathfrak{n}^*)$. Since we are only interested in dynamically possible sequences, we may look at the smaller set of $\mathfrak{n}^*$ which are in fact *null orbits*. We can see this as follows:

- Notice that if the state at time 1 was a member of the set $x_1^* + N$, then the state at time 2 is not only a member of $x_2^* + N$, but is in fact in $L(x_1^* + N)$.
- Therefore, the state at time 2 is in $L(x_1^* + N) \cap x_2^* + N$.
- Likewise, the state at time 3 is in $L(L(x_1^* + N) \cap x_2^* + N) \cap x_3^* + N$.

We may continue in this fashion, but instead let us simplify things by using the linearity of L.

- $L(x_1^* + N) = L(x_1^*) + L(N) = x_2^* + L(N)$
- Therefore, $L(x_1^* + N) \cap x_2^* + N = x_2^* + L(N) \cap N$.
- $L(L(x_1^* + N) \cap x_2^* + N) = L(x_2^* + L(N) \cap N)$.
- Therefore, $L(L(x_1^* + N) \cap x_2^* + N) \cap x_3^* + N = x_3^* + L(L(N) \cap N) \cap N$.
- And so on ...

Continuing this, we get

$$\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in [N, N \cap L(N), N \cap L(N \cap L(N)), ...]. \tag{4.4}$$

and defining

$$\begin{array}{ll} N_1 & \equiv N \\ N_2 & \equiv N \cap L(N_1) \\ N_3 & \equiv N \cap L(N_2) \\ \dots & \dots \\ N_T & \equiv N \cap L(N_{T-1}). \end{array} \tag{4.5}$$

we arrive at the requirement that null orbits $\mathfrak{n}^*$ (which may be added to the "real" orbit without changing the measurements) satisfy,

$$\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in [N_1, N_2, ..., N_T]. \tag{4.6}$$

Now assume that L is invertible. If $N_T = \{0\}$, it follows that $E_{\tilde{P}(\tilde{b})}$ has rank $= n$ and $N_A \cap N_{\tilde{P}} = \{0\}$.

The obvious question then is, "What is the smallest T for which $N_T = \{0\}$?" The approach we will take to answer this is to notice that the *dimension* of N_i steps down at a maximal rate when the intersections defining the N_i are all *transverse*.

We go into this in much more detail in section 4.6, but the intuitive idea is that two subspaces will typically intersect so as to have an intersection of minimal dimension – we expect that the sequence of intersections will in fact step down in dimension as fast as possible. How fast is this? If W_1 and W_2 are a pair of randomly chosen subspaces of $\mathbb{R}^n$ having dimension (or degrees of freedom) k_1 and k_2 respectively, then we would expect together, they would have $k_1 + k_2$ degrees of freedom. But of course this number can not exceed n. So the degrees of freedom in W_2 *not contained* in W_1 would (typically) be exactly $n - k_1$. The intersection $W_1 \cap W_2$ contains the degrees of freedom of W_2 which *are contained* in W_1. So $\dim(W_1 \cap W_2) + n - k_1 = k_2$. This leads us to:

Statement 4.4.1. *A typical intersection of subspaces W_1 and W_2 with $k_1 = \dim(W_1)$ and $k_2 = \dim(W_2)$, has the property that*

$$\dim(W_1 \cap W_2) = k_2 + k_2 - n \tag{4.7}$$

This leads simply to the fact that if the dimension of the null space is $n - d$ and somehow L is "typical" then $\dim(L(N) \cap N) = (n-d) + (n-d) - n = n - 2d$. Continuing, we get that the dimension of the intersection steps down by d each time. And of course this is what we would naively expect ... each measurement of d quantities reduces the degrees of freedom that we do not know by d. And this leads to the answer that $T = \lceil n/d \rceil$.

We begin section 4.6 by examining exactly how a T for which $E_{\tilde{P}(\tilde{b})}$ is full rank, can fail to exist. Quite simply, we can fail to have unique solutions when L preserves some subspace of N. This is also not surprising since in this case we have an entire subspace of $\tilde{X}$, each point of which is in fact a *null orbit* of L. We then show that typical dynamics L give us exactly the optimal T. Finally, we carefully conjecture a path to a similar result for the case in which neither P nor L are linear.

Statement 4.4.2 (Technical Overview Summary). *The problem of determining the object sequence from the sequence of measurements reduces – in the linear case – to the inversion of a particular matrix. In the noiseless case, existence of solutions is not in question. Uniqueness of a solution that is guaranteed to exist is an important question and the fact that any null sequence $\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in N^T$ can be added to any state sequence without changing the measurements illustrates the problem. Since we are assuming that the state sequence satisfies the dynamics, we find that null sequences which give non-uniqueness are in fact* null orbits. *Careful consideration of the dynamical constraints shows in fact that these* null orbits *are elements of a particular set of intersections. Since* transversality *of intersections is typical (in a sense to be defined more carefully below), we expect that the number of observations needed to get unique invertibility is in fact* $T = \lceil n/d \rceil$.

Remark 4.4.1. *Note that* $\mathfrak{n}^* = (\mathfrak{n}_1, \mathfrak{n}_2, ..., \mathfrak{n}_T) \in [N, N \cap L(N), N \cap L(N \cap L(N)), ...]$ *is actually more than the set of "null" solutions. It turns out to be small enough to meet our needs.*

Remark 4.4.2. *If* L *is not invertible, then* $N_T = \{0\}$ *does not imply that* $N_A \cap N_{\tilde{P}} = \{0\}$ *since* $N_A \cap N_{\tilde{P}} \in \{\textit{The Set } (L^{-1}(...L^{-1}(0)...)), ..., L^{-1}(0), 0) \cap N^T\}$. *More precisely, it can be all the orbits of points in* $L^{-T+1}(0) \cap N$. *Therefore, trivial* N_T *is a necessary but insufficient condition for the invertibility of the data* $\tilde{d}^*$ *when* L *is not invertible. Consider the following example.*

Example 4.4.2. *Suppose that* $X = \mathbb{R}^{20}$ *and that* $P : x \to (x_{20}) \in \mathbb{R}^1$. *This implies that the null space of* P *is* 19 *dimensional. Suppose that* L *is the nilpotent operator given by the* 20×20 *matrix*

$$\begin{bmatrix} 0 & 1 & 0 & 0 & ... & 0 \\ 0 & 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & & 0 & 0 \end{bmatrix}. \tag{4.8}$$

Suppose that $T = 21$. *Since* $L^{T-1} = 0$, *we have that* $N_T = \{0\}$. *Now since* $L^{-T+1}(0) \cap N = N$ *maps into* $L^{-T+2}(0) \cap N = L^{-T+2}(0)$ *and* $L^{-T+2}(0)$ *maps into* $L^{-T+3}(0) \cap N = L^{-T+3}(0)$ *and so on, we have that the orbits of every point in* $L^{-T+1}(0) \cap N = N$ *is an element of* $N_A \cap N_{\tilde{P}} = \{0\}$. *Therefore even though* $N_T = N_{21} = \{0\}$ *we still have a* 19 *dimensional* $N_A \cap N_{\tilde{P}} = \{0\}$*!*

Remark 4.4.3. *If the dynamics and projection are not linear, the ideas remain the same. First, let us look at the linear case a little differently. Notice that the* $x_i + N$ *are actually the* level sets *of* P *at* d_i. *These level sets are in fact simply translates of each other. Linearity implies that* $L(x_i + N) = L(x_i) + L(N)$ *and this permits us to look exclusively at iterates of the* N. *Moving to the nonlinear case, we now have level sets of* P *which are not simply translates of each other and we can no longer use linearity to decompose the operation of the dynamical operator which we will denote* F. *If we define the level sets* $N_{d_i} \equiv P^{-1}(d_i)$, *the intersections of interest are* $F(N_{d_1}) \cap N_{d_2}$, $F(F(N_{d_1}) \cap N_{d_2}) \cap N_{d_3}$, *and so on. We look at this in much greater detail in section 4.6.3.*

4.5 The Solution: Numerical Examples

We now give two examples in which we apply the technique described in section 4.4 to invert simulated sequences of (noiseless) radiographs using one view. Our purpose in this section is simply to demonstrate the procedure. We assume our object lies within a 10×10 pixelation and has constant density within each pixel, so the object space X is $\mathbb{R}^{100}$. We use the same initial condition with two different linear operators L_1 and L_2 which we describe below. In each case, the projection P sums the values down the columns of the pixelation. We use, in a sense, the largest parameterization of our object space; namely, we assume nothing about the object and seek to determine the value in each pixel. At the end of this section, we comment briefly on the poor numerical conditioning of these problems and indicate first steps taken to improve the numerics. This is a subject of current study.

To reiterate the procedure, first choose a basis $\{b_i\}_{i=1}^{100}$ for X. With L and P representing the dynamics and projection operators respectively, we build a $100 \times 10t$ matrix E where the ith row of E is $(Pb_i, PLb_i, PL^2 b_i \ldots, PL^{t-1}b_i) = \tilde{P}_t \tilde{b}_i$. As soon as t is large enough so that rank $E = 100$, we have a unique solution x for the equation $xE = \tilde{d}^*$. Since we know the dynamics L, we can then reconstruct the sequence $\tilde{x}^* = (x, Lx, L^2x, \ldots, L^{t-1}x)$.

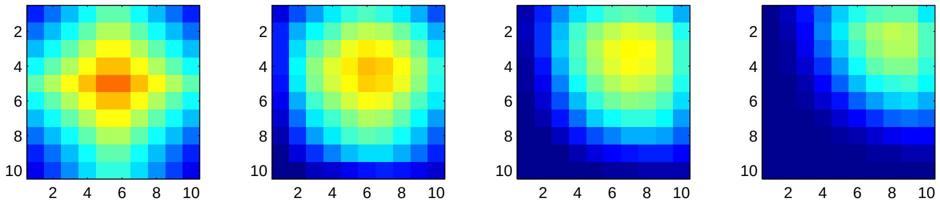
In each example, we chose the canonical basis $\{e_i\}_{i=1}^{100}$ for X where $e_i(j) = \delta_{ij}$ $(1 \leq j \leq 100)$. The first linear operator L_1 can be described as a combination of a diffusion and a shift. The best way to describe L_1 is as a two step process:

Diffuse spread the mass at any pixel to itself and the neighboring pixels: the contribution to $p_{new}(i,j)$ from $p_{old}(i,j)$ is simply $(1-D) * p_{old}(i,j)$ and the contribution to $p_{new}(i \pm 1, j \pm 1)$ from $p_{old}(i,j)$ is simply $f_{i \pm 1, j \pm 1} * D * p_{old}(i,j)$ where $\sum f_{i\pm 1, j \pm 1} = 1$.

Shift Now set $p_{new}(i,j) = p_{new}(i+1, j-1)$.

where we have ignored the diddling we must do at the boundaries and we are using $p_{old}()$ and $p_{new}()$ to denote pixel values. The effect of L_1 is pictured below for various times t.

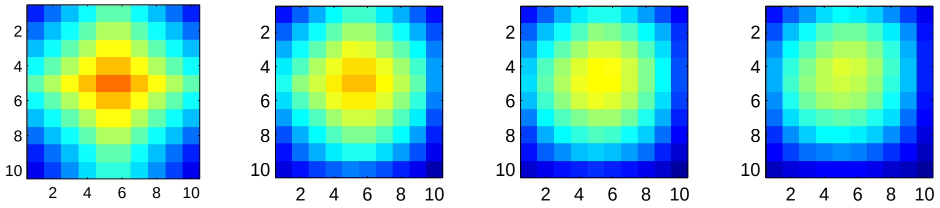
Here, the rank of E increased by 10 each time step, so we achieved rank $E = 100$ in the minimal number of steps and were able to solve for the initial condition x.



(a) x (b) L_1^2x (c) L_1^5x (d) L_1^9x

Figure 4.2: Initial condition and $L_1^t x$ for $t = 2, 5, 9$.

Our second operator, L_2, was a diffusion operator where the diffusion coefficient varied over the pixelation. We explain it in relation to the description of L_1 above. D (in item [Diffuse] above) is a function of the i, j-pixel, and for the i, j-pixel is chosen to be $\frac{4}{5}(i^3 j^2 10^{-5})^{1/4}$ (so the rate of diffusion was greatest in the lower right corner of the pixelation and was least in the upper left corner). The [Shift] part is skipped, but we must again fiddle with things at the boundary. The effect of L_2 is pictured below for a few times t.

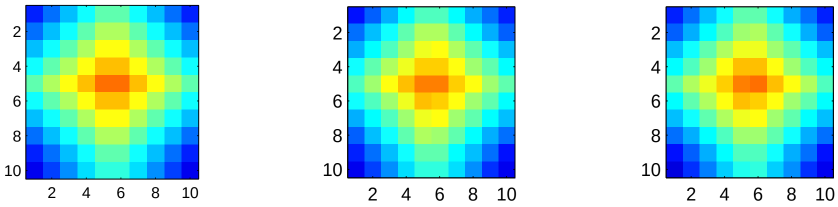


(a) x (b) L_2^2x (c) L_2^5x (d) L_2^9x

Figure 4.3: Initial condition and $L_2^t x$ for $t = 2, 5, 9$.

In this example, the rank of E again increased by 10 each step reaching 100 after 10 steps. Pictured below are the initial condition x and the reconstructions obtained by using the data sequence $(Px^*, PL_2x^*, PL_2^2x^*, \ldots, PL_2^t x^*)$ for $t = 9, 14$.

With regard to the numerical conditioning of these problems, we note that the condition numbers of the matrices E constructed using L_1 and L_2 were on the order of 10^{12} and 10^{11} respectively. By running the dynamics longer than the number of time steps required to achieve full rank, we were able to reduce the condition



(a) x^*

(b) $x, t = 9$

(c) $x, t = 14$

Figure 4.4: Initial condition and reconstructions using L_2 for $t = 9, 14$.

number in both cases. Namely, using L_1 for 15 time steps reduced the condition number of E to 10^{11}. But with L_2, using 12 time steps reduced the condition number of E to 10^9, and at 15 time steps the condition number reduced to 10^8. In both cases, extending beyond 15 steps gave no significant improvement.

Remark 4.5.1. *These relatively ugly condition numbers mean for all practical purposes, we have a null space to worry about. More specifically, the presence of even small amounts of noise will make conclusions about components of the object associated with the small singular values, meaningless.*

4.6 The Solution: Nitty-Gritty Details

The three subsections of this section contain technical details of the paper which we have described intuitively in section 4.4. Having read section 4.4, one could simply read the theorems and conjectures of this section and skip the proofs if so inclined and still understand the paper. With the warning that this section is more detailed and more demanding than the other sections of the paper, we invite the reader to dive in.

4.6.1 Transversality is (more than) enough

We now study how the dimension of N_k depends on k. What conditions imply that eventually $\dim(N_k)$ becomes 0 for some k? How prevalent are the L's for a fixed P that have $T^F \equiv \{$ minimal T such that $N_{T-1} = \{0\}$ $\} = \lceil n/d \rceil$. That is, how prevalent are the L's having the optimal reduction property?

Let us begin by reiterating the definitions of the N_k.

$$\begin{array}{ll} N_1 & \equiv N \\ N_2 & \equiv N \cap L(N_1) \\ N_3 & \equiv N \cap L(N_2) \\ \dots & \dots \\ N_T & \equiv N \cap L(N_{T-1}). \end{array} \tag{4.9}$$

If G and H are linear subspaces of J then the only situation stable to small perturbations is that the intersection of G and H has minimal dimension. That is, we expect $\dim(G \cap H) = \dim(G) + \dim(H) - \dim(J)$ where a non-positive result indicate the trivial intersection of dimension zero. If this is the case we shall say that G and H are transverse and will write this as $G \pitchfork H$.

Referring to the above definitions of the N_k's we see that these sets decrease in size at the maximum allowable rate if the intersections defining them are transverse. For example, if as above, $\dim(X) = n$ and $\dim(N) = p$, transverse intersections imply that the sequence of dimensions is p , $p+p-n$, $p+p+p-n-n$,... or if we note that $d = n - p$ then the sequence is p , $p-d$, $p-2d$, $p-3d$,... and we get that $\dim(N_{\lceil n/d \rceil}) = 0$ (Remember that we are assuming that P is full rank.)

Suppose that the intersections are not transverse. Then we still have the following lower bound on the rate at which the dimension of the N_i's decrease.

Theorem 4.6.1 (Minimal Reduction Theorem). *If there is no nontrivial subspace G of N such that $L(G) \subset G$ then for $i \leq \dim N + 1$, $\dim(N_i) \leq \dim(N) - i + 1$.*

Proof. Assuming for the moment, that $N_{i+1} \subseteq N_i$, we get that $\dim N_{i+1} \leq \dim N_i$. Then, if $\dim N_{i+1} = \dim N_i$, we conclude that $N_i = N_{i+1}$. By definition of N_{i+1}, we then have $N_i = N \cap L(N_i)$ so $N_i \subseteq L(N_i)$, hence $\dim N_i \leq \dim L(N_i)$. Since L is a linear operator, $\dim L(N_i) \leq \dim N_i$ holds as well, so $N_i = L(N_i)$. But N_i is a subspace of N, so it follows that $N_i = \{0\}$. This means that $\dim(N_{i+1}) = \dim(N_i)$ only if $N_i = \{0\}$ so that $\dim(N_{i+1}) \leq \dim(N_i) - 1$ if $\dim(N_i) \neq 0$. This implies our monotonically decreasing upper bound for the dimensions of the N_i.

To see that $N_{i+1} \subseteq N_i$: we use induction. We have $N_2 = N \cap L(N) \subseteq N = N_1$, so $N_2 \subseteq N_1$. If $N_{k+1} \subseteq N_k$, then $N_{k+2} = N \cap L(N_{k+1}) \subseteq N \cap L(N_k) = N_{k+1}$. □

Can one find an example of minimal reduction? Yes! Consider the case in which $X = \mathbb{R}^6$ and

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{4.10}$$

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.11}$$

Then we get that N is given by

$$N = \begin{bmatrix} 0 \\ 0 \\ x \\ x \\ x \\ x \end{bmatrix} \tag{4.12}$$

where the x's can be any value. This gives

$$N_1 = \begin{bmatrix} 0 \\ 0 \\ x \\ x \\ x \\ x \end{bmatrix} N_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ x \\ x \\ x \end{bmatrix} N_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ x \\ x \end{bmatrix} N_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ x \end{bmatrix} N_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.13}$$

4.6.2 Optimal Dynamics are generic

We now present and prove a theorem that addresses the point of how prevalent operators having the optimal reduction property are. In fact as the title of this section suggests, optimal L are generic. By generic we mean open and dense (as opposed to merely residual).

Definition 4.6.1 ($\mathfrak{T}$). *Identify the set of operators $\hat{L} = (L_1, L_2, ..., L_{T-1})$ with $\mathbb{R}^{(T-1)n^2}$. Now define $\mathfrak{T} \subset \mathbb{R}^{(T-1)n^2}$ to be all those $\hat{L} \in \mathbb{R}^{(T-1)n^2}$ such that*

$$L_i(N_i) \pitchfork N \quad \forall i. \tag{4.14}$$

Theorem 4.6.2 (Extended Linear Transverse Intersection Theorem). *$\mathfrak{T}$ is open and dense in $R^{(T-1)n^2}$.*

In the following proof, we will use d_A to denote $\dim(A)$. Since the set of all invertible $\hat{L}$ in $\mathfrak{T}$ is open and of full measure in $\mathbb{R}^{(T-1)n^2}$ we shall assume that $\hat{L}$ is invertible throughout the proof. We shall also use the fact that for invertible L, $L(A \cap L(B)) = L(A) \cap L^2(B)$.

Proof. We first observe that $M_1 \pitchfork M_2 \Leftrightarrow d_{P_{M_1^\perp}(M_2)} = \min(d_{M_1^\perp}, d_{M_2})$ or equivalently that $\text{rank}(P_{M_1^\perp} \circ P_{M_2}) = \min(d_{M_1^\perp}, d_{M_2})$. A little bit of thought is enough to convince oneself that $\text{rank}(P_{M_1{}^\perp} \circ P_{L(M_2)}) = \text{rank}(P_{M_1{}^\perp} \circ L \circ P_{M_2})$. Therefore we get that $M_1 \pitchfork L(M_2) \Leftrightarrow \text{rank}(P_{M_1{}^\perp} \circ L \circ P_{M_2}) = \min(d_{M_1{}^\perp}, d_{M_2})$.(Here we have used the invertibility of L to conclude that $d_{M_2} = d_{L(M_2)}$). Let us approach the problem a little more generally. We shall use the fact that $K \pitchfork L(M) \Leftrightarrow \text{rank}(P_{K^\perp} \circ L \circ P_M) = \min(d_{K^\perp}, d_M)$ to show that the set $\mathfrak{T}_*$ of all L in $\mathbb{R}^{n^2}$ such that $K \pitchfork L(M)$ is open and of full measure. Here K and M are linear subspaces of $\mathbb{R}^n$.

Define column vectors $p_{\cdot,i}$ for $i = 1, ..., d_{K^\perp}$ that are orthogonal to each other and span $K^\perp$. Likewise let $q_{\cdot,i}$ for $i = 1, ..., d_M$ be column vectors orthogonally spanning M. Define the n by n matrices P and Q as follows:

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,d_{K^\perp}} & 0 & \cdots & 0 \\ p_{2,1} & p_{2,2} & \cdots & p_{2,d_{K^\perp}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,d_{K^\perp}} & 0 & \cdots & 0 \end{pmatrix} \tag{4.15}$$

and

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,d_M} & 0 & \cdots & 0 \\ q_{2,1} & q_{2,2} & \cdots & q_{2,d_M} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ q_{n,1} & q_{n,2} & \cdots & q_{n,d_M} & 0 & \cdots & 0 \end{pmatrix}. \tag{4.16}$$

Then

$$P_{K^\perp} = P \circ P^T \tag{4.17}$$

and

$$P_M = Q \circ Q^T, \tag{4.18}$$

so that we get

$$P_{K^\perp} \circ L \circ P_M = P \circ P^T \circ L \circ Q \circ Q^T. \tag{4.19}$$

Now we note that

$$\operatorname{rank}(P \circ P^T \circ L \circ Q \circ Q^T) = \operatorname{rank}(P^T \circ L \circ Q). \tag{4.20}$$

To show that $\mathfrak{T}_*$ is open in $\mathbb{R}^{n^2}$, observe that

$$P^T \circ L \circ Q = \begin{array}{cc} & \begin{array}{cc} d_M & n-d_M \end{array} \\ \begin{array}{c} d_{K^\perp} \\ n-d_{K^\perp} \end{array} & \left(\begin{array}{cc} U_L & 0 \\ 0 & 0 \end{array}\right) \end{array} \tag{4.21}$$

and therefore

$$\operatorname{rank}(P^T \circ L \circ Q) = \min(d_{K^\perp}, d_M) \Leftrightarrow U_L \text{ is full rank.} \tag{4.22}$$

Note that U_L is a continuous function of U, i.e. $U_L : \mathbb{R}^{n^2} \to \mathbb{R}^{d_M \cdot d_{K^\perp}}$ is continuous (actually smooth). Assume without the loss of generality, that $d_{K^\perp} \geq d_M$. Let $\phi^{d_{K^\perp}}_{d_M}(U_L)$ be the d_M dimensional measure of regions in $\mathbb{R}^{K^\perp}$ applied to the parallelepiped with edges equal to the columns of U_L. Then $\mathfrak{T}_*$ is precisely equal to $(U_L)^{-1}((\phi^{d_{K^\perp}}_{d_M})^{-1}(\mathbb{R} \setminus \{0\}))$. Since both U_L and $\phi^{d_{K^\perp}}_{d_M}$ are continuous, we have that $\mathfrak{T}_*$ is open.

To show that $\mathbb{R}^{n^2} \setminus \mathfrak{T}_*$ has zero n^2-dimensional Lebesgue measure, we first introduce a change of coordinates. Define $\hat{P}$ to be an orthogonal matrix obtained by filling in the zero columns of P appropriately. Obtain $\hat{Q}$ from Q analogously. Then

$$\begin{aligned}
P^T \circ L \circ Q &= P^T \circ \hat{P} \circ \hat{P}^T \circ L \circ \hat{Q} \circ \hat{Q}^T \circ Q \tag{4.23}\\
&= \begin{pmatrix} I_{d_{K^\perp}} & 0 \\ 0 & 0 \end{pmatrix} \circ \hat{L}_L \circ \begin{pmatrix} I_{d_M} & 0 \\ 0 & 0 \end{pmatrix} \tag{4.24}\\
&= \hat{L}_L(\mathrm{ul}) \tag{4.25}\\
&= \text{upper left } (d_{K^\perp} \times d_M) \text{ block of } \hat{L}_L \tag{4.26}
\end{aligned}$$

where I_ζ is the identity matrix of dimension ζ and we have set $\hat{L}_L = \hat{P}^T \circ L \circ \hat{Q}$. So

$$\begin{aligned} \mathfrak{T}_* =& \{L |\ \text{rank}(\hat{L}_L(\text{ul})) = \min(d_{K^\perp}, d_M)\ \} \\ =& \{L |\ \hat{L}_L(\text{ul}) \text{ is full rank.}\} \end{aligned} \tag{4.27}$$

Since $\hat{P}$ and $\hat{Q}$ are orthogonal, we have that the n^2 dimensional Lebesgue measure of

$$\mathfrak{T}_* = \{L |\ \hat{L}_L(\text{ul}) \text{ is full rank}\} \tag{4.28}$$

and

$$\hat{\mathfrak{T}}_* \equiv \{\hat{L} |\ \hat{L}(\text{ul}) \text{ is full rank}\} \tag{4.29}$$

are equal.

We now prove that $\mathbb{R}^{n^2} \setminus \hat{\mathfrak{T}}_*$ has measure zero. Any $\hat{L}$ can be written as the block matrix with dimension of $\hat{L}_{ul}$ being $d_{K^\perp} \times d_M$.

$$\begin{pmatrix} \hat{L}_{ul} & \hat{L}_{ur} \\ \hat{L}_{ll} & \hat{L}_{lr} \end{pmatrix} \tag{4.30}$$

We can write $\hat{L}$ out in terms of elements as

$$\hat{L} = \begin{pmatrix} \hat{l}_{11} & \hat{l}_{12} & \dots & \hat{l}_{1n} \\ \hat{l}_{21} & \hat{l}_{22} & \dots & \hat{l}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{l}_{n1} & \hat{l}_{n2} & \dots & \hat{l}_{nn} \end{pmatrix}. \tag{4.31}$$

Now assume that $d_{K^\perp} \geq d_M$. Identify $\mathbb{R}^{n^2 - d_{K^\perp} \cdot d_M + (d_M - 1)\cdot(d_{K^\perp}+1)}$ with the elements of f_{ij} of an $n \times n$ matrix with the elements f_{d_M,d_M} , f_{d_M+1,d_M} , $\cdots$, $f_{d_{K^\perp},d_M}$ removed.

Next, define a mapping of $\phi : \mathbb{R}^{n^2 - d_{K^\perp} \cdot d_M + (d_M - 1)\cdot(d_{K^\perp}+1)} \to \mathbb{R}^{n^2}$ by

$$\hat{l}_{ij} = \begin{cases} f_{ij} & \text{for } (i,j) \ \notin \{i \leq d_{K^\perp} \text{ and } j = d_M\} \\ \sum f_{kd_M} f_{ik} & \text{for } (i,j) \ \in \{i \leq d_{K^\perp} \text{ and } j = d_M\} \end{cases} \tag{4.32}$$

which has, as its image precisely those matrices in which $\hat{L}_{ul}$ has column d_M that is the linear combination of the first $d_M - 1$ columns. If we redefine our mapping to get a series of completely analogous mappings, each of which has a column of $\hat{L}_{ul}$ being dependent on the other columns of $\hat{L}_{ul}$, then we end up with d_M such maps. Since each of these maps are smooth, and singular (the rank of the derivative is not equal to the dimension of the image space) Sard's theorem [65] tells us that the n^2-dimensional measure of the image of each map is zero. Therefore the union of the images also has measure zero. But this union is exactly $\mathbb{R}^{n^2} \setminus \hat{\mathfrak{T}}_*$. Since the case of $d_{K^\perp} < d_M$ is completely analogous, we have now shown that $\mathfrak{T}_*$ is open and of full measure in $\mathbb{R}^{n^2}$.

To complete the proof we let the operator change at each step so that $x_2 = L_1(x_1)$, $x_3 = L_2(x_2)$, and so on. Now we have an extended operator $\tilde{L} = (L_1, L_2, ..., L_{T-1}) \in (\mathbb{R}^{n^2})^{T-1}$. We will show that the set $\mathfrak{S} \equiv \{\tilde{L} | \; N \pitchfork L_i(N_i)$ for $i = 1, 2, ..., T-1\}$ is open and dense in $(\mathbb{R}^{n^2})^{T-1}$.

Define $\mathbb{C}_1$ to be the open subset of full measure in $\mathbb{R}^{n^2}$ whose members, L_1, satisfy $N \pitchfork L_1(N_1)$. Choose a countable subset, $\mathbb{D}_1$, which is dense in $\mathbb{C}_1$. Now for each element D_1^k of $\mathbb{D}_1$, define $\mathbb{C}_2^k$ to be the open full measure subset of $\mathbb{R}^{n^2}$ such that $L \in \mathbb{C}_2^k$ implies that $N \pitchfork L(N \cap D_1^k(N_1))$, or equivalently, $N \pitchfork (L(N) \cap L \cdot D_1^k(N_1))$. Define $\mathbb{C}_2 \equiv \bigcap_k \mathbb{C}_2^k$. Since $\mathbb{C}_2$ is dense in $\mathbb{R}^{n^2}$ we can pick a countable $\mathbb{D}_2 \subset \mathbb{C}_2$ that is also dense in $\mathbb{R}^{n^2}$. We have that $L_1 \in \mathbb{D}_1$ and $L_2 \in \mathbb{D}_2$ implies that $N \pitchfork L_1(N_1)$ and $N \pitchfork L_2(N \cap L_1(N_1))$. Continuing this process we obtain $\mathbb{D}_i$ for $i = 1, 2, ..., T-1$ such that $(L_1, L_2, ..., L_{T-1}) \in \mathbb{D}_1 \times \mathbb{D}_2 \times ... \times \mathbb{D}_{T-1}$ implies that all the intersections are transverse, i.e. that $N \pitchfork L_i(N_i)$ for $i = 1, 2, ..., T-1$. We have therefore found a subset of $\mathfrak{S}$ which is dense in $(\mathbb{R}^{n^2})^{T-1}$.

Now we show that $\mathfrak{S}$ is open. (In what follows, we assume that $T(n - d_N) \leq n$. One should replace equations 4.37 to 4.39 by their equivalents with the subscripts and superscripts switched for each equation for which $j(n - d_N) > n$. One then considers the n dimensional volume of the column space in $j(n - d_N)$ space.) The requirement that each of the intersections are transverse is equivalent to the requirement that

$$\begin{aligned}
\dim(N \cap L_1(N)) &= d_N + d_N - n \\
\dim(N \cap L_2(N) \cap L_2 L_1(N)) &= 3d_N - 2n \\
\dim(N \cap L_3(N) \cap L_3 L_2(N) \cap L_3 L_2 L_1(N)) &= 4d_N - 3n \\
\vdots \quad &= \quad \vdots \\
\dim(N \cap L_{T-1}(N) \cap L_{T-1} L_{T-2}(N) \cap ... \cap L_{T-1} ... L_1(N)) &= T d_N - (T-1)n.
\end{aligned} \tag{4.33}$$

which in turn is equivalent to a requirement involving orthogonal complements, specifically that

$$\begin{array}{cc} & n \\ 2(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L_1^{-1} \end{bmatrix} \end{array} \tag{4.34}$$

$$\begin{array}{cc} & n \\ 3(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L_2^{-1} \\ rN^{\perp} \circ L_1^{-1} \circ L_2^{-1} \end{bmatrix} \end{array} \tag{4.35}$$

$$\vdots$$

$$\begin{array}{cc} & n \\ T(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L_{T-1}^{-1} \\ \vdots \\ rN^{\perp} \circ L_1^{-1} \circ ... \circ L_{T-2}^{-1} \circ L_{T-1}^{-1} \end{bmatrix} \end{array} \tag{4.36}$$

all have full rank where $rN^{\perp}$ is the matrix with rows equal to independent n-dimensional vectors spanning the linear subspace $N^{\perp}$. This last set of expressions follows from the fact that if M and K are linear subspaces of $\mathbb{R}^n$ then $(M \cap K)^{\perp} = \text{span}(M^{\perp}, K^{\perp})$ so that the matrices immediately above have rank $2(n-d_N), 3(n-d_N), ..., T(n-d_N)$ iff the previous intersections have dimensions $2d_N - n, 3d_N - 2n, ..., Td_N - (T-1)n$ respectively. But this last expression can be seen to be exactly those matrices which satisfy the equations

$$\phi^n_{2(n-d_N)}(\text{rows from first matrix}) \neq 0 \tag{4.37}$$

$$\phi^n_{3(n-d_N)}(\text{rows from second matrix}) \neq 0 \tag{4.38}$$

$$\vdots$$

$$\phi^n_{T(n-d_N)}(\text{rows from T - 1st matrix}) \neq 0 \tag{4.39}$$

where $\phi^l_j(vectors)$ measures the j-dimensional volume of the the parallelepiped spanned by the *vectors* in $\mathbb{R}^l$. But since the inverse operation is continuous on the set of invertible matrices and these volume functions are smooth, we have that the set of $(L_1, L_2, ..., L_{T-1})$ having the full rank property is open. Thus, $\mathfrak{S}$ is open in $(\mathbb{R}^{n^2})^{T-1}$. □

In the next section we conjecture an approach to the nonlinear case which uses the above derivation, but before we do this we show that the linear case can actually be made significantly simpler.

Theorem 4.6.3 (Linear Transverse Intersection Theorem). *If the set of operators L is identified with $\mathbb{R}^{n^2}$ and we define $\mathfrak{T} \subset \mathbb{R}^{n^2}$ to be all those $L \in \mathbb{R}^{n^2}$ such that*

$$L(N_i) \pitchfork N \quad \forall i. \tag{4.40}$$

Then $\mathfrak{T}$ is open and dense in R^{n^2}.

Proof. As was seen in the proof of the previous theorem, the transversality requirement and the fact that we are considering the case of $L_i = L$ for all i, reduces to

$$\begin{array}{cc} & n \\ 2(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L^{-1} \end{bmatrix} \end{array} \tag{4.41}$$

$$\begin{array}{cc} & n \\ 3(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L^{-1} \\ rN^{\perp} \circ L^{-2} \end{bmatrix} \end{array} \tag{4.42}$$

$$\vdots$$

$$\begin{array}{cc} & n \\ T(n-d_N) & \begin{bmatrix} rN^{\perp} \\ rN^{\perp} \circ L^{-1} \\ \vdots \\ rN^{\perp} \circ L^{-(T-1)} \end{bmatrix} \end{array} \tag{4.43}$$

all having full rank. But this is equivalent to another full rank condition as follows. Let $cN^{\perp}$ be the transpose of $rN^{\perp}$. In other words, while $rN^{\perp}$ are the orthogonal row vectors that span compliment of the null space, $cN^{\perp}$ are the column vectors that span the "same" space. This condition is equivalent to the following. For a dense and open set of L:

$$\dim(S_k \equiv \operatorname{span}(cN^{\perp}, L \circ cN^{\perp}, ..., L^{k-1} \circ cN^{\perp})) = \min(k \cdot d_{cN^{\perp}}, n) \tag{4.44}$$

That is, for an open and dense set of L the sequence of subspaces S_k generated by the iterates $L^j \circ cN^{\perp}$ $j = 1, 2, ..., k-1$, are of maximal dimension.

This fact is well known. For lack of a reference I give a proof here. Without loss of generality let $cN^{\perp}$ be the k left most columns of the $n \times n$ identity matrix. Then the dimension of S_k is the rank of the matrix formed by taking the k left columns of I, followed by the k left columns of L, followed by the k left columns of L^2, and so on.

Pick the upper left matrix minor and compute it's determinate this will give a polynomial in $l_{11}, l_{12}, l_{21}, ...l_{nn}$ which we want to show is nonzero except on an open and dense set. As long as the polynomial is nonzero at one point then we are done since this implies that the set of zeros occupies a submanifold of R^{n^2} that is at most, $n^2 - 1$ dimensional. (So we even have more! The set of "good" L has full measure and is open.)

To show that the determinant of the upper left matrix minor is nonzero at a point, we consider $L =$ permutation matrix that shifts everything to the left k clicks. This gives us the identity matrix in the upper left matrix minor and so the determinant in question evaluates to 1. □

Remark 4.6.1. *As the above proof shows, $\mathfrak{T}$ is in fact open and full measure. This improves the result from one stated which implies stable approximation by $\hat{L}$ having the optimal reduction property, to one that implies this AND the improbability of non-optimal reduction. Further improvements would involve the characterization of $\mathfrak{T}_\epsilon \equiv \mathfrak{T} \cap \{L \in \mathbb{R}^{n^2} | cond(L) \quad < \quad 1/\epsilon\}$.*

Remark 4.6.2. *The above proof also works with slight modifications to prove theorem 4.6.2, but the proof we give leads to a conjectured proof for the case of nonlinear dynamics, and so seems more useful even if it is more cumbersome.*

4.6.3 Extension to the Nonlinear Case

The above theorem is extendible to the nonlinear case as follows. Actually, we are conjecturing such an extension in what follows. It should be noted that the nonlinear "extension" does not imply the linear theorem.

In this section the state space (object space) will be M, a compact manifold of dimension m. The projection operator will be a smooth function $P : M \to \mathbb{R}^d$ and the dynamics will be given by F_i's which map M to itself diffeomorphically. Instead of using linearity to get a fixed null space, we find the "null" space we are now interested in is a level set of P. These sets can change in nontrivial ways as the point in the range of P changes. We now want to know about intersections of these "null" sets with images of other of the "null" sets under F.

The transversality theorem found on page 74 of [40] implies that the set $\mathfrak{T}$ of $f \in C^r(M)$ which map a submanifold K of M back into M to intersect another submanifold N transversely, is open and dense. (For compact M the topology is nice, see [40] for details.) What we need is a bit more complicated.

Let $\mathfrak{F}$ be the quasi-stratification of M into the level-sets $\mathfrak{F}_x$, $x \in R^d$ of P. Let N be the union of a finite number of (not-necessarily injectively) immersed submanifolds of dimension $\leq n$ whose self-intersections are transverse in the sense that the tangent spaces of the "participants" in the intersection span the largest possible subspace of $T_{i_x}M$ where i_x is a point of self-intersection.

Conjecture 4.6.1. *For an open and dense set of $F \in D^\infty(M)$*

$$dim(I_x) \leq max(n-d, 0) \;\; \forall x \in \mathbb{R}^d \tag{4.45}$$

where $I_x \equiv (F(N) \cap \mathfrak{F}_x)$ and,

$$I_x \text{ is a finite union of stably immersed submanifolds} \tag{4.46}$$

.

This permits us to conclude that, for any initial point $x_0 \in X$ and dense $\tilde{F} = (F_1, ..., F_{T-1})$, the intersection obtained by $T = \lceil m/d \rceil$ measurements will have dwindled to a finite set of points, call it S. The next measurement ($T = \lceil m/d \rceil + 1$) will generically have precisely one point in the intersection of the set S and the level set corresponding to the next measurement. (We use the same argument as we used in the linear case to get a product of dense sets $\mathfrak{D}_1 \times ... \times \mathfrak{D}_{T-1}$.)

Now, if indeed the $\lceil m/d \rceil + 1$"th" intersection is a single point, then the mapping $G : x \in M \to (Px, PF_1x, PF_2F_1x, ..., PF_{T-1}...F_1x)$ has only one point in the inverse image of $G(x_0)$. Since the point we have "found" (the conjecture above) comes from stable intersections this should guarantee that the point $G(x_0)$ in fact has a neighborhood in which G is invertible. This should in turn guarantee that there is an open neighborhood B_ϵ of G in $C^\infty(M)$ such that $H \in B_\epsilon$ implies $H^{\leftarrow}(H(x_0))$ is a single point. We then use the fact that a small neighborhood of $\tilde{F}$ maps into this small neighborhood under the mapping $\tilde{J} \to (P, PJ_1, ..., PJ_{T-1}...J_1)$ for $J \in D^\infty(M)$.

We have arrived at our second conjecture.

Conjecture 4.6.2. *If M is a compact smooth manifold of dimension m, P is a smooth function mapping M to $\mathbb{R}^d$, x_0 is a particular point in M, $T = \lceil m/d \rceil + 1$, $\mathfrak{D} \equiv (\mathbb{D})^{T-1}$ is the $T-1$-fold product of the space of smooth diffeomorphisms from M to M $(D^\infty(M))$ and we define $H_\delta \equiv (P, P\delta_1, ..., P\delta_{T-1}...\delta_1) : M \to \mathbb{R}^{dT}$, where $\delta = (\delta_1, ..., \delta_{T-1}) \in \mathfrak{D}$, then the set $\mathfrak{O}$ of $\delta \in \mathfrak{D}$ such that*

$$H_\delta^{\leftarrow}(H_\delta(x_0)) = \{x_0\} \tag{4.47}$$

is open and dense in $\mathfrak{D}$.

4.7 Relation to Known Results

The results obtained above are known as observability results in control theory and phase space reconstructions (delay coordinate embeddings) in dynamical systems. Our results are different in that we consider variations of the dynamics with the observation function kept fixed whereas other results either assume that the dynamics are fixed and the observation function changes or that both the dynamics and the observation function is variable, see [96, 81, 1, 5, 92]. While Aeyels [1] does consider the case where the observation function is fixed and the dynamics are variable he does so for vector fields (not maps). He is also looking at the case where he wants all initial points to be recoverable from the sequence of measurements and this requires 2n+1 , 1-dim measurements to recover the n-dim initial points. Similar comments apply to the comparison to Stark's more recent paper [92]. Our minimal reduction theorem is a more precise version of the well known theorem in control theory that states that if the observability matrix is not full rank then no number of measurements can give you full information on the state of the system and if it is full rank then you need at most n measurements (of any dimension) of a system that has an n-dimensional state space. In the preparation of this paper we became aware of theorem 5.3.13 in [68] which, together with the usual observability theorem, is equivalent to our minimal reduction theorem.

4.8 Summary and Discussion

In this paper we examined the use of dynamics in the inversion of projection data obtained at a sequence of times. The main results confirm that for any fixed measurement projection and generic dynamics, we can simply combine the number of measurements into one large super-measurement which we invert to obtain the state we are trying to reconstruct. A following paper will deal with some aspects of the stochastic or noisy case of reconstruction from projections using dynamics.

What we have established is only a first step in the direction leading to the fruitful combination of dynamics and measured data. Many variants of the proposed underlying tomography problem lead to the abstract problems identical or similar to the one we have begun to examine. For example, one might know the dynamics up to some set of parameters which we try to obtain – together with the state vector – from the measurements. It seems to us that there are at least two important directions to go next. One is the examination of the present formulation in the presence of noise. This will bring us much closer to “real” situations in that the ubiquity of noise makes certain problems which are well posed in the noiseless case, ill-posed in the presence of noise. The second direction is the attack of a very carefully chosen concrete problem involving dynamics that we understand analytically or at least numerically. This will invariably involve certain toy-like characteristics which should nevertheless be useful for the approach to the large, more realistic problems.

Natural questions that arise include:

1) How is the problem of reconstruction from a sequence of projections related to the reconstruction of a 3-dim object from a spatial sequence of slices? This arises when one wants to interpolate a set of CAT scans to generate a 3-dim density image. A related problem arises when one is trying to compress a video movie by using some clever interpolation in the uncompress process.

2) Can we construct an efficient algorithm for reconstruction in the case of non-linear projections and/or dynamics? Actually, even the “noiseless” linear case, while conceptually simple, is not trivial in the high dimensional case, where by “noiseless” we imply the lack of measurement noise (but *include* the presence of “noise” induced by computation and approximation). Extraordinarily large condition numbers are pervasive and so any error , like roundoff, soon overwhelms you. Future papers will begin to address issues such as these which are involved in the dynamically constrained reconstruction of objects from *noisy projections.*

3) If one has a set of measurements, how does one use

A) knowledge of the underlying dynamics (possibly incomplete) and

B) freedom to choose the object (state space) parameterization

in such a way as to get a well posed inverse problem that wastes as little of the prior and measured information as possible? That is, how does one use all the prior and measured information in the generation of the final reconstruction? Even if we are using all the information, there are different ways of distributing remaining uncertainty about the reconstructed object. What freedom do we have to move these inevitable uncertainties around to different parts of the object?

4) In the high dimensional case, the above questions are incredibly difficult to answer. Can we find approximate answers? Can we determine how good these approximate answers are? For example, can we obtain bounds on the amount of information that our parameterization/reconstruction/use-of--priors wastes? It seems likely that ideas from the fields of function approximation/data modeling and machine learning will be useful here. Tradeoff's between model bias versus model variance is an example of one issue that is pertinent here and is also a topic in machine learning.

5) Suppose we do the whole analysis with ϵ fattened null spaces. (I.E. instead of looking at the dimension of the intersections of N and the $L^i(N)$'s, we consider the volume and shape of the corresponding intersections defined using the ϵ-neighborhood of N.) In this case, what sort of shape and volume do we get for the final intersection (which before, was a single point)? This is along the same lines as remark 4.6.1 and closely related to the idea behind Kalman filtering. (In fact with the correct viewpoint, this "is" Kalman filtering, even though Kalman filtering usually assumes Gaussian statistics and not the uniform distributions suggested here.)

Chapter 5

Models of Dynamics and the Entropy Gap

Kevin R. Vixie [1]

5.1 Entropy and Models of Data Streams: An Introduction

Jorma Rissanen [71] said that the three main tasks of signal processing are, prediction, compression, and estimation and that the greatest of these is estimation. For a process Y that produces strings $\{y\}_1^T \equiv (y_1, y_2, \ldots, y_T)$, suppose that one obtains by some estimation procedure a *model*, θ. By model[2] we mean a family of functions that map strings, $\{y\}_1^T$, to probabilities, $\theta\left(\{y\}_1^T\right)$

With such a model, optimal performance at the other tasks is straight forward in principle. For a given estimation procedure that adjusts the parameters θ of a model to fit measured data, we sometimes want to know how well the procedure works on data generated by a process not in the model class at all. In this paper we take up the task of measuring model fidelity.

[1] The research for this chapter was done in collaboration with Andrew M. Fraser.

[2] We will be a bit careless with our use of θ. In addition to referring to the entire model generated from measurements, it will sometimes refer to the model probability function, sometimes to the approximate dynamics that the model implicitly defines, and sometimes to the state space also defined implicitly by the model. The specific intentions in any particular occurrence will be clear from the context.

We propose a measure of model fidelity based on the notion of *relative entropy*. The task of modeling some (unknown) system given measurements from that system is in general a difficult one. Since it is much easier in practice to accurately determine the relative entropy of a proposed model to the true process than it is to actually fit a good model (at least in the case of chaotic dynamical systems), we have chosen a less daunting task, but one that we feel has been neglected.

In the remainder of this section, we introduce our notation, lay out the assumed context and the propose the measure of model fidelity which we will call the *gap*, G. Sections 5.2 and 5.3 look at the relationship between Lyapunov exponents and entropy – a relationship which we exploit in the definition of G. Section 5.4 is a closer look at the notion of Sinai-Ruelle-Bowen (SRB) measures, a concept on which the previous two sections depend in a crucial way. We finally return to our proposed measure in section 5.5 where we carefully study the gap's exact meaning, calculation, and approximation. (Approximation is of interest due to the constraints of finite precision calculations). We close with a section of conclusions and comments on future work.

5.1.1 Notation.

In order to maintain coherence throughout the paper, we will use the following notation.

X will be the state space, x or x_i are individual elements of X and $\{x\}_1^n \equiv (x_1, ..., x_n)$ is a sequence of states: usually $X = R^k$ but sometimes X will be a compact k-dimensional manifold.

$F : X \to X$ will be the mapping or function which governs the dynamics in the state space.

μ will be a probability measure that is invariant under F.

$M : X \to Y$ will be a mapping from the state space to the measurement space. We call $y = M(x)$ the measurement of x by M.

Y**,** y **or** y_i **and** $\{y\}_1^n \equiv (y_1, ..., y_n)$ – will be the space of measurements, individual measurements and a sequence of measurements, respectively. Y will typically be $\mathbb{R}^m$, $m < k$.

$\mathcal{B}$ and β or β_i – will be a partition of the state space X and individual elements of $\mathcal{B}$, respectively.

$\tilde{}$ will denote the finite precision (floating point) version of a quantity, function or set. For example, the finite precision version of F will be $\tilde{F}$.

θ [3] will be a model that approximates aspects of F, μ, and M.

In addition to the usual definition, conjecture, theorem, etc., we will also use **Statement**'s when we want to give distillations of concepts, intuitive statements of meaning or the practical uses of a concept.

Other notation will be established as needed throughout the rest of the paper.

5.1.2 Overview

In many cases our understanding of the world is derived from a series of measurements. The time series that results from such a series of measurements may be either scalar or vector. Typically, the dimension of the measurements is less than the dimension of the underlying state space.

Formalizing this, our measurements $\{y\}_1^n = (y_1, y_2, ..., y_n)$, $y_i \in \mathbb{R}^m$, for $i = 1, ..., n$ are generated from a series of states (of the real world) $x_1, x_2, ..., x_n \in \mathbb{R}^K$ by the action of a measurement operator $M : \mathbb{R}^k \to \mathbb{R}^m$, $k \geq m$ so that $(y_1, y_2, ..., y_n) = (M(x_1), M(x_2), ..., M(x_n))$. The series of states $\{x\}_1^n$ is dictated by a function $F : x_{i+1} = F(x_i)$. Data driven modeling is the attempt to get back to F and X from the measurements or data $\{y\}_1^n$. The idea is schematically illustrated in Fig. 5.1. The model, which we will call θ, can then be used to reach a number of seemingly different goals such as prediction of the future based on the past, compression for optimal transmission, interpolation of sparsely sampled data and qualitative understanding of the system that generated the data. The view we take in this paper makes all these goals look very similar. Aside from issues relating to the finiteness of resources (which may in the end make a huge difference!), each of the goals above is met when F can somehow be obtained. In this paper we concentrate on measuring model fidelity when the data used to build

[3] As mentioned in the footnote to the first paragraph of this chapter, we will be a bit sloppy with notation when it comes to probabilities. More specifically, we shall denote the model and the probabilities dictated by the model by θ.

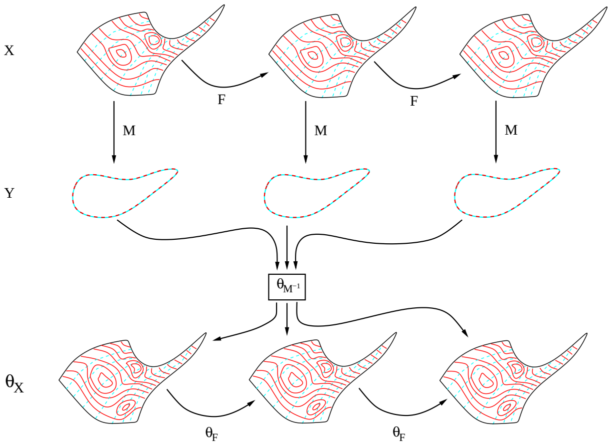


Figure 5.1: The big picture: A state space, Dynamics, Measurements, and a Model. θ_F is the model of F, $\theta_{M^{-1}}$ is the model's attempted reconstruction of X to θ_X.

the model comes from a known system. Such is the case when one is concentrating on the modeling process and is therefore attempting to eliminate as many of the unknowns as possible from the process of model validation.

So how does one measure the fidelity of a model? Is our model successful in capturing the data and does it have the correct implicit assumptions? Is the model a good approximation to the underlying system? These questions are the ones addressed in this paper.

In order to get at these questions we make some assumptions. We assume that we have some procedure for building a model based on data and that we wish to know how well the procedure is doing. In order to measure this performance, we will assume a known system F is available which can be used to generate test data and calculate quantities such as the Lyapunov exponents directly (via the Benettin procedure [8, 9] for example). In this way, model performance can be dissected and studied in great detail.

Under these assumptions, we will define a relative entropy gap and propose using it to measure model fidelity.

5.1.3 The Gap

Definition 5.1.1 (Entropy of a partition). *Given a probability measure μ, the entropy of a partition $\mathcal{B}$ is*

$$H(\mathcal{B}) = -\sum_{b\in\mathcal{B}} \mu(b)\log(\mu(b)) \tag{5.1}$$

Definition 5.1.2 (Partition refinement). *Given two partitions $\mathcal{A}$ and $\mathcal{B}$, the refinement $\mathcal{A}\vee\mathcal{B}$ is defined to be the coarsest partition β such that every element of β is contained wholly in some component of $\mathcal{A}$ and some component of $\mathcal{B}$. This can easily be seen to be the partition generated by the intersections of elements of $\mathcal{A}$ and $\mathcal{B}$.*

Definition 5.1.3 (Joint entropy). *Similarly, given a probability measure μ, and two partitions $\mathcal{A}$ and $\mathcal{B}$, the joint entropy is*

$$H(\mathcal{A},\mathcal{B}) = -\sum_{c\in\mathcal{A}\vee\mathcal{B}} \mu(c)\log(\mu(c)) \tag{5.2}$$

By applying the *set* map F^{-1} one may obtain images of a partition. We use the notation

$$\{\mathcal{B}\}_n^F = \mathcal{B}\vee F^{-1}\mathcal{B}\vee F^{-2}\mathcal{B}\vee\ldots\vee F^{-n+1}\mathcal{B} \tag{5.3}$$

to indicate partition generated by all the intersections of elements of the partitions $\mathcal{B}, F^{-1}\mathcal{B}, ..., F^{-n+1}\mathcal{B}$.

Definition 5.1.4 (Entropy rate). *We define the entropy rate of a map F, a probability measure μ that is invariant under F, and a partition $\mathcal{B}$ as*

$$h(\mathcal{B},F,\mu) = \lim_{n\to\infty}\frac{1}{n}H\left(\{\mathcal{B}\}_n^F\right) \tag{5.4}$$

Definition 5.1.5 (Kolmogorov entropy). *The Kolmogorov entropy is defined as the supremum of entropy rates over all partitions*

$$h_K(F,\mu) = \sup_{\mathcal{B}} h(\mathcal{B},F,\mu) \tag{5.5}$$

The Kolmogorov entropy is also sometimes called the KS (Kolmogorov Sinai) entropy, the metric invariant entropy or the metric entropy. Given two probability measures, μ and θ, we now define *cross entropies.* The idea is that we believe μ is *true* and θ comes from an approximate model.

Definition 5.1.6 (Cross entropy). *Given probability measures μ and θ, the cross entropy of a partition $\mathcal{B}$ is*

$$H(\mathcal{B}, \mu : \theta) = -\sum_{b \in \mathcal{B}} \mu(b) \log(\theta(b)). \tag{5.6}$$

Definition 5.1.7 (Cross entropy rate). *Given probability measures μ and θ, the cross entropy rate of a partition $\mathcal{B}$ is*

$$h(\mathcal{B}, F, \mu : \theta) = \lim_{n \to \infty} \frac{1}{n} H\left(\{\mathcal{B}\}_n^F, \mu : \theta \right). \tag{5.7}$$

Definition 5.1.8 (Relative entropy rate). *Given probability measures μ and θ, the relative entropy rate of a partition $\mathcal{B}$ is*

$$d_{\mathcal{B},F}(\mu||\theta) = h(\mathcal{B}, F, \mu : \theta) - h(\mathcal{B}, F, \mu). \tag{5.8}$$

We now come to the definition of the *gap*, G that we are proposing as a measure of model fidelity.

Definition 5.1.9. *The* gap *(relative entropy gap), $G(\mathcal{B}, F, \mu, \theta)$ is given by*

$$G(\mathcal{B}, F, \mu, \theta) \equiv h(\mathcal{B}, F, \mu : \theta) - h_L(F, \mu) \tag{5.9}$$

where h_L is the sum of the positive Lyapunov exponents of F. When there is no danger of confusion, we will abbreviate $G(\mathcal{B}, F, \mu, \theta)$ to G.

Since $h_L \geq h_K$, with equality when the invariant measure μ is smooth in the unstable directions (see section 5.2 below), we have that

$$h(\mathcal{B}, F, \mu) \leq h_K(F, \mu) \leq h_L(F, \mu)$$

$$\Rightarrow G(\mathcal{B}, F, \mu, \theta) \leq h(\mathcal{B}, F, \mu : \theta) - h(\mathcal{B}, F, \mu) = d(\mu||\theta)$$

$\Rightarrow$ a large positive gap implies a large relative entropy rate which in turn means there is room for improvement of the model.

The idea of a relative entropy distance can be understood as follows.

Statement 5.1.1 (Intuitive understanding of relative entropy). *There at least a couple of ways to look at relative entropy.*

- *Relative entropy can be seen as a measure of "distance" that makes sense because of Jensen's inequality which in turn is equivalent to the fact that convex sets always lie to one side of supporting hyperplanes.*
- *Relative entropy is also the exponential penalty that one pays for using the wrong probability distribution when building a system (code) for the purposes of data compression.*

A very nice reference for both of these is Cover and Thomas' book [24].

In the next section we look at the relation of Lyapunov exponents and entropy more closely.

5.2 The Pesin and Ruelle relations

If one examines how partitions of a state space "refine themselves" under iterates of a diffeomorphism, it becomes natural to ask if the spreading measured by the positive Lyapunov exponents might not have a direct relationship to the Kolmogorov entropy of the same diffeomorphism. The answer is yes; there is in fact a close relationship. The exact nature of that relationship has been examined carefully in the work of Ruelle, Pesin, Ledrappier and Young, and others which we review here.

5.2.1 Preliminary definitions

Recall from section 5.1.3 that the Kolmogorov entropy of a map/invariant measure pair $\{F, \mu\}$ is given by

$$h_K(F, \mu) = \sup_{\mathcal{B}} \lim_{n\to\infty} \frac{1}{n} H\left(\mathcal{B} \vee F^{-1}\mathcal{B} \vee F^{-2}\mathcal{B} \vee \ldots \vee F^{-n+1}\mathcal{B}\right) \tag{5.10}$$

Statement 5.2.1 (Kolmogorov Entropy Picture). *Very roughly, one may think of the Kolmogorov entropy as the average log of the number of pieces the partition $\mathcal{B} \vee F^{-1}\mathcal{B} \vee F^{-2}\mathcal{B} \vee \ldots \vee F^{-n+1}\mathcal{B}$ goes to under the next iteration (mapping by F^{-1}) and intersection with $\mathcal{B}$. If F is invertible, then $H(F\mathcal{B}) = H(F)$ for any partition $\mathcal{B}$ so that we obtain*

$$h_K(F, \mu) = \sup_{\mathcal{B}} \lim_{n\to\infty} \frac{1}{n} H\left(\mathcal{B} \vee F\mathcal{B} \vee \ldots \vee F^{n-1}\mathcal{B}\right) \tag{5.11}$$

which is the Kolmogorov entropy of F^{-1}. (This will turn out to have consequences when we discuss Pesin's relation.) Continuing with the invertible case, if at time $t = n$ we "know" $\mathcal{F} \equiv \mathcal{B} \vee F\mathcal{B} \vee \ldots \vee F^{n-1}\mathcal{B}$, meaning we know which tiny piece of this partition we are in, then at time $t = n+1$ we will "know" $\mathcal{F} \vee F\mathcal{F}$ and the average log number of pieces one element of $\mathcal{F}$ splits into is $H(\mathcal{B} \vee F\mathcal{F}|F\mathcal{F}) = H(\mathcal{B}|F\mathcal{F}) = H(\mathcal{B}|F\mathcal{B} \vee \ldots \vee F^n\mathcal{B})$ which converges to $h_K(F, \mu)$ as n goes to ∞.

Definition 5.2.1 (Lyapunov spectrum). *Let F be a differentiable function or map which maps an k-dimensional manifold X to itself. Let μ be an F-invariant probability measure. Let v be any element of $TX(x)$, the tangent space of X at x. We define the Lyapunov spectrum of F at x as follows: The Lyapunov spectrum is the collection of values*

$$\lambda_{F,x}(v) \equiv \lim_{n\to\infty} \frac{1}{n} \log ||DF^n(v)|| \tag{5.12}$$

as v ranges over $TX(x)$. Oseledec's theorem ([111, 51, 62]) tells us that in fact, for μ almost every x, this limit exists $\forall v \in TX(x)$ and it takes on exactly k (not necessarily distinct) values, $\lambda_1(x), ..., \lambda_k(x)$. Furthermore, there are k, 1-dimensional subspaces of $TX(x)$ $E_i(x)$, i=1,...,k such that $TX(x) = E_1(x) \oplus ... \oplus E_k(x)$ and $\lambda_{F,x}(v) = \lambda_i(x)$ for every $v \in E_i(x)$.

Statement 5.2.2 (Lyapunov Picture). *One might imagine a small parallelepiped with edges aligned with the $E_i(x)$. Now tile a small neighborhood of x with these parallelepipeds. Let $\mathcal{I}$ be the set of i such that $\lambda_i > 0$. Then, on average, one step in the iteration of F will cause one parallelepiped to occupy $e^{\sum_{i\in\mathcal{I}} \lambda_i}$ other parallelepipeds (not necessarily densely). That is, if one looks at the intersection of the map-forward of the single parallelepiped with the tiling, there will be $e^{\sum_{i\in\mathcal{I}} \lambda_i}$ elements of the tiling that are intersected. Now actually, one should look at $D^N F$ with sufficiently high N ... since any given single iteration of the mapping may not look at all like the Lyapunov exponents say it should. But this presents no real problem since we can look at F^N for sufficiently high N and repeat the procedure. (In that case we get that "individual" iterates of the map F^N look like the Lyapunov exponents say it should with exponents $N\lambda_1, ..., N\lambda_k$.)*

Finally we define the concept of absolute continuity. It is actually very simple.

Definition 5.2.2. *Let λ and μ be two measures on the space/σ-algebra pair $(X, \mathcal{X})$. λ is said to be absolutely continuous with respect to (a.c.w.r.t) μ if*

$$\mu(\mathcal{A}) = 0 \Rightarrow \lambda(\mathcal{A}) = 0, \tag{5.13}$$

where $\mathcal{A} \in \mathcal{X}$, the σ-algebra of measurable sets for both λ and μ.

A practical result of absolute continuity can be stated in the following way.

Statement 5.2.3 (absolute continuity of measures). *If λ is a.c.w.r.t. μ then anything that can be measured with λ can be measured with μ. The precise way in which this can be done is given by the Radon-Nickodym Theorem (see section 3.2 of [30] for example) which says that if λ is a.c.w.r.t. $\mu \Rightarrow$ there is an μ-integrable f such that $\lambda(\mathcal{A}) = \int_{\mathcal{A}} f d\mu$ for every set $\mathcal{A}$ which λ can measure.*

5.2.2 Ruelle's Inequality

In 1978 Ruelle [76] published a paper proving the following inequality relating Lyapunov exponents and the Kolmogorov entropy.

Theorem 5.2.1 (Ruelle, 1978). *Let F be a C^1 (not necessarily invertible) mapping of a compact manifold X to itself. Let $\chi(x) \equiv \sum_{i|\lambda_i(x)>0} \lambda_i(x)$ where the $\lambda_i(x)$ are the Lyapunov exponents of F at x. Then for any μ in the set of measures invariant under F we have that*

$$h_K \leq \int_X \chi d\mu. \tag{5.14}$$

Note that if μ is ergodic then $\chi(x)$ (which is F-invariant) is constant μ a.e. and we get that $h_K \leq \chi$.

We will look at the reason for the inequality in more detail in the next section, but we will try to anticipate the resolution in the following explanation.

Statement 5.2.4. *Statement 5.2.1 informs us that (roughly) Kolmogorov entropy equals the log of the number of elements that a single element of the partition redistributes itself to (on average), and statement 5.2.2 says that the sum of the positive Lyapunov exponents tells us the same thing, so it seems reasonable to believe that Ruelle's inequality is actually an equality. And in fact, Pesin showed that under the assumption of a smooth invariant measure, we do in fact have strict equality.*

To understand how a strict inequality might arise, recall that in statement 5.2.2 the number of pieces a small parallelepiped spreads out to occupy when mapped *forward with DF was taken to be the number of pieces that this small region mapped forward to under F.* But *if the measure is not evenly distributed among those target pieces, then this number could – according to the invariant measure – be less. (Remember that the Lyapunov products do not know about the invariant measure!) And since the "number of pieces" is actually the factor by which the number of pieces $B(n)$ increases as n increases by one, with $B(n)$ – in the smooth measure case – given by*

$$B(n) \approx \Pi_{i|\lambda_i>0}(e^{n\lambda_i}) \tag{5.15}$$

we might expect something like

$$B(n) \approx \Pi_{i|\lambda_i>0}(e^{nd_i\lambda_i}) \tag{5.16}$$

for $0 < d_i < 1$ in the case in which the dimension of the measure is less than 1 in the direction of λ_i.

This loss of dimension is the reason why Ruelle's inequality cannot, without further assumptions, be strengthened into an equality.

5.2.3 Pesin's Formula

In 1977 Pesin showed [67] that under certain assumptions on the invariant measure, Ruelle's inequality is in fact an equality. Before stating the theorem, we remind the reader that the Riemannian measure is that measure obtained from the volume associated with the Riemannian metric and that we often abbreviate "ν is absolutely continuous with respect to μ" by $\nu << \mu$. Also, by saying that μ and λ are equivalent we mean that $\nu << \mu$ and $\mu << \nu$ are both true.

Theorem 5.2.2 (Pesin's Formula). *If F is a C^2 diffeomorphism of X, a compact Riemannian manifold X, to itself and μ is an F invariant (probability) measure which is equivalent to the Riemannian measure, then*

$$h_K = \int_X \chi d\mu. \tag{5.17}$$

So we see that in fact if we have a measure that is not fractal (i.e. in the language of statement 5.2.4, if we have a measure which does not "lose dimension"), we can strengthen Ruelle's inequality to an equality. In the next section we find the answer to the following question,

Question 2. *What are the weakest assumptions under which Pesin's relation holds?*

We will find that one needs "good behavior" from the measure in the directions that correspond to the positive Lyapunov exponents – a result that we hope is not too surprising to the reader at this point.

5.3 The Definitive Clarification by Ledrappier and Young

In this section, as indicated in the last part of the previous section, we look at the reasons that Ruelle's inequality cannot generally be strengthened into an equality, as well as precisely what is needed, in the form of assumptions on the measure, to get equality. The careful dissection and definitive settlement of these questions was published by Ledrappier and Young in 1985 [54, 55]. A very nice exposition of those results and in fact of many parts of the ergodic theory of differentiable dynamical systems can be found in notes with that very title by Lai-Sang Young [111].

Standing assumptions for this section are that

- X is a smooth compact Riemannian manifold,
- F is a C^2 diffeomorphism of X to itself, and
- μ is an F invariant Borel probability measure on X.[4]

We now proceed to the first of two main results in the pair of 1985 papers by Ledrappier and Young [54, 55].

[4] A Borel measure is a measure whose σ-algebra includes all the open sets. Often it is the smallest such σ-algebra which is complete. Complete means that if $\mu(A) = 0$ and $B \subset A$ then B is also in the σ-algebra and of course it's measure is zero.

5.3.1 The strengthening of Pesin's formula

We begin by explaining several concepts needed to understand the first theorem of Ledrappier and Young. We will attempt to give precise intuitions for the concepts. For the technical details see section 1.3 of [111], section 1 of [54], and the articles by Rohlin [72, 73].

The first concept is that of the unstable foliation of a diffeomorphism F mapping a compact Riemannian manifold to itself. We define the unstable foliation as the collection of

$$W^u(x) \equiv \{y \in X \,|\, \limsup_{n\to\infty} \frac{1}{n} \log(d(F^{-n}x, F^{-n}y)) < 0\}. \tag{5.18}$$

where x ranges over all of X and $d(\cdot,\cdot)$ is the distance induced by the Riemannian metric. We will call the entire collection of such subsets of X, W^u. It turns out that this defines a partition of X into immersed C^2 sub-manifolds.

The second concept is that of a (measurable) partition which is subordinate to W^u. Let $\zeta(x)$ be the element of a partition ζ that contains x. If ζ is a measurable partition of X, then we shall say that it is subordinate to W^u if for μ a.e. x, $\zeta(x)$ is contained in $W^u(x)$ and $\zeta(x)$ contains an an open neighborhood of x in $W^u(x)$.

The third concept we need to introduce before the theorem is that of a system of conditional measures with respect to ζ. This is a bit more tricky in detail, but intuitively, it is quite simple. Essentially, if we have a measure μ on X and a measurable partition ζ, then the system of measures $\mu_{\zeta(x)}$ is said to be a system of conditional measures of μ with respect to ζ if each $\mu_{\zeta(x)}$ is the normalized restriction of μ to $\zeta(x)$. The technicality comes in because it would not be unusual for $\mu(\zeta(x)) = 0\ \forall x$. Due to this possibility, the technically correct requirements are that $\mu_{\zeta(x)}$ be a probability measure for all x and that for all measurable $E \subset X$, $\mu(E) = \int \mu_{\zeta(x)}(E)\, d\mu(x)$.

Finally, we shall say that μ has absolutely continuous conditional measures on the unstable manifolds (W^u) if for every measurable partition ζ subordinate to W^u we have that $\mu_{\zeta(x)} << \nu_{W^u(x)}$ for a.e. x, where $\nu_{W^u(x)}$ is the Riemannian measure on $W^u(x)$ inherited from the Riemannian measure on X by virtue of the fact that W^u is an immersed sub-manifold of X. We shall abbreviate "absolutely continuous conditional measure" by a.c.c.m.

Now we are ready for the theorem.

Theorem 5.3.1 (Ledrappier-Young-1 1985). *If μ is an invariant measure of F, a C^2 diffeomorphism of a compact Riemannian manifold X to itself, then*

$$\mu \text{ has a.c.c.m.'s on } W^u \Leftrightarrow h_K(F) = \int \chi(x)\, d\mu. \tag{5.19}$$

5.3.2 The Definitive Clarification

Now we turn to the final clarification of the relation between Lyapunov exponents and entropy. In this result Ledrappier and Young answered the question, "What is the precise relation between entropy and the Lyapunov spectrum for any C^2 diffeomorphism mapping a compact smooth manifold to itself?". For all the details see the original paper by Ledrappier and Young [55].

We will first state the theorem for the case in which μ is ergodic. Again we will need to introduce a few concepts in order for the theorem to make sense.

First we need to redefine the $\lambda_i(x)$ and the $E_i(x)$ that arose in the definition of Lyapunov exponents above. The $\lambda_i(x)$ will now be the *distinct* values that the exponents take on so that $\lambda_1(x) > \lambda_2(x) > ... > \lambda_r(x)$. The $E_i(x)$ will now be the subspaces of $TF(x)$ associated with these redefined $\lambda_i(x)$. Consequently, the "new" E_i are sums of the old 1-dimensional $E_i(x)$. If we let m_i be the multiplicity of $\lambda_i(x)$, then we have that $m_i = \dim E_i(x)$ and the $\sum m_i = k$, where k is the dimension of X.

We can now define the W^i-manifolds, the nested family of unstable foliations of F. Now since the Lyapunov exponents are constant a.e. μ, we can unambiguously define $\kappa_p = \max\{i|\lambda_i > 0\}$. For $i \leq \kappa_p$ we define the i"th unstable foliation as the collection of

$$W^i(x) \equiv \{y \in X|\limsup_{n\to\infty} \frac{1}{n} \log(d(F^{-n}x, F^{-n}y)) \leq -\lambda_i\}. \tag{5.20}$$

where x ranges over all of X and $d(\cdot,\cdot)$ is the distance induced by the Riemannian metric. We will call the entire collection of such subsets of X, W^i. This defines a partition of X into immersed C^2 sub-manifolds. These sub-manifolds are tangent to sums of the E_i; $W^i(x)$ is tangent to $E_1(x) \oplus ... \oplus E_i(x)$ for $i \leq \kappa_p$.

As before, μ defines conditional measures on the leaves of the W^i. We can then compute the Hausdorff dimension of these measures. For ergodic μ this dimension is constant across leaves so that there is one dimension for each W^i and we shall call this the dimension of μ on W^i.

Now we are ready for the ergodic version of the theorem.

Theorem 5.3.2 (Ledrappier-Young-2.1 1985). *Let F be a C^2 diffeomorphism of X which is compact and smooth and μ be an ergodic F-invariant Borel probability measure on X. Let $\lambda_1, ..., \lambda_r$ be the* distinct *Lyapunov exponents of F. Let δ_i be the dimension of μ on the W^i-manifolds. Then, for $1 \leq i \leq \kappa_p$ there are numbers γ_i with $0 \leq \gamma_i \leq \dim E_i$ such that*

$$\delta_i = \sum_{j \leq i} \gamma_j \tag{5.21}$$

for $i = 1, ..., \kappa_p$ and

$$h_K(F) = \sum_{i \leq \kappa_p} \lambda_i \gamma_i. \tag{5.22}$$

Just before we move on to the non-ergodic case, a definition in preparation for that theorem.

Definition 5.3.1 ($\kappa_p(x)$ and Γ_i). *Let F be a C^2 diffeomorphism of X which is compact and smooth and μ be an F-invariant Borel probability measure on X. Let the number of distinct, positive Lyapunov exponents at x be denoted $\kappa_p(x)$. Because we do not assume the (F, μ) is ergodic, then $\kappa_p(x)$ need not be constant μ almost everywhere. Define $\Gamma_i \equiv \{x | \kappa_p(x) \geq i\}$. In words, Γ_i is the set of x which have at least i distinct, positive Lyapunov exponents. (Since $\kappa_p(x)$ is no longer a constant almost everywhere, the domain of the various functions must be adjusted accordingly: we will use Γ_i for this purpose.)*

Now the non-ergodic case. Remember that any non-ergodic invariant measure has a natural partition into ergodic components. On each of those ergodic components, the previous theorem works so that (modulo some details!) we get

Theorem 5.3.3 (Ledrappier-Young-2.2 1985). *Let F be a C^2 diffeomorphism of X which is compact and smooth and μ be an F-invariant Borel probability measure on X. Let $\lambda_1(x), ..., \lambda_{r(x)}(x)$ be the* distinct *Lyapunov exponents of F. Let $\delta_i : \Gamma_i \to \mathbb{R}$ be the dimension of μ on the $W^i(x)$-manifolds. Then, for $1 \leq i \leq \kappa_p(x)$ there exist measurable functions $\gamma_i : \Gamma_i \to \mathbb{R}$ with $0 \leq \gamma_i(x) \leq \dim E_i(x)$ such that*

$$\delta_i = \sum_{j \leq i} \gamma_j(x) \tag{5.23}$$

for $i = 1, ..., \kappa_p(x)$ *and*

$$h_K(F) = \int \sum_{i \le \kappa_p(x)} \lambda_i(x)\gamma_i(x)\ d\mu(x). \tag{5.24}$$

The main difference to notice is that we end up with an integral over the all the elements of the partition (into ergodic components) which should not be surprising.

5.3.3 Why Ruelle's Inequality is not an Equality

We now come back to the question of why Ruelle's inequality is not an equality. Recall that Ruelle's inequality didn't assume invertibility, while the last three theorems have assumed F to be a C^2 diffeomorphism. In exchange for the loss of generality incurred by assuming F is a C^2 diffeomorphism, we were able to state in theorems 5.3.2 and 5.3.3 precisely *how* equality is *not* attained in Ruelle's inequality.

Paraphrasing those results, a strict inequality occurs when the measure conditioned on the unstable foliation is singular (w.r.t. the Lebesgue measure conditioned on the same foliation). The rate of increase in uncertainty (i.e. the entropy rate) should therefore be measured by the sum $\sum \lambda_i \gamma_i$ instead of the strictly greater quantity $\sum \lambda_i \dim E_i$, where the fact that the invariant measure is fractal in one of the λ_i directions is reflected in the corresponding γ_i being strictly less than $\dim E_i$.

The example shown in figure 5.2 illustrates how one can end up with a strict inequality. We back away from our assumption of invertibility for the example. What we see is that when the dimension of the invariant measure is fractal in the unstable direction, we find that the entropy is reduced – and by the precise amount ($\frac{\log(2)}{\log(3)}$) predicted (*for diffeomorphisms*) by Ledrappier and Young.

5.4 SRB measures

As we discovered in the preceding section, Pesin's relation holds precisely when the invariant measure under investigation is absolutely continuous with respect to the volume measure *on the unstable foliation.* L.S. Young calls these measures SRB measures, but a glance through the literature shows that SRB (also refered to as

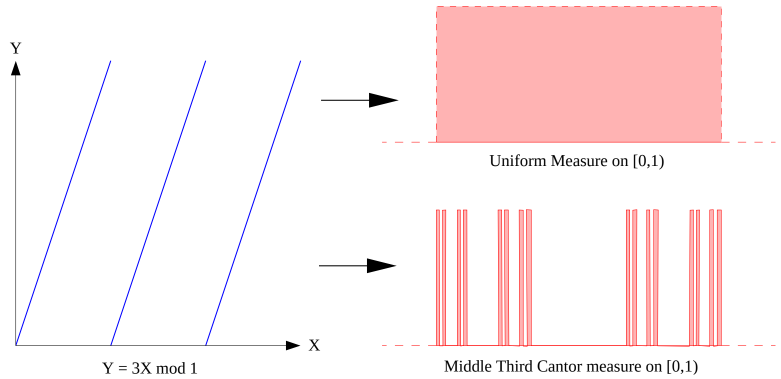


Figure 5.2: $3x \mod 1$ on the uniform density versus $3x \mod 1$ on the middle third Cantor set. In both cases the Lyapunov exponent is $\lambda = \log(3)$, but $h(\mu_{\text{Cantor}}) = \log(2)$ while $h(\mu_{\text{uniform}}) = \log(3)$.

SBR) measures appear to have several different meanings. This section is intended to shed light on that situation.

5.4.1 Three definitions

In the following the dynamics are governed by $x_{i+1} = F(x_i)$.

Definition 5.4.1. *We will say that a measure μ is an acSRB measure if μ is* absolutely continuous *with respect to the Lebesgue measure on the unstable foliation of F.*

This is what was originally called an SBR and then SRB measure by L.S. Young.

Definition 5.4.2. *We will say that a measure μ is an eSRB measure if, for a set of x with* positive Lebesgue measure,

$$\lim_{n->\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(F^i(x)) = \int \phi \, d\mu \tag{5.25}$$

where ϕ is any function in a suitable class of test functions.

This can be understood to mean that the weak limit of the measure concentrated on a single orbit from time 0 to n exists $(n \to \infty)$ AND that limit is "find-able" since we get the same limit for a set of initial conditions x with positive Lebesgue measure. *The key point* – and the thing that the Birkhoff ergodic theorem doesn't give you – is that the set of initial x for which equation 5.25 holds has non-zero *Lebesgue* measure.

Definition 5.4.3. *We will call μ the sSRB measure if we find it as the limit as $\epsilon \to 0$ of a* stochastic *system obtained by adding "noise" with small parameter ϵ.*

5.4.2 Relationships between acSRB, eSRB, and sSRB.

The first of the results concerning the relationship between the three different definitions was contained in a paper by Sinai in 1972, in which he showed that for Anosov systems (defined below), there was a unique acSRB measure supported on the attractor and that this measure was also eSRB. This was followed by two results extending Sinai's work to the Axiom-A case (defined below). The first was by published by Ruelle in 1976 (but submitted in 1973) and the second was published by Bowen and Ruelle in 1975 (but submitted in 1974). Before proceeding to the statement of the pertinent results, we review several definitions.

Definition 5.4.4 (Uniform Hyperbolicity). *Suppose that Λ is a closed F-invariant subset of X and that we can find linear subspaces, E_x^+ and E_x^-, of TF_x for every $x \in \Lambda$ such that $TF_x = E_x^+ \oplus E_x^-$, $\dim E_x^+ + \dim E_x^- = \dim X$, and E_x^+ and E_x^- depend continuously on x. Suppose also that $F(E_x^+) = E_{F(x)}^+$ and $F(E_x^-) = E_{F(x)}^-$. If we can also find constants $C > 0, \lambda > 1$ such that*

$$\begin{aligned} ||T_x F^{-n}(v)||_{F^{-n}(x)} &\leq C\lambda^{-n}||v|| \text{ if } v \in E_x^+ &(5.26)\\ ||T_x F^{n}(v)||_{F^{n}(x)} &\leq C\lambda^{-n}||v|| \text{ if } v \in E_x^- &(5.27)\end{aligned}$$

for all $n \geq 0$, then we say that Λ is a hyperbolic or uniformly hyperbolic invariant set.

Let us put this a bit more informally:

Statement 5.4.1 (Uniform Hyperbolicity). *Uniform hyperbolicity simply means that at each point x on the attractor we can split the tangent space T_xX (what the derivative operates on) into a sum of spaces, E_x^+ and E_x^-, which are F-invariant : $DF(E_x^+) = E_{F(x)}^+$ and $DF(E_x^-) = E_{F(x)}^-$. These subspaces have the special property that on the E^+, DF is expanding at a rate which is globally bounded below and on*

E^-, DF is contracting at a rate which is globally bounded above. Since everything is either expanding or contracting at rates which are globally bounded away from 1, the system is uniformly *hyperbolic.*

If the whole manifold X is uniformly hyperbolic then F is called an *Anosov diffeomorphism.* If the non-wandering set [5] is uniformly hyperbolic and the set of periodic points is dense in the non-wandering set, then F is called an *Axiom-A diffeomorphism.* That Anosov ⇒ Axiom-A follows from the Anosov Closing Lemma, see page 75 of [13].

Remark 5.4.1. *There is a bit of confusion in how the term* hyperbolic *is used. Sometimes a system is said to be hyperbolic if it's Lyapunov spectrum does not include zero. Such systems are also termed non-uniformly hyperbolic. A strict interpretation of even this term is misleading since non-uniformly hyperbolic is used for systems which* may or may not *be hyperbolic. A better term would be not-necessarily-hyperbolic, but non-uniform hyperbolicity has become standard. One can also find hyperbolic being used synonymously with uniformly hyperbolic.*

We are now ready to state the theorem.

Theorem 5.4.1 (Sinai-Ruelle-Bowen [88, 75, 14]). *Suppose F is a C^2 diffeomorphism on an n-dimensional manifold. Suppose A is an Axiom-A attractor* [6] *with basin of attraction W. Then,*

- *There exists a unique acSRB measure with support in A.*
- *There is a subset $\tilde{W} \subset W$, such that $W \backslash \tilde{W}$* [7] *has Lebesgue measure* 0 *and,*

$$\lim_{n->\infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi(F^i(x)) = \int \phi(x) d\mu(x) \tag{5.28}$$

  for all x in $\tilde{W}$ and ϕ is any function in a suitable class of test functions. In other words, μ is an eSRB measure.

[5] Non-wandering set $\equiv$ $\{x \in X| \ \forall$ nbhd's $\mathcal{B}_x$ of x, $\mathcal{B}_x \cap F^n(\mathcal{B}_x) \neq \emptyset$ infinitely often in $n\}$.

[6] An Axiom-A attractor is a set Λ such that 1) $F : \Lambda = \Lambda$ (i.e. Λ is an invariant set under F), 2) F is uniformly hyperbolic on Λ, 3) There is a compact neighborhood of Λ, U such that $F(U) \subset U$ and $\Lambda = \cap_{n \geq 0} F^n(U)$, and 4) F is topologically transitive on Λ. F is topologically transitive on W if for any two open sets A and B in W, there exists an n such that $T^{-n} A \cap B \neq \emptyset$.

[7] $W \backslash \tilde{W}$ is defined to be $\{x \in W | x \notin \tilde{W}\}$.

The relation of acSRB and sSRB has been investigated by Kifer, Young and Liu. The first result, published by Kifer in 1974 [52] established that *stochastic dynamical systems with parameter* ϵ derived from Axiom-A systems have stationary measures which converge to the acSRB measure of the Axiom-A system as ϵ goes to 0. Very roughly, one should think of an operator O_F^ϵ that maps forward a measure μ in the following way.

Let ρ be the density of μ with respect to the Lebesgue measure. Define $\psi_{x^*}^\epsilon = \psi_{x^*}^\epsilon(x, x^*)$ to be a "bump" function with support in the epsilon ball centered on x^*, where $\int \psi_{x^*}^\epsilon(x, x^*)dx = 1$. We can think of $\psi_{F(x^*)}^\epsilon(y, F(x^*))$ as the probability density for $y = F_S^\epsilon(x^*)$, (where F_S^ϵ is the stochastic version of F with parameter ϵ), given we begin, before application of F_S^ϵ at x^*. We can now define $O_F^\epsilon(\rho)(y) = \int \psi_{x^*}^\epsilon(y, F(x^*))\rho(x^*)dx^*$. This operator O_F^ϵ has stationary densities which we might call ρ_F^ϵ. The result can be restated. For an Axiom-A F, $\rho_F^\epsilon \to \mu$ as $\epsilon \to 0$ where μ is the acSRB measure of F. (Section IV.H of [28] contains an exposition of this.) Young came to similar results for *random dynamical systems* in a paper published in 1986 [110]. More recently, Liu and collaborators have continued this line of investigation, see [58, 57, 59].

Pugh and Shub have shown that in fact one needs less than Anosov or Axiom A assumptions to get that the acSRB measure is an eSRB measure. More precisely,

Theorem 5.4.2 (Pugh and Shub, 1989 [69]). *If μ is an ergodic acSRB measure with no zero Lyapunov exponents, then there exists a set of x with positive Lebesgue measure such that,*

$$\frac{1}{n}\sum_{i=0}^{n-1} \phi(F^i(x)) \to \int \phi d\mu. \tag{5.29}$$

In other words, μ is an eSRB measure.

5.4.3 Examples and results for specific systems

In a paper published in 1993 [7] Benedicks and Young showed that the Hénon system possesses a unique acSRB measure for a set of parameter values with positive Lebesgue measure. It follows from Pugh and Shub's theorem that the acSRB measure is also an eSRB measure. More recently, Tucker obtained that the Lorenz system has an eSRB measure and gave an affirmative answer to the question of whether or not the numerical simulations of Lorenz's system are "real" [97]. It appears that Tucker at least comes into the vicinity of an acSRB measure since the proof of existence uses the work of Viana's in which the measure – with the stable directions *quotiented out* – is absolutely continuous w.r.t the Lebesgue measure (see lemma 4.8 on page 96 of [99]). This is of course not a proof that the eSRB measure is an acSRB measure.

There have been a fair number of other results on the existence of SRB measures for systems with various assumptions about the rates of expansion. See [3, 4, 12, 19, 20, 23, 22, 26, 29, 41, 42, 44, 45, 47, 56, 79, 84, 85, 94, 98].

An example of how we may fail to get an acSRB was given by Bowen in [13]. As illustrated in figure 5.3 below the only invariant density which comes via an ergodic average as in the case of an eSRB measure is the Dirac δ measure concentrated at p_2. Note that even though the measure one gets from averages is not acSRB it is of course eSRB!

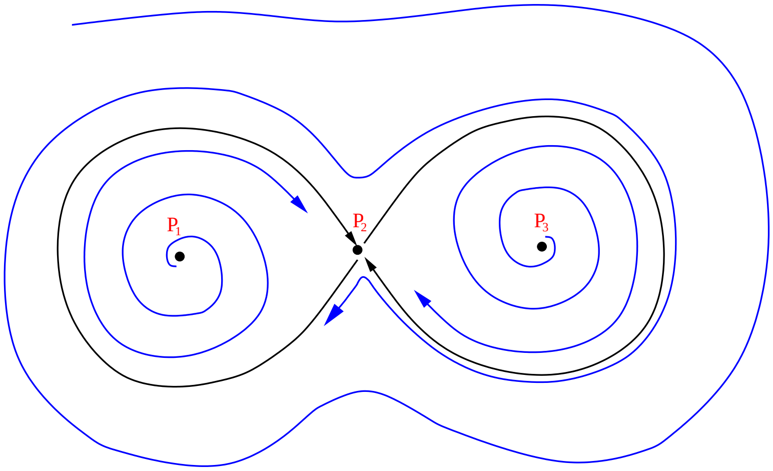


Figure 5.3: Bowen's ∞ counterexample: A flow in the plane with sources at P_1 and P_3, and a hyperbolic saddle at P_2. The unstable manifolds of P_2 connect up with the stable manifolds of P_2 (homoclinic connections). The whole figure ∞ is globally attracting. This is an example of an eSRB measure which *is not* an acSRB measure. The eSRB measure is the Dirac δ measure concentrated at P_2 – any initial point except P_1 and P_3 ends spending asymptotically longer and longer times in arbitrarily small neighborhoods of P_2. But this eSRB measure has dimension 0 in the unstable direction and so it cannot be absolutely continuous w.r.t. the Lebesgue measure on the unstable manifold at P_2.

5.4.4 Summary: SRB measures

In summary, there are three main definitions of an SRB measure which are equivalent in the case of axiom A (and therefore Anosov) systems. Except for the result of Pugh and Shub [69], the relation of these different definitions has yet to be worked out for other systems. Other notions of *generalized SRB measures* have been defined. See for example [77].

In the case of specific systems we have several results proving the existence of SRB measures, most notably the results mentioned above for the Hénon and Lorenz systems. Finally, that there are obstructions to the existence of SRB measures can be seen in Bowen's simple counterexample.

5.5 Back to measuring model fidelity

We turn again to the question of modeling performance and the use of the relative entropy gap G as a measure of that performance. The practical utility of G is connected to the answers to several questions.

1. What does the "real" (perfect computation, asymptotic valued) G tell us about the relation of the model and the system which generated the data?

2. How does finite precision affect these calculations? (What possible errors could we make in our inferences about the model-system relation due to the fact that our calculations are finite precision?)

3. How could one numerically estimate G?

4. What effect does finite data have on our result? In other words, what can we say about convergence rates?

5.5.1 The Entropy Gap: What Does it Mean?

Let $\lambda_1 > \lambda_2 > ... > \lambda_r$ be the distinct Lyapunov exponents of F, n_i be the dimensions associated with these distinct λ_i, $p = \max\{i|\lambda_i > 0\}$, and E_i be the subspaces associated with the λ_i. Then the Kolmogorov entropy of the system computed via the Lyapunov exponents can be written as follows

$$h_K = \sum_{\lambda_i | i \leq p} \lambda_i \gamma_i \tag{5.30}$$

where $0 \leq \gamma_i \leq n_i$, and $\gamma_i = n_i$ if μ is absolutely continuous in the λ_i directions and

$$h_{L(F)} \equiv \sum_{\lambda_{i|i \leq p}} \lambda_i n_i. \tag{5.31}$$

Given a dynamical system F, a partition $\mathcal{B}$, and a model θ, we may estimate $h_{L(F)}$ and $h(\mathcal{B}, F, \mu : \theta)$(see section 5.5.4). A fundamental measure of the fidelity of a model θ to a true processes μ is relative entropy rate $d(\mu || \theta)$. The relative entropy rate:

- Gives the cost in bits per time step of doing a data compression based on the wrong model. See [24] pages 89-90.
- Bounds the wealth doubling rate in gambling or investment. See [24] pages 127 and 129.
- Gives error exponents for classifiers. See [24] page 312 and the original paper by Chernoff [21].

Using equations 5.30 and 5.31 and the fact that the entropy rate is the sup (over measurable partitions) of the entropy rates computed on partitions, we get

$$\begin{aligned} h_{L(F)} &\geq h_K & (5.32) \\ h_K &\geq h(\mathcal{B}, F, \mu) & (5.33) \end{aligned}$$

and since

$$d_{\mathcal{B},F}(\mu || \theta) = h(\mathcal{B}, F, \mu : \theta) - h(\mathcal{B}, F, \mu) \tag{5.34}$$

we arrive at

$$d_{\mathcal{B},F}(\mu || \theta) \geq h(\mathcal{B}, F, \mu : \theta) - h_{L(F)} = \text{The Gap}, \tag{5.35}$$

which permits us to conclude that a large gap implies a large relative entropy rate.

If the gap is zero, the relative entropy rate is $d_{\mathcal{B},F}(\mu||\theta) = (h_{L(F)} - h_K) + (h_K - h(\mathcal{B}, F, \mu))$, which we conjecture[8] means:

- the model is missing the fractal aspects of the measure along the unstable foliation, and/or
- the partition $\mathcal{B}$ is not generating (i.e. $h_K > h(\mathcal{B}, F, \mu)$).

If models that capture such fractal aspects are deemed to be unacceptably complex or sensitive in an application, then a zero gap indicates that the model may be optimal[9] within the acceptable class.

5.5.2 Finite Precision: Lyapunov Exponents

We now consider the effects of finite precision on the calculation and meaning of the Lyapunov spectrum. We begin with notation and assumptions.

First, let us denote the (floating point) finite precision representation of $\mathbb{R}^k$ by $\tilde{\mathbb{R}}^k_{\epsilon,K}$ where it is understood that $\epsilon = 1/2^{\{\text{number of bits in mantissa}\}}$ and K is the number of bits in the exponent. To simplify notation we will define $A \equiv \tilde{\mathbb{R}}^k_{\epsilon,K}$. Let $\tilde{F}$ be the rounded version (on A) of F. When we speak of derivatives of $\tilde{F}$ we mean the finite precision representations of DF which we denote $\widetilde{DF}$. Note that this is not the same thing as $D\tilde{F}$ which isn't even defined (A is discrete and finite!).

There are difficulties inherent in the floating point representation of a diffeomorphism F. The first is that $\tilde{F}$ is almost always *NOT* one-to-one. This can be seen by considering what happens when the diffeomorphism maps from a region with one set of finite precision exponents to another with a different set of finite precision exponents. (We have a finite precision exponent for each coordinate or dimension.) Then if the derivative has a magnitude that is smaller than one in a direction corresponding to an increasing exponent, we are forced to map (round) many points in the domain to one point in the range. See figure 5.4.

[8] A caution is warranted by a result given in [87] of two *different* processes for which the relative entropy rate is equal to zero!

[9] If the partition implicit in the measurement is not generating, then the "best" model can give a negative gap even without using fractal models and a zero gap would indicate there is still room for improvement.

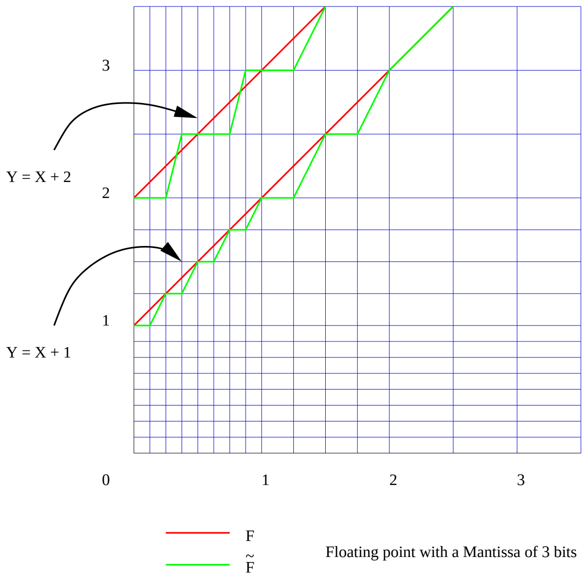


Figure 5.4: Example of loss of injectivity. This figure shows the graphs of $y = x+1$ and $y = x+2$ as well as their finite precision representations when the mantissa has 3 bits. (**Note: This is not the standard floating point representation!** We have simplified things by assuming numbers are positive. And the set of numbers we can represent are exactly those one obtains by looking at $BBBx2^E$ where B = 0 or 1 and E is any positive or negative integer. By considering only positive numbers we avoid having to think about a sign bit.)

The second difficulty is that moving from F to $\tilde{F}$ cannot be viewed as a C^1 small perturbation of F even though it is small C^0. Let us be a bit more careful here and reduce the set A to the subset of A made up of the periodic orbits and fixed points – we will call this subset A_p. On this set, $\tilde{F}$ is injective. Even under this reduced mapping, we find that we could not have gotten there from F by a C^1 small perturbation. Being a bit more precise, suppose that F^* is some diffeomorphism which agrees with $\tilde{F}$ on A_p. While we can make $||F - F^*|| \approx \epsilon$, typically $||DF - DF^*|| \approx 1$ cannot be avoided. This is illustrated in figure 5.5 which shows a circle diffeomorphism in which the finite precision representation induces $\mathcal{O}(\epsilon)$ perturbations to the mapping, but $\mathcal{O}(1)$ perturbations to the derivative (which can be seen with the help of the mean value theorem). We have chosen a uniform discretization for simplicity. The problem with making the meaning of calculated Lyapunov exponents precise is that the finite precision system $\tilde{F}$ is a map from A – a discrete and finite set – to itself. What exactly do the derivatives $\widetilde{DF}$ tell

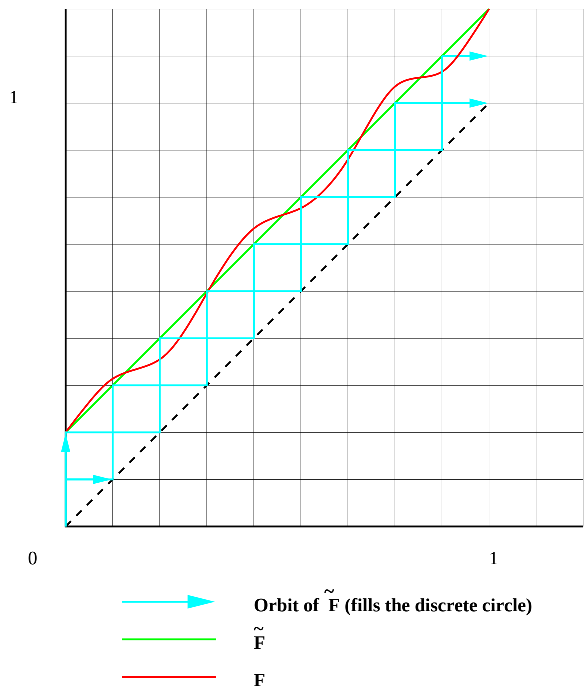


Figure 5.5: Example of a finite precision perturbation for which $||F - F^*||_{C^0} \approx \epsilon$, but $||DF - DF^*|| \approx 1$ for all possible extensions F^* of $\tilde{F}$.

us about $\tilde{F}$? And since every orbit is eventually periodic then the finite precision system $\tilde{F}$ has entropy equal to 0.

There are at least two routes to explore:

1 we can try to show that the Lyapunov exponents calculated from the finite precision system $\tilde{F}$ converge to those of F as $\epsilon \to 0$ and

2 we can *fix* ϵ and try to say something about the existence of an infinite precision system $\tilde{F}^* : \mathbb{R}^k \to \mathbb{R}^k$ which is an extension of $\tilde{F}$ whose Lyapunov exponents and entropy are given by the calculations done on the finite precision system $\tilde{F}$.

We will now formulate two results following these two approaches.

Convergence as $\epsilon \to 0$.

What can we say about the asymptotic convergence of Lyapunov exponents as round-off errors go to zero? We answer this with a theorem asserting that asymptotically, things behave well. But first a warning. Practically speaking, asymptotic results can be misleading to the unwary since roundoff errors $\leq 2^{-100,000,000,000}$ might be required for 2 digits of accuracy in the result! Though it is true that things are often better than this, we will use an estimate of orbital divergence guaranteeing (roughly) that getting the Lyapunov exponents to within ϵ will require roundoff errors less than $\delta = \exp(-N\alpha)$ where N can be large and $\alpha > 0$. Therefore in our case "For sufficiently small roundoff error" means vanishingly small roundoff error!

Definition 5.5.1. *Let F be a C^2 diffeomorphism of a compact k-dimensional manifold M to itself. Let $\tilde{F}$ be a finite precision representation of F on M. Then* **roundoff** $\equiv$ *supremum of the round-off error* $= \sup_{x \in M}(|F(x) - \tilde{F}(x)|)$.

Theorem 5.5.1. *Let F be a C^2 diffeomorphism of a compact k-dimensional manifold M to itself. Let μ be an invariant probability measure of F. Then given an $\epsilon > 0$ we may choose $\delta(\epsilon)$, $N(\epsilon)$, and W_ϵ such that for any finite precision representation $\tilde{F}$ with* **roundoff** $< \delta(\epsilon)$ *we have:*

- $\forall x \in W_\epsilon$, *the finite time Lyapunov exponents given by*

$$\tilde{\Lambda}(x) \equiv \{\lambda \in \mathbb{R} | \lambda = \frac{1}{N(\epsilon)} \log ||\widetilde{DF}^{N(\epsilon)}(x)(v)|| \quad \textit{for some} \quad v \in TM_x\} \tag{5.36}$$

  are within ϵ of the Lyapunov exponents of F calculated at any y in a $\delta(\epsilon)$ neighborhood of x.

- $\mu(W_\epsilon) \geq 1 - \epsilon$.

Proof: By "the Lyapunov spectrum α is within ϵ of Lyapunov spectrum β" we will mean that $\max\{|\lambda_1^\alpha - \lambda_1^\beta|, ..., |\lambda_k^\alpha - \lambda_k^\beta|\} \leq \epsilon$, where we have reverted, for the purpose of this sentence, to exactly k, not necessarily distinct Lyapunov exponents which are arranged in descending order.

Define W_n to be the set of $x \in M$ such that the Lyapunov spectrum computed from the first n iterates of F with initial condition x has converged to within ϵ_1 of the true Lyapunov spectrum of F. Oseledec's theorem assures us that $W = \cup_1^\infty W_n$ is a set of full measure (i.e. $\mu(W) = 1$. Let $N(\epsilon_1, \epsilon_2)$ be the first natural number k such that $\mu(W_k) \geq 1 - \epsilon_2$. Since F is Lipschitz on M with Lipschitz constant C_F for some $C_F > 0$, then we can use the mean value theorem for multidimensional spaces to get that orbits computed with "mistakes" bounded by

$$\delta_1(\delta_2, n) \equiv \frac{\delta_2}{(C_F^{n+1} - 1)/(C_F - 1)} \tag{5.37}$$

stay within δ_2 of each other for at least n steps. We can also conclude that since F is C^2, that there is a $\delta_3(\epsilon_3, n)$ such that if the sequence $\{y\}_1^n \equiv \{y_1, y_2, ..., y_n\} \in M$ is within $\delta_3(\epsilon_3, n)$ of an orbit of F, call it $\{x\}_1^n$ (i.e. $x_{i+1} = F^i(x_i)$ for $i = 1, ..., n-1$ and $||y_i - x_i|| \leq \delta_3(\epsilon_3, n)$ for $i = 1, ..., n$), the n-time Lyapunov spectrum computed along $\{y\}_1^n$ is within ϵ_3 of the n-time Lyapunov spectrum calculated along $\{x\}_1^n$. Finally, there is a $\delta_4(n, \epsilon_4)$ such that if the elements of a sequence of $m \times m$ matrices $\{A_1, ...A_n\}$ are all within $\delta_4(n, \epsilon_4)$ of the elements of another sequence of matrices $\{B_1, ..., B_n\}$, then

$$||\Pi_{i=1}^n A_i - \Pi_{i=1}^n B_i||_2 \leq \epsilon_4. \tag{5.38}$$

This follows from the equivalence of norms in finite dimensional spaces AND the fact that we can restrict ourselves to matrices in some bounded subset of $\mathbb{R}^{m^2}$. Finally, since the norms of the n-fold products of derivative matrices are bounded **away** from zero we may choose ϵ_4 such that if $||\Pi_{i=1}^n A_i - \Pi_{i=1}^n B_i||_2 \leq \epsilon_4(\epsilon_5)$, then $\log ||\Pi_{i=1}^n A_i \; v|| - \log ||\Pi_{i=1}^n B_i \; v||_2 < \epsilon_5$ for $||v|| = 1$.

Now we begin to put all these pieces together. Choosing ϵ_1 and ϵ_2 we get a set $W_{N(\epsilon_1,\epsilon_2)}$ with $\mu(W_{N(\epsilon_1,\epsilon_2)}) > 1 - \epsilon_2$ such that for every $x \in W_{N(\epsilon_1,\epsilon_2)}$, the $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum calculated at x is within ϵ_1 of the true spectrum at x. Next, requiring that **roundoff** $\leq \tilde{\delta_1} \equiv \delta_1(\delta_3(\epsilon_3, N(\epsilon_1, \epsilon_2)), N(\epsilon_1, \epsilon_2))$ will force the actual $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum to be within ϵ_3 of the $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum computed on the finite precision orbit, *but with exact derivatives*. Requiring that **roundoff** $\leq \tilde{\delta_2} \equiv \delta_4(N(\epsilon_1, \epsilon_2), \epsilon_4(\epsilon_5))$ ensures that the computed $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum is within ϵ_5 of the $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum computed on the finite precision orbit with exact derivatives.

Condensing things once more, choosing **roundoff** $\leq \min\{\tilde{\delta_1}, \tilde{\delta_2}\}$, we get that the computed $N(\epsilon_1, \epsilon_2)$-time Lyapunov spectrum is within $\epsilon_1 + \epsilon_3 + \epsilon_5$ of the true spectrum on a subset of M with measure $\geq 1 - \epsilon_2$. Since $\epsilon_1, \epsilon_2, \epsilon_3$, and ϵ_5 were chosen independently of each other we are done. **QED.**

Remark 5.5.1. *The impracticality mentioned before the proof comes from the exponentially small δ_1. If our system is uniformly hyperbolic then we can use a shadowing orbit that does not – a-priori – suffer from a constraint on the numerical errors that is exponentially small. In the next subsection, we consider what sort of result we need for* practical *computation of Lyapunov exponents to be useful.*

Remark 5.5.2. *We have not tried to optimize this result. We actually only need errors that are small in exponent. I.E. I could be looking at $e^{n\epsilon}$ instead of ϵ since I am really interested in something like the behavior of $\frac{1}{n}\log(\mathbf{quantity}(n))$ instead of simply $\mathbf{quantity}(n)$. But working with exponents* would not *remove the problem with the vanishing of δ_1.*

Remark 5.5.3. *When considering the computation of Lyapunov exponents on a finite precision machine, Bochi's recent proof of a theorem put forward by Mañé in 1983 [61] gives us a pause. Here is the theorem (for which Mañé never published a proof).*

Theorem 5.5.2 (Bochi-Mañé [61, 63, 11]). *Let M be a compact Riemannian 2-dimensional smooth manifold and μ be the normalized area. Denote by $Diff^1_\mu(M)$ the set of all μ preserving C^1 diffeomorphisms with the C^1 topology. Then there exists a residual subset $\mathcal{R} \subset Diff^1_\mu(M)$ such that every $f \in \mathcal{R}$ is either Anosov or*

$$\lambda^+(f,x) = \lim_{n\to+\infty} \frac{1}{n}\log||Df^n_x|| = 0 \tag{5.39}$$

for μ almost every x. ($\lambda^+(f,x) \equiv$ largest Lyapunov exponent.)

Note that this result DOES NOT imply that close to every non-Anosov area-preserving diffeomorphism there is one with zero Lyapunov exponents. It DOES imply that any C^1-open set of non-Anosov, area-preserving diffeomorphisms would have a dense subset of area-preserving diffeomorphisms with zero exponents. Consequently, the existence of a non-zero Lyapunov exponent diffeomorphism in this open set, would permit us to conclude that the Lyapunov exponents are not *continuous in F (i.e. small changes in the mapping F, can lead to jumps in the spectrum of exponents).*

Remark 5.5.4. *A recent result of Sauer's [82] warns against assuming that even though a system might not be shadow-able the time averages come out OK. He argues in fact, that for a test function ϕ, one expects $\langle\phi\rangle_{computed} - \langle\phi\rangle_{true} = K\delta^h$ where $\langle\ \rangle$ indicates integration against the "natural measure", i.e. time averages.*

Here K might be zero (things may work out) but if K is not zero, then he gives an argument (not a proof yet) that divergence is governed by the strength of the non-hyperbolicity h, and the size of the numerical errors δ (which we are calling **roundoff***). This result underlines the importance of the stopping time prescribed in theorem 5.5.1.*

The existence of an extension for a fixed ϵ.

Having already observed the difficulties associated with the finite precision representations of diffeomorphisms, we can see that wishing for some sort metric "conjugacy" between the original map F and some extension $\tilde{F}^*$ of the finite precision representation $\tilde{F}$ is probably too much to ask for. Instead, we might simply ask for an extension for which the Lyapunov exponents and entropy are identical (or close) to those calculated from $\tilde{F}$.

We now conjecture a desired result – one that needs to be true for a calculation using a finite precision representation to have meaning. We use the notation introduced in the first part of this subsection. Recall that **roundoff** $\equiv$ supremum of the round-off error $= \sup_{x \in M}(|F(x) - \tilde{F}(x)|)$, that A is the finite precision approximation of $\mathbb{R}^k$ and that A_p is the subset of A made up of the periodic orbits and fixed points of $\tilde{F}$.

Conjecture 5.5.1. *Suppose that F is a diffeomorphism mapping $\mathbb{R}^k$ to itself with an invariant measure μ which is also ergodic. Suppose further that $\tilde{F}$ is the finite precision (floating point) representation mapping $A \equiv \tilde{\mathbb{R}}^k_{\epsilon,K}$ to itself. Denote the restriction of $\tilde{F}$ to A_p by $\tilde{F}_p$. Let B be a closed ball centered on the origin in $\mathbb{R}^k$ big enough to contain A_p in the interior. Let $c\Lambda_p$ be the convex hull[10] of the Lyapunov spectra of $\tilde{F}_p$ – we may get a different spectrum for each periodic orbit in A_p. Then:*

1 *There exists a C^2 diffeomorphism $\tilde{F}^*_p$ of B* onto *B such that $||\tilde{F}_p(x) - \tilde{F}^*_p(x)|| \leq$* **roundoff** *and $||\widetilde{DF}_p(x) - D\tilde{F}^*_p|| \leq$* **roundoff** *for all $x \in A_p$.*

2 *The Lyapunov spectra of $\tilde{F}^*_p$, denote them by $\Lambda^*_p(x)$, lies in a* **roundoff** *neighborhood of $c\Lambda_p$ for all $x \in A_p$.*

[10] The *convex hull* of a point set A is the smallest convex set containing A. Equivalently, the convex hull of A is the intersection of all convex sets containing A.

5.5.3 Finite Precision: Cross Entropy

So far we have addressed finite precision effects on the computation of the $h_{L(F)}$ part of $h(\mathcal{B}, F, \mu : \theta) - h_{L(F)}$. What of the cross entropy of the model computed along the system's orbit? That is what of $h(\mathcal{B}, F, \mu : \theta)$? How does finite precision affect the computation of this quantity?

Recall the definition of entropy rate, (Eqn. 5.4). For a stationary stochastic process with finite alphabet $\mathcal{A}$ we get

$$h = - \lim_{n \to \infty} \frac{1}{n} \sum_{\{a\}_1^n} \mu\left(\{a\}_1^n\right) \log\ \mu\left(\{a\}_1^n\right) \tag{5.40}$$

$$= - \lim_{n \to \infty} \sum_{\{a\}_1^n} \mu\left(\{a\}_1^n\right) \log\ \mu\left(a_n \middle| \{a\}_1^{n-1}\right) \tag{5.41}$$

$$= - \lim_{n \to \infty} \sum_{\{a\}_{-n+1}^0} \mu\left(\{a\}_{-n+1}^0\right) \log\ \mu\left(a_0 \middle| \{a\}_{-n+1}^{-1}\right) \tag{5.42}$$

where the last line follows from stationarity.

Suppose we have one (long) symbol sequence. Can we use this process sample to compute the entropy rate? One would hope so since summing over all possible sequences to compute the entropy rate is computationally out of reach. And in fact, this is exactly what the theorems of Shannon-McMillan (SM) and Shannon-McMillan-Breiman[11] (SMB) assures us we can do: we can compute the entropy rate by using single symbol sequences that are long enough. Actually SM is implied by SMB, but we will quote both of them here.

Theorem 5.5.3 (Shannon-McMillan). *For an ergodic stationary process with a finite alphabet $\mathcal{A}$, probability measure μ, and entropy rate equal to h, for any $\epsilon > 0$ there is an n_0 and for all $n \geq n_0$, there is a set of outcomes $T_n \in \mathcal{A}^n$ such that:*

$$\text{(a) } \mu(T_n) \geq 1 - \epsilon \tag{5.43}$$

$$\text{(b) } \forall \{a\}_n^T \in T_n,\ \ \exp(-n(h+\epsilon)) \leq \mu(\{a\}_1^n) \leq \exp(-n(h-\epsilon)). \tag{5.44}$$

Breiman's generalization says

[11] It should be noted that it is these theorems which form the conceptual basis for the coding of symbol sequences from ergodic sources. The theorems say, in effect, that coding systems exist which code symbol sequences emanating from an ergodic source with bit sequences of (asymptotic) length equal to the number of symbols in the sequence times the binary entropy rate of the ergodic source.

Theorem 5.5.4 (Shannon-McMillan-Breiman [15, 16]).

$$\lim_{n\to\infty} -\frac{1}{n}\log\left(\mu(\{a\}_1^n)\right) = h \textit{ with probability one.} \tag{5.45}$$

We would like to have theorems like the two above for cross entropy. Following Eqn. 5.8, if we have a stationary process with a probability measure μ and a finite alphabet $\mathcal{A}$ then the cross entropy rate for a second probability measure θ is

$$h(\mu:\theta) = \lim_{n\to\infty} \frac{1}{n} H\left(\{A\}_1^n, \mu:\theta\right), \tag{5.46}$$

where

$$H\left(\{A\}_1^n, \mu:\theta\right) = -\sum_{\{a\}_1^n} \mu\left(\{a\}_1^n\right) \log\left(\theta\left(\{a\}_1^n\right)\right). \tag{5.47}$$

So, we would like a theorem like

"Desired Theorem" 5.5.1 (Shannon-McMillan for Cross-Entropy). *For an ergodic stationary process with a finite alphabet $\mathcal{A}$, probability measure μ, and cross entropy rate for a second measure $h(\mu:\theta)$, for any $\epsilon > 0$ there is an n_0 and for all $n \geq n_0$, there is a set of outcomes $T_n \in \mathcal{A}^n$ such that:*

$$\text{(a) } \mu(T_n) \geq 1-\epsilon \tag{5.48}$$

$$\text{(b) } \forall \{a\}_1^n \in T_n,\ \exp(-n(h(\mu:\theta)+\epsilon)) \leq \theta(\{a\}_1^n) \leq \exp(-n(h(\mu:\theta)-\epsilon)). \tag{5.49}$$

Instead of proving this generalization of the original Shannon-McMillan theorem, we will prove a version of the Shannon-McMillan-Breiman theorem valid for cross entropy. We make assumptions on θ which, in the usual SMB case follow from stationarity and ergodicity. The assumptions on θ imply that:

- the present becomes asymptotically independent of the past as the past becomes more distant, i.e. $g_n \equiv -\log\theta(a_0|a_{-1},...,a_{-n})$ converge μ a.e.
- and we can apply the dominated convergence theorem, i.e. $E(\sup_k |g_k|) < \infty$.

We believe that one should be able to get the result assuming only that θ is stationary, in which case we would need to allow the cross entropy rate to be equal to $+\infty$. See remark 5.5.5 which follows the proof of the theorem. Now the theorem.

Theorem 5.5.5 (Shannon-McMillan-Breiman for Cross Entropy). *If we assume that the random variables $g_n \equiv -\log\theta(a_0|a_{-1},...,a_{-n})$ converge μ a.e. and they are bounded above by an (L_1) integrable function (i.e. $E(\sup_k|g_k|) < \infty$) then we have that*

$$\lim_{n\to\infty} -\frac{1}{n}\log\left(\theta(\{a\}_1^n)\right) = h(\mu:\theta) \textit{ with probability one.} \tag{5.50}$$

Proof: We prove this using the generalized ergodic theorem of Breiman's [15].

Theorem 5.5.6 (Breiman). *Let T be a metrically transitive one-to-one measure preserving transformation of the probability space $(\Omega, \mathcal{B}, \mu)$ onto itself. Let $g_0(\omega), g_1(\omega), ...$ be a sequence of measurable functions on Ω converging a.s. to the function $g(\omega)$ such that $E(\sup_k|g_k|) < \infty$. Then*

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k\omega) = Eg \ \textit{ a.s.} \tag{5.51}$$

Define $g_n \equiv -\log\theta(a_0|\{a\}_{-n}^{-1})$. Let T be the shift map. By assumption we satisfy the conditions of Breiman's ergodic theorem. Therefore we have that

$$-\frac{1}{n}\log\theta(a_1,...,a_n) = -\frac{1}{n}\sum_{k=1}^{n} g_k(T^k(\omega)) = Eg \ \text{ a.s..} \tag{5.52}$$

The dominated convergence theorem then permits us to observe that

$$Eg = \lim_{n\to\infty} Eg_n \tag{5.53}$$

and since

$$h(\mu:\theta) = \lim_{n\to\infty}\tfrac{1}{n}\textstyle\sum_{k=1}^{n} Eg_k \tag{5.54}$$

$$= \lim_{n\to\infty} Eg_n \tag{5.55}$$

we are done. **QED.**

Remark 5.5.5. *If θ is Markov of any order then $g_k = g_N$ for all $k \geq N$ where N is the Markov order. We then get both conditions on g_k satisfied. Thus θ being Markov of any order implies that the SMB for cross-entropy holds. This makes one think of taking a general stationary θ and approximating it by Markov processes. As noted before the theorem, we will need to permit the entropy rate to be $= \infty$. This is work for the future. (Actually, this is part of what Algoet and Cover do in their "Sandwich" proof of SMB [2].)*

Remark 5.5.6. *We went from the stochastic process characterized by θ to a measure preserving transformation without much comment. (See Breiman [17] section 6.2 for more on these alternate but equivalent representations.)*

Next we address briefly the practical computation of the Lyapunov exponents and the loglikelihoods discussed so far.

5.5.4 Practical Computation

Essentially, the Benettin procedure [8, 9] is a method for computing all the Lyapunov exponents. One way to *think* of the procedure is that one simply picks a large enough N, then computes the matrix $DF^N(x) = DF(F^{N-1}(x)) \circ DF(F^{N-2}(x) \circ ... \circ DF(F(x)) \circ DF(x)$. Next the SVD of this matrix is computed:

$$DF^N(x) = O_l \circ S \circ O_r^t. \tag{5.56}$$

The singular values of $DF^N(x)$ lie along the diagonal of S and $\frac{1}{N}$ times the logs of those singular values gives the finite time Lyapunov spectrum. The right singular vectors (columns of O_r) form a basis for the Lyapunov subspaces. How this is actually *done* in the algorithm is different. We refer the reader to the details in [8, 9].

We are assuming throughout that one can compute the log likelihoods with ease so that the only remaining question is, "What is the rate of convergence of the log likelihood estimates?" This will be briefly addressed in the next subsection.

5.5.5 Convergence Rates

What can we say about convergence rates? In complete generality, not much. We now list existing results concerning the calculation of Lyapunov exponents and entropy.

Convergence of Lyapunov Exponent Calculations

Very little has been written about the (rigorous) convergence of numerical routines which calculate Lyapunov exponents. A paper by Brian Hunt [43] is one of the few examples in which the Lyapunov exponents are calculated and rigorous bounds given. A paper of Mera and Moran [64] proves the asymptotic convergence of the Eckmann and Ruelle algorithm in the case of exact arithmetic (not finite precision). In the Eckmann and Ruelle algorithm one has at least two limiting processes to worry about – the convergence of the tangent map approximations (estimated from measurements) and convergence of the finite time Lyapunov exponents to the Lyapunov exponents. Since Oseledec's theorem eliminates worry about convergence of the finite time Lyapunov exponents, one must deal with the convergence of the tangent map approximations.

In addition to the papers by Benettin et.al. [8, 9] in which they assume the system (F) is available and in which they use the Gram-Schmidt decomposition, there is the original paper by Wolf et.al. [106] which initiated a series of other papers all on the computation of Lyapunov exponents from data. Slightly later, Eckmann et.al. [27] published a study of an alternative to the Wolf et.al. algorithm. This algorithm, first suggested in [28], uses the QR decomposition and is the object of study in Mera and Moran's paper mentioned above.

The methods tend to divide along the lines of {discrete/continuous}, {computations done in the state space/computations done from measurements of the state-space} and method of computation based on {Gram-Schmidt/QR-decomp./SVD}. The studies of convergence are either asymptotic and rigorous or empirical comparisons of various methods.

For other algorithms and studies of convergence etc. see the following fairly complete list of references [6, 10, 25, 33, 35, 36, 37, 46, 49, 50, 66, 70, 74, 78, 80, 93, 103, 104, 105, 108, 109, 112].

Convergence of Entropy Calculations

The question of "How fast can the cross-entropy rate be calculated?" is identical to the question, "What is the rates of convergence in the Generalized ergodic theorem of Breiman's?". We know of nothing along these lines for the convergence of cross entropy rates, but here is what's known for on the convergence of entropy estimates (where we are not concerned with the convergence of the random variables g_k found in theorem 5.5.5 above.)

We quote from page 165 of Paul Shields book "The ergodic theory of discrete sample paths" [86] (which we recommend as a reference):

> Many of the processes studied in classical probability theory, such as i.i.d processes, Markov processes, and finite-state processes have exponential rates of convergence for frequencies of all orders and for entropy, but some standard processes, such as renewal processes, do not have exponential rates. In the general ergodic case, no uniform rate is possible, that is, given any convergence rate for frequencies or for entropy, there is an ergodic process that does not satisfy the rate.

Exactly how these results translate in the case of cross-entropy estimation still needs to be studied.

5.6 Summarizing Cartoons, Questions and Synopsis

In view of the fact that this chapter has been rather long and involved we will now paint a *simplified*, summarizing picture.

In this chapter, we proposed and studied a measure of model fidelity based on the notion of *relative entropy rate*. More specifically, we did so for those models built on the basis of data generated by a known, underlying system. This will typically be the case in the testing phase of model building. That is, if we are testing a model building procedure, we will do this with data collected from a *known* system. We begin with a system and a set of measurements derived from that system (data). We build a model based on that data. Then we compute the loglikelihood and the positive Lyapunov exponents in order to obtain the final number, the *Gap*. See figure 5.6 for a visual summary.

What we wanted from our calculations was the relative entropy rate, $h(\mu||\theta)$. And in fact, under special conditions "The Gap" $= h(\mu||\theta)$. Under somewhat more general conditions this equivalence disappears and we retain only that the computed Gap is a lower bound to $h(\mu||\theta)$, i.e "The Gap" $< h(\mu||\theta)$. Some of the steps outlined in figure 5.6 can be augmented to show questions: see figure 5.7.

Under the most realistic conditions, the fundamental questions behind the questions shown in figure 5.7 involve the smoothness of the invariant measure and the relation of finite computations to their ideal goals.

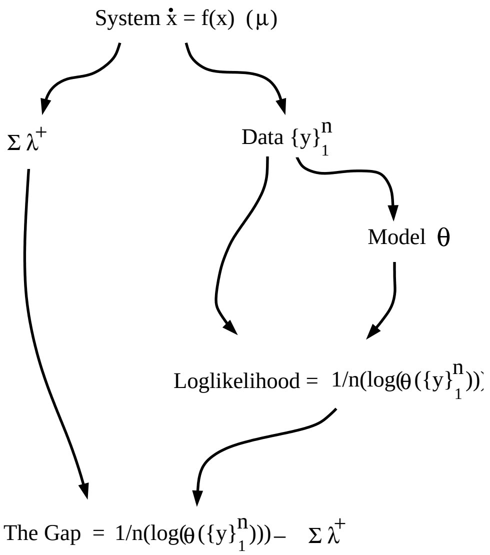

System $\dot{x} = f(x)$ (μ)
$\Sigma \lambda^{+}$
Data $\{y\}_1^n$
Model θ
Loglikelihood = $1/n(\log(\theta(\{y\}_1^n)))$
The Gap = $1/n(\log(\theta(\{y\}_1^n))) - \Sigma \lambda^{+}$

Figure 5.6: Illustrative summarizing cartoon: The dependencies and computed quantities.

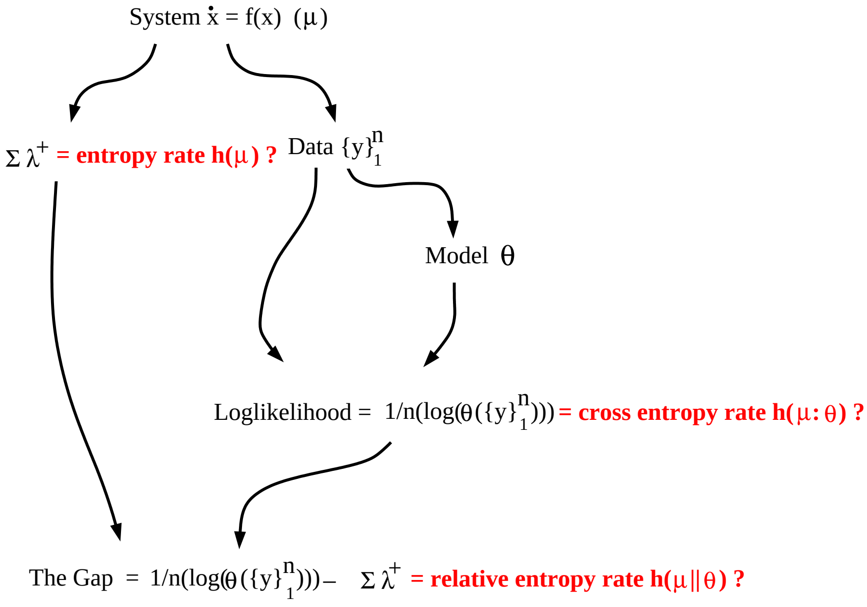

System $\dot{x} = f(x)$ (μ)
$\Sigma \lambda^{+}$ = entropy rate h(μ) ?
Data $\{y\}_1^n$
Model θ
Loglikelihood = $1/n(\log(\theta(\{y\}_1^n)))$ = cross entropy rate h(μ: θ) ?
The Gap = $1/n(\log(\theta(\{y\}_1^n))) - \Sigma \lambda^{+}$ = relative entropy rate h(μ||θ) ?

Figure 5.7: Illustrative summarizing cartoon with questions

Is μ an SRB measure ? Whether or not the measure is smooth along the unstable manifolds is the key question as demonstrated by the work of Leddrapier and Young.

Does acSRB=eSRB or =sSRB? If we can prove another version of "SRB" for our invariant measure μ say eSRB or sSRB, can we show that this implies that μ is in fact acSRB as needs to to be the case for our purposes?

Convergence? We have finite data and finite computing resources at our disposal. Always. So how close do we get to the asymptotic result? Very little is known about bounds on convergence rates.

Finite Precision Effects? Even if we can run an algorithm as long as desired, there is still the fact that we are doing the calculation on a finite precision model of the real number system. How does this effect our final answer? This is another question that is very difficult to answer in all but the simplest cases.

Does the SMB theorem hold for our model? Since we are using a generalized version of the Shannon-McMillan-Breiman theorem to get that the log-likelihood converges to the cross entropy rate, are we sure that the model process and underlying system satisfy the theorem's assumptions? Can we prove a more general theorem?

5.6.1 Synopsis

In the end, we are left with two question areas – one fairly narrow and the other broad – which must be considered in order to understand what our computations mean and what they might say about the relationship of our model to the underlying system.

Is μ smooth along the unstable directions? The question of whether or not a particular system has an SRB measure is a very difficult one. The result of this fact is that if we are using a system for which we are unsure of the smoothness of the invariant measure in the unstable direction, we may be under-estimating the measure of divergence between the model and system.

Do our calculations approximate the infinite? The question of rigorous convergence rates is also a very difficult one. Effort should be put into these questions. This is involved since we have the following dimensions to the computational question:

1 Does our system/model satisfy the assumptions of our computational methods? (e.g. the SMB theorem.)

2 Finite precision vs. Infinite precision.

3 Finite Data vs. Infinite Data.

4 Finite Compute time vs. Infinite Compute time.

5.7 Conclusions

In this paper we consider the use of entropy, cross-entropy, and the Lyapunov spectrum for the characterization of model fidelity. A gap G, previously pointed to in [31], was carefully defined and discussed. The use of Lyapunov exponents to obtain the KS entropy of the system generating the data necessitates consideration of SRB measures. Due to the variety of definitions of SRB and the level of technical details connected with them we took a careful look at the area. The result is that the G can be zero without this indicating that the model perfectly reproduces the system. We presented some indication of why this may be an acceptable. Nevertheless, if the gap is positive, this does indicate there is room for improvement. Further work might include:

1 A deeper look at conjecture 5.5.1 (This is the conjecture that Lyapunov exponents computed in finite precision are representative of some nearby system).

2 Another look at the Shannon-McMillan-Breiman theorem for cross-entropy (in particular, we think it is true for arbitrary stationary processes as long as we permit the limit to be $+\infty$).

3 Studies of convergence rates for the various computations above.

4 A look at how much the assumption that F is a C^2 diffeomorphism can be weakened and still retain the second Ledrappier-Young result which asserts that the entropy equals the "fractally" weighted sum of Lyapunov exponents.

Chapter 6

Afterword

What remains is to follow some of the questions opened up in the work presented in this dissertation. In particular, I am interested in the following questions, one or two of which I intend to explore in the near future.

From chapter 3: How sensitive is the test for aliasing based on sampling stationarity? Can we characterize the loss of sampling stationarity so as to be able to differentiate between 1) extremely under-sampled, 2) almost correctly sampled, and 3) correctly sampled signals? How does the test depend on the amount of data used to make the classification?

From chapter 3: Can we characterize, in some very practical way, the precise conditions a signal must satisfy in order for it to have *sampling stationarity*?

From chapter 4: Can the conjectured route for the nonlinear case (see sec. 4.6.3) be followed to a theorem as suggested?

From chapter 4: What sorts of results can one get under the addition of measurement noise?

From chapter 5: What are the weakest assumptions on F that still permit us to obtain theorems 5.3.2 and 5.3.3. (These are the theorems which tell us that the entropy equals the averaged, "fractally" weighted sum of positive Lyapunov exponents.)

From chapter 5: What result along the lines of Conjecture 5.5.1 can we prove? (This is the conjecture that, roughly, Lyapunov exponents computed in finite precision are representative of some nearby system.)

From chapter 5: Can we prove that a general stationary process θ satisfies the conditions assumed in the Shannon-McMillan-Breiman theorem for cross entropy (Theorem 5.5.5)? (With perhaps the relaxation of boundedness ... we would permit the cross entropy rate to be infinite.)

Bibliography

[1] Dirk Aeyels. Generic observability of differentiable systems. *SIAM Journal of Control and Optimization*, 19(5):595–603, september 1981.

[2] Paul H. Algoet and Thomas M. Cover. A sandwich proof of the Shannon-McMillan-Breiman theorem. *Ann. Probab.*, 16(2):899–909, 1988.

[3] José F. Alves, Christian Bonatti, and Marcelo Viana. SRB measures for partially hyperbolic systems whose central direction is mostly expanding. *Invent. Math.*, 140(2):351–398, 2000.

[4] José Ferreira Alves. SRB measures for non-hyperbolic systems with multidimensional expansion. *Ann. Sci. École Norm. Sup. (4)*, 33(1):1–32, 2000.

[5] M. Balde and P. Jouan. Genericity and observability of control-affine systems. *ESAIM: Control, Optimization, and Calculus of Variations*, 3:345–359, October 1998.

[6] György Barna and Ichiro Tsuda. A new method for computing Lyapunov exponents. *Phys. Lett. A*, 175(6):421–427, 1993.

[7] Michael Benedicks and Lai-Sang Young. Sinaĭ-Bowen-Ruelle measures for certain Hénon maps. *Invent. Math.*, 112(3):541–576, 1993.

[8] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15:9–20, 1980.

[9] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and

for hamiltonian systems; a method for computing all of them. part 2: Numerical application. *Meccanica*, 15:21–30, 1980.

[10] B. S. Berger and M. Rokni. Least squares approximation of Lyapunov exponents. *Quart. Appl. Math.*, 47(3):505–508, 1989.

[11] Jairo Bochi. Genericity of zero lyapunov exponents. Technical report, IMPA, Brazil, 2001. http://www.preprint.impa.br/Shadows/SERIE_A/2001/04.html.

[12] Christian Bonatti and Marcelo Viana. SRB measures for partially hyperbolic systems whose central direction is mostly contracting. *Israel J. Math.*, 115:157–193, 2000.

[13] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms.* Springer-Verlag, Berlin, 1975. Lecture Notes in Mathematics, Vol. 470.

[14] Rufus Bowen and David Ruelle. The ergodic theory of Axiom A flows. *Invent. Math.*, 29(3):181–202, 1975.

[15] Leo Breiman. The individual ergodic theorem of information theory. *Annals of Mathematical Statistics*, 28(3):809–811, September 1957.

[16] Leo Breiman. Correction to "The individual ergodic theorem of information theory". *Annals of Mathematical Statistics*, 31(3):809–810, September 1960.

[17] Leo Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics, 1992. ISBN 0-89871-296-3.

[18] David R. Brillinger and Murray Rosenblatt. Computation and interpretation of k-th order spectra. In *Spectral Analysis Time Series (Proc. Advanced Sem., Madison, Wis., 1966)*, pages 189–232. John Wiley, NEw York, 1967.

[19] Jérôme Buzzi. Absolutely continuous S.R.B. measures for random Lasota-Yorke maps. *Trans. Amer. Math. Soc.*, 352(7):3289–3303, 2000.

[20] Eleonora Catsigeras and Heber Enrich. SRB measures of certain almost hyperbolic diffeomorphisms with a tangency. *Discrete Contin. Dynam. Systems*, 7(1):177–202, 2001.

[21] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507, 1952.

[22] N. Chernov. Statistical properties of piecewise smooth hyperbolic systems in high dimensions. *Discrete Contin. Dynam. Systems*, 5(2):425–448, 1999.

[23] N. Chernov, R. Markarian, and S. Troubetzkoy. Conditionally invariant measures for Anosov maps with small holes. *Ergodic Theory Dynam. Systems*, 18(5):1049–1073, 1998.

[24] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 1991.

[25] A. G. Darbyshire and D. S. Broomhead. Robust estimation of tangent maps and Liapunov spectra. *Phys. D*, 89(3-4):287–305, 1996.

[26] M. Denker and S. Rohde. On Hausdorff measures and SBR measures for parabolic rational maps. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 9(9):1763–1769, 1999. Discrete dynamical systems.

[27] J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto. Liapunov exponents from time series. *Phys. Rev. A (3)*, 34(6):4971–4979, 1986.

[28] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Modern Phys.*, 57(3, part 1):617–656, 1985.

[29] Heber Enrich. A heteroclinic bifurcation of Anosov diffeomorphisms. *Ergodic Theory Dynam. Systems*, 18(3):567–608, 1998.

[30] Gerald B. Folland. *Real Analysis: Modern Techniques and their Applications.* Wiley-Interscience, 1999. ISBN 0-471-31716-0.

[31] Andrew M. Fraser. Chaos and detection. *Physical Review E*, 53(5):4514–4523, May 1996.

[32] G. Frazer, A. Reilly, and B. Boashash. The bispectral aliasing test. In *Proceedings IEEE Workshop on Higher-Order Statistics*, pages 332–335, June 1993.

[33] Gary Froyland, Kevin Judd, and Alistair I. Mees. Estimation of Lyapunov exponents of dynamical systems using a spatial average. *Phys. Rev. E (3)*, 51(4, part A):2844–2855, 1995.

[34] Claude Gasquet and Patrick Witomski. *Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets.*, volume 30 of *Texts in Applied Mathematics.* Springer-Verlag, 1999.

[35] Karlheinz Geist, Ulrich Parlitz, and Werner Lauterborn. Comparison of different methods for computing Lyapunov exponents. *Progr. Theoret. Phys.*, 83(5):875–893, 1990.

[36] Ramazan Gencay and W. Davis Dechert. An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system. *Phys. D*, 59(1-3):142–157, 1992.

[37] Y. Guivarc'h and A. Raugi. Products of random matrices: convergence theorems. In *Random matrices and their applications (Brunswick, Maine, 1984)*, pages 31–54. Amer. Math. Soc., Providence, RI, 1986.

[38] Melvin J. Hinich and Hagit Messer. On the principal domain of the discrete bispectrum of a stationary signal. *IEEE Transactions on Signal Processing*, 43(9):2130–2134, 1995.

[39] Melvin J. Hinich and Murray A. Wolinsky. A test for aliasing using bispectral analysis. *Journal of the American Statistical Association*, 83:499–502, June 1988.

[40] Morris W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics.* Springer, 1976. corrected sixth printing, 1997.

[41] Hu Yi Hu and Lai-Sang Young. Nonexistence of SBR measures for some diffeomorphisms that are "almost Anosov". *Ergodic Theory Dynam. Systems*, 15(1):67–76, 1995.

[42] Huyi Hu. Conditions for the existence of SBR measures for "almost Anosov" diffeomorphisms. *Trans. Amer. Math. Soc.*, 352(5):2331–2367, 2000.

[43] Brian R. Hunt. Estimating invariant measures and Lyapunov exponents. *Ergodic Theory Dynam. Systems*, 16(4):735–749, 1996.

[44] Michael Jakobson and Sheldon Newhouse. On the structure of non-hyperbolic attractors. In *Dynamical systems and chaos, Vol. 1 (Hachioji, 1994)*, pages 103–111. World Sci. Publishing, River Edge, NJ, 1995.

[45] Michael Jakobson and Sheldon Newhouse. Asymptotic measures for hyperbolic piecewise smooth mappings of a rectangle. *Astérisque*, (261):xii, 103–159, 2000. Géométrie complexe et systèmes dynamiques (Orsay, 1995).

[46] T. M. Janaki and Govindan Rangarajan. Lyapunov exponents for continuous-time dynamical systems. *J. Indian Inst. Sci.*, 78(4):267–274, 1998.

[47] Esa Järvenpää. A SRB-measure for globally coupled circle maps. *Nonlinearity*, 10(6):1435–1469, 1997.

[48] J.R.Higgins. *Sampling Theory in Fourier and Signal Analysis.* Oxford University Press, 1996.

[49] James B. Kadtke, Jeffrey Brush, and Joachim Holzfuss. Global dynamical equations and Lyapunov exponents from noisy chaotic time series. In *Proceedings of the Second Workshop on Measures of Complexity and Chaos (Bryn Mawr, PA, 1992)*, volume 3, pages 607–616, 1993.

[50] G. Karch and Walter V. Wedig. Computational methods for Lyapunov exponents and invariant measures. In *Nonlinear dynamics and stochastic mechanics*, pages 463–477. CRC, Boca Raton, FL, 1995.

[51] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems.* Cambridge, 1995.

[52] Yuri I. Kifer. Small random perturbations of certain smooth dynamical systems. *Math. USSR-Izv.*, 8:1083–1108, 1975.

[53] J.W. Ioup L.A. Pflug, G.E. Ioup and R.L. Field. Properties of higher-order correlations and spectra for bandlimited, deterministic transients. *J. Acoustic Soc. Am.*, 91(2):975–988, Feb. 1992.

[54] F. Ledrappier and L.-S. Young. The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin's entropy formula. *Ann. of Math. (2)*, 122(3):509–539, 1985.

[55] F. Ledrappier and L.-S. Young. The metric entropy of diffeomorphisms. II. Relations between entropy, exponents and dimension. *Ann. of Math. (2)*, 122(3):540–574, 1985.

[56] Weigu Li and Meirong Zhang. Existence of SRB measures for expanding maps with weak regularity. *Far East J. Dyn. Syst.*, 2:75–97, 2000.

[57] Pei-Dong Liu. Random perturbations of Axiom A basic sets. *J. Statist. Phys.*, 90(1-2):467–490, 1998.

[58] Pei-Dong Liu and Min Qian. *Smooth ergodic theory of random dynamical systems.* Springer-Verlag, Berlin, 1995.

[59] Pei-Dong Liu and Hong-Wen Zheng. Stochastic stability of generalized SRB measures of Axiom A basic sets. *Proc. Amer. Math. Soc.*, 128(12):3541–3545 (electronic), 2000.

[60] Youming Liu. A distributional sampling theorem. *SIAM Journal of Mathematical Analysis*, 27(4):1153–1157, July 1996.

[61] Ricardo Mañé. Oseledec's theorem from the generic viewpoint. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 1259–1276, Warszawa, Poland, August 1983.

[62] Ricardo Mañé. *Ergodic theory and differentiable dynamics.* Springer-Verlag, Berlin, 1987. Translated from the Portuguese by Silvio Levy.

[63] Ricardo Mañé. Lyapunov exponents of generic area preserving diffeomorphisms. In F. Ledrappier, J. Lewowicz, and S. Newhouse, editors, *Proceddings of the International confernece on dynamical systems*, volume 362 of *Pitman Research Notes in Mathematics*, pages 110–119, Montevideo, Uruguay, 1995. Longman.

[64] M. Eugenia Mera and Manuel Morán. Convergence of the Eckmann and Ruelle algorithm for the estimation of Liapunov exponents. *Ergodic Theory Dynam. Systems*, 20(2):531–546, 2000.

[65] John W. Milnor. *Topology From The Differentiable Viewpoint.* University Press of Virginia, Charlottesville, 1965. Eighth printing, 1990.

[66] Gunter Ochs. Stability of Oseledets spaces is equivalent to stability of Lyapunov exponents. *Dynam. Stability Systems*, 14(2):183–201, 1999.

[67] Ja. B. Pesin. Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat. Nauk*, 32(4 (196)):55–112, 287, 1977.

[68] Jan Willem Polderman and Jan C. Willems. *Introduction to Mathematical Systems Theory: A Behavioral Approach*, volume 26 of *Texts in Applied Mathematics.* Springer, 1998.

[69] Charles Pugh and Michael Shub. Ergodic attractors. *Transactions of the American Mathematical Society*, 312(1):1–54, 1989.

[70] K. Ramasubramanian and M. S. Sriram. A comparative study of computation of Lyapunov spectra with different algorithms. *Phys. D*, 139(1-2):72–86, 2000.

[71] Jorma Rissanen. Universal coding, information, prediction, and estimation. *IEEE Trans. Inform. Theory*, 30(4):629–636, 1984.

[72] V. A. Rohlin. On the fundamental ideas of measure theory. *Amer. Math. Soc. Translation*, 1952(71):55, 1952.

[73] V. A. Rohlin. Lectures on the entropy theory of transformations with invariant measure. *Uspehi Mat. Nauk*, 22(5 (137)):3–56, 1967. Translated in Russian Mathematical Surveys.

[74] Michael T. Rosenstein, James J. Collins, and Carlo J. De Luca. A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D*, 65(1-2):117–134, 1993.

[75] David Ruelle. A measure associated with axiom-A attractors. *Amer. J. Math.*, 98(3):619–654, 1976.

[76] David Ruelle. An inequality for the entropy of differentiable maps. *Bol. Soc. Brasil. Mat.*, 9(1):83–87, 1978.

[77] David Ruelle. Differentiation of SRB states. *Comm. Math. Phys.*, 187(1):227–241, 1997.

[78] M. Sano and Y. Sawada. Measurement of the Lyapunov spectrum from a chaotic time series. *Phys. Rev. Lett.*, 55(10):1082–1085, 1985.

[79] E. A. Sataev. Ergodic properties of the Belykh map. *J. Math. Sci. (New York)*, 95(5):2564–2575, 1999. Dynamical systems. 7.

[80] Shinichi Sato, Masaki Sano, and Yasuji Sawada. Practical methods of measuring the generalized dimension and the largest Lyapunov exponent in high-dimensional chaotic systems. *Progr. Theoret. Phys.*, 77(1):1–5, 1987.

[81] Tim Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3/4):579–616, 1991.

[82] Timothy D. Sauer. Trajectory averages in dynamical systems. preprint, May 2001. Presented at the 2001 SIAM dynamical systems meeting.

[83] I. Scharfer and Hagit Messer. The bispectrum of sampled data: Part 1 - detection of the sampling jitter. *IEEE Trans. Sig. Proc.*, 41:296–312, Jan. 1993.

[84] J. Schmeling. A dimension formula for endomorphisms—the Belykh family. *Ergodic Theory Dynam. Systems*, 18(5):1283–1309, 1998.

[85] J. Schmeling and S. Troubetzkoy. Dimension and invertibility of hyperbolic endomorphisms with singularities. *Ergodic Theory Dynam. Systems*, 18(5):1257–1282, 1998.

[86] Paul C. Shields. *The ergodic theory of discrete sample paths.* American Mathematical Society, Providence, RI, 1996.

[87] Paul C. Shields. The interactions between ergodic theory and information theory. *IEEE Trans. Inform. Theory*, 44(6):2079–2093, 1998. Information theory: 1948–1998.

[88] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[89] Ya. G. Sinai. *Introduction to Ergodic Theory.* Mathematical Notes. Princeton University Press, 1976.

[90] K. T. Smith, D.C. Solomon, and S. L. Wagner. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bulletin of the American Mathematical Society*, 83:1227–70, 1977.

[91] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, volume 6 of *Texts in Applied Mathematics.* Springer, 2nd edition, 1998.

[92] Jaroslav Stark. Delay embeddings for forced systems. I. deterministic forcings. *Journal of Nonlinear Science*, 9:255–332, 1999.

[93] David E. Stewart. A new algorithm for the SVD of a long product of matrices and the stability of products. *Electron. Trans. Numer. Anal.*, 5(June):29–47 (electronic), 1997.

[94] Naoya Sumi. Diffeomorphisms approximated by Anosov on the 2-torus and their SBR measures. *Trans. Amer. Math. Soc.*, 351(8):3373–3385, 1999.

[95] A. Swami. Pitfalls in polyspectra. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume IV, pages 97–100, 1993.

[96] Floris Takens. Detecting strange attractors in turbulence. In *Dynamic Days at Warwick*, number 898 in Lectue Notes in Mathematics. Springer Verlag, 1981.

[97] Warwick Tucker. The Lorenz attractor exists. *C. R. Acad. Sci. Paris Sér. I Math.*, 328(12):1197–1202, 1999.

[98] Raúl Ures. Hénon attractors: SBR measures and Dirac measures for sinks. In *International Conference on Dynamical Systems (Montevideo, 1995)*, pages 214–219. Longman, Harlow, 1996.

[99] Marcelo Viana. Stochastic dynamics of deterministic systems. In *21st Brazillian Math. Colloquium*, number 21 in Brazillian Math. Colloquium. IMPA, 1997. http://www.impa.br/ viana/#downloads.

[100] Kevin R. Vixie and Gary L. Sandine. Reconstruction from projections using dynamics: Non-stochastic case. *LANL e-print archive, http://arXiv.org*, 2001. Link to paper http://arXiv.org/abs/math.DS/0101037.

[101] Kevin R. Vixie, David E. Sigeti, and Murray Wolinsky. Detection of aliasing in persistent signals. Technical report, Los Alamos National Laboratory, 1999. http://xxx.lanl.gov/abs/chao-dyn/9905021.

[102] Kevin R. Vixie, Murray Wolinsky, and David Sigeti. The bispectral aliasing test: A clarification and some key examples. In M. Deriche, B. Boashash, and W. W. Boles, editors, *Proceedings of the Fifth International Symposium on Signal Processing and its Applications (ISSPA'99)*, Brisbane, Australia, 1999. Signal Processing Research Centre, Queensland University of Technology. http://xxx.lanl.gov/abs/chao-dyn/9906020.

[103] Hubertus F. von Bremen, Firdaus E. Udwadia, and Wlodek Proskurowski. An efficient QR based method for the computation of Lyapunov exponents. *Phys. D*, 101(1-2):1–16, 1997.

[104] W. V. Wedig. Lyapunov exponents and invariant measures of dynamic systems. In *Bifurcation and chaos: analysis, algorithms, applications (Würzburg, 1990)*, pages 367–376. Birkhäuser, Basel, 1991.

[105] A. Wolf and J. A. Vastano. Intermediate length scale effects in Lyapunov exponent estimation. In *Dimensions and entropies in chaotic systems (Pecos River Ranch, N.M., 1985)*, pages 94–99. Springer, Berlin, 1986.

[106] Alan Wolf, Jack B. Swift, Harry L. Swinney, and John A. Vastano. Determining Lyapunov exponents from a time series. *Phys. D*, 16(3):285–317, 1985.

[107] Murray Wolinsky. Invitation to the bispectrum. Technical report, University of Texas: Applied Research Laboratories, 1988.

[108] Dachuan Xiao and Qing He. A novel algorithm for calculating the smallest Liapunov exponent. *J. Franklin Inst. B*, 334(4):653–658, 1997.

[109] Shao Qing Yang, Xin Hua Zhang, and Chang An Zhao. A robust method for estimating the largest Lyapunov exponent. *Acta Phys. Sinica*, 49(4):636–640, 2000.

[110] Lai-Sang Young. Stochastic stability of hyperbolic attractors. *Ergodic Theory and Dynamical Systems*, 6(2):311–319, 1986.

[111] Lai-Sang Young. Ergodic theory of differentiable dynamical systems. In *Real and complex dynamical systems (Hillerød, 1993)*, pages 293–336. Kluwer Acad. Publ., Dordrecht, 1995. see http://www.cims.nyu.edu/˜lsy/expository.html.

[112] X. Zeng, R. A. Pielke, and R. Eykholt. Extracting Lyapunov exponents from short time series of low precision. *Modern Phys. Lett. B*, 6(2):55–75, 1992.