## CLARKSON UNIVERSITY

# Learning Features of Dynamical Systems by Data Driven Analysis Methods 

A Dissertation by<br>Sudam Surasinghe<br>Department of Mathematics<br>Clarkson Center for Complex Systems Science

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Mathematics

April 28, 2022

The committee below have examined the thesis dissertation entitled "Learning Features of Dynamical Systems by Data Driven Analysis Methods" presented by Sudam Surasinghe, a candidate for the degree of Doctor of Philosophy (Mathematics), and have certified that it is worthy of acceptance.

4126/2022
Date


## Abstract

We study the data-driven analysis methods for dynamical systems and focus on the theories in machine learning, manifold learning, and data sciences. Learning causation and learning global dynamics are two major field we will address in this project.

Causal inference is a fundamental question in all science and the most popular methods in this area are probabilistic in nature, including the Nobel prize-winning work on Granger-causality, and also the recently highly popular transfer entropy. In this thesis we propose a new measure to identify causal variables using a geometric approach and connect transfer entropy with the new measure. Beyond the conceptual advancement, a geometric description of causality further allows for new and efficient computational methods of causal inference.

Trending tool for learning dynamics, Dynamic Mode decomposition (DMD) can be used to separate spatial and temporal components of time series data. The DMD approximates the linear Koopman operator on a projected space. In the spirit of Johnson-Lindenstrauss lemma, we will use a random projection to estimate the DMD modes in a reduced dimensional space. In practical applications, snapshots are in a high-dimensional observable space and the DMD operator matrix is massive. Hence, computing DMD with the full spectrum is expensive, so our main computational goal is to estimate the eigenvalue and eigenvectors of the DMD op-
erator in a projected domain. We generalize the current algorithm to estimate a projected DMD operator. We focus on a powerful and simple random projection algorithm that will reduce the computational and storage costs. While, clearly, a random projection simplifies the algorithmic complexity of a detailed optimal projection, as we will show, the results can generally be excellent, nonetheless, and the quality could be understood through a well-developed theory of random projections. We will demonstrate that modes could be calculated for a low cost by the projected data with sufficient dimension. This simple and computationally efficient random projection improves aspects of the DMD, the Extended DMD (EDMD), and also kernelized DMD approximations of the Koopman operator.

## Acknowledgements

I am deeply indebted to my advisor, Prof. Erik Bollt. His valuable advice, guidance, suggestions, experience, wisdom, and support have contributed greatly to my academic development. His great mentorship over the past six years has helped me improve as a mathematician and achieve my career goals. I could not have hoped for a better advisor.

I would like to thank Ioannis Kevrekidis for his generous time, feedback, and interest in the "geometric causal inference" project. I also would like to thank Dr. Marko Budišić for discussions about "Gibraltar" Project and for encouragement. I would like to thank oceanographers, Dr. Larry Pratt and Dr. Jose Sanchez-Garrido for sharing data and valuable information to apply our theories in the field of oceanography. I would like to express my deepest appreciation to the members of my thesis committee, Professors Erik Bollt, Marko Budišić, Prashant Athavale, Brian Helenbrook, and Jianhua Zhang for their helpful career advice and suggestions in general. I'm extremely grateful to my friends, Sulalitha Priyankara and Katie Bakewell for helping me to stabilize in the USA.

Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents Swarna Surasinghe and Indra Malani, whose love and guidance are with me in whatever I pursue. They are the ultimate role models. I would like to thank my brother, Madusha Surasinghe,
and parent-in-law, Titus Pathirana and Mangalika Pathirana for their encouragement. I also would like to thank my daughter, Buddhara Surasinghe for understanding my busy life. Most importantly, I cannot begin to express my thanks to my loving and supportive wife, Raji Pathirana who provides unending inspiration.

## Contents

1 Introduction ..... 1
1.1 The Historical Background ..... 2
1.1.1 Information Flow ..... 2
1.1.2 Approximating Koopman Operator ..... 3
1.2 Research Problems and Contributions ..... 5
1.2.1 Geometry of Information Flow ..... 5
1.2.2 Randomized Dynamic Mode Decomposition ..... 7
2 Transfer Entropy ..... 17
2.1 Quantifying Information Flows ..... 17
2.2 Transfer Entropy for Deterministic Systems ..... 20
2.2.1 Calculating Conditional differential Entropy ..... 22
2.2.2 Numerical Demonstration of Boundedness Issue ..... 26
3 Relating Geometric Measures to Causal Inference ..... 38
3.1 Geometric Interpretation of Causal Inference ..... 38
3.2 Relating Transfer Entropy to a Geometric Bound ..... 41
4 Geometry of Information Flow ..... 55
4.1 Relating the Information Flow as Geometric Orientation of Data ..... 56
4.1.1 Measure Causality by Correlation Dimension ..... 57
4.2 Results and Discussion ..... 59
4.2.1 $\quad$ Synthetic Data: $X^{\prime}=g(X)+b Y+C$ ..... 60
4.2.2 Synthetic Data: $X^{\prime}=a X+b Y$ with $a \neq 0$ ..... 62
4.2.3 Synthetic Data: Nonlinear Cases ..... 63
4.2.4 Application Data ..... 65
5 Dynamic Mode Decomposition on Projected Space ..... 75
5.1 Koopman Operator ..... 76
5.2 Dynamic Mode Decomposition ..... 79
5.2.1 DMD on Projected Space ..... 81
6 Randomized Projection Learning Method for Dynamic Mode Decompo-
sition ..... 94
6.1 Randomized Dynamic Mode Decomposition ..... 94
6.1.1 Random Projection ..... 95
6.1.2 Random Projection for DMD ..... 96
6.2 Results and Discussion ..... 98
6.2.1 Logistic Map ..... 98
6.2.2 Toy Example: Demonstrates the Variable Separation and Iso-lating Dynamics100
6.2.3 Gulf of Mexico ..... 104
7 Conclusion and Future Problems ..... 117
7.1 Geometric Causal Inference ..... 117
7.2 Randomized Dynamic Mode Decomposition ..... 119
7.3 Future Work ..... 121
7.3.1 Causal inference with multiple time series . . . . . . . . . . . 121
7.3.2 Learning Transfer Operators by Kernel Density Estimation. . 122
7.3.3 Estimating Koopman operator with Random Projection. . . . 123

Appendices 133

A On the Asymmetric Spaces Transfer Operators 134

## Chapter 1

## Introduction

Data science is currently the most trending topic in academia. Traditionally, data from many disciplines have been obtained and used to verify a given hypothesis. The advancement of technology makes data abundant and abundantly collected in every single experiment at a low-cost [33]. Insight from this data can be obtained beyond the verification of the hypothesis. It can be used to learn underlying features and behaviors.

Mining special features from large-scale time series data is a vital part of current research because it enhances the short-term and long-term predictions about the system. Research in feature learning of time-series data includes the fast-growing information theory and Koopman theory. However, there is an overall lack of research regarding the geometric aspect of the information flow. Also, there are gaps in efficient and accurate models to approximate Koopman modes. This research aims to address two main parts. First, we build geometric tools to measure the information flow (which is discussed in our article [48] ). Next, we discuss the random projection-based methods to efficiently and accurately approximate the Koopman modes (which is discussed in our article [49] ). This chapter will pro-
vide an introduction to the study by first discussing the background and context, followed by the research problem, and contribution.

### 1.1 The Historical Background

### 1.1.1 Information Flow

Causation Inference is perhaps one of the most fundamental concepts in science, underlying questions such as "what are the causes of changes in observed variables". Identifying, indeed even defining causal variables of a given observed variable is not an easy task, and these questions date back to the Greeks [15, [54]. This includes important contributions from more recent luminaries such as Russel [40], and from philosophy, mathematics, probability, information theory, and computer science. We have written that [9], "a basic question when defining the concept of information flow is to contrast versions of reality for a dynamical system. Either a subcomponent is closed or alternatively, there is an outside influence due to another component".

Claude Granger's Nobel prize [20] winning work leading to Granger Causality (see also Wiener [53]) formulates causal inference as a concept of quality of forecasts. That is, we ask, does system $X$ provide sufficient information regarding forecasts of future states of the system $X$, or are there improved forecasts with observations from system $Y$ ? We declare that $X$ is not closed, as it is receiving influence (or information) from system $Y$ when data from $Y$ improve forecasts of $X$. Such a reduction of uncertainty perspective of causal inference is not identical to the interventionists' concept of allowing perturbations and experiments to decide what changes indicate influences. This data-oriented philosophy of causal inference is
especially appropriate when (1) the system is a dynamical system of some form producing data streams in time, and (2) a score of influence may be needed. In particular, contrasting forecasts is the defining concept underlying Granger Causality (G-causality), and it is closely related to the concept of information flow as defined by transfer entropy [6, 43], which can be proved as a nonlinear version of Granger's otherwise linear (ARMA) test [4]. In this spirit, we find methods such as Convergent Cross-Mapping method (CCM) [45], and causation entropy (CSE) [47] to disambiguate direct versus indirect influences [2, 8, 25, 31, 39, 44, 46, 47].

On the other hand, closely related to information flow are concepts of counterfactuals: "what would happen if ..." [19] that are foundational questions for another school leading to the highly successful Pearl "Do-Calculus" built on a specialized variation of Bayesian analysis [35]. These are especially relevant for nondynamical questions (inputs and outputs occur once across populations), such as a typical question of the sort, "why did I get fat" may be premised on inferring probabilities of removing influences of saturated fats and chocolates. However, with concepts of counterfactual analysis in mind, one may argue that Granger is less descriptive of causation inference, but rather more descriptive of information flow. In fact, there is a link between the two notions for so-called "settable" systems under a conditional form of exogeneity [51, 52]. This thesis focuses on the information flow perspective, which is causation as it relates to G-causality.

### 1.1.2 Approximating Koopman Operator

Koopman operator [26] theory provides great utility in data-driven analysis and high dimensional non-linear dynamical systems. Hence, the Koopman theory is becoming a popular alternative formalism for the dynamical systems [3]. Koopman operator is linear in the infinite-dimensional space of observables. This oper-
ator can be used to decompose the spatial and temporal components of the system. Numerical methods known as dynamic mode decomposition (DMD), extended DMD, and kernel DMD, approximate the Koopman operator on projected function space.

The DMD algorithm is introduced by Schmid, P. J. (2010) and is currently known as the standard dynamic mode decomposition. He introduces this method in the field of fluid dynamics to exact dynamic information from flow fields that are either generated from numerical simulation or physical experiments. It has become popular among the community of fluid mechanics because it is similar to the Proper Orthogonal Decomposition (POD) methods. POD and DMD methods are data-driven equation-free methods and in contrast to the POD algorithm, DMD methods can isolate the spatial features related to specific frequencies. The DMD thrived after realizing its connection[[] mazic, rowly] to the Koopman theory. Tu et. al.(2014) improve the DMD algorithm to a larger class of datasets, including non-sequential time series.

The gap between the Koopman spectrum and the DMD spectrum is filled by Williams et. al.(2015) extending the DMD into a larger function space. He later introduces the kernel method which is a computationally feasible alternative to EDMD. The Kernel method uses sets of scalar observables that are implicitly defined by Mercer's kernel. They also proved EDMD and KDMD approximates the Koopman eigendecomposition. Depending on the chosen kernel, the dimension of the observable space can be increased without increasing the calculation or storage complexity.

### 1.2 Research Problems and Contributions

### 1.2.1 Geometry of Information Flow

The traditional narrative of information flow is in terms of comparing stochastic or deterministic dynamical systems in probabilistic terms. However, highly popular probabilistic method transfer entropy suffer from boundedness issue when data arise from a deterministic system (see section ??). Even though the geometry is perhaps a natural place to describe a dynamical system, there is a huge lack of research in the geometric interpretation of the information flow. Thus, we will try to provide a bridge between concepts of causality as information flow to the underlying geometry. This connection will empower to boost the accessible data analysis tools in the field of information theory. For example, manifold learning methods such as diffusion maps can be used to identify the underlying geometry of the data which may be used to find the causal inference. This thesis offers a unifying description for interpreting geometric formulations of causation together with traditional statistical or information-theoretic interpretations and develops a new measure for the information flow.

We analyze connections between information flow by transfer entropy to geometric quantities that describe the orientation of underlying functions of a corresponding dynamical system. In the course of this analysis, we have needed to develop a new "asymmetric transfer operator" (asymmetric Frobenius-Perron operator) evolving ensemble densities of initial conditions between spaces whose dimensionalities do not match. With this, we proceed to give a new exact formula for transfer entropy, and from there we are able to relate this Kullback-

Leibler divergence-based measure directly to other more geometrically relevant divergences, specifically total variation divergence and Hellinger divergence, by Pinsker's inequality. This leads to a succinct upper bound of the transfer entropy by quantities related to a more geometric description of the underlying dynamical system. Then, we present a new measure in the spirit of G-causality that is more directly motivated by geometry. This measure is developed in terms of the classical fractal dimension concept of correlation dimension.

The main connections and corresponding sections of this part of the work are summarized as a dichotomy: Geometry and Causation (information flow structure) as described in Figure (1.1). Our contribution in this paper is as follows:

- We present analytical and data-driven approach to identify causality by geometric methods, and thus also a unifying perspective.
- We show that a derivative (if it exists) of the underlining function of the time series has a close relationship to the transfer entropy (Section 3.2).
- Correlation dimension can be used as a measurement for dynamics of a dynamical system. We will show that this measurement can be used to identify the causality (Section 4).
- We provide a new tool called geoC to identify the causality by geometric terms (Section 4).


Figure 1.1: The relationship of causation and geometry.

### 1.2.2 Randomized Dynamic Mode Decomposition

Since the Koopman operator is an infinite-dimensional operator, current research is mostly focused on lifting up the dimension of the observable space. However, due to the limited resources, the final calculation needs to be achieved in a lowdimensional space. Even though DMD methods are implicitly projecting the observable space, there is a lack of generalized explicit formalism regarding this issue. In addition, standard existing DMD algorithms reliance on a singular value decomposition (SVD) of the snapshot matrix. However, direct SVD calculations of such matrices can be quite memory-consuming and computationally intensive. One can notice that SVD calculations of snapshot matrices in big projects would require the use of supercomputers or days of computation.

This thesis will discuss the explicit generalization of the projection of observable space. To allow processing on small-scale computers and in a shorter time frame, we propose developing an algorithm based on a randomized projection [13], which is used to reduce the dimension of observable space in DMD, which we call rDMD. In order to utilize and carefully analyzed the rDMD, we will use the Johnson-Lindenstrauss lemma. It is clear that a random projection is simple as compared to a detailed optimal projection method, but our analysis and examples demonstrate, nonetheless, the quality and efficiency.

Strong theoretical support from Johnson-Lindenstrauss(JL) lemma [24] makes the random projection method reliable and has extensive utilization in the field of data science. According to the JL lemma, if data points lie in a sufficiently highdimensional space, then those data points may be projected into a sufficiently lowdimensional space while approximately preserving the distance of the data points. Furthermore, the projection can be done by a random matrix, which makes al-
gorithms based on the JL lemma both past and simple. Hence, this tool is more powerful and adopted heavily in data science.

JL lemma-based random projection and SVD-based projection can be used to project $N$ dimensional data into a lower dimension $L \ll N$. Data matrix $X_{N \times M}$ can be projected (by random projection) into a lower dimension ( $L$ ) subspace as $X_{L}:=R X$, where $R$ is a random matrix with unit length. Hence, the random projection is very simple because it relies only on matrix multiplication. Moreover, computational complexity is $\mathcal{O}(M L N)$, while SVD has computational complexity $\mathcal{O}\left(N M^{2}\right)$ when $M<N$ [5]. We will use the random projection to project highdimensional snapshot matrices into a manageable low-dimensional space. From a theoretical perspective, the dimensions of the input and output spaces in the Koopman operator can be reduced by the random projection method; thus, reducing the storage and computational cost of the DMD algorithm.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309. DOI: $10.1063 / 1.5031109$, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https ://www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138. URL: https://link.aps.org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL: https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079. eprint: https://science.sciencemag. org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag. org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi . org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 2

## Transfer Entropy

This chapter will discuss a currently famous probabilistic method known as transfer entropy for measuring causality. First, we will introduce the transfer entropy as a measure of difference between two probability distribution. Then we will focus on information flow between deterministic dynamical systems. Boundedness issues of calculating transfer entropy arise with these deterministic Systems will also be discussed. Finally, we will provide example results to demonstrate the issue. This boundedness issue is a one of the major motivation for considering geometric measures. We will adopt standard notations in the probability theory in this part of the discussion. Also, we will use lower case letter $(x)$ as a value of its upper case letter $(X)$ random variable.

### 2.1 Quantifying Information Flows

For now, we assume that $x, y$ are real valued scalars. We use a shorthand notation, $x:=x_{n}, x^{\prime}:=x_{n+1}$ for any particular time $n$, where the prime ( ${ }^{\prime}$ ) notation denotes "next iterate". Likewise, let $z=(x, y)$ denote the composite variable, and its future
composite state, $z^{\prime}$. Consider the simplest of cases, where there are two coupled dynamical systems written as discrete time maps,

$$
\begin{align*}
x^{\prime} & =f_{1}(x, y)  \tag{2.1}\\
y^{\prime} & =f_{2}(x, y) \tag{2.2}
\end{align*}
$$

Identifying the causal effect of $y$ on $x$ hinges on the contrast between two alternative versions of the possible origins of $x^{\prime}$ and is premised on deciding one of the following two cases: Either

$$
\begin{equation*}
x^{\prime}=f_{1}(x), \quad \text { or } \quad x^{\prime}=f_{1}(x, y) \tag{2.3}
\end{equation*}
$$

is descriptive of the actual function $f_{1}$. The definition of transfer entropy [6, 7, 43] $T_{y \rightarrow x}$ is defined to decide this question by comparing the deviation from a proposed Markov property,

$$
\begin{equation*}
p\left(x^{\prime} \mid x\right) \stackrel{?}{=} p\left(x^{\prime} \mid x, y\right) \tag{2.4}
\end{equation*}
$$

The transfer entropy, measuring the influence of coupling from variables $y$ onto the future of the variables $x$, denoted by $x^{\prime}$ is defined as Kullback-Leibler (KL) divergence of conditional probability $p\left(x^{\prime} \mid x\right)$ from $p\left(x^{\prime} \mid x, y\right)$. Assuming the outcome spaces may be continuous, the differential entropy version of a KullbackLeibler divergence definition for transfer entropy can be stated. For this purposes, Kullback-Leibler divergence can be stated as in definition 2.1.1.

Definition 2.1.1 (Kullback-Leibler Divergence). Let outcome space $\Omega$ have a measure $\mu$ with probability measures $P_{1}$ and $P_{2}$ are absolutely continuous to $\mu$, so that $p_{1}=\frac{d P_{1}}{d \mu}$ and $p_{2}=\frac{d P_{2}}{d \mu}$, then Kullback-Leibler divergence of $P_{2}$ from $P_{1}$ is defined
by

$$
\begin{aligned}
D_{K L}\left(P_{1} \| P_{2}\right) & =\int_{\Omega} p_{1} \log \frac{p_{1}}{p_{2}} d \mu \\
& =\int_{\Omega} p_{1} \log p_{1} d \mu-\int_{\Omega} p_{1} \log p_{2} d \mu \\
& =-h\left(P_{1}\right)-\int_{\Omega} p_{1} \log p_{2} d \mu
\end{aligned}
$$

where $h\left(p_{1}\right)=-\int_{\Omega} p_{1} \log p_{1} d \mu$ is the differential entropy of $P_{1}$.

We will allow the abuse of notation to write the KL-divergence in terms of the pdf's as the arguments, $D_{K L}\left(p_{1} \| p_{2}\right)$. In the case of $p_{1}(x), p_{2}(x)$ are probability density functions of a continuous random variable $X, \mathrm{KL}$ divergence of $p_{2}$ form $p_{1}$ is,

$$
\begin{equation*}
D_{K L}\left(p_{1} \| p_{2}\right)=\mathbb{E}_{p_{1}}\left[\log \left(\frac{p_{1}}{p_{2}}\right)\right]=\int p_{1}(x) \log \left(\frac{p_{1}(x)}{p_{2}(x)}\right) d x \tag{2.5}
\end{equation*}
$$

Furthermore, if $p_{1}\left(x^{\prime} \mid x\right), p_{2}\left(x^{\prime} \mid x\right)$ are conditional probability density functions of random variables $\left(X, X^{\prime}\right)$ then the KL divergence,

$$
\begin{equation*}
D_{K L}\left(p_{1}\left(x^{\prime} \mid x\right) \| p_{2}\left(x^{\prime} \mid x\right)\right)=\iint p_{1}\left(x^{\prime}, x\right) \log \left(\frac{p_{1}\left(x^{\prime} \mid x\right)}{p_{2}\left(x^{\prime} \mid x\right)}\right) d x^{\prime} d x \tag{2.6}
\end{equation*}
$$

Hence the definition of transfer entropy [6, 7, 43] is given by:

$$
\begin{equation*}
T_{y \rightarrow x}=D_{K L}\left(p\left(x^{\prime} \mid x, y\right) \| p\left(x^{\prime} \mid x\right)\right) \tag{2.7}
\end{equation*}
$$

and, it can be explained by conditional entropy as:

$$
\begin{align*}
T_{y \rightarrow x}= & \iiint p\left(x, y, x^{\prime}\right) \log \left(\frac{p\left(x^{\prime} \mid x, y\right)}{p\left(x^{\prime} \mid x\right)}\right) d x d y d x^{\prime}  \tag{2.8}\\
= & \iiint p\left(x, y, x^{\prime}\right) \log p\left(x^{\prime} \mid x, y\right) d x d y d x^{\prime} \\
& -\iint p\left(x, x^{\prime}\right) \log p\left(x^{\prime} \mid x\right) d x d x^{\prime} \\
= & -h\left(X^{\prime} \mid X, Y\right)+h\left(X^{\prime} \mid X\right) \\
= & h\left(X^{\prime} \mid X\right)-h\left(X^{\prime} \mid X, Y\right)
\end{align*}
$$

where $h\left(X^{\prime} \mid X\right)=-\mathbb{E}\left[\log p\left(X^{\prime} \mid X\right)\right]$ and $h\left(X^{\prime} \mid X, Y\right)=-\mathbb{E}\left[\log p\left(X^{\prime} \mid X, Y\right)\right]$.
The Kullback-Leibler divergence used here contrasts these two possible explanations of the process generating $x^{\prime}$. Since $D_{K L}$ may be written in terms of mutual information, the units are as any entropy, bits per time step. Notice that we have overloaded the notation writing $p\left(x^{\prime} \mid x\right)$ and $p\left(x^{\prime} \mid x, y\right)$. Our practice will be to rely on the arguments to distinguish functions as otherwise different (likewise) distinguishing cases of $f_{1}(x)$ versus $f_{1}(x, y)$.

### 2.2 Transfer Entropy for Deterministic Systems

Considering the evolution of deterministic dynamical system $x$ as a stochastic process [6, 30], we may write a probability density function in terms of all those variables that may be relevant, $p\left(x, y, x^{\prime}\right)$. Contrasting the role of the various input variables requires us to develop a new singular transfer operator between domains that do not necessarily have the same number of variables. Notice that the definition of transfer entropy (Equation (2.7)) seems to rely on the absolute continuity of the joint probability density $p\left(x, y, x^{\prime}\right)$. However, that joint distribution of $p(x, y, f(x, y))$ is generally not absolutely continuous, noticing its support is
$\left\{(x, y, f(x, y)):(x, y) \in \Omega_{x} \times \Omega_{y} \subseteq \mathbb{R}^{2}\right\}$, a measure 0 subset of $\mathbb{R}^{3}$. Therefore, the expression $h(f(X, Y) \mid X, Y)$ is not well defined as a differential entropy and hence there is a problem with transfer entropy. We expand upon this important detail in the upcoming subsection. To guarantee existence, we interpret these quantities by convolution to smooth the problem. Adding an "artificial noise" with standard deviation parameter $\epsilon$ allows definition of the conditional entropy at the singular limit $\epsilon$ approaches to zero, and likewise the transfer entropy follows.

The probability density function of the sum of two continuous random variables $(U, Z)$ can be obtained by convolution, $P_{U+Z}=P_{U} * P_{Z}$. Random noise ( $Z$ with mean $\mathbb{E}(Z)=0$ and variance $\mathbb{V}(Z)=C \epsilon^{2}$ ) added to the original observable variables regularizes, and we are interested in the singular limit, $\epsilon \rightarrow 0$. We assume that $Z$ is independent of $X, Y$. In experimental data from practical problems, we argue that some noise, perhaps even if small, is always present. Additionally, noise is assumed to be uniform or normally distributed in practical applications. Therefore, for simplicity of the discussion, we mostly focused on those two distributions. With this concept, Transfer Entropy can now be calculated by using $h\left(X^{\prime} \mid X, Y\right)$ and $h\left(X^{\prime} \mid X\right)$ when

$$
\begin{equation*}
X^{\prime}=f(X, Y)+Z \tag{2.9}
\end{equation*}
$$

where now we assume that $X, Y, Z \in \mathbb{R}$ are independent random variables and we assume that $f: \Omega_{x} \times \Omega_{y} \rightarrow \mathbb{R}$ is a component-wise monotonic (we will consider the monotonically increasing case for consistent explanations, but one can use monotonically decreasing functions in similar manner) continuous function of $X, Y$ and $\Omega_{x}, \Omega_{y} \subseteq \mathbb{R}$.

### 2.2.1 Calculating Conditional differential Entropy

Calculation of transfer entropy depends on the conditional probability. Hence, we will first focus on conditional probability. Since for any particular values $x, y$ the function value $f(x, y)$ is fixed, we conclude that $X^{\prime} \mid x, y$ is just a linear function of Z. We see that

$$
\begin{equation*}
p_{X^{\prime} \mid X, Y}\left(x^{\prime} \mid x, y\right)=\operatorname{Pr}\left(Z=x^{\prime}-f(x, y)\right)=p_{Z}\left(x^{\prime}-f(x, y)\right), \tag{2.10}
\end{equation*}
$$

where $p_{Z}$ is the probability density function of $Z$.
Note that the random variable $X^{\prime} \mid x$ is a function of $(Y, Z)$. To write $U+Z$, let $U=f(x, Y)$. Therefore, convolution of densities of $U$ and $Z$ gives the density function for $p\left(x^{\prime} \mid x\right)$ (See Section 2.2 .2 for examples). Notice that a given value of the random variable, say $X=\alpha$, is a parameter in $U$. Therefore, we will denote $U=f(Y ; \alpha)$. We will first focus on the probability density function of $U, p_{U}(u)$, using the Frobenius-Perron operator,

$$
\begin{equation*}
p_{U}(u)=\sum_{y: u=f(y ; \alpha)} \frac{p_{Y}(f(y ; \alpha))}{\left|f^{\prime}(f(y ; \alpha))\right|} . \tag{2.11}
\end{equation*}
$$

In the multivariate setting, the formula is extended similarly interpreting the derivative as the Jacobian matrix, and the absolute value is interpreted as the absolute value of the determinant. Denote $\mathbf{Y}=\left(Y_{1}, \Upsilon_{2}, \ldots, Y_{n}\right), \mathbf{g}(\mathbf{Y} ; \alpha)=$ $\left(g_{1}, g_{2}, \ldots, g_{n}\right)$ and $U=f(\alpha, \mathbf{Y}):=g_{1}(\mathbf{Y} ; \alpha) ;$ and the vector $\mathbf{V}=\left(V_{1}, V_{2}, \ldots, V_{n-1}\right) \in$ $\mathbb{R}^{n-1}$ such that $V_{i}=g_{i+1}(\mathbf{Y}):=Y_{i+1}$ for $i=1,2, \ldots, n-1$. Then, the absolute value of the determinate of the Jacobian matrix is given by: $\left|J_{g}(\mathbf{y})\right|=\left|\frac{\partial g_{1}(\mathbf{y} ; \alpha)}{\partial y_{1}}\right|$. As an aside, note that $J$ is lower triangular with diagonal entries $d_{i i}=1$ for $i>1$. The
probability density function of $U$ is given by

$$
\begin{equation*}
p_{U}(u)=\int_{S} p_{\mathbf{Y}}\left(g^{-1}(u, \mathbf{v} ; \alpha)\right)\left|\frac{\partial g_{1}}{\partial y_{1}}\left(g^{-1}(u, \mathbf{v} ; \alpha)\right)\right|^{-1} d \mathbf{v}, \tag{2.12}
\end{equation*}
$$

where $S$ is the support set of the random variable $\mathbf{V}$.
Since the random variable $X^{\prime} \mid x$ can be written as a sum of $U$ and $Z$, we find the probability density function by convolution as follows:

$$
\begin{equation*}
p_{X^{\prime} \mid x}\left(x^{\prime} \mid x\right)=\int p_{U}(u) p_{Z}\left(x^{\prime}-u\right) d u . \tag{2.13}
\end{equation*}
$$

Now, the conditional differential entropy $h(Z \mid X, Y)$ is in terms of these probability densities. It is useful that translation does not change the differential entropy, $h_{\epsilon}(f(X, Y)+Z \mid X, Y)=h(Z \mid X, Y)$. In addition, $Z$ is independent from $X, Y$, $h(Z \mid X, Y)=h(Z)$. Now, we define

$$
\begin{equation*}
h(f(X, Y) \mid X, Y):=\lim _{\epsilon \rightarrow 0^{+}} h_{\epsilon}(f(X, Y)+Z \mid X, Y) \tag{2.14}
\end{equation*}
$$

if this limit exists.
We consider two scenarios: (1) $Z$ is a uniform random variable or (2) $Z$ is a Gaussian random variable. If it is uniform in the interval $[-\epsilon / 2, \epsilon / 2]$, then the differential entropy is $h(Z)=\ln (\epsilon)$. If specifically, $Z$ is Gaussian with zero mean and $\epsilon$ standard deviation, then $h(Z)=\frac{1}{2} \ln \left(2 \pi e \epsilon^{2}\right)$. Therefore, $h_{\epsilon}(f(X, Y)+Z \mid X, Y) \rightarrow$ $-\infty$ as $\epsilon \rightarrow 0^{+}$in both cases. Therefore, $\left.h(f(X, Y) \mid X, Y)\right)$ is not finite in this definition (Equation (2.14) as well. Thus, instead of calculating $X^{\prime}=f(X, Y)$, we need
to use a noisy version of data $X^{\prime}=f(X, Y)+Z$. For that case,

$$
h\left(X^{\prime} \mid X, Y\right)=h(Z)= \begin{cases}\ln (\epsilon) ; & Z \sim U(-\epsilon / 2, \epsilon / 2)  \tag{2.15}\\ \frac{1}{2} \ln \left(2 \pi e \epsilon^{2}\right) ; & Z \sim \mathcal{N}\left(0, \epsilon^{2}\right)\end{cases}
$$

where $U(-\epsilon / 2, \epsilon / 2)$ is the uniform distribution in the interval $[-\epsilon / 2, \epsilon / 2]$, and $\mathcal{N}\left(0, \epsilon^{2}\right)$ is a Gaussian distribution with zero mean and $\epsilon$ standard deviation.

Now, we focus on $h\left(X^{\prime} \mid X\right)$. If $X^{\prime}$ is just a function of $X$, then we can similarly show that: if $X^{\prime}=f(X)$, then

$$
h(f(X)+Z \mid X)=h(Z)= \begin{cases}\ln (\epsilon) ; & Z \sim U(-\epsilon / 2, \epsilon / 2)  \tag{2.16}\\ \frac{1}{2} \ln \left(2 \pi e \epsilon^{2}\right) ; & Z \sim \mathcal{N}\left(0, \epsilon^{2}\right)\end{cases}
$$

In addition, notice that, if $X^{\prime}=f(X, Y)$, then $h\left(X^{\prime} \mid X\right)$ will exist, and most of the cases will be finite. However, when we calculate $T_{y \rightarrow x}$, we need to use the noisy version to avoid the issues in calculating $h\left(X^{\prime} \mid X, Y\right)$. We will now consider the interesting case $X^{\prime}=f(X, Y)+Z$ and calculate $h\left(X^{\prime} \mid X\right)$. We require $p_{X^{\prime} \mid X}$ and Equation (2.13) can be used to calculate this probability. Let us denote $I:=$ $\int p_{U}(u) p_{Z}\left(x^{\prime}-u\right) d u$; then,

$$
\begin{align*}
h_{\epsilon}\left(X^{\prime} \mid X\right) & =\iint I p_{X}(x) \ln (I) d x^{\prime} d x  \tag{2.17}\\
& =\int p_{X}(x) \int I \ln (I) d x^{\prime} d x \\
& =\mathbb{E}_{X}(Q),
\end{align*}
$$

where $Q=\int I \ln (I) d x^{\prime}$. Notice that, if $Q$ does not depend on $x$, then $h\left(X^{\prime} \mid X\right)=$ $Q \int p_{X} d x=Q$ because $\int p_{X} d x=1$ (since $p_{x}$ is a probability density function).

Therefore, we can calculate $h_{\epsilon}\left(X^{\prime} \mid X\right)$ by four steps. First, we calculate the density function for $U=f(x, Y)$ (by using Equation (2.11) or (2.12). Then, we calculate $I=p_{X^{\prime} \mid X}$ by using Equation (2.13). Next, we calculate the value of $Q$, and finally we calculate the value of $h_{\epsilon}\left(X^{\prime} \mid X\right)$.

Thus, the transfer entropy from $y$ to $x$ follows in terms of comparing conditional entropies,

$$
\begin{equation*}
T_{y \rightarrow x}=h\left(X^{\prime} \mid X\right)-h\left(X^{\prime} \mid X, Y\right) \tag{2.18}
\end{equation*}
$$

This quantity is not well defined when $X^{\prime}=f(X, Y)$, and therefore we considered the $X^{\prime}=f(X, Y)+Z$ case. This interpretation of transfer entropy depends on the parameter $\epsilon$, as we define

$$
\begin{equation*}
T_{y \rightarrow x}:=\lim _{\epsilon \rightarrow 0^{+}} T_{y \rightarrow x}(\epsilon)=\lim _{\epsilon \rightarrow 0^{+}} h_{\epsilon}\left(X^{\prime} \mid X\right)-h_{\epsilon}\left(X^{\prime} \mid X, Y\right) \tag{2.19}
\end{equation*}
$$

if this limit exists.
Note that

$$
T_{y \rightarrow x}= \begin{cases}\lim _{\epsilon \rightarrow 0^{+}} h(Z)-h(Z)=0 ; & X^{\prime}=f(X)  \tag{2.20}\\ \infty ; & X^{\prime}=f(X, Y) \neq f(X)\end{cases}
$$

Thus, we see that a finite quantity is ensured by the noise term. We can easily find an upper bound for the transfer entropy when $X^{\prime}=f(X, Y)+Z$ is a random variable with finite support (with all the other assumptions mentioned earlier) and suppose $Z \sim U(-\epsilon / 2, \epsilon / 2)$. First, notice that the uniform distribution maximizes entropy amongst all distributions of continuous random variables with finite support. If $f$ is component-wise monotonically increasing continuous function, then
the support of $X^{\prime} \mid x$ is $\left[f\left(x, y_{\text {min }}\right)-\epsilon / 2, f\left(x, y_{\text {min }}\right)+\epsilon / 2\right]$ for all $x \in \Omega_{x}$. Here, $y_{\text {min }}$ and $y_{\max }$ are minimum and maximum values of $Y$. Then, it follows that

$$
\begin{equation*}
h_{\epsilon}\left(X^{\prime} \mid X\right) \leq \ln \left(\left|f\left(x_{\max }, y_{\max }\right)-f\left(x_{\max }, y_{\min }\right)+\epsilon\right|\right), \tag{2.21}
\end{equation*}
$$

where $x_{\max }$ is the maximum $x$ value. We see that an interesting upper bound for transfer entropy follows:

$$
\begin{equation*}
T_{y \rightarrow x}(\epsilon) \leq \ln \left(\left|\frac{f\left(x_{\max }, y_{\max }\right)-f\left(x_{\max }, y_{\min }\right)}{\epsilon}+1\right|\right) \tag{2.22}
\end{equation*}
$$

### 2.2.2 Numerical Demonstration of Boundedness Issue

In this section, we will focus on analytical results and numerical estimators for conditional entropy and transfer entropy for specific examples (see Figure 2.1, 2.2 ). As we discussed in previous sections starting with ??, computing the transfer entropy for $X^{\prime}=f(X, Y)$ has technical difficulties due to the singularity of the quantity $h\left(X^{\prime} \mid X, Y\right)$. First, we will consider the calculation of $h\left(X^{\prime} \mid X\right)$ for $X^{\prime}=$ $f(X, Y)$, and then we will discuss the calculation for noisy data. In the following examples, we assumed that $X, Y$ are random variables such that $X, Y \stackrel{i i d}{\sim} U([1,2])$. A summary of the calculations for a few examples are listed in Table 2.1.

Table 2.1: Conditional entropy $h\left(X^{\prime} \mid X\right)$ for $X^{\prime}=f(X, Y)$, for specific parametric examples listed, under the assumption that $X, Y \stackrel{i i d}{\sim} U([1,2])$.

| $f(X, Y)$ | $h\left(X^{\prime} \mid X\right)$ |
| :---: | :---: |
| $g(X)+b Y$ | $\ln (b)$ |
| $g(X)+b Y^{2}$ | $\ln (8 b)-5 / 2$ |
| $g(X)+b \ln (Y)$ | $\ln \left(\frac{b e}{4}\right)$ |

We will discuss the transfer entropy with noisy data because making $h\left(X^{\prime} \mid X, Y\right)$
well defined requires absolute continuity of the probability density function $p\left(x, y, x^{\prime}\right)$. Consider, for example, the problem form $X^{\prime}=g(X)+b Y+C$, where $X, Y$ are uniformly distributed independent random variables over the interval $[1,2]$ (the same analysis can be extend to any finite interval) with $b$ being a constant, and $g$ a function of random variable $X$. We will also consider $C$ to be a random variable, which is distributed uniformly on $[-\epsilon / 2, \epsilon / 2]$. Note that it follows that $h\left(X^{\prime} \mid X, Y\right)=\ln \epsilon$. To calculate the $h\left(X^{\prime} \mid X\right)$, we need to find the conditional probability $p\left(X^{\prime} \mid x\right)$ and observe that $X^{\prime} \mid x=U+C$, where $U=g(x)+b Y$. Therefore,

$$
p_{U}(u)= \begin{cases}\frac{1}{b} & ; g_{1}(x)+b \leq X^{\prime} \leq g_{1}(x)+2 b  \tag{2.23}\\ 0 & ; \text { otherwise }\end{cases}
$$

and

$$
p_{X^{\prime} \mid X}\left(X^{\prime} \mid x\right)= \begin{cases}\frac{x^{\prime}+\epsilon / 2-g(x)}{b \epsilon} & ; g(x)-\epsilon / 2 \leq X^{\prime} \leq g(x)+\epsilon / 2  \tag{2.24}\\ \frac{1}{b} & ; g(x)+\epsilon / 2 \leq X^{\prime} \leq b+g(x)-\epsilon / 2 \\ \frac{-x^{\prime}+\epsilon / 2+g(x)+b}{b \epsilon} & ; b+g(x)-\epsilon / 2 \leq X^{\prime} \leq b+g(x)+\epsilon / 2 \\ 0 & ; \text { otherwise }\end{cases}
$$

By the definition of transfer entropy, we can show that

$$
\begin{equation*}
h\left(X^{\prime} \mid X\right)=\ln b+\frac{\epsilon}{2 b} \tag{2.25}
\end{equation*}
$$

and hence transfer entropy of this data are given by

$$
T_{y \rightarrow x}(\epsilon ; b)= \begin{cases}\ln \frac{b}{\epsilon}+\frac{\epsilon}{2 b} ; & b \neq 0  \tag{2.26}\\ 0 ; & b=0\end{cases}
$$

Therefore, when $b=0$, the transfer entropy $T_{y \rightarrow x}=\ln \epsilon-\ln \epsilon=0$. In addition, notice that $T_{y \rightarrow x}(\epsilon ; b) \rightarrow \infty$ as $\epsilon \rightarrow 0$. Therefore, convergence of the numerical estimates is slow when $\epsilon>0$ is small (see Figure 2.2).

(a) Examples for $X^{\prime}=g(X)+b Y$. The left figure shows results for $g(X)=X$ and the right shows results for $g(X)=X^{2}$.


(b) Examples for $X^{\prime}=g(X)+b Y^{2}$. The left figure shows results for $g(X)=X$ and the right shows results for $g(X)=e^{x}$.

Figure 2.1: Conditional entropy $h\left(X^{\prime} \mid X\right)$. Note that these numerical estimates for the conditional entropy by the KSG method [28], converge (as $N \rightarrow \infty$ ) to the analytic solutions (see Table 2.1).


Figure 2.2: Numerical results and analytical results for transfer entropy $T_{y \rightarrow x}(\epsilon ; b)$ to the problem $X^{\prime}=X+b Y+\epsilon$. Transfer entropy vs. $\epsilon$ shows in (a) for fixed $b$ value. (b) and (c) show the behavior of the transfer entropy for $b$ values with fixed $\epsilon$ values. Notice that convergence of numerical solution is slow when epsilon is small.

Boundedness issue noticed in this section motivate us to finding a alternative description of the information flow in dynamical systems. As we discussed in the introduction section geometry is perhaps a natural place to describe the dynamics.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309. DOI: $10.1063 / 1.5031109$, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https: //www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138. URL: https://link.aps.org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL:https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079, eprint: https://science.sciencemag.org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag . org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi . org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 3

## Relating Geometric Measures to Causal Inference

Now, we focus on the geometric interpretation of the information flow. As we seen in the section 2.1, causality question can be stated as equation (2.3). Transfer entropy anser this question indirectly by conditional probabilities (eq. 2.4). Our aim is to define the causality question more directly using the geometric measures. Furthermore, we can relate the geometric measures with probabilities and even transfer entropy. This chapter will relate the probability and transfer entropy with the geometric measures.

### 3.1 Geometric Interpretation of Causal Inference

Consider that the coupling structure between variables may be characterized by the directed graph illustrated in Figure 3.1.


Figure 3.1: A directed graph presentation of the coupling stucture questions corresponding to Equations (2.1) and (2.2).

In one time step, without loss of generality, we may decide on equation (2.3), the role of $y$ on $x^{\prime}$, based on $T_{y \rightarrow x}>0$, exclusively in terms of the details of the argument structure of $f_{1}$. This is separate from the reverse question of $f_{2}$ as to whether $T_{x \rightarrow y}>0$. In geometric terms, assuming $f_{1} \in C^{1}\left(\Omega_{1}\right)$, it is clear that, unless the partial derivative $\frac{\partial f_{1}}{\partial y}$ is zero everywhere, then the $y$ argument in $f_{1}(x, y)$ is relevant. This is not a necessary condition for $T_{y \rightarrow x}>0$, which is a probabilistic statement, and almost everywhere is sufficient.

For further analysis, consider a manifold of points $\left(x, y, x^{\prime}\right) \in X \times Y \times X^{\prime}$ as the graph over $\Omega_{1}$, which we label $\mathcal{M}_{2}$. In the following, we assume $f_{1} \in$ $C^{1}\left(\Omega_{1}\right), \Omega_{1} \subset X \times Y$. Our primary assertion here is that the geometric aspects of the set $\left(x, y, x^{\prime}\right)$ projected into $\left(x, x^{\prime}\right)$ distinguishes the information flow structure. Refer to Figure 3.2 for notation. Let the level set for a given fixed $y$ be defined,

$$
\begin{equation*}
L_{y}:=\left\{\left(x, x^{\prime}\right): x^{\prime}=f(x, y), y=\text { constant }\right\} \in \Omega_{2}=X \times X^{\prime} \tag{3.1}
\end{equation*}
$$



Figure 3.2: $\Omega_{2}=X \times X^{\prime}$ manifold and $L_{y}$ level set for (a) $x^{\prime}=f_{1}(x)=-0.005 x^{2}+$ 100, (b) $x^{\prime}=f_{1}(x, y)=-0.005 x^{2}+0.01 y^{2}+50$. The dimension of the projected set of ( $x, x^{\prime}$ ) depends on the causality as just described. Compare to Figure 3.3 and Equation (3.8).

When these level sets are distinct, then the question of the relevance of $y$ to the outcome of $x^{\prime}$ is clear:

- If $\frac{\partial f_{1}}{\partial y}=0$ for all $(x, y) \in \Omega_{1}$, then $L_{y}=L_{\tilde{y}}$ for all $y, \tilde{y}$.

Notice that, if the $y$ argument is not relevant as described above, then $x^{\prime}=f_{1}(x)$ better describes the associations, but if we nonetheless insist to write $x^{\prime}=f_{1}(x, y)$, then $\frac{\partial f_{1}}{\partial y}=0$ for all $(x, y) \in \Omega_{1}$. The converse is interesting to state explicitly,

- If $L_{y} \neq L_{\tilde{y}}$ for some $y, \tilde{y}$, then $\frac{\partial f_{1}}{\partial y} \neq 0$ for some $(x, y) \in \Omega_{1}$, and then $x^{\prime}=$ $f_{1}(x)$ is not a sufficient description of what should really be written $x^{\prime}=$ $f_{1}(x, y)$. We have assumed $f_{1} \in C^{1}\left(\Omega_{1}\right)$ throughout.


### 3.2 Relating Transfer Entropy to a Geometric Bound

Noting that transfer entropy and other variations of the G-causality concept are expressed in terms of conditional probabilities, we recall that

$$
\begin{equation*}
p\left(x^{\prime} \mid x, y\right) p(x, y)=p\left(x, y, x^{\prime}\right) \tag{3.2}
\end{equation*}
$$

Again, we continue to overload the notation on the functions $p$, the details of the arguments distinguishing to which of these functions we refer.

Now, consider the change of random variable formulas that map between probability density functions by smooth transformations. In the case that $x^{\prime}=f_{1}(x)$ (in the special case that $f_{1}$ is one-one), then

$$
\begin{equation*}
p\left(x^{\prime}\right)=\frac{p(x)}{\left|\frac{d f_{1}}{d x}(x)\right|}=\frac{p\left(f_{1}^{-1}\left(x^{\prime}\right)\right)}{\left|\frac{d f_{1}}{d x}\left(f_{1}^{-1}\left(x^{\prime}\right)\right)\right|} \tag{3.3}
\end{equation*}
$$

In the more general case, not assuming one-one-ness, we get the usual FrobeniusPerron operator,

$$
\begin{equation*}
p\left(x^{\prime}\right)=\sum_{x: x^{\prime}=f_{1}(x)} p\left(x, x^{\prime}\right)=\sum_{x: x^{\prime}=f_{1}(x)} \frac{p(x)}{\left\lvert\, \frac{d f_{1}(x)}{d x}(x)\right.}, \tag{3.4}
\end{equation*}
$$

in terms of a summation over all pre-images of $x^{\prime}$. Notice also that the middle form is written as a marginalization across $x$ of all those $x$ that lead to $x^{\prime}$. This FrobeniusPerron operator, as usual, maps densities of ensembles of initial conditions under the action of the map $f_{1}$.

Comparing to the expression

$$
\begin{equation*}
p\left(x, x^{\prime}\right)=p\left(x^{\prime} \mid x\right) p(x) \tag{3.5}
\end{equation*}
$$

we assert the interpretation that

$$
\begin{equation*}
p\left(x^{\prime} \mid x\right):=\frac{1}{\left|\frac{d f_{1}}{d x}(x)\right|} \delta\left(x^{\prime}-f_{1}(x)\right) \tag{3.6}
\end{equation*}
$$

where $\delta$ is the Dirac delta function. In the language of Bayesian uncertainty propagation, $p\left(x^{\prime} \mid x\right)$ describes the likelihood function, if interpreting the future state $x^{\prime}$ as data, and the past state $x$ as parameters, in a standard Bayes description, $p$ (data $\mid$ parameter $) \times p$ (parameter). As usual for any likelihood function, while it is a probability distribution over the data argument, it may not necessarily be so with respect to the parameter argument.

Now, consider the case where $x^{\prime}$ is indeed nontrivially a function with respect to not just $x$, but also with respect to $y$. Then, we require the following asymmetric space transfer operator, which we name here an asymmetric Frobenius-Perron operator for smooth transformations between spaces of dissimilar dimensionality:

Theorem 3.2.1 (Asymmetric Space Transfer Operator). If $x^{\prime}=f_{1}(x, y)$, for $f_{1}$ : $\Omega_{1} \rightarrow \mathrm{Y}$, given bounded open domain $(x, y) \in \Omega_{1} \subset \mathbb{R}^{2 d}$, and range $x^{\prime} \in \mathrm{Y} \subset \mathbb{R}^{d}$, and $f_{1} \in C^{1}\left(\Omega_{1}\right)$, and the Jacobian matrices, $\frac{\partial f_{1}}{\partial x}(x, y)$, and $\frac{\partial f_{1}}{\partial y}(x, y)$ are not both rank deficient at the same time, then taking the initial density $p(x, y) \in L^{1}\left(\Omega_{1}\right)$, the following serves as a transfer operator mapping asymmetrically defined densities $P: L^{1}\left(\Omega_{1}\right) \rightarrow$ $L^{1}(\mathrm{Y})$

$$
\begin{equation*}
p\left(x^{\prime}\right)=\sum_{(x, y): x^{\prime}=f_{1}(x, y)} p\left(x, y, x^{\prime}\right)=\sum_{(x, y): x^{\prime}=f_{1}(x, y)} \frac{p(x, y)}{\left|\frac{\partial f_{1}}{\partial x}(x, y)\right|+\left|\frac{\partial f_{1}}{\partial y}(x, y)\right|} \tag{3.7}
\end{equation*}
$$

The proof of this is in Appendix A. Note also that, by similar argumentation, one can formulate the asymmetric Frobenius-Perron type operator between sets of dissimilar dimensionality in an integral form.

Corollary 3.2.1.1 (Asymmetric Transfer Operator, Kernel Integral Form). Under the same hypothesis as Theorem 3.2.1. we may alternatively write the integral kernel form of the expression,

$$
\begin{align*}
P: L^{2}\left(\mathbb{R}^{2}\right) \rightarrow & L^{2}(\mathbb{R})  \tag{3.8}\\
p(x, y) & \left.\mapsto \quad p^{\prime}\left(x^{\prime}\right)=P[p](x, y)\right] \\
& = \\
& =\int_{L_{x^{\prime}}} p\left(x, y, x^{\prime}\right) d x d y=\int_{L_{x^{\prime}}} p\left(x^{\prime} \mid x, y\right) p(x, y) d x d y \\
& =\int_{L_{x^{\prime}}} \frac{1}{\frac{\partial f_{1}}{\partial x}(x, y)\left|+\left|\frac{\partial f_{1}}{\partial y}(x, y)\right|\right.} p(x, y) d x d y \tag{3.9}
\end{align*}
$$

This is in terms of a line integration along the level set, $L_{x^{\prime}}$. See Figure 3.3.

$$
\begin{equation*}
L_{x^{\prime}}=\left\{(x, y) \in \Omega_{1}: f(x, y)=x^{\prime} \text { a chosen constant. }\right\} \tag{3.10}
\end{equation*}
$$

In Figure 3.3, we have shown a typical scenario where a level set is a curve (or it may well be a union of disjoint curves), whereas, in a typical FP-operator between sets of the same dimensionality, generally the integration is between preimages that are usually either singletons, or unions of such points, $p^{\prime}\left(x^{\prime}\right)=\int \delta(s-$ $f(x)) p(s) d s=\sum_{x: f(x)=x^{\prime}} \frac{p(x)}{|D f(x)|}$.


Figure 3.3: The asymmetric transfer operator, Equation (3.8), is written in terms of intefration over the level set, $L_{x^{\prime}}$ of $x^{\prime}=f_{1}(x, y)$ associated with a fixed value $x^{\prime}$, Equation (3.10).

Contrasting standard and the asymmetric forms of transfer operators as described above, in the next section, we will compute and bound estimates for the transfer entropy. However, it should already be apparent that, if $\frac{\partial f_{1}}{\partial y}=0$ in probability with respect to $p(x, y)$, then $T_{y \rightarrow x}=0$.

Comparison to other statistical divergences reveals geometric relevance: Information flow is quite naturally defined by the KL-divergence, in that it comes in the units of entropy, e.g., bits per second. However, the well-known Pinsker's inequality [36] allows us to more easily relate the transfer entropy to a quantity that has a geometric relevance using the total variation, even if this is only by an inequality estimate.

Recall that Pinsker's inequality [36] relates random variables with probability distributions $p_{1}$ and $p_{2}$ over the same support to the total variation and the KLdivergence as follows:

$$
\begin{equation*}
0 \leq \frac{1}{2} T V\left(P_{1}, P_{2}\right) \leq \sqrt{D_{K L}\left(P_{1} \| P_{2}\right)} \tag{3.11}
\end{equation*}
$$

written as probability measures $P_{1}, P_{2}$. The total variation distance between probability measures is a maximal absolute difference of possible events,

$$
\begin{equation*}
T V\left(P_{1}, P_{2}\right)=\sup _{A}\left|P_{1}(A)-P_{2}(A)\right| \tag{3.12}
\end{equation*}
$$

but it is well known to be related to $1 / 2$ of the L1-distance in the case of a common dominating measure, $p_{1}(x) d \mu=d P_{1}, p_{2}(x) d \mu=d P_{2}$. In this work, we only need absolute continuity with respect to Lebesgue measure, $p_{1}(x)=d P_{1}(x), p_{2}(x)=$ $d P_{2}(x)$; then,

$$
\begin{equation*}
T V\left(P_{1}, P_{2}\right)=\frac{1}{2} \int\left|p_{1}(x)-p_{2}(x)\right| d x=\frac{1}{2}\left\|p_{1}-p_{2}\right\|_{L^{1}} \tag{3.13}
\end{equation*}
$$

here with respect to Lebesgue measure. In addition, we write $D_{K L}\left(P_{1} \| P_{2}\right)=$ $\int p_{1}(x) \log \frac{p_{1}(x)}{p_{2}(x)} d x$; therefore,

$$
\begin{equation*}
\frac{1}{2}\left\|p_{1}-p_{2}\right\|_{L^{1}}^{2} \leq \int p_{1}(x) \log \frac{p_{1}(x)}{p_{2}(x)} d x \tag{3.14}
\end{equation*}
$$

Thus, with the Pinsker inequality, we can bound the transfer entropy from below by inserting the definition Equation (2.7) into the above:

$$
\begin{equation*}
0 \leq \frac{1}{2}\left\|p\left(x^{\prime} \mid x, y\right)-p\left(x^{\prime} \mid x\right)\right\|_{L^{1}}^{2} \leq T_{y \rightarrow x} \tag{3.15}
\end{equation*}
$$

The assumption that the two distributions correspond to a common dominating measure requires that we interpret $p\left(x^{\prime} \mid x\right)$ as a distribution averaged across the same $p(x, y)$ as $p\left(x^{\prime} \mid x, y\right)$. (Recall by definition [11] that $\lambda$ is a common dominating measure of $P$ and $Q$ if $p(x)=d P / d \lambda$ and $q(x)=d Q / d \lambda$ describe corresponding densities). For the sake of simplification, we interpret transfer entropy relative to
a uniform initial density, $p(x, y)$, for both entropies of Equation (2.18). With this assumption, we interpret

$$
\begin{equation*}
0 \leq \frac{1}{2}\left\|\frac{1}{\left|\frac{\partial f_{1}}{\partial x}(x, y)\right|+\left|\frac{\partial f_{1}}{\partial y}(x, y)\right|}-\frac{1}{\left|\frac{d f_{1}}{d x}(x)\right|}\right\|_{L^{1}\left(\Omega_{1}, p(x, y)\right)}^{2} \leq T_{y \rightarrow x} \tag{3.16}
\end{equation*}
$$

In the special case that there is very little information flow, we would expect that $\left|\frac{\partial f_{1}}{\partial y}\right|<b \ll 1$, and $b \ll\left|\frac{\partial f_{1}}{\partial x}\right|$, almost every $x, y$; then, a power series expansion in small $b$ gives

$$
\begin{equation*}
\frac{1}{2}\left\|\frac{1}{\left|\frac{\partial f_{1}}{\partial x}(x, y)\right|+\left|\frac{\partial f_{1}}{\partial y}(x, y)\right|}-\frac{1}{\left|\frac{d f_{1}}{d x}(x)\right|}\right\|_{L^{1}\left(\Omega_{1}, p(x, y)\right)}^{2} \approx \frac{\operatorname{Vol}\left(\Omega_{1}\right)}{2} \frac{<\left|\frac{\partial f_{1}}{\partial y}\right|>^{2}}{<\left|\frac{\partial f_{1}}{\partial x}\right|>^{4}} \tag{3.17}
\end{equation*}
$$

which serves approximately as the TV-lower bound for transfer entropy where have used the notation $<\cdot>$ to denote an average across the domain. Notice that, therefore, $\delta\left(p\left(x^{\prime} \mid x, y\right), p\left(x^{\prime} \mid x\right)\right) \downarrow$ as $\left|\frac{\partial f_{1}}{\partial y}\right| \downarrow$. While Pinsker's inequality cannot guarantee that $T_{y \rightarrow x} \downarrow$, since TV is only an upper bound, it is clearly suggestive. In summary, comparing inequality Equation (3.16) to the approximation 3.17) suggests that, for $\left|\frac{\partial f_{1}}{\partial y}\right| \ll b \ll\left|\frac{\partial f_{1}}{\partial x}\right|$, for $b>0$, for a.e. $x, y$, then $T_{y \rightarrow x} \downarrow$ as $b \downarrow$.

Now, we change to a more computational direction of this story of interpreting information flow in geometric terms. With the strong connection described in the following section, we bring to the problem of information flow between geometric concepts to information flow concepts, such as entropy, it is natural to turn to studying the dimensionality of the outcome spaces, as we will now develop.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309 . DOI: 10.1063/1.5031109, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https: //www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138, URL: https://link.aps . org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL: https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https: //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079. eprint: https://science.sciencemag. org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag. org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi. org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991.
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844.
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 4

## Geometry of Information Flow

Now, we will explore numerical estimation aspects of transfer entropy for causation inference in relationship to geometry as described theoretically in the previous chapter, and we will compare this numerical approach to geometric aspects.

As theory suggests, there is a strong relationship between the information flow (causality as measured by transfer entropy) and the geometry, encoded for example in the estimates leading to Equation (3.17). The effective dimensionality of the underlying manifold as projected into the outcome space is a key factor to identify the causal inference between chosen variables. Indeed, any question of causality is in fact observer dependent. To this point, suppose $x^{\prime}$ only depends on $x, y$ and $x^{\prime}=f(x, y)$, where $f \in C^{1}\left(\Omega_{1}\right)$. We noticed that (Section ??) $T_{y \rightarrow x}=0 \Longleftrightarrow \frac{\partial f}{\partial y}=$ $0, \forall(x, y) \in \Omega_{1}$. Now, notice that $\frac{\partial f}{\partial y}=0, \forall(x, y) \in \Omega_{1} \Longleftrightarrow x^{\prime}=f(x, y)=f(x)$. Therefore, in the case that $\Omega_{1}$ is two-dimensional, then $\left(x, x^{\prime}\right)$ would be a onedimensional, manifold if and only if $\frac{\partial f}{\partial y}=0, \forall(x, y) \in \Omega_{1}$. See Figure 3.2. With these assumptions,

$$
T_{y \rightarrow x}=0 \Longleftrightarrow\left(x, x^{\prime}\right) \text { data lie on a1D manifold. }
$$

### 4.1. RELATING THE INFORMATION FLOW AS GEOMETRIC ORIENTATION OF DATA56

Likewise, for more general dimensionality of the initial $\Omega_{1}$, the story of the information flow between variables is in part a story of how the image manifold is projected. Therefore, our discussion will focus on estimating the dimensionality in order to identify the nature of the underlying manifold. Then, we will focus on identifying causality by estimating the dimension of the manifold, or even more generally of the resulting set if it is not a manifold but perhaps even a fractal. Finally, this naturally leads us to introduce a new geometric measure for characterizing the causation, which we will identify as $G e o_{y \rightarrow x}$.

### 4.1 Relating the Information Flow as Geometric Orientation of Data

For a given time series $x:=x_{n} \in \mathbb{R}^{d_{1}}, y:=y_{n} \in \mathbb{R}^{d_{2}}$, consider the $x^{\prime}:=x_{n+1}$ and contrast the dimensionalities of $\left(x, y, x^{\prime}\right)$ versus $\left(x, x^{\prime}\right)$, in order to identify that $x^{\prime}=f(x)$ or $x^{\prime}=f(x, y)$. Thus, in mimicking the premise of Granger causality, or likewise of Transfer entropy, contrasting these two versions of the explanations of $x^{\prime}$, in terms of either $(x, y)$ or $x$, we decide the causal inference, but this time, by using only the geometric interpretation. First, we recall how fractal dimensionality evolves under transformations, [42].

Theorem 4.1.1 ([42]). Let A be a bounded Borel subset of $\mathbb{R}^{d_{1}}$. Consider the function $F: A \rightarrow \mathbb{R}^{d_{1}} \times \mathbb{R}^{d_{1}}$ such that $F(x)=\left(x, x^{\prime}\right)$ for some $x^{\prime} \in \mathbb{R}^{d_{1}}$. The correlation dimension $D_{2}(F(A)) \leq d_{1}$, if and only if there exists a function $f: A \rightarrow \mathbb{R}^{d_{1}}$ such that $x^{\prime}=f(x)$ with $f \in C^{1}(A)$.

The idea of the arguments in the complete proof found in Sauer et. al., [42], are as follows. Let $A$ be bounded Borel subset of $\mathbb{R}^{d_{1}}$ and $f: A \rightarrow \mathbb{R}^{d_{1}}$ with $f \in C^{1}(A)$.

Then, $D_{2}(f(A))=D_{2}(A)$, where $D_{2}$ is the correlation dimension [41]. Note that $D_{2}(A) \leq d_{1}$. Therefore, $D_{2}(F(A))=D_{2}(A) \leq d_{1}$, with $F: A \rightarrow \mathbb{R}^{d_{1}} \times \mathbb{R}^{d_{1}}$ if and only if $F(x)=(x, f(x))$.

Now, we can describe this dimensional statement in terms of our information flow causality discussion, to develop an alternative measure of inference between variables. Let $\left(x, x^{\prime}\right) \in \Omega_{2} \subset \mathbb{R}^{2 d_{1}}$ and $\left(x, y, x^{\prime}\right) \in \Omega_{3} \subset \mathbb{R}^{2 d_{1}+d_{2}}$. We assert that there is a causal inference from $y$ to $x$, if $\operatorname{dim}\left(\Omega_{2}\right)>d_{1}$ and $d_{1}<\operatorname{dim}\left(\Omega_{3}\right) \leq d_{1}+d_{2}$, (Theorem 4.1.1). In this paper, we focus on time series $x_{n} \in \mathbb{R}$ which might also depend on time series $y_{n} \in \mathbb{R}$, and we will consider the geometric causation from $y$ to $x$, for $(x, y) \in A \times B=\Omega_{1} \subset \mathbb{R}^{2}$. We will denote geometric causation by GeoC $C_{y \rightarrow x}$ and assume that $A, B$ are Borel subsets of $\mathbb{R}$. Correlation dimension is used to estimate the dimensionality. First, we identify the causality using the dimensionality of on $\left(x, x^{\prime}\right)$ and $\left(x, y, x^{\prime}\right)$. Say, for example, that $\left(x, x^{\prime}\right) \in \Omega_{2} \subset$ $\mathbb{R}^{2}$ and $\left(x, y, x^{\prime}\right) \in \Omega_{3} \subset \mathbb{R}^{3}$; then, clearly we would enumerate a correlation dimension causal inference from $y$ to $x$, if $\operatorname{dim}\left(\Omega_{2}\right)>1$ and $1<\operatorname{dim}\left(\Omega_{3}\right) \leq 2$ (Theorem 1).

### 4.1.1 Measure Causality by Correlation Dimension

As we have been discussing, the information flow of a dynamical system can be described geometrically by studying the sets (perhaps they are manifolds) $X \times X^{\prime}$ and $X \times Y \times X^{\prime}$. As we noticed in the last section, comparing the dimension of these sets can be interpreted as descriptive of information flow. Whether dimensionality be estimated from data or by a convenient fractal measure such as the correlation dimension $\left(D_{2}().\right)$, there is an interpretation of information flow when contrasting $X \times X^{\prime}$ versus $X \times Y \times X^{\prime}$, in a spirit reminiscent of what is done with transfer entropy. However, these details are geometrically more to the point.

### 4.1. RELATING THE INFORMATION FLOW AS GEOMETRIC ORIENTATION OF DATA58

Here, we define $\operatorname{GeoC}_{y \rightarrow x}$ (geometric information flow) by GeoC(.|.) as conditional correlation dimension.

Definition 4.1.1 (Conditional Correlation Dimensional Geometric Information Flow). Let $\mathcal{M}$ be the manifold of data set $\left(X_{1}, X_{2}, \ldots, X_{n}, X^{\prime}\right)$ and let $\Omega_{1}$ be the data set $\left(X_{1}, X_{2}, \ldots, X_{n}\right)$. Suppose that the $\mathcal{M}, \Omega_{1}$ are bounded Borel sets. The quantity

$$
\begin{equation*}
\operatorname{GeoC}\left(X^{\prime} \mid X_{1}, \ldots, X_{n}\right):=D_{2}(\mathcal{M})-D_{2}\left(\Omega_{1}\right) \tag{4.1}
\end{equation*}
$$

is defined as "Conditional Correlation Dimensional Geometric Information Flow". Here, $D_{2}($.$) is the usual correlation dimension of the given set, [16-18].$

Definition 4.1.2 (Correlation Dimensional Geometric Information Flow). Let $x:=$ $x_{n}, y=y_{n} \in \mathbb{R}$ be two time series. The correlation dimensional geometric information flow from $y$ to $x$ as measured by the correlation dimension and denoted by $G e o C_{y \rightarrow x}$ is given by

$$
\begin{equation*}
\operatorname{GeoC}_{y \rightarrow x}:=\operatorname{GeoC}\left(X^{\prime} \mid X\right)-\operatorname{GeoC}\left(X^{\prime} \mid X, Y\right) \tag{4.2}
\end{equation*}
$$

A key observation is to notice that, if $X^{\prime}$ is a function of $\left(X_{1}, X_{2}, \ldots, X_{n}\right)$, then $D_{2}(\mathcal{M})=D_{2}\left(\Omega_{1}\right)$; otherwise, $D_{2}(\mathcal{M})>D_{2}\left(\Omega_{1}\right)$ (Theorem 1). If $X$ is not influenced by $y$, then $\operatorname{GeoC}\left(X^{\prime} \mid X\right)=0, \operatorname{GeoC}\left(X^{\prime} \mid X, Y\right)=0$ and therefore $G e o C_{y \rightarrow x}=$ 0 . In addition, notice that $\operatorname{Geo}_{y \rightarrow x} \leq D_{2}(X)$, where $X=\left\{x_{n} \mid n=1,2, \ldots\right\}$. For example, if $x_{n} \in \mathbb{R}$, then $\operatorname{GeoC}_{y \rightarrow x} \leq 1$. Since we assume that influence of any time series $z_{n} \neq x_{n}, y_{n}$ to $x_{n}$ is relatively small, we can conclude that GeoC $y_{y \rightarrow x} \geq 0$, and, if $x^{\prime}=f(x, y)$, then $\operatorname{GeoC}\left(X^{\prime} \mid X, Y\right)=0$. Additionally, the dimension $\left(\operatorname{GeoC}\left(X^{\prime} \mid X\right)\right)$ in the $\left(X, X^{\prime}\right)$ data scores how much additional (other than $X$ ) information is needed to describe the $X^{\prime}$ variable. Similarly, the dimen-
sion $\operatorname{GeoC}\left(X^{\prime} \mid X, Y\right)$ in the $\left(X, Y, X^{\prime}\right)$ data describes how much additional (other than $X, Y)$ information is needed to define $X^{\prime}$. However, when the number of data points $N \rightarrow \infty$, the value $\operatorname{GeoC}_{y \rightarrow x}$ is not negative (equal to the dimension of $X$ data). Thus, theoretically, GeoC identifies a causality in the geometric sense we have been describing.

### 4.2 Results and Discussion

Now, we present specific examples to contrast the transfer entropy with our proposed geometric measure to further highlight the role of geometry in such questions. Table 4.1 provides a summary of our numerical results. We use synthetic examples with known underlining dynamics to understand the accuracy of our model. Calculating transfer entropy has theoretical and numerical issues (see chapter. 2) for those chosen examples while our geometric approach accurately identifies the causation. We use the correlation dimension of the data because data might be fractals. Using a Hénon map example, we demonstrate that fractal data will not affect our calculations. Furthermore, we use a real-world application that has a positive transfer entropy to explain our data-driven geometric method. Details of these examples can be found in the following subsections.

Table 4.1: Summary of the results. Here, we experiment our new approach by synthetics and real world application data.

| Data | Transfer Entropy (Section 2.2.2) | Geometric Approach |
| :---: | :---: | :---: |
| $\begin{aligned} & \text { Synthetic: } \quad \mathrm{f}(\mathrm{x}, \mathrm{y})=a X+ \\ & b Y+C, a, b, c \in \mathbb{R} \end{aligned}$ | Theoretical issues can be noticed. Numerical estimation have boundedness issues when $b \ll$ 1. | Successfully identify the causation for all the cases ( $100 \%$ ). |
| Synthetic: $\mathrm{f}(\mathrm{x}, \mathrm{y})=a g_{1}(\mathrm{X})+$ $b g_{2}(Y)+C, a, b, c \in \mathbb{R}$ | Theoretical issues can be noticed. Numerical estimation have boundedness issues when $b \ll$ 1. | Successfully identify the causation for all the cases ( $100 \%$ ). |
| Hénon map: use data set invariant under the map. | special case of $a X^{2}+b Y+C$ with $a=-1.4, b=c=1$. Estimated transfer entropy is positive. | Successfully identify the causation. |
| Application: heart rate vs. breathing rate | Positive transfer entropy. | Identify positive causation. It also provides more details about the data. |

### 4.2.1 Synthetic Data: $X^{\prime}=g(X)+b Y+C$

Now, we focus on quantifying the geometric information flow by comparing dimensionalities of the outcomes' spaces. We will contrast this to the transfer entropy computations for a few examples of the form $X^{\prime}=g(X)+b Y+C$.

To illustrate the idea of geometric information flow, let us first consider a simple example, $x^{\prime}=a x+b y+c$. If $b=0$, we have $x^{\prime}=f(x)$ and, when $b \neq 0$, we have the $x^{\prime}=f(x, y)$ case. Therefore, dimensionality of the data set $\left(x^{\prime}, x\right)$ will change with parameter $b$ (see Figure 4.2). When the number of data points $N \rightarrow \infty$ and $b \neq 0$, then $\operatorname{GeoC}_{y \rightarrow x} \rightarrow 1$. Generally, this measure of causality depends on the
value of $b$, but also the initial density of initial conditions.
In this example, we contrast theoretical solutions with the numerically estimated solutions (Figure 4.1). Theoretically, we expect

$$
T_{y \rightarrow x}=\left\{\begin{array}{ll}
0 & ; b=0 \\
\infty & ; b \neq 0
\end{array} \text { as } \quad N \rightarrow \infty .\right.
$$

In addition, the transfer entropy for noisy data can be calculated by Equation (2.26).


Figure 4.1: Geometric information flow vs. Transfer entropy for $X^{\prime}=b Y$ data.


Figure 4.2: Manifold of the data $\left(x^{\prime}, x\right)$ with $x^{\prime}=b y$ and $y$ is uniformly distributed in the interval $[0,1]$. Notice that, when (a) $b=0$, we have a 1D manifold, (b) $b \neq 0$ we have 2D manifold, in the $\left(x^{\prime}, x\right)$ plane.

### 4.2.2 Synthetic Data: $X^{\prime}=a X+b Y$ with $a \neq 0$

The role of the initial density of points in the domain plays an important role in how the specific information flow values are computed depending on the measure used. To illustrate this point, consider the example of a unit square, $[0,1]^{2}$, that is uniformly sampled, and mapped by

$$
\begin{equation*}
X^{\prime}=a X+b Y, \text { with } a \neq 0 \tag{4.3}
\end{equation*}
$$

This fits our basic premise that $\left(x, y, x^{\prime}\right)$ data embeds in a 2D manifold, by ansatz of Equations (2.1) and (4.3), assuming for this example that each of $x, y$ and $x^{\prime}$ are scalar. As the number of data point grows, $N \rightarrow \infty$, we can see that

$$
G e o C_{y \rightarrow x}= \begin{cases}0 & ; b=0 \\ 1 & ; b \neq 0\end{cases}
$$

because $\left(X, X^{\prime}\right)$ data are on 2D manifold iff $b \neq 0$ (numerical estimation can be seen in Figure 4.3p). On the other hand, the conditional entropy $h\left(X^{\prime} \mid X, Y\right)$ is not defined, becoming unbounded when defined by noisy data. Thus, it follows that transfer entropy shares this same property. In other words, boundedness of transfer entropy depends highly on the $X^{\prime} \mid X, Y$ conditional data structure, while, instead, our geometric information flow measure highly depends on $X^{\prime} \mid X$ conditional data structure. Figure 4.3k demonstrates this observation with estimated transfer entropy and analytically computed values for noisy data. The slow convergence can be observed, Equation (2.26), Figure 2.2.


Figure 4.3: (a) shows the geometric information flow and (b) represents the Transfer entropy for $x^{\prime}=x+b y$ data. The figures show the changes with parameter $b$. We can notice that the transfer entropy has similar behavior to the geometric information flow of the data.

### 4.2.3 Synthetic Data: Nonlinear Cases

Now, consider the Hénon map,

$$
\begin{align*}
& x^{\prime}=1-1.4 x^{2}+y  \tag{4.4}\\
& y^{\prime}=x
\end{align*}
$$

as a special case of a general quadratic relationship, $x^{\prime}=a x+b y^{2}+c$, for discussing how $x^{\prime}$ may depend on $(x, y) \in \Omega_{1}$. Again, we do not worry here if $y^{\prime}$ may or may not depend on $x$ and or $y$ when deciding dependencies for $x^{\prime}$. We will discuss two cases, depending on how the $(x, y) \in \Omega_{1}$ data are distributed. For the first case, assume $(x, y)$ is uniformly distributed in the square, $[-1.5,1.5]^{2}$. The second and dynamically more realistic case will assume that $(x, y)$ lies on the invariant set (the strange attractor) of the Hénon map. The geometric information flow is shown for both cases in Figure 4.4. We numerically estimate the transfer entropy for both cases, which gives $T_{y \rightarrow x}=2.4116$ and 0.7942 , respectively. (However, recall that the first case for transfer entropy might not be finite analytically,
and there is slow numerical estimation-see Table 4.2).
Table 4.2: Hénon Map Results. Contrasting geometric information flow versus transfer entropy in two different cases, 1st relative to uniform distribution of initial conditions (reset each time) and 2nd relative to the natural invariant measure (more realistic).

| Domain | GeoC | $T_{y \rightarrow x}$ |
| :---: | :---: | :---: |
| $[-1.5,1.5]^{2}$ | 0.90 | 2.4116 |
| Invariant Set | 0.2712 | 0.7942 |


(a) $\left(x, y, x^{\prime}\right)$ data for Hénon Map.


Figure 4.4: Consider the Hénon map, Equation $\left(4.4\right.$, within the domain $[-1.5,1.5]^{2}$ and the invariant set of Hénon map. (a) the uniform distribution case (green) as well as the natural invariant measure of the attractor (blue) are shown regarding the $\left(x, y, x^{\prime}\right)$ data for both cases; $(\mathbf{b})$ when $(x, y) \in[-1.5,1.5]^{2}$, notice that $\operatorname{GeoC}_{y \rightarrow x}=0.9$, and (c) if $(x, y)$ is in an invariant set of Hénon map, then GeoC $y_{y \rightarrow x}=0.2712$.

### 4.2.4 Application Data

Moving beyond bench-marking with synthetic data, we will contrast two measures of information flow in a real world experimental data set. Consider heart rate $\left(x_{n}\right)$ vs. breathing rate $\left(y_{n}\right)$ data (Figure 4.5) as published in [22, 37], consisting of 5000 samples. Correlation dimension of the data $X$ is $D_{2}(X)=1.00$, and $D_{2}\left(X, X^{\prime}\right)=1.8319>D_{2}(X)$. Therefore, $X^{\prime}=X_{n+1}$ depends not only on $x$, but also on an extra variable (Theorem 4.1.1). In addition, correlation dimension of the data $(X, Y)$ and $\left(X, Y, X^{\prime}\right)$ is computed $D_{2}(X, Y)=1.9801$ and $D_{2}\left(X, Y, X^{\prime}\right)=2.7693>D_{2}(X, Y)$, respectively. We conclude that $X^{\prime}$ depends on extra variable(s) other that $(x, y)$ (Theorem 4.1.1) and the correlation dimension geometric information flow, $\mathrm{GeoC}_{y \rightarrow x}=0.0427$, is computed by Equations 4.2) and (4.1). Therefore, this suggests the conclusion that there is a causal inference from breathing rate to heart rate. Since breathing rate and heart rate share the same units, the quantity measured by geometric information flow can be described without normalizing. Transfer entropy as estimated by the KSG method ([28]) with parameter $k=30$ is $T_{y \rightarrow x}=0.0485$, interestingly relatively close to the GeoC value. In summary, both measures for causality $(G e o C, T)$ are either zero or positive together. It follows that there exists a causal inference.

Table 4.3: Heart rate vs. breathing rate data-contrasting geometric information flow versus transfer entropy in breath rate to heart rate.

$$
\begin{array}{cc}
\hline \text { GeoC }_{y \rightarrow x} & T_{y \rightarrow x} \\
\hline 0.0427 & 0.0485
\end{array}
$$



Figure 4.5: Result for heart rate $\left(x_{n}\right)(\mathbf{a}, \mathbf{c})$ vs. breathing rate $\left(y_{n}\right)$ data $(\mathbf{b}, \mathbf{d})$. The top row is the scatter plot of the data, and the second row represents the dimension of the data.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616.
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309 . DOI: 10.1063/1.5031109, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https: //www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138, URL: https://link.aps . org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL: https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079, eprint: https://science.sciencemag.org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag. org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi. org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991.
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 5

## Dynamic Mode Decomposition on Projected Space

This chapter and the next are focusing on the understanding the global dynamics by observable time series data. To this purpose, we will study the numerical estimator DMD for Koopman operator. Furthermore we will generalize the current DMD algorithm as a operator on projected observable space. That generalization will lead to introduce the random projection methods into this field and we will develop a computationally efficient randomized DMD algorithm.

In this chapter we will discuss the background theory of DMD operator. Then we will focus on estimating the spectrum of the Koopman operator by DMD. Finally we will generalize the current DMD algorithm which compute the Koopman spectrum on projected observable space.

### 5.1 Koopman Operator

Since DMD approximate the eigenpairs of the Koopman operator, we will first review the underlining Koopman operator theory. Consider a discrete-time dynamical system

$$
\begin{equation*}
x_{n+1}=S\left(x_{n}\right) \tag{5.1}
\end{equation*}
$$

where $S: \mathcal{M} \rightarrow \mathcal{M}$ and $\mathcal{M}$ is a finite dimensional manifold. (If we have a differential equation or continuous time dynamical system, the flow map can be considered.) The variable $x$ is often recognized as a state variable and $\mathcal{M}$ as phase space. The associated Koopman operator is described as the evaluation of observable functions (Figure 5.1) $\psi: \mathcal{M} \rightarrow \mathbb{R}$ in function space $\mathcal{F}$. Instead of analyzing the individual trajectories in phase space, the Koopman operator operates on the observations [10, 26, 38, 50].


Figure 5.1: This figure shows the behavior of the Koopman operator $\mathcal{K}$ in observable space $\mathcal{F}$ associated with a dynamical system $S$. The Koopman operator evaluates the observable $\psi$ at downstream or future $x^{\prime}=x_{n+1}=s\left(x_{n}\right)$.

Definition 5.1.1 (Koopman operator [26]). The Koopman operator $\mathcal{K}$ for a map $S$ is defined as the following composition,

$$
\begin{array}{r}
\mathcal{K}: \mathcal{F} \rightarrow \mathcal{F} \\
\psi \mapsto \mathcal{K}[\psi]=\psi \circ S \tag{5.2}
\end{array}
$$

on the function space $\mathcal{F}$.

It is straightforward, to prove [26],

$$
\begin{equation*}
\mathcal{K}\left[a \psi_{1}+b \psi_{2}\right]=a\left(\psi_{1} \circ S\right)+b\left(\psi_{2} \circ S\right)=a \mathcal{K}\left[\psi_{1}\right]+b \mathcal{K}\left[\psi_{2}\right] \tag{5.3}
\end{equation*}
$$

for $\psi_{1}, \psi_{2} \in \mathcal{F}$ and $a, b \in \mathbb{C}$ and, therefore, the Koopman operator is linear on $\mathcal{F}$. This is an interesting and important property of the operator because the associated map $S$ most probably will be non-linear. Even though the operator is associated with a map that evolves in a finite dimensional space, $\mathcal{F}$ the function space in which the operator acts on could possibly be an infinite dimensional. This is the trade-off between costs for the linearity [10].

Spectral analysis of the Koopman operator can be used to decompose the dynamics, which is the key success in the DMD. Assuming the spectrum of the Koopman operator $\mathcal{K}$ is given by

$$
\begin{equation*}
\mathcal{K} \psi_{i}(x)=\lambda_{i} \psi_{i}(x) \quad i=1,2,3, \ldots \tag{5.4}
\end{equation*}
$$

then vector-valued observables $\boldsymbol{g}: \mathcal{M} \rightarrow \mathbb{R}^{N}$ (or $\mathbb{C}^{N}$ ) can be represented by

$$
\begin{equation*}
\boldsymbol{g}(\boldsymbol{x})=\sum_{i=1}^{\infty} \psi_{i}(\boldsymbol{x}) \boldsymbol{\phi}_{i} \tag{5.5}
\end{equation*}
$$

where $\phi_{i} \in \mathbb{R}^{N}$ (or $\mathbb{C}^{N}$ ) are the vector coefficients of the expansion and called "Koopman modes"(here we assumed that components of $g$ lie within the span of the eigenfunctions of $\mathcal{K}$ ). Note that the observable value at time $n+1$ is given by

$$
\begin{equation*}
\boldsymbol{g}\left(\boldsymbol{x}_{\boldsymbol{n}+\mathbf{1}}\right)=\sum_{i=1}^{\infty} \lambda_{i}^{n} \psi_{i}\left(\boldsymbol{x}_{\mathbf{0}}\right) \boldsymbol{\phi}_{\boldsymbol{i}} . \tag{5.6}
\end{equation*}
$$

This decomposition can be used to separate the spacial and time components of the dynamical system and can be used to isolate the specific dynamics.

### 5.2 Dynamic Mode Decomposition

The dynamic mode decomposition is a data-driven method to estimate the Koopman modes from numerical or experimental data [38]. Suppose dynamics are governed by Equation (5.1) for any state $\boldsymbol{x}$ and vector valued measurements are given by observable $g(x) \in \mathbb{R}^{N}$. For a given set of data

$$
\begin{aligned}
& X=\left[\begin{array}{llll}
g\left(x_{0}\right) & g\left(x_{1}\right) & \ldots & g\left(x_{M-1}\right)
\end{array}\right], \\
& Y=\left[\begin{array}{llll}
y_{0} & y_{1} & \ldots & y_{M-1}
\end{array}\right]
\end{aligned}
$$

where $y_{i}=\boldsymbol{g}\left(\boldsymbol{s}\left(\boldsymbol{x}_{\boldsymbol{i}}\right)\right)$, the Koopman modes and eigenvalues of the Koopman operator can be estimated through solving the least-squares problem

$$
\begin{equation*}
\mathbb{K}=\underset{K}{\arg \min }\|K X-Y\|_{F}^{2}=\underset{K}{\arg \min } \sum_{i=0}^{M-1}\left\|K g\left(\boldsymbol{x}_{\boldsymbol{i}}\right)-\boldsymbol{y}_{\boldsymbol{i}}\right\|_{2}^{2} \tag{5.7}
\end{equation*}
$$

and $\mathbb{K}=Y X^{\dagger}$ (here $X^{\dagger}$ is the pseudo-inverse of $X$ ) is defined as the "Exact DMD" operator [50]. The eigenvalue $(\hat{\lambda})$ of $\mathbb{K}$ is an approximation of an eigenvalue $(\lambda)$ of $\mathcal{K}$; the corresponding right eigenvector $(\hat{\phi})$ is called the DMD mode and approximates the Koopman mode $(\phi)$. Then the observable value $\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t}))$ at time $t$ can be modeled as

$$
\begin{equation*}
\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t}))=\sum_{i=1}^{r} \psi_{i}\left(\boldsymbol{x}_{\mathbf{0}}\right) \hat{\boldsymbol{\phi}}_{i} \hat{\lambda}_{i}^{t} \tag{5.8}
\end{equation*}
$$

where $r$ is the number of selected DMD modes and demonstrates the finite dimensional approximation for vector-valued observable $\boldsymbol{g}$ under the Koopman operator.

Based on this decomposition, data matrices can be expressed as

$$
\begin{align*}
X_{N \times M} & =\Phi_{N \times r} T_{r \times M}  \tag{5.9}\\
Y_{N \times M} & =\Phi_{N \times r} \Lambda_{r \times r} T_{r \times M}
\end{align*}
$$

where $\Phi=\left[\psi_{1}\left(x_{0}\right) \boldsymbol{\phi}_{1} \psi_{2}\left(x_{0}\right) \boldsymbol{\phi}_{2} \ldots \psi_{r}\left(x_{0}\right) \boldsymbol{\phi}_{r}\right], T$ is a Vandermonde matrix with $T_{i j}=\lambda_{i}^{j-1}$ for $i=1,2, \ldots r, j=1,2, \ldots, M$ and $\Lambda=\operatorname{diag}\left\{\lambda_{1}, \lambda_{2}, \ldots, \lambda_{r}\right\}$. Note that with the above decomposition $\mathbb{K}=Y X^{\dagger}=\Phi \Lambda T T^{\dagger} \Phi^{\dagger}$. We will suppose $\mathbb{K}$ has distinct eigenvalues $\lambda_{i}$, columns of $X$ are linearly independent and $r \leq$ $M$. In practical applications, we are expected to fully understand the data set by relatively few ( $r \ll M$ ) modes. This can be considered one of the dimension reduction steps of the algorithm. Additionally, the dimension of columns of the data matrix need to be reduced.

In practice, the columns of data matrix $X$ (and $Y$ ) are constructed by the snapshot matrices of spatial observable data. More often, those snapshots lie in a high-dimensional space $\mathbb{R}^{N}\left(N \gg 1\right.$ and roughly $\mathcal{O}\left(10^{15}\right)$ to $\mathcal{O}\left(10^{10}\right)$ ), but the number of snapshots or time steps $(M)$ are small and often it is $\mathcal{O}\left(10^{3}\right)$ to $\mathcal{O}\left(10^{1}\right)$ [12]. Hence, computing the spectrum of matrix $\mathbb{K}$ by direct SVD is computationally intensive, even though most of the eigenvalues will be zero. Method-ofsnapshot, parallel version of SVD, or randomized SVD can be used to attack this difficulty [32, 34]. In this project, we use a more simple randomized method by generalizing the DMD algorithm. We can project our data matrices $X, Y$ into a low-dimensional space $R^{L}$ with $r \leq L \leq M \ll N$; therefore, we need to estimate the spectrum of $\mathbb{K}$ based on the computation on the projected space. Our proposed randomized DMD method is focused on this dimension reduction step.

### 5.2.1 DMD on Projected Space

Computational and storage costs of DMD can be reduced by projecting data into a low-dimensional observable space. Let $P \in \mathbb{R}^{L \times N}$ be any rank $L$ projection matrix, then dimension of data matrices $X, Y \in \mathbb{R}^{N \times M}$ can be reduced to $L \times M$ by the projection $X_{L}=P X, Y_{L}=P Y$. The DMD operator on the projected space (see Figure 5.2) is given by,

$$
\begin{equation*}
\hat{\mathbb{K}}=\underset{K \in \mathbb{R}^{L \times L}}{\arg \min }\left\|K X_{L}-Y_{L}\right\|_{F}^{2}=\underset{K \in \mathbb{R}^{L \times L}}{\arg \min }\|K P X-P Y\|_{F}^{2} \tag{5.10}
\end{equation*}
$$

and $\hat{\mathbb{K}}=P Y(P X)^{\dagger}$. Therefore

$$
\begin{equation*}
\hat{\mathbb{K}}=P Y(P X)^{\dagger}=P Y X^{\dagger} P^{\dagger}=P \mathbb{K} P^{\dagger} \tag{5.11}
\end{equation*}
$$

where $\mathbb{K}=Y X^{\dagger}$ is the DMD operator on the original space.


Figure 5.2: DMD operator on the projected space. This figure shows the relationship between the DMD operator on original space and DMD on projected space. The operator $\hat{K}$ on the projected space is defined in Equation 5.10 and can be calculated by Equation (5.10).

Proposition 5.2.0. Some eigenpairs $(\lambda, \phi)$ of $\mathbb{K}$ can be obtain by $\left(\lambda_{L}, \phi_{L}\right)$ of projected $D M D \mathbb{K}$ with $\lambda=\lambda_{L}$ and $\phi=P^{\dagger} \phi_{L}$.

Proof. Let $\left(\lambda_{L}, \phi_{L}\right)$ be an eigenpair of $\hat{\mathbb{K}}$. Then $\hat{\mathbb{K}} \phi_{L}=\lambda_{L} \phi_{L}$ and by Equation (5.11), $P \mathbb{K} P^{\dagger} \phi_{L}=\lambda_{L} \phi_{L}$. Now let $P^{\dagger} \phi_{L}=\phi$, then $\phi_{L}=P \phi$ because $P P^{\dagger}=I$. Hence, $P \mathbb{K} P^{\dagger} \phi_{L}=P \mathbb{K} \phi=\lambda_{L} P \phi$ and $P\left(\mathbb{K} \phi-\lambda_{L} \phi\right)=0$. Since $\mathbb{K} \phi-\lambda_{L} \phi=0$ is a solution to the above equation, $\lambda_{L}$ is an eigenvalue and the corresponding eigenvector is $\phi=P^{\dagger} \phi_{L}$ of $\mathbb{K}$.

In other words, we can lift up the dimension of eigenvectors in a projected space by $P^{\dagger}$ to obtain an eigenvector in the original data space. However to avoid the direct calculation of the pseudo-inverse of the projection matrix, we can calculate the eigenvector in the output space $Y_{L}$ of the DMD operator and lift up the
vector into the original output space $Y$. We can easily show that $\hat{\phi}=Y(P X)^{\dagger} \phi_{L}$ is an eigenvector (see equation (5.12)) of $\mathbb{K}$ for corresponding non-zero eigenvalues.

$$
\begin{align*}
\mathbb{K} \hat{\phi} & =\mathbb{K} Y(P X)^{\dagger} \phi_{L}=\mathbb{K} Y X^{\dagger} P^{\dagger} \phi_{L}  \tag{5.12}\\
& =\mathbb{K} \mathbb{K} P^{\dagger} \phi_{L} \\
& =\mathbb{K} \lambda P^{\dagger} \phi_{L} \\
& =\lambda Y X^{\dagger} P^{\dagger} \phi_{L} \\
& =\lambda Y(P X)^{\dagger} \phi_{L} \\
& =\lambda \hat{\phi} .
\end{align*}
$$

Moreover, notice, $P \hat{\phi}=P Y(P X)^{\dagger} \phi_{L}=\hat{\mathbb{K}} \phi_{L}$ and, therefore, $\hat{\phi}$ estimates the eigenvector on the output space $Y$. A detailed view of this lifting operator is shown in Figure 5.3. It provides the relationship of the lifting operator with the DMD operator acting on any general observable vector $z=\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t})) \in \mathbb{R}^{N}$.


Figure 5.3: The figure shows the projecting operator $P$ and DMD related lifting operator $Y(P X)^{\dagger}=\mathbb{K} P^{\dagger}$, which should be used in DMD algorithms. Instead of using $P^{\dagger}$ as the lifting operator, $Y(P X)^{\dagger}$ can be used for efficient calculations. Moreover, notice that $\hat{z}_{L}^{\prime}:=P \hat{z}^{\prime}=\hat{\mathbb{K}} z_{L}$.

Next, the focus moves to the spatial-temporal decomposition of the projected data matrices by spectrum of the DMD operator. Note that the observable value $\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t}))$ at time $t$ can be modeled as $\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t}))=\sum_{i=1}^{r} \psi_{i}\left(x_{\mathbf{0}}\right) P^{\dagger}\left(\boldsymbol{\phi}_{L}\right)_{i} \hat{\lambda}_{i}^{t}$ and similar to the Equation (5.9), data can be decomposed as

$$
\begin{align*}
X_{N \times M} & =P_{N \times L}^{\dagger} \tilde{\Phi}_{L \times r} T_{r \times M}  \tag{5.13}\\
Y_{N \times M} & =P_{N \times L}^{\dagger} \tilde{\Phi}_{L \times r} \Lambda_{r \times r} T_{r \times M} .
\end{align*}
$$

This decomposition leads to $\mathbb{K}=Y X^{\dagger}=P^{\dagger} \tilde{\Phi} \Lambda T T^{\dagger} \tilde{\Phi}^{\dagger} P$ and if $r \leq L$ all the nonzero eigenvalues and corresponding eigenvectors of $\mathbb{K}$ can be constructed by the projected DMD operator. Further, Equation (5.13) can be use to isolate the spatial profile of interesting dynamical compotes, such as attractors, periodic behaviors, etc.

Based on the choice of the projection matrix, we have alternative ways to estimate the spectrum of the DMD operator.

Remark 1 (Projection by SVD). A commonly used projection matrix is based on SVD of the input matrix $X=U \Sigma V^{*}$ and the projection matrix is chosen to be $P=U^{*}$, here, * represents the conjugate transpose of a matrix. Using Equation (5.11) and SVD of X, the operator on the projected space can be formulated as $\hat{\mathbb{K}}=U^{*} Y V \Sigma^{-1}$.

Remark 2 (Standard DMD and Exact DMD). Let eigenpair of an SVD-based $\hat{\mathbb{K}}=$ $U^{*} Y V \Sigma^{-1}$ be given by $\left(\lambda, \phi_{L}\right)$. In a standard DMD (Reference Schmid's paper and Tu's paper) use the eigenvector $P^{\dagger} \phi_{L}=U \phi_{L}$ to estimate eigenvectors of $\mathbb{K}$. On the other hand, in the exact DMD (reference Tu's paper) this eigenvector is estimated by $Y(P X)^{\dagger} \phi_{L}=$ $Y V \Sigma^{-1} \phi_{L}$.

Remark 3. One can also use $Q R$ decomposition-based projection methods. $Q R$ decomposition of the input-output snapshot data matrix $[X Y]$ is used in some existing methods [50], and so our philosophy of random projection methods used here will greatly improve efficiency, but, nonetheless, with quality that is controlled in an analytically rigorous way.

In this study, we propose a simple random projection-based method to estimate the spectrum of the DMD operator.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309 . DOI: 10.1063/1.5031109, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: 10.1103 / PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi . org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https ://www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138, URL: https://link.aps . org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL: https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079, eprint: https://science.sciencemag.org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag . org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi . org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 6

## Randomized Projection Learning

## Method for Dynamic Mode

## Decomposition

In this chapter, we will discuss more details about the estimation of Koopman spectrum using a simple randomized DMD algorithm. Note that the current methods are based on the singular value decomposition of the data matrix $X$ to construct a projection, and our proposed algorithm is based on the random projection method to project data into a low-dimensional space.

### 6.1 Randomized Dynamic Mode Decomposition

Last chapter(7), we generalized currently used DMD algorithms as an operator on projected input and output observable spaces. Now, we discuss the simplest possible yet powerful projector namely random projection which can be used to project data into a low-dimensional space.

### 6.1.1 Random Projection

The random projection method is based on the Johnson-Lindenstrauss lemma, which can be seen as a direct consequence of the consentration of measure phenomenon [14]. Furthermore, this results is at the heart of many data science applications [23]. In this thesis, we use a random matrix $R$ generated by a Gaussian distribution, such that each element $r_{i j} \stackrel{i i d}{\sim} \mathcal{N}(0,1)$ with a normalize column is of unit length.

Theorem 6.1.1 (Johnson-Lindenstrauss lemma [24]). For any $0<\epsilon<1$ and any integer $M>1$, let $L$ be a positive integer, such that $L \geq L_{0}$ with $L_{0}=\frac{C \ln M}{\epsilon^{2}}$, where $C$ is a suitable constant ( $C \approx 8$ in practice, $C=2$ is good enough). Then for any set $X$ of $M$ data points in $\mathbb{R}^{N}$, there exists a map $f: \mathbb{R}^{N} \rightarrow \mathbb{R}^{L}$ such that for all $x_{1}, x_{2} \in X$,

$$
(1-\epsilon)\left\|x_{1}-x_{2}\right\|^{2} \leq\left\|f\left(x_{1}\right)-f\left(x_{2}\right)\right\|^{2} \leq(1+\epsilon)\left\|x_{1}-x_{2}\right\|^{2}
$$

Theorem 6.1.2 (Random Projection [13]). For any $0<\epsilon, \delta<\frac{1}{2}$ and positive integer $N$, there exists a random matrix $R$ of size $L \times N$ such that for $L \geq L_{0}$ with $L_{0}=\frac{C \ln (1 / \delta)}{\epsilon^{2}}$. and for any unit-length vector $x \in R^{N}$

$$
\operatorname{Pr}\left\{\left\|\|R x\|^{2}-1 \mid>\epsilon\right\} \leq \delta\right.
$$

or

$$
\operatorname{Pr}\left\{\left|\|R x\|^{2}-1\right|>\epsilon\right\} \leq e^{-C L \epsilon^{2}}
$$

A low rank approximation for both $X, Y$ can be found using the random projection method. Notice that both these matrices have $M$ points from $N$ dimensional observable space and, therefore, we can use random projection matrix $R$ of size
$L \times N$ with the $L \geq \frac{C \ln N}{\epsilon^{2}}$, which provides $\epsilon-$ isometry to $\mathbb{R}^{N}$. (See Figure 5.2 for details).

### 6.1.2 Random Projection for DMD

Our suggested randomized dynamic mode decomposition(rDMD) is based on the random projection applied to the theory of DMD on a projected space. We can reduce the dimension of the data matrix $X, Y$ in DMD by using a random projection matrix $R_{L \times N}$. In other words, we construct a projection matrix $P$ discussed in Section 5.2 .1 as a random matrix $R$ whose columns have unit lengths and entries that are selected independently and identically from a probability distribution. Therefore, the rDMD matrix on the projected space is given by $\hat{\mathbb{K}}=R Y(R X)^{\dagger}$, and if an eigenpair of $\hat{\mathbb{K}}$ is given by $\left(\lambda, \phi_{L}\right)$, then the eigenpair of $\mathbb{K}$ is given $\left(\lambda, Y(R X)^{\dagger} \phi_{L}\right)$. Algorithm 1 represents the major steps needed to estimate the eigenvalues and corresponding eigenvectors of the DMD operator with the random projection method.In addition, Figure 6.1 shows the details of the inputoutput variables of the algorithm, spatiotemporal decomposition of the data, and how to use the eigendecomposition of the Koopman operator to isolate and interpret the spatial features of a dynamical system.

The calculation of the projection matrix of a standard or exact DMD algorithm based on the SVD of the snapshot matrix $X$ is needed to store a full high-resolution data matrix, which leads to memory issues. Our proposed rDMD algorithm can avoid these storage issues, because low-dimensional matrices $X_{L}, Y_{L}$ obtained by matrix multiplications only need to store one row and one column of each matrix at a time. Additionally, this algorithm reduces the computational cost, since we only need to calculate the pseudo-inverse of comparatively lower dimensional matrix. The choice of the distribution of $R$ can further reduce the computational cost [1].

Furthermore, one time step forecasting error for any given snapshot by using the rDMD algorithm can be bounded by using the JL theory (see theorem 6.1.3).

Algorithm 1 Randomized DMD (rDMD). Figure 6.1 shows the details of the input and output variable.
Data: $X, Y \in \mathbb{R}^{N \times M}$
Input: $\epsilon$
$L_{0}=\frac{C \ln M}{\epsilon^{2}}$;
Choose $L$ such that $L \geq L_{0}$;
Construct a random matrix $R=\frac{1}{\sqrt{L}}\left(r_{i j}\right) \in \mathbb{R}^{L \times N}$ such that $r_{i j} \sim \mathcal{N}(0,1)$;
Calculate $X_{L}:=R X, Y_{L}:=R Y$;
Calculate $\hat{\mathbb{K}}=Y_{L} X_{L}^{+}$;
$\left[\Lambda \Phi_{L}\right]=\operatorname{eigs}(\hat{\mathbb{K}})$;
Result: $\operatorname{diag}(\Lambda), Y X_{L}^{\dagger} \Phi_{L}$


Figure 6.1: The figure summarize the rDMD algorithm, its input data, and the output variables. This also explains how to use the eigendecomposition of the Koopman operator to isolate and interpret the spatial features of a dynamical system. This figure is new from the previous manuscript.

Theorem 6.1.3 (Error Bound). Let $z=\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t})), z^{\prime}=\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{t}+\mathbf{1})) \in \mathbb{R}^{N}$. Error bound of estimating $z^{\prime}$ by using the $r D M D$ as ${z^{\prime}}^{\prime}=Y X^{\dagger} R^{\dagger} R z$ is given by

$$
\begin{equation*}
E\left(z^{\prime} ; L\right):=\left\|z^{\prime}-\hat{z}^{\prime}\right\| \leq \frac{\left\|R z^{\prime}-\hat{\mathbb{K}} R z\right\|}{1-\epsilon}:=U B \tag{6.1}
\end{equation*}
$$

with at least the probability of $\mathcal{O}\left(1 / M^{2}\right)$ for any $0<\epsilon<1$ with $L>\frac{C \log (M)}{\epsilon^{2}}$.
Proof. Since $\hat{\mathbb{K}}=R Y X^{\dagger} R^{\dagger}$, the rDMD acts on the projected vector, which can be rearranged as $\hat{\mathbb{K}} R z=R \hat{z}^{\prime}$. Therefore,

$$
\left\|R z^{\prime}-\hat{\mathbb{K}} R z\right\|=\left\|R z^{\prime}-R \hat{z}^{\prime}\right\|
$$

Now we can apply the JL theory to attain the desired error bound.

$$
(1-\epsilon)\left\|z^{\prime}-\hat{z}^{\prime}\right\| \leq\left\|R z^{\prime}-R \hat{z}^{\prime}\right\|=\left\|R z^{\prime}-\hat{\mathbb{K}} R z\right\| .
$$

Hence $\left\|z^{\prime}-\hat{z}^{\prime}\right\| \leq \frac{\left\|R z^{\prime}-\hat{\mathbb{K}} R z\right\|}{1-\epsilon}$.

### 6.2 Results and Discussion

In this section, we demonstrate the theory of rDMD with a few examples. The first two examples consider the computation for known dynamics and demonstrate the error analysis. The final example demonstrates application in the field of oceanography and isolates the interesting features by rDMD, compering the resulting modes with the exact DMD results.

### 6.2.1 Logistic Map

We first consider a dataset of 300 snapshots from a logistic map,

$$
x_{n+1}=a x_{n}\left(1-x_{n}\right)
$$

with $a=3.56994$. In this case, all initial conditions will converge to a period256 orbit. Therefore the rank of the snapshot matrix with relatively high samples
should be 256 . We forecasted the data by using the rDMD method and then analyzed the error of the prediction and compared it with the theoretical upper bound. With $N=5000$ initial conditions and $M=300$ samples, the dimension $L$ of the projecting space can be chosen as $L \geq \frac{C \ln (300)}{\epsilon^{2}} \approx \frac{34.22}{\epsilon^{2}}$ when $C=6$. rDMD with projection into a 50 dimensional space can accurately forecast the time series data. (Figure 6.2 shows the original vs. predicted data for one trajectory.) Furthermore, Figure 6.3 demonstrates the bound of the error of the forecast explained in Equation (6.1) and how the error relates to the distortion parameter $\epsilon$ (Figure 6.3a) and the dimension of the projected space (Figure 6.3b). Since the rank of the snapshot matrix is 256 , any $L \geq 256$ will perform very accurately. This example validates the error bound we discussed in Equation (6.1) and the error of the prediction depends on the error exhibited by the projected DMD operator and the distortion parameter ( $\epsilon$ or the projected dimension) from the JL theory.


Figure 6.2: (a) Shows the predicted data using the rDMD algorithm with projected dimension $L=50$ compared to the original data from logistic map $x_{n+1}=$ $3.56994 x_{n}\left(1-x_{n}\right)$ for initial condition $x(0)=0.4967$. Further, (b) shows the estimated error and theoretical upper bounds (Equation (6.1)) for some projected dimension $L$, and this example validates the theoretical bound.


Figure 6.3: This shows the prediction error of the logistic map $x_{n+1}=$ $3.56994 x_{n}\left(1-x_{n}\right)$ by rDMD and its theoretical upper bounds(Equation 6.1)). (a) Represents the error with respect to the distortion $\epsilon$ and (b) shows the error with the dimension of the projected space that will guarantee the bound for this example.

### 6.2.2 Toy Example: Demonstrates the Variable Separation and Isolating Dynamics

To demonstrate the variable separation and to isolate the spatial structures based on the time dynamics, we consider a toy example (motivated by [29]),

$$
\begin{equation*}
z(x, t)=\sum_{j=1}^{20} j \operatorname{sech}(0.1 x+j) e^{i \gamma_{j} t}=\sum_{j=1}^{20} \Phi_{j}(x) T_{j}(t) \tag{6.2}
\end{equation*}
$$

where $\gamma_{j}$ 's are constants, and let $\Phi_{j}(x)=j \operatorname{sech}(0.1 x+j)$ and $T_{j}(t)=e^{i \gamma_{j} t}$.
Comparing this Equation (6.2) with decomposition Equation (5.8), the rDMD algorithm is expected to isolate 20 periodic modes by rDMD algorithm. The data set (snapshot matrix) for this problem is constructed by $N=20,000$ spatial grid points and $M=5001$ temporal grid points with $\gamma_{j}=j$ (see Figure 6.4). As discussed in the previous section, if $L \geq 20$ then those expected modes can be isolated
and there exist eigenvalues $\lambda_{j}$ of an rDMD operator, such that

$$
\omega_{j}=\ln \left(\lambda_{j}\right)=j i
$$

for $j=1,2, \ldots, 20$ (see Figure 6.5). Furthermore, we expect corresponding rDMD modes equal to spatial variables of the model, such that

$$
b_{j}\left(x_{0}\right) \phi_{j}(x)=j \operatorname{sech}(0.1 x+j)=\Phi_{j}(x)
$$



Figure 6.4: [-20]Original dataset constructed by Equation 6.2. (a) shows the $\operatorname{Re}(z(x, t))$ plot for all $x, t$ values at $20,000 \times 5000$ grid points. (b) represents the time series plot for two initial conditions. (c) provides the snapshots of few different time points. Our goal is separate and isolate spatial variables $\Phi_{j}(x)=$ $j \operatorname{sech}(0.1 x+j)$ and $T_{j}(t)=e^{i j t}$ from the given data constructed by $z(x, t)$.


Figure 6.5: (a) Shows the absolute error for estimated eigenvalues from rDMD and the exact DMD when the dimension of the projected space $\mathrm{L}=20$. (b) Shows the absolute error for the estimated $10^{\text {th }}$ DMD mode by rDMD and exact DMD methods. In this case, both methods have very accurate results and error is less than $10^{-10}$.

As expected, we noticed that the calculated modes have negligible error when the dimension of projected space $L \geq r=20$. Figure 6.5 shows the absolute error of eigenvalues and DMD modes. All modes behave similarly; here, we present mode 10 for demonstration purposes by the SVD-based exact DMD method and the random projection based rDMD method. Notice that errors of both methods are less than $10^{-10}$ when $L \geq r=20$.

Further, we examine the case when the projected dimension $L=17<r=20$ and compare the results of rDMD with the exact DMD. We notice that both methods demonstrate similar errors and rDMD is almost as good as the SVD projectionbased exact DMD (see Figures 6.6 and 6.7). When the number of actual modes $(r)$ is larger than the dimension of the projected space $(L)$, the projected DMD operator only estimates the $L$ number of modes, leading to both truncation errors and errors for eigenpair estimation based on the projected DMD operator. The $L<r$ case can
be modeled as,

$$
\begin{equation*}
z(x, t)=\sum_{j=1}^{L} \hat{b}_{j} \hat{\phi}_{j}(x) e^{\hat{\omega}_{j} i t}+E_{L+1} \tag{6.3}
\end{equation*}
$$

where $E_{L+1}=\sum_{L=j+1}^{m} b_{j} \phi_{j}(x) e^{\omega_{j} i t}$ is the truncated error that also affects the estimation process of eigenpairs. Therefore, if $L<r$, then there exists an error in eigenvalues and eigenvectors calculated by any method based on the projected DMD. However, this example demonstrates that rDMD can provide the results as good as the SVD projection-based method with very low computational costs (See Table 6.1.

| Method | Projected by | Computational Time(s) |
| :--- | :---: | :---: |
| Exact DMD | SVD | 521.09 |
| rDMD | Random Projection | 2.35 |

Table 6.1: Computational costs for the SVD-based exact DMD and random projection-based rDMD method for the data simulated by Equation (6.2). Computational costs of SVD for the high-dimensional snapshot matrix is relatively larger than random projection.


Figure 6.6: (a) Compares the eigenvalues $\omega_{j}=\ln \lambda_{j}$ calculated from rDMD (random projection $(R P)$ with $L=17$ ) and exact DMD (SVD projection with $L=17$ ) methods with the expected true values $\gamma_{j}=j i$. Here, $L=17<20$ is the dimension of the projected space. (b) Shows the absolute error for the estimated eigenvalues from rDMD and exact DMD.


Figure 6.7: (a) Compares the modes $b_{10} \phi_{10}(x)$ calculated from rDMD (random projection $(R P)$ with $\mathrm{L}=17$ ) and exact DMD (SVD projection with $\mathrm{L}=17$ ) methods with the expected true values $\Phi_{10}(x)=10 \operatorname{sech}(0.1 x+10)$. (b) Shows the absolute error for estimated values from rDMD and exact DMD.

### 6.2.3 Gulf of Mexico

In this example, we consider the data from HYbrid Coordinate Ocean Model (HYCOM) [21], which simulates the ocean data around the Gulf of Mexico. We used hourly surface velocity component $(u, v)$ with $1 / 25^{0}$ spatial resolution $(N=541 \times 347$ grid point data for 10 days ( 240 h and $M=239$ ). Understanding the dynamics from the oceanographic data is an interesting application of DMD because those dynamics can be decomposed by tidal constituents. Hence, we are expected to isolate the dynamics associated with the tidal period; in other words, the final DMD mode selection is based on the period $P_{i}=2 \pi / \operatorname{Im}\left(\ln \left(\lambda_{i}\right)\right)$ of the modes(see Table 6.2). We constructed the snapshot matrix

$$
X=\left[\begin{array}{l}
u  \tag{6.4}\\
v
\end{array}\right]
$$

by stacking the snapshots of velocity components $(u, v)$ in each column to perform the DMD analysis.

Figure 6.8 shows that most of the eigenvalues calculated from the SVD-based exact DMD and random projection-based rDMD are in agreement. Furthermore, eigenvalues that isolated the specific dynamics are almost equal. Additionally, Figures $6.9-6.11$ show the spacial profile of those modes from the exact DMD and rDMD methods. Moreover, each mode clearly isolate the interesting oceanographic features (see Table 6.2) and both methods provide almost the same spacial structures(see Figures 6.9 6.11) as expected.

Table 6.2: DMD modes for the Gulf of Mexico data set. Modes are selected based on the association to the tidal periods.

| Mode | Period (h) <br> DMD <br> rDMD | Associated Feature |  |
| :---: | :---: | :---: | :--- |
|  | $\infty$ | $\infty$ | Gulf stream around the GOM (see Figure 6.9) |
| 2 | 12.47 | 12.47 | Semi-diurnal tidal constituents (see Figure 6.10) |
| 3 | 23.85 | 24.56 | Diurnal tidal constituents (see Figure 6.11) |
| 4 | 6.08 | 6.07 | Second harmonic to semi-diurnal tidal constituents (see <br> Figure 6.11) |
| 5 | 4.16 | 4.17 | Third harmonic to semi-diurnal tidal constituents (see <br> Figure 6.11) |



Figure 6.8: Eigenvalues $\lambda_{i}$ calculated from the exact DMD and rDMD methods. (a) Full spectrum of the two methods with projected space dimension $L=239$ and (b) shows the first five modes. The mode selection is based on the comparison of the tidal periods with period of the DMD modes.


Figure 6.9: This figure compares the (a) DMD and (b) rDMD background mode identified by data from the Gulf of Mexico (GOM). This background mode captures the ocean current passing through the GOM.


Figure 6.10: This figure compares the (a) DMD and (b) rDMD mode associated with the M2 tidal frequency. This mode capture the "red tides".


Figure 6.11: (a)-(c) represent the exact DMD modes 3, 4, and 5 and (d)-(e) show the rDMD modes 3,4 , and 5 . Mode 3 is a diurnal mode with period 23.85 h for the exact DMD case and 24.56 h for the rDMD case. Modes 4 and 5 are associated with the second and third harmonic of semi-diurnal tidal constituents, respectively.

Notice that the dimension of the snapshot matrix is $375,454 \times 239$ and the SVD calculation of this matrix is more costly for both computation and storage. On the
other hand, random projection only performs by the matrix multiplication, which could be done at a relatively low cost. Hence, we achieve almost the same results by using the random projection method, at relatively lower computational and storage costs.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309. DOI: $10.1063 / 1.5031109$, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https : //www . hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10 . 1103/PhysRevE. 69.066138. URL: https://link.aps.org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL:https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett.85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079, eprint: https://science.sciencemag.org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag . org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi. org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991.
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Chapter 7

## Conclusion and Future Problems

In this thesis, I demonstrated the understanding global dynamics of the dynamical system by using the time series data. This dissertation specialized in two folds, namely identify the causation by geometric measures and reduce the computational cost of dynamic mode decomposition by using the random projection methods. In this chapter I will summarize the details of the each part and then I will discuss the problems which I will attempt to solve in the future.

### 7.1 Geometric Causal Inference

We have developed here a geometric interpretation of information flow as a causal inference as usually measured by a positive transfer entropy, $T_{y \rightarrow x}$. Our interpretation relates the dimensionality of an underlying manifold as projected into the outcome space and summarizes the information flow. Furthermore, the analysis behind our interpretation involves standard Pinsker's inequality that estimates entropy in terms of total variation, and, through this method, we can interpret the production of information flow in terms of details of the derivatives describing
relative orientation of the manifolds describing inputs and outputs (under certain simple assumptions).

A geometric description of causality allows for new and efficient computational methods for causality inference. Furthermore, this geometric perspective provides a different view of the problem and facilitates the richer understanding that complements the probabilistic descriptions. Causal inference is weaved strongly throughout many fields and the use of transfer entropy has been a popular black box tool for this endeavor. Our method can be used to reveal more details of the underling geometry of the data-set and provide a clear view of the causal inference. In addition, one can use the hybrid method of this geometric aspect and existing other methods in their applications.

We provided a theoretical explanation (part I: Mathematical proof of the geometric view of the problem) and numerical evidence (part 2: A data-driven approach for mathematical framework) of a geometric view for the causal inference. Our experiments are based on synthetic (toy problems) and practical data. In the case of synthetic data, the underlining dynamics of the data and the actual solution to the problem are known. For each of these toy problems, we consider a lot of cases by setting a few parameters. Our newly designed geometric approach can successfully capture these cases. One major problem may be if data describes a chaotic attractor. We prove theoretically (Theorem 4.1.1) and experimentally (by Hénon map example: in this toy problem, we also know actual causality) that correlation dimension serves to overcome this issue. Furthermore, we present a practical example based on heart rate vs. breathing rate variability, which was already shown to have positive transfer entropy, and here we relate this to show positive geometric causality.

Furthermore, we have pointed out that transfer entropy has analytic conver-
gence issues when future data $\left(X^{\prime}\right)$ are exactly a function of current input data $(X, Y)$ versus more generally $\left(X, Y, X^{\prime}\right)$. Therefore, referring to how the geometry of the data can be used to identify the causation of the time series data, we develop a new causality measurement based on a fractal measurement comparing inputs and outputs. Specifically, the correlation dimension is a useful and efficient way to define what we call correlation dimensional geometric information flow, $\mathrm{GeoC}_{y \rightarrow x}$. The $\mathrm{GeoC}_{y \rightarrow x}$ offers a strongly geometric interpretable result as a global picture of the information flow. We demonstrate the natural benefits of $G e o C_{y \rightarrow x}$ versus $T_{y \rightarrow x}$, in several synthetic examples where we can specifically control the geometric details, and then with a physiological example using heart and breathing data.

### 7.2 Randomized Dynamic Mode Decomposition

We also demonstrated that our rDMD can achieve very accurate results with lowdimensional data embedded in a high-dimensional observable space. Recent analytic technology from the concepts of high-dimensional geometry of data, and concentration of measure, have shown-surprising, if not initially intuitively-that even random projection methods can be quite powerful and capable. Here, in the setting of DMD methods approximating and projecting the action of a Koopman operator, we show that randomized projection can be developed and analyzed rigorously by the Johnson-Lindenstrauss theorem formalism, showing a powerful and simple approach.

We provided a theoretical framework and experimental results to address those issues raised from SVD-based methods by introducing our new rDMD algorithm. The theoretical framework is based on generalizing the SVD-based concept as a projection of high-dimensional data into a low-dimensional space. We proved that
eigenpairs of DMD in the original space can be estimated by using any rank $L$ projection matrix $P$. Being able to estimate eigenpairs allowed us to use the powerful and simple Johnson-Lindenstrauss lemma and the random projection method, allowing us to project data with matrix multiplication. Therefore, our proposed random projection-based DMD (rDMD) can estimate eigenpairs of the DMD operator with low storage and computational costs. Further, the error of the estimation can be controlled by choosing the dimension of the projected space; we demonstrated this error bound through the "logistic map" example.

DMD promises the separation of the spatial and time variables from data. Hence, we experimentally demonstrated how well the rDMD algorithm performed this task by a toy example. Notice that the number of those isolated modes ( $m$ ) are relatively (i.e., to spatial and temporal resolution) low in practical applications. If $m \ll M$, then the rank of the data matrix is much lower, and those eigenvalues and vectors of interest can be estimated accurately by projecting data into the much lower dimensional space $L \geq m$. The SVD projection-based exact DMD method still needs to calculate the SVD of a high-dimensional (roughly $10^{10} \times 10^{3}$ ) data matrix, while rDMD only requires multiplying the data matrix by a much lower-dimensional projection matrix. Furthermore, we noticed that both exact and random DMD methods experience similar errors. However random projection is much faster and needs less space for the calculations. We also demonstrate that practical applications provide similar results by using oceanographic data from the Gulf of Mexico.

Since the size of the DMD matrix is enormous in those applications (this could be roughly $10^{10} \times 10^{10}$ ), the eigenpairs must be estimated by projecting data into the low-dimensional space. Estimating eigenvalues and eigenvectors of a DMD operator using a high-dimensional snapshot data matrix (in applications, this could
be $10^{10} \times 10^{3}$ ) with existing SVD-based methods is expensive. The computational efficiency of the rDMD led to a new path of the current Koopman analysis. It allows using more observable variables in the data matrix without need for much extra computational power. Hence, state variables and more non-linear terms can be used in an analysis, with low costs, to improve the Koopman modes. The JL theory can be adopted further into the field of numerical methods of the Koopman theory. As a next step, we can use the random projection concept in the extended DMD and kernel DMD methods.

### 7.3 Future Work

### 7.3.1 Causal inference with multiple time series

We would like to identify the causal variable simultaneously from the pool of time series data by using geometric measures. The geometry of the data set will provide an easier calculation in the causality identification process and a better understanding of the causal variables. We consider the variable $X_{j}$ as interested observable which need to find the causal variables among the set of variables $V=\left\{X_{1}, X_{2}, \ldots, X_{m}\right\}$. The dynamical system $X_{j}$ can be written as a discrete time map

$$
\begin{equation*}
X_{j}^{\prime}=f\left(X_{1}, X_{2}, \ldots, X_{m}\right) \tag{7.1}
\end{equation*}
$$

However some of those variables might not be influence to the future state of $X_{j}$ and we need to identify the actual set of causal variables $S=\left\{X_{j_{1}}, \ldots, X_{j_{k}}\right\} \subseteq V$
for $X_{j}$. Thus we need to identify the variables $X_{j_{i}}$ such that

$$
\begin{equation*}
X_{j}^{\prime}=f\left(X_{j_{1}}, \ldots, X_{j_{k}}\right) \text { and } k \leq m \tag{7.2}
\end{equation*}
$$

In this case we call $X_{j_{i}}$ causes $X_{j}$ and can be represented using Fig. 7.1.). In this


Figure 7.1: Causal diagram or network representation of $X_{j_{i}}$ causes $X_{j}$.
project, we will only going to find a causal variables but not the function $f$. In other words we are showing there exits such a function. We will use the fractal correlation dimension to measure causation and find the causal diagram among the given set variables. All connections in an entire causal diagram with $m$ variable can be identified by $m^{2}$ comparisons of correlation dimensions.

We also intend to bring another special data analysis manifold learning tool "diffusion map" to investigate the causal relationship among the given time seties data. This method will be able to use in non-fractal data sets and it can be incorporate to find the exact causal relationship $(f)$ between variables just using the data.

### 7.3.2 Learning Transfer Operators by Kernel Density Estimation

To infer transfer operators from data is usually take as a classical problem that hinges on the Ulam method. The usual description is in terms of projection onto basis functions that are characteristic functions supported over a fine grid of rectangles, that we have previously called the Ulam-Galerkin method when taken in terms of finite time. We describe that the same problem can be understood by statistical density estimation formalism. In these terms, the usual Ulam-Galerkin
approach is density estimation by the histogram method. This perspective allows us other methods. However, this is not the only popular method of density estimation, and we will point out inherent efficiencies available by the popular kernel density estimation method, and this general phrasing of the problem allows for analysis of bias and variance, toward a discussion of the mean square error for example.

### 7.3.3 Estimating Koopman operator with Random Projection

We are investigating how to incorporate the random projection concept in the "Extended DMD" (EDMD) and "kernel DMD" (KDMD) methods to estimate the Koopman spectrum more efficiently. Approximation of an EDMD operator is based on the choice of projected space $\left(\mathcal{F}_{N}\right)$. While the data matrices in EDMD method very large, the matrix computation in a high dimensional even infinite dimensional observable space can be achieved indirectly and much more efficiently using a kernel function. However, dimension of a given data set also depends on the number of sampling points $(M)$. Analytic EDMD can discussed(see [27]) as $M \rightarrow \infty$, but in practical applications $M$ is finite.

Kernel trick can be use to lift up the dimension $N$ (can be infinite ) of observable space which could be grater than $M$, and we can project that data in to $L \ll N$ dimensional ( $L$ only depends on $M$ ) projected space $\mathcal{F}_{L}$ using a dot product preserving(https://doi.org/10.1016/j.patrec.2016.03.031) random projection. Hence we can use two important tools, kernel methods and random projection methods to improve the DMD algorithm with efficient calculations.

We will generalize the kernel method by choosing an empirical kernel and which approximate the Koopman operator as a finite rank operator. Current kernel method uses observables evaluated at a specific set of data points to construct
the Gram matrices. If input space is a reproducing kernel Hilbert space then any $N$ data points can be used to reconstruct a $N$-dimensional Hilbert space by using an experimental kernel, which allows the use of any $N$ datapoint to construct the Gram matrices and approximate the Koopman operator and it generalizes the current kernel DMD algorithm. Further, we will show that any random $r(\ll N)$ vectors can be used to approximate the N -dimensional space and hence the Koopman operator. This will reduce the computational complexity. Therefore, we are planning to provide a method to reduce the dimension of space of observables without losing much information and generalize the exciting extended and kernel DMD methods.

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$.
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309. DOI: $10.1063 / 1.5031109$, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https://doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https: //www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138, URL: https://link.aps . org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL:https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: 10.1103/PhysRevLett. 85.461. URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079, eprint: https://science.sciencemag.org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag. org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi. org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

## Appendices

## Appendix A

## On the Asymmetric Spaces Transfer

## Operators

In this section we prove Theorem 3.2.1 concerning a transfer operator for smooth transformations between sets of perhaps dissimilar dimensionality. In general, the marginal probability density can be found by integrating (or summation in the case of a discrete random variable) to marginalize the joint probability densities. When $x^{\prime}=f(x, y)$, the joint density $\left(x, y, x^{\prime}\right)$ is non-zero only at points on $x^{\prime}=$ $f(x, y)$. Therefore, $\rho\left(x^{\prime}\right)=\sum_{(x, y): x^{\prime}=f(x, y)} \rho\left(x, y, x^{\prime}\right)$ and notice that $\rho\left(x, y, x^{\prime}\right)=$ $\rho\left(x^{\prime} \mid x, y\right) \rho(x, y)$ (By Bayes theorem). Hence, $\rho\left(x^{\prime}\right)=\sum_{(x, y): x^{\prime}=f(x, y)} \rho\left(x^{\prime} \mid x, y\right) \rho(x, y)$ and we only need to show the following claims. We will discuss this by two cases. First, we consider $x^{\prime}=f(x)$ and then we consider more general case $x=f(x, y)$. In higher dimensions we can consider similar scenarios of input and output variables, and correspondingly the trapezoidal bounding regions would need to be specified in which we can analytically control the variables.

Proposition A.0.0. Let $X \in \mathbb{R}$ be a random variable with probability density function $\rho(x)$. Suppose $\rho(x), \rho(. \mid x)$ are Radon-Nikodym derivatives (of induced measure with
respect to some base measure $\mu$ ) which is bounded above and bounded away from zero. In addition, let $x^{\prime}=f(x)$ for some function $f \in C^{1}(\mathbb{R})$. Then,

$$
\rho\left(x^{\prime} \mid X=x_{0}\right)=\lim _{\epsilon \rightarrow 0} d_{\epsilon}\left(x^{\prime}-f\left(x_{0}\right)\right)
$$

where $d_{\epsilon}\left(x^{\prime}-f\left(x_{0}\right)\right)=\left\{\begin{array}{ll}\frac{1}{2 \epsilon\left|f^{\prime}\left(x_{0}\right)\right|} & ;\left|x^{\prime}-f\left(x_{0}\right)\right|<\epsilon\left|f^{\prime}\left(x_{0}\right)\right| \\ 0 & ; \text { otherwise }\end{array}\right.$.
Proof. Let $1 \gg \epsilon>0$ and $x \in I_{\epsilon}=\left(x_{0}-\epsilon, x_{0}+\epsilon\right)$. Since $\rho$ is a Radon-Nikodym derivative with bounded above and bounded away from zero, $\rho\left(I_{\epsilon}\right)=\int_{I_{\epsilon}} \frac{d \rho}{d \mu} d \mu \geq$ $\frac{m}{2 \epsilon}$ where $m$ is the infimum of the Radon-Nikodym derivative. Similarly $\rho\left(I_{\epsilon}\right) \leq \frac{M}{2 \epsilon}$ where $M$ is the supremum of the Radon-Nikodym derivative. In addition, $\mid x^{\prime}-$ $f\left(x_{0}\right)|\approx| f^{\prime}\left(x_{0}\right)| | x-x_{0} \mid$ for $x \in I_{\epsilon}$. Therefore, $x^{\prime} \in\left(f\left(x_{0}\right)-\epsilon\left|f^{\prime}\left(x_{0}\right)\right|, f\left(x_{0}\right)+\right.$ $\left.\epsilon\left|f^{\prime}\left(x_{0}\right)\right|\right)=I_{\epsilon}^{\prime}$ when $x \in I_{\epsilon}$. Hence, $\rho\left(x^{\prime} \mid x \in I_{\epsilon}\right)=\rho\left(x^{\prime} \in I_{\epsilon}^{\prime}\right)$ and $\frac{m}{2 \epsilon\left|f^{\prime}\left(x_{0}\right)\right|} \leq$ $\rho\left(x^{\prime} \mid x \in I_{\epsilon}\right) \leq \frac{M}{2 \epsilon\left|f^{\prime}\left(x_{0}\right)\right|}$. Therefore, $\rho\left(x^{\prime} \mid X=x_{0}\right)=\lim _{\epsilon \rightarrow 0} d_{\epsilon}\left(x^{\prime}-f\left(x_{0}\right)\right)$

Proposition A.0.0. Let $X, Y \in \mathbb{R}$ be random variables with joint probability density function $\rho(x, y)$. Suppose $\rho(x, y)$ and $\rho(. \mid x, y)$ are Radon-Nikodym derivatives (of induced measure with respect to some base measure $\mu$ ) which is bounded above and bounded away from zero. In addition, let $x^{\prime}=f(x, y) \in \mathbb{R}$ for some function $f \in C^{1}(\mathbb{R})$. Then,

$$
\rho\left(x^{\prime} \mid X=x_{0}, Y=y_{0}\right)=\lim _{\epsilon \rightarrow 0} d_{\epsilon}\left(x^{\prime}-f\left(x_{0}, y_{0}\right)\right)
$$

where $d_{\epsilon}\left(x^{\prime}-f\left(x_{0}, y_{0}\right)\right)= \begin{cases}\frac{1}{2 \epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right)} & ;\left|x^{\prime}-f\left(x_{0}, y_{0}\right)\right|<\epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right. \\ 0 & ; \text { otherwise }\end{cases}$
Proof. Let $1 \gg \epsilon>0$ and $A_{\epsilon}=\left\{(x, y) \mid x \in\left(x_{0}-\epsilon, x_{0}+\epsilon\right), y \in\left(y_{0}-\epsilon, y_{0}+\epsilon\right)\right.$. Since $\rho$ is a Radon-Nikodym derivative with bounded above and bounded away
from zero, $\rho\left(A_{\epsilon}\right)=\int_{A_{\epsilon}} \frac{d \rho}{d \mu} d \mu \geq \frac{m}{4 \epsilon^{2}}$ where $m$ is the infimum of the Radon-Nikodym derivative. Similarly, $\rho\left(A_{\epsilon}\right) \leq \frac{M}{4 \epsilon^{2}}$ where $M$ is the supremum of the Radon-Nikodym derivative. In addition, $\left|x^{\prime}-f\left(x_{0}, y_{0}\right)\right| \approx\left|f_{x}\left(x_{0}, y_{0}\right)\right|\left|x-x_{0}\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\left|y-y_{0}\right|$ for $(x, y) \in A_{\epsilon}$. Therefore, $x^{\prime} \in\left(f\left(x_{0}, y_{0}\right)-\epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right), f\left(x_{0}, y_{0}\right)+\right.$ $\left.\epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right)\right)=I_{\epsilon}^{\prime}$ when $(x, y) \in A_{\epsilon}$. Hence, $\rho\left(x^{\prime} \mid(x, y) \in A_{\epsilon}\right)=$ $\rho\left(x^{\prime} \in I_{\epsilon}^{\prime}\right)$ and $\frac{m}{2 \epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right)} \leq \rho\left(x^{\prime} \mid x \in I_{\epsilon}\right) \leq \frac{M}{2 \epsilon\left(\left|f_{x}\left(x_{0}, y_{0}\right)\right|+\left|f_{y}\left(x_{0}, y_{0}\right)\right|\right)}$. Therefore, $\rho\left(x^{\prime} \mid X=x_{0}, Y=y_{0}\right)=\lim _{\epsilon \rightarrow 0} d_{\epsilon}\left(x^{\prime}-f\left(x_{0}, y_{0}\right)\right)$

If $f$ only depends on $x$, then the partial derivative of $f$ with respect to $y$ is equal to zero and which leads to the same result as A.0.0

## Bibliography

[1] Dimitris Achlioptas. "Database-friendly random projections: Johnson-Lindenstrauss with binary coins". In: Journal of Computer and System Sciences 66.4 (2003). Special Issue on PODS 2001, pp. 671-687.
[2] Abd AlRahman R AlMomani, Jie Sun, and Erik Bollt. "How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification". In: arXiv preprint arXiv:1905.08061 (2019).
[3] Hassan Arbabi. Introduction to koopman operator theory of dynamical systems. Jan. 2020. URL: https://www.mit.edu/~arbabi/research/KoopmanIntro. pdf.
[4] Lionel Barnett, Adam B. Barrett, and Anil K. Seth. "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: Phys. Rev. Lett. 103 (23 Dec. 2009), p. 238701. DOI: 10.1103/PhysRevLett. 103.238701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.238701.
[5] Ella Bingham and Heikki Mannila. "Random Projection in Dimensionality Reduction: Applications to Image and Text Data". In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. San Francisco, California: Association for Computing Machinery, 2001, pp. 245-250.
[6] E. Bollt and N. Santitissadeekorn. Applied and Computational Measurable Dynamics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2013. DOI: $10.1137 / 1.9781611972641$
[7] E.M. Bollt. "Synchronization as a process of sharing and transferring information". In: International Journal of Bifurcation and Chaos 22 (Dec. 2012). DOI: 10.1142/S0218127412502616
[8] Erik M Bollt, Jie Sun, and Jakob Runge. "Introduction to Focus Issue: Causation inference and information flow in dynamical systems: Theory and applications". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075201.
[9] Erik M. Bollt. "Open or closed? Information flow decided by transfer operators and forecastability quality metric". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 28.7 (2018), p. 075309. DOI: $10.1063 / 1.5031109$, eprint: https://doi.org/10.1063/1.5031109. URL: https://doi.org/10.1063/1. 5031109 ,
[10] Erik M. Bollt. "Geometric considerations of a good dictionary for Koopman analysis of dynamical systems: Cardinality, "primary eigenfunction," and efficient representation". In: Communications in Nonlinear Science and Numerical Simulation 100 (2021), p. 105833.
[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.
[12] Kevin K. Chen, Jonathan H. Tu, and Clarence W. Rowley. Variants of Dynamic Mode Decomposition: Boundary Condition, Koopman, and Fourier Analyses. Apr. 2012.
[13] Sanjoy Dasgupta. "Experiments with Random Projection". In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. UAI '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 143-151.
[14] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of the JohnsonLindenstrauss lemma". In: International Computer Science Institute, Technical Report 22.1 (1999), pp. 1-5.
[15] Andrea Falcon. "Aristotle on Causality". In: The Stanford Encyclopedia of Philosophy. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University, 2019.
[16] Peter Grassberger. "Generalized dimensions of strange attractors". In: Physics Letters A 97.6 (1983), pp. 227-230. ISSN: 0375-9601. DOI: https: //doi. org/ 10.1016/0375-9601(83) 90753-3. URL: http://www.sciencedirect.com/ science/article/pii/0375960183907533.
[17] Peter Grassberger and Itamar Procaccia. "Characterization of Strange Attractors". In: Phys. Rev. Lett. 50 (5 Jan. 1983), pp. 346-349. DOI: $10.1103 /$ PhysRevLett.50.346. URL:https://link.aps.org/doi/10.1103/PhysRevLett. 50.346
[18] Peter Grassberger and Itamar Procaccia. "Measuring the strangeness of strange attractors". In: Physica D: Nonlinear Phenomena 9.1 (1983), pp. 189-208. ISSN: 0167-2789. DOI: https://doi. org/10.1016/0167-2789(83) 90298-1. URL: http://www.sciencedirect.com/science/article/pii/0167278983902981.
[19] Ned Hall. "Two Concepts of Causation". In: Causation and Counterfactuals. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, 2004, pp. 225276.
[20] David F. Hendry. "The Nobel Memorial Prize for Clive W. J. Granger". In: Scand. J. Econ 106.2 (2004), pp. 187-213.
[21] HYCOM + NCODA Gulf of Mexico $1 / 25^{\circ}$ Analysis. 2021. eprint: https ://www. hycom.org/data/gomu0pt04/expt-90pt1m000. URL: https://www.hycom. org/data/gomu0pt04/expt-90pt1m000.
[22] Y. Ichimaru and GB. Moody. "Development of the polysomnographic database on CD-ROM". In: Psychiatry and Clinical Neurosciences 53 (1999), pp. 175-177.
[23] Laurent Jacques. A Quantized Johnson Lindenstrauss Lemma: The Finding of Buffon's Needle. 2015. arXiv: 1309.1507 [cs.IT].
[24] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of lipschitz maps into Banach spaces.
[25] Pileun Kim, Jonathan Rogers, Jie Sun, and Erik Bollt. "Causation entropy identifies sparsity structure for parameter estimation of dynamic systems". In: Journal of Computational and Nonlinear Dynamics 12.1 (2017), p. 011008.
[26] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". In: Proceedings of the National Academy of Sciences 17.5 (1931), pp. 315-318.
[27] Milan Korda and Igor Mezić. "On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator". In: Journal of Nonlinear Science 28.2 (Nov. 2017), pp. 687-710.
[28] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. "Estimating mutual information". In: Phys. Rev. E 69 (6 June 2004), p. 066138. DOI: 10. 1103/PhysRevE.69.066138. URL: https://link.aps.org/doi/10.1103/ PhysRevE.69.066138.
[29] Nathan J. Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. Dynamic mode decomposition: data-driven modeling of complex systems. Society for industrial and applied mathematics., 2016.
[30] A. Lasota and M.C. Mackey. Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences. Springer New York, 2013. ISBN: 9781461242864. URL: https://books.google.com/books?id=5gHaBwAAQBAJ.
[31] Warren M Lord, Jie Sun, Nicholas T Ouellette, and Erik M Bollt. "Inference of causal information flow in collective animal behavior". In: IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016), pp. 107116.
[32] Shane A McQuarrie, Cheng Huang, and Karen E Willcox. "Data-driven reducedorder models via regularised Operator Inference for a single-injector combustion process". In: Journal of the Royal Society of New Zealand 51.2 (2021), pp. 194-211.
[33] Francisco J. Montáns, Francisco Chinesta, Rafael Gómez-Bombarelli, and J. Nathan Kutz. "Data-driven modeling and learning in science and engineering". In: Comptes Rendus Mécanique 347.11 (2019). Data-Based Engineering Science and Technology, pp. 845-855. ISSN: 1631-0721. DOI: https://doi.org/10.1016/
j.crme.2019.11.009. URL: https://www.sciencedirect.com/science/ article/pii/S1631072119301809.
[34] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. "Sparsitypromoting algorithms for the discovery of informative Koopman-invariant subspaces". In: Journal of Fluid Mechanics 917 (2021).
[35] Judea Pearl. "Bayesianism and Causality, or, Why I Am Only a Half-Bayesian". In: Foundations of Bayesianism. Ed. by David Corfield and Jon Williamson. Kluwer Academic Publishers, 2001, pp. 19-36.
[36] Mark S Pinsker. "Information and information stability of random variables and processes". In: Dokl. Akad. Nauk SSSR (1960).
[37] DR. Rigney, AL. Goldberger, WC. Ocasio, Y. Ichimaru, GB. Moody, and RG. Mark. "Multi-channel physiological data: description and analysis". In: Time Series Prediction: Forecasting the Future and Understanding the Past (1993), pp. 105129.
[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. "Spectral analysis of nonlinear flows". In: Journal of Fluid Mechanics 641 (2009), pp. 115-127.
[39] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marı, et al. "Inferring causation from time series in Earth system sciences". In: Nature communications 10.1 (2019), pp. 1-13.
[40] Bertrand Russell. "I.-On the Notion of Cause". In: Proceedings of the Aristotelian Society 13.1 (June 1913), pp. 1-26.
[41] Tim. Sauer and James A. Yorke. "Are the dimensions of a set and its image equal under typical smooth functions?" In: Ergodic Theory and Dynamical Systems 17 (1997), pp. 941-956.
[42] Tim. Sauer, James A. Yorke, and Martin. Casdagli. "Embedology". In: Statistical Physics 65 (1991), pp. 579-616.
[43] Thomas Schreiber. "Measuring Information Transfer". In: Phys. Rev. Lett. 85 (2 July 2000), pp. 461-464. DOI: $10.1103 /$ PhysRevLett. 85.461 . URL: https : //link.aps.org/doi/10.1103/PhysRevLett.85.461.
[44] Amila Sudu Ambegedara, Jie Sun, Kerop Janoyan, and Erik Bollt. "Informationtheoretical noninvasive damage detection in bridge structures". In: Chaos: An Interdisciplinary Journal of Nonlinear Science 26.11 (2016), p. 116312.
[45] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. "Detecting Causality in Complex Ecosystems". In: Science 338.6106 (2012), pp. 496-500. ISSN: 0036-8075. DOI: $10.1126 /$ science. 1227079. eprint: https://science.sciencemag. org/content/338/6106/ 496.full.pdf. URL: https://science.sciencemag. org/content/338/ 6106/496
[46] J. Sun, D. Taylor, and E. Bollt. "Causal Network Inference by Optimal Causation Entropy". In: SIAM Journal on Applied Dynamical Systems 14.1 (2015), pp. 73-106. DOI: $10.1137 / 140956166$. eprint: https://doi.org/10.1137/ 140956166. URL: https://doi.org/10.1137/140956166.
[47] Jie Sun and Erik M. Bollt. "Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings". In: Physica D: Nonlinear Phenomena 267 (2014). Evolving Dynamical Networks, pp. 49-57. ISSN: 0167-2789. DOI: https://doi. org/10.1016/j.physd.2013.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0167278913001991.
[48] Sudam Surasinghe and Erik M Bollt. "On Geometry of Information Flow for Causal Inference". In: Entropy 22.4 (2020), p. 396.
[49] Sudam Surasinghe and Erik M Bollt. "Randomized Projection Learning Method for Dynamic Mode Decomposition". In: Mathematics 9.21 (2021), p. 2803.
[50] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. "On dynamic mode decomposition: Theory and applications". In: Journal of Computational Dynamics 1.2 (2014), pp. 391-421.
[51] H. White, K. Chalak, and X. Lu. "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: JMRL: Workshop and Conference Proceedings 12 (2011), pp. 1-29.
[52] Halbert White and Karim Chalak. "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning". In: J. Mach. Learn. Res. 10 (Dec. 2009), pp. 1759-1799. ISSN: 1532-4435. URL: http://dl. acm.org/citation.cfm?id=1577069.1755844
[53] Norbert Wiener. "The theory of prediction". In: Mathematics for the Engineer (1956).
[54] C. J. F. Williams. ""Aristotle's Physics, Books I and II", Translated with Introduction and Notes by W. Charlton". In: Mind 82.n/a (1973), p. 617.

