Information bottleneck learns dominant transfer operator eigenfunctions in dynamical systems

Matthew S. Schmitt[1*] **Maciej Koch-Janusz**[1,2,3*] **Michel Fruchart**[1,4]
Daniel S. Seara[1] **Vincenzo Vitelli**[1]
[1]University of Chicago [2]University of Zurich
[3]Haiqu, Inc. [4]ESPCI, Laboratoire Gulliver
*Contributed equally
{schmittms, dsseara, vitelli}@uchicago.edu
maciej@haiqu.ai
michel.fruchart@espci.fr

Abstract

A common task across the physical sciences is that of model reduction: given a high-dimensional and complex description of a full system, how does one reduce it to a small number of important collective variables? Here we investigate model reduction for dynamical systems using the information bottleneck framework. We show that the optimal compression of a system's state is achieved by encoding spectral properties of its transfer operator. After demonstrating this in analytically-tractable examples, we show our findings hold also in variational compression schemes using experimental fluids data. These results shed light into the latent variables in certain neural network architectures, and show the practical utility of information-based loss functions.

1 Introduction

An exhaustive description of a physical system is usually impractical due to the sheer volume of information involved. One often seeks to simplify it, so that it becomes tractable in practice. This procedure, known as model reduction, appears in many guises depending on the system studied. As an example, a chemical reaction may be described by tracking all the positions and velocities of the participating atoms. A much simpler description can be obtained by focusing on discrete chemical species: the reaction is then modelled by a single continuous variable parameterizing the path between these states. This description can be further simplified by considering only discrete transitions between the chemical species [1, 2, 3].

It is of great interest to what extent the model reduction process can be automatized by computers, where reduced models may be used to speed up expensive simulations or extract collective variables in complex systems. Recent developments include the use of so-called Koopman operators that effectively linearize the dynamics, combined with deep learning techniques such as autoencoders to reconstruct and compress the dynamics from observations [4, 5, 6, 7, 8].

Here, we cast model reduction as an information-theoretic problem of finding a lossy compression scheme whose variables maximize the mutual information with the future state – an approach known as the Information Bottleneck (IB) [9, 10]. Using explicit analytical calculations, we connect this approach to the operator-theoretic formalism of dynamical systems and show that the optimal compression reflects spectral properties of the transfer operator. We then show that this connection still holds for so-called variational IB, where the compression is achieved by a neural network. These findings reveal that even when trained on noisy and high-dimensional data, latent variables in these neural networks can be surprisingly interpretable.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

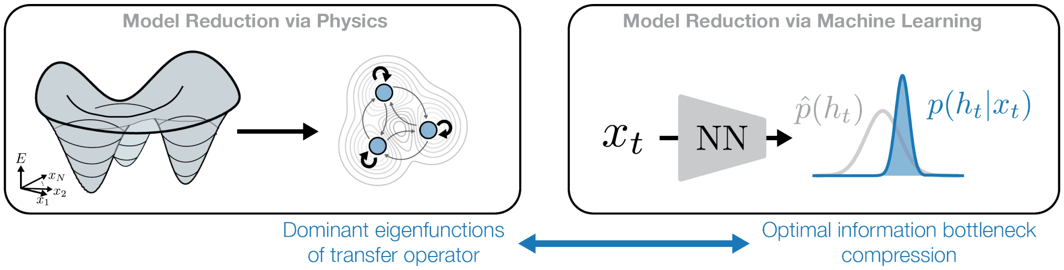

Figure 1: Model reduction in physical systems can be performed by selecting the slowest-decaying eigenfunctions of the transfer operator. One can alternatively frame the problem as finding a compression of the state x which is maximally informative of the future. We show that these two approaches are related.

Related Work Several works have used the information bottleneck framework to understand physical systems, for example in the context of the real-space renormalization group and the identification of order parameters [11, 12, 13, 14], for quantum systems [15], or in the study of time series compression [16, 17, 18]. Conversely, physics-inspired approaches have been used to study the bifurcation structure of information bottleneck solutions [19, 20, 21]. We apply a similar machinery to understand IB restricted to dynamical systems, allowing us to directly interpret the encoder's dependence on physical quantities.

Independent of information theory, attempts have been made to use neural networks to directly learn approximate linearizations of the transfer operator [5] and compress dynamical systems with (variational) autoencoder architectures [8]. Our approach shows that the nice properties of such compressions arise organically when using information-theoretic loss functions.

2 Optimal encoders reflect spectral content of the transfer operator

To connect the familiar quantities of dynamical systems with those of information theory, we start by formulating model reduction as a compression problem. We describe the state of a dynamical system by a random variable: let $p_{X_t}(x)$ be the probability of finding a system in a state x at time t (X_t denotes the corresponding random variable). We wish to identify a reduced description of the system in terms of so-called relevant variables h (H_t is the corresponding random variable). A reduction of x into h can be understood as a probabilistic encoding $p_{H_t|X_t}(h|x)$ [aka $p(h|x)$] which gives the probability of attributing the value h to the state x. For discrete h, this can be understood as a soft partition of the state space of x.

For $p(h|x)$ to be a *useful* reduction, the resulting variable h should contain just enough information to predict the state of the system in the future. To formalize this prescription, we use the so-called Information Bottleneck framework [9]: the encoding $p_{H|X}$ is chosen to minimize the information bottleneck (IB) Lagrangian

$$\mathcal{L}_{\text{IB}}[p_{H_t|X_t}] = I(X_t, H_t) - \beta I(X_{t+\Delta t}, H_t). \tag{1}$$

This objective finds an encoding which discards as much information about the current state x_t as possible while retaining information about the future state $x_{t+\Delta t}$, as proposed in Ref.[10].

We connect the properties of the optimal encoder with those of the dynamics of X_t by considering the operator U governing its evolution, known as the transfer, or Perron-Frobenius, operator:

$$p_{X_{t+\Delta t}}(x) = [U p_{X_t}](x). \tag{2}$$

Purely deterministic systems can be handled in this framework by taking measurement uncertainty into account. The information contained in X_t and $X_{t+\Delta t}$ is, for reversible dynamics, exactly the same, but only in the limit of infinitely fine resolution.

Starting from the form of the optimal encoder computed in [9] and using the formal spectral decomposition of the transfer operator we find that the encoder can be expressed as

$$p^*_\beta(h|x) = \frac{1}{\mathcal{N}(x)}\, p^*_\beta(h)\, \exp\left[\beta \sum_n \mathrm{e}^{\lambda_n \Delta t} \phi_n(x) f_n(h)\right], \tag{3}$$

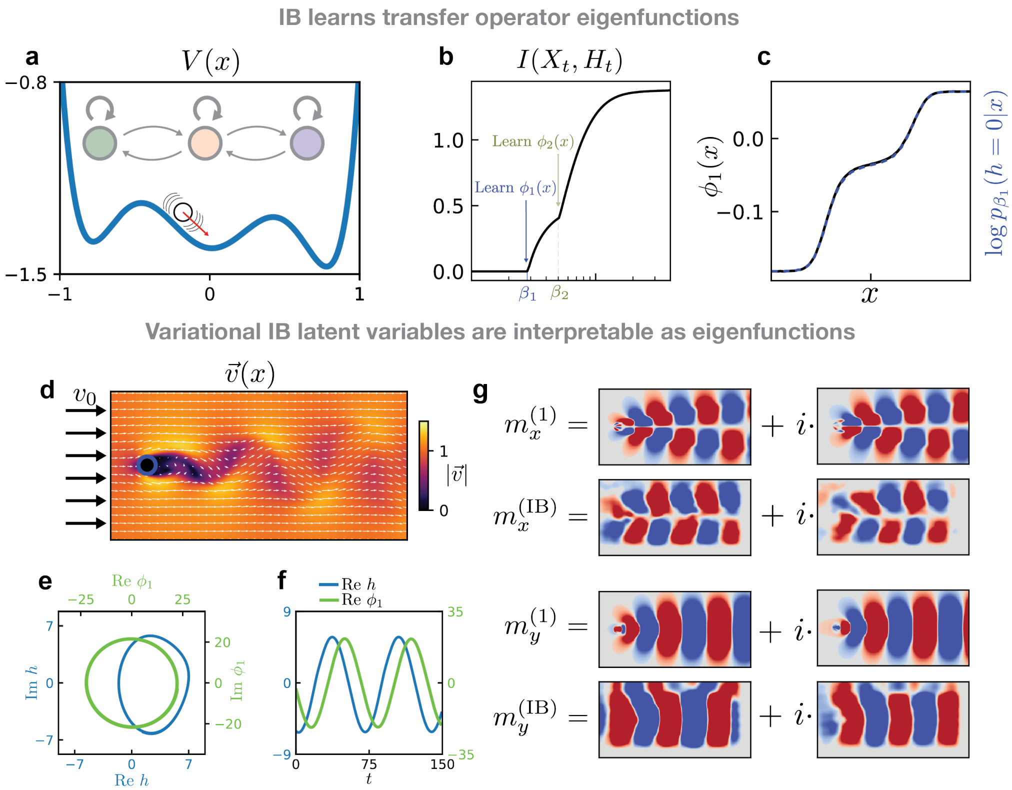

Figure 2: **IB learns eigenfunctions of the adjoint transfer operator** (a) Brownian particle oscillating in a triple well potential. (b) Information contained in an optimal encoding changes in discrete steps as the compression parameter β is varied. (c) At the first transition, the deviations of the encoder from a uniform encoder are approximately given by $p(h|x) \sim \exp(\beta e^{\lambda_1 \Delta t}\phi_1(x) f_1(h))$. (d) Fluid flow past a disk-shaped obstacle exhibits periodic vortex shedding in a so-called von Kármán street. (e) Dynamics in latent space (blue) compared to the evolution of the dominant Koopman mode amplitudes obtained via DMD (green). (f) Time evolution of one component of the latent variable (h_1, blue) and Koopman mode amplitude (green). (g) Comparison between the first Koopman mode ($\vec{m}^{(1)} = \frac{\partial \phi_1}{\partial \vec{v}}$) and those extracted from VIB ($\vec{m}^{(\mathrm{IB})} = \frac{\partial h}{\partial \vec{v}}$).

where λ_n and $\phi_n(x)$ are the eigenvalues and left eigenvectors of U, and $f_n(h)$ are independent of x.

When $\beta < 1$, Eq. (3) is minimized by a trivial encoder $p(h|x) = p(h)$. In this case, $\mathcal{L}_{\mathrm{IB}} = 0$ and no information passes through the bottleneck. As β is increased past a critical value β_{crit}, information is suddenly allowed into the relevant variables. Near this transition, the form of the encoder can be understood by perturbatively expanding in β, similar as in [19, 21]. Immediately after this transition and for long times Δt, the optimal encoding depends on x only through the first eigenfunction $\phi_1(x)$,

$$p^*_\beta(h|x) \sim \exp\left[\beta e^{\lambda_1 \Delta t}\phi_1(x) f_1(h)\right]. \tag{4}$$

Additional terms in the spectral expansion are included at subsequent IB transitions as β is further increased (see Fig.2b). However, their contribution to the relevant information depends on factors such as the time horizon Δt and the spectral gap in the system.

3 Numerical examples

We illustrate the above results by considering a Brownian particle trapped in a confining triple-well potential (Fig. 2a). In the overdamped limit, the state of the particle is completely determined by its position $X_t \in \mathbb{R}$, which we compress into a discrete variable $H_t \in \{0, 1, ..., N_H - 1\}$. To exactly solve the IB optimization problem we numerically approximate the transfer operator using an Ulam approximation [22, 23]. In Fig. 2b we show how the encoded information changes with β, exhibiting discrete jumps at β_1 and β_2. At these jumps, the encoder gains a dependence on x through the first and second eigenfunctions $\phi_1(x)$ and $\phi_2(x)$ of the transfer operator. This can be understood from the stability of the uniform encoder: at β_1, it becomes unstable to perturbations by $\phi_1(x)$, which is reflected in the form of the optimal encoder (Fig. 2c).

Exactly solving the IB objective is difficult as it requires knowledge of the exact conditional distribution $p(x_{t+\Delta t}|x_t)$ which is intractable in practice. The IB optimization problem can, however, be solved approximately by finding an encoder which minimizes a tractable upper bound on the Lagrangian Eq. 1. We achieve this by bounding the terms as

$$I(X_t; H_t) \leq D_{\mathrm{KL}}(p(h|x) \| \hat{p}(h)) \tag{5}$$

$$I(X_{t+\Delta t}; H_t) \geq I_{\mathrm{NCE}}(X_{t+\Delta t}; H_t), \tag{6}$$

where $\hat{p}(h)$ is a variational approximation to the marginal $p(h)$ [24] and I_{NCE} is a so-called noise-contrastive estimate of the mutual information [25]. The Kullback-Leibler divergence D_{KL} is obtained analytically by using a Gaussian ansatz for $p(h|x)$ and letting the marginal $\hat{p}(h)$ be a spherical unit-variance Gaussian [24]. Concretely, encoded variables H_t are sampled from $p(H_t|X_t)$ by computing

$$h_t = f_W(x_t) + \sigma_W(x_t)\eta, \tag{7}$$

where f_W and σ_W deterministic functions modeled by neural networks with parameters (weights) W, and η is a Gaussian random variable with unit variance. This procedure is called variational IB (VIB). The benefit of this variational method is that the loss function can be computed directly from samples, and access to the full distribution $p(x_t, x_{t+\Delta t})$ is not required.

Our above results provide an interpretation of VIB latent variables even for high-dimensional systems, which we illustrate by considering the flow of a fluid past a cylinder (Fig. 2d). The system is characterized by a high-dimensional velocity field $\vec{v}(x) \in \mathbb{R}^{2\times N_{\text{pixels}}}$, where $N_{\text{pixels}} \sim \mathcal{O}(10^5)$. At Reynold's number Re ≈ 150, the fluid undergoes periodic vortex shedding behind the cylinder, forming what is known as a von Kármán street. In this regime, the slowest varying functions of the state variable $\vec{v}(\vec{x})$ are those capturing the oscillatory wake which does not decay in time. These oscillations are captured in the VIB latent variables $[h_0, h_1]$ which are periodic in time (Fig. 2e, f).

In this system, eigenfunctions of the adjoint transfer operator are linear functions of the state variable, $\phi_n[\vec{v}] = \langle \vec{v}(\vec{x}), \vec{m}^{(n)}(\vec{x})\rangle$, where $\vec{m}^{(n)}$ is the n-th "Koopman mode". We compute these modes using dynamic mode decomposition (DMD) [27, 28]. Note that these modes are given by gradients of the eigenfunctions $\partial\phi_n/\partial\vec{v}(\vec{x})$. This suggests that we can study the learned function $h[\vec{v}] = h_0[\vec{v}] + i\, h_1[\vec{v}]$ by examining its gradients with respect to the input, $\partial h/\partial\vec{v}(\vec{x})$.

Based on our analytical results, we expect gradients of the VIB encoder to reflect the dominant characteristics of the subleading Koopman mode $\vec{m}^{(1)}$. This is borne out in Fig. 2g, suggesting that VIB not only recovers the essential oscillatory nature of the dynamics, but does so by learning the correct slowly varying functions of the state variable given by the Koopman eigenfunctions. We emphasize that VIB, which is a deep neural network, learns the true Koopman eigenfunction rather than an arbitrary function with the correct periodicity, as could be obtained from the flow velocity at one well-chosen pixel.

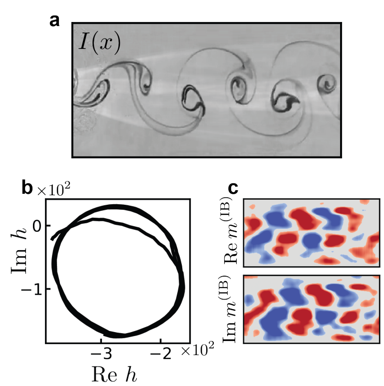

Figure 3: (a) Experimentally-imaged von Kármán street extracted from Ref. [26]. (b) Trajectory of learned encoding variables in the latent space. (c) Koopman mode $m^{(\mathrm{IB})}$, derived as in Fig. 2.

Finally, we apply VIB to a von Kármán street experiment found in the Youtube video [26]. The vortex street is visualized by a dye injected at the site of the obstacle, which is immersed in water flowing uniformly to the right (Fig. 3a). Again, we find that VIB learns clear cyclical dynamics of the latent variables (Fig. 3b), and gradients of the latent variables are similar to those in (Fig. 3c).

4 Conclusion

Here we have characterized the connection between optimal model reduction, phrased in terms of information theory, and the spectral content of the transfer operator. We showed that this connection applies even when using approximate variational methods on real data. This suggests that information theoretic objectives provide a natural path towards physical interpretability of latent variables in deep neural networks.

Acknowledgments and Disclosure of Funding

M.S.S. acknowledges support from a MRSEC-funded Graduate Research Fellowship (DMR-2011854). M.K.-J. gratefully acknowledges financial support from the European Union's Horizon 2020 programme under Marie Sklodowska-Curie Grant Agreement No. 896004 (COMPLEX ML). D.S.S. acknowledges support from a MRSEC-funded Kadanoff–Rice fellowship and the University of Chicago Materials Research Science and Engineering Center (DMR-2011854). V.V. acknowledges support from the Simons Foundation, the Complex Dynamics and Systems Program of the Army Research Office under grant W911NF19-1-0268, the National Science Foundation under grant DMR-2118415 and the University of Chicago Materials Research Science and Engineering Center, which is funded by the National Science Foundation under award no. DMR-2011854.

References

[1] Daniel T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992. ISSN 0378-4371. doi: https://doi.org/10.1016/0378-4371(92)90283-V. URL https://www.sciencedirect.com/science/article/pii/037843719290283V.

[2] Bree B Aldridge, John M Burke, Douglas A Lauffenburger, and Peter K Sorger. Physicochemical modelling of cell signalling pathways. *Nature Cell Biology*, 8(11):1195–1203, November 2006.

[3] Tomasz Lipniacki, Beata Hat, James R. Faeder, and William S. Hlavacek. Stochastic effects and bistability in t cell receptor signaling. *Journal of Theoretical Biology*, 254(1):110–122, 2008. ISSN 0022-5193. doi: https://doi.org/10.1016/j.jtbi.2008.05.001. URL https://www.sciencedirect.com/science/article/pii/S0022519308002208.

[4] Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, may 2022. doi: 10.1137/21M1401243.

[5] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning koopman invariant subspaces for dynamic mode decomposition. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3a835d3215755c435ef4fe9965a3f2a0-Paper.pdf.

[6] Pantelis R Vlachas, Georgios Arampatzis, Caroline Uhler, and Petros Koumoutsakos. Multiscale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence*, 4(4):359–366, April 2022.

[7] Tailin Wu, Takashi Maruyama, and Jure Leskovec. Learning to accelerate partial differential equations via latent global evolution, 2022. arXiv:2206.07681.

[8] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52(1):477–508, 2020. doi: 10.1146/annurev-fluid-010719-060214. URL https://doi.org/10.1146/annurev-fluid-010719-060214.

[9] Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method, 2000. arXiv:0004057.

[10] Felix Creutzig, Amir Globerson, and Naftali Tishby. Past-future information bottleneck in dynamical systems. *Phys. Rev. E*, 79:041925, Apr 2009. doi: 10.1103/PhysRevE.79.041925. URL https://link.aps.org/doi/10.1103/PhysRevE.79.041925.

[11] Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14(6):578–582, Jun 2018. ISSN 1745-2481. doi: 10.1038/s41567-018-0081-4. URL https://doi.org/10.1038/s41567-018-0081-4.

[12] Patrick M. Lenggenhager, Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Optimal renormalization group transformation from information theory. *Phys. Rev. X*, 10:011037, Feb 2020. doi: 10.1103/PhysRevX.10.011037. URL https://link.aps.org/doi/10.1103/PhysRevX.10.011037.

[13] Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Statistical physics through the lens of real-space mutual information. *Phys. Rev. Lett.*, 127:240603, Dec 2021. doi: 10.1103/PhysRevLett.127.240603. URL https://link.aps.org/doi/10.1103/PhysRevLett.127.240603.

[14] Adam G Kline and Stephanie E Palmer. Gaussian information bottleneck and the non-perturbative renormalization group. *New Journal of Physics*, 24(3):033007, mar 2022. doi: 10.1088/1367-2630/ac395d. URL https://dx.doi.org/10.1088/1367-2630/ac395d.

[15] Zhelun Zhang and Yi-Zhuang You. Observing schrödinger's cat with artificial intelligence: Emergent classicality from information bottleneck, 2023. arXiv:2306.14838.

[16] Felix Creutzig and Henning Sprekeler. Predictive coding and the slowness principle: An information-theoretic approach. *Neural Computation*, 20(4):1026–1041, 2008. doi: 10.1162/neco.2008.01-07-455.

[17] Dedi Wang and Pratyush Tiwary. State predictive information bottleneck. *The Journal of Chemical Physics*, 154(13):134111, 04 2021. ISSN 0021-9606. doi: 10.1063/5.0038198. URL https://doi.org/10.1063/5.0038198.

[18] Vedant Sachdeva, Thierry Mora, Aleksandra M. Walczak, and Stephanie E. Palmer. Optimal prediction with resource constraints using the information bottleneck. *PLOS Computational Biology*, 17(3):1–27, 03 2021. doi: 10.1371/journal.pcbi.1008743. URL https://doi.org/10.1371/journal.pcbi.1008743.

[19] Albert Parker, Tomáš Gedeon, and Alexander Dimitrov. Annealing and the rate distortion problem. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002. URL https://proceedings.neurips.cc/paper_files/paper/2002/file/ccbd8ca962b80445df1f7f38c57759f0-Paper.pdf.

[20] Tailin Wu, Ian Fischer, Isaac L. Chuang, and Max Tegmark. Learnability for the information bottleneck. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 1050–1060. PMLR, 22–25 Jul 2020. URL https://proceedings.mlr.press/v115/wu20b.html.

[21] Vudtiwat Ngampruetikorn and David J. Schwab. Perturbation theory for the information bottleneck. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=A2HvBPoSBMs.

[22] S.M. Ulam. *Problems in Modern Mathematics*. Science Editions, 1960. URL https://books.google.com/books?id=J_zQjw593VUC.

[23] Tien-Yien Li. Finite approximation for the frobenius-perron operator. a solution to ulam's conjecture. *Journal of Approximation Theory*, 17(2):177–186, 1976. ISSN 0021-9045. doi: https://doi.org/10.1016/0021-9045(76)90037-X. URL https://www.sciencedirect.com/science/article/pii/002190457690037X.

[24] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. *CoRR*, abs/1612.00410, 2016. URL http://arxiv.org/abs/1612.00410.

[25] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019. arXiv:1807.03748.

[26] Jacob Albright (user jacobalbright3585). Flow Visualization in a Water Channel, 2017. URL https://www.youtube.com/watch?v=30_aADFVL9M. [Accessed 03-Sept-2023].

[27] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009. doi: 10.1017/S0022112009992059.

[28] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. doi: 10.1017/S0022112010001217.