

Nonlinear correlations underlie linear response and causality

Gabriele Di Antonio^{1,2,3,*} and Gianni Valerio Vinci^{1,†}¹National Center for Radiation Protection and Computational Physics, *Istituto Superiore di Sanità*, 00161 Rome, Italy²Research Center “Enrico Fermi”, 00184 Rome, Italy³Engineering Department, “Roma Tre” University of Rome, 00146 Rome, Italy

(Received 11 October 2024; accepted 21 July 2025; published 31 October 2025)

The inference of causal relationships among observed variables is a pivotal, long-standing problem in the scientific community. An intuitive method for quantifying these causal links involves examining the response of one variable to perturbations in another. The fluctuation-dissipation theorem elegantly connects this response to the correlation functions of the unperturbed system, thereby bridging the concepts of causality and correlation. However, this relationship becomes intricate in nonlinear systems, where knowledge of the invariant measure is required but elusive, especially in high-dimensional spaces. In this study, we establish a link between generalized correlation functions with the response function. This connection provides an alternative method for computing the inference of causal interactions, even when the invariant measure is unknown. We validate our theoretical framework by applying it to several nonlinear systems, showing convergence and consistency for increasing number of measured observables and samples. Finally, we discuss a significant interplay between the resulting causal network and the relevant timescales of the system.

DOI: [10.1103/vjvt-gdnl](https://doi.org/10.1103/vjvt-gdnl)

Introduction. Inference of causal relations between measured quantities is a central and old open problem in science. The network of causal links between available degrees of freedom is essential to build effective models of reality. As the old adage goes “correlations do not imply causation,” however, a proper use of statistical information can go a long way in predicting causal relationships. For example, in their seminal work Wiener [1] and Granger [2] proposed that Y is causally related to X if the forecasting of X based on its previous values increases when the information of Y is added. Another promising concept was proposed by Liang and Kleeman [3] and later refined in Refs. [4–7]. The authors defined the information flow between different degrees of freedom in the context of stochastic dynamical systems. An alternative approach, also based on forecasting, is rooted in the embedding theory of dynamical systems pioneered by Takens [8]. Despite the fact that both approaches have intrinsic limitations, they have been massively adopted in virtually all fields of science [9–12]. There is, however, another possible definition of causality commonly accepted in physics. We say that X is causally related to Y if a small perturbation on the latter has a significant effect on the former. The quantity that measures this effect is commonly referred to as the response function $R_{Y \rightarrow X}(t)$. While estimating causation

via intervention is certainly a good idea, practical limitations often hinder our ability to clearly perturb certain degrees of freedom and observe the system’s response. This issue can be bypassed thanks to fluctuation-dissipation theorem [13] that relates the response to generalized correlation functions of the unperturbed system. Building on this concept, Baldovin *et al.* [14] have proposed to measure causality as time integral of $R_{Y \rightarrow X}(t)$ to account for potential time-lagged causal effects. When the system is linear, the response is proportional to the standard cross-correlation functions between all possible pairs of degrees of freedom (see also Refs. [15–17]). For general nonlinear systems, the fluctuation-dissipation theorem can still be applied but requires the knowledge of the functional form of the invariant probability density functions $p_s(\mathbf{x})$ [13], which is nontrivial to estimate for high-dimensional problems. In this Letter, we apply the Koopman formalism [18,19] and show that by lifting the system to higher dimensions, introducing extra degrees of freedom, a simple and general relationship between response and correlations among the extended set of variables can be recovered, analogous to what occurs in linear systems, providing a handy framework for inferring causal interactions among the system’s variables. The connections between Koopman theory and response functions have recently been studied in several ways, including the following works [20–23].

We begin by establishing a relation between the correlations computed in the Koopman operator basis and the response function. We first apply our findings to an instructive nonlinear example where the high-dimensional lifting can be carried out analytically. We test our approach on various dynamical systems and we explore the convergence and constraints for increasing dimensionality. Finally, utilizing the computed $R(t)$ we use our results to infer causality as proposed in Ref. [14].

*Contact author: gabriele.diantonio@cref.it†Contact author: gianni.vinci@iss.it

Koopman operator and generalized linear response. In the following, we will consider the generic stochastic Markovian dynamical systems of the form

$$d\mathbf{X} = a(\mathbf{X})dt + b(\mathbf{X})d\mathbf{W}, \quad (1)$$

where $\mathbf{X} \in R^N$ is an N -dimensional stochastic process, $a(\mathbf{x}) \in R^N$ is the drift term, and $b(\mathbf{X}) \in R^{N \times S}$ is the diffusion term, while $d\mathbf{W} \in R^S$ are independent Wiener noise sources [24]. We also assume natural boundary conditions on a domain $\mathcal{D}$. For these systems, the Koopman operator $\mathcal{K}$ governs the time evolution of observables f :

$$\mathcal{K}^t f(\mathbf{X}_0) = E[f(\mathbf{X}_t)|\mathbf{X}_0], \quad (2)$$

where E indicates the expectation value, conditioned to $\mathbf{X}(t=0) = \mathbf{X}_0$. Note that for systems like Eq. (1), the associated infinitesimal generator $\mathcal{L}f = \lim_{t \rightarrow 0} (\mathcal{K}^t f - f)/t$ is the backward Kolmogorov operator [19,25] whose eigenvalues λ_n dictate the relevant timescales of the dynamics and are also involved in the first passage time problem [26]. Since it is in general difficult to find the eigenfunctions $\mathcal{L}\phi_n(\mathbf{x}) = \lambda_n \phi_n(\mathbf{x})$, we can resort to a pseudospectral approach [27] that in the Koopman formalism is typically referred to as Extended Dynamic Mode Decomposition [28]. Consider a set of d observables, or basis functions, $\boldsymbol{\psi}(\mathbf{x}) = \{\psi_k(\mathbf{x})\}_{k=1}^d$ that ideally span a subspace containing the state of the system at all times. We choose the basis to be orthogonal, and normalized, with respect to the inner product defined by the measure $w(\mathbf{x})$, i.e., $\int_{\mathcal{D}} \psi_i(\mathbf{x})\psi_j(\mathbf{x})w(\mathbf{x})d\mathbf{x} = \delta_{ij}$. Then, the matrix $\mathbf{K}(t) = e^{\mathcal{L}t}$, where $L_{nm} = \int_{\mathcal{D}} \psi_n(\mathbf{x})\mathcal{L}\psi_m(\mathbf{x})d\mathbf{x}$, approximates the action of the Koopman operator, i.e.,

$$E[\boldsymbol{\psi}(\mathbf{X}_t)|\mathbf{X}_0] \approx \mathbf{K}(t)\boldsymbol{\psi}(\mathbf{X}_0). \quad (3)$$

While for $d \rightarrow \infty$ convergence is guaranteed and $\mathbf{L}$ becomes isospectral to $\mathcal{L}$, in practical application controlling spectral pollution due to finite d is still an open problem, which we will not address here [29,30]. Notably, if we chose as basis function the eigenfunctions of $\mathcal{L}$, $K_{nm}(t) = e^{\lambda_n t} \delta_{nm}$. In the following, we will define the matrix $\boldsymbol{\alpha}$ that relates physical degree of freedom with the observables by the best fit of $\mathbf{x}$ by $\boldsymbol{\psi}$, i.e., $\mathbf{x} \approx \boldsymbol{\alpha}\boldsymbol{\psi}(\mathbf{x})$. Then, expected values can be expressed as follows:

$$E[\mathbf{X}_t|\mathbf{X}_0] \approx \boldsymbol{\alpha}\mathbf{K}(t)\boldsymbol{\psi}(\mathbf{X}_0). \quad (4)$$

Here, we will treat $\boldsymbol{\psi}$ as independent degrees of freedom, which we will refer to as virtual to distinguish them from the physical degree of freedom $\mathbf{x}$, similarly to the auxiliary response field introduced in statistical physics [31]. For the response function, we follow Ref. [14] and we will indicate with $\{\cdot\}_p$ the system perturbed in the initial condition. The virtual response seen on the physical variable x_j due to a perturbation in the virtual variable $\psi_i(\mathbf{X}_0) \rightarrow \psi_i(\mathbf{X}_0) + \epsilon$ is defined as follows:

$$\tilde{R}_{\psi_i \rightarrow x_j}(t) = \lim_{\epsilon \rightarrow 0} \frac{\langle X_j(t) \rangle_p - \langle X_j(t) \rangle}{\epsilon}, \quad (5)$$

where the brackets include the average with respect to the distribution of the initial conditions [32]. From Eq. (4) follows the relation of the virtual response to the Koopman matrix:

$$\tilde{R}_{\psi_i \rightarrow x_j}(t) \approx [\boldsymbol{\alpha}\mathbf{K}(t)]_{ji}. \quad (6)$$

Clearly, Eq. (6) cannot be used in practice because there is no way to perturb such virtual degrees of freedom without changing $\mathbf{X}(t)$; however, as we will show, this perturbation is not needed. In the same way, we can look at the effect of perturbing $x_i(0) \rightarrow x_i(0) + \epsilon$, and finally uncover the relation between the standard response function and Eq. (6):

$$R_{x_i \rightarrow x_j}(t) \approx \sum_{m=1}^d \left\langle \frac{\partial \psi_m(\mathbf{x})}{\partial x_i} \right|_{\mathbf{X}_0} \tilde{R}_{\psi_m \rightarrow x_j}(t). \quad (7)$$

The expansion in the dimensionality of the system, by including the virtual degrees of freedom, has a linearizing effect, which is also the major appeal of the Koopman formalism [33]. In turn, this suggests that we can relate the virtual response to correlations of the unperturbed system. In fact, applying $\boldsymbol{\alpha}$ to Eq. (4) and using Eq. (6), we find

$$E[\mathbf{X}_t|\mathbf{X}_0]\boldsymbol{\psi}(\mathbf{X}_0)^T \approx \tilde{\mathbf{R}}_{\boldsymbol{\psi} \rightarrow \mathbf{x}}(t)\boldsymbol{\psi}(\mathbf{X}_0)\boldsymbol{\psi}(\mathbf{X}_0)^T.$$

By averaging over the initial condition, we find a fluctuation-dissipation-type relation:

$$\tilde{\mathbf{R}}_{\boldsymbol{\psi} \rightarrow \mathbf{x}}(t) \approx \mathbf{C}_{\mathbf{x}\boldsymbol{\psi}}(t)\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}(0)^{-1}, \quad (8)$$

where $\mathbf{C}_{fg}(t) = \langle f(\mathbf{X}_t)g(\mathbf{X}_0)^T \rangle$ are covariance matrices in case of null averages. It is worth noting that Eq. (8) does not stand on the semigroup condition, providing an autonomous estimate of $\mathbf{K}(t)$ for every time point, which enhances the robustness of the method by mitigating error propagation. Finally, Eq. (8) can be used to find the relation between the linear response of the system and unperturbed correlations with respect to virtual degrees of freedom, one of the main results of this Letter:

$$\mathbf{R}(t) \approx \mathbf{C}_{\mathbf{x}\boldsymbol{\psi}}(t)\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}(0)^{-1} \langle \nabla \boldsymbol{\psi} \rangle, \quad (9)$$

where $\langle \nabla \boldsymbol{\psi} \rangle_{ij} = \langle \frac{\partial \psi_i(\mathbf{x})}{\partial x_j} |_{\mathbf{X}_0} \rangle$ and $R_{ij} = R_{x_j \rightarrow x_i}$. Notably the proportionality of the linear response of systems like Eq. (1) to nonstandard correlations can be inferred starting from the fact that under rather general assumption [13],

$$R_{x_i \rightarrow x_j}(t) = - \left\langle X_t^{(j)} \frac{\partial \ln[p_s(\mathbf{x})]}{\partial x^{(i)}} \right|_{\mathbf{X}_0} \right\rangle. \quad (10)$$

In the case in which the considered invariant subspace contains the functions $\nabla \ln[p_s(\mathbf{x})]$ (i.e., $\exists \boldsymbol{\beta}; \nabla \ln[p_s(\mathbf{x})] \approx \boldsymbol{\psi}(\mathbf{x})\boldsymbol{\beta}^T$), we can rewrite Eq. (10) as

$$\mathbf{R}(t) \approx \mathbf{C}_{\mathbf{x}\boldsymbol{\psi}}(t)\boldsymbol{\beta}^T. \quad (11)$$

Contrary to Eq. (9), where we need the stationary expectation values of the virtual degree of freedom that can be sampled via Monte Carlo methods, to apply Eq. (10) the knowledge of functional form of $p_s(\mathbf{x})$ is required, which is not always easy to get especially for high-dimensional systems. The relation between Eqs. (9) and (11) is more clear if we chose as weight of the inner product the stationary distribution, i.e., $w(\mathbf{x}) = p_s(\mathbf{x})$. Then, in this case, the normalization condition implies that $\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}(0) = \mathcal{I}$ is the identity matrix, while from the

definition of β and integration by parts it follows that

$$\begin{aligned}\beta_{ij} &= - \int_{\mathcal{D}} \psi_i(\mathbf{x}) \frac{\partial_j p_s(\mathbf{x})}{p_s(\mathbf{x})} w(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{D}} \frac{\partial \psi_i(\mathbf{x})}{\partial x_j} p_s(\mathbf{x}) d\mathbf{x} = \langle \nabla \psi \rangle_{ij}.\end{aligned}$$

Hence, Eqs. (11) and (9) are, in this case, equivalent.

Linear response in nonlinear systems. For a system with linear drift and constant diffusion, i.e., a multidimensional Ornstein-Uhlenbeck process [24], the choice $\psi_k(\mathbf{x}) = x_k$ is sufficient to fully capture the response. In fact, in this case, the stationary distribution $p_s(\mathbf{x})$ is Gaussian, and from Eq. (10), it follows that $\mathbf{R}(t) = \mathbf{C}_{\mathbf{xx}}(t)\mathbf{C}_{\mathbf{xx}}(0)^{-1}$. Recently, in Ref. [14], it has been shown that the fluctuation-dissipation formula for linear systems can sometimes yield satisfactory results even in the context of nonlinear dynamics. This is possible only when the stationary distribution is well approximated by a Gaussian for the reasons mentioned above. On the other hand, our result (9) is the natural generalization that applies for more broad cases paying the cost of extending “virtually” the dimensionality of the systems. In the Koopman literature, the proposed generalization parallels the extension of the Dynamic Mode Decomposition in Ref. [28]. We note, however, that analogous conclusions can be achieved without invoking the Koopman formalism at all. To do so, we can use a generalization of the regression theorem, an important result of stochastic calculus [24] that has found several applications in quantum mechanics [34]. The theorem is generally presented for stochastic systems with linear drift and constant diffusion, but it can be generalized also to multiplicative noise as long as the drift component remains linear. For a d -dimensional process $\mathbf{X}$ such that

$$d\mathbf{X} = \mathbf{A}\mathbf{X}dt + b(\mathbf{X})d\mathbf{W},$$

one can show that independently of the specific choice of $b(\mathbf{X})$, the time-dependent evolution of the vector of observables $\mathbf{m}(t) = \{E[X_{i,t}|\mathbf{X}_0]\}_{i=1}^d$ is simply given by

$$\mathbf{m}(t) = e^{\mathbf{A}t}\mathbf{m}(0). \quad (12)$$

Now we go back to the general system (1) and suppose we have found a collection of virtual degrees of freedom, i.e., functions $Z_n = f(\mathbf{X})$, such that the extended system with $\tilde{\mathbf{X}} = (\mathbf{X} \cdots \mathbf{Z})$ has an approximately linear drift:

$$d\tilde{\mathbf{X}} \approx \mathbf{A}\tilde{\mathbf{X}}dt + b(\tilde{\mathbf{X}})d\mathbf{W}.$$

By applying the regression theorem (12), the time-dependent expectation values follow from which relation (8) can be obtained and used to finally derive the response functions of the true degrees of freedom.

We begin with an instructive example in which incorporating a broader set of observables is crucial for accurately estimating the correct linear response. This example is particularly valuable as it allows for analytical calculations. Specifically, we consider the following nonlinear stochastic differential equation:

$$\begin{aligned}d\mathbf{X} &= \mathbf{G}_x\mathbf{X}dt + \sigma_x d\mathbf{W}_x, \\ d\mathbf{Y} &= \mathbf{G}_y\mathbf{Y}dt + \mathbf{G}_{yx}\mathbf{X} \odot \mathbf{X} + \sigma_y d\mathbf{W}_y\end{aligned} \quad (13)$$

where $\odot$ indicates the element-wise product. We have two coupled systems, where the term $\mathbf{G}_{yx}\mathbf{X} \odot \mathbf{X}$ introduces nonlinear interactions between the variables. When considering the case in which $\mathbf{G}_x$ is a diagonal matrix, this system can be transformed into a linear one by introducing a new set of observables defined as $\mathbf{Z} = \mathbf{X} \odot \mathbf{X}$, leading to the following evolution:

$$\begin{aligned}d\mathbf{Z} &= 2\mathbf{G}_x\mathbf{Z}dt + \sigma_x^2 dt + 2\sigma_x\mathbf{X} \odot d\mathbf{W}_x, \\ d\mathbf{X} &= \mathbf{G}_x\mathbf{X}dt + \sigma_x d\mathbf{W}_x, \\ d\mathbf{Y} &= \mathbf{G}_y\mathbf{Y}dt + \mathbf{G}_{yx}\mathbf{Z} + \sigma_y d\mathbf{W}_y.\end{aligned} \quad (14)$$

Thanks to the regression theorem (12), these systems admit an exact finite-size Koopman matrix, i.e., $\mathbf{K}(t) = e^{\mathbf{A}t}$, where $\mathbf{A}$ is the matrix of the linear drift in Eq. (14). Then using Eq. (7), we can derive analytically the exact response function of Eq. (13). We analyze the agreement of Eq. (9) with the exact linear response curves for two sets of observables: (1) Hermite functions up to order 1, referred to as H_1 ; and (2) Hermite functions up to order 2, referred to as H_2 , and compare both also to the response function sampled by perturbing the initial conditions, i.e., using definition (5) and referred to as sampled response. In Fig. 1(a), an example comparison between these three cases and the exact response function is shown. The average error [see Fig. 1(b)] associated with H_2 scales with the number of realizations in the same way as the sampled response, while H_1 is order of magnitude worse. This happens because the stationary distribution $p_s(\mathbf{x})$ is asymmetric; thus, a linear approximation, implicit in H_1 , will not work. However, we note that the increase of the standard deviation σ_y in Eq. (1) mitigates the influence of X on Y , thereby reducing the asymmetry of its distribution and leading the system to behave more linearly. This effect is clearly shown plotting the average error versus σ_y [Fig. 1(c)], with the accuracy of H_1 approaching that of the sampled response as σ_y increases. From a computational standpoint, Eq. (9) has been evaluated using a pseudoinverse operation for enhanced efficiency. Specifically, let $\mathbf{V}(t)$ represent the matrix whose columns correspond to the system variables at time t , with rows reflecting their values across different realizations of noise, and similarly, let $\mathbf{P}(t)$ be the corresponding matrix for the set of functions ψ , then the virtual response can be estimated as follows:

$$\tilde{\mathbf{R}}_{\psi \rightarrow x}(t) = [\mathbf{V}(t)\mathbf{P}^T(0)][\mathbf{P}(0)\mathbf{P}^T(0)]^{-1} = \mathbf{V}(t)\mathbf{P}^\dagger(0),$$

where $\{\cdot\}^\dagger$ denotes the pseudoinverse operation. For the example we have used so far, a basis of functions containing polynomials up to the second order is sufficient to achieve perfect agreement between Eq. (9) and the exact response function. Clearly, this is not true in general. We selected three prototypical stochastic systems in two dimensions (x, y) and computed the response of variable x to a perturbation of y , i.e., $R_{y \rightarrow x}(t)$. We considered a gradient system with four stable states [35]:

$$\begin{aligned}dX &= -\partial_X V(X, Y)dt + \sigma dW_x, \\ dY &= -\partial_Y V(X, Y)dt + \sigma dW_y,\end{aligned} \quad (15)$$

with $V(x, y) = (x - 1)^2 + (y - 0.5)^2 + 0.1xy$. For a case with nonpolynomial drift, a network of two interacting neural

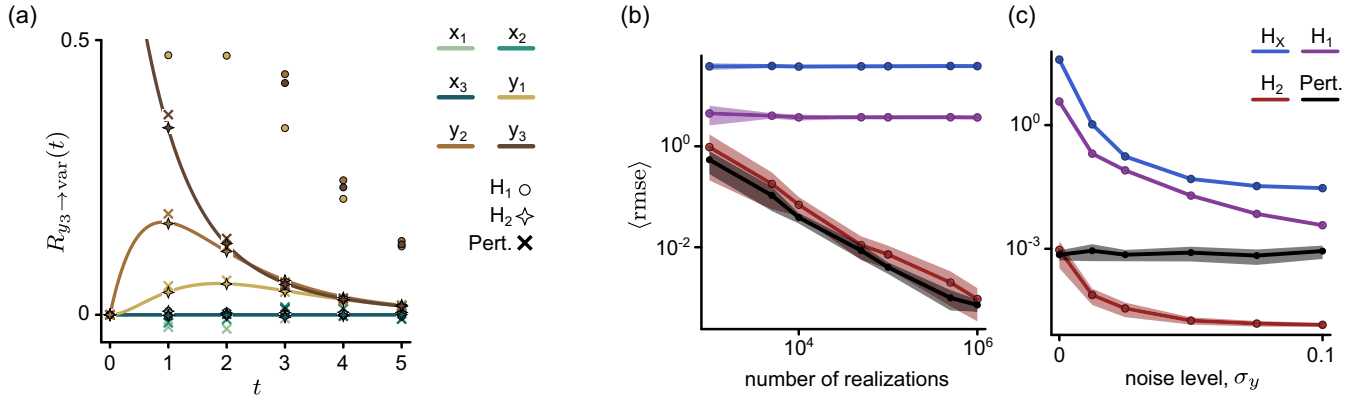


FIG. 1. (a) Various estimations of the response function in relation to the perturbing variable y_3 for a six-dimensional example system as described in Eq. (13). The solid line represents the ground truth, while the individual points denote the definitions of perturbation and the fluctuation correlation formula presented in Eq. (9), specifically for the first-order H_1 and second-order H_2 Hermite polynomial observables. Averages over 10^6 experiment realizations for system parameters $\sigma_x = 0.1$ and $\sigma_y = 0.025$. (b), (c) The average reconstruction error of response curves across all dimensions and time steps, estimated using root-mean-square error, for the same system depicted in panel (a). Curves are estimated by perturbation limit and from Eq. (9) using observables $\psi_k(\mathbf{x}) = x_k$, named H_x , and Hermite basis functions H_1 and H_2 . This analysis considers, respectively, (b) varying numbers of process realizations ($\sigma_y = 0.0$) and (c) different levels of noise σ_y (10^6 experiments). The shaded ribbon area illustrates the standard deviation resulting from ten complete analysis repetitions. The time resolution is set at $dt = 0.01$ and perturbation amplitude $\epsilon = 0.01$, with the stochastic Heun method for integration.

populations is given by

$$\begin{aligned} dX &= [\tanh(J_{11}X + J_{12}Y) - X]dt + \sigma dW_x, \\ dY &= [\tanh(J_{21}X + J_{22}Y) - Y]dt + \sigma dW_y, \end{aligned} \quad (16)$$

and finally, a Stuart-Landau stochastic oscillator is given as follows:

$$\begin{aligned} dX &= [-a(X^2 + Y^2)(bY + X) + aX - Y]dt + \sigma dW_x, \\ dY &= [-a(X^2 + Y^2)(bX + Y) + aY + X]dt + \sigma dW_y. \end{aligned} \quad (17)$$

Despite the qualitative differences in the stationary distributions of the three systems (see Fig. 2), augmenting the

function library with second-order polynomials (H_2) yielded no improvement over the linear case (H_1) for any of them. At the same time, however, including polynomials of order 3 (H_3) allows us to fully capture the response function for all cases. Notably, this approach yields significantly fewer terms than methods based on spectral estimation [22]. The reason behind this behavior can be promptly understood for systems like Eq. (15). In fact, for gradient systems, the generalized Fluctuation-Dissipation (F.D.) theorem (10) gives us

$$R_{y \rightarrow x}(t) = 2\langle X(t)\partial_y V(X_0, Y_0) \rangle / \sigma^2.$$

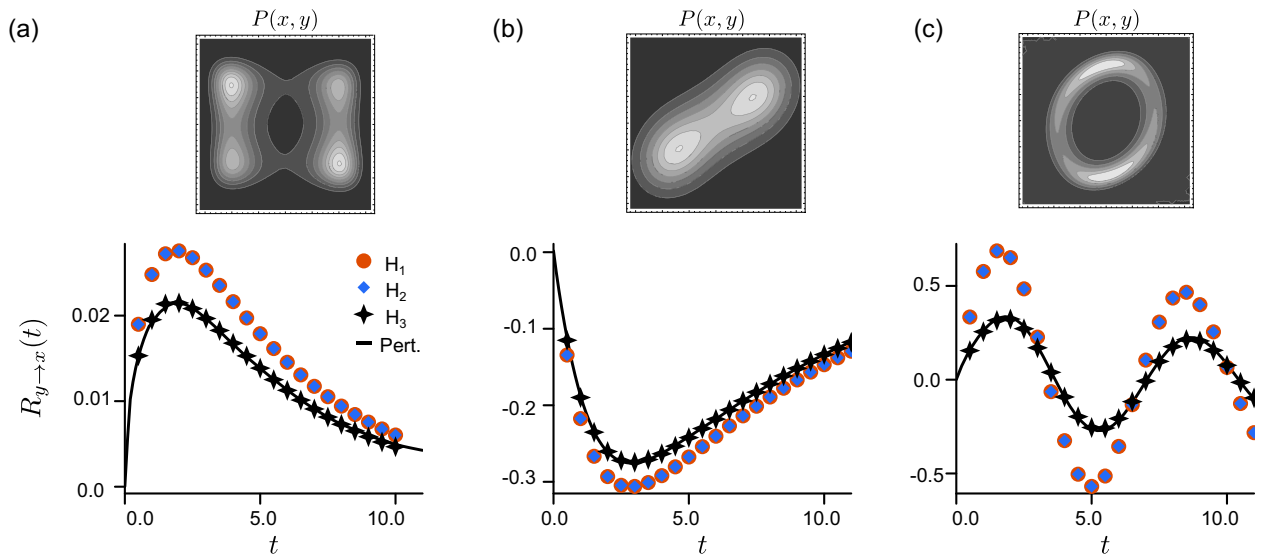


FIG. 2. In the top row, we show the stationary distribution for system (15) in panel (a), system (16) in panel (b), and system (17) in panel (c). In the bottom row, the response functions (in black) are computed from the time-dependent solution of the Fokker-Planck equation, while the scatter points represent the result of Eq. (9) for different basis functions H_1 , H_2 , and H_3 , as shown in Fig. 1. The parameters are $\alpha = 0.1$, $a = 1$, $b = -0.3$, $J_{11} = 0.9$, $J_{12} = 0.5$, $J_{21} = 0.3$, $J_{22} = 0.6$, and $\sigma = 0.3$.

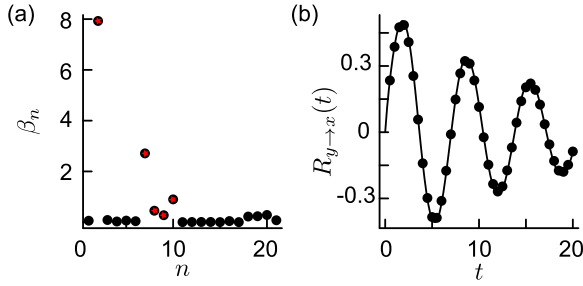


FIG. 3. (a) Weight coefficients $\beta_n = \beta_{n,2}$ (where 1 and 2 are the indexes for variables x and y , respectively) for the Stuart-Landau system (17). Dominant terms (highlighted in red) were selected as basis functions for the response function estimation. (b) Comparison between the estimated response function (derived from the reduced basis) and the exact analytical result.

In our case, $\partial_y V(x, y) = \alpha x + (y^2 - 1/2)y$, which shows that the response is fully predicted once we consider the simple basis $\{x, y, y^3\}$. However, finding a good basis is generally a nontrivial problem. In the following section, we will address the problem of selecting the dictionary $\psi_n(\mathbf{x})$ in a principled way for general stochastic systems.

Principled selection of the dictionary of basis functions. We will occasionally use Einstein notation for the sums involved. Equation (9) suggests us a way to estimate the optimal base from empirical data. For all the two-dimensional systems considered above, the response of a variable x due to a perturbation of y is approximated by

$$R_{y \rightarrow x}(t) \approx \langle X \psi_n(\vec{X}_0) \rangle \beta_{n,2}.$$

We defined

$$\beta_{n,j} = [\mathbf{C}_{\psi\psi}(0)^{-1}]_{nk} \langle \partial_j \psi_k(\vec{X}_0) \rangle,$$

where $(\partial_x, \partial_y) = (\partial_1, \partial_2)$. All the integrals involved in the weights β are with respect to the stationary distribution, thus can be generally computed empirically. As in all spectral approaches β tells us the most relevant “modes” in the expansion, it allows us to optimally tune the dictionary of basis function $\psi(\vec{x})$. A concrete example is shown in Fig. 3, where we computed $\beta_{n,2}$ with the usual basis of Hermite polynomials for the stationary state of Stuart-Landau system (17). As shown in Fig. 3(a), the only functions with a nonnegligible β are

$$\psi(x, y) = \{H_1(x), H_1(y), H_1(x)H_2(y), H_2(x)H_1(y), H_3(x), H_3(y)\}.$$

As expected, using this reduced base, we get an excellent agreement in the response function [see Fig. 3(b)]. The optimal choice of functions in general is the one that diagonalizes the matrix $\mathbf{C}_{\psi\psi}(0)$, or in other terms we should find a set of functions that is orthogonal with respect to the stationary measure $d\mu = P(\mathbf{x})d\mathbf{x}$. This is an old idea that dates back to the theory of polynomial chaos first introduced by Wiener and subsequent formal results such as the Kosambi-Karhunen-Loève theorem [36]. Even though the core concepts remained the same, these spectral expansion approaches, under the large umbrella of the Koopman formalism, have found recently renovated interest thanks to the

widespread computational power and the development of a powerful machine-learning algorithm [35]. We note, however, that typically in all these cases the focus is on learning a large set of approximated eigenfunctions and eigenvalues of the Koopman operator from empirical measurements, which is undoubtedly challenging especially for high-dimensional systems. In this work, however, we are interested only in specific response functions. For example, we have seen that for system (17) a basis of just five elements predicts perfectly $R_{y \rightarrow x}(t)$, but clearly, by no means this small dictionary is sufficient to represent a large collection of the true eigenfunctions of the Koopman operator. So far, we have avoided questions related to optimal size of the dictionary of $\psi(\mathbf{x})$. A first reasonable step that allows us to find a sparse dictionary is to compute β from a large set of functions and sort the most relevant elements with respect to the entries of $|\beta_n|$. We need, however, some guiding principle to properly truncate this sorted dictionary. It is possible to derive several analytical constraints that can guide this selection and optimization of the dictionary. In fact, by comparing our expression (9) with the F.D. theorem (10), we find the central consistency relation:

$$\beta \psi(\mathbf{x}) \approx -\nabla \ln(p_s(\mathbf{x})). \quad (18)$$

By multiplying by an arbitrary function $f(\mathbf{x})$ and integrating with respect to the invariant measure, we arrive at the empirically testable relation

$$\beta \langle \psi(\mathbf{x}) f(\mathbf{x}) \rangle \approx \langle \nabla f(\mathbf{x}) \rangle \quad (19)$$

that can be used as an error estimation for the optimal choice of the basis. The arbitrary function $f(\mathbf{x})$ can be tailored to specific problem at hand. Notably, for the specific case $f(\mathbf{x}) = \psi(\mathbf{x})$, Eq. (19) translates to

$$\mathbf{C}_{\psi\psi}(0)\beta \approx \langle \nabla \psi \rangle.$$

The error

$$\delta_i = \|\mathbf{C}_{\psi\psi}(0)\beta_i - \langle \partial_i \psi \rangle\| \quad (20)$$

is minimized by the linear regression $\beta = \mathbf{C}_{\psi\psi}(0)^{-1} \langle \nabla \psi \rangle$, which consistently is also the definition of β based on Eq. (9).

We can find further constraints introducing the primitives of the dictionary with respect to the j th variable $\Psi_{n,j}(\mathbf{x})$, i.e.,

$$\frac{d\Psi_{n,j}(\mathbf{x})}{dx_j} = \psi_n(\mathbf{x}).$$

Finally, we obtain the following approximation from Eq. (18):

$$\sum_n \beta_{n,j} \langle \Psi_{n,j}(\mathbf{x}) \rangle \approx - \int p_s(\mathbf{x}) \ln[p_s(\mathbf{x})] d\mathbf{x}. \quad (21)$$

The right-hand side is the Shannon entropy S of the stationary measure. Thus, the proper choice of basis function must minimize also the error on the entropy:

$$S_j = \sum_n \beta_{n,j} \langle \Psi_{n,j}(\mathbf{x}) \rangle \approx S.$$

Albeit estimating the stationary entropy from empirical measures is in general a challenging problem [37], several numerical approaches have been proposed in the recent years [38]. As a proof of concept, we will use the gradient

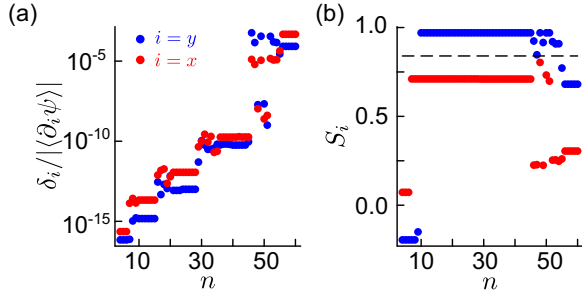


FIG. 4. (a) Normalized residual error for variables x and y in the gradient system (15), plotted against dictionary size. (b) Estimated entropy from Eq. (21) as a function of dictionary size. The dashed black line denotes the exact theoretical entropy computed from the stationary distribution.

system (15) for which the stationary distribution, hence the entropy S , is known analytically. As we increase the size of the dictionary, we see in Fig. 4(a) that the residual error increases, due to the increasing numerical error in the correlation matrix inversion. For this reason, it is difficult to use this constraint alone as stop criteria for the convergence of the response. On the contrary, the estimated entropy for both x and y quickly saturates close to the exact value when the cubic terms are introduced in the basis dictionary in accordance with the observation we made in the previous section. Even if we do not have in general access to the true value of S , the constraints (21) apply for all degrees of freedom; thus, it must hold in general that

$$S_i - S_j \approx 0. \quad (22)$$

We tested this idea for the challenging case of the Lorenz system with additional uncorrelated white noise, as in Ref. [16]:

$$\begin{aligned} dX &= [\sigma(Y - X)]dt + \sqrt{2\epsilon}dW_x, \\ dY &= [X(\rho - Z) - Y]dt + \sqrt{2\epsilon}dW_y, \\ dZ &= [XY - \beta Z]dt + \sqrt{2\epsilon}dW_z, \end{aligned} \quad (23)$$

with the parameters of the standard chaotic case ($\sigma, \rho, \beta = 10, 28, 8/3$) and noise $\epsilon = 8.0$ [39]. With the aim of reproducing the response function $R_{y \rightarrow x}(t)$, we sampled $\mathbf{C}_{\psi\psi}(\mathbf{0})$ and $\langle \nabla \psi \rangle$ using a large basis of function with polynomials up to order 11. Sorting the resulting β by absolute value we observe in Fig. 5(a), we see a sort of continuum of β that makes it difficult to select an *a priori* threshold for the dictionary size. The distance from the exact response function, obtained by perturbing the dynamics, and the one estimated using our approach quickly drop as more and more elements are included in the dictionary [see Fig. 5(b)]. At the same time, however, we see that after a plateau adding more elements in the dictionary results in larger errors in the response estimate. Also, in this case, the residual error alone [Fig. 5(c)] is not enough to decide when to stop because it increases on average with the dictionary size. On the other hand, the difference in entropy estimations gives very clear signals [see Fig. 5(d)], especially when there are too many elements of the dictionary. To show this, we selected three different sizes, i.e., $n = 63$ (red), $n = 256$ (green), and $n = 293$ (blue), which correspond, respectively, to two local minima and a local maximum of $|S_x - S_y|$. Convergence in the response is already found at $n = 63$ [see Fig. 5(e)], as correctly signaled. This also applies for the much larger basis of $n = 256$. On the contrary, for the

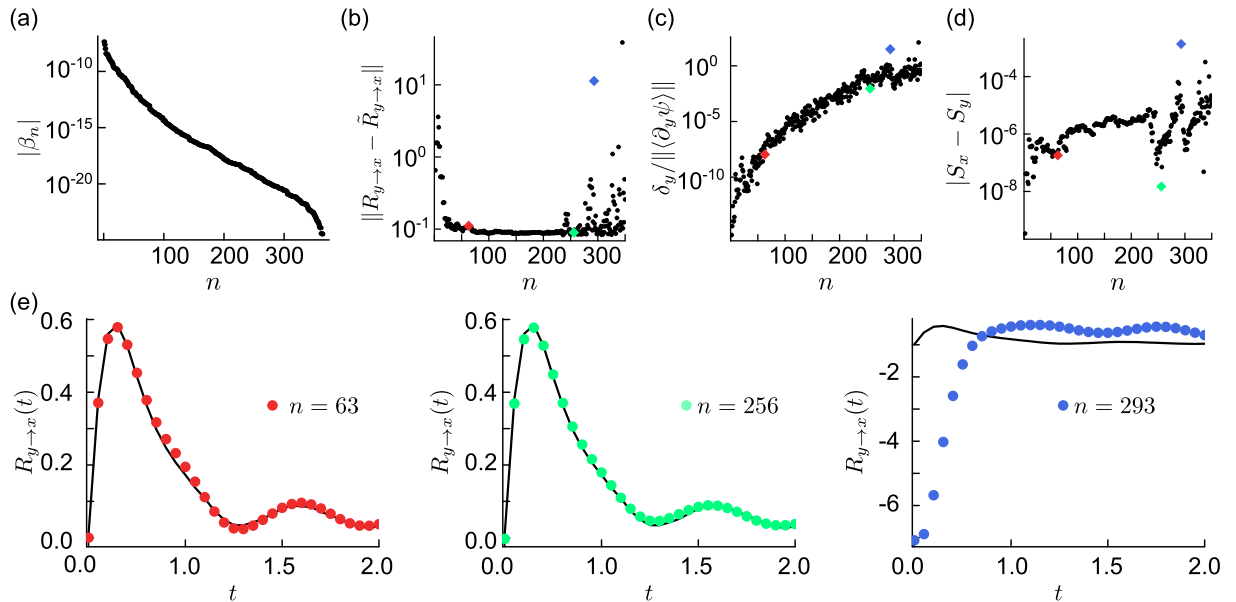


FIG. 5. (a) Sorted absolute values of coefficients $|\beta_n|$ for the stochastic Lorenz system. We are considering here only the weights of the response of the perturbation of y to x . If we indicate 1, 2, and 3 as the indexes for variables x , y , and z , respectively, β_n in figure would be in matrix notation $\beta_{n,2}$. (b) Relative error between exact response and n -element dictionary estimates. (c) Residual error response approximations. (d) Difference between estimated entropy values as in Eq. (22). (e) Comparison of exact response (black curve) and approximated responses (colored scatter points) across three dictionary sizes.

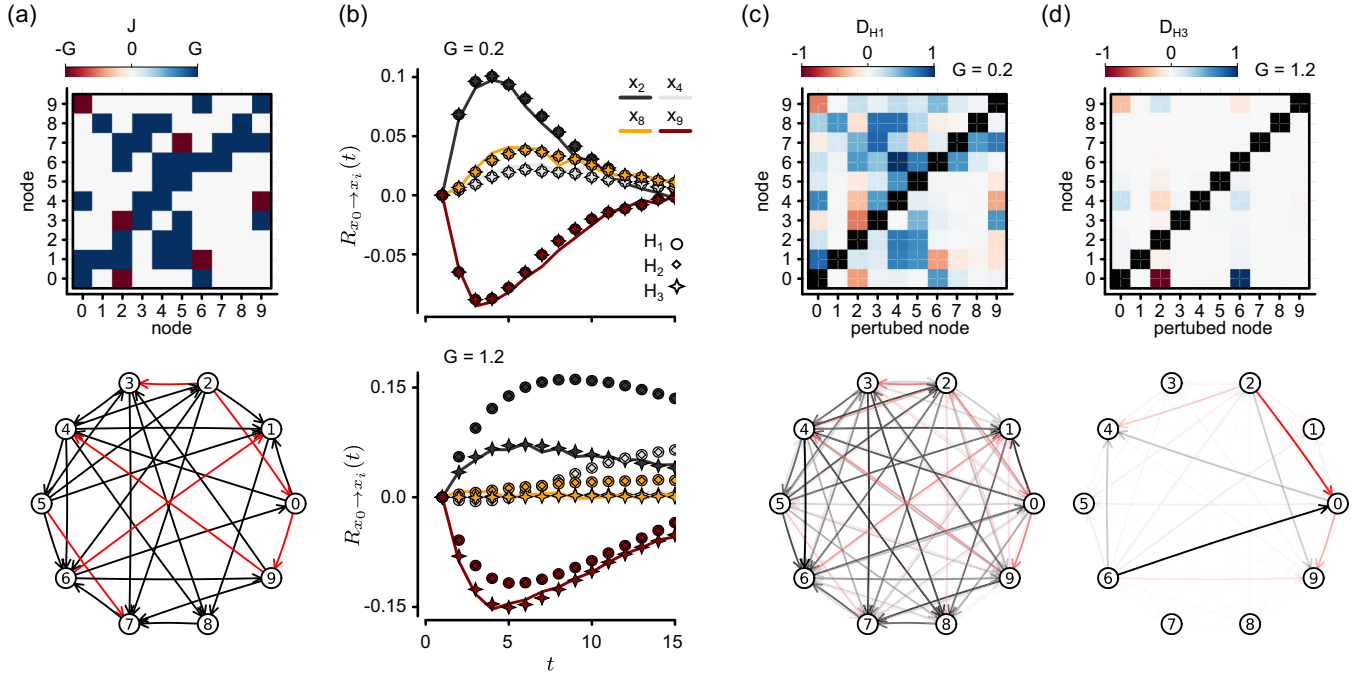


FIG. 6. (a) Top: Synaptic matrix for the ten-dimensional example system described in Eq. (26). Here, the parameter G scales the connection strengths, thereby modulating the system's nonlinearity. Bottom: The corresponding directed graph, where red arrows indicate connections with negative weights. (b) Response curves following perturbations to the first node (x_1) for two different levels of nonlinearity. Solid lines show the estimates obtained via the perturbative approach, while scatter points represent the estimates from Eq. (9) using observables from the Hermite basis functions H_1 , H_2 , and H_3 . Estimated causality matrices (top panels) and their associated graphs (bottom panels) for parameter values $G = 0.2$ (c) and $G = 1.2$ (d). In the graphs, red edges denote negative causal dependences, and the transparency of each edge reflects the intensity of the causal connection. The system is simulated for 5000 steps, the time resolution is set at $dt = 0.01$ and perturbation amplitude $\epsilon = 0.01$, with the stochastic Heun method for integration.

case $n = 293$, where the entropy difference is maximal, the error in the response function estimate is also large. In conclusion, while more work still needs to be done to verify the usefulness of these constraints for other systems, we observe that the convergence of the self-consistent error on the entropy $|S_x - S_y|$ follows the convergence of the error on the response. As such, it might be used as a proxy for the latter to establish criteria for avoiding spectral pollution.

All our observations, including the self-consistent constraints that can guide the dictionary optimization, open the door to an effective way to empirically estimate causality between different degrees of freedom from empirical measurement of unperturbed stochastic systems, as we will see in the next section.

From response to causality. It is possible to use the response function as a measure of interventional causality. As done in Ref. [14], we can measure the strength of the causal dependence of the variable x_i on the variable x_j as the total response that the former has consequently a perturbation of the latter, i.e.,

$$D_{ij} = \int_0^\infty R_{x_j \rightarrow x_i}(t) dt, \quad (24)$$

then, using our result (9), we have

$$\mathbf{D} \approx \left[\int_0^\infty \mathbf{C}_{\mathbf{x}\psi}(t) dt \right] \mathbf{C}_{\psi\psi}(0)^{-1} \langle \nabla \psi \rangle, \quad (25)$$

thus relating correlation and causality. The better the estimation of the response function, the better the estimation of the causal link between variables will be. It is also noteworthy that by selecting the eigenfunctions ϕ of $\mathcal{L}$ as the basis, we derive

$$D_{ij} \approx - \sum_k \alpha_{ik} \frac{\langle \partial_j \phi_k \rangle}{\lambda_k},$$

where λ_k denotes the eigenvalues of $\mathcal{L}$. In essence, this indicates that the causal relationship D_{ij} can be interpreted as a summation of contributions from x_j across all the eigenmodes. These are quantified as $\langle \partial_j \phi_k \rangle$ and are weighted by the significance of each eigenmode, i.e., the associated timescale $1/\lambda_k$. To investigate the behavior of this causality measure in complex high-dimensional systems, we consider the prototype of a biologically inspired network of neurons [40,41]. The system's state is represented by the vector $\mathbf{X}(t)$, whose element $X_i(t)$ corresponds to the activity of the i th unit, evolving according to the nonlinear dynamics:

$$d\mathbf{X} = [\tanh(G\mathbf{J}\mathbf{X}) - \mathbf{X}]dt + \sigma d\mathbf{W}, \quad (26)$$

where $\mathbf{J} \in \mathbb{R}^{N \times N}$ is the synaptic matrix governing the recurrent input currents and $\mathbf{W}(t)$ is an N -dimensional uncorrelated Wiener process. This dynamical framework also serves as the foundation for several modern machine-learning approaches [42–44]. For our test set, we consider $N = 10$ nodes, sparsely connected with a synaptic strength G [Fig. 6(a)]. These connections may be excitatory (positive) or inhibitory

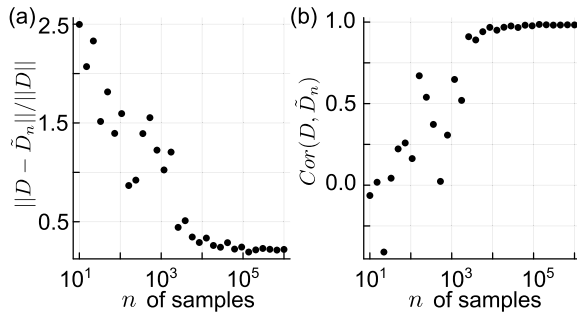


FIG. 7. Relative distance measured by the Frobenius norm of the real matrix D and the one estimated from n samples, $\tilde{D}_n$, is shown in panel (a), while element-wise correlation is reported in panel (b).

(negative), influencing the dynamics. The factor $G \in \mathbb{R}$ modulates the recurrent current, thereby controlling the level of nonlinearity in the system. Although the connection pattern remains invariant, the system's response can vary significantly based on the value of G . As expected, for small values of G , the system behaves nearly linearly, and the first-order observables H_1 can adequately capture the linear response. For larger values of G , the third-order term of the hyperbolic tangent function becomes crucial for describing the system's behavior, necessitating the inclusion of higher-order terms, as shown in Fig. 6(b). In cases with low nonlinearity, the causality links closely resemble the synaptic matrix [Fig. 6(c)]. However, as nonlinearity increases, the information in the connectivity matrix diminishes. As seen in Fig. 6(d), for high values of G , the primary causal links reduce in number, further amplified by the nonlinearity. As a result, paths of higher lengths become crucial, and a node without a direct connection may exert a stronger influence than nodes with direct connections. For example, in Fig. 6(d), we observe an inhibitory influence of node "6" on node "9," despite the positive connection, and an excitatory influence emerges on node "4." All our results are implicitly based on the capabilities of estimating expectation values from a large collection of samples. In practical context, however, the number of samples is usually limited; thus, it is important to verify the convergence of the causality matrix as the sample size is increased. In Fig. 7, we show that for system (26) already 10^4 samples are enough to converge to the right D . While the rate of convergence depends strongly on the specific systems under examination, this shows that even for such high-dimensional systems, a small sample size might be enough to infer proper causal relations.

Conclusions and future directions. By establishing a link between the response function and the Koopman operator, we derived an alternative way to compute the response function for a general class of nonlinear stochastic systems. Our result is rooted in a spectral approach that, albeit quite old [45,46], is now facing renewed interest and widespread adoption [33,47]. Besides all the possible issues that are common to causality inference in general, such as non-Markovianity [48] and nonstationarity, the major limitation of our result is in the proper choice of the library of basis functions. In fact, while convergence can be formally guaranteed when the size of the library goes to infinity, in practice the best would be to find the smallest set of functions possible to avoid spectral

pollution and overfitting. A small size in the library is also essential in systems of high dimensions to avoid computational cost. We found several self-consistent constraints that strongly guide dictionary selection and optimization, as we showed for several examples. While we will leave further investigation of related issues to a future work, we would like to point toward interesting opportunities offered by the great approximation capabilities of machine-learning algorithms and deep neural networks. For example, in Refs. [33,49–51] the authors showed that already a simple architecture like an autoencoder is capable of finding, by minimizing the loss function

$$\|E[\psi(\mathbf{X}_t)|\mathbf{X}_0] - \mathbf{K}(t)\psi(\mathbf{X}_0)\|_2$$

or its variants, a satisfactory low-dimensional representation of the Koopman operator for the chaotic Lorenz dynamical system and other nonlinear examples. The constraints we found can equally be used as loss function in the same settings. The advantage of using simple architectures is that, once the weights of the network are known, the matrix $\partial_{x_i}\psi_j(\mathbf{x})$ can be computed analytically. In general, it is true that the challenges inherent in applying the Koopman framework—or, more broadly, any pseudospectral decomposition—also affect our approach. However, the regression theorem and our examples demonstrate that it may be possible, in principle, to identify a small number of virtual degrees of freedom that approximately linearize the drift. This does not imply that we have fully reconstructed the entire spectrum of the Koopman operator, which is clearly infinite. Nonetheless, our results suggest that it might still be feasible to achieve highly accurate estimates of the response functions and consequently of causal links. In other words, an incomplete reconstruction of the Koopman operator can still lead to an exact reproduction of the response functions for appropriately chosen degrees of freedom—a point that should be considered when dealing with high-dimensional systems, as we have shown here for several examples. The causal measure explored in this Letter is highly adaptable to various extensions. Depending on the timescale of interest, one can limit the response integral to different time ranges, looking, for example, for short-term and long-term causal links. Furthermore, dynamical systems can exhibit different behaviors depending on the basin of attraction considered. By sampling trajectory data from specific regions of the state space, the analysis can be further tailored to particular conditions. The concept of information flow between degrees of freedom, first introduced in Ref. [3], has recently found applications in causal detection across various fields, ranging from neuroscience [52] to climate change [53] and quantum physics [54]. This approach, in its most general form, requires knowledge of the drift and diffusion terms, as well as the marginal stationary distributions. In fact, the equations defining the information flow share some similarities with the generalized Fluctuation-Dissipation Theorem and for this reason in practical cases suffer from the same limitations of the latter that originally motivated our work. Recently [5], a linear approximation of the drift terms—akin to the H_1 assumption in our notation—has been proposed. While the direct connection between information flow and the response function, if it exists, has yet to be fully understood, we anticipate that our approach could be readily applied to this

quantity as well; for an approach in this direction, we refer to Ref. [55]. Importantly, it enables a blind, data-driven analysis that can potentially overcome the limitations inherent in linear assumptions—limitations that we have already explored here in the context of the response function. Being able to infer the causal network of a set of measured variables without the need of perturbing the system is certainly powerful. Thus, given the general premises and the data-driven nature of our results, we expect that this approach will be a valuable tool and will find several applications in multiple domains of science.

Acknowledgments. We are grateful to Maurizio Mattia, Guido Gigante, Roberto Benzi, Angelo Vulpiani, and Marco Baldovin for their valuable suggestions and comments on an earlier version of this manuscript. Work funded by NextGenerationEU and MUR (PNRR-M4C2I1.3) project MNESYS (PE0000006-DD 1553 11.10.2022) and project EBRAINS-Italy (IR0000011-DD 101 16.6.2022).

Appendix A: The Koopman matrix for system (14). Consider the general system with a linear drift and arbitrary diffusion and natural boundary conditions:

$$dX_i = A_{ij}X_j dt + b_{ij}(X)dW_j, \quad (A1)$$

where we used Einstein notation for repeated indices and dW_j are independent Wiener noise sources. Defining the diffusion matrix as $D = b^T b/2$, the evolution in time of probability density $p(X, t|X_0)$ is governed by the following Fokker-Planck equation [25]:

$$\partial_t p = -A_{ij}\partial_i(x_j p) + \partial_i\partial_j(D_{ij}p). \quad (A2)$$

Multiplying (A2) by x_k , and taking the average $\langle \cdot \rangle$, and applying integration by parts, the diffusion part is zero even in the multiplicative case, since

$$\int x_k \partial_i \partial_j D_{ij} p = \int \partial_i \partial_j (x_k) D_{ij} p = 0.$$

Thus, the moments follow the linear dynamics:

$$\partial_t E[X_{i,t}|X_0] = A_{ij}E[X_{j,t}|X_0], \quad (A3)$$

a result commonly referred to as regression theorem [24]. The solution can be written using the exponential matrix

$$E[X_{i,t}|X_0] = e^{At}X_0.$$

Thus, following the definition, the Koopman matrix for this set of observable is $K(t) = e^{At}$.

Appendix B: Response function from Koopman decomposition. Consider a set of d observables $\{\psi_k(\mathbf{x})\}_{k=1}^d$ ideally spanning a subspace containing the state of the system at all times and that admits the family of $d \times d$ matrices $K(t)$ that approximate the action of the Koopman operators K^t .

The average observable evolution in this space, with initial condition X_0 , is described as

$$E[\psi(X_t)|X_0] \approx K(t)\psi(X_0).$$

Under the semigroup assumption, it would hold the relation $K(t) = e^{Lt}$, with L approximating the action of the infinitesimal Koopman generator $\mathcal{L}$. Note that for $d \rightarrow \infty$,

convergence is guaranteed and L becomes isospectral to $\mathcal{L}$. In the particular case in which the observables ψ_k are the linear evolving eigenfunctions ϕ_k , K would be the diagonal matrix of the associated eigenvalues λ_k .

Given the coordinates of the state of the system in this space $\mathbf{x} = \alpha\psi$, the expected value of the associated stochastic process up to time t is

$$E[X_{i,t}|X_0] = \sum_{n=1}^d \sum_{m=1}^d \alpha_{in} K_{nm}(t) \psi_m(X_0).$$

The effect of perturbing the function (virtual variable) $\psi_i[X_0] \rightarrow \psi_i[X_0] + \epsilon_\psi$ on the physical variable x_j is then described as

$$E_p[X_{j,t}|X_0] = E[X_{j,t}|X_0] + \tilde{R}_{\psi_i \rightarrow x_j}(t) \epsilon_\psi,$$

where we introduce $\tilde{R}_{\psi_i \rightarrow x_j}(t)$ as the virtual response function defined by

$$\tilde{R}_{\psi_i \rightarrow x_j}(t) = \lim_{\epsilon_\psi \rightarrow 0} \frac{\langle X_j(t) \rangle_p - \langle X_j(t) \rangle}{\epsilon_\psi} = [\alpha K(t)]_{ji},$$

where the brackets include the average with respect to the distribution of the initial conditions. Note that in case of ψ_k eigenfunctions and semigroup condition, it holds $\tilde{R}_{\psi_i \rightarrow x_j}(t) = \alpha_{ji} e^{\lambda_i t}$.

Let us now look at physical perturbations. Let us consider the perturbation $X_{i,0} \rightarrow X_{i,0} + \epsilon$. By expanding $\psi_k[X_t]$ around the unperturbed system, we find

$$\begin{aligned} E_p[X_{j,t}|X_0] &= E[X_{j,t}|X_0] + \epsilon \sum_{m=1}^d \left. \frac{\partial \psi_m(\mathbf{x})}{\partial x_i} \right|_{X_0} [\alpha K(t)]_{jm} \\ &= E[X_{j,t}|X_0] + \epsilon \sum_{m=1}^d \left. \frac{\partial \psi_m(\mathbf{x})}{\partial x_i} \right|_{X_0} \tilde{R}_{\psi_m \rightarrow x_j}(t). \end{aligned}$$

Averaging over the initial condition distribution, by the linear response definition we find a relation that links the physical to the virtual response function:

$$R_{x_i \rightarrow x_j}(t) = \sum_{m=1}^d \left\langle \left. \frac{\partial \psi_m(\mathbf{x})}{\partial x_i} \right|_{X_0} \right\rangle \tilde{R}_{\psi_m \rightarrow x_j}(t).$$

Upon substituting the expression for the virtual response from Eq. (8), we obtain the main result:

$$R_{x_i \rightarrow x_j}(t) = \sum_{m=1}^d \left\langle \left. \frac{\partial \psi_m(\mathbf{x})}{\partial x_i} \right|_{X_0} \right\rangle [C_{\mathbf{x}\psi}(t) C_{\psi\psi}(0)^{-1}]_{jm}.$$

- [1] N. Wiener, The theory of prediction, in *Modern Mathematics for the Engineer*, edited by E. F. Beckenbach (McGraw-Hill, New York, 1956).
- [2] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**, 424 (1969).
- [3] X. S. Liang and R. Kleeman, Information transfer between dynamical system components, *Phys. Rev. Lett.* **95**, 244101 (2005).
- [4] X. S. Liang, Unraveling the cause-effect relation between time series, *Phys. Rev. E* **90**, 052150 (2014).
- [5] X. S. Liang, Normalized multivariate time series causality analysis and causal graph reconstruction, *Entropy* **23**, 679 (2021).
- [6] X. S. Liang, Information flow within stochastic dynamical systems, *Phys. Rev. E* **78**, 031113 (2008).
- [7] X. S. Liang, Information flow and causality as rigorous notions *ab initio*, *Phys. Rev. E* **94**, 052201 (2016).
- [8] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 2003).
- [9] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls, Causal inference for time series, *Nat. Rev. Earth Environ.* **4**, 487 (2023).
- [10] A. K. Seth, A. B. Barrett, and L. Barnett, Granger causality analysis in neuroscience and neuroimaging, *J. Neurosci.* **35**, 3293 (2015).
- [11] G. V. Vinci and R. Benzi, Economic complexity: Correlations between gross domestic product and fitness, *Entropy* **20**, 766 (2018).
- [12] M. A. Javarone, G. Di Antonio, G. V. Vinci, R. Cristodaro, C. J. Tessone, and L. Pietronero, Disorder unleashes panic in bitcoin dynamics, *J. Phys.: Complexity* **4**, 045002 (2023).
- [13] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Fluctuation–dissipation: Response theory in statistical physics, *Phys. Rep.* **461**, 111 (2008).
- [14] M. Baldovin, F. Cecconi, and A. Vulpiani, Understanding causation via correlations and linear response theory, *Phys. Rev. Res.* **2**, 043436 (2020).
- [15] M. Baldovin, F. Cecconi, A. Provenzale, and A. Vulpiani, Extracting causation from millennial-scale climate fluctuations in the last 800 kyr, *Sci. Rep.* **12**, 15320 (2022).
- [16] M. Baldovin, L. Caprini, and A. Vulpiani, Handy fluctuation-dissipation relation to approach generic noisy systems and chaotic dynamics, *Phys. Rev. E* **104**, L032101 (2021).
- [17] D. Lucente, A. Baldassarri, A. Puglisi, A. Vulpiani, and M. Viale, Inference of time irreversibility from incomplete information: Linear systems and its pitfalls, *Phys. Rev. Res.* **4**, 043103 (2022).
- [18] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, Spectral analysis of nonlinear flows, *J. Fluid Mech.* **641**, 115 (2009).
- [19] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, and C. Schütte, Data-driven approximation of the Koopman generator: Model reduction, system identification, and control, *Physica D* **406**, 132416 (2020).
- [20] M. Santos Gutiérrez, V. Lucarini, M. D. Chekroun, and M. Ghil, Reduced-order models for coupled dynamical systems: Data-driven methods and the Koopman operator, *Chaos* **31**, 053116 (2021).
- [21] M. S. Gutiérrez and V. Lucarini, On some aspects of the response to stochastic and deterministic forcings, *J. Phys. A: Math. Theor.* **55**, 425002 (2022).
- [22] N. Zagli, M. Colbrook, V. Lucarini, I. Mezić, and J. Moroney, Bridging the gap between Koopmanism and response theory: Using natural variability to predict forced response, *arXiv:2410.01622*.
- [23] V. Lucarini, Interpretable and equation-free response theory for complex systems, *arXiv:2502.07908*.
- [24] C. Gardiner, *Stochastic Methods* (Springer, Berlin, 2009), Vol. 4.
- [25] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, Heidelberg, 1996).
- [26] G. V. Vinci and M. Mattia, Escape time in bistable neuronal populations driven by colored synaptic noise, *Phys. Rev. Res.* **7**, 023172 (2025).
- [27] B. Shizgal, *Spectral Methods in Chemistry and Physics. Applications to Kinetic Theory and Quantum Mechanics. Scientific Computation* (Springer, Dordrecht, 2015).
- [28] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, *J. Nonlinear Sci.* **25**, 1307 (2015).
- [29] M. J. Colbrook, L. J. Ayton, and M. Szőke, Residual dynamic mode decomposition: Robust and verified Koopmanism, *J. Fluid Mech.* **955**, A21 (2023).
- [30] M. J. Colbrook, B. Roman, and A. C. Hansen, How to compute spectra with error control, *Phys. Rev. Lett.* **122**, 250201 (2019).
- [31] U. C. Täuber, *Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior* (Cambridge University Press, Cambridge, UK, 2014).
- [32] $\langle f(x) \rangle = \int_{\mathcal{D}} f(x) p(x, t | x_0) p_s(x_0) dx dx_0$.
- [33] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, Modern Koopman theory for dynamical systems, *SIAM Rev.* **64**, 229 (2022).
- [34] C. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics* (Springer Science & Business Media, Cambridge, UK, 2004).
- [35] S. Klus, A. Bittracher, I. Schuster, and C. Schütte, A kernel-based approach to molecular conformation analysis, *J. Chem. Phys.* **149**, 244109 (2018).
- [36] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, NY, 2003).
- [37] A. Vulpiani, F. Cecconi, and M. Cencini, *Chaos: From Simple Models to Complex Systems. Series on Advances in Statistical Mechanics* (World Scientific Publishing, Singapore, 2009).
- [38] A. Namdari and Z. Li, A review of entropy measures for uncertainty quantification of stochastic processes, *Adv. Mech. Eng.* **11**, 1687814019857350 (2019).
- [39] We verified that our results hold also for the case of small noise $\epsilon \sim 0.1$, but convergence requires almost double the number of dictionary functions, likely due the singular nature of the stationary measure.
- [40] A. Treves, Mean-field analysis of neuronal spike dynamics, *Network* **4**, 259 (1993).
- [41] M. Mattia and P. Del Giudice, Population dynamics of interacting spiking neurons, *Phys. Rev. E* **66**, 051917 (2002).

- [42] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [43] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach, *Phys. Rev. Lett.* **120**, 024102 (2018).
- [44] G. Di Antonio, T. Gili, A. Gabrielli, and M. Mattia, Linearizing and forecasting: A reservoir computing route to digital twins of the brain, *bioRxiv* 2025.
- [45] B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, *Proc. Natl. Acad. Sci. USA* **17**, 315 (1931).
- [46] J. von Neumann, Zur operatorenmethode in der klassischen mechanik, *Ann. Math.* **33**, 587 (1932).
- [47] M. Budišić, R. Mohr, and I. Mezić, Applied Koopmanism, *Chaos* **22**, 047510 (2012).
- [48] C. Capone, G. Gigante, and P. Del Giudice, Spontaneous activity emerging from an inferred network model captures complex spatio-temporal dynamics of spike data, *Sci. Rep.* **8**, 17056 (2018).
- [49] B. Lusch, J. N. Kutz, and S. L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, *Nat. Commun.* **9**, 4950 (2018).
- [50] D. J. Alford-Lago, C. W. Curtis, A. T. Ihler, and O. Issan, Deep learning enhanced dynamic mode decomposition, *Chaos* **32**, 033116 (2022).
- [51] S. E. Otto and C. W. Rowley, Linearly recurrent autoencoder networks for learning dynamics, *SIAM J. Appl. Dyn. Syst.* **18**, 558 (2019).
- [52] D. T. Hristopulos, A. Babul, S. Babul, L. R. Brucar, and N. Virji-Babul, Disrupted information flow in resting-state in adolescents with sports related concussion, *Front. Hum. Neurosci.* **13**, 419 (2019).
- [53] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorrioz, and X. S. Liang, On the causal structure between CO₂ and global temperature, *Sci. Rep.* **6**, 21691 (2016).
- [54] B. Yi and S. Bose, Quantum Liang information flow as causation quantifier, *Phys. Rev. Lett.* **129**, 020501 (2022).
- [55] C. A. Pires, D. Docquier, and S. Vannitsem, A general theory to estimate information transfer in nonlinear systems, *Physica D* **458**, 133988 (2024).