

# Multidimensional measures of impulsively driven stochastic systems based on the Kullback-Leibler distance

Saar Rahav and Shaul Mukamel

*Department of Chemistry, University of California, Irvine, California 92697, USA*

(Received 13 October 2009; published 17 March 2010)

By subjecting a dynamical system to a series of short pulses and varying several time delays, we can obtain multidimensional characteristic measures of the system. Multidimensional Kullback-Leibler response function (KLRF), which are based on the Kullback-Leibler distance between the initial and final states, are defined. We compare the KLRF, which are nonlinear in the probability density, with ordinary response functions obtained from the expectation value of a dynamical variable, which are linear. We show that the KLRF encode different level of information regarding the system's dynamics. For overdamped stochastic dynamics two-dimensional KLRF shows a qualitatively different variation with the time delays between pulses, depending on whether the system is initially in a steady state or in thermal equilibrium.

DOI: [10.1103/PhysRevE.81.031116](https://doi.org/10.1103/PhysRevE.81.031116)

PACS number(s): 05.40.-a, 89.70.Cf, 02.50.Ey

## I. INTRODUCTION

One of the most common ways to investigate the properties of a dynamical system is to study how it responds to controlled external perturbations. The response of a system to a weak perturbing field is related to its equilibrium fluctuations by the celebrated fluctuation dissipation relation [1,2]. The response provides a direct measure of system dynamics and fluctuations.

In a time-domain response measurement one uses a series of impulsive perturbations (Fig. 1) and records some property of the system as a function of their time delays. Impulsive perturbations make it possible to study the free-dynamical evolution of the system during the time delays unmasked by the time profile of the perturbing field. Furthermore, the joint dependence on several time delays can be used to separate the contributions of different dynamical pathways. Due to its dependence on multiple time delays this method is termed multidimensional.

The response of a system is typically measured by the expectation value of some operator [1–3]. This is a linear functional of the system probability density (or the density matrix for quantum systems). Multidimensional response have had considerable success in nonlinear spectroscopy due to the ability to control and shape optical fields. Applications range from spin dynamics in NMR [4], vibrational dynamics of proteins in infrared systems and electronic energy transfer in photosynthetic complexes as probed by visible pulses [5,6]. These span a broad range of time scales from milliseconds to femtoseconds.

Several nonlinear functionals of probability densities, which have interesting physical interpretations, are known. One such quantity is the Von-Neuman entropy  $\tilde{S}(\rho) = -\text{Tr } \rho \ln \rho$ . A related quantum nonlinear measure called the concurrence serves as a measure of quantum entanglement [7,8]. The Kullback-Leibler distance (KLD) or relative entropy,  $\text{Tr } \rho_0 \ln \rho_0/\rho$ , which compares one probability distribution to another, is a nonlinear measure that had been found useful in many applications. This paper aims at developing multidimensional measures based on the KLD.

Numerous applications of the KLD differ in the probability distributions involved. The ratio of the probability of a

stochastic path and its reverse at a steady state has been connected to a change in entropy [9–12]. For an externally driven system a similar quantity was found to be related to the work done on the system [13]. As a result, the KLD which compares the path distribution to a distribution of reversed paths is a measure of the lack of reversibility of a thermodynamical process.

For distributions in phase-space (as opposed to path space), the KLD between the density of a driven system and the density of a reversed process [14] or the distance between the driven density and the corresponding equilibrium density [15] (for the same value of parameters) were shown to be bounded by dissipated work in the process. The transfer of information through a stochastic resonance is quantified by the KLD between the probability distributions with and without the external input [16,17]. The ability of neuronal networks to retain information about past events was characterized by the Fisher information [18], which is closely related to the KLD between a distribution and the one obtained from it by a small perturbation.

We shall examine the response of a system to impulsive perturbations which drive it out of a stationary (steady state or equilibrium) state. The KLD between the distribution before and after the perturbation does not correspond to an entropy or work. However, since it compares the perturbed and unperturbed densities, it characterizes how “easy” it is to drive the system away from its initial state. In ordinary response theory, one compared the expectation values of some operator taken over the perturbed and unperturbed probability density. This depends on the specific properties of the

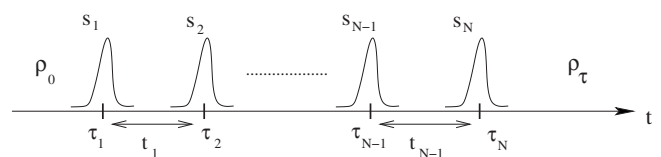


FIG. 1. Heuristic schematic of the process studied. A system is prepared at an initial probability distribution  $\rho_0$  and subjected to impulsive perturbations. The  $k$ th pulse is centered at  $\tau_k$  and its strength is denoted by  $s_k$ .  $t_j = \tau_{j+1} - \tau_j$  are the time intervals between successive pulses.

observed operator. The KLD is a more robust measure for the effect of impulsive perturbations on the probability density. By expanding the KLD in the perturbation strength we obtain a hierarchy of Kullback-Leibler response functions (KLRFs). These differ qualitatively from the hierarchy of ordinary response functions (ORFs), since they are nonlinear in the probability density. The KLRF serve as a type of measures characterizing the dynamics and encoding different information than the ORF. For example, the second order KLRF, which we connect to the Fisher-information, is found to exhibit qualitatively different dependence on the time delays, depending on whether the system is perturbed out of a steady state or out of thermal equilibrium. This is in contrast to the corresponding ORF. The fluctuation dissipation relation [1,2], which is linear in the density matrix, can also distinguish between systems driven out of equilibrium and out of a steady state. The KLD offers a different window into this aspect.

The structure of the paper is as follows. In Sec. II we describe the multidimensional measures and present the two hierarchies of ORF and KLRF response functions. These are then calculated using a formal perturbation theory in the coupling strength to the external perturbation. In Sec. III we show that for systems undergoing overdamped stochastic dynamics the nonlinear KLRF are naturally described using a combination of the stochastic dynamics and its dual dynamics. In Sec. IV we extend the results of Sec. III to discrete Markovian systems with a finite number of states. Our results are discussed in Sec. V.

## II. MULTIDIMENSIONAL MEASURES FOR NONLINEAR RESPONSE BASED ON THE KULLBACK-LEIBLER DISTANCE

We consider a system initially at a stationary state (either equilibrium or a steady state), which is perturbed by a series of short pulses, as depicted in Fig. 1. The probability distribution describing the driven system at time  $\tau$ ,  $\rho(\tau)$ , depends parametrically on  $s_i$ , the strength of the  $i$ th pulse, as well as on the time differences between the pulses  $t_i \equiv \tau_{i+1} - \tau_i$  with  $\tau_{n+1} = \tau$ .

Ordinary response theory focuses on the expectation value of some observable,

$$\langle A \rangle(t) = \text{Tr}[\rho(t)A], \quad (1)$$

and its dependence on the parameters  $s_i, \tau_i$ . The lowest ORFs are

$$\mathcal{R}_i^{(1)}(\tau - \tau_i) \equiv \left. \frac{\partial \langle A \rangle}{\partial s_i} \right|_{s=0}, \quad (2)$$

$$\mathcal{R}_{ij}^{(2)}(\tau - \tau_i, \tau_i - \tau_j) \equiv \left. \frac{\partial^2 \langle A \rangle}{\partial s_i \partial s_j} \right|_{s=0}, \quad (3)$$

$$\mathcal{R}_{ijk}^{(3)}(\tau - \tau_i, \tau_i - \tau_j, \tau_j - \tau_k) \equiv \left. \frac{\partial^3 \langle A \rangle}{\partial s_i \partial s_j \partial s_k} \right|_{s=0}, \quad (4)$$

and so forth. The time differences in Eqs. (2)–(4) can be expanded in terms of the time delays between the pulses,

$\tau - \tau_i = \sum_{k=i}^n t_k$ .  $\mathcal{R}^{(j)}$  are used to investigate various properties of the unperturbed dynamics, such as the existence of excited modes and the relaxation back to a steady state.

Here, we focus on different, but closely related quantity. Instead of studying an expectation value of an observable, we focus on a quantity that compares the perturbed and unperturbed probability distributions. The KLD, also known as the relative entropy, is defined as

$$\mathcal{D}(\rho_0 \parallel \rho) \equiv \text{Tr} \rho_0 \ln \frac{\rho_0}{\rho} = \left\langle \ln \frac{\rho_0}{\rho} \right\rangle_0. \quad (5)$$

The KLD vanishes when the two distributions are equal  $\mathcal{D}(\rho \parallel \rho) = 0$  and is positive otherwise,  $\mathcal{D}(\rho' \parallel \rho) > 0$  for  $\rho'(x) \neq \rho(x)$  [19]. Note that the KLD is not a true distance since  $\mathcal{D}(\rho' \parallel \rho) \neq \mathcal{D}(\rho \parallel \rho')$  and it does not satisfy the triangle inequality.

The KLD measures the dissimilarity between two distributions. It had found many applications in the field of information theory [19]. For instance, the mutual information between two random variables  $x, y$  is  $\mathcal{D}[P(x, y) \parallel P(x)P(y)]$ , where  $P(x, y)$  is the joint distribution while  $P(x)$  [ $P(y)$ ] is the marginal distribution of  $x$  ( $y$ ). In the present application the Kullback-Leibler distance is used as a measure for the deviation of the driven system from its initial state.

In a manner similar to the definition of the ORF, we define a KLRF hierarchy by taking derivatives of the KLD with respect to pulse strengths and displaying them with respect to the time delays

$$\mathcal{Q}_i^{(1)}(\tau - \tau_i) \equiv \left. \frac{\partial \mathcal{D}}{\partial s_i} \right|_{s=0} = \left\langle \left. \frac{\partial}{\partial s_i} \ln \frac{\rho_0}{\rho} \right|_{s=0} \right\rangle_0, \quad (6)$$

$$\mathcal{Q}_{ij}^{(2)}(\tau - \tau_i, \tau_i - \tau_j) \equiv \left. \frac{\partial^2 \mathcal{D}}{\partial s_i \partial s_j} \right|_{s=0} = \left\langle \left. \frac{\partial^2}{\partial s_i \partial s_j} \ln \frac{\rho_0}{\rho} \right|_{s=0} \right\rangle_0, \quad (7)$$

$$\begin{aligned} \mathcal{Q}_{ijk}^{(3)}(\tau - \tau_i, \tau_i - \tau_j, \tau_j - \tau_k) &\equiv \left. \frac{\partial^3 \mathcal{D}}{\partial s_i \partial s_j \partial s_k} \right|_{s=0} \\ &= \left\langle \left. \frac{\partial^3}{\partial s_i \partial s_j \partial s_k} \ln \frac{\rho_0}{\rho} \right|_{s=0} \right\rangle_0. \end{aligned} \quad (8)$$

All the derivatives are calculated at  $s=0$ , and we have used the relation  $\lim_{s \rightarrow 0} \rho(\tau) = \rho_0$ . This is also true for all other  $s$  derivatives in the following. To keep the notation simple we will not state this explicitly. Higher order KLRF are defined similarly. It is important to note that the ORF are linear in  $\rho$ , whereas the KLD are nonlinear. We thus expect the KLD to carry qualitatively different information about the dynamics.

The second derivative [Eq. (7)], known as the Fisher Memory (or information) matrix, plays an important role in information theory since the Cramer-Rao inequality means that it is a measure of the minimum error in estimating the value of a parameter of a distribution [19]. The Fisher information has been used recently to analyze the survival of information in stochastic networks [18].

Conservation of probability implies that the first KLRF vanishes

$$Q_i^{(1)}(\tau - \tau_i) = - \int dx \rho_0 \frac{\partial}{\partial s_i} \ln \rho = - \int dx \frac{\partial}{\partial s_i} \rho = 0. \quad (9)$$

The second derivative, the Fisher memory matrix, is given by

$$Q_{ij}^{(2)}(\tau - \tau_i, \tau_i - \tau_j) = \int dx \rho_0^{-1} \frac{\partial \rho}{\partial s_i} \frac{\partial \rho}{\partial s_j} = \left\langle \frac{\partial \ln \rho}{\partial s_i} \frac{\partial \ln \rho}{\partial s_j} \right\rangle_0. \quad (10)$$

A straightforward calculation allows us to recast the third order derivative of  $\mathcal{D}$  in terms of products of lower order derivatives

$$Q_{ijk}^{(3)}(\tau - \tau_i, \tau_i - \tau_j, \tau_j - \tau_k) = \left\langle \frac{\partial \ln \rho}{\partial s_i} \frac{\partial^2 \ln \rho}{\partial s_j \partial s_k} \right\rangle_0 + \left\langle \frac{\partial \ln \rho}{\partial s_j} \frac{\partial^2 \ln \rho}{\partial s_i \partial s_k} \right\rangle_0 + \left\langle \frac{\partial \ln \rho}{\partial s_k} \frac{\partial^2 \ln \rho}{\partial s_i \partial s_j} \right\rangle_0 + \left\langle \frac{\partial \ln \rho}{\partial s_i} \frac{\partial \ln \rho}{\partial s_j} \frac{\partial \ln \rho}{\partial s_k} \right\rangle_0. \quad (11)$$

In what follows the derivatives will be calculated perturbatively in  $s_j$ . It is important to note that the  $N$ th derivative of  $\ln \rho$  has contributions from interaction with at most  $N$  pulses. The contribution from the linear component, which interacts with  $N$  pulses, has the same structure of the perturbation theory for observables (which is also linear). However, since  $\ln \rho$  is a non linear function of  $\rho$  the  $N$ th derivative contains a nonlinear contribution which is a product of lower order contributions for  $\rho$ . The KLRF encode qualitatively different information about the system dynamics in comparison to the ORF.

The time evolution of the probability distribution is given by

$$\frac{\partial \rho}{\partial \tau} = - \hat{\mathcal{L}}(\tau) \rho. \quad (12)$$

This formal equation is quite general and can describe either Hamiltonian (Unitary) or stochastic dynamics where the operator  $\hat{\mathcal{L}}$  will accordingly be the Liouville or the Fokker-Planck operator.

For a system subjected to a time-dependent weak perturbation we can write

$$\hat{\mathcal{L}}(\tau) = \hat{\mathcal{L}}_0 + \hat{\mathcal{L}}'(\tau), \quad (13)$$

where we assume that the unperturbed system is time independent and is initially in a steady state,  $\rho_0$ , so that  $\hat{\mathcal{L}}_0 \rho_0 = 0$ . We consider an impulsive perturbation of the form

$$\hat{\mathcal{L}}'(\tau) = - \sum_{i=1}^n s_i \delta(\tau - \tau_i) \hat{\mathcal{L}}_A, \quad (14)$$

where  $\hat{\mathcal{L}}_A$  describes the action of a pulse on the probability distribution and  $s_i$  is the overall strength of the  $i$ th pulse.

Using these definitions, the state of the system at time  $\tau$  can be expanded as a power series in the number of interactions with the pulses

$$\begin{aligned} \rho(\tau) &= \rho_0 + \delta \rho_1 + \delta \rho_2 + \dots \\ &= \rho_0 + \sum_{i=1}^n s_i \mathcal{S}^{(1)}(x; \tau - \tau_i) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^i s_i s_j \left(1 - \frac{\delta_{ij}}{2}\right) \mathcal{S}^{(2)}(x; \tau - \tau_i, \tau_i - \tau_j) + \dots \end{aligned} \quad (15)$$

The partial corrections for the density,  $\mathcal{S}^{(j)}$ , appearing in Eq. (15), contain all the information necessary for computing both the KLRF and ORF. They are given by

$$\mathcal{S}^{(1)}(x; t) \equiv \int dx_1 dx_2 \mathcal{U}_0(x, x_2; t) \mathcal{L}_A(x_2, x_1) \rho_0(x_1), \quad (16)$$

$$\begin{aligned} \mathcal{S}^{(2)}(x; t', t) &\equiv \int dx_1 \dots dx_4 \mathcal{U}_0(x, x_4; t') \mathcal{L}_A(x_4, x_3) \\ &\quad \times \mathcal{U}_0(x_3, x_2; t) \mathcal{L}_A(x_2, x_1) \rho_0(x_1). \end{aligned} \quad (17)$$

Here  $\hat{\mathcal{U}}_0(\tau - \tau') = \exp[-(\tau - \tau') \hat{\mathcal{L}}_0]$  is the free propagator of the unperturbed system. Conservation of probability requires that  $\int dx \delta \rho_i(x) = 0$ , which, in turn, means that  $\int dx \mathcal{S}^{(j)} = 0$ .

Equations (15)–(17) can be used to calculate the logarithmic derivatives, which then determine the KLRFs. We will only need the first two logarithmic derivatives, which are given by

$$\frac{\partial \ln \rho}{\partial s_i} = \frac{1}{\rho_0} \frac{\partial \delta \rho_1}{\partial s_i} = \frac{1}{\rho_0(x)} \mathcal{S}^{(1)}(x; \tau - \tau_i) \quad (18)$$

and

$$\begin{aligned} \frac{\partial \ln \rho}{\partial s_i \partial s_j} &= \frac{1}{\rho_0} \frac{\partial^2 \delta \rho_2}{\partial s_i \partial s_j} - \frac{1}{\rho_0^2} \frac{\partial \delta \rho_1}{\partial s_i} \frac{\partial \delta \rho_1}{\partial s_j} \\ &= \frac{1}{\rho_0(x)} \mathcal{S}^{(2)}(x; \tau - \tau_i, \tau_i - \tau_j) \\ &\quad - \frac{1}{\rho_0^2(x)} \mathcal{S}^{(1)}(x; \tau - \tau_i) \mathcal{S}^{(1)}(x; \tau - \tau_j). \end{aligned} \quad (19)$$

To calculate the ORF, we substitute Eq. (15) into Eq. (1), resulting in

$$\begin{aligned} \langle A \rangle_\tau &= \langle A \rangle_0 + \sum_{n=1}^{\infty} \int_{\tau_0}^{\tau} d\tau_n \int_{\tau_0}^{\tau_n} d\tau_{n-1} \dots \int_{\tau_0}^{\tau_2} d\tau_1 s(\tau_n) \dots s(\tau_1) \\ &\quad \times \mathcal{R}^{(n)}(\tau - \tau_n, \tau_n - \tau_{n-1}, \dots, \tau_2 - \tau_1). \end{aligned} \quad (20)$$

It is interesting to compare the KLRF with the ORF. Calculate  $Q^{(j)}$  and  $\mathcal{R}^{(j)}$  for  $j$  perturbing pulses. To leading order, we find

$$\mathcal{Q}_1^{(1)}(t_1) = 0,$$

$$\mathcal{R}^{(1)}(t_1) = \int dx A(x) \mathcal{S}^{(1)}(x; t_1). \quad (21)$$

At the next order, we compare the Fisher information to the second-order ORF,

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \int dx \rho_0^{-1}(x) \mathcal{S}^{(1)}(x; t_1 + t_2) \mathcal{S}^{(1)}(x; t_2) \quad (22)$$

and

$$\mathcal{R}^{(2)}(t_2, t_1) = \int dx A(x) \mathcal{S}^{(2)}(x; t_2, t_1). \quad (23)$$

The nondiagonal elements of  $\mathcal{Q}_{ij}^{(2)}$  depend on the two delay times. Expressions for the third order response functions are given in Appendix B 1.

Both  $\mathcal{R}^{(j)}$  and  $\mathcal{Q}^{(j)}$  depend on the same set of  $j$  time intervals with some important differences.  $\mathcal{Q}^{(1)}$  vanishes, while the linear response  $\mathcal{R}^{(1)}$  does not.  $\mathcal{Q}^{(2)}$  and  $\mathcal{R}^{(2)}$  have a different structure:  $\mathcal{R}^{(2)}$  can be calculated from the second order correction to the density (or  $\mathcal{S}^{(2)}$ ) while  $\mathcal{Q}^{(2)}$  is determined from a product of  $\mathcal{S}^{(1)}$ s describing the first-order interaction with different pulses. This difference reflects the non-linear dependence of the KLRF on  $\rho$  and also applies to higher orders.

We have chosen to use  $\mathcal{D}(\rho_0 \parallel \rho)$  as the measure for the effect of the perturbations.  $\mathcal{D}(\rho \parallel \rho_0)$  would have been equally suitable. However, as discussed in Appendix A, the leading order of both KLDs in the strength of the perturbation, i.e., their Fisher information, coincide. Therefore all the following results pertaining to the Fisher information would hold for either choice.

### III. APPLICATION TO OVERDAMPED STOCHASTIC DYNAMICS

In the following we use the formal results of Sec. II to calculate the leading order ORF and KLRF for a system undergoing overdamped stochastic dynamics. We show that the Fisher information is related to a forward-backward stochastic process. The backward part is driven by the  $\rho_0$ -dual process, which will be simply referred to as the dual in what follows. The Fisher information is found to exhibit qualitatively different properties for systems perturbed from equilibrium or from a steady state. We also use the eigenfunctions and eigenvalues of the dynamics to derive explicit expressions for several low-order ORF and KLRF.

In stochastic dynamics the probability density plays the role of a reduced density matrix, which depends on a few collective coordinates. In this reduced description the entropy  $-\text{Tr } \rho \ln \rho$  typically increases with time. This should be contrasted with a description which includes all the degrees of freedom, where the dynamics is unitary and the entropy does not change in time. For completeness unitary dynamics is discussed in Appendix A.

#### A. Fisher information for systems perturbed out of equilibrium vs steady state

The Fisher information can be represented in terms of the dual stochastic dynamics. This interesting property reflects its non-linear dependence on  $\rho$ . We examine a stochastic dynamics of several variables  $x_j$  given by

$$\frac{dx_i}{dt} = F_i(\mathbf{x}) + \xi_i(t; \mathbf{x}). \quad (24)$$

Here we use the Ito stochastic calculus.<sup>1</sup> The noise terms are assumed to be Gaussian with

$$\langle \xi_i(t) \xi_j(t') \rangle = G_{ij} \delta(t - t'). \quad (25)$$

with  $G(\mathbf{x})$  as a symmetric positive definite matrix. While for many systems this matrix does not depend on the coordinate,  $G_{ij} = \frac{2}{\gamma} k_B T \delta_{ij}$ , this assumption will not be used in what follows.

Equation (24) is equivalent to the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = - \sum_i \frac{\partial}{\partial x_i} F_i \rho + \sum_{ij} \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} G_{ij} \rho = - \hat{\mathcal{L}}_0 \rho. \quad (26)$$

In what follows we present the dual dynamics, which can be loosely thought as the time reversed dynamics: it have the same steady state, but with reversed steady-state current. We consider the current density

$$J_i = F_i \rho - \sum_j \frac{1}{2} \frac{\partial}{\partial x_j} G_{ij} \rho. \quad (27)$$

The Fokker-Planck equation can be written in terms of the current,

$$\frac{\partial \rho}{\partial t} = - \sum_i \frac{\partial J_i}{\partial x_i}. \quad (28)$$

The steady state is the solution of

$$\hat{\mathcal{L}}_0 \rho_0 = 0. \quad (29)$$

We write

$$\rho_0 \equiv e^{-\phi_0}, \quad (30)$$

which defines  $\phi_0$ . For systems at equilibrium  $\phi_0$  is simply the potential. However, this is not the case for general steady states. The steady-state current can be written as

$$\begin{aligned} J_i^{(0)} &= F_i \rho_0 - \sum_j \frac{1}{2} \frac{\partial}{\partial x_j} G_{ij} \rho_0 \\ &= \frac{1}{2} \sum_j G_{ij} \left( \sum_k 2G_{jk}^{-1} F_k - \sum_{kl} G_{jk}^{-1} \frac{\partial G_{kl}}{\partial x_l} + \frac{\partial \phi_0}{\partial x_j} \right) \rho_0. \end{aligned} \quad (31)$$

After some algebra, the generator of the stochastic dynamics

<sup>1</sup>The Ito and Stratonovich calculus offer two different recipes of interpreting Eq. (24). Both methods are equally viable as long as they are used in a consistent manner. Details can be found in Ref. [22].



can be written in terms of the steady-state density and currents [20]

$$\hat{\mathcal{L}}_0 \rho = \sum_i \frac{\partial}{\partial x_i} e^{\phi_0} J_i^{(0)} \rho - \frac{1}{2} \sum_{ij} \frac{\partial}{\partial x_i} G_{ij} e^{-\phi_0} \frac{\partial}{\partial x_j} e^{\phi_0} \rho. \quad (32)$$

The dual dynamics is defined by

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^d \rho, \quad (33)$$

with

$$\hat{\mathcal{L}}^d = \rho_0 \hat{\mathcal{L}}_0^\dagger \rho_0^{-1}. \quad (34)$$

A straightforward calculation gives

$$\hat{\mathcal{L}}^d \rho = \sum_i \frac{\partial}{\partial x_i} F_i^d \rho - \sum_{ij} \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} G_{ij} \rho, \quad (35)$$

with

$$F_i^d = -F_i + \sum_j e^{\phi_0} \frac{\partial}{\partial x_j} G_{ij} e^{-\phi_0}. \quad (36)$$

It is a simple matter to verify that the dual dynamics has the same steady state as the original one, but the steady-state currents have opposite signs. It can be simulated by integrating the Ito stochastic equation;

$$\frac{dx_i}{dt} = F_i^d(\mathbf{x}) + \xi_i(t; \mathbf{x}). \quad (37)$$

The dual dynamics reverses the non conservative forces in the system. This relates the joint probability to go from one place to another in the original dynamics to the joint probability of the reversed sequence of events in the dual dynamics [20]

$$P_0(\mathbf{x}', t_1 | \mathbf{x}) \rho_0(\mathbf{x}) = P^d(\mathbf{x}, t_1 | \mathbf{x}') \rho_0(\mathbf{x}'). \quad (38)$$

The left-hand side of Eq. (38) is the joint steady-state probability to first the system at  $x$  and at  $x'$  after a time  $t_1$ . The right hand side is the joint probability of the reversed sequence of events but for a modified dynamics. When this modified dynamics is the dual these joint probabilities become equal.

We next turn to discuss the system's response to a series of impulsive perturbation. We assume that the perturbation is of the form

$$\hat{\mathcal{L}}_A \rho = \frac{1}{\gamma} \sum_i \frac{\partial}{\partial x_i} \frac{\partial A}{\partial x_i} \rho, \quad (39)$$

with  $A(x)$  as a potential field perturbing the system. Using Eq. (16) we have

$$S_1(\mathbf{x}_1; t_1) = \int d\mathbf{x}_0 P_0(\mathbf{x}_1, t_1 | \mathbf{x}_0) \frac{1}{\gamma} B(\mathbf{x}_0) \rho_0(\mathbf{x}_0), \quad (40)$$

where

$$B(\mathbf{x}) \equiv \sum_i \frac{\partial^2 A}{\partial x_i^2} - \frac{\partial A}{\partial x_i} \frac{\partial \phi_0}{\partial x_i}. \quad (41)$$

With the help of Eq. (38), we obtain

$$S_1(\mathbf{x}_1, t_1) = \int d\mathbf{x}_0 \frac{1}{\gamma} B(\mathbf{x}_0) P^d(\mathbf{x}_0, t_1 | \mathbf{x}_1) \rho_0(\mathbf{x}_1). \quad (42)$$

We now have all the tools needed to compare the ORF and KLRf for overdamped stochastic dynamics. The leading-order response function is given by

$$\mathcal{R}_1(t_1) = \frac{1}{\gamma} \int d\mathbf{x}_1 d\mathbf{x}_0 A(\mathbf{x}_1) P_0(\mathbf{x}_1, t_1 | \mathbf{x}_0) B(\mathbf{x}_0) \rho_0(\mathbf{x}_0), \quad (43)$$

while  $\mathcal{Q}^{(1)} = 0$ . At the next order, we have

$$\begin{aligned} \mathcal{Q}_{12}^{(2)}(t_2, t_1) &= \frac{1}{\gamma^2} \int d\mathbf{x}_2 d\mathbf{x}_1 d\mathbf{x}_0 B(\mathbf{x}_2) P^d(\mathbf{x}_2, t_2 | \mathbf{x}_1) \\ &\quad \times P_0(\mathbf{x}_1, t_1 + t_2 | \mathbf{x}_0) B(\mathbf{x}_0) \rho_0(\mathbf{x}_0) \end{aligned} \quad (44)$$

and

$$\begin{aligned} \mathcal{R}_2(t_2, t_1) &= \frac{1}{\gamma^2} \sum_i \int d\mathbf{x}_2 d\mathbf{x}_1 d\mathbf{x}_0 A(\mathbf{x}_2) P_0(\mathbf{x}_2, t_2 | \mathbf{x}_1) \\ &\quad \times \frac{\partial}{\partial x_{1i}} \left[ \frac{\partial A}{\partial x_{1i}} P_0(\mathbf{x}_1, t_1 | \mathbf{x}_0) \right] B(\mathbf{x}_0) \rho_0(\mathbf{x}_0). \end{aligned} \quad (45)$$

Some insight into the structure of different response functions can be gained by representing them as ensemble averages over stochastic trajectories. The first-order response function can be simulated directly using stochastic trajectories of the original dynamics. The appearance of a derivative of the conditional probability complicates the direct simulation of  $\mathcal{R}_2$ . It may be possible to circumvent this difficulty using the finite field method, where one combined simulations with and without a finite but small perturbation [21].  $\mathcal{Q}_{12}^{(2)}$  can be simulated with trajectories which follow the original dynamics for time  $t_1 + t_2$  and then the dual dynamics for time  $t_2$ .

Systems at equilibrium are self-dual, allowing to substitute  $\int d\mathbf{x}_1 P^d(\mathbf{x}_2, t_2 | \mathbf{x}_1) P_0(\mathbf{x}_1, t_1 + t_2 | \mathbf{x}_0) = P_0(\mathbf{x}_2, t_1 + 2t_2 | \mathbf{x}_0)$  in Eq. (44). As a result the Fisher information only depends on a single time variable  $t_1 + 2t_2$ . This is in contrast to systems which are perturbed from a nonequilibrium steady state, whose Fisher information is qualitatively different and is a two-dimensional function of  $t_1$  and  $t_2$ .

For the self-dual case  $\mathcal{Q}_{12}^{(2)}$  has the same structure as  $\mathcal{R}^{(1)}$ , up to a replacement of  $A(x)$  with  $B(x)$ . However, in the general, non-self-dual case, the structure of  $\mathcal{Q}_{12}^{(2)}$  is manifestly different than that of  $\mathcal{R}^{(1)}$  and  $\mathcal{R}^{(2)}$ .

## B. Eigenfunction expansion of the Fisher information

An alternative approach for the calculation of the Fisher information, as well as higher-order KLRf, uses eigenfunction expansions for the density  $\rho$ . It will be sufficient to examine a simple one dimensional model where  $\hat{\mathcal{L}}_0$  is a

Fokker-Planck operator and the perturbation is given by Eq. (39).

The propagation of the unperturbed system can be described in terms of the eigenvalues and eigenfunctions of  $\hat{\mathcal{L}}_0$ . The right eigenfunctions satisfy

$$\hat{\mathcal{L}}_0 \rho_n = \Lambda(n) \rho_n. \quad (46)$$

Similarly, the left eigenfunctions,  $q_n$ , satisfy  $q_n \mathcal{L}_0 = \Lambda(n) q_n$ . It is assumed that the right and left eigenfunctions constitute a biorthogonal system, that is,

$$\int dx q_n(x) \rho_m(x) = \delta_{nm}. \quad (47)$$

Any probability density can be expanded in terms of right eigenfunctions,

$$\rho(x) = \sum_n C_n \rho_n(x), \quad (48)$$

with

$$C_n = \int dx q_n(x) \rho(x). \quad (49)$$

We consider systems perturbed out of thermal equilibrium, with a probability density  $\rho_0(x)$ . In this case the left eigenvalues are simply related to the right eigenvalues. For variable  $x$  which is even with respect to time reversal, this relation takes the form [22]

$$\rho_n(x) = q_n(x) \rho_0(x). \quad (50)$$

As a result, one can use only the left eigenfunctions, which in this case satisfy the orthogonality condition

$$\int dx q_n(x) q_m(x) \rho_0(x) = \delta_{nm}. \quad (51)$$

We note that  $q_0(x)=1$  and  $\Lambda(0)=0$  correspond to the equilibrium distribution.

Our goal is to calculate the response of the system to a perturbation with a coordinate dependent operator  $A(x)$ . Several types of integrals appear repeatedly in the calculation, and it will be convenient to introduce an appropriate notation. One comes from the need to decompose the probability distribution into eigenstates after each interaction with a pulse

$$\begin{aligned} \mathcal{B}_{mn} &\equiv \int dx q_m(x) \frac{\partial}{\partial x} \left[ \frac{\partial A}{\partial x} q_n(x) \rho_0(x) \right] \\ &= - \int dx \frac{\partial q_m}{\partial x} \frac{\partial A}{\partial x} q_n(x) \rho_0(x), \end{aligned} \quad (52)$$

where integration by parts was used in the second equality. [We assume that  $\rho_0(x)$  falls off fast enough to eliminate boundary terms.] The calculation of the response also involves an evaluation of the average of an observable, which, in the current setting, leads to integrals of the form

$$\mathcal{C}_{An} \equiv \int dx A(x) q_n(x) \rho_0(x). \quad (53)$$

It is straightforward to calculate  $\mathcal{S}^{(1)}$  and  $\mathcal{S}^{(2)}$  with the help of Eq. (52). We find

$$\mathcal{S}_1(x; t_1) = \sum_n \mathcal{B}_{n0} e^{-\Lambda(n)t_1} q_n(x) \rho_0(x) \quad (54)$$

and

$$\mathcal{S}^{(2)}(x; t_2, t_1) = \sum_{mn} \mathcal{B}_{mn} \mathcal{B}_{n0} e^{-\Lambda(n)t_1 - \Lambda(m)t_2} q_m(x) \rho_0(x). \quad (55)$$

We now calculate the first few ORF and KLRF. The first-order response functions is

$$\mathcal{R}_1(t_1) = \sum_n \mathcal{C}_{An} \mathcal{B}_{n0} e^{-\Lambda(n)t_1}. \quad (56)$$

Similarly, the second-order response function is given by

$$\mathcal{R}_2(t_2, t_1) = \sum_{mn} \mathcal{C}_{Am} \mathcal{B}_{mn} \mathcal{B}_{n0} e^{-\Lambda(n)t_1 - \Lambda(m)t_2}. \quad (57)$$

$\mathcal{R}^{(2)}$  should be compared to the KLRF of the same order, namely, the Fisher information

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \sum_n \mathcal{B}_{n0}^2 e^{-\Lambda(n)[t_1+2t_2]}, \quad (58)$$

which is calculated with the help of Eqs. (22) and (51). Again, the Fisher information of systems perturbed out of equilibrium depends only on  $t_1+2t_2$ .

For a system initially in a steady state the Fisher information is

$$\begin{aligned} \mathcal{Q}_{12}^{(2)}(t_2, t_1) &= \sum_{nm} \mathcal{B}_{n0} \mathcal{B}_{m0} e^{-\Lambda(n)[t_1+t_2] - \Lambda(m)t_2} \\ &\quad \times \int dx \rho_0^{-1}(x) \rho_n(x) \rho_m(x). \end{aligned} \quad (59)$$

However, since the relation Eq. (50) does not hold in this case, the integral in Eq. (59) does not vanish for  $n \neq m$ , and the Fisher information becomes a two-dimensional function of  $t_1$  and  $t_2$ .

$\mathcal{Q}_{123}^{(3)}$  and  $\mathcal{R}^{(3)}$  are calculated in Appendix B 2, where it is shown that  $\mathcal{Q}_{123}^{(3)}$  is a three-dimensional function of all its time variables. This qualitatively different signature of systems initially at steady state vs equilibrium is unique to the Fisher information due to its quadratic dependence on  $\delta\rho$ . It does not apply to higher-order quantities, such as  $\mathcal{Q}^{(3)}$ .

The eigenfunctions for a simple example, of an harmonic oscillator with an exponential perturbation, are presented in Appendix C.

#### IV. MASTER EQUATIONS, DUAL DYNAMICS, AND THE FISHER INFORMATION FOR DISCRETE SYSTEMS

The description of the KLRF in terms of a combination of the regular stochastic dynamics and its dual holds also for

Markovian systems with a finite number of states. Below we derive a simple expression for the Fisher information of a stochastic jump process in terms of the dual dynamics of the original process.

**A. Dual dynamics of discrete Markovian systems**

Consider a system with a finite number of states, undergoing a Markovian stochastic jump process described by the master equation

$$\dot{\mathbf{P}} = \mathcal{R}\mathbf{P}. \tag{60}$$

$\mathbf{P}$  is a vector of probabilities to find the system in its states and  $\mathcal{R}$  is the transition rate matrix [3]. Its off diagonal elements are positive,  $\mathcal{R}_{ij} > 0$  for  $i \neq j$ , and express the rate of transitions from state  $j$  to  $i$ , given that the system is at  $j$ . The diagonal elements satisfy  $\sum_i \mathcal{R}_{ii} = 0$ . We further assume that there exist a unique steady state,  $\mathbf{P}^{(s)}$ , satisfying

$$\mathcal{R}\mathbf{P}^{(s)} = 0, \tag{61}$$

and that at this steady state there is a nonvanishing probability to find the system at each of the states  $i$ ,  $P_i^{(s)} \neq 0$ . The master equation is one of the simplest models for irreversible stochastic dynamics. Below we briefly describe some of its relevant properties, such as the backward equation and its dual dynamics.

One can define an evolution operator

$$\mathbf{P}(t) = \mathcal{U}_t \mathbf{P}(0), \tag{62}$$

which satisfy the equation of motion

$$\frac{\partial}{\partial t} \mathcal{U}_t = \mathcal{R}\mathcal{U}_t, \tag{63}$$

with the initial condition  $\mathcal{U}_0 = 1$ .

For this model a dynamical variable  $\mathbf{A}$  is a vector with a value corresponding to each state of the system. Its expectation value is given by

$$\langle \mathbf{A} \rangle_t \equiv \mathbf{A} \cdot \mathbf{P}(t) = \sum_i A_i P_i(t). \tag{64}$$

Instead of calculating this average by propagating  $\mathbf{P}$ , one can define a time-dependent dynamical variables,  $\mathbf{A}(t)$ , such that  $\langle \mathbf{A} \rangle_t \equiv \mathbf{A}(t) \cdot \mathbf{P}(0)$ . This is the analog of the Schrödinger and Heisenberg pictures in quantum mechanics. The equation of motion for  $\mathbf{A}(t)$  is easily shown to be

$$\frac{\partial}{\partial t} \mathbf{A}(t) = \mathcal{R}^\dagger \mathbf{A}(t). \tag{65}$$

This equation is known as the backward equation, which turns to be related to the dual dynamics. (The backward equation is also written as a function of an *initial* rather than a final time.)  $\mathcal{R}$  and  $\mathcal{R}^\dagger$  have the same eigenvalues. However, the roles of the right and left eigenvectors are interchanged and  $\mathbf{A}(t)$  decays to a uniform vector with at long times.

Let us now define a diagonal matrix  $\mathbf{\Pi}$ , using  $\Pi_{ii} = P_i^{(s)}$ . The dual evolution is then defined as

$$\mathcal{R}^d \equiv \mathbf{\Pi} \mathcal{R}^\dagger \mathbf{\Pi}^{-1}. \tag{66}$$

The fact that  $\mathbf{\Pi}$  is built from the steady state of  $\mathcal{R}$ ,  $\mathbf{P}^s$ , guarantees that  $\mathcal{R}^d$  is a physically reasonable rate matrix, that is, that it satisfies  $\sum_j \mathcal{R}_{ij}^d = 0$ . The dual dynamics describes a physically allowed process which has the same steady state as the original process it was derived from. However, at this steady state the dual currents have opposite signs compared to the steady state currents of the original dynamics. A process is self-dual, that is,  $\mathcal{R}^d = \mathcal{R}$ , if and only if it satisfies detailed balance. Self-duality is therefore related to being in thermal equilibrium.

**B. Perturbation theory and the Fisher information**

Consider a system subjected to several impulsive perturbations

$$\mathcal{R}(\tau) = \mathcal{R}_0 + \sum_\alpha s_\alpha \mathcal{R}_A \delta(\tau - \tau_\alpha), \tag{67}$$

where  $\mathcal{R}_A$  corresponds to some physical perturbation  $A$ . Here the free evolution is given by

$$\mathcal{U}_t^0 = e^{\mathcal{R}_0 t}. \tag{68}$$

Similarly the evolution during an impulsive perturbation can be described by

$$\mathbf{P}(\tau_\alpha + \epsilon) = e^{s_\alpha \mathcal{R}_A} \mathbf{P}(\tau_\alpha - \epsilon), \tag{69}$$

where  $\epsilon$  is arbitrarily small.

We expand the exponent in Eq. (69) and collect all terms of the same order in  $s_\alpha$ . The expression for the Fisher information, following Eq. (22), is

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \sum_i \frac{1}{P_i^s} (\mathcal{U}_{t_2}^0 \mathcal{R}_A \mathbf{P}^s)_i (\mathcal{U}_{t_1+t_2}^0 \mathcal{R}_A \mathbf{P}^s)_i. \tag{70}$$

This expression can be simplified by writing one of the propagators in terms of the dual process using the relation

$$e^{\mathcal{R}_0^\dagger t} \mathbf{\Pi}^{-1} = \mathbf{\Pi}^{-1} e^{\mathcal{R}_0 t}. \tag{71}$$

After some algebraic manipulations, we find

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \sum_i (\mathcal{R}_A^d e^{\mathcal{R}_A^d t_2} e^{\mathcal{R}_0(t_1+t_2)} \mathcal{R}_A \mathbf{P}^s)_i. \tag{72}$$

In Eq. (72) we have defined  $\mathcal{R}_A^d \equiv \mathbf{\Pi} \mathcal{R}_A^\dagger \mathbf{\Pi}^{-1}$ .  $\mathcal{R}_A^d$  is not the dual of  $\mathcal{R}_A$  since  $\mathbf{\Pi}$  is not composed of the eigenvector of  $\mathcal{R}_A$  which corresponds to a vanishing eigenvalues. ( $\mathbf{\Pi}$  is composed of the eigenvalue of  $\mathcal{R}_0$ .)

Equation (72) is the analog of Eq. (44) for discrete systems. It demonstrates that for this model the fisher information can be simulated using a combination of the ordinary process and its dual. However one must also include a (possibly artificial) dual perturbation,  $\mathcal{R}_A^d$ , which can nevertheless be computed from the known physical one. As before, Eq. (72) shows that the Fisher information, for self-dual systems, is a function of  $t_1 + 2t_2$ .

As a side remark, the model considered here has only even “degrees of freedom” under time reversal. More general models may also include odd degrees of freedom, such as

momenta. In that case one may speculate that there would be anti-self-dual systems whose dual turns out to be the time reversed dynamics. This would lead to a Fisher information depending on  $t_1$  alone, as is the case for unitary dynamics.

## V. DISCUSSION

In this work we have studied a system driven out of its steady state by a sequence of impulsive perturbations. We have defined a set of measures for the response of the system to the perturbation, the KLRF, which are given by the series expansion of the Kullback-Leibler distance between the perturbed and unperturbed probability distributions.

At each order the KLRF and ORF depend on the same time differences between the pulses. However, there are important differences stemming from the nonlinear dependence of the KLRF on  $\rho$ . The expression for the KLRF, for instance Eqs. (22) and (B2) reveal quantities which can be simulated using several trajectories which end at the same point. We have shown that a simpler, but equivalent description exists. It uses the dual dynamics which allows to “run some of the trajectories backward.” This description is especially appealing for the Fisher information. Instead of viewing the Fisher information as composed over sum of pair of trajectories joined at their end point one can view it as an average of contributions of a single forward-backward trajectory.

Another difference between the KLRF and the ORF has to do with the appearance of derivatives of conditional probabilities, which for *deterministic* systems would correspond to groups of very close trajectories. These first appear in  $\mathcal{S}^{(2)}$ , see for instance Eqs. (45) for  $\mathcal{R}^{(2)}$ . The nonlinear character of the KLRF means that such terms appear in comparatively higher order of the perturbation theory. For example,  $\mathcal{S}^{(2)}$  contributes to  $\mathcal{R}^{(2)}$  but not to the Fisher information. Instead it first contributes to  $\mathcal{Q}_{123}^{(3)}$ .

We have demonstrated that the Fisher information behaves in a qualitatively different way depending on whether the system is perturbed from equilibrium or from an out-of-equilibrium steady state. For the classically stochastic systems considered here we have seen that in the former case  $\mathcal{Q}_{12}^{(2)}(t_1, t_2) = f(t_1 + 2t_2)$ . That is, the Fisher information is a one-dimensional function of the two time delays. This qualitative difference results from the self-duality of equilibrium dynamics, which is another expression of the principle of detailed balance.  $\mathcal{Q}_{123}^{(3)}$  does not show a similar reduction in dimension [see Eq. (B6)]. This property is special to the Fisher information.

We have focused on the properties of the Fisher information for overdamped stochastic dynamics. Do other types of systems exhibit similar behavior? In Appendix A we consider deterministic Hamiltonian systems. In that case the unitary dynamics results in a Fisher information which depends only on the time  $t_1$ . It will be of interest to study how (not overdamped) stochastic systems bridge between the overdamped and unitary limits.

We showed that the KLRF can serve as a useful measure characterizing the system’s dynamics. They encode information which differs from the information encoded in the ORF. This is demonstrated by the ability of the Fisher information

to distinguish between systems perturbed out of equilibrium or out of a nonequilibrium steady state. We expect other useful properties of the KLRF to be revealed by further studies.

## ACKNOWLEDGMENTS

One of us (S.M.) wishes to thank Eran Mukamel for most useful discussions. The support of the National Science Foundation (Grant No. CHE-0745892) and the National Institutes of Health (Grant No. GM-59230) is gratefully acknowledged.

## APPENDIX A: UNITARY DYNAMICS

For completeness, in this appendix we discuss the application of nonlinear response theory to deterministic Hamiltonian systems. We use simple examples to clarify the relation between the response of a system and the Kullback-Leibler distance. The Hamiltonian of the system is assumed to be of the form  $\mathcal{H} = \mathcal{H}_0 - s(\tau)A$ , where  $A$  is the perturbation. We start with some general comments stemming from the fact that the dynamics in phase space is an incompressible flow.

### 1. General considerations

The state of a classical system is described by its phase-space density  $\rho$ . It is important to note that this is the full probability density, which includes information on all the degrees of freedom. Let us denote the propagator of the classical trajectories by  $h$  so that  $\mathbf{x}(\tau) = h_\tau(\mathbf{x}(0))$  where  $\mathbf{x}$  denotes a phase-space point (all coordinates and momenta). Similarly, the propagator for the probability distribution is denoted by  $\mathcal{U}(\tau)$ . Liouville theorem tells us that phase-space volumes do not change in time. As a result, there is a simple relation between the phase-space densities at different times,

$$\rho(\mathbf{x}, \tau) = \mathcal{U}(\tau)\rho(\mathbf{x}, 0) = \rho(h_{-\tau}(\mathbf{x}), 0) = \rho(\mathbf{x}(0), 0). \quad (\text{A1})$$

The density is simply transported with the dynamics in phase space. The density at  $h_\tau(\mathbf{x})$  at time  $\tau$  is equal to the density at  $\mathbf{x}$  at time zero.

This property of the phase-space dynamics have an interesting consequence. Any integral whose integrand depends locally on  $\rho$  alone is time independent since the values of the integrand are just transported around by the dynamics. An interesting example is the entropy function

$$\tilde{\mathcal{S}}(\tau) \equiv - \int d\mathbf{x} \rho(\mathbf{x}, \tau) \ln \rho(\mathbf{x}, \tau). \quad (\text{A2})$$

Let us change the integration variable to  $\mathbf{x}_0 = h_{-\tau}(\mathbf{x})$ , which is just the phase-space point which would flow to  $\mathbf{x}$  after a time  $\tau$ . Liouville theorem assures us that the Jacobian of the transformation is unity and therefore



$$\tilde{S}(\tau) = - \int d\mathbf{x}_0 \rho(\mathbf{x}_0, 0) \ln \rho(\mathbf{x}_0, 0) \equiv \tilde{S}_0, \quad (\text{A3})$$

where we have used Eq. (A1). It is clear that this entropy does not depend on time.

Equation (A3) is a result of the unitary evolution in phase space. It connects with all the intricate problems related to the emergence of macroscopic irreversibility out of microscopic reversible dynamics. This deep problem is beyond the scope of the current paper. Such problems are circumvented when one uses a reduced probability density, whose dynamics is irreversible to begin with.

Unitary dynamics lead to an interesting result for the reverse Kullback-Leibler distance

$$\mathcal{D}(\rho \parallel \rho_0) \equiv \text{Tr} \rho \ln \frac{\rho}{\rho_0} = \text{Tr} \rho \ln \rho - \text{Tr} \rho \ln \rho_0. \quad (\text{A4})$$

We have seen that the first term is a constant of the motion. Assuming a Hamiltonian system initially in equilibrium  $\rho_0 = e^{-\beta \mathcal{H}_0} / Z$ . We have

$$\mathcal{D}(\rho \parallel \rho_0) = -\tilde{S}_0 + \text{Tr} \rho \beta (\mathcal{H}_0 - \mathcal{F}_0) = \beta \text{Tr} (\rho - \rho_0) \mathcal{H}_0, \quad (\text{A5})$$

where  $\mathcal{F}_0 = -k_B T \ln Z$  is the free energy of the initial equilibrium state.  $\mathcal{D}(\rho \parallel \rho_0)$  is therefore linear in  $\rho$ , and thus it is equivalent to a calculation of response functions.

The general comments above raise two interesting points. First, the fact that terms such as  $\text{Tr} \rho \ln \rho$  are constant means that their expansion in powers of pulse strengths  $s_i$  turns out to have only the constant term, all other terms in the expansion must vanish.

The second point of interest has to do with the relation between the Kullback-Leibler distance and its reverse. Generally,  $\mathcal{D}(\rho \parallel \rho_0) \neq \mathcal{D}(\rho_0 \parallel \rho)$ . We are interested in systems perturbed by a series of pulses and compare the initial and final distributions. In that case we can use unitarity to change variables from the phase-space points at the final time to the points at the initial time which are connected to it by the dynamics. This gives

$$\begin{aligned} \mathcal{D}(\rho \parallel \rho_0) &= \int d\mathbf{x} \rho(\mathbf{x}, \tau) \ln \frac{\rho(\mathbf{x}, \tau)}{\rho_0(\mathbf{x})} = \int d\mathbf{x}_0 \rho_0(\mathbf{x}_0) \ln \frac{\rho_0(\mathbf{x}_0)}{\tilde{\rho}(\mathbf{x}_0, 0)} \\ &= \mathcal{D}(\rho_0 \parallel \tilde{\rho}). \end{aligned} \quad (\text{A6})$$

Here  $\rho_0(\mathbf{x}(\tau)) = \tilde{\rho}(\mathbf{x}_0, 0)$ , that is,  $\tilde{\rho}$  is the density that would evolve to the equilibrium density under the influence of the pulses. It is not equal to  $\rho(\mathbf{x}, \tau)$ , which tells us that the distance and its reverse are not equal.

While in general the distance and its reverse are not equal, when the distance between the distributions is small, it is easy to show that their leading order expansion in  $\delta\rho = \rho - \rho_0$  is the same. Loosely speaking

$$\mathcal{D}(\rho_0 \parallel \rho) = \frac{1}{2} \int d\mathbf{x} \left( \frac{\delta\rho^2}{\rho_0} + O(\delta\rho^3) \right), \quad (\text{A7})$$

$$\mathcal{D}(\rho \parallel \rho_0) = \frac{1}{2} \int d\mathbf{x} \left( \frac{\delta\rho^2}{\rho_0} + O(\delta\rho^3) \right). \quad (\text{A8})$$

We can deduce that the Fisher information, which is determined by this leading order, could be obtained from a calculation of response for systems with unitary evolution. (Since it could also be obtained from the reverse distance.)

Our last general point is also related to unitarity, but has to do with the Fisher information. According to Eq. (22) the Fisher information is built from two partial densities. (More accurately, density differences.) These evolve with respect to the same Hamiltonian  $\mathcal{H}_0$  between interactions with the pulses. As a result, the integral in Eq. (22) do not depend on  $t_2$ , for the same reason that caused the entropy  $\tilde{S}$  to be time independent. We find that

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \mathcal{Q}_{12}^{(2)}(t_1), \quad (\text{A9})$$

as a result of the unitarity of the phase-space dynamics. Similar independence on the final time interval would also appear for higher-order terms in the expansion of the Kullback-Leibler distance.

## 2. Single Harmonic oscillator

The simplest system that can serve as an example is a single Harmonic oscillator

$$\mathcal{H}_0 = \frac{P^2}{2M} + \frac{1}{2} M \Omega^2 Q^2. \quad (\text{A10})$$

We take the perturbation to be

$$A(Q) = \alpha e^{-Q/Q_0}. \quad (\text{A11})$$

For this system it is trivial to solve for the free evolution

$$Q(\tau) = Q(0) \cos \Omega \tau + \frac{P(0)}{M\Omega} \sin \Omega \tau,$$

$$P(\tau) = P(0) \cos \Omega \tau - \Omega M Q(0) \sin \Omega \tau. \quad (\text{A12})$$

This relation can be inverted, expressing  $P(0), Q(0)$  in terms of  $P(\tau), Q(\tau)$

$$Q(0) = Q(\tau) \cos \Omega \tau - \frac{P(\tau)}{M\Omega} \sin \Omega \tau,$$

$$P(0) = \Omega M Q(\tau) \sin \Omega \tau + P(\tau) \cos \Omega \tau. \quad (\text{A13})$$

Equation (A13) will be useful when one propagates probability distributions in time.

The equilibrium distribution of the Harmonic oscillator is a Gaussian

$$\rho_0 = \frac{2\pi k_B T}{\Omega} e^{-\beta P^2/2M - \beta M \Omega^2 Q^2/2}. \quad (\text{A14})$$

We also note that

$$\mathcal{L}_A \rho_0 = \{A, \rho_0\} = \frac{\beta \alpha}{M Q_0} P e^{-Q/Q_0} \rho_0, \quad (\text{A15})$$

is the linear order correction for the density just after interaction with one pulse.

To calculate  $\mathcal{S}^{(1)}$  we need to propagate Eq. (A15). With the help of Eqs. (A1) and (A13) we find

$$\begin{aligned} \mathcal{S}^{(1)}(Q, P; t_1) &= \frac{\beta \alpha}{M Q_0} (\Omega M Q \sin \Omega t_1 + P \cos \Omega t_1) \\ &\times e^{-1/Q_0 [Q \cos \Omega t_1 - P/M \Omega \sin \Omega t_1]} \rho_0(Q, P). \end{aligned} \quad (\text{A16})$$

In the derivation we have used the fact that  $\rho_0$  is invariant under the evolution with respect to  $\mathcal{H}_0$ .

To calculate  $\mathcal{S}^{(2)}$  we operate with  $\mathcal{L}_A$  on  $\mathcal{S}^{(1)}$  and then propagate the resulting correction for the density for a time interval  $t_2$ . A straightforward calculation leads to

$$\begin{aligned} \mathcal{S}^{(2)}(Q, P; t_2, t_1) &= \mathcal{U}(t_2) \mathcal{L}_A \mathcal{S}^{(1)} \\ &= -\frac{\beta \alpha^2}{M Q_0^2} \left\{ \cos \Omega t_1 + [\Omega M Q \sin \Omega(t_1 + t_2) + P \cos \Omega(t_1 + t_2)] \right. \\ &\quad \times \left[ -\frac{\beta}{M} (\Omega M Q \sin \Omega t_2 + P \cos \Omega t_2) + \frac{1}{Q_0 M \Omega} \sin \Omega t_1 \right] \left. \right\} \\ &\quad \times \exp \left\{ -\frac{Q}{Q_0} [\cos \Omega t_2 + \cos \Omega(t_1 + t_2)] + \frac{P}{M \Omega Q_0} [\sin \Omega t_2 + \sin \Omega(t_1 + t_2)] \right\} \rho_0. \end{aligned} \quad (\text{A17})$$

We now turn to calculate the first two ORF, using Eqs. (21) and (23). The calculation of  $\mathcal{R}^{(1)}$  is cumbersome but straightforward.

$$\mathcal{R}^{(1)}(t_1) = \int dQ dP A(Q) \mathcal{S}^{(1)}(Q, P; t_1) = -\frac{\alpha^2}{\Omega M Q_0^2} \sin \Omega t_1 \exp \left[ \frac{1}{\beta M \Omega^2 Q_0^2} (1 + \cos \Omega t_1) \right]. \quad (\text{A18})$$

The calculation of  $\mathcal{R}^{(2)}$  is more involved, and we only include the final result

$$\mathcal{R}^{(2)}(t_2, t_1) = \frac{\alpha^3}{M^2 \Omega^2 Q_0^4} \sin \Omega t_2 [\sin \Omega t_1 + \sin \Omega(t_1 + t_2)] \exp \left[ \frac{1}{M \beta \Omega^2 Q_0^2} \left\{ \frac{3}{2} + \cos \Omega t_1 + \cos \Omega t_2 + \cos \Omega(t_1 + t_2) \right\} \right]. \quad (\text{A19})$$

We would like to compare these response functions to the KLRF and in particular to the Fisher information, which can be calculated using Eq. (22). The calculation can be simplified by using the classical coordinates right after the second interaction with the pulse as integration variables. Due to unitarity, the Fisher information only depends on the time difference  $t_1$ , see also the discussion in the previous subsection.

We get

$$\begin{aligned} \mathcal{Q}_{12}^{(2)}(t_2, t_1) &= \frac{\beta \alpha^2}{M Q_0^2} \left[ \cos \Omega t_1 - \frac{1}{\beta M \Omega^2 Q_0^2} \sin^2 \Omega t_1 \right] \\ &\quad \times \exp \left[ \frac{1}{\beta M \Omega^2 Q_0^2} (1 + \cos \Omega t_1) \right]. \end{aligned} \quad (\text{A20})$$

This expression seems similar to the first-order response function. One can indeed show that they are related by

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) = \beta \frac{\partial}{\partial t_1} \mathcal{R}^{(1)}(t_1). \quad (\text{A21})$$

It will be interesting to check whether this expression could be generalized to any Hamiltonian system (with unitary dynamics).

## APPENDIX B: THIRD ORDER ORF AND KLRF

In the main text we have calculated the ORF and KLRF to second order. In the appendix we present the third-order quantities,  $\mathcal{R}^{(3)}$  and  $\mathcal{Q}_{123}^{(3)}$ .

### 1. Perturbation theory

Following the calculations performed in Sec. II, the third-order response functions are given by

$$\mathcal{R}^{(3)}(t_3, t_2, t_1) = \int dx A(x) \mathcal{S}^{(3)}(x; t_3, t_2, t_1) \quad (\text{B1})$$

and

$$\begin{aligned} \mathcal{Q}_{123}^{(3)}(t_3, t_2, t_1) = & \int dx \{ \rho_0^{-1}(x) [\mathcal{S}^{(1)}(x; t_1 + t_2 + t_3) \mathcal{S}^{(2)}(x; t_3, t_2) \\ & + \mathcal{S}^{(1)}(x; t_2 + t_3) \mathcal{S}^{(2)}(x; t_3, t_1 + t_2) \\ & + \mathcal{S}^{(1)}(x; t_3) \mathcal{S}^{(2)}(x; t_3 + t_2, t_1)] \\ & - 2\rho_0^{-2}(x) \mathcal{S}^{(1)}(x; t_1 + t_2 + t_3) \\ & \times \mathcal{S}^{(1)}(x; t_2 + t_3) \mathcal{S}^{(1)}(x; t_3) \}, \end{aligned} \quad (\text{B2})$$

where

$$\begin{aligned} \mathcal{S}^{(3)}(x; t_3, t_2, t_1) \equiv & \int dx_1 \cdots dx_6 \mathcal{U}_0(x, x_6; t_3) \mathcal{L}_A(x_6, x_5) \\ & \times \mathcal{U}_0(x_5, x_4; t_2) \mathcal{L}_A(x_4, x_3) \\ & \times \mathcal{U}_0(x_3, x_2; t_1) \mathcal{L}_A(x_2, x_1) \rho_0(x_1). \end{aligned} \quad (\text{B3})$$

## 2. Eigenvalue expansions

Here we present expressions for the third order ORF and KLRF in term of the eigenvalues and eigenfunctions of the stochastic dynamics. At high orders the nonlinear character of the contributions to the Kullback-Leibler distance result in integrals with products of several eigenfunctions. Here we will only need the one with three eigenfunctions

$$\mathcal{J}_{nml} \equiv \int dx q_n(x) q_m(x) q_l(x) \rho_0(x). \quad (\text{B4})$$

The third order ORF is given by

$$\mathcal{R}^{(3)}(t_3, t_2, t_1) = \sum_{lmn} \mathcal{C}_{Al} \mathcal{B}_{lm} \mathcal{B}_{mn} \mathcal{B}_{n0} e^{-\Lambda(n)t_1 - \Lambda(m)t_2 - \Lambda(l)t_3}. \quad (\text{B5})$$

This ORF should be compared to the third order KLRF, which is calculated using Eq. (B2). We find

$$\begin{aligned} \mathcal{Q}_{123}^{(3)}(t_3, t_2, t_1) = & \sum_{mn} \mathcal{B}_{m0} \mathcal{B}_{mn} \mathcal{B}_{n0} \{ e^{-\Lambda(n)t_2} e^{-\Lambda(m)[t_1+t_2+2t_3]} \\ & + e^{-\Lambda(n)[t_1+t_2]} e^{-\Lambda(m)[t_2+2t_3]} \\ & + e^{-\Lambda(n)t_1} e^{-\Lambda(m)[t_2+2t_3]} \} \\ & - 2 \sum_{nml} \mathcal{B}_{n0} \mathcal{B}_{m0} \mathcal{B}_{l0} \mathcal{J}_{nml} e^{-\Lambda(n)[t_1+t_2+t_3]} \\ & \times e^{-\Lambda(m)[t_2+t_3]} e^{-\Lambda(l)t_3}. \end{aligned} \quad (\text{B6})$$

The expression for  $\mathcal{Q}_{123}^{(3)}$ , presented in Eq. (B6), has three terms in which the orthogonality condition (51) has been used, pointing to a reduction in dimension in the time dependence of this specific term. However, the time combinations in these terms are all different. In addition, the fourth term in Eq. (B6) clearly depends on all its time variables. We conclude that in contrast to the Fisher information, the higher-

order KLRF depend on all their time variables. The reduction in dimension is therefore specific for the Fisher information. It results from the fact that it is built out of a single product of two density corrections.

## APPENDIX C: THE OVERDAMPED HARMONIC OSCILLATOR

In this appendix we consider a simple example of an overdamped harmonic oscillator with a potential  $1/2M\Omega Q^2$ , with a perturbing potential  $A(Q) = \alpha e^{-Q/Q_0}$ . For this system it is possible to write explicit expressions for the eigenvalues and eigenfunctions of the Fokker-Planck operator, as well as to perform several of the integral, defined in Sec. III. In this case

$$\hat{\mathcal{L}}_0 \rho = -\frac{1}{\gamma} \frac{\partial}{\partial Q} \left[ M\Omega^2 Q \rho + k_B T \frac{\partial}{\partial Q} \rho \right], \quad (\text{C1})$$

$$\hat{\mathcal{L}}_A \rho = -\frac{\alpha}{Q_0 \gamma} \frac{\partial}{\partial Q} [e^{-Q/Q_0} \rho]. \quad (\text{C2})$$

This model has been studied extensively [22]. The equilibrium density is given by

$$\rho_0(Q) = \sqrt{\frac{M\Omega^2}{2\pi k_B T}} e^{-M\Omega^2/2k_B T Q^2}. \quad (\text{C3})$$

The left eigenfunctions and the eigenvalues of this model are

$$q_n(Q) = \sqrt{\frac{1}{2^n n!}} H_n \left( Q \sqrt{\frac{M\Omega^2}{2k_B T}} \right), \quad (\text{C4})$$

$$\Lambda(n) = \frac{M\Omega^2}{\gamma} n. \quad (\text{C5})$$

Here,  $H_n$  are the Hermit polynomials.

The following properties of the Hermit polynomials

$$\frac{dH_n}{dx} = 2n H_{n-1}(x), \quad (\text{C6})$$

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}, \quad (\text{C7})$$

together with the exponential form of  $A(Q)$ , allow to compute some of the integrals defined in Eqs. (52) and (B4) explicitly.

For instance,

$$\begin{aligned} \mathcal{C}_{An} = & \alpha \int dQ e^{-Q/Q_0} q_n(Q) \rho_0(Q) \\ = & \alpha \sqrt{\frac{1}{2^n n!}} \int dy e^{-(y/Q_0) \sqrt{2k_B T/M\Omega^2}} H_n(y) e^{-y^2}. \end{aligned} \quad (\text{C8})$$

One can substitute Eq. (C7) for the Hermit polynomial and use integration by parts. This leads to

$$C_{An} = \alpha(-1)^n \sqrt{\frac{1}{2^n n!}} Q_0^{-n} \left(\frac{2k_B T}{M\Omega^2}\right)^{n/2} \int dy e^{-y^2} \times e^{-(y/Q_0)\sqrt{2k_B T/M\Omega^2}}. \quad (\text{C9})$$

This is a Gaussian integral which is easily evaluated to give

$$C_{An} = (-1)^n \alpha \sqrt{\frac{\pi}{2^n n!}} Q_0^{-n} \left(\frac{2k_B T}{M\Omega^2}\right)^{n/2} e^{k_B T/2Q_0^2 M\Omega^2}. \quad (\text{C10})$$

One can also easily calculate the integrals

$$\begin{aligned} \mathcal{B}_{n0} &= - \int dQ \frac{\partial q_n}{\partial Q} \frac{\partial A}{\partial Q} \rho_0(Q) \\ &= \frac{2n}{\sqrt{2^n n!}} \frac{\alpha}{\gamma Q_0} \sqrt{\frac{M\Omega^2}{2\pi k_B T}} \int dy H_{n-1}(y) e^{-y^2} e^{-(y/Q_0)\sqrt{2k_B T/M\Omega^2}} \end{aligned} \quad (\text{C11})$$

by the same technique. One finds

$$\mathcal{B}_{n0} = (-1)^{n-1} \frac{2n}{\sqrt{2^n n!}} \frac{\alpha}{\gamma Q_0^n} \left(\frac{2k_B T}{M\Omega^2}\right)^{n/2-1} e^{k_B T/2Q_0^2 M\Omega^2}. \quad (\text{C12})$$

Comparing Eqs. (C10) and (C12), we see that for this model

$$\mathcal{B}_{n0} = -n \frac{M\Omega^2}{\gamma k_B T \sqrt{\pi}} C_{An}. \quad (\text{C13})$$

As a result, there is a simple relation between the Fisher information and the first-order response function,

$$\mathcal{Q}_{12}^{(2)}(t_2, t_1) \propto \left. \frac{\partial \mathcal{R}_1}{\partial t_1} \right|_{t_1 \rightarrow t_1 + 2t_2}. \quad (\text{C14})$$

This relation is a special result for this model, and is not expected to hold for other systems.

One can also obtain explicit results for  $\mathcal{B}_{nm}$  with nonvanishing indices. However, the calculation and the result are quite cumbersome and are omitted. We were not able to calculate  $\mathcal{J}_{nml}$  explicitly.

- 
- [1] H. B. Callen and T. A. Welton, *Phys. Rev.* **83**, 34 (1951).  
[2] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Mechanics II: Nonequilibrium Statistical Mechanics* (Springer, New York, 1998).  
[3] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 2007).  
[4] R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon Press, Oxford, 1998).  
[5] S. Mukamel, D. Abramavicius, L. Yang, W. Zhuang, I. V. Schweigert, and D. V. Voronine, *Acc. Chem. Res.* **42**, 553 (2009).  
[6] D. Abramavicius, B. Palmieri, D. Voronine, F. Sanda, and S. Mukamel, *Chem. Rev.* **109**, 2350 (2009).  
[7] W. K. Wootters, *Phys. Rev. Lett.* **80**, 2245 (1998).  
[8] T. Yu and J. H. Eberly, *Science* **323**, 598 (2009).  
[9] J. L. Lebowitz and H. Spohn, *J. Stat. Phys.* **95**, 333 (1999).  
[10] C. Maes and K. Netočný, *J. Stat. Phys.* **110**, 269 (2003).  
[11] A. Gomez-Marín, J. M. R. Parrondo, and C. Van den Broeck, *Phys. Rev. E* **78**, 011107 (2008).  
[12] D. Andrieux *et al.*, *J. Stat. Mech.: Theory Exp.* (2008), P01002.  
[13] G. E. Crooks, *J. Stat. Phys.* **90**, 1481 (1998).  
[14] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, *Phys. Rev. Lett.* **98**, 080602 (2007).  
[15] S. Vaikuntanathan and C. Jarzynski, *EPL* **87**, 60005 (2009).  
[16] I. Goychuk and P. Hänggi, *Eur. Phys. J. B* **69**, 29 (2009).  
[17] A. Neiman, B. Shulgin, V. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund, *Phys. Rev. Lett.* **76**, 4299 (1996).  
[18] S. Ganguli, D. Huh, and H. Sompolsky, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18970 (2008).  
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley & Sons, Hoboken, 2006).  
[20] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, *J. Stat. Mech.: Theory Exp.* (2006), P08001.  
[21] S. Mukamel, V. Khidekel, and V. Chernyak, *Phys. Rev. E* **53**, R1 (1996).  
[22] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).