Bounds on the rates of statistical divergences and mutual information via stochastic thermodynamics

Jan Karbowski [*]
Institute of Applied Mathematics and Mechanics, Department of Mathematics, Informatics, and Mechanics, University of Warsaw, ul. Banacha 2, 02-097 Warsaw, Poland

(Received 9 August 2023; revised 14 April 2024; accepted 18 April 2024; published 24 May 2024)

Statistical divergences are important tools in data analysis, information theory, and statistical physics, and there exist well-known inequalities on their bounds. However, in many circumstances involving temporal evolution, one needs limitations on the rates of such quantities instead. Here, several general upper bounds on the rates of some f-divergences are derived, valid for any type of stochastic dynamics (both Markovian and non-Markovian), in terms of information-like and/or thermodynamic observables. As special cases, the analytical bounds on the rate of mutual information are obtained. The major role in all those limitations is played by temporal Fisher information, characterizing the speed of global system dynamics, and some of them contain entropy production, suggesting a link with stochastic thermodynamics. Indeed, the derived inequalities can be used for estimation of minimal dissipation and global speed in thermodynamic stochastic systems. Specific applications of these inequalities in physics and neuroscience are given, which include the bounds on the rates of free energy and work in nonequilibrium systems, limits on the speed of information gain in learning synapses, as well as the bounds on the speed of predictive inference and learning rate. Overall, the derived bounds can be applied to any complex network of interacting elements, where predictability and thermodynamics of network dynamics are of prime concern.

DOI: 10.1103/PhysRevE.109.054126

I. INTRODUCTION

Statistical divergences, or distances, known as f-divergences, are commonly used to quantify the difference between two probability distributions [1,2]. The most popular special cases of these divergences are the Renyi divergence [1], the Tsallis divergence [3], and the Kullback-Leibler (KL) divergence [4], which is a limiting case of the former two (for a review, see [5]). In statistical physics, KL and Tsallis divergences have prominent roles and have been shown to relate to information gain and other important physical quantities, such as entropy production, work, and other observables [6–9]. In computer science, and recently in machine learning, KL has been used, among other things, in assessing coding accuracy and efficiency [10,11]. Moreover, f-divergences have many applications in classic information theory [12], and in the emerging field of information geometry [13]. There are many inequalities relating different types of divergences and inequalities bounding them from above [2,14]. However, virtually all of these relations and bounds apply only to static (stationary) situations. Since physical quantities generally depend on time, probability distributions describing them are often time-dependent. Consequently, in real physical systems, statistical divergences can also change in time, and their variability can provide important information about the predictability of a probabilistic system's dynamics. It is known that for isolated stochastic systems with Markov dynamics (either of master equation or Fokker-Planck equation types), all f-divergences decrease monotonically with time between a time-dependent state probability distribution and its equilibrium distribution [1,15–19], which can be interpreted as a loss of information in autonomously relaxing systems [10]. However, no such simple relation exists for statistical divergences between two arbitrary time-dependent distributions.

The goal of this paper is to shed some light on more general conditions of this type by determining the fundamental bounds on the rates of popular f-divergences for arbitrary probability distributions, and to relate these bounds to known observables. The obtained bounds may have practical applications, as it is often difficult to calculate exactly the rates of statistical divergences for a system at hand. Moreover, and more importantly, such limits may have a conceptual meaning, especially with regard to stochastic and information thermodynamics (e.g., see [8,9,20–22]), when we interpret stochastic f-divergences as generalized information gains [10]. Indeed, the inequalities found here for the rates of statistical divergences have a similar flavor to several inequalities discovered recently in stochastic thermodynamics linking physical observables with information, entropy production, and the speed of global dynamics [23–27].

In this work, two types of bounds on the rates of Tsallis, Renyi, and Kullback-Leibler divergences are derived. The first type is very general and consists of two inequalities related solely to kinematic characteristics. The second type is more restrictive, as it applies only to Markov dynamics for probabilities obeying a master equation, and it consists of four inequalities involving both kinematic and thermodynamic observables. These results are then used to derive

[*]jkarbowski@mimuw.edu.pl

general bounds on the rate of mutual information between two stochastic variables. As an example, a driven one-step Markov process is used to illustrate the sharpness and ranking of the obtained bounds. These bounds can be used naturally for determining lower speed limits on stochastic dynamics and minimal dissipation. Various other, more specific, applications are also presented, ranging from physics to neuroscience. These applications involve the limits on the rates of free energy in nonequilibrium thermodynamic systems, as well as the bounds on the speed of information gain, predictive inference, and learning rate in neural systems.

II. PRELIMINARIES

A. Statistical divergences and useful relations

Consider a physical system that has internal states labeled by index n, and which can be described by two time-dependent probability distributions $p_n(t)$ and $q_n(t)$. Although it is not essential for the arguments below, it is convenient to think about $q_n(t)$ as a true (reference) probability distribution of the system stochastic dynamics, and about $p_n(t)$ as its estimation or prediction. Before we introduce f-divergences, let us first define a helpful quantity, which can be called the α-coefficient $C_\alpha(p||q)$ between the distributions p and q (also known as the Chernoff α-coefficient or divergence [28]),

$$C_\alpha(p||q) = \left\langle \left(\frac{p}{q}\right)^{\alpha-1} \right\rangle_p, \tag{1}$$

where α is an arbitrary real number, and the symbol $\langle (p/q)^{\alpha-1}\rangle_p = \sum_n p_n(t)(\frac{p_n(t)}{q_n(t)})^{\alpha-1}$, which means averaging with respect to probability distribution p. Note that equivalently $C_\alpha = \langle (p/q)^\alpha\rangle_q$, which implies $C_\alpha(p||p) = 1$ for all α, and also $C_0(p||q) = C_1(p||q) = 1$. (The focus is on discrete states, but the results are also valid for continuous variables through replacing sums by integrals, and such transformations are done below occasionally.) The α-coefficient provides a core for basic f-divergences, and thus it can be of interest in itself.

Two major f-divergences, Tsallis T_α and Renyi R_α, between p and q distributions are expressed in terms of the α-coefficient as [5]

$$T_\alpha(p||q) = \frac{C_\alpha(p||q) - 1}{\alpha - 1} \tag{2}$$

and

$$R_\alpha(p||q) = \frac{\ln C_\alpha(p||q)}{\alpha - 1}, \tag{3}$$

which implies a simple relationship between them as $R_\alpha = \ln(1 + (\alpha-1)T_\alpha)/(\alpha-1)$. When $T_\alpha(p||q)$ and $R_\alpha(p||q)$ are close to 0, then the probability distribution p approximates or predicts the distribution q very well.

In the limit $\alpha \mapsto 1$, Tsallis and Renyi divergences both tend to KL divergence D_{KL}, known also as relative entropy [10], i.e., $T_1 = R_1 = D_{\mathrm{KL}}(p||q) = \sum_n p_n(t) \ln \frac{p_n(t)}{q_n(t)}$. For $\alpha = 1/2$, we obtain $T_{1/2}$ as a Hellinger distance, while for $\alpha = 2$ our T_2 is the Pearson χ^2-divergence.

In our derivations, we will also need two inequalities. The first is the stochastic version of the generalized Hölder inequality, which for m arbitrary stochastic variables $X_1, X_2, \ldots, X_m$ takes the form [29]

$$\langle |X_1 \cdots X_m|\rangle \leqslant \langle |X_1|^{1/\lambda_1}\rangle^{\lambda_1} \cdots \langle |X_m|^{1/\lambda_m}\rangle^{\lambda_m}, \tag{4}$$

where λ_i are positive real numbers such that $\sum_{i=1}^m \lambda_i = 1$, the symbol $|\cdots|$ means the absolute value, and $\langle\cdots\rangle$ denotes averaging with respect to some probability distribution. The equality in Eq. (4) is achieved when there are proportionalities between all the rescaled variables, i.e., when $|X_i|^{1/\lambda_i} = c_i|X_1|^{1/\lambda_1}$ for every $i = 2, \ldots, m$, where c_i are some positive (possibly time-dependent) coefficients. When $m = 2$ and $\lambda_1 = \lambda_2 = 1/2$, Eq. (4) becomes a classic Cauchy-Schwarz inequality.

The second useful inequality relates arithmetic and geometric means to the so-called logarithmic mean of two positive numbers x and y, and it reads [30]

$$\sqrt{xy} \leqslant \frac{x - y}{\ln(x) - \ln(y)} \leqslant \frac{x + y}{2}. \tag{5}$$

With Eqs. (1)–(5), we have all the necessary ingredients to derive the upper bounds on the rates of f-divergences.

B. Rates of statistical divergences.

Because of the relations (1)–(3), the rates of Tsallis and Renyi divergences can be expressed in terms of the rate of the α-coefficient. For that reason, and because calculations are a little easier for C_α, below we focus on the temporal rate of C_α and its bounds. The bounds for the rates of T_α and R_α are obtained as straightforward extensions of the bounds on dC_α/dt.

The temporal rate of $C_\alpha(p||q)$ can be written as

$$\begin{aligned}\frac{dC_\alpha}{dt} &= \alpha \sum_n \dot{p}_n[(p_n/q_n)^{\alpha-1} - C_\alpha] \\ &\quad - (\alpha - 1)\sum_n \dot{q}_n[(p_n/q_n)^\alpha - C_\alpha] \\ &\leqslant \left|\frac{dC_\alpha}{dt}\right| \leqslant |\alpha| \left\langle \left|\frac{\dot{p}}{p}[(p/q)^{\alpha-1} - C_\alpha]\right|\right\rangle_p \\ &\quad + |\alpha - 1| \left\langle \left|\frac{\dot{q}}{q}[(p/q)^\alpha - C_\alpha]\right|\right\rangle_q\end{aligned} \tag{6}$$

where $\dot{p}_n = dp_n/dt$, and similarly for $\dot{q}_n$. The notation $\langle\cdots\rangle_p$ means averaging with respect to distribution p_n. In both bracket terms, we subtracted C_α for convenience (see below), but this trick does not change the result of summation, as $\sum_n \dot{p}_n = \sum_n \dot{q}_n = 0$. In the last inequality we used a well-known relation $x + y \leqslant |x + y| \leqslant |x| + |y|$ for arbitrary real numbers x, y.

Equation (6) enables us to write the upper limit on the rate of the Kullback-Leibler divergence $D_{\mathrm{KL}}(p||q)$, since $D_{\mathrm{KL}} = \lim_{\alpha\mapsto 1}(C_\alpha - 1)/(\alpha - 1)$ and $D_{\mathrm{KL}} = T_1$. We have

$$\begin{aligned}\left|\frac{dD_{\mathrm{KL}}}{dt}\right| &\leqslant \left\langle \left|\frac{\dot{p}}{p}\right| \left|\ln\left(\frac{p}{q}\right) - D_{\mathrm{KL}}\right|\right\rangle_p + \left\langle \left|\frac{\dot{q}}{q}\right| \left|\left(\frac{p}{q}\right) - 1\right|\right\rangle_q, \\ dD_{KL}/dt &\leqslant \langle |\dot{p}/p| |\ln(p/q) - D_{\mathrm{KL}}|\rangle_p + \left\langle \left|\frac{d\ln(p/q)}{dt}\right|\right\rangle_p,\end{aligned} \tag{7}$$

where we used the fact that $\lim_{\alpha\mapsto 1}[(p/q)^{\alpha-1}-1]/(\alpha-1)=\ln(p/q)$.

Our goal in the next sections is to find upper bounds on dC_α/dt and $dD_{\rm KL}/dt$, which in effect is equivalent to determining the limits on the averages in Eqs. (6) and (7). Having the bounds on dC_α/dt, it is easy to obtain the upper limits on the rates of Tsallis and Renyi divergences, since from Eqs. (2) and (3) it follows that $dC_\alpha/dt=(\alpha-1)dT_\alpha/dt$ and $dC_\alpha/dt=(\alpha-1)e^{(\alpha-1)R_\alpha}dR_\alpha/dt$.

C. Temporal and relative Fisher information

In the derivation below, we will use the so-called temporal Fisher information, which is defined for probability distribution p_n as [23,31]

$$I_F(p)=\sum_n p_n\left(\frac{\dot p_n}{p_n}\right)^2\equiv\langle(\dot p/p)^2\rangle_p, \tag{8}$$

and analogically for the distribution q_n. Note that the role of the control/external parameter is played here by the time. The quantity $I_F(p)$ is usually interpreted as a square of the speed of global system dynamics described by the distribution p_n [23,31]. For example, if p_n is a Poisson distribution with time-dependent intensity parameter ν, i.e., $p_n=(\nu^n/n!)e^{-\nu}$, then the temporal Fisher information is $I_F(p)=(\dot\nu)^2/\nu$, where $\dot\nu$ is the temporal derivative of ν. This result indicates that $\sqrt{I_F(p)}$ is proportional to the absolute speed of changes in the intensity parameter.

By a direct extension, we can define relative temporal Fisher information $F(p||q)$ between two probability distributions p_n and q_n as

$$F(p||q)=\left\langle\left(\frac{d\ln(p/q)}{dt}\right)^2\right\rangle_p\equiv\langle(\dot p/p-\dot q/q)^2\rangle_p. \tag{9}$$

This is a definition of the relative temporal Fisher information, where time is the control parameter. Definition (9) is the generalization of more standard relative Fisher information with a nontemporal control parameter [18,32,33], which, however, has not received much attention in physics. $F(p||q)$ can be interpreted as a measure of the relative speeds of system dynamics described by two different distributions p_n and q_n. Additionally, $F(p||q)=0$ if and only if $p_n(t)=q_n(t)$ for all n. To get a more intuitive understanding of $F(p||q)$, let us take again the Poisson distribution for p and q with time-dependent intensity parameters ν_1 and ν_2, respectively. Then it can be shown that $F(p||q)=\nu_1[(\dot\nu_1/\nu_1)-(\dot\nu_2/\nu_2)]^2+(\dot\nu_2/\nu_2)^2(\nu_1-\nu_2)^2$. This means that $F(p||q)$ is zero at any given time only when both intensity parameters and their speeds are equal.

III. GENERAL KINEMATIC BOUNDS ON THE RATES OF DIVERGENCES

In this section, we derive upper bounds on the rates of T_α, R_α, and $D_{\rm KL}$, which we call the kinematic bounds.

A. Limits on rates of divergences via Fisher information

1. Rates of α-coefficient and Tsallis and Renyi divergences

Application of Eq. (4) for $m=2$ and $\lambda_1=\lambda_2=1/2$, with $X_1=\frac{\dot p}{p}$ and $X_2=[(\frac{p}{q})^{\alpha-1}-C_\alpha]$ for the first term on the right in the last line of Eq. (6), and similarly for the second term in that line, yields

$$\left|\frac{dC_\alpha}{dt}\right|\leqslant|\alpha|\sqrt{\left\langle\left(\frac{\dot p}{p}\right)^2\right\rangle_p}\sqrt{\langle[(p/q)^{\alpha-1}-C_\alpha]^2\rangle_p}$$
$$+|\alpha-1|\sqrt{\left\langle\left(\frac{\dot q}{q}\right)^2\right\rangle_q}\sqrt{\langle[(p/q)^{\alpha}-C_\alpha]^2\rangle_q}. \tag{10}$$

The ratios $\langle(\dot p/p)^2\rangle_p$ and $\langle(\dot q/q)^2\rangle_q$ can be identified with temporal Fisher information as in Eq. (8). The final step is to note that

$$\langle[(p/q)^{\alpha-1}-C_\alpha]^2\rangle_p=C_{2\alpha-1}-C_\alpha^2,$$
$$\langle[(p/q)^{\alpha}-C_\alpha]^2\rangle_q=C_{2\alpha}-C_\alpha^2. \tag{11}$$

Interestingly, the above averages correspond to variances of $(p_n/q_n)^{\alpha-1}$ and $(p_n/q_n)^\alpha$ around C_α averaged with respect to either p_n or q_n distributions.

After these substitutions, the general upper bound on dC_α/dt is given by

$$\left|\frac{dC_\alpha}{dt}\right|\leqslant|\alpha|\sqrt{I_F(p)}\sqrt{C_{2\alpha-1}-C_\alpha^2}$$
$$+|\alpha-1|\sqrt{I_F(q)}\sqrt{C_{2\alpha}-C_\alpha^2}. \tag{12}$$

The right-hand side of this inequality is a kinematic limit set on the dynamics of the α-coefficient, and it is called the bound B1 hereafter.

After using transformations in Eqs. (2) and (3), the inequality (12) allows us to find the upper bounds on the rates of Tsallis and Renyi divergences, dT_α/dt and dR_α/dt. They are given by

$$\left|\frac{dT_\alpha}{dt}\right|\leqslant|\alpha|\sqrt{I_F(p)}\sqrt{\frac{2}{(\alpha-1)}(T_{2\alpha-1}-T_\alpha)-T_\alpha^2}$$
$$+\sqrt{I_F(q)}\sqrt{(2\alpha-1)T_{2\alpha}-2(\alpha-1)T_\alpha-(\alpha-1)^2T_\alpha^2}, \tag{13}$$

and

$$\left|\frac{dR_\alpha}{dt}\right|\leqslant\frac{|\alpha|}{|\alpha-1|}\sqrt{I_F(p)}\sqrt{e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)}-1}$$
$$+\sqrt{I_F(q)}\sqrt{e^{[(2\alpha-1)R_{2\alpha}-2(\alpha-1)R_\alpha]}-1}. \tag{14}$$

Equations (12)–(14) constitute the first major result of this paper. They imply that the temporal rates of the Tsallis and Renyi divergence are bounded by the products of the global rates of system dynamics and various nonlinear combinations of associated divergences. It is good to keep in mind that inequalities (12)–(14) have a general character that is independent of the nature of dynamics of probabilities, i.e., valid for both Markovian and non-Markovian dynamics. Moreover, the core Eq. (12) is structurally similar to the upper bound on the average rate of a stochastic observable, recently investigated [23,24].

In a special case when the distribution q is the steady-state distribution of p, i.e., $q = p_\infty$, we have temporal Fisher information $I_F(q) = 0$, and Eqs. (13) and (14) simplify considerably. For instance, for $\alpha = 2$, from Eq. (13) we obtain the bound on the rate of Pearson divergence as $|dT_2/dt| \leqslant 2\sqrt{I_F(p)}\sqrt{2(T_3 - T_2) - T_2^2}$.

2. An example: Weakly time-dependent exponential distributions

To gain insight about the limit set on the rate of the α-coefficient in Eq. (12), it is instructive to analyze an explicit example. Consider two time-dependent continuous distributions $p_x(x) = \nu_1 e^{-\nu_1 x}$ and $q_x(x) = \nu_2 e^{-\nu_2 x}$, where ν_1 and ν_2 are time-dependent but positive. Furthermore, let us assume that $\nu_1(t) = \nu + \epsilon(\Delta + r_1(t))$ and $\nu_2(t) = \nu + \epsilon r_2(t)$, where the parameter $\epsilon \ll 1$, ν and Δ are time-independent, and $r_1(t)$ and $r_2(t)$ are arbitrary time-dependent functions. This parametrization means that $\nu_1(t)$ and $\nu_2(t)$, and their rates, differ only slightly ($\sim\epsilon$) at any moment of time. For this example, we can find explicit simple expressions for all quantities in Eq. (12) to the lowest order in ϵ (see Appendix A). From this it follows that the left-hand side of Eq. (12) is equal to $|\alpha(\alpha-1)|(\epsilon/\nu)^2|r_1 - r_2 + \Delta||\dot{r}_1 - \dot{r}_2| + O(\epsilon^3)$, and the right-hand side is $|\alpha(\alpha-1)|(\epsilon/\nu)^2|r_1 - r_2 + \Delta|(|\dot{r}_1| + |\dot{r}_2|) + O(\epsilon^3)$, where $\dot{r}_1, \dot{r}_2$ are time derivatives of r_1, r_2. Thus the difference between the two sides is set only by the difference between $|\dot{r}_1 - \dot{r}_2|$ and $(|\dot{r}_1| + |\dot{r}_2|)$, i.e., by the relative speeds of r_1 and r_2. The equality of both sides in Eq. (12) is achieved at the moments when $\dot{r}_1$ and $\dot{r}_2$ have the opposite signs, irrespective of the functional dependence of $r_1(t)$ and $r_2(t)$.

3. Rate of Kullback-Leibler divergence

From the first line of Eq. (7) and the above considerations, the upper bound on the rate $dD_{\rm KL}/dt$ immediately follows:

$$\left|\frac{dD_{\rm KL}}{dt}\right| \leqslant \sqrt{I_F(p)}\sqrt{\langle \ln^2(p/q)\rangle_p - D_{\rm KL}^2} + \sqrt{I_F(q)}\sqrt{T_2}. \quad (15)$$

The bound (15) involves the mean of the logarithm square, i.e., $\langle \ln^2(p/q)\rangle_p$, which may be difficult to compute in many practical situations. Therefore, it is good to have an upper limit on the logarithm in terms of other divergences. Such a limit is provided by Eq. (5) [see also Eq. (C4)], which implies for probabilities p and q:

$$\langle \ln^2(p/q)\rangle_p \leqslant \left\langle \left(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}}\right)^2 \right\rangle_p = \left\langle \frac{p}{q}\right\rangle_p - 1 = T_2. \quad (16)$$

Consequently, $dD_{\rm KL}/dt$ is restricted also by

$$\left|\frac{dD_{\rm KL}}{dt}\right| \leqslant \sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}^2} + \sqrt{I_F(q)}\sqrt{T_2}. \quad (17)$$

Obviously, the limit set by Eq. (15) is tighter than the one present in Eq. (17). Moreover, the prominent role in the bound (17) is played by Pearson divergence T_2.

B. Limits on the rates of divergences via relative Fisher information

The rate dC_α/dt in Eq. (6) can be equivalently expressed as

$$\frac{dC_\alpha}{dt} = \left\langle \frac{\dot{q}}{q}\left[\left(\frac{p}{q}\right)^\alpha - C_\alpha\right]\right\rangle_q + \alpha\left\langle \left(\frac{p}{q}\right)^{\alpha-1}[\dot{p}/p - \dot{q}/q]\right\rangle_p$$
$$\leqslant \sqrt{I_F(q)}\sqrt{C_{2\alpha} - C_\alpha^2} + |\alpha|\sqrt{F(p||q)}\sqrt{C_{2\alpha-1}}, \quad (18)$$

where we used the Cauchy-Schwarz inequality and the definition (9) for relative temporal Fisher information. The last line of Eq. (18) is another kinematic limit set on the dynamics of the α-coefficient, and it is called the bound B2 below.

Equation (18) gives us the upper bounds on the rate of Tsallis divergence:

$$\left|\frac{dT_\alpha}{dt}\right| \leqslant \sqrt{I_F(q)}\sqrt{\frac{(2\alpha-1)T_{2\alpha} - 2(\alpha-1)T_\alpha}{(\alpha-1)^2} - T_\alpha^2}$$
$$+ \frac{|\alpha|}{|\alpha-1|}\sqrt{F(p||q)}\sqrt{2(\alpha-1)T_{2\alpha-1} + 1}, \quad (19)$$

and on the rate of Renyi divergence:

$$\left|\frac{dR_\alpha}{dt}\right| \leqslant \frac{\sqrt{I_F(q)}}{|\alpha-1|}\sqrt{e^{(2\alpha-1)R_{2\alpha} - 2(\alpha-1)R_\alpha} - 1}$$
$$+ \frac{|\alpha|}{|\alpha-1|}\sqrt{F(p||q)}e^{(\alpha-1)(R_{2\alpha-1} - R_\alpha)}. \quad (20)$$

By the same token, from the second line of Eq. (7), the rate of Kullback-Leibler divergence is limited by

$$\left|\frac{dD_{\rm KL}}{dt}\right| \leqslant \sqrt{I_F(p)}\sqrt{\langle \ln^2(p/q)\rangle - D_{\rm KL}^2} + \sqrt{F(p||q)}$$
$$\leqslant \sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}^2} + \sqrt{F(p||q)}, \quad (21)$$

where for the second term on the right we used Eq. (4) for $m = 2$ with $X_1 = 1$, $X_2 = \frac{d\ln(p/q)}{dt}$, and $\lambda_1 = \lambda_2 = 1/2$.

Note that the rates dT_α/dt, dR_α/dt, $dD_{\rm KL}/dt$ and their bounds in Eqs. (13), (14), (15), (17), and (19)–(21) are all zero if $p_n(t) = q_n(t)$ for all n, regardless of the temporal dependence of these probabilities. This is because of the properties: $T_\alpha(p||p) = R_\alpha(p||p) = F(p||p) = 0$ for all α.

IV. KINEMATIC-THERMODYNAMIC BOUNDS FOR MARKOV PROCESSES

In this section, we derive bounds on dC_α/dt, dT_α/dt, dR_α/dt, and $dD_{\rm KL}/dt$ that involve both kinematic and thermodynamic variables. To do this, we need to assume that the dynamics of both probability distributions, p_n and q_n, are Markovian and represented by master equations [34]

$$\dot{p}_n = \sum_k (w_{nk}p_k - w_{kn}p_n),$$
$$\dot{q}_n = \sum_k (v_{nk}q_k - v_{kn}q_n), \quad (22)$$

where w_{kn} and v_{kn} are corresponding transition rates for jumps from n to k states. These aggregate transition rates can be composed of several subtransitions corresponding to distinct underlying physical processes labeled by s, i.e., $w_{kn} = \sum_s w_{kn}^{(s)}$ and $v_{kn} = \sum_s v_{kn}^{(s)}$ [7].

A. Bounds of the first kind

1. Rate of α-coefficient

Consider the term $\langle |\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|\rangle_p$ in the last two lines of Eq. (6). The same steps can be taken for the other term in that equation, and hence they are omitted here. First, we make the following decomposition: $\langle |\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|\rangle_p = \langle |\dot{p}/p|^{1/3}|\dot{p}/p|^{2/3}|(p/q)^{\alpha-1} - C_\alpha|\rangle_p$. Second, we apply the Hölder inequality (4) in the latter average with $X_1 = |\dot{p}/p|^{1/3}$, $X_2 = |\dot{p}/p|^{2/3}$, and $X_3 = |(p/q)^{\alpha-1} - C_\alpha|$, and for $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. As a result, we obtain

$$\langle |\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|\rangle_p \leqslant (I_F(p)\langle|\dot{p}/p|\rangle_p\langle|(p/q)^{\alpha-1} - C_\alpha|^3\rangle_p)^{1/3}. \quad (23)$$

The term $\langle|\dot{p}/p|\rangle_p$ can be limited in two different ways (see Appendix B). One way involves internal activity A_p in the system described by distribution p,

$$\langle|\dot{p}/p|\rangle_p \leqslant 2A_p, \quad (24)$$

where A_p, which also can be called the global average escape rate, is defined as [26,27,35]

$$A_p = \frac{1}{2}\sum_{nk}(w_{nk}p_k + w_{kn}p_n) \equiv \sum_n \overline{w}_n p_n = \langle\overline{w}\rangle_p, \quad (25)$$

where $\overline{w}_n = \sum_k w_{kn}$ is the total escape rate from state n.

The second way of bounding $\langle|\dot{p}/p|\rangle_p$ is through (see Appendix B)

$$\langle|\dot{p}/p|\rangle_p \leqslant \sqrt{2\dot{S}_p}\sqrt{A_p}, \quad (26)$$

where $\dot{S}_p$ is the coarse-grained entropy production rate in the system, with the distribution p, defined as [7,34,36]

$$\dot{S}_p = \frac{1}{2}\sum_{nk}(w_{nk}p_k - w_{kn}p_n)\ln\frac{w_{nk}p_k}{w_{kn}p_n}. \quad (27)$$

The inequality (26) is similar to the so-called speed limit relation found for the evolution of Markov thermodynamic systems [25].

The term $\langle|(p/q)^{\alpha-1} - C_\alpha|^3\rangle_p$ in Eq. (23) is bounded by [see Eq. (C1) in Appendix C]

$$\langle|(p/q)^{\alpha-1} - C_\alpha|^3\rangle_p \leqslant C_{3\alpha-2} - C_\alpha C_{2\alpha-1}. \quad (28)$$

Combining Eqs. (23), (24), (26), and (28), we obtain either

$$\langle |\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|\rangle_p \leqslant (2I_F(p)A_p[C_{3\alpha-2} - C_\alpha C_{2\alpha-1}])^{1/3}, \quad (29)$$

or

$$\langle |\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|\rangle_p \leqslant (I_F(p)\sqrt{2\dot{S}_pA_p}[C_{3\alpha-2} - C_\alpha C_{2\alpha-1}])^{1/3}. \quad (30)$$

Analogical inequalities can be obtained for the remaining term in the last line of Eq. (6), with appropriately defined $\dot{S}_q$ and A_q for the distribution q. Taking all that into account leads to the limits on the rate of the α-coefficient, as

$$\left|\frac{dC_\alpha}{dt}\right| \leqslant |\alpha|(2I_F(p)A_p(C_{3\alpha-2} - C_\alpha C_{2\alpha-1}))^{1/3} + |\alpha-1|(2I_F(q)A_q(C_{3\alpha} - C_\alpha C_{2\alpha}))^{1/3}, \quad (31)$$

which is a strictly kinematic bound with the right-hand side called from now on the bound B3, and

$$\left|\frac{dC_\alpha}{dt}\right| \leqslant |\alpha|(I_F(p)\sqrt{2\dot{S}_pA_p}[C_{3\alpha-2} - C_\alpha C_{2\alpha-1}])^{1/3} + |\alpha-1|(I_F(q)\sqrt{2\dot{S}_qA_q}[C_{3\alpha} - C_\alpha C_{2\alpha}])^{1/3}, \quad (32)$$

which represents a mixed kinematic-thermodynamic bound called the bound B4. These two equations constitute the third major result of this paper. They mean that for Markov dynamics dC_α/dt can be bounded not only by the global rate of system dynamics $[I_F(p), I_F(q)]$, but also by average activities (A_p, A_q), and/or the thermodynamic entropy production rates $(\dot{S}_p, \dot{S}_q)$. This generally suggests that the kinematic characteristics of the stochastic system are at least as important as the entropic (energetic) characteristics, in agreement with the notions in Ref. [37]. Furthermore, the restriction to Markov dynamics makes the bounds in Eqs. (31) and (32) less general than the bounds in Eqs. (12) and (18).

2. Rates of Tsallis and Renyi divergences

The inequalities in Eqs. (31) and (32) allow us to write the corresponding kinematic and thermodynamic bounds on the rate of Tsallis and Renyi divergences. For dT_α/dt we get the following inequality:

$$\left|\frac{dT_\alpha}{dt}\right| \leqslant \frac{|\alpha|}{|\alpha-1|^{2/3}}(I_F(p)\sqrt{A_p\Psi_p}[3T_{3\alpha-2} - 2T_{2\alpha-1}[1+(\alpha-1)T_\alpha] - T_\alpha])^{1/3} + (I_F(q)\sqrt{A_q\Psi_q}[(3\alpha-1)T_{3\alpha} - (2\alpha-1)T_{2\alpha}[1+(\alpha-1)T_\alpha] - (\alpha-1)T_\alpha])^{1/3}, \quad (33)$$

and for dR_α/dt we have

$$\left|\frac{dR_\alpha}{dt}\right| \leqslant \frac{|\alpha|}{|\alpha-1|}(I_F(p)\sqrt{A_p\Psi_p}[e^{3(\alpha-1)(R_{3\alpha-2}-R_\alpha)} - e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)}])^{1/3} + (I_F(q)\sqrt{A_q\Psi_q}[e^{(3\alpha-1)R_{3\alpha}-3(\alpha-1)R_\alpha} - e^{(2\alpha-1)R_{2\alpha}-2(\alpha-1)R_\alpha}])^{1/3}, \quad (34)$$

where Ψ_γ, with index γ either p or q, is

$$\Psi_\gamma = \begin{cases} 4A_\gamma, & \text{K bound,} \\ 2\dot{S}_\gamma, & \text{KT bound,} \end{cases} \tag{35}$$

with K and KT denoting, respectively, purely kinematic and mixed kinematic-thermodynamic bounds.

Equations (33) and (34) are slightly more complicated than the basal Eqs. (31) and (32), chiefly by the presence of various combinations of Tsallis and Renyi divergences of different order. However, in the case when the distribution of q is a steady-state distribution of p, the terms proportional to $I_F(q)$ vanish, and Eqs. (33) and (34) take simpler forms. For instance, for $\alpha = 2$, we obtain the limit on the rate of Pearson divergence as $|dT_2/dt| \leqslant 2(I_F(p)\sqrt{A_p\Psi_p}[3T_4 - 2T_3(1+T_2) - T_2])^{1/3}$.

3. Rate of Kullback-Leibler divergence

Now we turn to the rate of Kullback-Leibler divergence, with Eq. (7) as the starting point. Applying the Hölder inequality in the same way as above, we get

$$\langle|\dot{p}/p||\ln(p/q) - D_{\rm KL}|\rangle_p \leqslant (I_F(p)\langle|\dot{p}/p|\rangle_p\langle|\ln(p/q) - D_{\rm KL}|^3\rangle_p)^{1/3} \tag{36}$$

and

$$\langle|\dot{q}/q||(p/q) - 1|\rangle_q \leqslant (I_F(q)\langle|\dot{q}/q|\rangle_q\langle|(p/q) - 1|^3\rangle_q)^{1/3}. \tag{37}$$

The terms $\langle|\dot{p}/p|\rangle_p$ and $\langle|\dot{q}/q|\rangle_q$ are restricted by Eqs. (24) and (26) or their analogs.

The bound on $\langle|(p/q) - 1|^3\rangle_q$ is obtained immediately from Eq. (C1), with the result

$$\langle|(p/q) - 1|^3\rangle_q \leqslant C_3 - C_2 = 2T_3 - T_2. \tag{38}$$

Estimating $\langle|\ln(p/q) - D_{\rm KL}|^3\rangle_p$ requires more transformations. With the help of Eq. (C5) in Appendix C, that term can be bounded by various α-coefficients as

$$\langle|\ln(p/q) - D_{\rm KL}|^3\rangle_p \leqslant e^{-3D_{\rm KL}/2}C_{5/2} - e^{-D_{\rm KL}/2}C_{3/2} - e^{D_{\rm KL}/2}C_{1/2} + e^{3D_{\rm KL}/2}C_{-1/2}. \tag{39}$$

Combining Eqs. (7), (24), (26), and (36)–(39), we obtain the limit on the rate of KL divergence,

$$\left|\frac{dD_{\rm KL}}{dt}\right| \leqslant (I_F(q)\sqrt{A_q\Psi_q}[2T_3 - T_2])^{1/3} + (I_F(p)\sqrt{A_p\Psi_p}[e^{-3D_{\rm KL}/2}C_{5/2} - e^{-D_{\rm KL}/2}C_{3/2} - e^{D_{\rm KL}/2}C_{1/2} + e^{3D_{\rm KL}/2}C_{-1/2}])^{1/3}, \tag{40}$$

where the quantities Ψ_p and Ψ_q are given by Eq. (35). As can be seen, apart from similar terms to those in Eqs. (33) and (34), the upper bound contains also various exponents of $D_{\rm KL}$. More broadly, one can interpret the kinematic-thermodynamic bounds in Eqs. (33), (34), and (40) that the predictability of the system dynamics is associated with its levels of dissipation and dynamical agitation. The smaller these two factors are, the better is the prediction of the dynamics.

B. Bounds of the second kind

In this section we derive a second, alternative, thermodynamic-kinematic bound on the rate of statistical divergences.

Consider the first line of Eq. (6). We can substitute for $\dot{p}_n$ and $\dot{q}_n$ in this equation their Master equation dynamics given by Eq. (22). This leads to

$$\left|\frac{dC_\alpha}{dt}\right| \leqslant |\alpha|\sum_{nk}|w_{nk}p_k - w_{kn}p_n||(p_n/q_n)^{\alpha-1} - C_\alpha| + |\alpha - 1|\sum_{nk}|v_{nk}q_k - v_{kn}q_n||(p_n/q_n)^\alpha - C_\alpha|. \tag{41}$$

The first term on the right proportional to $|\alpha|$ can be limited again in two different ways [see Eq. (B6) in Appendix B]:

$$\sum_{nk}|w_{nk}p_k - w_{kn}p_n||(p_n/q_n)^{\alpha-1} - C_\alpha| \leqslant \sqrt{\Psi_p/2}(\sqrt{I_F(p)} + 2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}\langle[(p/q)^{\alpha-1} - C_\alpha]^4\rangle_p^{1/4}, \tag{42}$$

where $\langle\overline{w}^2\rangle_p = \sum_n \overline{w}_n^2 p_n$ is the second moment of total escape rate.

Moreover, by a direct computation we have

$$\langle[(p/q)^{\alpha-1} - C_\alpha]^4\rangle_p = C_{4\alpha-3} - 4C_\alpha C_{3\alpha-2} + 6C_\alpha^2 C_{2\alpha-1} - 3C_\alpha^4. \tag{43}$$

Combining Eqs. (41)–(43), and applying the same reasoning for the second term in Eq. (41), we obtain two upper bounds on the rate of the α-coefficient depending on the value for Ψ:

$$\left|\frac{dC_\alpha}{dt}\right| \leqslant |\alpha|\sqrt{\Psi_p/2}(\sqrt{I_F(p)} + 2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}\left(C_{4\alpha-3} - 4C_\alpha C_{3\alpha-2} + 6C_\alpha^2 C_{2\alpha-1} - 3C_\alpha^4\right)^{1/4}$$
$$+ |\alpha - 1|\sqrt{\Psi_q/2}(\sqrt{I_F(q)} + 2\sqrt{\langle\overline{v}^2\rangle_q})^{1/2}\left(C_{4\alpha} - 4C_\alpha C_{3\alpha} + 6C_\alpha^2 C_{2\alpha} - 3C_\alpha^4\right)^{1/4}. \tag{44}$$

TABLE I. Summary of the inequalities for the rates of divergences. Ψ_γ is equal either $4A_\gamma$ or $2\dot{S}_\gamma$.

Bound	Divergence type	Equation
B1		
	Chernoff	$\lvert\dot{C}_\alpha\rvert \leqslant \lvert\alpha\rvert\sqrt{I_F(p)}\sqrt{C_{2\alpha-1}-C_\alpha^2}+\lvert\alpha-1\rvert\sqrt{I_F(q)}\sqrt{C_{2\alpha}-C_\alpha^2}$
	Tsallis	$\lvert\dot{T}_\alpha\rvert \leqslant \lvert\alpha\rvert\sqrt{I_F(p)}[\frac{2}{(\alpha-1)}(T_{2\alpha-1}-T_\alpha)-T_\alpha^2]^{1/2}+\sqrt{I_F(q)}[(2\alpha-1)T_{2\alpha}-2(\alpha-1)T_\alpha-(\alpha-1)^2T_\alpha^2]^{1/2}$
	Renyi	$\lvert\dot{R}_\alpha\rvert \leqslant \frac{\lvert\alpha\rvert}{\lvert\alpha-1\rvert}\sqrt{I_F(p)}[e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)}-1]^{1/2}+\sqrt{I_F(q)}[e^{[(2\alpha-1)R_{2\alpha}-2(\alpha-1)R_\alpha]}-1]^{1/2}$
	Kullback-Leibler	$\lvert\dot{D}_{\mathrm{KL}}\rvert \leqslant \sqrt{I_F(p)}\sqrt{T_2-D_{\mathrm{KL}}^2}+\sqrt{I_F(q)}\sqrt{T_2}$
B2		
	Chernoff	$\lvert\dot{C}_\alpha\rvert \leqslant \sqrt{I_F(q)}\sqrt{C_{2\alpha}-C_\alpha^2}+\lvert\alpha\rvert\sqrt{F(p\Vert q)}\sqrt{C_{2\alpha-1}}$
	Tsallis	$\lvert\dot{T}_\alpha\rvert \leqslant \sqrt{I_F(q)}[\frac{(2\alpha-1)T_{2\alpha}-2(\alpha-1)T_\alpha}{(\alpha-1)^2}-T_\alpha^2]^{1/2}+\frac{\lvert\alpha\rvert}{\lvert\alpha-1\rvert}\sqrt{F(p\Vert q)}[2(\alpha-1)T_{2\alpha-1}+1]^{1/2}$
	Renyi	$\lvert\dot{R}_\alpha\rvert \leqslant \frac{\sqrt{I_F(q)}}{\lvert\alpha-1\rvert}\sqrt{e^{(2\alpha-1)R_{2\alpha}-2(\alpha-1)R_\alpha}-1}+\frac{\lvert\alpha\rvert}{\lvert\alpha-1\rvert}\sqrt{F(p\Vert q)}e^{(\alpha-1)(R_{2\alpha-1}-R_\alpha)}$
	Kullback-Leibler	$\lvert\dot{D}_{\mathrm{KL}}\rvert \leqslant \sqrt{I_F(p)}\sqrt{T_2-D_{\mathrm{KL}}^2}+\sqrt{F(p\Vert q)}$
B3,B4		
	Chernoff	$\lvert\dot{C}_\alpha\rvert \leqslant \lvert\alpha\rvert(I_F(p)\sqrt{A_p\Psi_p}[C_{3\alpha-2}-C_\alpha C_{2\alpha-1}])^{1/3}+\lvert\alpha-1\rvert(I_F(q)\sqrt{A_q\Psi_q}[C_{3\alpha}-C_\alpha C_{2\alpha}])^{1/3}$
	Tsallis	$\lvert\dot{T}_\alpha\rvert \leqslant \frac{\lvert\alpha\rvert}{\lvert\alpha-1\rvert^{2/3}}(I_F(p)\sqrt{A_p\Psi_p}[3T_{3\alpha-2}-2T_{2\alpha-1}[1+(\alpha-1)T_\alpha]-T_\alpha])^{1/3}$
		$+(I_F(q)\sqrt{A_q\Psi_q}[(3\alpha-1)T_{3\alpha}-(2\alpha-1)T_{2\alpha}[1+(\alpha-1)T_\alpha]-(\alpha-1)T_\alpha])^{1/3}$
	Renyi	$\lvert\dot{R}_\alpha\rvert \leqslant \frac{\lvert\alpha\rvert}{\lvert\alpha-1\rvert}(I_F(p)\sqrt{A_p\Psi_p}[e^{3(\alpha-1)(R_{3\alpha-2}-R_\alpha)}-e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)}])^{1/3}$
		$+(I_F(q)\sqrt{A_q\Psi_q}[e^{(3\alpha-1)R_{3\alpha}-3(\alpha-1)R_\alpha}-e^{(2\alpha-1)R_{2\alpha}-2(\alpha-1)R_\alpha}])^{1/3}$
	Kullback-Leibler	$\lvert\dot{D}_{\mathrm{KL}}\rvert \leqslant (I_F(q)\sqrt{A_q\Psi_q}[2T_3-T_2])^{1/3}+(I_F(p)\sqrt{A_p\Psi_p})^{1/3}$
		$\times[e^{-3D_{\mathrm{KL}}/2}C_{5/2}-e^{-D_{\mathrm{KL}}/2}C_{3/2}-e^{D_{\mathrm{KL}}/2}C_{1/2}+e^{3D_{\mathrm{KL}}/2}C_{-1/2}]^{1/3}$
B5,B6		
	Chernoff	$\lvert\dot{C}_\alpha\rvert \leqslant \lvert\alpha\rvert\sqrt{\Psi_p/2}(\sqrt{I_F(p)}+2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}(C_{4\alpha-3}-4C_\alpha C_{3\alpha-2}+6C_\alpha^2C_{2\alpha-1}-3C_\alpha^4)^{1/4}$
		$+\lvert\alpha-1\rvert\sqrt{\Psi_q/2}(\sqrt{I_F(q)}+2\sqrt{\langle\overline{v}^2\rangle_q})^{1/2}(C_{4\alpha}-4C_\alpha C_{3\alpha}+6C_\alpha^2C_{2\alpha}-3C_\alpha^4)^{1/4}$
	Tsallis	
	Renyi	
	Kullback-Leibler	$\lvert\dot{D}_{\mathrm{KL}}\rvert \leqslant \sqrt{\Psi_p/2}(\sqrt{I_F(p)}+2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}(e^{-2D_{\mathrm{KL}}}C_3-4e^{-D_{\mathrm{KL}}}C_2+6-4e^{D_{\mathrm{KL}}}+e^{2D_{KL}}C_{-1})^{1/4}$
		$+\sqrt{\Psi_q/2}(\sqrt{I_F(q)}+2\sqrt{\langle\overline{v}^2\rangle_q})^{1/2}(3T_4-8T_3+6T_2)^{1/4}$

This equation is an alternative to Eqs. (31) and (32), though a little more complicated, and it combines a purely kinematic bound called B5 (for $\Psi = 4A$) with a mixed kinematic-thermodynamic bound called B6 (for $\Psi = 2\dot{S}$). Note that in the steady state for both probability distributions p and q, all the terms dC_α/dt, $\dot{S}_p$, $\dot{S}_q$, $I_F(p)$, and $I_F(q)$ are zero, but $\langle\overline{w}^2\rangle_p$ and $\langle\overline{v}^2\rangle_p$ are nonzero.

The corresponding bounds on the rates of Tsallis and Renyi divergences can be obtained straightforwardly from Eq. (44), using transformation in Eqs. (2) and (3). The resulting inequalities are similar to Eq. (44), although more elaborate due to a more complicated combination of α-coefficients. Below, instead, we provide an explicit bound on the rate of KL divergence, which takes the form

$$\left|\frac{dD_{\mathrm{KL}}}{dt}\right| \leqslant \sqrt{\Psi_p/2}\Big(\sqrt{I_F(p)}+2\sqrt{\langle\overline{w}^2\rangle_p}\Big)^{1/2} \times (\langle[\ln(p/q)-D_{\mathrm{KL}}]^4\rangle_p)^{1/4} + \sqrt{\Psi_q/2}\Big(\sqrt{I_F(q)}+2\sqrt{\langle\overline{v}^2\rangle_q}\Big)^{1/2} \times (3T_4-8T_3+6T_2)^{1/4}, \tag{45}$$

where $\langle[\,\ln(p/q)-D_{\mathrm{KL}}]^4\rangle_p$ can be bounded as [see Eq. (C6) in Appendix C]

$$\langle[\ln(p/q)-D_{\mathrm{KL}}]^4\rangle_p \leqslant e^{-2D_{\mathrm{KL}}}C_3-4e^{-D_{\mathrm{KL}}}C_2+6 - 4e^{D_{\mathrm{KL}}}+e^{2D_{\mathrm{KL}}}C_{-1}. \tag{46}$$

Together Eqs. (44)–(46) constitute the fourth major result of this work. They imply that the rate of information gain about the system dynamics is restricted by both thermodynamic and kinematic characteristics, both of the true system (probabilities q) and its estimator (probabilities p).

V. COMPARISON OF THE BOUNDS: ONE-STEP DRIVEN PROCESS

In Table I, all the derived bounds on the rates of statistical divergences are summarized.

Next, we check the quality of the six upper bounds on dC_α/dt, denoted by B1–B6 and represented by Eqs. (12), (18), (31), (32), and (44), respectively. We choose a specific example of a stochastic dynamical system known as the one-step Markov jump process (known also as the birth-and-death process) with $N+1$ states [38]. We consider two versions of this system: one driven by periodic stimulation, and another

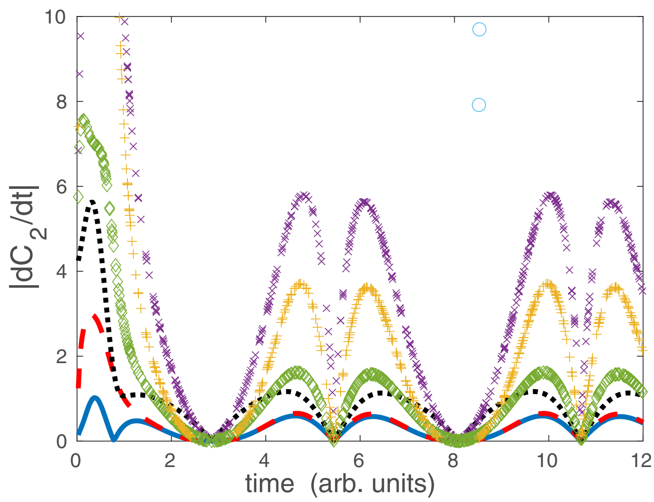

FIG. 1. Rate of α-coefficient in comparison to its various upper bounds as functions of time for $\alpha = 2$, corresponding to Pearson divergence. The solid line (blue) corresponds to the exact value of $|dC_2/dt|$ and was computed from Eq. (6). Upper bounds on $|dC_2/dt|$, i.e., B1–B6 are shown as a dashed (red) line for B1; a dotted (black) line for B2; crosses (purple) for B3; diamonds (green) for B4; circles (light blue) for B5; and pluses (yellow) for B6. Note that the best estimates for $|dC_2/dt|$ are given by the kinematic bounds B1 (dashed line) and B2 (dotted line), but the former is closer to the actual value of $|dC_2/dt|$. The bound B5 provides a poor estimate and is mostly out of scale. Parameters used are $a = a_0 = 3.0$, $b_0 = 1.0$, $g = 0.7$, $\omega = 1.2$, and $N = 9$.

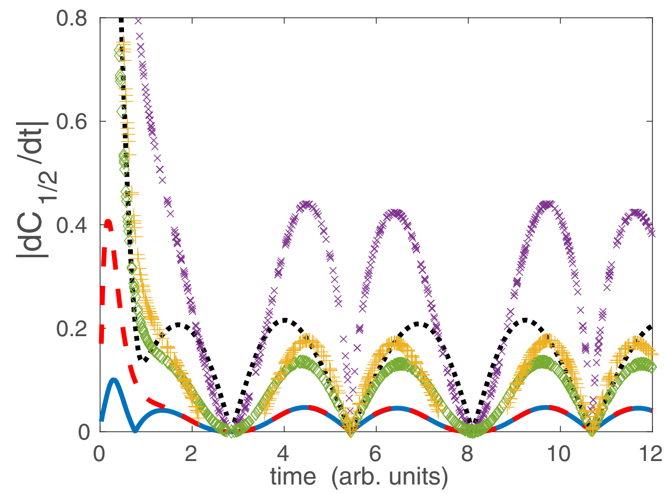

FIG. 2. The same as in Fig. 1 but for $\alpha = 1/2$, corresponding to Hellinger distance. Again the bound B1 gives the best estimate, but the accuracy for some other bounds is different from that in Fig. 1. Notably, the bounds B4 (diamonds) and B6 (pluses) provide often better estimates than the bound B2 (dotted line). The same labels used and parameters are as in Fig. 1.

relaxing to its steady state. For the driven case, the probability p_n of being in state n is described by the following master equation:

$$\dot{p}_n = w_{n,n-1}p_{n-1} + w_{n,n+1}p_{n+1} - (w_{n-1,n} + w_{n+1,n})p_n,$$

for $n = 1, \dots, N-1$, and for the boundary probabilities we have $\dot{p}_0 = w_{0,1}p_1 - w_{1,0}p_0$ and $\dot{p}_N = w_{N,N-1}p_{N-1} - w_{N-1,N}p_N$, with the transition rates $w_{n-1,n} = a_0 n$, $w_{n+1,n} = b(t)(N-n)$, where the time-dependent oscillating rate $b(t) = b_0(1 + g[\cos(\omega t) + 1])$. The parameters a_0 and b_0 are, respectively, the amplitudes of the downhill and uphill transitions, and oscillations of $b(t)$ are controlled by amplitude g and frequency ω.

For the relaxing case, we have the same structure of the master equation as above, but we denote the corresponding probabilities as q_n, with the time-independent transition rates $v_{n-1,n} = w_{n-1,n}$ and $v_{n+1,n} = b_0(N-n)$. In both cases, the same initial condition on the probabilities was used, i.e., $p_i(0) = q_i(0) = \frac{1}{N+1}$ for all $i = 0, 1, \dots, N$, which means that initially all the states are equally likely.

For this system we compute numerically the α-coefficient $C_\alpha(p||q)$ and its time derivative, as well as all B1–B6 bounds on dC_α/dt. Overall, the best estimate for $|dC_\alpha/dt|$ is provided by the bound B1, and the discrepancy between the two is very small as time progresses (Figs. 1 and 2). The bounds B2, B4, and B6 compete for second place, but their ranking changes dynamically. Their mutual relationship depends also on the order α (compare Figs. 1 and 2). The kinematic bounds B3 and B5 are rather weak, especially B5, which does not fit into the scale of Figs. 1 and 2. This suggests that purely kinematic bounds for Markov processes obeying the master equation do not capture well the rates of statistical divergences.

The superiority of the general kinematic bound B1 [Eq. (12)] follows from two facts. The first is that its derivation involves a minimal number of mathematical transformations, i.e., fewer consecutive inequalities on the way are required, and thus fewer inaccuracies are introduced. The second reason is more subtle and it concerns the number of constraints on the physical variables appearing in the bounds, which have to be satisfied to make the bounds good estimates of $|dC_\alpha/dt|$. The larger the number of constraints, the less likely the bound will be reached. To be more specific, let us compare the bound B1 with the bounds B3 and B4. The bound B1 is derived from Hölder inequality [Eq. (4)] for $m = 2$, i.e., only two variables are involved. Consequently, this inequality becomes an equality when only one constraint relating the two variables is satisfied ($|X_1|^{1/\lambda_1} \sim |X_2|^{1/\lambda_2}$). The equality in the Hölder inequality corresponds to the saturation of the bound B1. On the other hand, the bounds B3 and B4 [Eqs. (31) and (32)] are derived from Eq. (4) with $m = 3$ variables. To achieve equality in Eq. (4) in this case requires two constraints on these three variables ($|X_1|^{1/\lambda_1} \sim |X_2|^{1/\lambda_2} \sim |X_3|^{1/\lambda_3}$), which is much more restrictive on the dynamics of these variables than in the former case. As a result, the bounds B3 and B4 are more difficult to reach, and their values deviate significantly from the actual value of $|dC_\alpha/dt|$.

It is also interesting to consider why the bound B3 is much less accurate than the bound B4, given that they both follow from a similar derivation scheme with $m = 3$ in the Hölder inequality. Since the bounds B3 and B4 differ only by the factor Ψ_γ in Eq. (35), the fact that B3 gives much larger values than B4 means that entropy production $\dot{S}_\gamma$ is smaller than activity A_γ. This seems reasonable because $\dot{S}_\gamma$ can be close to 0 for systems close to equilibrium, while activity

A_γ is always strictly positive and can be large regardless of the distance from equilibrium [37]. Thus, in this particular case, having more information about the system (both activity and entropy production rather than activity only), is more advantageous and produces a tighter bound. This is similar to the case of the improved thermodynamic uncertainty relation [22,27]. However, this is not a general rule, as the case of B1 bound versus B3, B4 bounds shows. For the former bound we have less specific information about the system, and yet that bound is shown to perform the best.

VI. APPLICATIONS OF INEQUALITIES FOR THE RATES OF DIVERGENCES

The above inequalities for the upper bounds of various divergences can be used in different circumstances encountered in physics and interdisciplinary research. Having the bound on the module of divergence rate $\dot{D}_\alpha$ $(= dD_\alpha/dt)$, where D_α is either Tsallis T_α or Renyi divergence R_α, allows us to find the discrepancy between D_α at different time moments. In particular, because of the general relationship

$$D_\alpha(T) - D_\alpha(0) = \int_0^T \dot{D}_\alpha dt \leqslant \sqrt{T}\sqrt{\int_0^T |\dot{D}_\alpha|^2 dt}, \quad (47)$$

and because of the upper bounds on $|\dot{D}_\alpha|$, we can estimate the maximal difference between the values of divergences at initial and some later arbitrary time T.

It is worthwhile to stress that the basic inequalities on the rates of statistical divergences [Eqs. (12)–(15)] have a very similar structure to the inequalities for the average rates of stochastic observables [23,24]. Both of them follow from the Cauchy-Schwarz inequality and contain temporal Fisher information. In our case, the role of the observable is played by the statistical divergence, which can be interpreted as a generalized information gain.

A. Minimal speed and entropy production in terms of the rates of statistical divergences

In recent years, different speed limits on stochastic thermodynamics in different systems have been found [22,25,39–43]. Similarly, there has been an interest in determining minimal entropy production during stochastic evolution [22,44–49]. Here, we provide alternative lower limits on the speed of dynamical systems and their entropy production, using statistical divergences.

In a particular case when divergence $D_\alpha(p|p_\infty)$ is between the time-dependent system's probability distribution p and its steady-state distribution p_∞, the divergence $D_\alpha(p|p_\infty)$ can be interpreted as generalized information gain in relation to its steady state. Consequently, the rate of divergence $\dot{D}_\alpha(p|p_\infty)$ can be thought of as the speed of information gain away from the steady state.

The speed of global system dynamics can be defined as a square root of temporal Fisher information, i.e., $\sqrt{I_F(p)}$. Thus, Eqs. (13) and (14) provide lower bounds on the speed of system evolution through either Tsallis or Renyi divergences, or the α-coefficient, as

$$\sqrt{I_F(p)} \geqslant \frac{|\dot{C}_\alpha(p||p_\infty)|}{|\alpha|\sqrt{C_{2\alpha-1} - C_\alpha^2}}, \quad (48)$$

$$\sqrt{I_F(p)} \geqslant \frac{|\dot{T}_\alpha(p||p_\infty)|}{|\alpha|\sqrt{\frac{2}{(\alpha-1)}(T_{2\alpha-1} - T_\alpha) - T_\alpha^2}}, \quad (49)$$

and

$$\sqrt{I_F(p)} \geqslant \frac{|\alpha-1||\dot{R}_\alpha(p||p_\infty)|}{|\alpha|\sqrt{e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)} - 1}}. \quad (50)$$

These inequalities imply that the minimal speed of the system's stochastic dynamics is set by the rate of generalized information gain in this system.

Similarly, we can provide lower bounds on the entropy production rate in stochastic Markov systems by inverting Eqs. (32)–(34). As before, by considering statistical divergences between the time-dependent distribution p and its steady-state form p_∞, we obtain the following inequalities for $\dot{S}_p$:

$$\dot{S}_p \geqslant \frac{\dot{C}_\alpha(p||p_\infty)^6}{2\alpha^6 A_p I_F(p)^2[C_{3\alpha-2} - C_\alpha C_{2\alpha-1}]^2}, \quad (51)$$

and via the rates of Tsallis and Renyi divergences

$$\dot{S}_p \geqslant \frac{(\alpha-1)^4 \dot{T}_\alpha(p||p_\infty)^6}{2\alpha^6 A_p I_F(p)^2} \times \{3T_{3\alpha-2} - 2T_{2\alpha-1}[1 + (\alpha-1)T_\alpha] - T_\alpha\}^{-2} \quad (52)$$

and

$$\dot{S}_p \geqslant \frac{[(\alpha-1)\dot{R}_\alpha(p||p_\infty)]^6}{2\alpha^6 A_p I_F(p)^2} \times [e^{3(\alpha-1)(R_{3\alpha-2}-R_\alpha)} - e^{2(\alpha-1)(R_{2\alpha-1}-R_\alpha)}]^{-2}. \quad (53)$$

Equations (51)–(53) determine minimal dissipation for stochastic thermodynamic systems in terms of the rates of generalized information gains, the system's average activity, and its speed. As such, they are alternatives to the minimal limits on entropy production derived in other ways, and involving other quantities [22,44–49]. Interestingly, the minimal dissipation in the system is inversely proportional to the product of the system's activity and the fourth power of its speed. Thus, paradoxically, it is possible, in principle, to increase dissipation by decreasing activity A_p and global speed $\sqrt{I_F}$, as long as the rate of information gain is fixed.

Other, more specific, applications of the rates of divergences are provided below.

B. Applications in physics and biophysics

1. Limits on nonequilibrium rates of free energy and work in thermodynamics

Let us consider a physical system in contact with an environment (heat bath) at temperature T. Our goal is to find the bounds on the rates of available free energy and work associated with this system, which can be in thermal equilibrium or in nonequilibrium with the environment. In the first case, we describe the system by the probability distribution

$p_{eq,n}(t)$, that it is in state n at time t, while in the second, nonequilibrium case, we describe our system analogously by the probability distribution $p_n(t)$. The nonequilibrium version of the second law of thermodynamics for our system is [50]

$$\dot{W} - \dot{F} = k_B T \dot{S}, \tag{54}$$

where $\dot{W}$ is the rate of work performed on the system, $\dot{F}$ is the rate of nonequilibrium free energy, k_B is the Boltzmann constant, and $k_B T\dot{S}$ is the physical entropy production rate in energy units. Because of the presence of dissipation in the system, which mathematically means that $\dot{S} \geqslant 0$, we obtain the second law as $\dot{W} \geqslant \dot{F}$, or equivalently $\Delta W \geqslant \Delta F$. These inequalities indicate that the maximal useful work that can be extracted $(-\Delta W)$ is at most equal to the corresponding decrease in nonequilibrium free energy $(-\Delta F)$.

The time-dependent nonequilibrium free energy $F(t)$ is related to the time-dependent equilibrium free energy $F_{\rm eq}(t)$ by [8,50]

$$F(t) - F_{\rm eq}(t) = k_B T D_{\rm KL}(p||p_{\rm eq}), \tag{55}$$

which means that nonequilibrium free energy is always greater than the equilibrium one by the amount of information needed to specify the nonequilibrium state (quantified by the KL divergence between the distributions p_n and $p_{{\rm eq},n}$). The differences of the rates of these free energies, $\dot{F} - \dot{F}_{\rm eq} = k_B T \dot{D}_{\rm KL}(p||p_{\rm eq})$, are thus restricted by the bounds on KL divergence, given by Eqs. (15), (17), or (40) and (45). For example, using Eq. (17), we obtain bounds on the nonequilibrium free energy rate as

$$|\dot{F} - \dot{F}_{\rm eq}| \leqslant k_B T(\sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}} + \sqrt{I_F(p_{\rm eq})}\sqrt{T_2}). \tag{56}$$

This means that the speed with which free energy changes is limited by the speeds of global dynamics of nonequilibrium and equilibrium versions of the system [i.e., $I_F(p)$ and $I_F(p_{\rm eq})$], as well as by the Pearson and KL divergences between nonequilibrium and equilibrium distributions [i.e., $T_2(p||p_{\rm eq})$ and $D_{\rm KL}(p||p_{\rm eq})$].

Using Eqs. (54) and (56), we can also write the bounds on the rate of work performed on the system. We obtain

$$\begin{aligned} &-k_B T(\sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}} + \sqrt{I_F(p_{\rm eq})}\sqrt{T_2}) \\ &\leqslant \dot{W} - \dot{F}_{\rm eq} - k_B T\dot{S} \\ &\leqslant k_B T(\sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}} + \sqrt{I_F(p_{\rm eq})}\sqrt{T_2}). \end{aligned} \tag{57}$$

These inequalities allow us to find lower and upper bounds on the rates of extracted work from $(-\dot{W})$ or done on $(\dot{W})$ the thermodynamic system. In a particular case when the equilibrium probability distribution is time-independent, i.e., $\dot{p}_{{\rm eq},n} = 0$, we obtain a simpler formula for the maximal extracted work rate: $-\dot{W} \leqslant k_B T(-\dot{S} + \sqrt{I_F(p)}\sqrt{T_2 - D_{\rm KL}})$. That work rate is restricted not only by the entropy production rate but also by the global speed of the nonequilibrium state and the difference in Pearson and KL divergences.

2. *Overdamped particle in time-dependent potential versus "target" potential*

This example concerns kinematic bounds on the rates of divergences given by Eqs. (12)–(14). Consider a Brownian massless particle moving in 1D with trajectory $x(t)$ in a stochastic environment with a damping force $-kdx/dt$ (k is some positive constant). We study the motion of this particle in two different external time-dependent potentials, either $V_1(s_1(t))$ or $V_2(s_2(t))$, which are influenced by two time-dependent arbitrary signals $s_1(t)$ and $s_2(t)$. We call the potential V_2 the target or "desired" potential, and V_1 is the actual potential. Our goal is to study how fast the actual trajectory of the particle, corresponding to the V_1 potential, diverges from the target trajectory corresponding to the V_2 potential. For analytical tractability, we assume harmonic potentials in both cases, i.e., $V_i = \frac{1}{2}k\gamma[x - s_i(t)]^2$, where γ is the inverse of the (relaxation) time constant of the system. In this case, the signals $s_i(t)$ are the centers of the two potentials.

The equation of motion in both potentials is

$$\dot{x} = -\gamma[x - s_i(t)] + \sqrt{2\gamma\sigma^2}\eta(t), \tag{58}$$

where $i = 1, 2$ and it corresponds to the case with potential either V_1 or V_2, σ is the standard deviation of the noise in the system, and $\eta(t)$ is the δ-correlated Gaussian random variable related to the noise such that $\langle\eta(t)\rangle = 0$ and $\langle\eta(t)\eta(t')\rangle = \delta(t - t')$.

The trajectory $x(t)$ of the particle tries to follow the instantaneous value of the signal $s_i(t)$, but it is distracted by the noise, and thus the particle position is a stochastic variable. This particle dynamics can be described equivalently by the dynamics of the probability density of the particle position in terms of the Fokker-Planck equation as

$$\frac{\partial P_i(x,t)}{\partial t} = -\frac{\partial J_i(x,t)}{\partial x}, \tag{59}$$

with

$$J_i(x,t) = -\gamma[x - s_i(t)]P_i(x,t) - \gamma\sigma^2\frac{\partial P_i(x,t)}{\partial x}, \tag{60}$$

where $P_i(x,t)$ is the probability density of particle position in the potential related to signal $s_i(t)$, and $J_i(x,t)$ is the probability flux ($i = 1, 2$). Equation (59) can be exactly solved yielding [38]

$$P_i(x,t) = \frac{\exp\left(-\frac{\{x - e^{-\gamma t}[x_0 + \gamma g_i(t)]\}^2}{2\sigma^2(1 - e^{-2\gamma t})}\right)}{\sqrt{2\pi\sigma^2(1 - e^{-2\gamma t})}}, \tag{61}$$

where $g_i(t) = \int_0^t dt' e^{\gamma t'} s_i(t')$. The average value of particle position $x(t)$ in both potentials is $\langle x(t)\rangle = [x_0 + \gamma g_i(t)]e^{-\gamma t}$, and the variance is $\langle[x(t) - \langle x(t)\rangle]^2\rangle = \sigma^2(1 - e^{-2\gamma t})$.

Next, we calculate how the two probability densities $P_1(x,t)$ and $P_2(x,t)$ diverge as time progresses. For this we find a continuous version of the Chernoff α-coefficient given by Eq. (1), with $p(x,t) = P_1(x,t)$ and $q(x,t) = P_2(x,t)$, i.e., $C_\alpha(p||q) = \int dx\, p(x,t)[\frac{p(x,t)}{q(x,t)}]^{\alpha-1}$. We obtain

$$\begin{aligned} C_\alpha(p||q) = \exp\Bigg[&\frac{\alpha(\alpha-1)\gamma^2 e^{-2\gamma t}}{2\sigma^2(1 - e^{-2\gamma t})} \\ &\times\left(\int_0^t dt'\, e^{\gamma t'}[s_1(t') - s_2(t')]\right)^2\Bigg], \end{aligned} \tag{62}$$

from which we can determine the Tsallis and Renyi divergences for the Brownian particle. For example, the Renyi

$R_\alpha(p||q)$ divergence takes the form

$$R_\alpha(p||q) = \frac{\alpha\gamma^2 e^{-2\gamma t}\left(\int_0^t dt'\, e^{\gamma t'}[s_1(t') - s_2(t')]\right)^2}{2\sigma^2(1 - e^{-2\gamma t})}, \quad (63)$$

which means that $R_\alpha(p||q)$ measures the differences between the external signals $s_1(t)$ and $s_2(t)$ accumulated over time, appropriately weighted.

The rate at which the distributions $p(x,t)$ and $q(x,t)$ diverge is given by the derivative dC_α/dt and reads

$$\frac{dC_\alpha}{dt} = \frac{\alpha(\alpha-1)\gamma^2 e^{-\gamma t}}{\sigma^2(1 - e^{-2\gamma t})}\left(-\frac{\gamma e^{-\gamma t}}{(1 - e^{-2\gamma t})}[g_2(t) - g_1(t)]^2 + [s_2(t) - s_1(t)][g_2(t) - g_1(t)]\right). \quad (64)$$

Because dC_α/dt is a quadratic function of $[g_2(t) - g_1(t)]$, it has a maximum proportional to the square of the difference between the signals $s_1(t)$ and $s_2(t)$ (for $\alpha > 1$). More formally,

$$\frac{dC_\alpha}{dt} \leqslant \frac{\alpha(\alpha-1)\gamma C_\alpha}{4\sigma^2}[s_2(t) - s_1(t)]^2, \quad (65)$$

which leads to the maximal rate of Renyi divergence between the actual particle distribution and its target distribution,

$$\frac{dR_\alpha}{dt} \leqslant \frac{\alpha\gamma}{4\sigma^2}[s_2(t) - s_1(t)]^2. \quad (66)$$

To assess the bounds on the absolute value $|dC_\alpha/dt|$ and $|dR_\alpha/dt|$, we need to find the expressions for $C_{2\alpha} - C_\alpha^2$ and $C_{2\alpha-1} - C_\alpha^2$, as well as for temporal Fisher information. Using Eq. (62), it can be easily found that

$$C_{2\alpha} - C_\alpha^2 = C_\alpha^2\left(C_\alpha^{\frac{2\alpha}{(\alpha-1)}} - 1\right),$$
$$C_{2\alpha-1} - C_\alpha^2 = C_\alpha^2\left(C_\alpha^{\frac{2(\alpha-1)}{\alpha}} - 1\right). \quad (67)$$

The Fisher information $I_F(p)$ is given by

$$I_F(p) = \frac{\gamma^2}{(1 - e^{-2\gamma t})}\left(\frac{2e^{-2\gamma t}}{(e^{2\gamma t} - 1)} + \frac{[\langle x(t)\rangle - s_1(t)]^2}{\sigma^2}\right), \quad (68)$$

and similarly for $I_F(q)$. Equation (68) indicates that temporal Fisher information in this case is proportional (after a transient time) to the square of the discrepancy between the external signal and the average particle position, which tries to track it.

We can also compute the entropy production rate for this system. It is computed using the continuous formula [6]

$$\dot{S}_i = \int dx\, \frac{J_i(x,t)^2}{\gamma\sigma^2 P_i(x,t)}. \quad (69)$$

As a result, we obtain for entropy production a similar formula to the one for the temporal Fisher information. In particular, for the distribution $p(x,t)$ we have $\dot{S}_p$ as

$$\dot{S}_p = \gamma\left(\frac{e^{-2\gamma t}}{(e^{2\gamma t} - 1)} + \frac{[\langle x(t)\rangle - s_1(t)]^2}{\sigma^2}\right). \quad (70)$$

This formula indicates that the higher the discrepancy between the external signal and the average particle position, the larger is the entropy production rate.

Next, we address a question: how does the entropy production rate relate to the power dissipated in the particle system? A more general equation of motion for our particle, if it had mass m, is given by

$$m\ddot{x} = -\frac{\partial V_i(x,t)}{\partial x} - k\dot{x} + k\sqrt{2\gamma\sigma^2}\eta(t). \quad (71)$$

Both sides of this equation can be multiplied by the particle velocity $\dot{x}$, which after a simple rearrangement yields

$$\frac{dE_i}{dt} = \frac{\partial V_i(x,t)}{\partial t} - k\dot{x}^2 + k\sqrt{2\gamma\sigma^2}\dot{x}\eta(t), \quad (72)$$

where $E_i = \frac{1}{2}m\dot{x}^2 + V_i(x,t)$ is the mechanical energy of the system, and we used the relation $\frac{dV_i(x,t)}{dt} = \frac{\partial V_i(x,t)}{\partial t} + \frac{\partial V_i(x,t)}{\partial x}\dot{x}$. Equation (72) means that energy in the system is dissipated by three different factors: by a temporal decrease of the potential V_i, friction proportional to $k\dot{x}^2$, and noise proportional to $\dot{x}\eta(t)$. The average dissipated power, or energy rate, is $\langle\frac{dE_i}{dt}\rangle$, which yields

$$\left\langle\frac{dE_i}{dt}\right\rangle = -k\gamma^2([\langle x(t)\rangle - s_i(t)]^2 + \dot{s}_i[\langle x(t)\rangle - s_i(t)] - \sigma^2 e^{-2\gamma t}), \quad (73)$$

where we used the Novikov theorem [51] for the average $\langle\dot{x}\eta(t)\rangle = \frac{1}{2}\sqrt{2\gamma\sigma^2}$. Equation (73) shows that the effect of noise on the energy change is negligible after a transient time $\sim 1/\gamma$. The dominant contributions to $\langle\frac{dE_i}{dt}\rangle$ are proportional to the discrepancy between the external signal and the average particle position, which is similar to but not exactly the same as that in the formula for the entropy production rate [Eq. (70)]. The main difference between $\langle\frac{dE_i}{dt}\rangle$ and $\dot{S}_p$ is the term proportional to $\dot{s}_i$ (of any sign), which characterizes energy flux (either positive or negative) between the system and the environment. Finally, the form of Eq. (73) implies that the average energy rate is bounded from above by $\langle\frac{dE_i}{dt}\rangle \leqslant k\gamma^2[\sigma^2 e^{-2\gamma t} + \frac{1}{4}(\dot{s}_i)^2]$.

3. Memory bistable systems

Here, we analyze a memory switch that can be driven by time-dependent external factors, and it is motivated by biophysics of small molecules. We assume that this switch is a two-state system, described by Markov dynamics. Examples of such bistable memory switches are proteins (their activation and deactivation) and synapses in the brain [52–54].

Let us consider a two-state system with energies, respectively, E_1 and E_2 ($E_2 > E_1$), corresponding to states 1 and 2, that can be driven by a time-dependent chemical potential. The system is at temperature T, which plays the role of the noise, and there are stochastic jumps between states 1 and 2. We assume that the system is initially ($t = 0$) at equilibrium, and then a chemical potential $\mu(t)$ is turned on [$\mu(t) > 0$]. This enables the system to jump to the higher-energy state 2, thus acquiring new information above a thermal background (the system learns). Our goal is to determine the rate of this information gain, and its bounds in terms of physical quantities. The master equation corresponding to this situation is

$$\dot{p}_1 = w_{12}p_2 - w_{21}(t)p_1, \quad p_2 = 1 - p_1, \quad (74)$$

where the transition rate from state 2 to 1 is $w_{12} = e^{\beta\Delta E}/\tau$, with τ the timescale for the jumps, and the transition from state 1 to 2 is driven by the chemical potential $\mu(t)$ as $w_{21} = e^{-\beta[\Delta E-\epsilon\mu(t)]}/\tau$, with $\Delta E = E_2 - E_1 > 0$, $\beta = 1/(k_B T)$, and $\epsilon \ll 1$. The role of the chemical potentials is to lower the energy barrier so that the jumps to state 2 are more likely. Initially, $t = 0$, the system is at thermal equilibrium, and we have $p_1(0) = e^{\beta\Delta E}/[2\cosh(\beta\Delta E)]$.

Equation (74) can be solved exactly for an arbitrary form of $\mu(t)$. However, it is convenient to work in the limit of small ϵ to get analytical expressions for divergences and other physical quantities. We work to second order in ϵ, i.e., $p_1 = p_1^{(0)} + \epsilon p_1^{(1)} + \epsilon^2 p_1^{(2)} + O(\epsilon^3)$, and we obtain $p_1^{(0)}(t) = p_1(0)$ and

$$p_1^{(1)}(t) = -\frac{\beta}{\tau}e^{-w_0 t}\bigg[p_1(0)e^{-\beta\Delta E}\int_0^t ds\, \mu(s) + \frac{1}{\tau}\int_0^t ds\, e^{w_0 s}\int_s^t ds'\, \mu(s')\bigg], \tag{75}$$

where $w_0 = (\frac{2}{\tau})\cosh(\beta\Delta E)$. The form of $p_1^{(2)}$ does not appear in the expressions for the divergences and relevant physical quantities to ϵ^2 order, and thus it is not presented here explicitly.

The Chernoff α-coefficient, between nonequilibrium probabilities $p_1(t), p_2(t)$ and their equilibrium values $p_1(0), p_2(0)$, is given by

$$C_\alpha(p(t)||p(0)) = 1 + \epsilon^2\alpha(\alpha-1)\frac{\big(p_1^{(1)}\big)^2}{2p_1^{(0)}p_2^{(0)}} + O(\epsilon^3), \tag{76}$$

which gives us immediately that Tsallis and Renyi divergences are identical in this order:

$$T_\alpha(p(t)||p(0)) = R_\alpha(p(t)||p(0)) = \epsilon^2\frac{\alpha\big(p_1^{(1)}\big)^2}{2p_1^{(0)}p_2^{(0)}} + O(\epsilon^3). \tag{77}$$

They are both proportional to the square of nonequilibrium correction to the occupancy probability.

The rate of Tsallis and Renyi divergences is

$$\dot{T}_\alpha = \dot{R}_\alpha = -\frac{2\epsilon^2\alpha}{\tau}\cosh(\beta\Delta E)\big[e^{2\beta\Delta E}\beta\mu(t)p_1^{(1)} + 4\cosh^2(\beta\Delta E)\big(p_1^{(1)}\big)^2\big] + O(\epsilon^3), \tag{78}$$

and it is nonzero only if chemical potential is present, which corresponds to the detailed balance breaking in the system. When $\dot{T}_\alpha > 0$ (or $\dot{R}_\alpha > 0$), then the system is gaining information, whereas in the opposing case it is losing information.

Similar to the previous example, the rates $\dot{T}_\alpha$ and $\dot{R}_\alpha$ are quadratic in $p_1^{(1)}$, hence both of them are bounded from above by (for $\alpha > 0$)

$$\dot{T}_\alpha = \dot{R}_\alpha \leqslant \frac{\epsilon^2\alpha\beta^2 e^{4\beta\Delta E}\mu(t)^2}{8\tau\cosh(\beta\Delta E)} + O(\epsilon^3). \tag{79}$$

Thus, the speed of divergence from the equilibrium is limited by the square of the chemical potential and Boltzmann factors $e^{\beta\Delta E}$.

Alternatively, and more generally, one can use the bounds in Eqs. (13) and (14), but for that some thermodynamic and information quantities have to be determined first.

Temporal Fisher information is

$$I_F(p) = \frac{\epsilon^2\big[4\cosh^2(\beta\Delta E)p_1^{(1)} + \beta\mu(t)\big]^2}{16\tau^2\cosh^4(\beta\Delta E)} + O(\epsilon^3), \tag{80}$$

which means that the global speed of the system increases when the chemical potential is increasing.

Other physical quantities of interest are entropy production rates, $\dot{S}_p$, and average activity $A_p = \langle\overline{w}\rangle$. We find

$$\dot{S}_p = \frac{\epsilon^2\big[4\cosh^2(\beta\Delta E)p_1^{(1)} + \beta\mu(t)\big]^2}{2\tau\cosh(\beta\Delta E)} + O(\epsilon^3), \tag{81}$$

and for $A_p = w_{21}p_1 + w_{12}p_2$ we have

$$A_p = \frac{1}{\tau\cosh(\beta\Delta E)} + \frac{\epsilon\big[\beta\mu(t) - 2\sinh(2\beta\Delta E)p_1^{(1)}\big]}{2\tau\cosh(\beta\Delta E)} + O(\epsilon^2). \tag{82}$$

Equation (81) suggests that $\dot{S}_p$ grows quadratically with the chemical potential and grows more nonlinearly with the energy barrier ΔE. On the contrary, the average activity A_p decreases with ΔE.

Note that, to the leading order, we have a simple relationship between entropy production rate, Fisher information, and average activity:

$$\dot{S}_p = \frac{8I_F(p)}{\tau^2 A_p^3} + O(\epsilon). \tag{83}$$

This equation suggests that the speed of learning $\sqrt{I_F(p)}$ (acquiring new information) is proportional (with different powers) to the product of entropy production and average activity. Consequently, it seems possible to maintain the speed of learning while simultaneously decreasing dissipation and increasing internal activity.

C. Applications in neuroscience

1. Speed of gaining information during synaptic plasticity

It is believed that long-term information in real neural networks is encoded collectively in the synaptic weights [55–58]. Data from brain cortical networks suggest that synaptic weights are log-normally distributed, characterized by heavy tails [57,58]. Such distributions seem to be relatively stable during human development and adulthood [58].

In what follows, we want to determine the bounds on the speed of gaining information during synaptic learning. We assume that during synaptic plasticity, underlying learning in neural circuits, synaptic weights change their mean and standard deviation values, but they preserve their log-normal distributions. This assumption is consistent with the stability of weight distribution during the lifetime [58].

Let the initial probability density of synaptic weights w (before learning at time $t = 0$) be $\rho_0(w)$, and during the learning phase (at times $t > 0$) let it be $\rho(w, t)$. Thus, we have

$$\rho_0(w) = \frac{\exp\big(-[\ln(w) - m_0]^2/2\sigma_0^2\big)}{\sqrt{2\pi\sigma_0^2}w} \tag{84}$$

and

$$\rho(w,t)=\frac{\exp(-[\ln(w)-m(t)]^2/2\sigma^2(t))}{\sqrt{2\pi\sigma^2(t)}w}, \tag{85}$$

where m_0 and σ_0^2 are the mean and variance of logarithms of synaptic weights at $t=0$, and $m(t)$ and $\sigma^2(t)$ are the corresponding mean and variance of logarithms of synaptic weights for $t>0$.

The Chernoff α-coefficient between $\rho(w,t)$ and $\rho_0(w)$ can be found as

$$C_\alpha(\rho||\rho_0)=\frac{\sigma_0^\alpha\exp\left(\frac{\alpha(\alpha-1)[m(t)-m_0]^2}{2[\alpha\sigma_0^2-(\alpha-1)\sigma^2(t)]}\right)}{\sigma(t)^{\alpha-1}\sqrt{\alpha\sigma_0^2-(\alpha-1)\sigma^2(t)}}, \tag{86}$$

which is valid for $\alpha\sigma_0^2>(\alpha-1)\sigma^2(t)$, and $C_\alpha(\rho||\rho_0)=\infty$ for $\alpha\sigma_0^2\leqslant(\alpha-1)\sigma^2(t)$. This gives us, in the first case, the Renyi divergence

$$R_\alpha(\rho||\rho_0)=\frac{\alpha[m(t)-m_0]^2}{2\big[\alpha\sigma_0^2-(\alpha-1)\sigma^2(t)\big]}+\ln\frac{\sigma_0}{\sigma(t)}-\frac{\ln\big[1+(\alpha-1)\big[1-\sigma^2(t)/\sigma_0^2\big]\big]}{2(\alpha-1)}, \tag{87}$$

and the KL divergence

$$D_{\mathrm{KL}}(\rho||\rho_0)=\frac{[m(t)-m_0]^2}{2\sigma_0^2}+\ln\frac{\sigma_0}{\sigma(t)}-\frac{1}{2}\left[1-\frac{\sigma^2(t)}{\sigma_0^2}\right]. \tag{88}$$

The KL divergence is the standard information gain during synaptic plasticity [10,59], while the Renyi divergence is its generalization. Their rates yield the speeds of gaining information. The rate of R_α is somewhat complicated, but the rate of KL takes a simple form

$$\frac{dD_{KL}}{dt}=\frac{[m(t)-m_0]\dot{m}(t)}{\sigma_0^2}+\frac{\dot{\sigma}(t)}{\sigma(t)}\left[\frac{\sigma^2(t)}{\sigma_0^2}-1\right]. \tag{89}$$

The absolute values of both rates are bounded by the inequalities in Eqs. (14), (15), and (17), with the temporal Fisher information

$$I_F(\rho)=\frac{2\dot{\sigma}(t)^2+\dot{m}(t)^2}{\sigma^2(t)}. \tag{90}$$

This means that the speeds of gaining information during synaptic plasticity, while learning, are limited mostly by the speeds of changing the two parameters characterizing means and variances of synaptic weights.

An interesting question is, when is the bound in Eq. (15) for the plastic synapses saturated? To answer this question, we have to first determine $\langle\ln^2\frac{\rho(w,t)}{\rho_0(w)}\rangle_\rho$. It can be easily found as

$$\left\langle\ln^2\frac{\rho(w,t)}{\rho_0(w)}\right\rangle_\rho=D_{\mathrm{KL}}(\rho||\rho_0)^2+\frac{\big[\sigma^2(t)-\sigma_0^2\big]^2}{2\sigma_0^4}+\frac{\sigma^2(t)}{\sigma_0^4}[m(t)-m_0]^2. \tag{91}$$

With this, all the terms in Eq. (15) are given explicitly. After some arrangements, we find that the general inequality in Eq. (15) is equivalent to the following specific inequality:

$$0\leqslant\big\{2\sigma(t)\dot{\sigma}(t)[m(t)-m_0]-\dot{m}(t)\big[\sigma^2(t)-\sigma_0^2\big]\big\}^2. \tag{92}$$

This implies that the bound in Eq. (15) is saturated if the right-hand side of Eq. (92) is 0, which is satisfied when the ratio $[m(t)-m_0]/[\sigma^2(t)-\sigma_0^2]=\mathrm{const}$. In other words, the mean and variance of the logarithm of synaptic weights must change in a coordinated manner, which is a very restrictive condition on the stochastic dynamics of synapses.

2. *Predictive inference*

The bounds presented in Eqs. (12)–(15) could also be used in predicting the future behavior of a stochastic dynamical system. In particular, the brain neural networks have to often make predictions about some external signal, which is somehow important for the organism possessing that brain [8,60–63]. Let the external time-dependent sensory signal be x_t. Neurons in the sensory cortex of the brain try to predict the value of the signal $x_{t'}$ in future times $t'>t$ [60,62,63], using some internal dynamical variable m_t that relates to neural and synaptic activities. One can think about m_t as some sort of “memory” variable, which keeps the information about the past of the signal x_t up to the time t, in a compressed manner. In the simplest situation, neural activity m_t tries to predict the signal at the nearest future, i.e., to estimate the value $x_{t+\Delta t}$, where Δt is small. The key in this estimate is two conditional probabilities: $p(x_{t+\Delta t}|x_t)$ and $p(x_{t+\Delta t}|m_t)$. The former is the probability of the jump in the signal value from time t to time $t+\Delta t$, and the latter is the probability of the signal at time $t+\Delta t$ given the value m_t of the memory variable at time t. Thus, the first conditional probability describes a natural temporal evolution of the external signal, whereas the second is the estimate of this evolution given the knowledge of neural activity m_t.

The goodness of predictability can be quantified by KL divergence between actual external dynamics $p(x_{t+\Delta t}|x_t)$ and its prediction $p(x_{t+\Delta t}|m_t)$, i.e. [63],

$$D_{\mathrm{KL}}[p(x_{t+\Delta t}|x_t)||p(x_{t+\Delta t}|m_t)]=\int dx_{t+\Delta t}\;p(x_{t+\Delta t}|x_t)\ln\frac{p(x_{t+\Delta t}|x_t)}{p(x_{t+\Delta t}|m_t)}. \tag{93}$$

The smaller the value of D_{KL}, the better the memory variable predicts the external dynamics. The rate of D_{KL} measures how fast the prediction can deteriorate. In this sense, Eqs. (15) and (17) provide bounds on the speed of predictability degradation. These bounds are determined to a large extent by the temporal Fisher information: $I_F(p(x_{t+\Delta t}|x_t))$, which gives the square of the speed of transitions in the external signal, and $I_F(p(x_{t+\Delta t}|m_t))$, which yields the speed of external dynamics given the instantaneous value of the memory variable.

Below we analyze a specific example of predictive inference, in which one can obtain explicit formulas for all relevant variables appearing in Eqs. (15) and (17). Let us consider the following dynamics for x_t and m_t:

$$\dot{x}=-\gamma(x-\lambda_t)+\sqrt{2\gamma\sigma^2}\eta(t), \tag{94}$$

$$\dot{m}=-\gamma_m(m-x)+\sqrt{2\gamma_m\sigma_m^2}\eta_m(t), \tag{95}$$

where λ_t is some external dynamical variable governing the trajectory x_t, γ and γ_m are inverses of time constants for the external and internal systems, and σ and σ_m are standard deviations of the Gaussian noise terms η and η_m, as in Eq. (58). For simplicity, let us consider the case $\gamma_m/\gamma \gg 1$, which corresponds to the situation in which the dynamics of internal variable m_t is much faster than the external x_t. The discretized version of Eqs. (94) and (95) in this limit has the following forms:

$$x_{t+\Delta t} = x_t - \gamma(x_t - \lambda_t)\Delta t + \sqrt{2\gamma\sigma^2}\eta_t\Delta t, \tag{96}$$

$$x_t \approx m_t - \sqrt{\frac{2\sigma_m^2}{\gamma_m}}\eta_{m,t}. \tag{97}$$

Note that the memory variable m_t differs in this limit from the actual external variable x_t only by an appropriately rescaled noise term.

The relevant conditional probabilities can be found from Eqs. (96) and (97) as [16]

$$p(x_{t+\Delta t}|x_t) = \frac{\exp\left(-\frac{[x_{t+\Delta t}-x_t+\gamma(x_t-\lambda_t)\Delta t]^2}{4\gamma\sigma^2\Delta t}\right)}{\sqrt{4\pi\gamma\sigma^2\Delta t}}, \tag{98}$$

$$p(x_t|m_t) = \sqrt{\frac{\gamma_m\Delta t}{4\pi\sigma_m^2}}\exp\left(-\frac{\gamma_m\Delta t}{4\sigma_m^2}[x_t - m_t]^2\right), \tag{99}$$

where Eq. (98) is valid for a small time interval Δt. The remaining conditional probability $p(x_{t+\Delta t}|m_t)$ of interest is found from the relation

$$p(x_{t+\Delta t}|m_t) = \int dx_t\; p(x_{t+\Delta t}|x_t)p(x_t|m_t), \tag{100}$$

which after a straightforward calculation yields

$$\begin{aligned} p(x_{t+\Delta t}|m_t) &= \sqrt{\frac{\gamma_m\Delta t}{4\pi[\gamma\gamma_m\sigma^2(\Delta t)^2 + \sigma_m^2]}} \\ &\times \exp\left(-\frac{\gamma_m\Delta t[x_{t+\Delta t} - m_t + \gamma(x_t - \lambda_t)\Delta t]^2}{4[\gamma\gamma_m\sigma^2(\Delta t)^2 + \sigma_m^2]}\right). \end{aligned} \tag{101}$$

Note that $p(x_{t+\Delta t}|m_t)$ and $p(x_{t+\Delta t}|x_t)$ become identical, with substitution $m_t \leftrightarrow x_t$, if $\gamma_m \mapsto \infty$.

Equations (93), (98), and (101) allow us to find the KL divergence $D_{\mathrm{KL}}[p(x_{t+\Delta t}|x_t)||p(x_{t+\Delta t}|m_t)]$ as

$$\begin{aligned} &D_{\mathrm{KL}}[p(x_{t+\Delta t}|x_t)||p(x_{t+\Delta t}|m_t)] \\ &= \frac{1}{2}\ln\left(1 + \frac{\sigma_m^2}{\gamma\gamma_m\sigma^2(\Delta t)^2}\right) + \frac{\gamma_m\Delta t(x_t - m_t)^2 - 2\sigma_m^2}{4[\gamma\gamma_m\sigma^2(\Delta t)^2 + \sigma_m^2]}, \end{aligned} \tag{102}$$

and its temporal derivative as

$$\frac{dD_{\mathrm{KL}}}{dt} = \frac{\gamma_m\Delta t(x_t - m_t)\dot{x}_t}{2[\gamma\gamma_m\sigma^2(\Delta t)^2 + \sigma_m^2]}. \tag{103}$$

The rate of KL divergence in this case is proportional to the speed of change in the external variable x_t (it does not depend on the speed of m_t because $\dot{m}_t \approx 0$ in the limit $\gamma_m/\gamma \gg 1$), and to the instant difference between x_t and m_t. Consequently, the inference of the external signal improves if the rate of D_{KL} divergence decreases, which takes place when $\dot{x}_t$ and $(x_t - m_t)$ have opposite signs. For example, if the external signal slows down ($\dot{x}_t < 0$), then the prediction internal variable m_t should be smaller than the signal x_t to get closer to it in the next instant of time, i.e., to improve the prediction.

Equation (103) is the exact form of the speed of divergence between the true external dynamics x_t and its estimate using internal variable m_t. Alternatively, one can provide the bounds on dD_{KL}/dt using Eqs. (15) and (17). For this, one needs to determine the temporal Fisher information, defined as

$$I_F(p(x_{t+\Delta t}|x_t)) = \int dx_{t+\Delta t}\; p(x_{t+\Delta t}|x_t)\left[\frac{\dot{p}(x_{t+\Delta t}|x_t)}{p(x_{t+\Delta t}|x_t)}\right]^2 \tag{104}$$

and

$$I_F(p(x_{t+\Delta t}|m_t)) = \int dx_{t+\Delta t}\; p(x_{t+\Delta t}|m_t)\left[\frac{\dot{p}(x_{t+\Delta t}|m_t)}{p(x_{t+\Delta t}|m_t)}\right]^2, \tag{105}$$

which describe the speed of transitions in the external signal and the speed of predictions of that signal, respectively. After some simple algebra, one can find both Fisher informations as

$$I_F(p(x_{t+\Delta t}|x_t)) = \frac{\gamma\Delta t(\dot{x}_t - \dot{\lambda}_t)^2}{2\sigma^2} \tag{106}$$

and

$$I_F(p(x_{t+\Delta t}|m_t)) = \frac{\gamma_m\Delta t(\dot{x}_t)^2}{2[\gamma\gamma_m\sigma^2(\Delta t)^2 + \sigma_m^2]}, \tag{107}$$

and both of them depend on the speed of the external signal $\dot{x}_t$. Note that for $\gamma_m/\gamma \gg 1$ the second Fisher information $I_F(p(x_{t+\Delta t}|m_t))$ dominates over the first one, since generally $\gamma\Delta t \ll 1$. This means that the speed of prediction is much larger than the speed of the external signal, which is consistent with our initial assumption.

VII. BOUNDS ON THE RATE OF MUTUAL INFORMATION

Mutual information $I(x, y)$ between two stochastic variables x, y with joint probability $p_{xy}(x, y)$, and marginal probabilities $p_x(x)$, $p_y(y)$, can be defined as KL divergence between probabilities p and q given by $p = p_{xy}(x, y)$ and $q = p_x(x)p_y(y)$. More precisely, $I(x, y) = D_{\mathrm{KL}}(p_{xy}||p_xp_y)$.

A. General kinematic bounds

The general kinematic bound on the rate of mutual information $I(x, y)$, irrespective of the type of systems dynamics, follows from Eqs. (15) and (17), and takes the form

$$\begin{aligned} \left|\frac{dI(x,y)}{dt}\right| &\leqslant \sqrt{I_F(p_{xy})}\sqrt{\left\langle \ln^2\frac{p_{xy}}{p_xp_y}\right\rangle - I(x,y)^2} \\ &+ \sqrt{I_F(p_x) + I_F(p_y)}\sqrt{\left\langle\frac{p_{xy}}{p_xp_y}\right\rangle - 1} \\ &\leqslant \sqrt{I_F(p_{xy})}\sqrt{C_2^{xy} - 1 - I(x,y)^2} \\ &+ \sqrt{I_F(p_x) + I_F(p_y)}\sqrt{C_2^{xy} - 1}, \end{aligned} \tag{108}$$

where C_2^{xy} is the Chernoff coefficient, i.e., $C_2^{xy} = \langle \frac{p_{xy}}{p_x p_y} \rangle$, and averaging is done with respect to the joint probability p_{xy}. We used the fact that Fisher information of the product of probabilities decomposes into the sum, i.e., $I_F(p_x p_y) = I_F(p_x) + I_F(p_y)$. The upper bound on dI/dt is thus constrained by the global dynamical rates of the whole system x, y and of the subsystems x and y (it is called here bound B_I1). These rates are appropriately rescaled by the degrees of mutual correlations between variables x and y.

B. Examples: Weakly correlated systems

In this section we check the quality of the bound represented by Eq. (108) for two examples of weakly coupled systems X and Y, with continuous state variables x and y, respectively.

1. Bivariate Gaussian distribution

Consider the system (X, Y) of weakly correlated variables x and y with Gaussian joint probability density

$$p_{xy}(x, y) = \frac{\exp(-[\overline{x}, \overline{y}]\Sigma^{-1}[\overline{x}, \overline{y}]^T)}{2\pi\sqrt{\det(\Sigma)}}, \tag{109}$$

where $[\overline{x}, \overline{y}]$ is a two-dimensional vector with the components $\overline{x} = x - \mu_x$, and $\overline{y} = y - \mu_y$, where μ_x and μ_y denote mean values of x and y. The symbol Σ is a 2×2 covariance matrix with elements $\Sigma_{11} = \sigma_x^2$, $\Sigma_{22} = \sigma_y^2$, and $\Sigma_{12} = \Sigma_{21} = \epsilon r_0(t)\sigma_x\sigma_y$, where $\epsilon \ll 1$. We assume that only the correlation coefficient r_0 is time-dependent.

Consequently, marginal distributions $p_x(x) = \frac{\exp(-\overline{x}^2/2\sigma_x^2)}{\sqrt{2\pi\sigma_x^2}}$ and $p_y(y) = \frac{\exp(-\overline{y}^2/2\sigma_y^2)}{\sqrt{2\pi\sigma_y^2}}$ are time-independent, which implies that temporal Fisher information for these distributions $I_F(p_x) = I_F(p_y) = 0$.

For small ϵ the joint density (109) separates into a product of the marginal distributions as

$$p_{xy} = p_x p_y \Bigg(1 + \frac{\epsilon r_0}{\sigma_x\sigma_y}\overline{xy} + \frac{1}{2}\left(\frac{\epsilon r_0}{\sigma_x\sigma_y}\right)^2 \times \left[2(\overline{xy})^2 + (\sigma_x\sigma_y)^2 - \sigma_y^2\overline{x}^2 - \sigma_x^2\overline{y}^2\right] + O(\epsilon^3)\Bigg), \tag{110}$$

which allows us to compute all the variables in Eq. (108) to the lowest order in ϵ.

Thus, mutual information $I(x, y)$ is [10]

$$I(x, y) = -\tfrac{1}{2}\ln(1 - \epsilon^2 r_0(t)^2) = \tfrac{1}{2}\epsilon^2 r_0(t)^2 + O(\epsilon^3), \tag{111}$$

and temporal Fisher information for the joint density in Eq. (109) is

$$I_F(p_{xy}) = \left\langle \left(\frac{\dot{p}_{xy}}{p_x p_y}\right)^2 \right\rangle = \epsilon^2 \dot{r}_0(t)^2 + O(\epsilon^3), \tag{112}$$

where $\dot{r}_0(t) = dr_0/dt$. Moreover, we have

$$\left\langle \ln^2\left(\frac{p_{xy}}{p_x p_y}\right) \right\rangle = \epsilon^2 r_0(t)^2 + O(\epsilon^3). \tag{113}$$

With Eqs. (111)–(113) we can find the left- and right-hand sides of Eq. (108). It turns out that both sides of this equation are equal to each other in the lowest order of ϵ, with the values $\epsilon^2 r_0(t)|\dot{r}_0| + O(\epsilon^3)$ each. This implies that for weakly correlated Gaussian variables the bound on the rate of mutual information in Eq. (108) is saturated if σ_x and σ_y are time-independent. However, the situation is more complicated if the variances of x and y are time-dependent. In this case, much depends on how big are the rates of σ_x and σ_y. If they are small, of the order $\sim\epsilon$, then the bound in Eq. (108) is close to saturation, but if they are large, $\sim O(1)$, then the left-hand side is much smaller (by the factor proportional to ϵ) than the right-hand side. This follows from the observation that in this case all Fisher information, $I_F(p_{xy})$, $I_F(p_x)$, and $I_F(p_x)$, would be of order $O(1)$.

2. Bivariate non-Gaussian distribution

In this case, we choose the joint probability density (normalized) for the weakly correlated system (X, Y) in the form

$$p_{xy}(x, y) = \frac{(\kappa_1\kappa_2)^2[1 + \epsilon r(t)xy]}{\kappa_1\kappa_2 + \epsilon r(t)} e^{-(\kappa_1 x + \kappa_2 y)}, \tag{114}$$

where κ_1, κ_2 are some positive constants, $\epsilon \ll 1$, and $r(t)$ is a time-dependent positive parameter characterizing the degree of correlations between X and Y. Note that for $\epsilon = 0$ the systems X and Y become decoupled. The marginal probability densities p_x, p_y are time-dependent and read

$$p_x(x) = \frac{\kappa_1^2\kappa_2\left[1 + \epsilon r(t)\frac{x}{\kappa_2}\right]}{\kappa_1\kappa_2 + \epsilon r(t)} e^{-\kappa_1 x},$$
$$p_y(y) = \frac{\kappa_1\kappa_2^2\left[1 + \epsilon r(t)\frac{y}{\kappa_1}\right]}{\kappa_1\kappa_2 + \epsilon r(t)} e^{-\kappa_2 y}. \tag{115}$$

As in the Gaussian case, all the relevant variables present in Eq. (108) can be computed analytically for this case as a series expansion in the small parameter ϵ (see Appendix D). Consequently, mutual information $I(x, y)$ to the lowest order in ϵ is

$$I(x, y) = \frac{\epsilon^2}{2}\left(\frac{r(t)}{\kappa_1\kappa_2}\right)^2 + O(\epsilon^3). \tag{116}$$

Temporal Fisher information $I_F(p_{xy})$ and $I_F(p_x), I_F(p_y)$ takes the forms

$$I_F(p_{xy}) = 3\epsilon^2\left(\frac{\dot{r}(t)}{\kappa_1\kappa_2}\right)^2 + O(\epsilon^3),$$
$$I_F(p_x) = I_F(p_y) = \epsilon^2\left(\frac{\dot{r}(t)}{\kappa_1\kappa_2}\right)^2 + O(\epsilon^4), \tag{117}$$

where $\dot{r}(t) = dr/dt$. Additionally, we have

$$\left\langle \ln^2\left(\frac{p_{xy}}{p_x p_y}\right) \right\rangle = \epsilon^2\left(\frac{r(t)}{\kappa_1\kappa_2}\right)^2 + O(\epsilon^3),$$
$$\left\langle \frac{p_{xy}}{p_x p_y} \right\rangle = 1 + \epsilon^2\left(\frac{r(t)}{\kappa_1\kappa_2}\right)^2 + O(\epsilon^3). \tag{118}$$

Insertion of Eqs. (116)–(118) into Eq. (108) gives us the left-hand side, $|dI(x, y)/dt|$, equal to $(\frac{\epsilon}{\kappa_1\kappa_2})^2 r(t)|\dot{r}(t)|$, whereas both right-hand sides are identical to the lowest order and equal to $(\sqrt{2}+\sqrt{3})(\frac{\epsilon}{\kappa_1\kappa_2})^2 r(t)|\dot{r}(t)|$. This means that, for weakly correlated systems, the bound on the rate of mutual information in Eq. (108) is greater than its actual absolute value by a factor of 3.1, which is a worse performance than for the Gaussian case. Thus, comparing the two examples, it is clear that the degree of accuracy in estimation of $|dI(x,y)/dt|$ is model-dependent.

C. Kinematic-thermodynamic bounds for Markov processes

In this section, we provide upper bounds on the rates of mutual information for more restricted dynamics of composite Markov systems with interacting systems X and Y. First we discuss the dynamics of such systems and then give the bounds on the rate of mutual information between X and Y.

1. Master equation for composite systems

For the composite system (X, Y) containing two interacting subsystems X and Y, we can formulate the Master equation similar to Eq. (22),

$$\dot{p}_{xy} = \sum_{x',y'} (\mathbf{W}(xy|x'y')p_{x'y'} - \mathbf{W}(x'y'|xy)p_{xy}), \tag{119}$$

where p_{xy} is the joint probability that the system (X, Y) is in the state (x, y), and $\mathbf{W}(x'y'|xy)$ is the global transition rate from state (x, y) to state (x', y').

The average activity A_{xy} of the whole system (XY) corresponds to the first moment of the total escape rate and is given by [[26,27]; compare Eq. (25)]

$$A_{xy} \equiv \langle \overline{\mathbf{W}} \rangle_{xy} = \sum_{x,y} \overline{\mathbf{W}}(xy)p_{xy}, \tag{120}$$

where the total escape rate from state (xy) is $\overline{\mathbf{W}}(xy) = \sum_{x',y'} \mathbf{W}(x'y'|xy)$.

The second moment of the total escape rate of the composite system is defined as

$$\langle \overline{\mathbf{W}}^2 \rangle_{xy} = \sum_{x,y} \overline{\mathbf{W}}(xy)^2 p_{xy}, \tag{121}$$

and the system entropy production rate $\dot{S}_{xy}$ is

$$\dot{S}_{xy} = \frac{1}{2}\sum_{xy,x'y'} (\mathbf{W}(xy|x'y')p_{x'y'} - \mathbf{W}(x'y'|xy)p_{xy}) \times \ln \frac{\mathbf{W}(xy|x'y')p_{x'y'}}{\mathbf{W}(x'y'|xy)p_{xy}}. \tag{122}$$

In a particular case when the joint probability is separable, i.e., $p_{xy} = p_x p_y$, we denote the above quantities as $A^{(0)}_{xy} \equiv \langle \overline{\mathbf{W}} \rangle^{(0)}_{xy}$ for average activity, $\langle \overline{\mathbf{W}}^2 \rangle^{(0)}_{xy}$ for the second moment of the total escape rate, and $\dot{S}^{(0)}_{xy}$ as the entropy production rate.

For the so-called bipartite systems [64–66], the global transition rates $\mathbf{W}(x'y'|xy)$ in Eq. (119) can be written in a more explicit form as $\mathbf{W}(x'y'|xy) = w^y_{x'x}\delta_{yy'} + w^x_{y'y}\delta_{xx'}$. In this formula, $w^y_{x'x}$, $w^x_{y'y}$ are transition rates in the subsystems X and Y, respectively, which depend on the actual state in the neighboring system. Here, one considers only single transitions in either of the subsystems ($x \mapsto x'$ or $y \mapsto y'$), neglecting double simultaneous transitions, i.e., $(x, y) \mapsto (x', y')$, as they are less likely. With the above decomposition of $\mathbf{W}(x'y'|xy)$ one can write the master equation (119) as

$$\dot{p}_{xy} = \sum_{x'} w^y_{xx'}p_{x'y} + \sum_{y'} w^x_{yy'}p_{xy'} - \left(\sum_{x'} w^y_{x'x} + \sum_{y'} w^x_{y'y}\right)p_{xy}.$$

Moreover, also A_{xy}, $\langle \overline{\mathbf{W}}^2 \rangle_{xy}$, and $\dot{S}_{xy}$ can be expressed in terms of local transition rates $w^y_{x'x}$, $w^x_{y'y}$, which may be useful [64–66] (see also below).

2. The rates of mutual information for Markov processes

For the Markov dynamics represented by the Master equation (119), the thermodynamic-kinematic bound on the rate of mutual information between variables x and y can be deduced from Eq. (40), and reads

$$\left|\frac{dI(x,y)}{dt}\right| \leqslant \left([I_F(p_x) + I_F(p_y)]\sqrt{A^{(0)}_{xy}\Psi^{(0)}_{xy}}\left[C^{xy}_3 - C^{xy}_2\right]\right)^{1/3} + \left(I_F(p_{xy})\sqrt{A_{xy}\Psi_{xy}}\left[e^{-3I/2}C^{xy}_{5/2} - e^{-I/2}C^{xy}_{3/2} - e^{I/2}C^{xy}_{1/2} + e^{3I/2}C^{xy}_{-1/2}\right]\right)^{1/3}, \tag{123}$$

where the Chernoff coefficient $C^{xy}_\alpha = \langle (\frac{p_{xy}}{p_x p_y})^{\alpha-1}\rangle$, and averaging is done with respect to p_{xy}. Additionally, the symbols Ψ_{xy} and $\Psi^{(0)}_{xy}$ are analogs of Eq. (35) with the joint probability p_{xy} and its separable variant $p_{xy} = p_x p_y$, respectively.

An alternative kinematic-thermodynamic bound on the rate of mutual information, coming from Eqs. (45) and (46), is

$$\left|\frac{dI(x,y)}{dt}\right| \leqslant \sqrt{\Psi_{xy}/2}\left(\sqrt{I_F(p_{xy})} + 2\sqrt{\langle \overline{\mathbf{W}}^2 \rangle_{xy}}\right)^{1/2}\left(e^{-2I}C^{xy}_3 - 4e^{-I}C^{xy}_2 - 4e^I + e^{2I}C^{xy}_{-1} + 6\right)^{1/4} + \sqrt{\Psi^{(0)}_{xy}/2}\left(\sqrt{I_F(p_x) + I_F(p_y)} + 2\sqrt{\langle \overline{\mathbf{W}}^2 \rangle^{(0)}_{xy}}\right)^{1/2}\left(C^{xy}_4 - 4C^{xy}_3 + 6C^{xy}_2 - 3\right)^{1/4}. \tag{124}$$

The inequalities (123) and (124), called here bounds B_I2, B_I3 and B_I4, B_I5, respectively, imply that for Markov dynamical physical systems the rates at which we can gain information about one variable (X) by observing another correlated variable (Y) can be restricted in several ways, but the speeds of global and internal dynamics always appear in

TABLE II. Summary of the inequalities for the rates of MI.

Bound type	Equation
Kinematic:	
B_I1	$\lvert\dot{I}_{xy}\rvert \leqslant \sqrt{I_F(p_{xy})}\sqrt{C_2^{xy} - 1 - I_{xy}^2} + \sqrt{I_F(p_x) + I_F(p_y)}\sqrt{C_2^{xy} - 1}$
Kinematic-thermodynamic:	
B_I2, B_I3	$\lvert\dot{I}_{xy}\rvert \leqslant ([I_F(p_x) + I_F(p_y)]\sqrt{A_{xy}^{(0)}\Psi_{xy}^{(0)}}[C_3^{xy} - C_2^{xy}])^{1/3}$ $+(I_F(p_{xy})\sqrt{A_{xy}\Psi_{xy}}[e^{-3I_{xy}/2}C_{5/2}^{xy} - e^{-I_{xy}/2}C_{3/2}^{xy} - e^{I_{xy}/2}C_{1/2}^{xy} + e^{3I_{xy}/2}C_{-1/2}^{xy}])^{1/3}$
B_I4, B_I5	$\lvert\dot{I}_{xy}\rvert \leqslant \sqrt{\Psi_{xy}^{(0)}/2}(\sqrt{I_F(p_x) + I_F(p_y)} + 2\sqrt{\langle\overline{\mathbf{W}}^2\rangle_{xy}^{(0)}})^{1/2}(C_4^{xy} - 4C_3^{xy} + 6C_2^{xy} - 3)^{1/4}$ $+\sqrt{\Psi_{xy}/2}(\sqrt{I_F(p_{xy})} + 2\sqrt{\langle\overline{\mathbf{W}}^2\rangle_{xy}})^{1/2}(e^{-2I_{xy}}C_3^{xy} - 4e^{-I_{xy}}C_2^{xy} - 4e^{I_{xy}} + e^{2I_{xy}}C_{-1}^{xy} + 6)^{1/4}$

those limitations. This suggests that the speeds of information transfer between two systems are limited primarily by global and local speeds of system transformations. Since the speed of local dynamics is related to activity A and the entropy production rate, this also means that the rate of MI is constrained by thermodynamics, which is in line with previous conclusions [67]. Moreover, out of the three bounds on $dI(x,y)/dt$ represented by Eqs. (108), (123), and (124), the most accurate is the bound given by Eq. (108). This follows from the fact that this bound is the most general, as it is independent of the type of system dynamics (either Markovian or non-Markovian). This is also supported by the results in Figs. 1 and 2, where the bound B1 [corresponding to Eq. (108)] is very close to the actual value of $|dC_\alpha/dt|$.

It is important to note that the bounds on the rate of MI given by Eqs. (108), (123), and (124) are a new type of bounds, and they are different from other existing bounds on mutual information or its rate [67–69]. The inequalities derived here for the rate of MI are collected in Table II.

VIII. APPLICATION OF INEQUALITIES FOR THE RATES OF MI: BIPARTITE SENSING AND LANDAUER LIMIT

For bipartite systems, with X describing an external variable and Y an internal variable, one can decompose the rate of mutual information, $\dot{I}_{xy} \equiv dI(x,y)/dt$, between X and Y into the so-called information flows $\dot{I}_x$ and $\dot{I}_y$ as [66,70]

$$\dot{I}_{xy} = \dot{I}_x + \dot{I}_y, \tag{125}$$

where $\dot{I}_x = [I(x_{t+dt}, y_t) - I(x_t, y_t)]/dt$ and $\dot{I}_y = [I(x_t, y_{t+dt}) - I(x_t, y_t)]/dt$ with $dt \mapsto 0$. The rate $\dot{I}_x$ can be interpreted as the rate of change of mutual information between the two subsystems that is due only to the dynamics of X, and similarly for $\dot{I}_y$. One can also represent both $\dot{I}_x$ and $\dot{I}_y$ by the specific transition rates in both subsystems [66]:

$$\dot{I}_x = \sum_{x>x',y} \left(w_{xx'}^y p_{x'y} - w_{x'x}^y p_{xy}\right) \ln \frac{p(y|x)}{p(y|x')} \tag{126}$$

and

$$\dot{I}_y = \sum_{y>y',x} \left(w_{yy'}^x p_{xy'} - w_{y'y}^x p_{xy}\right) \ln \frac{p(x|y)}{p(x|y')}. \tag{127}$$

In the case of sensing the external variable X by the internal variable Y, and with no feedback from Y to X, the information flows $\dot{I}_x$ and $\dot{I}_y$ are also referred to in the literature as the nonpredictive information rate (or nostalgia rate) and the learning rate, respectively [8,67]. One can think about the variable Y as a molecular sensor or as activity of a neural network learning the dynamics of the external variable X. Thus the rate of information $\dot{I}_{xy}$ between internal and external dynamics is the sum of the learning rate ($\dot{I}_y$ dynamics of Y about external signal X) and the rate of nonpredictive information ($\dot{I}_x$). Using the general bounds derived in Sec. VII on the rate of mutual information, we can find the bounds on the learning rate in terms of the nostalgia rate. Since $|\dot{I}_x + \dot{I}_y| = |\dot{I}_{xy}| \leqslant B_I$, where B_I is any of the three bounds on the rate of mutual information [Eqs. (108), (123), and (124)], we obtain the following general bounds on the learning rate $\dot{I}_y$:

$$-B_I - \dot{I}_x \leqslant \dot{I}_y \leqslant B_I - \dot{I}_x. \tag{128}$$

Thus, the learning rate $\dot{I}_y$ about the external signal is bounded from below and above by the bounds involving temporal Fisher information and other kinematic and thermodynamic characteristics.

On the other hand, the nostalgia rate $\dot{I}_x$ in this case can be bounded as [8,71]

$$0 \leqslant -\dot{I}_x \leqslant \dot{S}(Y|X) - \frac{\dot{Q}_y}{k_B T}, \tag{129}$$

where the conditional entropy rate $\dot{S}(Y|X) = \frac{-d}{dt}\sum_{x,y} p_{xy} \ln p(y|x)$, and $\dot{Q}_y$ is the heat flow between the internal system Y and the thermal environment where $\dot{Q}_y = -k_B T \sum_{y>y',x} (w_{yy'}^x p_{xy'} - w_{y'y}^x p_{xy}) \ln \frac{w_{yy'}^x}{w_{y'y}^x}$.

Combining Eqs. (128) and (129), we obtain the limits on the learning rate $\dot{I}_y$ in terms of the thermodynamic quantities and the derived bounds B_I,

$$-B_I \leqslant \dot{I}_y \leqslant B_I + \dot{S}(Y|X) - \frac{\dot{Q}_y}{k_B T}. \tag{130}$$

This inequality implies that the maximal learning rate of the internal system about the external variable is limited by the internal system entropy production [which is $\dot{S}(Y|X) - \frac{\dot{Q}_y}{k_B T}$] and the upper bound on the rate of MI between the two subsystems.

Equation (130) can also be used in the context of information erasure and corresponding heat generation [8,72]. The right-hand side of Eq. (130) can be equivalently written as

$$-\frac{\dot{Q}_y}{k_B T} \geqslant \dot{I}_{\rm er} + (\dot{I}_y - B_I), \qquad (131)$$

where $\dot{I}_{\rm er} = -\dot{S}(Y|X)$ and it can be interpreted as the rate of Landauer erasure of the learned information about X, which is exactly the rate of decrease of the conditional entropy of Y about X [63]. Inequality (131) is another lower bound on the heat generation during continuous erasure of memory, though not in terms of nostalgia [as in Eq. (129)] but using the learning rate. However, it should be noted that the bound (131) is less sharp than the bound (129).

IX. CONCLUSIONS

In this work, two types of upper bounds on the rates of statistical divergences were obtained. The first type of bound, with two different inequalities [bounds B1 and B2 represented by Eqs. (12) and (18)], is very general, independent of the type of system dynamics, and it relates to the system's global speed via temporal Fisher information. The second type, with four different inequalities [bounds B3–B6 represented by Eqs. (31), (32), and (44)], is less general, applying only to Markov systems. The second type of limitation involves either purely kinematic variables (speeds of global dynamics and average activities) or a mixture of kinematic and thermodynamic variables, with the presence of entropy production rate characterizing dissipation in the system. Generally, the first type of limitation is tighter than the second, as shown by the numerical example with the one-step Markov process. The restrictions on the rates of divergences were also used to derive general bounds on the rates of mutual information between two stochastic variables, with arbitrary time varying probability distributions. Since statistical divergences can be thought of as generalized information gains [10], the present work suggests a link with information thermodynamics [21,22], and it is also related to recent applications of majorization in thermodynamics [73]. In particular, this work as well as [73] both provide a complementary view of out-of-equilibrium thermodynamic systems in a coherent manner in terms of information-theoretic and physical variables. Additionally, the ideas presented here may open new avenues in interdisciplinary research, e.g., by connecting some areas in computer science, or general computing either electronic or biological, to the physics of operating algorithms (e.g., [8,74]). Moreover, the derived bounds on the rates of mutual information might be useful estimates for information flow in real neurons [75], artificial deep neural networks [76], molecular circuits [77], or other systems where exact values are difficult to obtain [68] and require heavy numerical calculations [67,69]. Finally, it is worth mentioning that there are also other types of limits possible on the rates of statistical divergences, but the goal here was to have bounds that can be related clearly to the known physical observables.

ACKNOWLEDGMENTS

The work was supported by the Polish National Science Centre (NCN) Grant No. 2021/41/B/ST3/04300.

APPENDIX A: WEAKLY TIME-DEPENDENT EXPONENTIAL DISTRIBUTIONS APPLIED TO THE BOUND IN EQ. (12).

The α-coefficient and its rate can be computed exactly for the time-dependent exponential distributions. In particular, taking $p_x(x) = \nu_1 e^{-\nu_1 x}$ and $q_x(x) = \nu_2 e^{-\nu_2 x}$, with $\nu_1(t) = \nu + \epsilon[\Delta + r_1(t)]$ and $\nu_2(t) = \nu + \epsilon r_2(t)$, and the small parameter $\epsilon \ll 1$, yields

$$C_\alpha = \frac{(\nu_1/\nu_2)^\alpha}{1 + \alpha[(\nu_1/\nu_2) - 1]} = 1 + \frac{\alpha(\alpha-1)}{2}(\epsilon/\nu)^2(r_1 - r_2 + \Delta)^2 + O(\epsilon^3) \qquad \text{(A1)}$$

and

$$\frac{dC_\alpha}{dt} = \alpha(\alpha-1)(\epsilon/\nu)^2(r_1 - r_2 + \Delta)(\dot{r}_1 - \dot{r}_2) + O(\epsilon^3). \qquad \text{(A2)}$$

Note that C_α is finite provided $1 + \alpha[(\nu_1/\nu_2) - 1] > 0$. Otherwise it is infinite.

Temporal Fisher information is given by

$$I_F(p_x) = \left(\frac{\epsilon \dot{r}_1}{\nu}\right)^2 + O(\epsilon^3),$$
$$I_F(q_x) = \left(\frac{\epsilon \dot{r}_2}{\nu}\right)^2 + O(\epsilon^3). \qquad \text{(A3)}$$

Equations (A1)–(A3) allow us to find both sides in Eq. (12) in this case to the lowest order, as discussed in the main text.

APPENDIX B: UPPER LIMIT ON $\langle|\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|^z\rangle_p$

In this Appendix we show how to calculate the bound on $\langle|\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|^z\rangle_p$, where the exponent $z = 0$ or $z = 1$.

We can generally write

$$\langle|\dot{p}/p||(p/q)^{\alpha-1} - C_\alpha|^z\rangle_p = \sum_n |\dot{p}_n||(p/q)^{\alpha-1} - C_\alpha|^z \leqslant \sum_{nk} |w_{nk}p_k - w_{kn}p_n||(p/q)^{\alpha-1} - C_\alpha|^z, \qquad \text{(B1)}$$

where we used the master equation in Eq. (22) and the familiar relation $|x + y| \leqslant |x| + |y|$. Below we restrict the last line in Eq. (B1) in two different ways: one corresponding to a kinematic bound, and the second to a mixed kinematic-thermodynamic bound.

1. Kinematic bound

Let us decompose the last line in Eq. (B1) as

$$\begin{aligned}\sum_{nk}|w_{nk}p_k-w_{kn}p_n||(p/q)^{\alpha-1}-C_\alpha|^z\\ &=\sum_{nk}\sqrt{|w_{nk}p_k-w_{kn}p_n|}\sqrt{|w_{nk}p_k-w_{kn}p_n|[(p/q)^{\alpha-1}-C_\alpha]^{2z}}\\ &\leqslant\sum_{nk}\sqrt{(w_{nk}p_k+w_{kn}p_n)}\sqrt{(w_{nk}p_k+w_{kn}p_n)[(p/q)^{\alpha-1}-C_\alpha]^{2z}}\leqslant\sqrt{\sum_{nk}(w_{nk}p_k+w_{kn}p_n)}\\ &\quad\times\sqrt{\sum_{nk}(w_{nk}p_k+w_{kn}p_n)[(p/q)^{\alpha-1}-C_\alpha]^{2z}}=\sqrt{2A_p}\sqrt{\sum_n(\dot p_n+2\overline{w}_np_n)[(p/q)^{\alpha-1}-C_\alpha]^{2z}},\end{aligned}\tag{B2}$$

where we used first the known relation $|w_{nk}p_k-w_{kn}p_n|\leqslant(w_{nk}p_k+w_{kn}p_n)$, and then the Cauchy-Schwarz inequality. In the last line, we used the master equation (22) for the second factor. Additionally we introduced two quantities A_p and $\overline{w}_n$, which are, respectively, the average activity and the total escape rate from state n. The average activity A_p is defined as

$$A_p=\frac{1}{2}\sum_{nk}(w_{nk}p_k+w_{kn}p_n)\equiv\sum_n\overline{w}_np_n,\tag{B3}$$

from which we also have that the total escape rate from state n is $\overline{w}_n=\sum_k w_{kn}$.

For $z=0$ the last line in Eq. (B2) simplifies to $2A_p$, since $\sum_n\dot p_n=0$. This means that

$$\sum_{nk}|w_{nk}p_k-w_{kn}p_n|\leqslant 2A_p,\tag{B4}$$

which in combination with Eq. (B1) leads to Eq. (24) in the main text.

For $z=1$ we decompose the last factor in the last line of Eq. (B2) into the sum of two terms: one involving $\dot p$ and the second $\overline{w}_n$, and then we use the Cauchy-Schwarz inequality in both terms. Applying that procedure yields

$$\begin{aligned}\sum_n(\dot p_n+2\overline{w}_np_n)[(p_n/q_n)^{\alpha-1}-C_\alpha]^2&=\langle(\dot p/p)[(p/q)^{\alpha-1}-C_\alpha]^2\rangle_p+2\langle\overline{w}[(p/q)^{\alpha-1}-C_\alpha]^2\rangle_p,\\ &\leqslant(\sqrt{I_F(p)}+2\sqrt{\langle\overline{w}^2\rangle_p})\sqrt{\langle[(p/q)^{\alpha-1}-C_\alpha]^4\rangle_p},\end{aligned}\tag{B5}$$

where $\langle\overline{w}^2\rangle_p=\sum_n\overline{w}_n^2p_n$. This means that

$$\sum_{nk}|w_{nk}p_k-w_{kn}p_n||(p/q)^{\alpha-1}-C_\alpha|\leqslant\sqrt{2A_p}(\sqrt{I_F(p)}+2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}\langle[(p/q)^{\alpha-1}-C_\alpha]^4\rangle_p^{1/4},\tag{B6}$$

which corresponds to Eq. (42) in the main text.

2. Kinematic-thermodynamic bound

Now we decompose the last line in Eq. (B1) in a different way, which will allow us to introduce also entropy production rate. We have the following sequence of inequalities:

$$\begin{aligned}\sum_{nk}|w_{nk}p_k-w_{kn}p_n||(p_n/q_n)^{\alpha-1}-C_\alpha|^z&=\sum_{nk}\sqrt{(w_{nk}p_k-w_{kn}p_n)\ln\frac{w_{nk}p_k}{w_{kn}p_n}}\sqrt{\frac{(w_{nk}p_k-w_{kn}p_n)}{\ln\frac{w_{nk}p_k}{w_{kn}p_n}}}\\ \times|(p_n/q_n)^{\alpha-1}-C_\alpha|^z&\leqslant\sqrt{\sum_{nk}(w_{nk}p_k-w_{kn}p_n)\ln\frac{w_{nk}p_k}{w_{kn}p_n}}\sqrt{\sum_{nk}\frac{(w_{nk}p_k-w_{kn}p_n)}{\ln(w_{nk}p_k/w_{kn}p_n)}[(p_n/q_n)^{\alpha-1}-C_\alpha]^{2z}},\end{aligned}\tag{B7}$$

where we used in the second line the Cauchy-Schwarz inequality. The first factor in the last line of Eq. (B7) is the coarse-grained entropy production rate $\dot S_p$, i.e. [7,34,36],

$$\dot S_p=\frac{1}{2}\sum_{nk}(w_{nk}p_k-w_{kn}p_n)\ln\frac{w_{nk}p_k}{w_{kn}p_n}.\tag{B8}$$

It is worth mentioning that coarse-grained entropy production $\dot S_p$ satisfies inequality $\dot S_p\leqslant\frac{1}{2}\sum_s\sum_{nk}(w^{(s)}_{nk}p_k-w^{(s)}_{kn}p_n)\ln(w^{(s)}_{nk}p_k/w^{(s)}_{kn}p_n)$, which means that $\dot S_p$ is the lower bound on the true physical entropy production caused by distinct microscopic processes [7]. However, in our case this is a sufficient estimation.

The second factor on the right in the last line of Eq. (B7) can be bounded by the logarithmic mean, as in Eq. (5):

$$\frac{(w_{nk}p_k - w_{kn}p_n)}{\ln(w_{nk}p_k) - \ln(w_{kn}p_n)} \leqslant \frac{1}{2}(w_{nk}p_k + w_{kn}p_n). \tag{B9}$$

Combining Eqs. (B7)–(B9), we obtain the following inequality:

$$\begin{aligned}\sum_{nk}|w_{nk}p_k - w_{kn}p_n||(p_n/q_n)^{\alpha-1} - C_\alpha|^z &\leqslant \sqrt{2\dot{S}_p}\sqrt{\sum_{nk}\frac{1}{2}(w_{nk}p_k + w_{kn}p_n)[(p_n/q_n)^{\alpha-1} - C_\alpha]^{2z}}\\ &= \sqrt{\dot{S}_p}\sqrt{\sum_n(\dot{p}_n + 2\overline{w}_n p_n)[(p_n/q_n)^{\alpha-1} - C_\alpha]^{2z}}.\end{aligned} \tag{B10}$$

For $z = 0$ this inequality implies

$$\sum_{nk}|w_{nk}p_k - w_{kn}p_n| \leqslant \sqrt{2\dot{S}_p}\sqrt{A_p}, \tag{B11}$$

which in combination with Eq. (B1) produces Eq. (26) in the main text.

For $z = 1$, after applying Eq. (B5), Eq. (B10) gives

$$\sum_{nk}|w_{nk}p_k - w_{kn}p_n||(p/q)^{\alpha-1} - C_\alpha| \leqslant \sqrt{\dot{S}_p}(\sqrt{I_F(p)} + 2\sqrt{\langle\overline{w}^2\rangle_p})^{1/2}\langle[(p/q)^{\alpha-1} - C_\alpha]^4\rangle_p^{1/4}, \tag{B12}$$

which corresponds to Eq. (42) in the main text.

APPENDIX C: UPPER LIMIT ON $\langle|X - \langle X\rangle|^3\rangle$ AND RELATED INEQUALITIES

In this Appendix we provide three different bounds on $\langle|X - \langle X\rangle|^3\rangle$, where X is some non-negative random variable, for which the first four moments exist. (We require non-negativity of X, since we deal with such cases in this paper.)

The first method generates the following inequality:

$$\langle|X - \langle X\rangle|^3\rangle \leqslant \langle X^3\rangle - \langle X\rangle\langle X^2\rangle, \tag{C1}$$

which can be justifying in a few steps as

$$\begin{aligned}\langle|X - \langle X\rangle|^3\rangle &= \langle(X - \langle X\rangle)^2|X - \langle X\rangle|\rangle\\ &\leqslant \langle(X - \langle X\rangle)^2(X + \langle X\rangle)\rangle\\ &= \langle X\rangle\langle(X - \langle X\rangle)^2\rangle + \langle X(X - \langle X\rangle)^2\rangle\end{aligned}$$

and by performing the averages. In the above, a simple inequality valid for $X \geqslant 0$ was used, namely $|X - \langle X\rangle| \leqslant X + \langle X\rangle$.

The second method generates a more complicated inequality,

$$\begin{aligned}\langle|X - \langle X\rangle|^3\rangle &\leqslant \sqrt{\langle X^2\rangle - \langle X\rangle^2}\\ &\quad\times\sqrt{\langle X^4\rangle + 6\langle X\rangle^2\langle X^2\rangle - 4\langle X\rangle\langle X^3\rangle - 3\langle X\rangle^4},\end{aligned} \tag{C2}$$

which follows from applying the Cauchy-Schwarz inequality as

$$\begin{aligned}\langle|X - \langle X\rangle|^3\rangle &= \langle(X - \langle X\rangle)^2|X - \langle X\rangle|\rangle\\ &\leqslant \sqrt{\langle(X - \langle X\rangle)^4\rangle}\sqrt{\langle(X - \langle X\rangle)^2\rangle},\end{aligned}$$

and executing some algebra under the square roots.

The third method uses Minkowski inequality, known in generality as $\langle|X + Y|^s\rangle^{1/s} \leqslant \langle|X|^s\rangle^{1/s} + \langle|Y|^s\rangle^{1/s}$. In our case, we obtain

$$\langle|X - \langle X\rangle|^3\rangle \leqslant (\langle X^3\rangle^{1/3} + \langle X\rangle)^3. \tag{C3}$$

It is clear that the upper bound in Eq. (C3) is larger than $\langle X^3\rangle$, and hence larger than the bound in Eq. (C1). This implies that Eq. (C1) provides a tighter bound than Eq. (C3). Moreover, the formula (C1) is simpler than the formula (C2), and it requires lower moments. For these reasons, we use Eq. (C1) for estimations in this paper. For example, Eq. (28) can be obtained by substituting $(p/q)^{\alpha-1}$ for X, and noting that $C_\alpha = \langle(p/q)^{\alpha-1}\rangle_p$.

Now let us find the upper limit on a related average, which appears in Eq. (39), i.e., $\langle|\ln(X) - \mu|^3\rangle$, where $\mu = \langle\ln(X)\rangle$, and μ can be negative. We have $\langle|\ln(X) - \mu|^3\rangle = \langle|\ln(X/e^\mu)|^3\rangle$. From Eq. (5) for arbitrary positive numbers x, y we have

$$|\ln(x/y)| \leqslant |\sqrt{x/y} - \sqrt{y/x}|, \tag{C4}$$

which implies for $x = X$ and $y = e^\mu$ the following series of inequalities:

$$\begin{aligned}\langle|\ln(X) - \mu|^3\rangle &\leqslant \langle|\sqrt{X/e^\mu} - \sqrt{e^\mu/X}|^3\rangle\\ &= \langle[\sqrt{X/e^\mu} - \sqrt{e^\mu/X}]^2|\sqrt{X/e^\mu} - \sqrt{e^\mu/X}|\rangle\\ &\leqslant \langle[\sqrt{X/e^\mu} - \sqrt{e^\mu/X}]^2(\sqrt{X/e^\mu} + \sqrt{e^\mu/X})\rangle\\ &= e^{-3\mu/2}\langle X^{3/2}\rangle - e^{-\mu/2}\langle X^{1/2}\rangle\\ &\quad - e^{\mu/2}\langle X^{-1/2}\rangle + e^{3\mu/2}\langle X^{-3/2}\rangle.\end{aligned} \tag{C5}$$

Similarly, we can estimate $\langle[\ln(X) - \mu]^4\rangle$, which appears in Eq. (43),

$$\begin{aligned}\langle[\ln(X) - \mu]^4\rangle &\leqslant \langle[\sqrt{X/e^\mu} - \sqrt{e^\mu/X}]^4\rangle\\ &= e^{-2\mu}\langle X^2\rangle - 4e^{-\mu}\langle X\rangle\\ &\quad + 6 - 4e^\mu\langle X^{-1}\rangle + e^{2\mu}\langle X^{-2}\rangle.\end{aligned} \tag{C6}$$

Obviously, the inequalities (C5) and (C6) require that several first few fractional moments (even of negative order) exist.

APPENDIX D: MUTUAL INFORMATION AND TEMPORAL FISHER INFORMATION FOR THE MODEL IN EQ. (114)

Mutual information $I(x, y)$ between x and y variables is $I(x, y) = \langle \ln(\frac{p_{xy}}{p_x p_y}) \rangle$, which for the model in Eq. (114) translates to

$$I(x, y) = \ln\left(1 + \frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + \langle \ln(1 + \epsilon r(t)xy)\rangle - \langle \ln(1 + \epsilon r(t)x/\kappa_2)\rangle - \langle \ln(1+\epsilon r(t)y/\kappa_1)\rangle. \quad \text{(D1)}$$

Next, we approximate the logarithms to the second order in ϵ, according to the formula $\ln(1 + x) = x - x^2/2 + O(x^3)$. After that step, the mutual information reads

$$I(x, y) = \epsilon r(t)\left(\left[\frac{1}{\kappa_1\kappa_2} + \langle xy\rangle - \frac{1}{\kappa_2}\langle x\rangle - \frac{1}{\kappa_1}\langle y\rangle\right] + \frac{1}{2}\epsilon r(t)\left[\frac{1}{\kappa_2^2}\langle x^2\rangle + \frac{1}{\kappa_1^2}\langle y^2\rangle - \langle x^2y^2\rangle - \frac{1}{(\kappa_1\kappa_2)^2}\right]\right) + O(\epsilon^3). \quad \text{(D2)}$$

The averages in Eq. (D2) are given to the lowest order by

$$\langle x\rangle = \frac{1}{\kappa_1}\left(1 + \frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2), \quad \text{(D3)}$$

$$\langle y\rangle = \frac{1}{\kappa_2}\left(1 + \frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2), \quad \text{(D4)}$$

$$\langle xy\rangle = \frac{1}{\kappa_1\kappa_2}\left(1 + 3\frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2), \quad \text{(D5)}$$

$$\langle x^2\rangle = \frac{2}{\kappa_1^2}\left(1 + 5\frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2), \quad \text{(D6)}$$

$$\langle y^2\rangle = \frac{2}{\kappa_2^2}\left(1 + 5\frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2), \quad \text{(D7)}$$

$$\langle x^2y^2\rangle = \frac{4}{(\kappa_1\kappa_2)^2}\left(1 + 8\frac{\epsilon r(t)}{\kappa_1\kappa_2}\right) + O(\epsilon^2). \quad \text{(D8)}$$

Insertion of the averages in Eqs. (D3)–(D8) into Eq. (D2) with some algebra produces Eq. (116) in the main text.

Temporal Fisher information $I_F(p_x) = \langle (\dot{p}_x/p_x)^2\rangle$ can be obtained by finding the ratio $\dot{p}_x/p_x$, which is

$$\frac{\dot{p}_x}{p_x} = \epsilon\frac{\dot{r}(t)}{\kappa_2}\left[\frac{1}{\kappa_1}\left(-1 + \epsilon\frac{r(t)}{\kappa_1\kappa_2}\right) + x\left(1 - \epsilon\frac{r(t)x}{\kappa_2}\right)\right] + O(\epsilon^3). \quad \text{(D9)}$$

Averaging the square of this expression yields Eq. (117) in the main text. In a similar way, one can evaluate temporal Fisher information for the joint probability density p_{xy}.

[1] A. Renyi, On Measures of Entropy and Information, *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability* (University of California Press, Berkeley, CA, 1960), pp. 547–561.

[2] I. Csiszar, Information-type measures of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. **2**, 299 (1967).

[3] C. Tsallis, Generalized entropy-based criterion for consistent testing, Phys. Rev. E **58**, 1442 (1998).

[4] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[5] A. Cichocki and S.-I. Amari, Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarity, Entropy **12**, 1532 (2010).

[6] T. Hatano and S. Sasa, Steady-state thermodynamics of Langevin systems, Phys. Rev. Lett. **86**, 3463 (2001).

[7] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[8] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, Thermodynamics of prediction, Phys. Rev. Lett. **109**, 120604 (2012).

[9] G. Falasco, M. Esposito, and J.-C. Delvenne, Beyond thermodynamic uncertainty relations: nonlinear response, error-dissipation trade-offs, and speed limits, J. Phys. A **55**, 124002 (2022).

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, NJ, 2006).

[11] M. D. Reid and R. C. Williamson, Information, divergence and risk for binary experiments, J. Mach. Learn. Res. **12**, 731 (2011).

[12] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, IEEE Trans. Inform. Theor. **52**, 4394 (2006).

[13] S-I. Amari and H. Nagaoka, *Methods of Information Geometry* (Oxford University Press, Oxford, 2000).

[14] I. Sason and S. Verdu, f-Divergence inequalities, IEEE Trans. Inf. Theory **62**, 5973 (2016).

[15] T. Morimoto, Markov processes and the H-theorem, J. Phys. Soc. Jpn. **18**, 328 (1963).

[16] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989).

[17] L. Borland, A. R. Plastino, and C. Tsallis, Information gain within nonextensive thermostatistics, J. Math. Phys. **39**, 6490 (1998).

[18] T. Yamano, de Bruijn-type identity for systems with flux, Eur. Phys. J. B **86**, 363 (2013).

[19] A. N. Gorban, General H-theorem and entropies that violate the second law, Entropy **16**, 2408 (2014).

[20] G. E. Crooks, Measuring thermodynamic length, Phys. Rev. Lett. **99**, 100602 (2007).

[21] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[22] T. V. Vu and K. Saito, Thermodynamic unification of optimal transport: Thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits, Phys. Rev. X **13**, 011013 (2023).

[23] S. Ito and A. Dechant, Stochastic time evolution, information geometry and the Cramer-Rao bound, Phys. Rev. X **10**, 021056 (2020).
[24] S. B. Nicholson, L. P. Garcia-Pintos, A. del Campo, and J. R. Green, Time-information uncertainty relations in thermodynamics, Nat. Phys. **16**, 1211 (2020).
[25] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, Phys. Rev. Lett. **121**, 070601 (2018).
[26] T. Koyuk and U. Seifert, Thermodynamic uncertainty relations for time-dependent driving, Phys. Rev. Lett. **125**, 260604 (2020).
[27] V. T. Vo, T. V. Vu, and Y. Hasegawa, Unified thermodynamic-kinetic uncertainty relation, J. Phys. A **55**, 405004 (2022).
[28] A. Kolchinsky and B. D. Tracey, Estimating mixture entropy with pairwise distances, Entropy **19**, 361 (2017).
[29] H. Finner, A generalization of Hölder's inequality and some probability inequalities, Ann. Probab. **20**, 1893 (1992).
[30] B. C. Carlson, Some inequalities for hypergeometric functions, Proc. Am. Math. Soc. **17**, 32 (1966).
[31] B. R. Frieden, *Science from Fisher Information: A Unification*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2004).
[32] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. **173**, 361 (2000).
[33] T. Yamano, Phase space gradient of dissipated work and information: A role of relative Fisher information, J. Math. Phys. **54**, 113301 (2013).
[34] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, Rev. Mod. Phys. **48**, 571 (1976).
[35] M. Baiesi, C. Maes, and B. Wynants, Fluctuations and response of nonequilibrium states, Phys. Rev. Lett. **103**, 010602 (2009).
[36] C. Maes and K. Netocny, Time-reversal and entropy, J. Stat. Phys. **110**, 269 (2003).
[37] C. Maes, Frenesy: Time-symmetric dynamical activity in nonequilibria, Phys. Rep. **850**, 1 (2020).
[38] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).
[39] N. Margolus and L. B. Levitin, The maximum speed of dynamical evolution, Physica D **120**, 188 (1998).
[40] B. Shanahan, A. Chenu, N. Margolus, and A. del Campo, Quantum speed limits across the quantum-to-classical transition, Phys. Rev. Lett. **120**, 070401 (2018).
[41] R. Hamazaki, Speed limits for macroscopic transitions, PRX Quantum **3**, 020319 (2022).
[42] L. P. García-Pintos, S. B. Nicholson, J. R. Green, A. del Campo, and A. V. Gorshkov, Unifying quantum and classical speed limits on observables, Phys. Rev. X **12**, 011038 (2022).
[43] S. Deffner and S. Campbell, Quantum speed limits: From Heisenberg's uncertainty principle to optimal quantum control, J. Phys. A **50**, 453001 (2017).
[44] A. C. Barato and U. Seifert, Thermodynamic uncertainty relation for biomolecular processes, Phys. Rev. Lett. **114**, 158101 (2015).
[45] J. Li, J. M. Horowitz, T. R. Gingrich, and N. Fakhri, Quantifying dissipation using fluctuating currents, Nat. Commun. **10**, 1666 (2019).
[46] T. Van Vu, V. T. Vo, and Y. Hasegawa, Entropy production estimation with optimal current, Phys. Rev. E **101**, 042138 (2020).
[47] D. J. Skinner and J. Dunkel, Estimating entropy production from waiting time distributions, Phys. Rev. Lett. **127**, 198101 (2021).
[48] D. S. P. Salazar, Lower bound for entropy production rate in stochastic systems far from equilibrium, Phys. Rev. E **106**, L032101 (2022).
[49] A. Dechant, Minimum entropy production, detailed balance and Wasserstein distance for continuous-time Markov processes, J. Phys. A **55**, 094001 (2022).
[50] M. Esposito and C. van den Broeck, Second law and Landauer principle far from equilibrium, Europhys. Lett. **95**, 40004 (2011).
[51] E. A. Novikov, Functionals and the random-force method in turbulence theory, Sov. Phys. JETP **20**, 1290 (1965).
[52] P. Miller, A. M. Zhabotinsky, J. E. Lisman, and X-J. Wang, The stability of stochastic CaMKII switch: dependence on the number of enzyme molecules and protein turnover, PLoS Biol. **3**, e107 (2005).
[53] C. C. Petersen, R. C. Malenka, R. A. Nicoll, and J. J. Hopfield, All-or-none potentiation at CA3-CA1 synapses, Proc. Natl. Acad. Sci. (USA) **95**, 4732 (1998).
[54] J. Karbowski, Metabolic constraints on synaptic learning and memory, J. Neurophysiol. **122**, 1473 (2019).
[55] P. Dayan and L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, MA, 2000).
[56] D. B. Chklovskii, B. W. Mel, and K. Svoboda, Cortical rewiring and information storage, Nature (London) **431**, 782 (2004).
[57] H. Kasai, N. E. Ziv, H. Okazaki, S. Yagishita, and T. Toyoizumi, Spine dynamics in the brain, mental disorders, and artificial neural networks, Nat. Rev. Neurosci. **22**, 407 (2021).
[58] J. Karbowski and P. Urban, Information encoded in volumes and areas of dendritic spines is nearly maximal across mammalian brains, Sci. Rep. **13**, 22207 (2023).
[59] J. Karbowski and P. Urban, Cooperativity, information gain, and energy cost during early LTP in dendritic spines, Neural Comput. **36**, 271 (2024).
[60] W. Bialek, I. Nemenman, and N. Tishby, Predictability, complexity, and learning, Neural Comput. **13**, 2409 (2001).
[61] A. H. Lang, C. K. Fisher, T. Mora, and P. Mehta, Thermodynamics of statistical inference by cells, Phys. Rev. Lett. **113**, 148103 (2014).
[62] S. E. Palmer II, O. Marre, M. J. Berry, and W. Bialek, Predictive information in a sensory population, Proc. Natl. Acad. Sci. (USA) **112**, 6908 (2015).
[63] S. Still, Information bottleneck approach to predictive inference, Entropy **16**, 968 (2014).
[64] D. Hartich, A. C. Barato, and U. Seifert, Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation, J. Stat. Mech. (2014) P02016.
[65] G. Diana and M. Esposito, Mutual entropy production in bipartite systems, J. Stat. Mech. (2014) P04010.
[66] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, Phys. Rev. X **4**, 031015 (2014).
[67] A. C. Barato, D. Hartich, and U. Seifert, Information-theoretic versus thermodynamic entropy production in autonomous sensory networks, Phys. Rev. E **87**, 042104 (2013).
[68] L. Paninski, Estimation of entropy and mutual information, Neural Comput. **15**, 1191 (2003).
[69] A. Kraskov, H. Stogbauer, and P. Grassberger, Estimating mutual information, Phys. Rev. E **69**, 066138 (2004).

[70] A. E. Allahverdyan, D. Janzing, and G. Mahler, Thermodynamic efficiency of information and heat flow, J. Stat. Mech. (2009) P09011.

[71] J. Ehrich and D. A. Sivak, Energy and information flows in autonomous systems, Front. Phys. **11**, 1108357 (2023).

[72] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Develop. **5**, 183 (1961).

[73] T. Sagawa, *Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics* (Springer, Singapore, 2022).

[74] D. H. Wolpert, The stochastic thermodynamics of computation, J. Phys. A **52**, 193001 (2019).

[75] F. Rieke, D. Warland, R. de Ruyter, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1999).

[76] M. Gabrie, A. Manoel, C. Luneau *et al.*, Entropy and mutual information in models of deep neural networks, J. Stat. Mech. (2019) 124014.

[77] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, Science **334**, 354 (2011).