

Intrinsic irreversibility of Markovian chainsP. Di Porto,¹ B. Crosignani,² and E. DelRe^{1,3,*}¹*Dipartimento di Fisica, Università di Roma "La Sapienza," I-00185 Rome, Italy*²*Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA*³*IPCF-CNR, Università di Roma "La Sapienza," I-00185 Rome, Italy*

(Received 3 December 2013; published 13 May 2014)

We show that for a large class of stationary Markov processes the total variation distance between the final equilibrium distribution and that at a given time is a strongly monotonic vanishing function. We illustrate this for basic paradigmatic processes and discuss how, for systems susceptible to a canonical description, this can be interpreted as a statistical arrow of time that exists besides the standard decrease of free energy.

DOI: [10.1103/PhysRevE.89.052118](https://doi.org/10.1103/PhysRevE.89.052118)

PACS number(s): 02.50.Ga, 05.20.-y, 05.70.Ln

I. INTRODUCTION

Irreversibility emerges when we formulate a macroscopic description of natural phenomena [1]. In terms of statistical mechanics, this breaking of time symmetry is associated with an increase in entropy, for isolated systems, and a decrease in free energy in the canonical realm [2,3]. Since stochastic processes may or may not be susceptible to a canonical description, the question remains if a more general signature of irreversibility can be identified. Our focus here is on irreversibility for stochastic systems described by Markov chains, the most important random processes in physics, of which the canonical ones are a subclass [4].

We here report a strong monotonic convergence theorem of the total variation distance of probability distributions for relaxing stationary Markov processes. Our finding amounts to a new necessary condition for Markovian systems and hence increases our ability to infer from data the nature of underlying stochastic mechanisms. In those particular cases where a canonical formulation can be implemented, this necessary condition is independent of that of decreasing free energy.

Much about a macroscopic system can be detected by observing how it relaxes, i.e., how it passes from a given initial condition to one that no longer evolves with respect to the macroscopic observables we decide to use [5]. In standard relaxation experiments, the time evolution of a macroscopic quantity is detected, such as the magnetization or polarization in a crystal. In principle, an analysis of an ensemble of relaxing replicas will allow us to detect a far more fundamental quantity, i.e., how the actual probability distribution settles into its final equilibrium one. The question arises if some quantities can be formulated from the evolving distribution that have new properties and that can shed a different light on the physical mechanisms involved in the relaxation. As opposed to averaged quantities, such as magnetization, once we observe the underlying probability distributions, we can consider the way in which the “distance” between the probability distribution observed at a given instant of time and the final distribution vanishes as time goes by. The point is that the distance between distributions $\{p_m\}$, $\{q_m\}$, $m = 1, 2, \dots, N$, can be naturally defined in different ways [6–10], with relevant examples being Jeffreys distance [7] and, of course, the Euclidean dis-

tance, i.e., $[(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_N - q_N)^2]^{1/2}$. Interestingly, although these distances have the property of vanishing as the system relaxes to its final state and are such that the ensemble of probability distributions forms a metric space, none have been shown to converge to zero with a strong monotonic behavior. Strong monotonic convergence has been established for measures associated with free energy and its generalizations, such as directed divergence [7,8,11], which do not rank among metric distances.

II. MONOTONIC VANISHING OF D

Our result is achieved for the total variation distance defined as $D[p, q] \equiv D[\{p_m\}, \{q_m\}] = \frac{1}{2} \sum_{m=1}^N |p_m - q_m|$ [12], a measure used in characterizing the convergence behavior of Markov chains that has application in a variety of fields [13]. In previous studies, it has been shown that this distance obeys a *weakly* monotonic convergence theorem for ergodic Markov chains [14]. Hereafter, our stationary Markov process (SMP) is defined through the chains $p_m(\lambda + 1) = \sum_{n=1}^N a_{m,n} p_n(\lambda)$, ($m = 1, 2, \dots, N$), where $a_{m,n} \geq 0$ is the probability of a system in state n after λ steps to shift to state m after $\lambda + 1$ steps. Our basic finding is that a wide class of physically meaningful SMPs obeys the following theorem:

(a) There is only one equilibrium distribution $\{p_m^{(\text{eq})}\}$, i.e., a distribution for which

$$p_m^{(\text{eq})} = \sum_{n=1}^N a_{m,n} p_n^{(\text{eq})}. \quad (1)$$

(b) All $p_m^{(\text{eq})}$ turn out to be positive.

(c) The distance $D(\lambda) = \frac{1}{2} \sum_{m=1}^N |p_m^{(\text{eq})} - p_m(\lambda)|$ is a strongly monotonic vanishing function of λ , i.e., $D(\lambda_2) < D(\lambda_1)$ if $\lambda_2 > \lambda_1$ [15].

(d) If the transition matrix $\hat{M} \equiv \{a_{m,n}\}$ is not singular and we exclude the trivial case $P(\lambda = 0) = P^{(\text{eq})}$, the monotonic vanishing is asymptotic, i.e., $D(\lambda) > 0$ for all finite λ . By replacing λ with the continuous variable t , we have that

$$\lim_{t \rightarrow \infty} D(t) = 0 \quad (2)$$

in a monotonic asymptotic way. The theorem is explicitly demonstrated in Appendixes A and B (for point d).

*eugenio.delre@uniroma1.it

III. EXAMPLES OF PARTICULAR SMPs

To provide some insight and examples of the strong monotonic behavior of $D(t)$, we discuss three specific SMPs: (a) the random telegraph process, widely implemented in the description of dichotomic systems (such as spin and spinlike systems) [4], (b) the Rayleigh particle (an Ornstein-Uhlenbeck process) [4], and (c) the harmonically bound Brownian particle [16].

(a) *Random telegraph process.* A discrete SMP takes only the two values $y(t) = +1$, $y(t) = -1$, for which

$$p(y,t) = \frac{1}{2}[1 + \exp(-2\alpha t)]p(y,0) + \frac{1}{2}[1 - \exp(-2\alpha t)]p(-y,0), \quad (3)$$

where α represents the rate at which y switches from $+1$ to -1 and vice versa, with the asymptotic probabilities being $p^{(\text{eq})}(+1) = p^{(\text{eq})}(-1) = \frac{1}{2}$. Thus, we have

$$D(t) = \frac{1}{2} \exp(-2\alpha t) [|p(1,0) - p(-1,0)|], \quad (4)$$

which clearly exhibits the predicted monotonic vanishing.

(b) *Rayleigh particle.* The stochastic variable is the dimensionless velocity $\eta(\tau)$ of a (one-dimensional) Brownian particle starting with initial velocity $\eta(0) = 0$ at time $\tau = 0$, with the dimensionless time $\tau = \beta t$, where β is the velocity damping coefficient. The velocity density distribution reads [4]

$$p(\eta, \tau) = \left[\frac{\gamma(\tau)}{2\pi} \right]^{1/2} \exp[-\gamma(\tau)\eta^2/2], \quad (5)$$

where

$$\gamma(\tau) = \frac{1}{1 - \exp(-2\tau)}, \quad (6)$$

with the equilibrium distribution density being

$$p^{(\text{eq})}(\eta) = \frac{1}{(2\pi)^{1/2}} \exp(-\eta^2/2). \quad (7)$$

To proceed, we next note that D can be extended to the continuum: if the considered process is represented by a probability-density distribution $p(x)$, D becomes

$$D = \frac{1}{2} \int_{-\infty}^{+\infty} dx |p(x) - q(x)|. \quad (8)$$

If we shift to continuum, i.e., we use Eq. (8), Eqs. (5) and (7) yield

$$D(\tau) = \left(\frac{1}{8\pi} \right)^{1/2} \int_{-\infty}^{+\infty} d\eta |\exp(-\eta^2/2) - \gamma^{1/2} \exp[-\gamma(\tau)\eta^2/2]|. \quad (9)$$

Noting that the sign of the difference inside $|\dots|$ in Eq. (9) changes for $\eta = \pm\eta^*(\tau)$, with $\eta^*(\tau) = \{\ln \gamma(\tau)/[\gamma(\tau) - 1]\}^{1/2}$, the integral can easily be performed, thus leading to

$$D(\tau) = \text{Erf} \left[\frac{\eta^*(\tau)\gamma^{1/2}(\tau)}{\sqrt{2}} \right] - \text{erf} \left[\frac{\eta^*(\tau)}{\sqrt{2}} \right], \quad (10)$$

where erf is the error function. A plot of $D(\tau)$ showing its monotonic vanishing is given in Fig. 1.

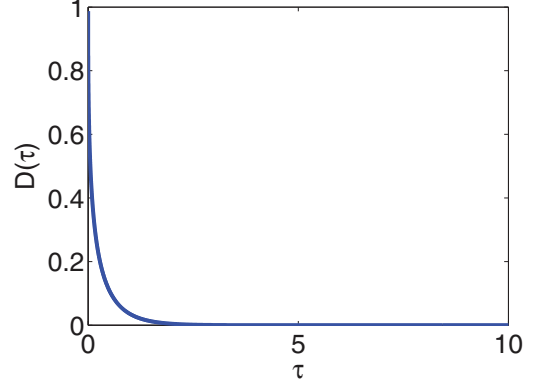


FIG. 1. (Color online) Monotonic vanishing of $D(\tau)$ for the Rayleigh particle.

(c) *Harmonically bound Brownian particle.* We consider a (one-dimensional) Brownian particle of mass M bound by a harmonic force, obeying the equation $du/dt = -\beta u + A(t) - \omega^2 x$, $u = dx/dt$, where $A(t)$ is Langevin's acceleration and ω is the circular frequency of the oscillator. We are dealing with two random variables x and u , so that, shifting to continuum,

$$D(t) = \frac{1}{2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} du |P^{(\text{eq})}(x,u) - P(x,u,t)|, \quad (11)$$

where $P(x,u,t)$ denotes the probability density of x and u at time t and $P^{(\text{eq})}(x,u)$ is its equilibrium asymptotic expression,

$$P^{(\text{eq})}(x,u) = \frac{M\omega}{2\pi kT} e^{-M\omega^2 x^2/2kT} e^{-Mu^2/2kT}. \quad (12)$$

In order to test the monotonic vanishing of $D(t)$, we consider two significant cases:

$$P(x,u,0) = \delta(x) \left[\frac{M}{2\pi kT} \right]^{1/2} e^{-Mu^2/2kT}, \quad (13)$$

which corresponds to an ensemble of oscillators starting from $x = 0$, with the initial velocity distribution $P(u_0)$ given by

$$P(u_0) = \left[\frac{M}{2\pi kT} \right]^{1/2} e^{-Mu_0^2/2kT}, \quad (14)$$

and

$$P(x,u,0) = \delta(x - x_0)\delta(u - u_0), \quad (15)$$

which corresponds to an ensemble of oscillators starting from initial position x_0 with initial velocity u_0 . In order to determine the explicit expressions of $P(x,u,t)$ relative to our examples, we have suitably extended the results of Ref. [16] (see Appendix C). The left panels of Figs. 2 and 3 show plots corresponding to Eqs. (13) and (15), respectively, and they exhibit the expected monotonic vanishing of $D(t)$. In both figures, we have $\beta_1 = (\beta^2 - 4\omega^2)^{1/2}$ (see Appendix C), and we have set $\beta = 1$, $kT/M = 1$, thus introducing suitable space and time units. It is worth noting that, if we replace in Eq. (11) $|P^{(\text{eq})}(x,u) - P(x,u,t)|$ with $|P^{(\text{eq})}(u) - P(u,t)|$, after tracing $P^{(\text{eq})}(x,u)$ and $P(x,u,t)$ over x , the vanishing of the “reduced” distance $D_u(t) = (1/2) \int_{-\infty}^{+\infty} du |P^{(\text{eq})}(u) - P(u,t)|$, where $P^{(\text{eq})}(u) = (M/2\pi kT)^{1/2} \exp(-Mu^2/2kT)$, is not necessarily monotonic (see the right panels in Figs. 2

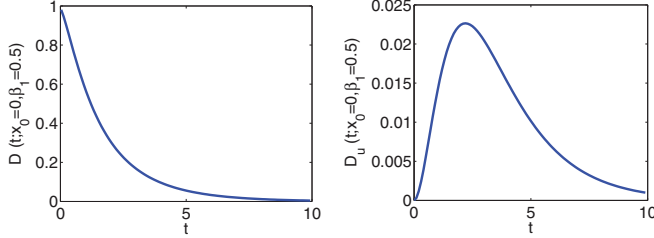


FIG. 2. (Color online) (left) Monotonic vanishing of $D(t)$ for the harmonically bound Brownian particle in the case corresponding to Eq. (13). (right) Nonmonotonic vanishing of $D_u(t)$ in the same case.

and 3). Therefore, the monotonic vanishing of the variation distance requires us to consider the whole set of variables defining the system state, i.e., in the present case, both x and u . In particular, Eq. (13) refers to the situation in which the Brownian particles start from the mechanical equilibrium position $x = 0$, with their velocity distribution having already achieved thermal equilibrium. Over time, the velocity distribution changes from its initial equilibrium configuration but eventually goes back to it, as shown in Fig. 3 (right panel). Thus, the behavior of the variation distances confirms a well-known feature of the harmonically bound particle: the velocity is not a Markov process, while the process composed of the velocity and position is Markovian. A clear physical insight of this property is described explicitly on page 76 of Ref. [4]: if we know the position and the velocity of the particle at time t , the probability density $P(x, u, t')$ of both quantities at a later time t' is determined in terms of the external force and temperature, and this implies the Markovian nature of the process. If we only know the particle velocity at time t and the external force depends on position, the incomplete information does not allow us to define the probability density of the velocity at later times. As a consequence, the process relative to the velocity alone is not Markovian in nature, as confirmed by Figs. 2 and 3. Intuitively, this is the random counterpart of the deterministic motion of a particle subject to a position-dependent force: if we know both velocity u and position x at time t , we can determine u at t' , but the knowledge of $u(t)$ alone does not allow us to foresee the velocity at later times.

We note that the monotonic vanishing of total variation distance applies also to situations in which the system size becomes important, to the extent that they are particular examples of stationary Markov processes (see, for instance,

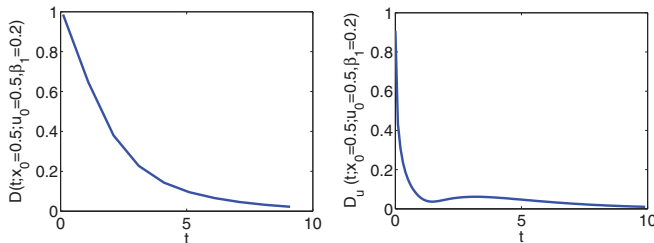


FIG. 3. (Color online) (left) Monotonic vanishing of $D(t)$ for the harmonically bound Brownian particle in the case corresponding to Eq. (15). (right) Nonmonotonic vanishing of $D_u(t)$ in the same case.

Ref. [17]). However, the effective usefulness of our criterion in describing the behavior of such irreversible processes is still an open question.

IV. DISCUSSION

To grasp the impact of our finding, consider, for example, the important debate on the inference of power laws in relaxing systems [18–21] or in the long tails of probability distributions, which suggests the existence of universal laws underlying many physical phenomena generally associated with complex systems and networks [22–25]. The basis for these inferences is the comparison of partial, i.e., nonequilibrium, distributions observed up to an experimentally accessible time t to model equilibrium distributions. A detection of $D(t)$ in these cases can help rule out that the underlying stochastic process is an SMP, even from very sparse data subsets [6]. This can, in some cases, form a quantitative basis for statements that the observed data have some truly fundamental complex nature.

In the particular case of canonical ensembles, we note that the strong monotonic convergence theorem constitutes a new necessary condition to be satisfied by an SMP, along with the strong monotonic decrease in free energy $F(t)$ given by $F(t) = U(t) - TS(t)$, where $U(t) = \sum_m p_m(t) E_m$, $S(t) = -k \sum_m p_m(t) \ln [p_m(t)]$, with k labeling Boltzmann's constant and E_m being the energy of the m th state or its generalization (Tsallis's free energy) [26,27]. From a mathematical point of view, the two conditions are mutually independent in that one could envisage situations in which the free energy decreases while the variation distance increases or vice versa (see Appendix D). This sets the stage for us to appreciate the more fundamental and far-reaching consequence: for a generic canonical ensemble in statistical mechanics, i.e., for an ensemble of $S \gg 1$ identical systems where each one is in contact with one out of S equivalent thermal baths at temperature T , the strong monotonic behavior of $D(t)$ constitutes a detectable time arrow that is inherently different from that associated with the time-dependent $F(t)$. Whereas the convergence of $F(t)$ is directly related to the notion that the total entropy, i.e., that of the system ensemble plus that of the bath ensemble, increases with time, the time arrow of $D(t)$ is an example of a statistical quantity that is *unrelated* to entropy.

Finally, we stress that the total variation distance $D(t)$ can be simpler to evaluate in some situations, while the free energy $F(t)$ can be preferable in other cases. In addition, there are systems, such as the random telegraph considered above, for which the temperature does not play any relevant role, excluding the use of free energy. In general, what matters is that the monotonic vanishing of $D(t)$ and the monotonic decreasing of $F(t)$ are *independent* criteria to test whether a given random process is a stationary Markov process (see Appendix D): if this is the case, both criteria must be independently satisfied.

In summary, given two configurations of a relaxing system P_1, P_2 , we may *a priori* assume that the first configuration in time is more distant than the second from the final configuration $P^{(eq)}$ if the distance is measured through the total variation distance $D(t)$. Thus, the strong convergence

of $D(t)$ is an intrinsic measure of irreversibility in stationary Markov processes.

ACKNOWLEDGMENT

We are grateful to A. Ciattoni for useful discussions.

APPENDIX A: MONOTONIC VANISHING OF THE SERIES $D(\lambda)$

Let us consider the set of equations obeyed by any equilibrium distribution $P^{(eq)}$,

$$\sum_{n=1}^N [a_{m,n} - \delta_{m,n}] p_n^{(eq)} = 0 \quad (m = 1, 2, \dots, N), \quad (\text{A1})$$

where $\delta_{m,n}$ is the Kronecker symbol, together with the relation

$$\sum_{m=1}^N [a_{m,n} - \delta_{m,n}] = \sum_{m=1}^N a_{m,n} - 1 = 0 \quad (n = 1, 2, \dots, N). \quad (\text{A2})$$

Equations (A2) imply that the determinant characterizing system (A1) possesses linearly dependent rows, so that it vanishes, which implies the existence of eigenvectors. Hereafter, we consider transition matrices \hat{M} fulfilling the two conditions

$$\sum_{m=1}^{\mathcal{M}} a_{m,n} < 1 \quad \text{for any } \mathcal{M} < N \quad (\text{A3})$$

for at least one $n \leq \mathcal{M}$ and

$$a_{m,n} > 0 \quad (n = 1, 2, \dots, N) \quad (\text{A4})$$

for at least one m . The above conditions are satisfied in most situations and possess an intuitive meaning in that they favor transitions among different regions in Γ space.

We now wish to show that no alternate-sign eigenvector of set (A1) exists and that all $p_m^{(eq)}$ are not vanishing. In fact, should such an eigenvector exist, we would have as a solution of (A1) (with a possible reordering of rows and columns)

$$p_1^{(eq)}, p_2^{(eq)}, \dots, p_{\mathcal{M}}^{(eq)} > 0, \quad (\text{A5})$$

$$p_{\mathcal{M}+1}^{(eq)}, p_{\mathcal{M}+2}^{(eq)}, \dots, p_N^{(eq)} \leq 0 \quad (\text{A6})$$

for a particular $\mathcal{M} < N$. We sum up the first \mathcal{M} equations of the set (A1), thus getting

$$\sum_{m=1}^{\mathcal{M}} \sum_{n=\mathcal{M}+1}^N a_{m,n} p_n^{(eq)} = \sum_{n=1}^{\mathcal{M}} \left[1 - \sum_{m=1}^{\mathcal{M}} a_{m,n} \right] p_n^{(eq)}. \quad (\text{A7})$$

Equations (A6) imply that the left-hand side of Eq. (A7) is not positive, while the right-hand side is positive, as a consequence of Eqs. (A5) and (A3). Therefore, no alternate-sign eigenvector of Eqs. (A1) exists because all $p_m^{(eq)}$ have a common sign. This implies that there is only one distribution satisfying Eqs. (A1) with $\sum_{m=1}^N p_m^{(eq)} = 1$ (all $p_m^{(eq)}$ being positive). In fact, should a different normalized distribution $\{q_m^{(eq)}\}$ exist, we would have an eigenvector $\{p_m^{(eq)} - q_m^{(eq)}\}$ with alternate signs, which is

impossible. Referring now to the obvious relations

$$p_1(\lambda), p_2(\lambda), \dots, p_N(\lambda) \geq 0, \quad \sum_{m=1}^N p_m(\lambda) = 1, \quad (\text{A8})$$

we wish to show that

$$\lim_{\lambda \rightarrow \infty} D(\lambda) = 0 \quad (\text{A9})$$

in a ‘‘strongly’’ monotonic way [$D(\lambda + 1) < D(\lambda)$], defining

$$D(\lambda) = \sum_{m=1}^N |p_m^{(eq)} - p_m(\lambda)|. \quad (\text{A10})$$

To this aim, we set

$$\eta_m(\lambda) = p_m^{(eq)} - p_m(\lambda), \quad (\text{A11})$$

so that, if we recall that $\sum_{m=1}^N a_{m,n} = 1$ ($n = 1, 2, \dots, N$),

$$D(\lambda) = \sum_{n=1}^N |\eta_n(\lambda)| = \sum_{m=1}^N \sum_{n=1}^N a_{m,n} |\eta_n(\lambda)|. \quad (\text{A12})$$

On the other hand, Eq. (A10) yields

$$D(\lambda + 1) = \sum_{m=1}^N |p_m^{(eq)} - p_m(\lambda + 1)|, \quad (\text{A13})$$

which leads to, keeping in mind Eq. (1),

$$D(\lambda + 1) = \sum_{m=1}^N \left| \sum_{n=1}^N a_{m,n} \eta_n(\lambda) \right|. \quad (\text{A14})$$

If we now compare Eq. (A12) with Eq. (A14) and remember that all $a_{m,n}$ are ≥ 0 , we obtain, in particular,

$$D(\lambda) - D(\lambda + 1) \geq \sum_{n=1}^N a_{m,n} |\eta_n(\lambda)| - \left| \sum_{n=1}^N a_{m,n} \eta_n(\lambda) \right|. \quad (\text{A15})$$

We note that, while the first term on the right-hand side of Eq. (A15) is made up by all nonnegative contributions, inside the sum of the second term there are both positive and negative contributions; in fact, all $a_{m,n}$ are positive [Eq. (A4)], and there are both positive and negative $\eta_n(\lambda)$ due to the relation

$$\sum_{n=1}^N \eta_n(\lambda) = 0. \quad (\text{A16})$$

As an immediate consequence,

$$D(\lambda) - D(\lambda + 1) > 0. \quad (\text{A17})$$

Therefore, $D(0), D(1), \dots, D(\lambda), \dots$ is a positive decreasing series. Obviously, if $D(0) = 0$, all terms vanish. If $D(0) > 0$, the decreasing series turns out to be indefinitely positive, provided the transition matrix \hat{M} is not singular (see Appendix B). In any case, the series possesses a limit $L \geq 0$. We wish to show that $L = 0$. To this aim, we observe that Eq. (A16) and the relation $\sum_{n=1}^N |\eta_n(\lambda)| \geq L$ imply that $\sum_{n=1}^N \eta_n(\lambda)$ is made up by two equal and opposite contributions, both $\geq L/2$ in magnitude. As a consequence, $\sum_{n=1}^N a_{m,n} \eta_n(\lambda)$ is made up by a positive contribution and a

negative contribution, both $\geq aL/2$ in magnitude, where $a > 0$ is the minimum value of $a_{m,n}$ ($n = 1, 2, \dots, N$) [see Eq. (A4)]. Therefore, $D(\lambda) - D(\lambda + 1) \geq aL$ for any λ [see Eq. (A15)], so that convergence implies $L = 0$.

As a result of the above issue, any Markov chain relative to our matrix \hat{M} tends to the equilibrium distribution $\{p_m^{(\text{eq})}\}$, which is made up by all positive elements. Thus, for any Markov chain there is a $\hat{\lambda}$ such that $p_\lambda > 0$ ($m = 1, 2, \dots, N$) for $\lambda > \hat{\lambda}$, which implies that all chains are ergodic. This property should ensure per se *weakly* monotonic vanishing of $D(\lambda)$, i.e., $D(\lambda + 1) \leq D(\lambda)$ [13,14,28]. To clarify this point, let us consider the two matrices

$$\hat{M}_1 = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \\ 2/3 & 1/3 & 0 \end{pmatrix}, \quad (\text{A18})$$

$$\hat{M} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/6 & 1/3 & 1/2 \\ 5/6 & 1/6 & 0 \end{pmatrix}, \quad (\text{A19})$$

which possess the common equilibrium distribution

$$\begin{pmatrix} p_1^{(\text{eq})} \\ p_2^{(\text{eq})} \\ p_3^{(\text{eq})} \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}. \quad (\text{A20})$$

If we start, e.g., with the initial distribution $p_1(\lambda = 0) = 1, p_2(\lambda = 0) = 0, p_3(\lambda = 0) = 0$, we obtain, for \hat{M}_1 , $D_1(\lambda = 0) = 2/3, D_1(1) = 2/3, D_1(2) = 1/3, D_1(3) = 1/9, D_1(4) = 1/9, \dots$, i.e., *weakly* monotonic vanishing, while, for \hat{M} , the series $D(\lambda)$ vanishes in a *strongly* monotonic way. The above examples show how Eq. (A4), which is satisfied in most physical situations, is expedient in assuring *strongly* monotonic vanishing of $D(\lambda)$. In turn, this essential condition implies that our clock never stops.

APPENDIX B: ASYMPTOTIC VANISHING OF THE SERIES $D(\lambda)$

We wish to see whether a $D(\lambda) > 0$ followed by $D(\lambda + 1) = 0$ exists. If \hat{M} is not singular, which permits an inverse matrix \hat{M}^{-1} , the answer is negative. In fact, should such a $D(\lambda)$ exist, we would have

$$\hat{M}^{-1}\{0, 0, \dots, 0\} = \{\eta_1(\lambda), \eta_2(\lambda), \dots, \eta_N(\lambda)\} \neq \{0, 0, \dots, 0\}, \quad (\text{B1})$$

which is clearly impossible. No general claim can be made if \hat{M} is singular.

APPENDIX C: MONOTONIC VANISHING OF $D(t)$ FOR THE HARMONICALLY BOUND BROWNIAN PARTICLE

We test here the monotonic vanishing of $D(t)$ for the harmonically bound Brownian particle. To this aim, we need the explicit expression of the probability density $P(x, u, t)$, a task which can be accomplished by generalizing the approach of Ref. [16]. In this way we obtain, in the particular case relative to an ensemble of particles all starting from position

x_0 with velocity u_0 at time $t = 0$,

$$P[x, u, t; x_0, u_0] = \frac{1}{2\pi[KG - H^2]^{1/2}} \exp \left(\frac{[-G(x-f)^2 + 2H(x-f)(u-g) - K(u-g)^2]}{2[KG - H^2]} \right), \quad (\text{C1})$$

where

$$f = f(t) = x_0 e^{-\frac{\beta t}{2}} \cosh\left(\frac{\beta_1 t}{2}\right) + \frac{(x_0 \beta + 2u_0)}{\beta_1} e^{-\frac{\beta t}{2}} \sinh\left(\frac{\beta_1 t}{2}\right), \quad (\text{C2})$$

$$g = g(t) = u_0 e^{-\frac{\beta t}{2}} \cosh\left(\frac{\beta_1 t}{2}\right) - \frac{(2x_0 \omega^2 + \beta u_0)}{\beta_1} e^{-\frac{\beta t}{2}} \sinh\left(\frac{\beta_1 t}{2}\right), \quad (\text{C3})$$

and

$$K = K(t) = \frac{kT}{M\omega^2} \left\{ 1 - \frac{e^{-\beta t}}{\beta_1^2} [2\beta^2 \sinh^2(\beta_1 t/2) + \beta\beta_1 \sinh(\beta_1 t) + \beta_1^2] \right\},$$

$$G = G(t) = \frac{kT}{M} \left\{ 1 - \frac{e^{-\beta t}}{\beta_1^2} [2\beta^2 \sinh^2(\beta_1 t/2) - \beta\beta_1 \sinh(\beta_1 t) + \beta_1^2] \right\},$$

$$H = H(t) = \frac{4\beta kT}{M\beta_1^2} e^{-\beta t} \sinh^2(\beta_1 t/2), \quad (\text{C4})$$

with

$$\beta_1 = (\beta^2 - 4\omega^2)^{1/2}. \quad (\text{C5})$$

Another interesting situation is that in which the initial velocity is not well determined but obeys a probability-density distribution $P(u_0)$, so that the time-dependent P reads

$$P(x, u, t; x_0) = \int_{-\infty}^{+\infty} du_0 P(x, u, t; x_0, u_0) P(u_0). \quad (\text{C6})$$

The variation distance corresponding to Eq. (C1) is given by

$$D(t; x_0, u_0) = \frac{1}{2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} du |P^{(\text{eq})}(x, u) - P(x, u, t; x_0, u_0)|, \quad (\text{C7})$$

while, for Eq. (C6),

$$D(t; x_0) = \frac{1}{2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} du |P^{(\text{eq})}(x, u) - P(x, u, t; x_0)|, \quad (\text{C8})$$

where $P^{(\text{eq})}(x, u)$ is given by Eq. (12).

APPENDIX D: MUTUAL INDEPENDENCE OF $D(t)$ AND $F(t)$

In order to prove the mutual independence of $D(t)$ and $F(t)$, we provide an example in which $D(t)$ decreases with time while $F(t)$ increases or vice versa. Consider a three-state system with energies (in units of $kT = 1$)

$$E_1 = E_2 = 0, \quad E_3 = 1, \quad (D1)$$

which corresponds to

$$p_1^{(eq)} = p_2^{(eq)} = \frac{e}{1 + 2e}, \quad p_3^{(eq)} = \frac{1}{1 + 2e}, \quad (D2)$$

and assume

$$p_1^{(1)} > p_2^{(1)} \geq p_1^{(eq)} = p_2^{(eq)} > p_3^{(eq)} > p_3^{(1)} \\ \times (p_1^{(1)} + p_2^{(1)} + p_3^{(1)} = 1) \quad (D3)$$

and

$$p_1^{(2)} = p_2^{(2)} = \frac{p_1^{(1)} + p_2^{(1)}}{2} + \epsilon, \quad p_3^{(2)} = p_3^{(1)} - 2\epsilon (\epsilon > 0). \quad (D4)$$

From Eqs. (D4), we have

$$D^{(2)} = D^{(1)} + 4\epsilon \rightarrow D^{(2)} > D^{(1)}, \quad (D5)$$

while, from Eqs. (D1) and (D4) ($F = \sum_{m=1}^3 p_m E_m + kT \sum_{m=1}^3 p_m \ln p_m$), we have

$$F^{(1)} = p_3^{(1)} + p_1^{(1)} \ln p_1^{(1)} + p_2^{(1)} \ln p_2^{(1)} + p_3^{(1)} \ln p_3^{(1)}, \quad (D6)$$

$$F^{(2)} = p_3^{(1)} - 2\epsilon + [p_1^{(1)} + p_2^{(1)} + 2\epsilon] \ln \left[\frac{p_1^{(1)} + p_2^{(1)}}{2} + \epsilon \right] \\ + [p_3^{(1)} - 2\epsilon] \ln [p_3^{(1)} - 2\epsilon], \quad (D7)$$

so that

$$F^{(2)} - F^{(1)} = \ln \left[\frac{p_1^{(1)} + p_2^{(1)}}{2} \right] (p_1^{(1)} + p_2^{(1)}) \\ - (p_1^{(1)} \ln p_1^{(1)} + p_2^{(1)} \ln p_2^{(1)}) + \mathcal{O}(\epsilon). \quad (D8)$$

Since $\ln \left[\frac{p_1^{(1)} + p_2^{(1)}}{2} \right] (p_1^{(1)} + p_2^{(1)}) < (p_1^{(1)} \ln p_1^{(1)} + p_2^{(1)} \ln p_2^{(1)})$, Eq. (D8) implies, for small enough ϵ ,

$$F^{(2)} < F^{(1)}. \quad (D9)$$

By inspecting Eqs. (D5) and (D9), we see that the distance D increases while the free energy F decreases or vice versa.

[1] H. B. Callen, *Thermodynamics* (Wiley, Hoboken, 1960).
 [2] H. D. Zeh, *The Physical Basis of the Direction of Time* (Springer, Berlin, 2007).
 [3] J. M. R. Parrondo, C. Van den Broeck, and R. Kawai, *New J. Phys.* **11**, 073008 (2009).
 [4] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2007).
 [5] S. Dattagupta, *Relaxation Phenomena in Condensed Matter Physics* (Academic, New York, 1987).
 [6] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *SIAM Ref.* **51**, 661 (2009).
 [7] J. K. Chung, P. L. Kannappan, C. T. Ng, and F. K. Sahoo, *J. Math. Anal. Appl.* **138**, 280 (1989).
 [8] S. Kullback and R. Leibler, *Ann. Math. Stat.* **22**, 79 (1951); S. Kullback, *Information Theory and Statistics* (Dover, New York, 1997).
 [9] B. Berckmoes, R. Lowen, and J. Van Casteren, *J. Math. Anal. Appl.* **374**, 412 (2011).
 [10] R. J. Budzyński, W. Kondrecki, and A. Królak, *Classical Quantum Gravity* **25**, 015005 (2008).
 [11] A. Plastino, A. R. Plastino, and H. G. Miller, *Phys. Lett. A* **235**, 129 (1997).
 [12] From its definition, the total variation distance ranges from 0 (if $p_m = q_m$ for all m) to 1 (if $q_m = 0$ for any $p_m > 0$ and $p_m = 0$ for any $q_m > 0$), and the ensemble of probability distributions is a metric space with respect to it, i.e., $D[p, q] = 0$ only if $p = q$, $D[p, q] = D[q, p]$, and $D[p, q] + D[q, r] \geq D[p, r]$.
 [13] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, RI, 2009).
 [14] S. Lalley (unpublished).
 [15] The vanishing of $D(\lambda)$ obviously implies that each $|p_m^{(eq)} - p_m(\lambda)|$ vanishes, but not necessarily in a monotonic way.
 [16] S. Chandrasekhar, *Rev. Mod. Phys.* **15**, 1 (1943).
 [17] K. Tomita and H. Tomita, *Prog. Theor. Phys.* **51**, 1731 (1974).
 [18] J. P. Bouchaud, *J. Phys.* **12**, 1705 (1992).
 [19] L. Leuzzi and T. M. Nieuwenhuizen, *Thermodynamics of the Glassy State* (Taylor and Francis, New York, 2008).
 [20] E. DelRe, B. Crosignani, P. Di Porto, and S. Di Sabatino, *Phys. Rev. E* **84**, 021112 (2011).
 [21] P. Di Porto, S. Di Sabatino, B. Crosignani, and E. DelRe, *J. Stat. Phys.* **153**, 479 (2013).
 [22] K. Christensen, L. Danon, T. Scanlon, and P. Bak, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2509 (2002).
 [23] F. Lillo and R. N. Mantegna, *Phys. Rev. E* **68**, 016119 (2003).
 [24] D. Sornette, in *Encyclopedia of Complexity and Systems Science* (Springer, Berlin, 2009), p. 7009.
 [25] A. Trabesinger, *Nat. Phys.* **8**, 13 (2012).
 [26] M. Shiino, *J. Phys. Soc. Jpn.* **67**, 3658 (1998).
 [27] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).
 [28] P. Diaconis, *Group Representations in Probability and Statistics* (Stanford University, Stanford, CA, 1987).