Infinite family of second-law-like inequalities

Carlos Pérez-Espigares,[1,*] Alejandro B. Kolton,[2,†] and Jorge Kurchan[3,‡]

[1]*Departamento de Electromagnetismo y Física de la Materia, Universidad de Granada, 18071 Granada, Spain*
[2]*CONICET, Centro Atómico Bariloche, 8400 San Carlos de Bariloche, Río Negro, Argentina*
[3]*PMMH-ESPCI, CNRS UMR 7636, 10 rue Vauquelin, 75005 Paris, France*

The probability distribution function for an out of equilibrium system may sometimes be approximated by a physically motivated "trial" distribution. A particularly interesting case is when a driven system (e.g., active matter) is approximated by a thermodynamic one. We show here that every set of trial distributions yields an inequality playing the role of a generalization of the second law. The better the approximation is, the more constraining the inequality becomes: this suggests a criterion for its accuracy, as well as an optimization procedure that may be implemented numerically and even experimentally. The fluctuation relation behind this inequality, a natural and practical extension of the Hatano-Sasa theorem, does not rely on the *a priori* knowledge of the stationary probability distribution.

I. INTRODUCTION

A recurring strategy applied to out of equilibrium systems is to represent the complex energy and dissipation sources by a bath with good equilibrium thermal properties. Two examples are the Edwards "thermodynamic" approach to granular matter [1] and recent developments for active matter (see Refs. [2–4] for recent examples), in which the combination of rapid energy bursts and friction is mimicked by a thermal bath. The aim of such pursuits is not necessarily to make the problem more easily solvable, but rather to cast it in a form that provides thermodynamic intuition and constraints. In this paper we derive some simple relations that help make this mapping more systematic and controlled. The method is based on the use of inequalities of the form of the second law, associated with each guess for the distribution function.

In these last two decades there has been a development of a family of relations valid for out of equilibrium systems [5], including the fluctuation theorem in its various forms [6–9] and the Jarzynski [10] and Crooks [11] relations. A later and extremely simple result is the Hatano-Sasa equality [12], which applies to systems that are continuously driven by time-dependent potentials inducing currents, so that even the stationary regimes are out of equilibrium. Their result is essentially a version of Jarzynski equality and the second principle, but with the energy replaced by the logarithm of the stationary distribution.

In this paper we derive an extension of the Hatano-Sasa theorem for Markovian systems, which has the practical advantage that it *does not rely on the a priori knowledge of the stationary probability distribution*. Instead, arbitrary "trial" smooth distributions can be used, thus allowing one to treat systems whose stationary distribution is either (i) too difficult to calculate, as generically occurs in out of equilibrium systems with many degrees of freedom, or (ii) unwieldy, as, for

instance, in the deterministic limit, where the nonequilibrium steady-state distributions are nonzero only over a fractal support. Our approach leads in particular to an inequality that can be used as a variational principle for improving, in a controlled way, physically motivated approximations to nonequilibrium steady-state distributions. The optimization procedure might be implemented numerically or even experimentally. As an illustration, we numerically approximate the stationary distribution of a paradigmatic nonequilibrium driven system with many degrees of freedom, the simple symmetric exclusion process in one dimension.

Just as in the case of the Hatano-Sasa equality [12,13], there is a "dual" (or adjoint) "backward" process that yields, when compared to the forward process of the original dynamics, a trajectory-dependent quantity playing the role of an entropy production that satisfies a form of the fluctuation theorem. For systems described by a Langevin dynamics, the dual backward process is obtained easily since it is given by a Langevin dynamics involving only additional *a priori known* external forces derived from the trial function itself. This remarkable property offers the possibility to explore numerically or even experimentally the interesting consequences of the associated detailed fluctuation relations, valid for systems that spontaneously relax to nonequilibrium steady states.

The organization of this paper is as follows. In Sec. II we review the derivation of the Hatano-Sasa fluctuation theorem [12]. After motivating a more general approach we provide in Sec. III A a first derivation of the integral version of our fluctuation relation along the same lines of the original Hatano-Sasa derivation. In Sec. III B we give a second, more general derivation, which yields the detailed version of the theorem (containing the integral version as a particular case), and in Sec. III C we discuss the physical interpretation of the dual dynamics behind it. In Sec. III D we discuss a family of inequalities that plays the role of the second law. In Sec. IV we propose an optimization procedure for approximating steady-state distributions. As an example, we apply it to the paradigmatic symmetric exclusion process in one dimension. In Sec. V we give a conclusion and perspectives.

---

*cpespigares@onsager.ugr.es
†koltona@cab.cnea.gov.ar
‡jorge@pmmh.espci.fr

II. THE HATANO-SASA RELATION

Consider a driven system with dynamic variables $\mathbf{x}$ with time-dependent external fields $\alpha(t)$ (e.g., shear rate, temperature gradient, etc.), with distribution $\rho(\mathbf{x},t)$ evolving through a generator $H(\alpha)$:

$$\dot{\rho}(\mathbf{x},t) = -H(\alpha(t))\,\rho(\mathbf{x},t). \tag{1}$$

Let us assume the dynamics admit, for every fixed value of the parameter α, a nonequilibrium steady state with distribution

$$\rho_{ss}(\mathbf{x};\alpha) = e^{-\phi(\mathbf{x};\alpha)}; \quad H(\alpha)\rho_{ss}(\mathbf{x};\alpha) = 0. \tag{2}$$

The Hatano-Sasa [12] result may be written

$$\left\langle e^{-\int_0^\tau dt\, \frac{\partial\phi(\mathbf{x};\alpha)}{\partial\alpha}\dot{\alpha}} \right\rangle_{\rho_{ss}(\mathbf{x};\alpha_1)} = 1, \tag{3}$$

which implies, by virtue of Jensen's inequality,

$$\left\langle \int_{\alpha_1}^{\alpha(\tau)} \frac{\partial\phi(\mathbf{x};\alpha)}{\partial\alpha}\,d\alpha \right\rangle_{\rho_{ss}(\mathbf{x};\alpha_1)} \geqslant 0. \tag{4}$$

The average $\langle\cdot\rangle$ is over all trajectories $\mathbf{x}(t)$ of duration τ, starting with an initial configuration chosen with the distribution $\rho_{ss}(\mathbf{x};\alpha_1)$ with $\alpha_1 \equiv \alpha(0)$. We shall refer to (3) and (4) as the Hatano-Sasa equality and inequality, respectively. In the particular case in which the stationary states $\rho_{ss}(\mathbf{x};\alpha)$ are Gibbs states we have

$$\rho_{ss}(\mathbf{x};\alpha) = \frac{e^{-\beta E(\mathbf{x},\alpha)}}{e^{-\beta F(\alpha)}}; \quad \phi(\mathbf{x};\alpha) = \beta[E(\mathbf{x},\alpha) - F(\alpha)], \tag{5}$$

and the Hatano-Sasa equality and inequality become the Jarzynski equality and the second law, respectively.

The proof is extremely simple. We start by decomposing the evolution in a large number M of time steps and compute, in operator (bra-ket) formalism, the quantity

$$\langle -|e^{-\frac{\tau}{M}H(\alpha_M)}\frac{\hat{\rho}_{ss}(\alpha_M)}{\hat{\rho}_{ss}(\alpha_{M-1})}\cdots\frac{\hat{\rho}_{ss}(\alpha_3)}{\hat{\rho}_{ss}(\alpha_2)}e^{-\frac{\tau}{M}H(\alpha_2)}$$
$$\frac{\hat{\rho}_{ss}(\alpha_2)}{\hat{\rho}_{ss}(\alpha_1)}e^{-\frac{\tau}{M}H(\alpha_1)}|\rho_{ss}(\alpha_1)\rangle = \langle -|\rho_{ss}(\alpha_M)\rangle = 1, \tag{6}$$

where $e^{-\frac{\tau}{M}H(\alpha_i)}|\rho_{ss}(\alpha_i)\rangle = |\rho_{ss}(\alpha_{i+1})\rangle$. We denote $\hat{\rho}(\alpha)$ as the operator such that $\langle\mathbf{x}|\hat{\rho}(\alpha)|\mathbf{x}'\rangle = \delta(\mathbf{x}-\mathbf{x}')\,\rho(\mathbf{x};\alpha)$ and $|\rho(\alpha)\rangle$ as the state such that $\rho(\mathbf{x};\alpha) \equiv \langle\mathbf{x}|\rho(\alpha)\rangle$. The state $|-\rangle$ corresponds to the flat distribution $\langle\mathbf{x}|-\rangle = 1$; i.e., the left eigenvector of $H(\alpha)$ has zero eigenvalue.

Now, using that the time step τ/M is small, we can write

$$\frac{\rho_{ss}(\mathbf{x};\alpha_{k+1})}{\rho_{ss}(\mathbf{x};\alpha_k)} \approx e^{-\frac{\tau}{M}\frac{\partial\phi(\mathbf{x};\alpha_k)}{\partial\alpha}\dot{\alpha}}. \tag{7}$$

Therefore (6) may be written as Eq. (3), and the result is proven. The exponential of the term $[-\int dt\,\frac{\partial\phi(\mathbf{x};\alpha)}{\partial\alpha}\dot{\alpha}]$, a functional of the trajectory, is thus weighted in (3) with the probability of each dynamical trajectory $\mathbf{x}(t)$ such that $\mathbf{x}(0)$ is sampled from $\rho_{ss}(\mathbf{x},\alpha_1)$.

In the Hatano-Sasa inequality (4), the equality holds in the quasistationary limit, when the probability distribution may be assumed to be, at each time t, the stationary one $\rho_{ss}(\mathbf{x};\alpha)$

corresponding to the value of α at that time:

$$\left\langle \int \frac{\partial\phi(\mathbf{x};\alpha)}{\partial\alpha}\,d\alpha \right\rangle = \int d\mathbf{x}\,d\alpha\,e^{-\phi(\mathbf{x};\alpha)}\frac{\partial\phi(\mathbf{x};\alpha)}{\partial\alpha}$$
$$= \int d\mathbf{x}[\rho_{ss}(\mathbf{x};\alpha_M) - \rho_{ss}(\mathbf{x};\alpha_1)] = 0. \tag{8}$$

This result is the generalization of the entropy change $S(\alpha_M) - S(\alpha_1) = \langle\int_0^\tau dt\,\dot{\mathbf{x}}\cdot\nabla\phi\rangle$ under reversible transformations, with $S(\alpha) \equiv -\int d\mathbf{x}\rho_{ss}(\mathbf{x};\alpha)\ln\rho_{ss}(\mathbf{x};\alpha)$ being the generalized Shannon entropy [12].

III. A MORE GENERAL APPROACH

The quantity $\phi(\mathbf{x};\alpha) = -\ln\rho_{ss}(\mathbf{x};\alpha)$ plays a role similar to the one of the energy function in a system with detailed balance, but it may become intractable as soon as we consider a driven system. The first difficulty is that it is, in general, impossible to obtain analytically. This is aggravated by the fact that in order to use (3) and (4), we need to know ρ_{ss} also where it is exponentially small. Another more serious problem arises from the fact that the function $\phi(\mathbf{x};\alpha)$ may only be small in a limited domain and very large everywhere else. An extreme form of this situation arises in the deterministic limit. Consider a noisy dynamics with a (Hoover [14]) thermostat:

$$\dot{q}_i = p_i,$$
$$\dot{p}_i = -\frac{\partial\mathcal{H}}{\partial q_i} + \underbrace{\gamma(t)p_i}_{\text{thermostat}} - \underbrace{f_i(\mathbf{q})}_{\text{forcing}} - \underbrace{\eta_i(\mathbf{t})}_{\text{noise}}, \tag{9}$$

where $\eta(\mathbf{t})$ is a Gaussian white noise of variance ϵ. Energy is conserved provided $\gamma(t) = \frac{(\mathbf{f}+\eta)\cdot\mathbf{p}}{\mathbf{p}^2}$. When there is forcing $\mathbf{f} \neq 0$, the stationary distribution is not flat. Indeed, in the limit of zero noise $\epsilon \to 0$, ρ_{ss} has, in fact, fractal support, and $\phi(\mathbf{x};\alpha)$ is infinity almost everywhere on the energy surface. If we attempt to apply the Hatano-Sasa inequality for a small noise amplitude in a process with varying α, because the region on the energy shell where $\phi(\mathbf{x};\alpha)$ is small is sparse and strongly dependent on α, almost all of the process takes place in regions in which $\phi(\mathbf{x};\alpha)$ is large: the trajectories are very far from quasistationary, and the Hatano-Sasa inequality, though true, becomes useless.

A similar situation arises when the potential is rapidly oscillating, as in vibrated granular matter, which we may think of as subjected to an oscillating gravity field. Here again, the system is always very far from the stationary situation corresponding to any instantaneous value of the field because it does not have the time to catch up with the oscillating stationary measure. And yet we still observe that rapidly vibrated granular matter behaves in a manner that resembles motion in contact with a heat bath and would expect some form of second law to apply in that case.

With the above motivations, we look for a more flexible approach. Instead of working with the true stationary distributions $\rho_{ss}(\mathbf{x};\alpha)$, we choose an arbitrary family of smooth functions as reference states $\rho_{\text{ref}}(\mathbf{x};\alpha)$ and the corresponding $\phi_{\text{ref}}(\mathbf{x};\alpha) \equiv -\ln\rho_{\text{ref}}(\mathbf{x};\alpha)$. In the following we derive an extension of the Hatano-Sasa integral and detailed fluctuation relations, using only these smooth functions.

A. Integral fluctuation theorem

In order to obtain a relation, we go through the same steps as in Sec. II. Starting from the initial distribution $\phi_{\text{ref}}(\mathbf{x}; \alpha_1)$, we compute, just as in (6),

$$\langle -|e^{-\frac{\tau}{M}H(\alpha_M)} \frac{\hat{\rho}_{\text{ref}}(\alpha_M)}{\tilde{\rho}_{\text{ref}}(\alpha_{M-1})} \cdots \frac{\hat{\rho}_{\text{ref}}(\alpha_2)}{\tilde{\rho}_{\text{ref}}(\alpha_1)} e^{-\frac{\tau}{M}H(\alpha_1)}|\rho_{\text{ref}}(\alpha_1)\rangle$$
$$= \langle -|\rho_{\text{ref}}(\alpha_M)\rangle = 1, \qquad (10)$$

but with $\tilde{\rho}(\alpha)$ being the operator associated with the state evolved by one time step $e^{-\frac{\tau}{M}H(\alpha)}|\rho(\alpha)\rangle$. We can thus write, for large M,

$$\frac{\tilde{\rho}_{\text{ref}}(\mathbf{x}; \alpha)}{\rho_{\text{ref}}(\mathbf{x}; \alpha)} \approx e^{\frac{\tau}{M}\varphi(\mathbf{x};\alpha)}, \qquad (11)$$

with

$$\varphi(\mathbf{x}; \alpha) \equiv -\frac{1}{\rho_{\text{ref}}(\mathbf{x}; \alpha)} \{H(\alpha)\,\rho_{\text{ref}}(\mathbf{x}; \alpha)\}. \qquad (12)$$

Here $H(\alpha)$ acts over the function $\rho_{\text{ref}}(\mathbf{x}; \alpha)$, so that it is in fact $\varphi(\mathbf{x}; \alpha) = -\langle \mathbf{x}|\frac{1}{\hat{\rho}_{\text{ref}}(\alpha)}H(\alpha)|\rho_{\text{ref}}(\alpha)\rangle$. We hence have

$$\frac{\rho_{\text{ref}}(\mathbf{x}; \alpha_{k+1})}{\tilde{\rho}_{\text{ref}}(\mathbf{x}; \alpha_k)} \approx e^{-\frac{\tau}{M}\frac{\partial \phi_{\text{ref}}(\mathbf{x};\alpha_k)}{\partial \alpha}\dot{\alpha} - \frac{\tau}{M}\varphi(\mathbf{x},\alpha)}, \qquad (13)$$

and we obtain a new equality, valid for all sets $\phi_{\text{ref}}(\mathbf{x}; \alpha)$,

$$\left\langle e^{-\int dt \frac{\partial \phi_{\text{ref}}(\mathbf{x};\alpha)}{\partial \alpha}\dot{\alpha} - \int dt\, \varphi(\mathbf{x};\alpha)}\right\rangle_{\rho_{\text{ref}}(\mathbf{x};\alpha_1)} = 1, \qquad (14)$$

which is the first main result of our paper. Defining

$$\mathcal{Y} \equiv \int dt\, \frac{\partial \phi_{\text{ref}}(\mathbf{x}; \alpha)}{\partial \alpha}\,\dot{\alpha} + \int dt\, \varphi(\mathbf{x}; \alpha), \qquad (15)$$

it can be simply written as $\langle e^{-\mathcal{Y}}\rangle = 1$. This integral fluctuation theorem is valid for any protocol $\alpha(t)$ and arbitrary times τ, like the Hatano-Sasa equality, to which it reduces if the reference state is chosen as $\rho_{\text{ref}}(\mathbf{x}; \alpha) = \rho_{ss}(\mathbf{x}; \alpha)$, but it holds for arbitrary smooth functions ρ_{ref} [15]. As we shall see, this immediately implies an inequality $\langle \mathcal{Y}\rangle \geqslant 0$ of the form of the second law.

B. Detailed fluctuation theorem

Just as in the case of the Hatano-Sasa relation, the result of Eq. (14) can be alternatively derived as a particular case of a detailed fluctuation theorem. We will use here a procedure that generalizes the one used for obtaining the detailed fluctuation theorem associated with the Hatano-Sasa theorem [12,13,16,17].

We are looking for a time-reversed form of the dynamics. Let us start by rewriting (10) as

$$\langle -|e^{-\frac{\tau}{M}H(\alpha_M)} \frac{\hat{\rho}_{\text{ref}}(\alpha_M)}{\tilde{\rho}_{\text{ref}}(\alpha_{M-1})} \cdots \frac{\hat{\rho}_{\text{ref}}(\alpha_2)}{\tilde{\rho}_{\text{ref}}(\alpha_1)} e^{-\frac{\tau}{M}H(\alpha_1)}|\rho_{\text{ref}}(\alpha_1)\rangle$$
$$= \langle \rho_{\text{ref}}(\alpha_M)|\frac{1}{\tilde{\rho}_{\text{ref}}(\alpha_{M-1})} \cdots \frac{\hat{\rho}_{\text{ref}}(\alpha_2)}{\tilde{\rho}_{\text{ref}}(\alpha_1)} e^{-\frac{\tau}{M}H(\alpha_1)}\hat{\rho}_{\text{ref}}(\alpha_1)|-\rangle$$
$$= \langle \rho_{\text{ref}}(\alpha_M)|\Pi_{k=1}^{M-1}\left[\frac{1}{\tilde{\rho}_{\text{ref}}(\alpha_k)} e^{-\frac{\tau}{M}H(\alpha_k)}\hat{\rho}_{\text{ref}}(\alpha_k)\right]|-\rangle.$$

We may now take the adjoint in order to reverse time:

$$\langle -|\Pi_{k=M-1}^{1}\left[\frac{1}{\tilde{\rho}_{\text{ref}}(\alpha_k)} e^{-\frac{\tau}{M}H(\alpha_k)}\hat{\rho}_{\text{ref}}(\alpha_k)\right]^{\dagger}|\rho_{\text{ref}}(\alpha_M)\rangle$$
$$= \langle -|\Pi_{k=M-1}^{1}\left[\frac{1}{\hat{\rho}_{\text{ref}}(\alpha_k)} e^{-\frac{\tau}{M}\{H(\alpha_k)+\varphi(\alpha_k)\}}\hat{\rho}_{\text{ref}}(\alpha_k)\right]^{\dagger}$$
$$|\rho_{\text{ref}}(\alpha_M)\rangle = \langle -|\Pi_{k=M-1}^{1}\left[e^{-\frac{\tau}{M}H^{adj}(\alpha_k)}\right]|\rho_{\text{ref}}(\alpha_M)\rangle.$$

This is a time-reversed dynamics with generator:

$$[H^{adj}(\alpha)]^{\dagger} \equiv \frac{1}{\hat{\rho}_{\text{ref}}(\alpha)} \{H(\alpha) + \varphi(\alpha)\}\,\hat{\rho}_{\text{ref}}(\alpha). \qquad (16)$$

We shall see below that it corresponds, in fact, to a Langevin process with a modified force field [see Eq. (27)].

In terms of the original and the adjoint dynamics, the evolution in time step τ/M is

$$P(\mathbf{x}'|\mathbf{x}; \alpha) \equiv \langle \mathbf{x}'|e^{-\frac{\tau}{M}H_\alpha}|\mathbf{x}\rangle, \qquad (17)$$

$$P^{adj}(\mathbf{x}'|\mathbf{x}; \alpha) \equiv \langle \mathbf{x}'|e^{-\frac{\tau}{M}H_\alpha^{adj}}|\mathbf{x}\rangle. \qquad (18)$$

The construction (16) tells us that, for each trajectory $\mathcal{T} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ with the initial condition chosen with probability $\rho_{\text{ref}}(\mathbf{x}_1; \alpha_1)$, there is a time-reversed (R) trajectory, with the initial condition chosen with probability $\rho_{\text{ref}}(\mathbf{x}_M; \alpha_M)$, and their respective weights are

$$\mathcal{P}[\mathcal{T}; \alpha] = \prod_{n=1}^{M-1} P(\mathbf{x}_{n+1}|\mathbf{x}_n; \alpha_n)\rho_{\text{ref}}(\mathbf{x}_1; \alpha_1)$$

and

$$[\mathcal{P}^{adj}[\mathcal{T}; \alpha]]^R = \prod_{n=1}^{M-1} P^{adj}(\mathbf{x}_n|\mathbf{x}_{n+1}; \alpha_n)\rho_{\text{ref}}(\mathbf{x}_M; \alpha_M).$$

We thus may define a quantity $\Xi[\mathcal{T}; \alpha]$ associated with each path, having an interpretation analogous to the *entropy production*,

$$\Xi[\mathcal{T}; \alpha] \equiv \ln \frac{\mathcal{P}[\mathcal{T}, \alpha]}{[\mathcal{P}^{adj}[\mathcal{T}, \alpha]]^R}. \qquad (19)$$

In the large M limit, it becomes

$$\Xi[\mathcal{T}; \alpha] \approx \int_0^\tau dt\, [\varphi(\mathbf{x}; \alpha) + \dot{\alpha}\partial_\alpha \phi_{\text{ref}}(\mathbf{x}, \alpha)]. \qquad (20)$$

It is clear from Eq. (19) that, in terms of Ξ, there is a detailed fluctuation theorem:

$$\langle \mathcal{O}[\mathcal{T}]e^{-\Xi[\mathcal{T},\alpha]}\rangle = \langle \mathcal{O}[\mathcal{T}]e^{-\int_0^\tau dt\, [\varphi(\mathbf{x};\alpha)+\dot{\alpha}\partial_\alpha \phi_{\text{ref}}(\mathbf{x},\alpha)]}\rangle$$
$$= [\langle \mathcal{O}[\mathcal{T}]\rangle^{adj}]^R, \qquad (21)$$

which is valid for an arbitrary functional $\mathcal{O}[\mathcal{T}]$ of the trajectory. The averages in (21) are performed with the real forward dynamics in the first term and with the time-reversed adjoint dynamics of Eq. (18) in the second term.

Equation (21) is a very general result. It represents a broad family of fluctuation theorems with a trajectory dependent "entropy production" of the form of Eq. (20), completely determined by the distributions $\rho_{\text{ref}}(\mathbf{x}; \alpha)$.

Clearly, choosing $\mathcal{O} = 1$ in this equation we get the integral fluctuation relation of Eq. (14). This detailed fluctuation

theorem, which can be used to derive a variety of Crooks-like relations, is the second main result of our paper.

C. Generalized dual (adjoint) dynamics

In order to give a simple physical interpretation of the dual dynamics let us now assume that our system is governed by a Langevin equation,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\alpha) + \boldsymbol{\xi}(t), \qquad (22)$$

with $\mathbf{f}(\mathbf{x};\alpha)$ being an arbitrary force (conservative or non-conservative) and $\boldsymbol{\xi}(t)$ being a Gaussian uncorrelated noise at temperature T, such that $\langle \boldsymbol{\xi}(t) \rangle = 0$ and $\langle \xi_n(t)\xi_m(t') \rangle = 2T\delta(t-t')\delta_{nm}$. This is associated with the Fokker-Planck process:

$$\frac{d\rho}{dt} = \boldsymbol{\nabla} \cdot [[T\boldsymbol{\nabla} - \mathbf{f}(\mathbf{x};\alpha)]\rho] = -H(\alpha)\rho. \qquad (23)$$

Using Eq. (12), φ is given in this case by

$$\varphi = -\boldsymbol{\nabla} - T\nabla^2\phi_{\text{ref}} + T\,|\boldsymbol{\nabla}\phi_{\text{ref}}|^2 + \mathbf{f} \cdot \boldsymbol{\nabla}\phi_{\text{ref}}. \qquad (24)$$

It is easy to check that, if $e^{-\phi_{\text{ref}}}$ is the stationary distribution, this expression vanishes.

The expression for a path probability is

$$\mathcal{P}[\mathcal{T};\alpha] \sim e^{-\frac{1}{4T}\int_0^\tau dt\,\{[\dot{\mathbf{x}}-\mathbf{f}(\mathbf{x};\alpha)]^2 + 4T\frac{\nabla\cdot\mathbf{f}}{2}\}}. \qquad (25)$$

The last term in the argument of the integral comes from the Stratonovich discretization scheme. Then, using equations (19) and (15) and time reversing in order to obtain the dynamical weight (that is, $[\mathcal{P}e^{-\mathcal{Y}}]^R = [[\mathcal{P}^{adj}]^R]^R = \mathcal{P}^{adj}$), we have

$$\mathcal{P}^{adj}[\mathbf{x};\alpha] \sim e^{-\frac{1}{4T}\int_0^\tau dt\,[(\dot{\mathbf{x}}+\mathbf{f}(\mathbf{x};\alpha))^2 + 4T(\varphi(\mathbf{x},\alpha) - \dot{\alpha}\partial_\alpha\phi_{\text{ref}}(\mathbf{x},\alpha) + \frac{\nabla\cdot\mathbf{f}}{2})]}$$

$$\sim e^{-\frac{1}{4T}\int_0^\tau dt\,[(\dot{\mathbf{x}}+\mathbf{f}(\mathbf{x};\alpha))^2 + 4T\varphi(\mathbf{x},\alpha) + 4T\dot{\mathbf{x}}\cdot\boldsymbol{\nabla}\phi_{\text{ref}}(\mathbf{x},\alpha) + 2T\boldsymbol{\nabla}\cdot\mathbf{f}]}, \qquad (26)$$

where in the last step we have dropped all reference to the boundary term, which is irrelevant for our present purposes.

Is there a Langevin equation associated with the weight of Eq. (26)? In order to answer such a question, we follow a procedure analogous to the one used recently in Ref. [16] for the standard dual dynamic weight. Plugging expression (24) into Eq. (26) leads to a simple expression,

$$\mathcal{P}^{adj}[\mathcal{T};\alpha] \sim e^{-\frac{1}{4T}\int_0^\tau dt\,[(\dot{\mathbf{x}}+\mathbf{f}+2T\boldsymbol{\nabla}\phi_{\text{ref}})^2 - 2T\boldsymbol{\nabla}\cdot[\mathbf{f}+2T\boldsymbol{\nabla}\phi_{\text{ref}}]]},$$

where we can clearly identify the action of the following Langevin equation (in Stratonovich scheme):

$$\dot{\mathbf{x}} = -\mathbf{f}(\mathbf{x};\alpha) - 2T\boldsymbol{\nabla}\phi_{\text{ref}}(\mathbf{x};\alpha) + \boldsymbol{\xi}(t). \qquad (27)$$

The dual (adjoint) dynamics corresponds to a Langevin process with opposite force and an additional external potential $\phi_{\text{ref}}(\mathbf{x};\alpha)$, which depends on the choice of ρ_{ref}.

All the results obtained so far reduce to the Hatano-Sasa results if we choose $\rho_{\text{ref}} = \rho_{ss}$, in which case, $\varphi = 0$, Ξ becomes the Hatano-Sasa functional $\mathcal{Y}_{HS} = \int_0^\tau dt\,\dot{\alpha}\partial_\alpha\phi(\mathbf{x},\alpha)$, and the extended dual dynamics becomes the well-known (†) standard dual dynamics [12,13,17,18], which in terms of transition probabilities reads $P^\dagger(\mathbf{x}|\mathbf{x}';\alpha) \equiv P(\mathbf{x}'|\mathbf{x};\alpha)\frac{\rho_{ss}(\mathbf{x},\alpha)}{\rho_{ss}(\mathbf{x}';\alpha)}$, as can easily be obtained from Eq. (18). The Langevin equation for the usual Hatano-Sasa dual dynamics (see, for instance, its

derivation in Ref. [16]) coincides with Eq. (27), replacing ϕ_{ref} by $\phi = -\ln\rho_{ss}$.

Finally, it is worth noting that the extended dual dynamics derived above has the advantage over the standard dual dynamics that all the forces are known, so that it might be implemented in practice, numerically or even experimentally, by applying appropriate external fields. It should thus be possible to verify, numerically or experimentally, the detailed fluctuation theorem of Eq. (19) as well as other *extended* Crooks-like [11] relations that easily follow from Eq. (19) and concern systems with nonequilibrium steady states. It is with this practical motivation that in the Appendix we derive the corresponding extended version of the three detailed fluctuation theorems of Esposito and Van der Broeck [13]. It would also be interesting to explore further the implications of the extended dual dynamics, generalizing the results based on dual dynamics approaches in Refs. [12,13,17,18].

D. Generalizations of the second law

As we did in Sec. II we us Jensen's inequality in Eq. (14) to obtain

$$\left\langle \int \frac{\partial\phi_{\text{ref}}(\mathbf{x};\alpha)}{\partial\alpha}d\alpha + \int dt\,\varphi(\mathbf{x};\alpha) \right\rangle_{\rho_{\text{ref}}(\mathbf{x};\alpha_1)} \geqslant 0. \qquad (28)$$

It is worth emphasizing that the time evolution implied in the brackets is the real dynamics with initial states drawn from the distribution $\rho_{\text{ref}}(\mathbf{x};\alpha_1)$. The relation is true for arbitrary $\rho_{\text{ref}}(\mathbf{x};\alpha)$; a bad choice only makes the inequality less constraining. This is the third main result of our paper and the central formula we will exploit for applications. The function $\varphi(\mathbf{x};\alpha)$ is a known, well-behaved extensive function of the dynamic variables, which vanishes if $\rho_{\text{ref}}(\mathbf{x};\alpha) = \rho_{ss}(\mathbf{x};\alpha)$. For example, for a Langevin process (22) it is given by Eq. (24).

If at constant α the system is able to converge to a stationary nonequilibrium regime, the inequality has to hold for large times such that the initial condition is forgotten. We thus get the stationary-state expectation:

$$\langle\varphi\rangle_{ss} = -\left\langle \frac{1}{\rho_{\text{ref}}}H\,\rho_{\text{ref}} \right\rangle_{ss} \geqslant 0 \quad \forall\rho_{\text{ref}}. \qquad (29)$$

This inequality is already implicit in the work of Lebowitz and Bergmann [19]. If we define $\dot{w}_{\text{ref}} = -Hw_{\text{ref}}$ with $w_{\text{ref}}(t = 0) = \rho_{\text{ref}}$, we can rewrite (29) as

$$\langle\varphi\rangle_{ss} = -\left[\frac{d}{dt}\int d\mathbf{x}\,\rho_{ss}\ln\left(\frac{\rho_{ss}}{w_{\text{ref}}(t)}\right)\right]_{t=0}$$

$$= -\left[\frac{d}{dt}D_{KL}(\rho_{ss}\,\|\,w_{\text{ref}}(t))\right]_{t=0} \geqslant 0 \qquad (30)$$

by virtue of the general result $\dot{D}_{KL}(w_1(t)\,\|\,w_2(t)) \leqslant 0$, valid for all times $t \geqslant 0$ with $w_1(t)$ and $w_2(t)$ being any two distributions evolving through H [19]. The positively defined Kullback-Leibler distance D_{KL} used above is often an actor in these problems; see [13,20–22].

For a purely Hamiltonian system $\dot{D}_{KL}(w_1(t)\,\|\,w_2(t)) = 0$ independently of w_1 and w_2: irreversibility in this case inescapably requires some form of coarse graining, which this method does not provide. Instead, in the case of a Langevin

process (23), a short computation [23] gives

$$\langle \varphi \rangle_{ss} = T \langle |\nabla(\phi_{\texttt{ref}} - \phi)|^2 \rangle_{ss} \geqslant 0, \qquad (31)$$

where $\phi = -\ln \rho_{ss}$. We have easy access to the left-hand side of the above equation numerically *or even experimentally* because we know $\phi_{\texttt{ref}}$ and the dynamics, but we do not have easy access to the right-hand side.

Let us consider now a system that is perturbed periodically, such as the granular system described above. Assume further that the system reaches, after a long time, a periodic state. We then have

$$\left\langle \oint \frac{\partial \phi_{\texttt{ref}}(\mathbf{x}; \alpha)}{\partial \alpha} d\alpha + \oint dt \, \varphi(\mathbf{x}; \alpha) \right\rangle \geqslant 0, \qquad (32)$$

where the time integral is over one cycle in the regime in which the distribution becomes periodic in time. If we make the further simplification that $\rho_{\texttt{ref}}$ is constant in time, we get

$$\left\langle \oint dt \, \varphi(\mathbf{x}; \alpha) \right\rangle \geqslant 0, \qquad (33)$$

where the dependence of φ on α comes from $H(\alpha)$.

IV. A VARIATIONAL SCHEME

The preceding section suggests an iterative variational procedure to optimize $\rho_{\texttt{ref}}$ at fixed α: (i) Propose a change to $\rho_{\texttt{ref}}(\mathbf{x})$, (ii) compute the new $\varphi = -\rho_{\texttt{ref}}^{-1} H \rho_{\texttt{ref}}$ (immediate), and (iii) run $\langle \varphi(\mathbf{x}) \rangle_{ss}$ and accept the change if the result is smaller. The resulting φ yields directly a second-law-like constraint, which is optimized. The optimization procedure we propose might be indeed implemented numerically or even experimentally to calculate, for instance, optimal effective interactions from steady-state measurements [24].

A. An application

As an illustrative and nontrivial example, we consider the simple symmetric exclusion process (SSEP), a one-dimensional lattice of L sites that are either occupied by a single particle or empty. A configuration at time t is defined by the vector of occupation numbers $\mathbf{n}(t) = [n_1(t), \ldots, n_L(t)]$ $[n_i(t) = 0, 1]$. Each particle in the bulk independently attempts to jump to an empty site to its right or to its left. At the left boundary each particle is injected at site 1 at rate α and removed from site 1 at rate γ, whereas at the right boundary particles are injected at site L at rate δ and removed from site L at rate β.

The choice of the rates α, γ, δ, and β corresponds to the system being in contact with infinite left and right reservoirs at densities $\rho_0 = \alpha/(\alpha + \gamma)$ and $\rho_1 = \delta/(\delta + \beta)$, respectively [25]. If $\rho_0 = \rho_1 = \rho$, the system is in equilibrium, and the distribution is of product form: $\rho_{eq}(\mathbf{n}) = \prod_{i=1}^{L} \rho^{n_i}(1 - \rho)^{1-n_i} = e^{\sum_{i=1}^{L} \mu n_i}/(1 + e^\mu)^L$, where $\mu = \ln[\rho/(1 - \rho)]$ is the chemical potential. As soon as $\rho_0 \neq \rho_1$, a current is established, and the problem becomes nontrivial, with long-range correlations. The evolution of the probability $\rho(\mathbf{n})$ of observing a configuration $\mathbf{n}$ is given by the master equation

$(n_k^+ = n_k + 1$ and $n_k^- = n_k - 1)$

$$\begin{aligned}
\frac{\partial \rho(\mathbf{n})}{\partial t} =& \sum_{k=1}^{L-1} \big[\delta_{n_k,1} \delta_{n_{k+1},0} \rho(\ldots, n_k^-, n_{k+1}^+, \ldots) \\
&+ \delta_{n_k,0} \delta_{n_{k+1},1} \rho(\ldots, n_k^+, n_{k+1}^-, \ldots) \\
&- (\delta_{n_k,1} \delta_{n_{k+1},0} + \delta_{n_k,0} \delta_{n_{k+1},1}) \rho(\ldots, n_k, n_{k+1}, \ldots) \big] \\
&+ \alpha \delta_{n_1,1} \rho(n_1^-, \ldots) + \gamma \delta_{n_1,0} \rho(n_1^+, \ldots) \\
&+ \delta \delta_{n_L,1} \rho(\ldots, n_L^-) + \beta \delta_{n_L,0} \rho(\ldots, n_L^+) \\
&- (\gamma \delta_{n_1,1} + \alpha \delta_{n_1,0} + \beta \delta_{n_L,1} + \delta \delta_{n_L,0}) \rho(n_1, \ldots, n_L).
\end{aligned} \qquad (34)$$

The full measure on the microscopic configurations in the steady state $\rho_{ss}(\mathbf{n})$ may be computed analytically through the so-called matrix method [26]. Here we propose an approximate form $\phi_{\texttt{ref}}(\mathbf{n}) = \sum_i h_i n_i + \sum_{i \neq j} J_{ij} n_i n_j$. Using the master equation, we evaluate $\varphi = \frac{1}{\rho_{\texttt{ref}}(\mathbf{n})} \frac{\partial \rho_{\texttt{ref}}(\mathbf{n})}{\partial t}$ [Eq. (12)] as

$$\begin{aligned}
\varphi =& \sum_{k=1}^{L-1} \big[\delta_{n_k,1} \delta_{n_{k+1},0} e^{h_k - h_{k+1} + \sum_{j \neq k, k+1} 2(J_{kj} - J_{k+1,j}) n_j} \\
&+ \delta_{n_k,0} \delta_{n_{k+1},1} e^{h_{k+1} - h_k - \sum_{j \neq k, k+1} 2(J_{kj} - J_{k+1,j}) n_j} \\
&- (\delta_{n_k,1} \delta_{n_{k+1},0} + \delta_{n_k,0} \delta_{n_{k+1},1}) \big] \\
&+ \alpha \delta_{n_1,1} e^{h_1 + \sum_{j \neq 1} 2 J_{1j} n_j} + \gamma \delta_{n_1,0} e^{-h_1 - \sum_{j \neq 1} 2 J_{1j} n_j} \\
&+ \delta \delta_{n_L,1} e^{h_L + \sum_{j \neq L} 2 J_{Lj} n_j} + \beta \delta_{n_L,0} e^{-h_L - \sum_{j \neq L} 2 J_{Lj} n_j} \\
&- (\gamma \delta_{n_1,1} + \alpha \delta_{n_1,0} + \beta \delta_{n_L,1} + \delta \delta_{n_L,0}).
\end{aligned} \qquad (35)$$

We compute the expectation value of this φ with the true SSEP dynamics and minimize with respect to the $[h_i, J_{ij}]$ using a suitable algorithm [27]. Clearly, for $\rho_0 = \rho_1 = \rho$ the system is in equilibrium, and we have for each site $h_i = h = -\mu = \ln[(1 - \rho)/\rho]$ and $J_{ij} = 0$ [see $\rho_{eq}(\mathbf{n})$ above]. Unlike the equilibrium case, as soon as $\rho_0 \neq \rho_1$, we obtain nonzero J_{ij} corresponding to the long-range correlations characteristic of the stationary nonequilibrium state; see Fig. 1. These correlations extend over macroscopic distances and reflect the intrinsic nonadditivity of nonequilibrium systems [25]. The optimized measure $\rho_{\text{opt}}(\mathbf{n}) = e^{-\phi_{\text{opt}}(\mathbf{n})}$ obtained with the $[h_i, J_{ij}]$ that minimize the expectation value of φ is not the exact solution of [26], but we have checked the quality
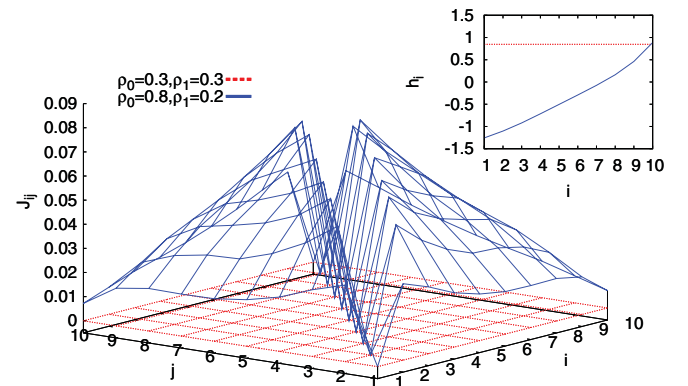

FIG. 1. (Color online) Optimized J_{ij} for the SSEP model with open boundaries for $\rho_0 = \rho_1$ (dotted red line) and for $\rho_0 \neq \rho_1$ (solid blue line). The inset shows similar results for the optimized h_i.

of the approximation by computing expectation values with this measure: this is most easily done with a Monte Carlo procedure with "energy" $\phi_{\text{opt}}(\mathbf{n})$. To do that one starts from a random initial configuration $\mathbf{n}(t=0)$ and evolves it with a Metropolis algorithm where the probability to go from a configuration $\mathbf{n}$ to a configuration $\mathbf{n}'$ in a single jump is $W(\mathbf{n} \rightarrow \mathbf{n}') = \min[\frac{\rho_{\text{opt}}(\mathbf{n}')}{\rho_{\text{opt}}(\mathbf{n})}, 1]$ (note that there are no reservoirs in this calculation). The configuration $\mathbf{n}'$ is the same as the configuration $\mathbf{n}$, except for the randomly chosen node k, which changes its value to $n'_k = 1 - n_k$. We then have

$$\frac{\rho_{\text{opt}}(\mathbf{n}')}{\rho_{\text{opt}}(\mathbf{n})} = \exp\left[(2n_k - 1)\left(h_k + 2\sum_{j \neq k} J_{kj}n_j\right)\right]. \quad (36)$$

Applying this dynamics we measured the steady-state density profile $\rho_i \equiv \langle n_i \rangle$ shown in Fig. 2 and compared it with the analytical result obtained using the exact stationary state measure $\rho_{ss}(\mathbf{n})$, which is (see [25])

$$\langle n_i \rangle = \frac{\rho_0\left(L + \frac{1}{\beta+\delta} - i\right) + \rho_1\left(i - 1 + \frac{1}{\alpha+\gamma}\right)}{L + \frac{1}{\alpha+\gamma} + \frac{1}{\beta+\delta} - 1}. \quad (37)$$

We also compared with the result obtained assuming local equilibrium considering no reservoirs at the boundaries and a spatially varying chemical potential, which is adjusted to maintain the same steady-state density profile (37). We then have that the local equilibrium measure is $\rho_{LE}(\mathbf{n}) = \prod_{i=1}^{L} e^{-h_i n_i}/(1 + e^{-h_i})$, where $h_i = -\mu_i = \ln[(1 - \langle n_i \rangle)/\langle n_i \rangle]$, with $\langle n_i \rangle$ given by (37). Notice that this local equilibrium measure for $\rho_0 \neq \rho_1$ turns into the equilibrium measure and by doing $\rho_0 = \rho_1$. In Fig. 2 we can see that there is perfect agreement with the exact analytical results for both $\rho_0 = \rho_1$ and $\rho_0 \neq \rho_1$ and, in the latter case, for the optimized measure and for the local equilibrium measure. We

FIG. 3. (Color online) Analytical and simulation results for the two-point correlation function $\langle n_i n_j \rangle_c$ in the steady state. Red squares and orange crosses correspond to the Monte Carlo procedure for densities $\rho_0 = 0.8$, $\rho_1 = 0.2$ using the optimized trial $\rho_{\text{opt}}(\mathbf{n})$ and the local equilibrium measure $\rho_{LE}(\mathbf{n})$, respectively, whereas blue diamonds are the analytical results; see Eq. (38). Purple triangles correspond to the Monte Carlo procedure for densities $\rho_0 = \rho_1 = 0.3$ using the equilibrium measure $\rho_{\text{eq}}(\mathbf{n})$, whereas black circles are the analytical results.

also measured the two-point correlation function $\langle n_i n_j \rangle_c \equiv \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$, obtaining the results shown in Fig. 3. Using again the exact measure $\rho_{ss}(\mathbf{n})$, one finds that the analytical prediction in the steady state for $1 \leqslant i < j \leqslant L$ is [25]

$$\langle n_i n_j \rangle_c = \frac{-(\rho_0 - \rho_1)^2\left(\frac{1}{\alpha+\gamma} + i - 1\right)\left(\frac{1}{\beta+\delta} + L - j\right)}{\left(\frac{1}{\alpha+\gamma} + \frac{1}{\beta+\delta} + L - 1\right)^2\left(\frac{1}{\alpha+\gamma} + \frac{1}{\beta+\delta} + L - 2\right)}. \quad (38)$$

For large L, introducing macroscopic coordinates $i = Lx$ and $j = Ly$, this becomes, for $x < y$, $\langle n_{Lx} n_{Ly} \rangle_c = -x(1 - y)(\rho_0 - \rho_1)^2/L$. As stated in [25], one may think that these weak, but long-range, correlations play no role in the macroscopic limit. However, they are responsible for a leading contribution in the variance of a macroscopic quantity such as the number of particles.

As expected, Fig. 3 shows how the Monte Carlo procedure fits exactly with the analytical results for $\rho_0 = \rho_1$ ($\langle n_i n_j \rangle_c = 0$) because we are using the equilibrium measure in which no spatial correlations are present. In addition, for $\rho_0 \neq \rho_1$ we see how the results obtained with the optimized measure are much closer to the exact analytical ones than those obtained with the local equilibrium assumption. This reflects the fact that with the optimized measure we are taking into account at least the two-site long-range correlations, which are not considered in the local equilibrium case. This shows that the physically motivated optimized trial is very good, at least regarding one- and two-point static spatial correlation functions. In this sense the out of equilibrium state of the SSEP can be thus fairly approximated by simple and intuitive quantities such as the local effective fields and two-site long-range interactions. Although more difficult to guess, one might of course add other terms to the trial function to improve the present agreement, for instance, higher order interaction terms, but the
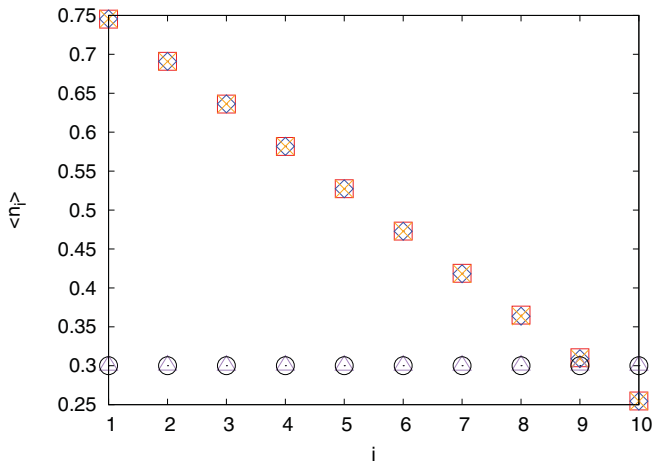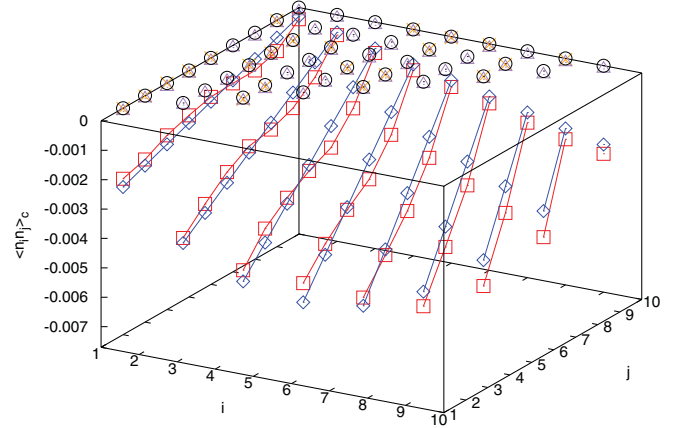

FIG. 2. (Color online) Analytical and simulation results for the steady-state density profile $\langle n_i \rangle$. Red squares and orange crosses correspond to the Monte Carlo procedure for densities $\rho_0 = 0.8$, $\rho_1 = 0.2$ using the optimized trial $\rho_{\text{opt}}(\mathbf{n})$ and the local equilibrium measure $\rho_{LE}(\mathbf{n})$, respectively, whereas blue diamonds are the analytical results; see Eq. (37). Purple triangles correspond to the Monte Carlo procedure for densities $\rho_0 = \rho_1 = 0.3$ using the equilibrium measure $\rho_{\text{eq}}(\mathbf{n})$, whereas black circles are the analytical results.

difficulty of the numerical minimization problem increases very rapidly.

V. CONCLUSION

We have derived an exact relation for Markovian systems that generalizes the Hatano-Sasa relation but does not rely on the *a priori* knowledge of the stationary probability distribution, but rather on arbitrary trial functions for the stationary distribution. More generally, we have derived the detailed version of the fluctuation relation by identifying a generalized form of dual (adjoint) dynamics, generating the backward process that yields a trajectory-dependent entropy production. For systems described by Langevin dynamics, we have showed that the dual dynamics is also governed by a simple Langevin dynamics, which may be expressed directly in terms of the trial functions. One may also obtain in this context a version of the three extended detailed fluctuation theorems of Ref. [13].

Our approach leads to an infinite family of inequalities that generalize the second law and suggests a variational principle for optimizing trial measures, in a quantitative and controlled way, to approximate nonequilibrium probability distributions. The optimization procedure we propose might be implemented numerically or even experimentally in order to infer nonequilibrium steady-state distributions in terms of intuitive physical quantities. To illustrate this, we have implemented this approximating scheme for the simple symmetric exclusion process in one dimension.

A particularly interesting case for applying this variational approach is active matter [28], where it has been proposed recently to represent the complex energy exchanges in the system by a bath with equilibrium-like properties (Refs. [2–4]). Another interesting and somewhat related system is that of current-driven vortices in superconductors with pinning. In this case the complex interplay of driving, quenched disorder, and vortex-vortex interactions yields a variety of nonequilibrium dynamical regimes and transitions that may sometimes be successfully described by effective temperatures [29]. At any rate, the important property of these approximations is that there is a second-law-type inequality associated with them.

APPENDIX: THREE EXTENDED DETAILED FLUCTUATION THEOREMS

Using the dual dynamics of of Eq. (27), it is straightforward to derive an extension for the three fluctuation theorems of Ref. [13] based, in our case, on smooth trial functions ρ_{ref} rather than the true steady-state measure ρ_{ss}.

We define three trajectory dependent "entropy productions" using time-reversal (R) and the dual (adjoint) dynamics:

$$\mathcal{S}[\mathcal{T},\alpha] \equiv \ln \mathcal{P}[\mathcal{T},\alpha] - \ln \mathcal{P}^R[\mathcal{T},\alpha], \qquad \text{(A1)}$$

$$\mathcal{Y}[\mathcal{T},\alpha] \equiv \ln \mathcal{P}[\mathcal{T},\alpha] - \ln [\mathcal{P}^{adj}[\mathcal{T},\alpha]]^R, \qquad \text{(A2)}$$

$$\mathcal{W}[\mathcal{T},\alpha] \equiv \ln \mathcal{P}[\mathcal{T},\alpha] - \ln \mathcal{P}^{adj}[\mathcal{T},\alpha], \qquad \text{(A3)}$$

so that by construction they satisfy the detailed fluctuation theorems,

$$\langle \mathcal{O}[\mathcal{T}] e^{-\mathcal{S}[\mathcal{T},\alpha]} \rangle = \langle \mathcal{O}[\mathcal{T}] \rangle^R, \qquad \text{(A4)}$$

$$\langle \mathcal{O}[\mathcal{T}] e^{-\mathcal{Y}[\mathcal{T},\alpha]} \rangle = [\langle \mathcal{O}[\mathcal{T}] \rangle^{adj}]^R, \qquad \text{(A5)}$$

$$\langle \mathcal{O}[\mathcal{T}] e^{-\mathcal{W}[\mathcal{T},\alpha]} \rangle = \langle \mathcal{O}[\mathcal{T}] \rangle^{adj}, \qquad \text{(A6)}$$

where again $\mathcal{O}$ is an arbitrary functional of the trajectory.

For a system satisfying a Fokker-Planck process (23), it is easy to show using Eqs. (25), (26), and (15) that

$$\partial_\tau \mathcal{S} = \beta \mathbf{f}(\mathbf{x};\alpha) \cdot \dot{\mathbf{x}}|_{t=\tau}, \qquad \text{(A7)}$$

$$\partial_\tau \mathcal{Y} = [\dot{\alpha}\partial_\alpha \phi_{\mathrm{ref}}(\mathbf{x},\alpha) + \varphi(\mathbf{x};\alpha)]|_{t=\tau}, \qquad \text{(A8)}$$

$$\partial_\tau \mathcal{W} = [\beta \mathbf{f}(\mathbf{x};\alpha) \cdot \dot{\mathbf{x}} - \dot{\alpha}\partial_\alpha \phi_{\mathrm{ref}}(\mathbf{x},\alpha) + \varphi(\mathbf{x};\alpha)]|_{t=\tau}, \qquad \text{(A9)}$$

with $\beta \equiv 1/T$. Therefore the total trajectory-dependent entropy production [8] may be written as

$$\partial_\tau \mathcal{S} = \partial_\tau \mathcal{Y} + \partial_\tau \mathcal{W} - 2\varphi. \qquad \text{(A10)}$$

The corresponding integral fluctuation theorems for $\mathcal{S}$, $\mathcal{Y}$, and $\mathcal{W}$, particular cases of Eqs. (A4), (A5), and (A6), respectively, read

$$\langle e^{-\mathcal{S}} \rangle = \langle e^{-\mathcal{Y}} \rangle = \langle e^{-\mathcal{W}} \rangle = 1. \qquad \text{(A11)}$$

The first fluctuation theorem in Eq. (A11) is the well-known integral fluctuation theorem for the total entropy production or Jarzynski relation, here generalized to arbitrary initial conditions and dynamics without detailed balance. The second one is the Hatano-Sasa extension discussed in detail in Sec. III, and the third one can be considered as an extension of the Speck and Seifert integral fluctuation theorem [30]. We note, however, that the interpretation of $\mathcal{Y}$ and $\mathcal{W}$ as the nonadiabatic $\mathcal{S}_{na}$ and adiabatic $\mathcal{S}_a$ entropy production contributions, respectively, is lost due to the presence of φ, which does not vanish in the adiabatic limit if $\rho_{\mathrm{ref}} \neq \rho_{\mathrm{ss}}$.

When $\rho_{\mathrm{ref}} = \rho_{\mathrm{ss}}$, $\phi_{\mathrm{ref}} = \phi$, and $\varphi = 0$, and the three extended fluctuation theorems reduce to the three fluctuation relations described in Ref. [13]. In this case, the dual dynamics becomes exactly the standard dual or adjoint (†) dynamics, and up to boundary terms we can identify $\mathcal{Y} = \mathcal{S}_{na}$, $\mathcal{W} = \mathcal{S}_a$. In addition, we note that Eq. (27) reduces to the standard dual dynamics in its Langevin form by the simple replacement $\phi_{\mathrm{ref}} \to \phi$. The dual dynamics given here is therefore the natural and probably simplest generalization of the standard dual dynamics that does not rely on the knowledge of ρ_{ss}.

[1] S. F. Edwards, in *Granular Matter: An Interdisciplinary Approach*, edited by A. Mehta (Springer, New York, 1994), p. 121, and references therein.

[2] J. Kurchan, J. Phys.: Condens. Matter **12**, 6611 (2000).

[3] D. Loi, S. Mossa, and L. F. Cugliandolo, Soft Matter (2011) (in press).

[4] S. Wang and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **108**, 15184 (2011).

[5] A broad review with an emphasis in experimental results and applications is F. Ritort, Séminaire Poincare **2**, 193 (2004), a more technical one specialized in the fluctuation relation is D. Evans and D. Searles, Adv. Phys. **51**, 1529 (2002); see also G. Gallavotti, *Statistical Mechanics: A Short Treatise*, Texts and Monographs in Physics Series (Springer, Berlin, 1999).

[6] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993); D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).

[7] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995); J. Stat. Phys. **80**, 931 (1995).

[8] J. Kurchan, J. Phys. A **31**, 3719 (1998).

[9] P. I. Hurtado, C. Pérez-Espigares, J. J. del Pozo, and P. L. Garrido, Proc. Natl. Acad. Sci. USA **108**, 7704 (2011).

[10] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); Phys. Rev. E **56**, 5018 (1997).

[11] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998); Phys. Rev. E **61**, 2361 (2000).

[12] T. Hatano and S. I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001); see also [31,32].

[13] M. Esposito and C. Van den Broeck, Phys. Rev. E **82**, 011143 (2010); C. Van den Broeck and M. Esposito, *ibid.* **82**, 011144 (2010); M. Esposito and C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010).

[14] W. G. Hoover, in *Molecular Dynamics*, Lecture Notes in Physics, edited by H. Araki *et al.*, Vol. 258 (Springer-Verlag, Heidelberg, 1986).

[15] This seems to be related to the expression, valid for systems that satisfy detailed balance, obtained in T. Sagawa and M. Ueda, Phys. Rev. Lett. **104**, 090602 (2010). Here the difference is that we are trying to approximate driven systems whose stationary measure we do not know.

[16] R. García-García, D. Domínguez, V. Lecomte, and A. B. Kolton, Phys. Rev. E **82**, 030104 (2010); R. García-García, V. Lecomte, A. B. Kolton, and D. Domínguez, J. Stat. Mech. (2012) P02009.

[17] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, J. Stat. Mech. (2006) P08001.

[18] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Phys. Rev. Lett. **87**, 040601 (2001); J. Stat. Phys. **107**, 635 (2002).

[19] J. L. Lebowitz and P. G. Bergmann, Ann. Phys. (NY) **1**, 1 (1957).

[20] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Phys. Rev. Lett. **98**, 080602 (2007); J. M. R. Parrondo, C. Van den Broeck, and R. Kawai, New J. Phys. **11**, 073008 (2009).

[21] N. Merhav and Y. Kafri, J. Stat. Mech. (2010) P12022.

[22] G. Verley, R. Chetrite, and D. Lacoste, J. Stat. Mech. (2011) P10025.

[23] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).

[24] See, for example, R. Monasson and S. Cocco, J. Comput. Neurosci. **31**, 199 (2011); S. Cocco, R. Monasson, and V. Sessak, Phys. Rev. E **83**, 051123 (2011).

[25] B. Derrida, J. L. Lebowitz, and E. R. Speer, J. Stat. Phys. **107**, 599 (2002).

[26] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, J. Phys. A **26**, 1493 (1993).

[27] J. Nocedal, Math. Comput. **35**, 773 (1980)

[28] C. Pérez-Espigares, A. B. Kolton, and J. Kurchan (in preparation).

[29] A. B. Kolton, R. Exartier, L. F. Cugliandolo, D. Dominguez, and N. Gronbech-Jensen, Phys. Rev. Lett. **89**, 227001 (2002).

[30] T. Speck and U. Seifert, J. Phys. A **38**, L581 (2005).

[31] Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. **130**, 29 (1998); see also S. Sasa and H. Tasaki, J. Stat. Phys. **125**, 125 (2006).

[32] The Hatano-Sasa and the Seifert [33] and Speck-Seifert relations [30] were very recently unified in three detailed fluctuation theorems through particular twofold decompositions of the total entropy production, either using single [13] or joint [16] probability distributions for the entropy production contributions.

[33] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).