Monotonicity of the dynamical activity

Christian Maes[1]**, Karel Netočný**[2] **and Bram Wynants**[3]

[1] Instituut voor Theoretische Fysica, KU Leuven, Belgium
[2] Institute of Physics AS CR, Prague, Czech Republic
[3] Institut de Physique Théorique, CEA-Saclay, France

E-mail: christian.maes@fys.kuleuven.be

Abstract

The Donsker–Varadhan rate function for occupation-time fluctuations has been seen numerically to exhibit monotone return to stationary non-equilibrium (Maes *et al* 2011 *Phys. Rev. Lett.* **107** 010601). That rate function is related to dynamical activity and, except under detailed balance, it does not derive from the relative entropy for which the monotonicity in time is well understood. We give a rigorous argument that the Donsker–Varadhan function is indeed monotone under the Markov evolution at large enough times with respect to the relaxation time, provided that a 'normal linear-response' condition is satisfied.

PACS numbers: 02.50.Ga, 05.20.Dd, 05.40.−a, 51.10.+y

(Some figures may appear in colour only in the online journal)

1. Physical motivation and main finding

Equilibrium under given thermodynamic conditions is characterized by minimizing the appropriate thermodynamic potential. For example, for a given temperature and pressure of the environment, an open system finds its equilibrium condition from minimizing the Gibbs free energy. That is why water has to boil at around 100 °C under atmospheric pressure, that is, the temperature above which the gas-phase of water gets a smaller Gibbs free energy. That minimization is also dynamically realized. Indeed, in many cases the second law of thermodynamics can be extended to become an H-theorem for the relaxation to equilibrium where the corresponding thermodynamic potential shows monotone decay to its equilibrium value. Close-to-equilibrium the situation then resembles the motion of a particle in a parabolic well with statistical forces derived from gradients in the equilibrium thermodynamic potential. That is indeed the framework of linear irreversible thermodynamics and the monotonicity in the approach to equilibrium relates essentially to the stability of the equilibrium phase.

The situation becomes much less clear when going to strong non-equilibrium regimes. Nature is full of stable non-equilibria for open systems in contact with different reservoirs, and yet no description in terms of thermodynamic potentials derived from energy and entropy

(possibly supplemented with few other macroscopic quantities) has been found. In fact, oscillations of the macroscopic condition are very well possible and can be effectively described by various models of dynamical systems. But even restricting ourselves to a stationary macroscopic condition, or to a stationary open system in contact with large reservoirs, no *physically* clear monotonicity has been observed in the approach to non-equilibrium. The numerical results observed in [12] open a new route that needs further study.

This paper addresses that general theoretical question, restricting to the simplest context of Markov evolutions as described by master or Fokker–Planck equations. It is true that there the relative entropy with respect to the stationary distribution is always monotone. Yet, the fact that the Lyapunov function is both well studied and not found directly related to known physical properties such as heat or work to which it relates under detailed balance. This paper looks elsewhere and we find a natural candidate for monotonicity to be the Donsker–Varadhan functional governing stationary dynamical fluctuations. It is physically related to the notion of dynamical activity (DA), as will be explained below, a concept that has grown in importance for the elucidation of relaxation behavior in kinematically constrained systems. In [12], we found that this functional is monotone in many examples. This paper adds a mathematical proof but we need an extra assumption: we show that this Donsker–Varadhan functional is monotone in the approach to stationary non-equilibrium under normal response behavior which is precisely stated as the sufficient condition in our main theorem 3.1. We also add various examples that clarify the nature of that sufficient condition. As a result, this paper is more technical than [12] and we now enter into more specific details.

Large deviation theory for Markov processes was developed by Donsker and Varadhan in 1975 [5]. We recall the main setting.

Consider an ergodic Markov jump process P_ρ with stationary probability law ρ and with transition rates $k(x, y)$ over a finite state space K. The empirical fraction of time that the system spends in state $x \in K$ over time-interval $[0, T]$ is

$$p_T(\omega, x) := \frac{1}{T} \int_0^T \delta_{x_s, x} \, \mathrm{d}s, \quad x \in K \tag{1.1}$$

where $\omega = (x_s, 0 \leqslant s < T)$ is the piecewise constant trajectory. By the assumed ergodicity, $p_T \to \rho$ for $T \uparrow +\infty$, P_ρ—almost surely. For the fluctuations around that law of large times, there is a principle of large deviations, abbreviated as

$$P_\rho[p_T \simeq \mu] \propto \mathrm{e}^{-T\mathcal{I}(\mu)}, \quad T \uparrow +\infty \tag{1.2}$$

for probability distributions μ on K, in the usual logarithmic and asymptotic sense $T \uparrow +\infty$; see e.g. [3, 4]. Equivalently, for all continuous functions f,

$$\lim_{T \uparrow +\infty} \frac{1}{T} \log \langle \mathrm{e}^{\int_0^T f(x_t) \, \mathrm{d}t} \rangle_\rho = \inf_\mu \left(\sum_x f(x) \mu(x) - \mathcal{I}(\mu) \right)$$

under the stationary expectation $\langle \cdot \rangle_\rho$.

$$\mathcal{I}(\mu) = \sup_{g > 0} \left(- \sum_x \frac{\mu(x)}{g(x)} \sum_y k(x, y) \left[g(y) - g(x) \right] \right) \tag{1.3}$$

over positive functions g.

Since the functional $\mathcal{I}$ is strictly convex with unique minimum reached at the stationary distribution it is mathematically natural to ask whether $\mathcal{I}(\mu_t)$ is also monotonically decaying to zero under the time evolution given by the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \mu_t(x) = \sum_{y \in K} [k(y, x)\mu_t(y) - k(x, y)\mu_t(x)], \quad \mu_0 = \mu$$

at least when μ (the initial condition) is sufficiently close to ρ. We found numerically in many cases that the answer is *yes*, where the close-to-stationarity is verified for large times t. Then, indeed μ_t gets sufficiently close to the stationary ρ. That was reported in [12]. A mathematical proof of monotonicity of the functional $\mathcal{I}(\mu_t)$ for large times t is lacking and this paper will still need an additional assumption, called 'normal linear response'. The normal linear response refers physically to the monotone decay of the linear response function, and mathematically it can be phrased as a sector condition on the backward generator L. The latter essentially means that the eigenvalues of L should be contained in a wedge of the complex plane with a sufficiently small angle. In particular, it will be easy to show that the monotonicity holds when the system satisfies the condition of detailed balance, and hence, by a continuity argument, the monotone return to steady non-equilibrium is also valid *around* detailed balance. Yet again, as we will see below, the monotonicity often continues to hold even beyond the linear regime around detailed balance.

We now turn to the more physical motivation. Recently, from the point of view of non-equilibrium statistical mechanics, there has been great interest in the dynamical fluctuation theory, and in the occupation statistics in particular. The Donsker–Varadhan functional is exactly governing these fluctuations as mentioned above. In that way, our result mirrors the monotone behavior of the relative entropy which is associated with the *static* fluctuations of the system [13]. Apart from this more abstract analogy, there are specific places where the functional $\mathcal{I}$ has turned up in non-equilibrium considerations. We know for example that close-to-equilibrium $\mathcal{I}$ is proportional to the excess in expected entropy production rate with respect to the stationary entropy production rate [11], which gives a fluctuation-based understanding of the minimum entropy production principle. Non-perturbatively, the functional $\mathcal{I}(\mu)$ is an excess in expected DA as we now explain.

Usually for jump processes, one calls DA the quantity defined on path-space that counts the number of jumps or transitions. Fixing any two distinct states $x,\ y$ the expected rate of jumping $x \to y$ when in x is of course $k(x, y)$. Therefore, under distribution μ, the expected number of jumps per unit time is given by

$$\xi(\mu) := \sum_x \mu(x) \sum_y k(x, y) = \frac{1}{2} \sum_{x,y} [\mu(x)k(x, y) + \mu(y)k(y, x)]$$

which is the symmetric counterpart of the current (formally with $k(x, x) \equiv 0$). Since no confusion arises here, in the following, we use the term *DA* for the functional $\xi(\mu)$ evaluating the expected value of the more commonly defined variable DA on path-space. The Donsker–Varadhan functional is the difference $\mathcal{I}(\mu) = \xi(\mu) - \xi_V(\mu)$, where ξ_V is defined like ξ but for modified rates $k_V(x, y) := k(x, y) \exp\{[V(y) - V(x)]/2\}$ with a potential V so that the jump process with rates $k_V(x, y)$ makes μ stationary. That will be explained in more detail in the next section, and will be made most explicit in formula (2.7). The $\mathcal{I}(\mu)$ is thus an excess in expected activity between the original dynamics and a modified dynamics for which μ is made stationary.

On a broader level, the functional $\mathcal{I}(\mu)$ refers to a combination of properties of a statistical mechanical system that are related to its reactivity and the ability to escape from its present state. DA, and more specifically its version as defined on path-space, has been studied in connection with glassy behavior and the glass transitions; kinetically constrained models show a reduced DA over an extensive number of states which leads to dynamical phase transitions [2, 6–8]. Finally, DA has appeared in fluctuation and response theory for steady non-equilibria [1, 14, 15]. The point is that as a function on trajectories, the DA is time symmetric and complements time-antisymmetric entropy fluxes whenever beyond the linear regime around

equilibrium. That is why it enters the non-equilibrium fluctuation structure as well as provides extra contributions to the fluctuation–response relations.

Concerning physical implications of the observed (and here partially proven) time monotonicity of the Donsker–Varadhan functional, we should again compare with the situation of relaxation to equilibrium. There the existence of a physically meaningful (Lyapunov) functional, which does not increase over time, remarkably restricts the collection of admissible relaxation processes. Here, we wish to proceed similarly to the strategy in equilibrium, but we argue that out of equilibrium it may be useful to start from the DA as a fundamental quantity instead of the entropy. This paper contains no final judgement on this proposal but only a mathematically rigorous argument that this remains a valid possibility. It is proven that there is indeed a general tendency to decrease the excess DA analogous to the arrow of time associated with the increase of entropy under equilibrium condition. This is also related to the largely open problem of non-equilibrium statistical forces: the monotonicity of the Donsker–Varadhan functional suggests that its gradient with respect to macroscopic parameters can play the role of non-equilibrium statistical forces mimicking Onsager's theory of hydrodynamic entropy production. For some related physical arguments supporting the fundamental role of noise and DA out of equilibrium see e.g. [10].

The next section specifies the mathematical set-up and the main definitions. Section 3 collects the main properties, with our result on monotone behavior. Section 4 discusses various specific examples far- and close-to-equilibrium. Proofs are collected in section 5 after which a final conclusion follows.

2. Set-up

As in the previous section, we consider a Markov jump process on a finite state space K with states $x, y, \ldots$ and transition rates $k(x, y)$. Probability distributions on K will be denoted by $\rho, \mu, \nu, \ldots$. The backward generator on functions f is

$$Lf(x) := \sum_{y \in K} k(x, y)[f(y) - f(x)]$$

and its transpose generates the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \mu_t(x) + \sum_{y \in K} j_{\mu_t}(x, y) = 0, \quad j_\nu(x, y) := k(x, y)\, \nu(x) - k(y, x)\, \nu(y) \quad (2.1)$$

for the evolution on probabilities μ_t starting from some initial $\mu_0 = \mu$ on K. We assume that the Markov process is irreducible with unique stationary probability distribution ρ, i.e. $\rho(x) > 0$ solves $\sum_y j_\rho(x, y) = 0$ for all $x \in K$.

We say that the dynamics satisfies detailed balance when there is a function U on K for which

$$k_e(x, y)\, \mathrm{e}^{-U(x)} = k_e(y, x)\, \mathrm{e}^{-U(y)}, \quad \rho_e(x) \propto \mathrm{e}^{-U(x)}. \quad (2.2)$$

Here and below we decorate the rates and the stationary law in that detailed balance case with the subscript 'e'. Then, the free energy functional

$$\mathcal{F}(\mu) := \sum_x \mu(x)U(x) + \sum_x \mu(x) \log \mu(x) \geqslant \mathcal{F}(\rho_e) = -\log \sum_x \exp[-U(x)]$$

satisfies the monotonicity $\mathcal{F}(\mu_t) \downarrow \mathcal{F}(\rho_e)$ as a function of time t. That is just a standard consequence of the general monotonicity of the relative entropy under stochastic transformations. However, the relation between the Shannon entropy $-\sum_x \mu(x) \log \mu(x)$ and physical notions as work or heat is mostly lost when far away from detailed balance.

A physically relevant alternative when moving away from detailed balance is to consider the instantaneous entropy production $\mathcal{E}(\mu)$, which for the given context is

$$
\begin{aligned}
\mathcal{E}(\mu) &:= \sum_{x,y} \mu(x) k(x,y) \log \frac{\mu(x) k(x,y)}{\mu(y) k(y,x)} \\
&= \frac{1}{2} \sum_{x,y} j_\mu(x,y) A_\mu(x,y), \qquad A_\mu(x,y) := \log \frac{\mu(x) k(x,y)}{\mu(y) k(y,x)},
\end{aligned}
\tag{2.3}
$$

as the product of 'fluxes' $j_\mu(x,y)$ and 'forces' $A_\mu(x,y)$ when the system's distribution is μ, reminiscent of irreversible thermodynamics—see e.g. [17] for more details.

We now introduce our main object. We embed the original dynamics into a larger family of processes with transition rates

$$
k_W(x,y) := k(x,y) \, \exp \frac{W(y) - W(x)}{2}
\tag{2.4}
$$

parameterized by functions W on K. These functions W are also called potentials. Here, we consider potentials that are directly connected with a probability distribution. What follows is a standard observation within the theory of large deviations, see e.g. section 3.1.2 in [3], but for self-consistency we give a full proof in section 5.1.

Proposition 2.1. *For an arbitrary probability distribution $\mu > 0$ there exists a potential $V = V_\mu$ on K such that μ is invariant under the modified dynamics with transition rates $k_V(x,y)$. The potential V_μ is unique up to an additive constant when the dynamics is irreducible.*

In other words, for arbitrary $\mu > 0$ we can always find a function V so that

$$
\sum_{y \in K} [k_V(x,y)\,\mu(x) - k_V(y,x)\,\mu(y)] = 0, \quad x \in K.
\tag{2.5}
$$

We can compare this with (1.3). Indeed, the Donsker–Varadhan large deviation functional can be written in terms of a potential W: taking $g = e^{W/2}$ in (1.3),

$$
\mathcal{I}(\mu) = \sup_W \sum_{x,y \in K} \mu(x) \, [k(x,y) - k_W(x,y)].
\tag{2.6}
$$

When the process is irreducible, cf proposition 2.1 and its proof in section 5.1, we then have

$$
\mathcal{I}(\mu) = \sum_{x,y \in K} \mu(x) \, [k(x,y) - k_V(x,y)] \quad \text{ with } V = V_\mu.
\tag{2.7}
$$

For physical motivation and as mentioned already in section 1, it is worth noting that $\mathcal{I}(\mu)$ is an excess or difference between the expected escape rates $\sum_x \mu(x) \sum_y k(x,y)$ and $\sum_x \mu(x) \sum_y k_V(x,y)$. Such an expected escape rate estimates the DA, i.e. the number of transitions per unit time in the process. We refer to the physics literature for further discussion [1, 2, 68, 15].

3. Main result

For simplicity in the following we always assume the irreducibility of continuous-time Markov processes with finite state space. Our main finding is that $\mathcal{I}(\mu_t)$ is monotone under the evolution (2.1) when close enough to stationarity, i.e. for large enough times t compared to the relaxation time, at least under some further and physically interpretable condition.

Define the real-space scalar product $(f,g) := \sum_x f(x) g(x) \rho(x)$ so that $(f, Lg) = (L^* f, g)$. Here, we have introduced the generator L^* of the time-reversed process:

$$
L^* f(x) := \sum_{y \in K} \frac{\rho(y) k(y,x)}{\rho(x)} [f(y) - f(x)].
$$

We write L_s for the symmetric part of the generator: $L_s := \frac{1}{2}(L + L^*)$, and

$$|||f||| := \max_{x,y} |f(x) - f(y)|$$

for the variation of a function f on K. Now comes the main result of the paper.

Theorem 3.1. *Suppose that there is a constant $c > 0$, so that $(L_s f, Lf) \geqslant c\,|||f|||^2$ for all functions f on K. Then, there is a time $t_o > 0$ so that for all initial probability distributions μ on K*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(\mu_t) \leqslant 0 \quad \text{for all times } t \geqslant t_o.$$

In section 5.3, it will be shown that the time t_o after which monotonicity sets in is of the order of the relaxation time (inverse of the exponential rate of convergence).

Since L and L_s have a bounded inverse on the functions f that have zero mean $\sum_x \rho(x) f(x) = 0$, the condition of theorem 3.1 in essence means to require that $(L_s f, Lf) > 0$ for all non-constant f. In fact, it is sufficient and more convenient to verify the inequality

$$(f, L^2 f) > 0 \tag{3.1}$$

which is usually called (a specific instance of) a sector condition. In particular, for L a normal operator this is equivalent to the assumption that all its non-zero eigenvalues $\lambda = -a + \mathrm{i}b \neq 0$ obey the inequality $\Re(\lambda^2) > 0$, i.e. $|b| < a$. Obviously, the condition (3.1) is fulfilled whenever the rates $k(x, y)$ satisfy detailed balance (2.2), $L_e = L_e^*$, see section 4.2 for further discussion. By a continuity argument, this also extends to dynamics where the detailed balance is only weakly violated.

A physical interpretation of the hypothesis for theorem 3.1 is in terms of the generalized susceptibility for the linear response around the stationary probability ρ. For a function B on K consider perturbed transition rates

$$k(\tau; x, y) := k(x, y) \, \mathrm{e}^{\frac{h_\tau}{2} [B(y) - B(x)]}, \qquad \tau \geqslant 0 \tag{3.2}$$

with small time-dependent amplitude h_τ, $|h_\tau| \leqslant \varepsilon$. It resembles (2.4) but now the perturbation is time dependent. That new time-dependent process is started from time zero in the distribution ρ which is stationary for $h \equiv 0$. At a later time $t > 0$, in the process with rates (3.2), when taking the expectation of a function G, we see the difference

$$\langle G(x_t) \rangle^h - \sum_x \rho(x) G(x) = \int_0^t \mathrm{d}\tau\, h_{t-\tau} \chi_{GB}(\tau) + O(\varepsilon^2)$$

which defines the generalized susceptibility χ_{GB}. There is an explicit formula extending the standard fluctuation–dissipation theorem, see e.g. [1]:

$$\chi_{GB}(t) := \frac{\delta}{\delta h_0} \langle G(x_t) \rangle^h |_{h=0} = -\frac{1}{2} \left[\frac{\mathrm{d}}{\mathrm{d}t} \langle B(x_0)\, G(x_t) \rangle_\rho + \langle LB(x_0)\, G(x_t) \rangle_\rho \right] \tag{3.3}$$

with right-hand expectations in the original stationary process P_ρ. Our next result gives an explicit expression for the zero-time susceptibility in the case $G = B$; see section 5.4 for a proof.

Proposition 3.2. *We have the identities*

$$\frac{\mathrm{d}}{\mathrm{d}t} \chi_{ff}(t) \Big|_{t=0} = \chi_{Lf,f}(0) = -(L_s f, Lf). \tag{3.4}$$

As a consequence, the hypothesis of theorem 3.1 can be rephrased as the condition $\chi_{ff}(t) \leqslant \chi_{ff}(0)$ for small enough $t > 0$. Such an equilibrium-like response behavior at initial times is called 'normal linear response' throughout the paper.

4. Examples

We illustrate the statements of the previous section by providing some examples.

4.1. Asymmetric diffusion on the ring

Consider a ring consisting of $N > 2$ sites, labeled $x = 1, 2, \ldots, N + 1 \equiv 1$. For a totally asymmetric random walker the only non-zero transition rates are of the form $k(x, x + 1) > 0$ with master equation (2.1) simplified to

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}t}(x) = \mu_t(x - 1)k(x - 1, x) - \mu_t(x)k(x, x + 1).$$

The stationary distribution is

$$\rho(x) = \frac{C}{k(x, x + 1)}, \quad C^{-1} = \sum_x \frac{1}{k(x, x + 1)}.$$

The corresponding generators of theorem 3.1 are

$$\begin{aligned} Lf(x) &= k(x, x + 1)[f(x + 1) - f(x)], \\ L_s f(x) &= \tfrac{1}{2}k(x, x + 1)[f(x + 1) + f(x - 1) - 2f(x)] \end{aligned}$$

so that the hypothesis of theorem 3.1 concerns

$$(L_s f, Lf) = \frac{C^2}{2} \sum_x \frac{1}{\rho(x)} [f(x + 1) - f(x)][f(x + 1) + f(x - 1) - 2f(x)]. \tag{4.1}$$

Homogeneous case. Let $k(x, x + 1) \equiv p$ for some given $p > 0$. In that case, (4.1) is bounded from below by the variation of f because, with $g(x) = f(x + 1) - f(x)$, $\sum_x g(x)[g(x) - g(x - 1)] = \tfrac{1}{2}\sum_x[g(x) - g(x - 1)]^2 \geqslant 0$ on the ring.

But we can also explicitly show that $\mathcal{I}(\mu_t)$ is monotone for all times $t \geqslant 0$ and starting from all possible $\mu > 0$. For this, we find the potential V_μ solving

$$0 = \mu(x)p\,\mathrm{e}^{[V(x+1)-V(x)]/2} - \mu(x - 1)p\,\mathrm{e}^{[V(x)-V(x-1)]/2}$$

or

$$\frac{V(x + 1) - V(x)}{2} = -\log \mu(x) + \frac{1}{N} \sum_y \log \mu(y).$$

Therefore, the Donsker–Varadhan functional (2.7) equals

$$\begin{aligned} \mathcal{I}(\mu) &= \sum_x p\mu(x)[1 - \mathrm{e}^{\frac{V(x+1)-V(x)}{2}}] \\ &= p - pN\left[\prod_{y=1}^N \mu(y)\right]^{\frac{1}{N}} \end{aligned}$$

in terms of the geometric mean of the $\mu(1), \mu(2), \ldots, \mu(N)$. The time derivative at $\mu_t = \mu$ is computed to be

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I}(\mu_t) = -p^2 \left[\prod_{y=1}^N \mu(y)\right]^{\frac{1}{N}} \sum_{x=1}^N \left(\frac{\mu(x - 1)}{\mu(x)} - 1\right)$$

which is non-positive by applying Jensen's inequality as

$$\log\left(\frac{1}{N} \sum_{x=1}^N \frac{\mu(x - 1)}{\mu(x)}\right) \geqslant \frac{1}{N} \sum_{x=1}^N \log \frac{\mu(x - 1)}{\mu(x)} = 0.$$
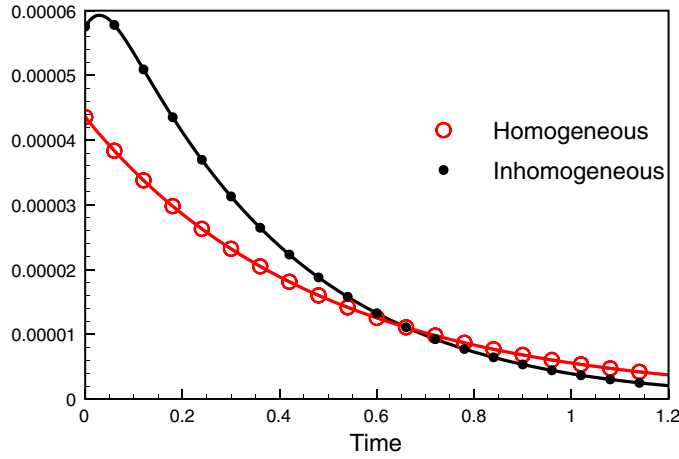
Figure 1. The functional $\mathcal{I}(\mu_t)$ as a function of time. The curve with open circles shows the evolution of $\mathcal{I}(\mu_t)$ for the case of a homogeneous stationary distribution. The black curve (closed smaller circles) represents the case where the stationary distribution is inhomogeneous: $\rho(1) = 1/91, \rho(2) = \rho(3) = \rho(4) = 30/91$.

In fact, by the same argument, the time derivative is strictly negative whenever $\mu \neq \rho$. Therefore, for homogeneous totally asymmetric walkers on a ring, we always have monotonicity of the geometric mean of the occupations $\mu_t(x)$, as a consequence of the monotonicity of the Donsker–Varadhan functional.

Inhomogeneous case. We look back at (4.1) which is now of the form $\sum_x \frac{1}{\rho(x)} g(x)[g(x) - g(x-1)]$, always with $g(x) = f(x) - f(x-1)$. An explicit computation of the time derivative of $\mathcal{I}(\mu_t)$ gives, with $\mu_t = \mu$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(\mu_t) = 2\epsilon^2 C^2 \sum_x \frac{V(x)}{\rho(x)} [V(x-1) - V(x)] + o(\epsilon^2) \tag{4.2}$$

with $V(x) = -h(x) + \frac{1}{N} \sum_x h(x)$ for $\mu(x) = \rho(x)[1 + \epsilon h(x)]$. We thus see how the condition in theorem 3.1 appears. Without any further condition and depending on the shape of the stationary distribution ρ, this time derivative (4.2) can be either positive or negative. For an example making $\mathcal{I}(\mu_t)$ non-montone at initial times, we take $N = 4$ with rates $k(1,2) = 30, k(2,3) = k(3,4) = k(4,1) = 1$, and μ determined from $\epsilon = 0.02$ and

$$V(1) = 1, V(2) = -3, V(3) = 0, V(4) = 2. \tag{4.3}$$

Then, the time derivative (4.2) is positive whenever

$$\frac{1}{\rho(1)} > \frac{12}{\rho(2)} + \frac{4}{\rho(4)}.$$

To visualize this example we show in figure 1 the result of a numerical computation of $\mathcal{I}(\mu_t)$ for this initial condition (4.3). Observe, however, even in this example, that after a short initial slip $\mathcal{I}(\mu_t)$ starts decreasing monotonically. In other words, while for all $\epsilon > 0$ there is a probability μ in the neighborhood of the stationary law ρ with variational distance $d(\mu, \rho) \leqslant \epsilon$, so that $\mathcal{I}(\mu_t)$ is not monotone at $\mu_t = \mu$, still $\mathcal{I}(\mu_t)$ decays monotonically to zero *eventually* (after a long enough time t). This also indicates that the hypothesis in theorem 3.1 is not at all necessary.

An example of a driven diffusion process on the ring was treated in [12]. There, in figure 1 of [12], we have seen that $\mathcal{I}(\mu_t)$ keeps decaying monotonically while the entropy production $\mathcal{E}(\mu_t)$, the continuum version of (2.3), oscillates in time.

4.2. Detailed balanced dynamics

Under detailed balance (2.2), we can take $g = \sqrt{\rho_e/\mu}$ in (1.3) to find equality with the Dirichlet form

$$\mathcal{I}_e(\mu) = -\sum_x \rho_e(x) \sqrt{\frac{\mu(x)}{\rho_e(x)}} \left(L_e \sqrt{\frac{\mu}{\rho_e}} \right)(x) = -\left(\sqrt{\frac{\mu}{\rho_e}}, L_e \sqrt{\frac{\mu}{\rho_e}} \right).$$

Furthermore, for $\mu_t = \mu$ at time t,

$$\dot{\mu}_t = \rho_e L_e \left(\frac{\mu}{\rho_e} \right)$$

so that we obtain the time derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}_e(\mu_t) = -\sum_x \frac{\rho_e(x)}{f(x)} (L_e f^2)(x) (L_e f)(x), \tag{4.4}$$

where we have abbreviated $f = \sqrt{\mu/\rho_e}$. Obviously, that time derivative is negative whenever $f - 1$ is sufficiently small (close-to-stationarity): with $f(x) = 1 + \epsilon h(x)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}_e(\mu_t) = -2\epsilon^2 \sum_x \rho_e(x)[(L_e h)(x)]^2 + o(\epsilon^2).$$

That is equivalent to what was mentioned before under theorem 3.1: the hypothesis there is always satisfied under detailed balance, and we see that the time t_0 will be of the order of the relaxation time, characterizing the uniform exponentially fast convergence to equilibrium. More explicit calculations reveal also that the derivative (4.4) is strictly negative for all $f > 0$ when $|K| \leqslant 3$.

On the other hand, we can check that the time derivative (4.4) may obtain positive values too. For convenience, we briefly consider a one-dimensional diffusion variant of the same problem, with the forward generator

$$L_e g = \frac{1}{\rho_e} \frac{\mathrm{d}}{\mathrm{d}x} \left(\rho_e \frac{\mathrm{d}g}{\mathrm{d}x} \right)$$

for smooth functions g. Let us take $\mu(x) = a/\rho_e(x)$ inside the interval $[1, \ell]$, where $\ell > 1$ and a is a normalization, and $\mu(x) \simeq \rho_e(x)$ very rapidly decaying to zero outside that same interval. Then, at that μ, (4.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}_e(\mu_t) \simeq -2 \int_1^\ell \mathrm{d}x f''(x) (\log f)''(x)$$

which will be positive e.g. when f is convex while $\log f$ is concave—for example, with $\rho_e(x) = c_\ell/x^2$ on $[1, \ell]$, we have $f(x) \sim x^2$, $\log f(x) \sim \log x$ on that same interval.

5. Proofs

5.1. Proof of proposition 2.1

For arbitrary $\mu > 0$, we consider the auxiliary functional

$$Y_\mu(W) := \sum_{x,y \in K} \mu(x) k_W(x, y) = \sum_{x,y \in K} \mu(x) k(x, y) \exp\left[\frac{W(y) - W(x)}{2} \right] \tag{5.1}$$

defined on all functions on K. Of course, since the value only depends on the differences $W(y) - W(x)$ we can as well take $W \in C_0(K)$, the collection of all functions that are equal to zero on a fixed 'root' $x_0 \in K$. This functional Y_μ is non-negative and convex

$$Y_\mu(\lambda W_1 + (1-\lambda)W_2) \leqslant \lambda Y_\mu(W_1) + (1-\lambda)Y_\mu(W_2) \tag{5.2}$$

by convexity of each contribution $\mu(x) k_W(x, y)$. Below in lemma 5.1, we prove that under the irreducibility assumption, Y_μ is actually *strictly* convex and that it attains inside $C_0(K)$ a unique minimum at some $W = W_\mu$. Hence, W_μ is also a minimizer (unique up to an additive constant) on the unconstrained space of all functions on K, implying that for all $x \in K$,

$$0 = \left.\frac{\delta Y_\mu}{\delta W(x)}\right|_{W=W_\mu} = \frac{1}{2} \sum_{y \in K} [\mu(y) k_W(y, x) - \mu(x) k_W(x, y)] \tag{5.3}$$

which is just the stationarity of μ for the dynamics with rates $k_W(x, y)$, i.e. $V = V_\mu$ of proposition 2.1 does exist and equals W_μ. This also proves formula (2.6) as

$$\mathcal{I}(\mu) = \sum_{x,y \in K} \mu(x) k(x, y) - Y_\mu(W_\mu) = \sup_V \sum_{x,y \in K} \mu(x) [k(x, y) - k_V(x, y)]. \tag{5.4}$$

A general reducible dynamics can be decomposed into irreducible components (including isolated sites) and for each of them the above argument holds true, i.e. the supremum on the right-hand side of (2.6) is attained on a function V_μ, which is also a solution of the inverse stationarity problem and which is unique up to a constant within each component.

Again turning to irreducible dynamics, we have

Lemma 5.1. *For any $\mu > 0$, $Y_\mu|_{C_0(K)}$ is strictly convex and has a unique minimum.*

Proof. By irreducibility, there exists a cyclic sequence of states $(x_0, x_1, \ldots, x_n = x_0)$ that covers the whole space K and such that for all consecutive pairs of states, $\mu(x_{i-1}) k(x_{i-1}, x_i) \geqslant \delta$ with some $\delta > 0$. If W_1 and W_2 are such that the relation (5.2) becomes an equality, then, for all $i = 1, \ldots, n$, $W_1(x_i) - W_1(x_{i-1}) = W_2(x_i) - W_2(x_{i-1})$, by using that the exponential is strictly convex. From $W_1(x_0) = W_2(x_0) = 0$ then follows $W_1 = W_2$, identically. This proves the strict convexity and hence the uniqueness of the minimum for $Y_\mu|_{C_0(K)}$.

To prove that the minimum exists, we consider the compact sets

$$C_0^a(K) := \{W \in C_0(K); |W(x)| \leqslant a \quad \text{for all } x\}, \quad a > 0$$

and define $M_\mu := Y_\mu(0) = \sum_{x,y \in K} \mu(x) k(x, y)$. By construction, for any $W \in C_0(K) \setminus C_0^a(K)$ there exists i such that $W(x_i) - W(x_{i-1}) > a/n$ and hence $Y_\mu(W) > \delta \, e^{a/(2n)}$. Fix now some a so that $\delta \, e^{a/(2n)} > M_\mu$. By compactness, Y_μ on the set $C_0^a(K)$ attains the minimum, which then coincides with the minimum of $Y_\mu|_{C_0(K)}$. $\qquad \square$

5.2. The map $\mu \mapsto V_\mu$

Most importantly, from the previous section, the map $\mu \mapsto V_\mu$ is a bijection when we think of the potential modulo a constant. Moreover V_μ depends smoothly on μ, and vice versa. In other words, the map $\mu \mapsto V_\mu$ is a diffeomorphism with variational distance $d(\mu, \rho)$ of the same order as V:

$$c_0 d(\mu, \rho) \leqslant |||V_\mu||| \leqslant c_1 d(\mu, \rho),$$

for constants $c_0, c_1 > 0$. That is really a consequence of the irreducibility of the finite Markov process, or see chapter 2 in [9].

Heuristically it suffices to understand the linearized map around $\mu = \rho$, $V_\rho = 0$ since the modified rates $k_{V_1+V_2}(x, y) = k_{V_1}(x, y) \exp\{V_2(y) - V_2(x)\}/2$ each time define an irreducible Markov process for each V_1. Writing $\mu = \rho(1 + \varepsilon h)$ for some function h with mean $\sum_x \rho(x)h(x) = 0$ and for small ε, we easily find $V_\mu = \varepsilon v + O(\varepsilon^2)$ with

$$L_s v = L^* h. \tag{5.5}$$

Note here that ρ is also invariant under the time-reversed process and under the (detailed balanced) process generated by L_s. Hence, $\sum_x \rho(x)L^*h(x) = 0$ and L^*h is in the domain of the (Drazin) pseudo-inverse $(L_s)^{-1}$, so that (5.5) has a unique solution (again up to a constant); in fact $h = 0$ if and only if $v = 0$.

The computation leading to (5.5) goes as follows. For all $x \in K$,

$$0 = \sum_y \left(\mu(y)k(y, x) \, \mathrm{e}^{\frac{V(x)-V(y)}{2}} - \mu(x)k(x, y) \, \mathrm{e}^{\frac{V(y)-V(x)}{2}} \right)$$

which by expanding the exponential directly yields the identity

$$L^* \left(\frac{\mu}{\rho} - 1 \right)(x) - L_s V_\mu(x) = w(x, \mu), \tag{5.6}$$

where (with $V = V_\mu$)

$$w(x, \mu) := \sum_y \left\{ \left(\frac{\mu(y)}{\rho(y)} - 1 \right) \frac{\rho(y)k(y, x)}{\rho(x)} \frac{V(x) - V(y)}{2} + \left(\frac{\mu(x)}{\rho(x)} - 1 \right) k(x, y) \frac{V(x) - V(y)}{2} \right\}$$

$$+ \sum_y \left\{ \frac{\mu(y)}{\rho(x)} k(y, x) - \frac{\mu(x)}{\rho(x)} k(x, y) \right\} \delta_V(x, y) \tag{5.7}$$

for $\delta_V(x, y) := \sum_{n=2} \left[\frac{V(x)-V(y)}{2} \right]^n \frac{1}{n!}$. Each difference

$$\left| \frac{\mu(x)}{\rho(x)} - 1 \right| \leqslant C_0 |||V_\mu|||, \quad x \in K$$

so that $|w(x, \mu)| \leqslant C_1 |||V_\mu|||^2$ for some constant C_1 when $|||V_\mu|||$ is sufficiently small.

5.3. Proof of theorem 3.1

We must take the time derivative of $\mathcal{I}(\mu_t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(\mu_t) = -\frac{1}{2} \sum_{x,y} \dot\mu_t(x)k(x, y)[V_t(y) - V_t(x) + 2\delta_V(y, x)]$$

$$+ \frac{1}{2} \sum_{x,y} \mu_t(x)k_{V_t}(x, y)[\dot V_t(x) - \dot V_t(y)] \tag{5.8}$$

where $\dot V_t(x) := \frac{\mathrm{d}}{\mathrm{d}t} V_{\mu_t}(x)$. The second line in (5.8) equals zero because per fixed time t, μ_t is stationary for the dynamics with rates $k_{V_t}(x, y)$. We thus have

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(\mu_t) = -\frac{1}{2} \sum_x \dot\mu_t(x) \left[LV_t(x) + 2 \sum_y k(x, y)\delta_V(y, x) \right].$$

Looking at the first term, we use that

$$\dot\mu_t(x) = \rho(x)L^* \left(\frac{\mu_t}{\rho} - 1 \right)(x) = \rho(x)L_s V_t(x) + \rho(x)w(x, \mu_t)$$

as introduced in (5.6). In other words, we have obtained

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(\mu_t) = -\frac{1}{2} Q(V_t) - \frac{1}{2} \sum_x \rho(x)w(x, \mu_t)LV_t(x) - \sum_x \dot\mu_t(x)k(x, y)\delta_{V_t}(y, x)$$

for the quadratic form $Q(f) := (Lf, L_s f) = (f, L^* L_s f) = (L_s Lf, f)$ which, from the hypothesis of theorem 3.1, is bounded from below by $c \, |||f|||^2$. Since $|||V_t||| \leqslant K \exp[-\gamma t]$ for some $K < \infty, \gamma > 0$, it suffices finally to realize that, at least for large enough times t,

$$\left| \frac{1}{2} \sum_x \rho(x) w(x, \mu_t) LV_t(x) + \sum_x \dot{\mu}_t(x) k(x, y) \delta_{V_t}(y, x) \right| \leqslant C |||V_t|||^3$$

for some $C < \infty$. That easily follows by applying uniform bounds such as $|LV(x)| \leqslant C_2 \, |||V|||$ and $|\delta_V(x, y)| \leqslant C_3 \, |||V|||^2$ for small enough V, combined with the previous estimate $|w(x, \mu)| \leqslant C_1 \, |||V_\mu|||^2$ making also $|\dot{\mu}_t(x)| \leqslant C_4 \, |||V_t|||$. That concludes the proof of theorem 3.1.

The proof above obviously gives an estimate of the time t_0 after which monotonicity surely sets in. Since $|||V_t|||$ is of the order $\exp(-\gamma t)$, $Q(V_t) > 0$ dominates the time derivative of the Donsker–Varadhan functional when $\gamma t \gg 1$, i.e. for times beyond the relaxation time to the non-equilibrium steady regime.

5.4. Proof of proposition 3.2

A straightforward computation gives, for all $t \geqslant 0$,

$$\chi_{ff}(t) = -\frac{1}{2} \left[\frac{d}{dt} \langle f(x_0) f(x_t) \rangle_\rho + \langle Lf(x_0) f(x_t) \rangle_\rho \right]$$
$$= -\frac{1}{2} \langle (L + L^*) f(x_0) f(x_t) \rangle_\rho \tag{5.9}$$

and hence

$$\left. \frac{d}{dt} \chi_{ff}(t) \right|_{t=0} = -(L_s f, Lf) = -Q(f). \tag{5.10}$$

Similarly, we look at the response of the observable Lf to find

$$\chi_{Lf,f}(t) = -\frac{1}{2} \left[\frac{d}{dt} \langle f(x_0) Lf(x_t) \rangle_\rho + \langle Lf(x_0) Lf(x_t) \rangle_\rho \right]$$
$$= -\frac{1}{2} \langle (L^* + L) f(x_0) Lf(x_t) \rangle_\rho$$
$$\chi_{Lf,f}(0) = -(L_s f, Lf). \tag{5.11}$$

In particular, the equality (5.10) = (5.11) shows that at equal times we can commute the time derivative and the derivative with respect to the perturbation.

6. Conclusion

The Donsker–Varadhan functional is related to the dynamical activity (DA) as recently studied in constructions of non-equilibrium statistical mechanics. We have given a necessary condition for its monotonicity under the master equation. The condition was called 'normal linear response,' as it requires the generalized susceptibility to initially decay in time. We have also given examples where that condition fails and where the DA starts out being non-monotone. It remains open to understand why in those models where that sufficient condition is violated the large-time behavior of the activity is still monotone as observed numerically.

References

[1] Baiesi M, Maes C and Wynants B 2009 Nonequilibrium linear response for Markov dynamics: I. Jump processes and overdamped diffusions *J. Stat. Phys.* **137** 1094–116

[2] Bodineau T and Toninelli C 2011 Activity phase transition for constrained dynamics arXiv:1101.1760v1 [cond-mat.stat-mech]

[3] Dembo A and Zeitouni O 1998 *Large Deviations Techniques and Applications* (New York: Springer)

[4] Deuschel J-D and Stroock D W 1989 *Large Deviations* (New York: Academic)

[5] Donsker M D and Varadhan S R 1975 Asymptotic evaluation of certain Markov process expectations for large time: I *Commun. Pure Appl. Math.* **28** 1–47

[6] Garrahan J P, Jack R L, Lecomte V, Pitard E, van Duijvendijk K and van Wijland F 2009 First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories *J. Phys. A: Math. Theor.* **42** 075007

[7] Garrahan J P, Sollich P and Toninelli C 2010 Kinetically constrained models arXiv:1009.6113

[8] Jack R, Garrahan J P and Chandler D 2006 Spacetime thermodynamics and subsystem observables in kinetically constrained models of glassy materials *J. Chem. Phys.* **125** 184509

[9] Kato T 1966 *Perturbation Theory for Linear Operators* (Berlin: Springer)

[10] Landauer R 1988 Motion out of noisy states *J. Stat. Phys.* **53** 233–48

[11] Maes C and Netočný K 2007 Minimum entropy production principle from a dynamical fluctuation law *J. Math. Phys.* **48** 053306

[12] Maes C, Netočný K and Wynants B 2011 Monotone return to steady nonequilibrium *Phys. Rev. Lett.* **107** 010601

[13] Maes C and Netočný K 2007 Static and dynamical nonequilibrium fluctuations *C. R. Phys.* **8** 591–7

[14] Maes C, Netočný K and Wynants B 2008 On and beyond entropy production the case of Markov jump processes *Markov Process. Rel. Fields* **14** 445–64

Maes C and Netočný K 2008 Canonical structure of dynamical fluctuations in mesoscopic nonequilibrium steady states *Europhys. Lett.* **82** 30003

[15] Nemoto T and Sasa S 2011 Thermodynamic formula for the cumulant generating function of time-averaged current *Phys. Rev.* E **84** 061113

[16] Perez-Espigares C, Kolton A B and Kurchan J 2012 An infinite family of second law-like inequalities *Phys. Rev.* E **85** 031135

[17] Schnakenberg J 1976 Network theory of microscopic and macroscopic behavior of master equation systems *Rev. Mod. Phys.* **48** 571–85