Fluctuation relations are identities, holding in non-equilibrium systems, that have attracted a lot of interest in the last 18 years. This is a series of 4 lectures discussing various aspects of such relations for stochastic equations modeling non-equilibrium processes.

**Lecture 1:** Transient fluctuation relations for Markov processes  
- Maes’ view of fluctuation relations  
- Jarzynski-Crooks-Hatano-Sasa relations for nonstationary Markov chains  
- Case of continuous time Markov processes

**Lecture 2:** 2nd Law of Stochastic Thermodynamics  
- Work, heat and entropy in stochastic thermodynamics  
- Fluctuation relations and the 2nd Law of Stochastic Thermodynamics  
- Finite time refinement of the 2nd Law and Landauer principle

**Lecture 3:** Fluctuation-dissipation relations  
- Jarzynski-Hatano-Sasa relation near stationary state  
- General Fluctuation-Dissipation Theorem  
- Green-Kubo formula for diffusions

**Lecture 4:** Large deviations and stationary fluctuation relations  
- Gallavotti-Cohen type fluctuation relations  
- Macroscopic fluctuation theory  
- A non-trivial example

---

**I. TRANSIENT FLUCTUATION RELATIONS FOR MARKOV PROCESSES**

**A. A bit of history**

The history of fluctuation relations may be traced back to the late seventies/early eighties papers by Bochkov-Kuzovlev [BK77, BK79, BK81] that were not remarked at the time. In the next development, in 1993 Evans-Cohen-Morriss observed in [ECM93] a symmetry in the distribution of fluctuations of microscopic pressure in a thermostatted particle system driven by external shear. Attempts to explain this symmetry on the theoretical ground led to the formulation of the Evans-Searles transient fluctuation relation [ES94] and of the Gallavotti-Cohen stationary fluctuation relation [GC95a, GC95b]. On the other hand, Jarzynski in 1997 proved in [J97a] a simple equality, which appeared to be closely related to the Bochkov-Kuzovlev one, see [J07a]. Originally formulated for deterministic non-equilibrium evolutions, the fluctuation relations were quickly extended to stochastic dynamics in [J97b, K98, LS99]. All these works attracted in late nineties and afterwards a widespread interest and led to an avalanche of papers. For reviews see [ES02, G08, M03, J11, SF12]. In these lectures, which are neither an

---

1 Extended version of lectures given at the Mathematics Department of Helsinki University, November 2012
exhaustive review of the subject, nor follow the thread of history, I shall discuss some topics related to fluctuation relations in which I was involved myself.

### B. Few trivial identities

As observed by Maes in [M99], in the stochastic setup, the fluctuation relations have their root in trivial identities comparing two probability measures $P(dx)$ and $P'(dx')$ absolutely continuous with respect to each other, where $x \mapsto x'$ is a, possibly trivial, involution. We shall denote by $E$ and $E'$ the expectations with respect to $P(dx)$ and $P'(dx')$, respectively. If $e^{-W(x)} = \frac{P(dx')}{P(dx)}$ is the Radon-Nikodym derivative of $P'(dx')$ with respect to $P(dx)$ then, trivially,

$$E e^{-W(x)} = 1$$

and, more generally,

$$E F(x) e^{-W(x)} = E' F(x').$$

In particular, taking $F(x) = f(W(x))$, we obtain

$$E f(W(x)) e^{-W(x)} = E' f(-W'(x))$$

where $W'(x) = -W(x')$ so that $e^{-W(x)} = \frac{P(dx')}{P(dx)}$. Eq. (1.3) implies the following relation between the probability distribution $\pi(W) dW$ of the random variable $W(x)$ with respect to $P(dx)$ and $\pi'(W) dW$ of the random variable $W'(x)$ with respect to $P'(dx)$:

$$e^{-W} \pi(W) = \pi'(-W).$$

### C. Application to Markov chains

We shall first apply the above tautological relations to discrete-time (nonstationary) Markov chain $(x_n)_{n=0}^{N+1}$ with space of states $\mathcal{X}$. Let us denote by $P_n(x, dy)$ the transition probabilities for the process and by $\mu_0(dx)$ the distribution of $x_0$. The probability space of the process may be taken as $\mathcal{X} = \mathcal{X}^{N+2}$ with the probability measure

$$P_{\mu_0}[dx] = \mu_0(dx_0) P_0(x_0, dx_1) \cdots P_N(x_N, dx_{N+1}),$$

where $x \equiv (x_0, \ldots, x_{N+1})$ (we shall denote the trajectories or functional dependence by square brackets). Suppose that $(x'_n)_{n=0}^{N+1}$ is another Markov chain with the same space of states $\mathcal{X}$ corresponding to transition probabilities $P'_n(x, dy)$ and the initial measure $\mu'_0(dx)$. It corresponds to the measure

$$P'_{\mu'_0}[dx] = \mu'_0(dx_0) P'_0(x_0, dx_1) \cdots P'_N(x_N, dx_{N+1}),$$

Let us now consider an involution $* : \mathcal{X} \to \mathcal{X}$ and let us extend it to $\mathcal{X}^{N+1}$ by combining it with the time reflection, so that for $x = (x_0, \ldots, x_{N+1})$,

$$x^* = (x_{N+1}, \ldots, x_0).$$

We may apply the scheme described at the beginning of the lecture to the case at hand defining

$$e^{-W_n[x]} = \frac{P'_{\mu'_0}[dx^*]}{P_{\mu_0}[dx]} = \frac{\mu'_0(dx^*_{N+1}) P'_0(x_0^*, dx^*_1) \cdots P'_N(x_N^*, dx^*_0)}{\mu_0(dx) P_0(x_0, dx_1) \cdots P_N(x_N, dx_{N+1})} = e^W \pi'_N(x^*),$$

inferring immediately that

$$E_{\mu_0} e^{-W_n[x]} = 1,$$

where $E_{\mu_0}$ stands for the expectation with respect to $P_{\mu_0}(dx)$ and that

$$e^{-\mathcal{W}} \pi_N(W) = \pi'_N(-W),$$

where $\pi_N(W)$ and $\pi'_N(W)$ is the probability density of $\mathcal{W}_N[x]$ and $\mathcal{W}'_N[x]$ with respect to the probability measure $P_{\mu_0}[dx]$ and $P'_{\mu'_0}[dx]$, respectively. Of course, we have to assume that the relative absolute continuity of $P_{\mu_0}[dx]$ and $P'_{\mu'_0}[dx^*]$ but this is assured if all initial and transition probability measures have positive densities with respect to a fixed measure $\lambda(dx)$ on $\mathcal{X}$ that we shall take to be the counting measure if $\mathcal{X}$ is discrete and the Lebesgue measure if $\mathcal{X} = \mathbb{R}^d$. 

2
D. Hatano-Sasa fluctuation relations for Markov chains

Suppose that \( \nu_n(dx) = e^{-\varphi_n(x)}\lambda(dx) \) are probability measures left invariant under transition probabilities \( P_n(x, dy) \):

\[
\int_X \nu_n(dx) P_n(x, dy) = \nu_n(dy). \tag{1.11}
\]

Take

\[
P'_n(x, dy) = \frac{P_n(y^*, dx^*)}{\nu_n^\ast(dx^*)} \nu_n^\ast(dy^*) \tag{1.12}
\]

for \( n^\ast \equiv N - n \), assuming again that the Radon-Nikodym derivative exists. \( P'_n(x, dy) \) are Markov transition probabilities and, for \( \nu'_n(dx) = \nu_n^\ast(dx^*) \),

\[
\int_X \nu'_n(dx) P'_n(x, dy) = \int_X \nu_n^\ast(dx^*) \frac{P_n(y^*, dx^*)}{\nu_n^\ast(dx^*)} \nu_n^\ast(dy^*) = \nu_n^\ast(dy^*) = \nu'_n(dy) \tag{1.13}
\]

(the integration is over \( x \)), so that the measures \( \nu'_n(dx) = e^{-\varphi'_n(x)}\lambda(dx^*) \) with \( \varphi'_n(x) = \varphi_n^\ast(x^*) \) are left invariant under the transition probabilities \( P'_n(x, dy) \). The process \( (x'_n) \) has the interpretation of a specific time-reversal of the original process \( (x_n) \). The Radon-Nikodym derivative (1.8) takes now the form

\[
e^{-\mathcal{W}_N[x]} = \mu'_0(dx^*_{N+1}) \nu_N(dx_N) \cdots \nu_1(dx_1) \mu_0(dx_0) = e^{\mathcal{W}_N(x^*)} \tag{1.14}
\]

and the identities (1.9) and (1.10) hold for any choice of the initial measures \( \mu_0 \) and \( \mu'_0 \) with densities

\[
\mu_0(dx) = \rho_0(x) d\lambda(x), \quad \mu'_0(dx) = \rho'_0(x) \lambda(dx^*). \tag{1.15}
\]

Explicitly,

\[
\mathcal{W}_N[x] = - \ln \mu'_0(dx^*_{N+1}) - \varphi_N(x_{N+1}) + \sum_{n=1}^{N} (\varphi_n(x_n) - \varphi_{n-1}(x_n)) + \ln \rho_0(x_0) + \varphi_0(x_0). \tag{1.16}
\]

In the particular case with the initial measures \( \mu_0 = \nu_0 \) and \( \mu'_0 = \nu'_0 \), the boundary contributions vanish so that

\[
\mathcal{W}_N[x] = \sum_{n=1}^{N} (\varphi_n(x_n) - \varphi_{n-1}(x_n)) \tag{1.17}
\]

and we obtain the relation

\[
\mathbb{E}_{\nu_0} e^{-\sum_{n=1}^{N} (\varphi_n(x_n) - \varphi_{n-1}(x_n))} = 1 \tag{1.18}
\]

which is the Markov chain version of the **Hatano-Sasa relation** [HS01]. Note that in special situations when the time reversed and the original process have the same law, relation (1.10) reduces to the property of the probability density \( \pi(W) \):

\[
e^{-W} \pi_N(W) = \pi_N(-W). \tag{1.19}
\]

E. Jarzynski and Crooks relations

Consider the special case when the transition probabilities satisfy the detailed balance relation with respect to Hamiltonians \( H_n(x) \) for inverse temperature \( \beta = \frac{1}{k_B T} \), i.e. when

\[
\lambda(dx) P_n(x, dy) = \lambda(dy) P_n(y, dx) e^{-\beta(H_n(y) - H_n(x))}. \tag{1.20}
\]
In that case, assuming that the partition functions
\[
Z_n = \int_X e^{-\beta H_n(x)} \lambda(dx)
\]  
(1.21)
are finite, the Gibbs states corresponding to Hamiltonians \( H_n \),
\[
\nu_n(dx) = Z_n^{-1} e^{-\beta H_n(x)},
\]  
(1.22)
are left invariant under the transition probabilities \( P_n(x, dx) \) and
\[
\varphi_n(x) = \beta (H_n(x) - F_n),
\]  
(1.23)
where \( F_n = -\beta^{-1}\ln Z_n \) are the free energies. The Hatano-Sasa relation (1.18) reduces in this case to the Markov-chain version of the Jarzynski equality [J97a]:
\[
\mathbb{E}_{\nu_0} e^{-\beta W_N[x]} = e^{-\beta \Delta \mathcal{F} N}
\]  
(1.24)
for \( \Delta \mathcal{F} N = F_N - F_0 \) and
\[
W_N[x] = \sum_{n=1}^{N} (H_n(x_n) - H_{n-1}(x_n)).
\]  
(1.25)
The quantity \( W_N \) may be interpreted as the work performed on the system, see more on that in the next lecture. By the Jensen inequality, identity (1.24) implies that
\[
\mathbb{E}_{\nu_0} W_N[x] \geq \Delta \mathcal{F} N
\]  
(1.26)
i.e. that the average work is bounded below by the change of the free energy between the initial and final times. The Jarzynski equality (1.24) contains, however, more information. For example, it implies that the probability of observing the trajectories with \( W_N[x] \leq \Delta \mathcal{F} N - a \) for positive \( a \) is exponentially small. Indeed,
\[
e^a \mathbb{E}_{\nu_0} 1\{W_N[x] \leq \Delta \mathcal{F} N - a\} = e^{\beta \Delta \mathcal{F} N} \mathbb{E}_{\nu_0} e^{-\beta (\Delta \mathcal{F} N - a)} 1\{W_N[x] \leq \Delta \mathcal{F} N - a\} \\
\leq e^{\beta \Delta \mathcal{F} N} \mathbb{E}_{\nu_0} e^{-\beta W_N[x]} 1\{W_N[x] \leq \Delta \mathcal{F} N - a\} \leq 1.
\]  
(1.27)
In the case with detailed balance, the time reversed transition probabilities are
\[
P_n^*(x, dy) = P_n(x^*, dy^*).
\]  
(1.28)
They preserve the Gibbs measures
\[
\nu_n^*(dx) = Z_n^{-1} e^{-\beta H_n^*(x)} \lambda(dx^*)
\]  
(1.29)
for \( H_n^*(x) = H_n(x^*) \) and \( Z_n^* = Z_n \). If
\[
W_N^*[x] = \sum_{n=1}^{N} (H_n^*(x_n) - H_{n-1}(x_n)) = -W_N[x^*],
\]  
(1.30)
and \( p_N(W) (p_N^*(W)) \) denotes the probability density of \( W_N[x] (W_N^*[x]) \) with respect to \( \mathcal{P}_N[dx] \) \( (\mathcal{P}_N^*[dx]) \) then relation (1.10) implies the Markov chain version of the Crooks relation [C99]
\[
e^{-\beta W} p_N(W) = e^{-\beta \Delta \mathcal{F} N} p_N^*(W).
\]  
(1.31)
The distribution function \( p_N^* \) on the right hand side may be replaced by \( p_N \) for time-reversible processes with \( P_n(x, dx) = P_n(x, dx) \).

The utility of Eq. (1.24) and (1.31) is that it permits to extract the free energy difference between initial and final Gibbs states in a nonstationary Markov chain with instantaneous detailed balance from the statistics of the work \( W_N \) performed on the system. Usually in thermodynamics, the free energy difference is equal to the work performed in a quasi-stationary process between the initial and final Gibbs state that requires long times. Note that here there is no assumption that the process has
to be close to stationary and although the initial state $\mu_0$ was assumed to be equal to the Gibbs one $\nu_0$, the final state $\mu_N$ (the distribution of $x_N$) is, as a rule different from $\nu_N$. Hence the interest of the above Jarzynski and Crooks identities for numerical calculations of the free energy differences of mesoscopic systems in different states, e.g. of DNA/RNA-hairpin stretching experiments and simulations, see FIG. 1. In particular, Eq. (1.31) shows that $\Delta_N F$ may be found as the value of $W$ for which $p(W) = p(-W)$.

FIG. 1: Work statistics in RNA stretching, from [R06]

F. Continuous time limit

Similar considerations apply to nonstationary continuous-time Markov processes. On the formal level, they may be obtained by taking the limit of discrete-time Markov processes when with the time-step $\epsilon$ tending to zero, with the total time interval $N\epsilon = \tau$ kept constant, the transition probabilities for $n$ such that $n\epsilon = t$ is fixed have the behavior

$$P_n(x, dy) = \epsilon w_t(x, dy) + (1 - \epsilon) \int_x w_t(x, dz) \delta_x(dy) + o(\epsilon).$$

(1.32)

The limiting continuous-time Markov process $(x_t)$ for $t \in [0, \tau]$ has the initial distribution $\mu_0(dx)$. Quantities $w(x, dy)$, that are defined modulo signed measures concentrated on the diagonal, may be distributional but are positive measures away from the diagonal and give there the transition rates of the continuous time Markov process. The backward generators of the limiting process defined by the identity

$$\frac{d}{dt} \mathbb{E}_{\mu_0} f(x_t) = \mathbb{E}_{\mu_0}(L_t f)(x_t)$$

(1.33)

are given by the formula

$$(L_t f)(x) = \int_x w_t(x, dy) f(y) - \left( \int_x w_t(x, dz) \right) f(x).$$

(1.34)

The transition probabilities $P_{s,t}(x, dy)$ of the process $(x_t)$ for $s \geq t$ evolve in time according to the equations

$$\partial_t P_{s,t}(x, dy) = -L_s(x) P_{s,t}(x, dy), \quad \partial_t P_{s,t}(x, dy) = L^*_t(y) P_{s,t}(x, dy),$$

(1.35)

where $L^*_t$ is the adjoint operator acting on measures, with the condition $P_{t,t}(x, dy) = \delta_x(dy)$. In the operator notation,

$$P_{s,t} = \exp \left[ \int_s^t L_\sigma d\sigma \right].$$

(1.36)
The time $t$ probability distributions $\mu_t(dx)$ of the process defined by

$$E_{\mu_0} f(x_t) = \int f(y) \mu_t(dy)$$

are given by the relation

$$\mu_t(dy) = \int \mu_0(dx) P_{0,t}(x,dy)$$

and evolve according to the equation

$$\partial_t \mu_t = L_t^* \mu_t.$$

On the other hand, the instantaneously invariant measures $\nu_t(dx)$ satisfy

$$L_t^* \nu_t = 0.$$

If the state space $\mathcal{X}$ is discrete, a continuous time Markov process jumps from $x(t) = x$ to $x(t+dt) = y \neq x$ with probability $w(x,dy)dt$ and otherwise stays at $x$. If $\mathcal{X} = \mathbb{R}^d$, the process is a diffusion with a drift or a jump process or a combination of both.

**Examples 1.** General diffusion process.

Consider a stochastic differential equation in $\mathbb{R}^d$

$$dx = X_{\alpha t}(x)dt + X_{\alpha t}(x) \circ dW_\alpha(t),$$

written with the Stratonovich convention indicated by symbol $\circ$, with arbitrary time-dependent vector fields $X_{\alpha t}$, $\alpha = 1, \ldots, A$ and independent standard one-dimensional Wiener processes $W_\alpha(t)$. Eq. (1.41), together with the initial distribution, defines a continuous-time Markov process with transition rates

$$w_t(x,dy) = (X_{\alpha t}(x) \cdot \nabla_x + \frac{1}{2}(X_{\alpha t}(x) \cdot \nabla_x)^2) \delta(x-y) dy$$

and the backward generator

$$L_t = X_{\alpha t} \cdot \nabla + \frac{1}{2} (X_{\alpha t} \cdot \nabla)^2.$$ 

The densities of the time $t$ measures $\mu_t(dx) = \rho_t(x) \lambda(dx)$ of the process evolve according to the Fokker-Planck equation

$$\partial_t \rho_t(x) = L_t^\dagger \rho_t(x) = -\nabla \cdot j_t(x),$$

where $L_t^\dagger$ is the formal adjoint of $L_t$ with respect to the Lebesgue measure $\lambda(dx)$ and

$$j_t(x) = \left(\rho_t X_{\alpha t} - \frac{1}{2}(\nabla \cdot (\rho_t X_{\alpha t})) X_{\alpha t}\right)(x) = (\rho_t \tilde{X}_{\alpha t} - D_t \nabla \rho_t)(x) \equiv j_{\rho_t}(x)$$

is the *probability current*, where

$$\tilde{X}_{\alpha t} = X_{\alpha t} - \frac{1}{2}(\nabla \cdot X_{\alpha t}) X_{\alpha t}, \quad D_t = \frac{1}{2}X_{\alpha t} \otimes X_{\alpha t}$$

(we employ the summation convention so $\alpha$ is summed over in the above formulas). If, following [N67], we introduce the *current velocity* $v_t(x)$ by the relation

$$j_t(x) = \rho_t(x) v_t(x)$$

then the Fokker-Planck equation (1.44) may be rewritten as the advection equation

$$\partial_t \rho_t(x) + \nabla (\rho_t v_t)(x) = 0,$$

which will be used a lot below. Note, nevertheless, that

$$v_t(x) = (\tilde{X}_{\alpha t} - D_t \nabla \ln \rho_t)(x)$$

(1.49)
where \( \rho_t \).

**Examples 2.** Langevin process.

A particular diffusion process in \( \mathbb{R}^d \) is given by the **Langevin equation**

\[
dx = M((-\nabla H_t)(x) + f_t(x)) \, dt + (2D)^{1/2} dW(t) \tag{1.50}
\]

where \( M \) (the mobility) is a matrix with nonnegative symmetric part, and \( D \) (the diffusivity) is a nonnegative matrix, \( H_t(x) \) is the time-dependent Hamiltonian, \( f_t(x) \) is a nonconservative force and \( W(t) \) the standard \( d \)-dimensional Wiener process. For simplicity, we shall take matrices \( M \) and \( D \) \( t\)- and \( x\)-independent here. The Markov process corresponding to Eq. (1.50) has the transition rates

\[
w(x, dy) = (-M(\nabla H_t)(x) + M f_t(x) + D \nabla_x) \cdot \nabla_x \delta(x - y) \, dy \tag{1.51}
\]

and the backward generator

\[
L_t = (-M(\nabla H_t) + M f_t(x) + D \nabla) \cdot \nabla. \tag{1.52}
\]

The probability current takes in this case the form

\[
j_t(x) = (-M(\nabla H_t)(x) + M f_t(x) - D \nabla) \rho_t(x) \tag{1.53}
\]

and the current velocity is

\[
v_t(x) = -M(\nabla H_t)(x) + M f_t(x) - D \nabla \ln \rho_t(x). \tag{1.54}
\]

One says that \( M \) and \( D \) satisfy the **Einstein relation** if

\[
M + M^t = 2\beta D. \tag{1.55}
\]

If this is the case and the nonconservative force \( f_t \) vanishes then the Gibbs measures

\[
\nu_t(dx) = Z_t^{-1} e^{-\beta H_t(x)} \lambda(dx) \tag{1.56}
\]

are instantaneously invariant.

**Examples 3.** (Einstein’s) Brownian motion.

In the even dimensional case set \( x = (q, p) \), where \( q \) is the position and \( p \) the momentum. Let

\[
M = \begin{pmatrix} 0 & -I \\ I & M^{-1} \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & M^{-2}D \end{pmatrix},
\]

\[
H_t(q, p) = \frac{1}{2} p \cdot m^{-1} p + V_t(x), \quad f_t(q, p) = \phi_t(x), 0, \tag{1.57}
\]

where \( M, D, m \) are half-dimension positive matrices, the first two commuting. Then the Langevin stochastic equation takes the form

\[
dq = m^{-1} p \, dt, \quad dp = \left( -M^{-1}m^{-1} p - (\nabla V)(q) + \phi_t(q) \right) dt + M^{-1}(2D)^{1/2} dW(t) \tag{1.58}
\]

which is the **underdamped** Langevin equation with the Hamiltonian dynamics accompanied by a friction and a random force. In this example, one uses the involution \( (q, p)^* = (q, -p) \) for the time reversal. The Einstein relation reads here

\[
M = \beta D \tag{1.59}
\]

aligning the friction coefficient \( M^{-1} \) with the diffusivity \( D \) (physically, they come from the same source: the interaction with the thermal environment). The case with \( V_t = 0, \phi_t = 0 \) describes the Einstein-Smoluchowski Brownian motion. In the limit \( m \to 0 \), the underdamped Langevin equation reduces to the **overdamped** one for \( q(t) \) which reads:

\[
dq = M(-\nabla V_t(q) + \phi_t) \, dt + (2D)^{1/2} dW(t) \tag{1.60}
\]

which has again the form of Eq. (1.50) with \( x = q = x^* \), \( M = M^t \) and \( D = D \).

**Examples 4.** Lévy process.
The Lévy jump process in \( \mathbb{R}^d \) corresponds to
\[
 w_t(x, dy) = w_t(d(y-x))
\]  
where \( w_t(dy) \) is a positive measure.

For the nonstationary continuous-time Markov process, the time-reversed process defined by analogy to the one for the discrete-time process corresponds to the transition rates
\[
 w_t'(x, dy) = \frac{w_t(y^*, dx^*)}{\nu_t'(dx^*)} \nu_t(dy^*) ,
\]  
where \( t^* \equiv \tau - t \) and to an initial measure \( \mu'_0 \). By a limiting argument, we may infer that the fluctuation relation carry over to the case of continuous-time Markov processes. In particular, writing
\[
 \mu_0(dx) \rho_0(x) \lambda(dx), \quad \mu_0'(dx) = \rho_0'(x) \lambda(dx^*),
\]  
for \( \varphi_t'(x) = \varphi_t^*(x^*) \) and
\[
 W_{\tau}[x] = -\ln \rho_0'(x^*) - \varphi_{\tau^*}(x^*) + \int_0^\tau \partial_t \varphi_t(x(t)) \, dt + \ln(\rho_0(x_0) + \varphi_0(x_0)) = -W_{\tau}[x^*],
\]  
where \( (x^*)_t = x^*_t \) for \( x = (x_t) \), we obtain the relations
\[
 E_{\mu_0} e^{-W_{\tau}[x]} = 1
\]  
and
\[
 e^{-W_{\tau}} \pi_{\tau}(W) = \pi_{\tau}^*(-W),
\]  
the continuous-time counterparts of Eqs. (1.9) and (1.10). In the particular case when \( \mu_0 = \nu_0 \) and \( \mu_0' = \nu_0' \), the expression (1.65) reduces to
\[
 W_{\tau}[x] = \int_0^\tau (\partial_t \varphi_t)(x(t)) \, dt
\]  
and we obtain the Hatano-Sasa equality [HS01]
\[
 E_{\nu_0} e^{-\int_0^\tau (\partial_t \varphi_t)(x_t) \, dt} = 1,
\]  
see Eq. (1.18), and, if the direct and the reversed process have the same law, also the identity
\[
 e^{-W_{\tau}} \pi_{\tau}(W) = \pi_{\tau}(-W)
\]  
that is the continuous-time version of the Eq. (1.19).

The detailed balance condition for the continuous-time Markov process is defined similarly as for the discrete time and takes the form
\[
 \lambda(dx) w_t(x, dy) = \lambda(dy) w_t(y, dx) e^{-\beta(H_t(y) - H_t(x))},
\]  
compare to Eq. (1.20). It implies again the instantaneous invariance of Gibbs measures
\[
 \nu_t(dx) = Z_t^{-1} e^{-\beta H_t(x)} \lambda(dx).
\]  
The transition rates for the time-reversed process take then the form
\[
 w_t'(x, dy) = w_t^*(x^*, dy^*).
\]
One obtains the continuum time versions of the Jarzynski equality (1.24) and the Crooks identity (1.31)

$$\begin{align*}
\mathbb{E}_{\nu_0} e^{-\beta W_{\tau}^{\ast}[x]} &= e^{-\beta \Delta \tau}, \\
 e^{-\beta W} p^r(W) &= e^{-\beta \Delta \tau} p^r_{\ast}(-W)
\end{align*}$$

where

$$W_{\tau}[x] = \int_0^\tau (\partial_t H_t)(x_t) \, dt$$

(1.75)

and $W'_{\tau}(x)$ is given by the similar formula with $H_t(x)$ replaced by $H'_t(x) = H_t(x^\ast)$ and where $p^r(W)dW$ $\langle p'_r(W)dW \rangle$ the probability density function of $W_{\tau}[x]$ $\langle W_{\tau}[x] \rangle$ in the direct (reversed) process, respectively.

G. Fluctuation relations for general diffusion processes

In [CG08], we proved general fluctuation relations for diffusion processes solving stochastic equations (1.41) on the time interval $[0, \tau]$. Upon a division of the drift field

$$X_{\alpha t}(x) = X^{\ast}_{\alpha t}(x) + X^-_{\alpha t}(x),$$

(1.76)

we defined the time-reversed process as the one corresponding to the drift and diffusion vector fields

$$X'^{\ast}_{\alpha t} = (X^+_{\alpha t},)^\ast - (X^-_{\alpha t},)^\ast, \quad X'^{-}_{\alpha t} = (X^+_{\alpha t},)^\ast,$$

(1.77)

where $X^\ast$ is the push forward of the vector field $X$ by an involution $x \mapsto x^\ast$, i.e.

$$(X^\ast)^i(x^\ast) = \frac{\partial x^i}{\partial x^j} X^j(x).$$

(1.78)

In other words, $X^{\ast}_{0^+}$ was chosen to transform by the pseudo-vector and $X^{\ast}_{0^-}$ by the vector rule under the time reversal. The choice of the rule for $X_{\alpha t}$ is immaterial. We showed, combining the Girsanov and Feynman-Kac formulas, that in this case

$$\frac{P'_{\mu_0}[dx^\ast]}{P_{\nu_0}[dx]} = e^{-W_{\tau}[x]}$$

(1.79)

for

$$W_{\tau}[x] = -\ln \rho_0(x^\ast_{0^+}) + \int_0^\tau J_t \, dt + \ln \rho_0(x_0) = -W'_{\tau}[x^\ast],$$

(1.80)

where

$$J_t = \tilde{X}^+_{\alpha t}(x_t) \cdot \nabla \varphi_t(x_t) \left( \frac{dx_t}{dt} - X^+_{\alpha t}(x_t) \right) - \nabla \cdot X^+_{\alpha t}(x_t),$$

(1.81)

in the notations of (1.46). For the sake of illustration, let us consider three particular time reversals.

1. Case (a)

First, we shall show that the case studied before is, indeed, a particular instance of such a general scheme. Upon taking

$$\tilde{X}^+_{\alpha t}(x) = -D_t(x) \nabla \varphi_t(x)$$

(1.82)

for $e^{-\Phi_t}$ such that $L^1 e^{-\Phi_t} = 0$, we obtain

$$J_t = -\nabla \varphi_t(x_t) \cdot \left( \frac{dx_t}{dt} - X^-_{\alpha t}(x_t) \right) - \nabla \cdot X^-_{\alpha t}(x_t) = -\nabla \varphi_t(x_t) \cdot \frac{dx_t}{dt}.$$

(1.83)
The last equality holds because
\[ 0 = e^{\psi_t} L_t^1 e^{-\psi_t} = e^{\psi_t} \partial_i \left( - X_{\alpha t}^1 e^{-\psi_t} + \frac{1}{2} X_{\alpha t}^j \partial_j \left( X_{\alpha t}^j e^{-\psi_t} \right) \right) \]
\[ = e^{\psi_t} \nabla \cdot \left( - X_{\alpha t}^1 e^{-\psi_t} + X_{\alpha t}^p e^{-\psi_t} \right) = - e^{\psi_t} \nabla \cdot \left( X_{\alpha t}^1 e^{-\psi_t} \right) = \left( \nabla \varphi_t \right) \cdot X_{\alpha t}^1 - \nabla \cdot X_{\alpha t}^1. \] (1.84)

Now, since
\[ - \int_0^\tau \nabla \varphi_t(x_t) \cdot \frac{dx_t}{dt} dt = - \varphi_r(x_r) + \varphi_0(x_0) + \int_0^\tau \left( \partial_t \varphi_t \right) (x_t) dt, \] (1.85)

functional (1.80) reduces in this case to expression (1.65).

2. Case (b)

In another important example that will be used below, let us take \( X_{\alpha t} = X_0 + Y_t \) with \( X_0 \) and \( X_{\alpha} \) time independent. Let \( d\nu(dx) = e^{-\varphi(x)} \lambda(dx) \) be the invariant measure for \( Y_t = 0 \). Taking
\[ \tilde{X}_{\alpha t}^+ = - \mathcal{D}(x) \nabla \varphi(x) \] (1.86)
and proceeding as before, we obtain
\[ J_t = - \nabla \varphi(x_t) \cdot \left( \sigma \frac{dx_t}{dt} - Y_t(x_t) \right) - \nabla \cdot Y_t(x_t). \] (1.87)

3. Case (c)

Finally, suppose that we split the drift vector field taking
\[ \tilde{X}_{\alpha t}^+(x) = 0. \] (1.88)

In this case,
\[ J_t = - \nabla \cdot X_{\alpha t}(x_t). \] (1.89)

The latter expression makes sense also in the limit of deterministic dynamical processes with \( X_0 = 0 \) when \( X_{\alpha t} = 0 \) and when \( J_t \) becomes the phase-space contraction rate, giving rise in this case to the Evans-Searles fluctuation relation [ES94, ES02].

Example 5. Consider the underdamped Langevin dynamics of an anharmonic chain, with phase-space points \( x = (q_0, p_0, \ldots, q_L, p_L) \), \( q_i, p_i \in \mathbb{R}^d \), that is governed by the stochastic equations
\[ dq_i = m^{-1} p_i dt, \quad dp_i = \left( - M_i^{-1} m^{-1} p_i - \nabla_{q_i} V(q) \right) + (2 \beta_i^{-1} M_i^{-1})^{1/2} dW_i(t) \] (1.90)
with scalar \( m, M_i, \beta_i > 0 \) and with the potential
\[ V(q) = \frac{k}{2} \sum_{i=1}^L (q_i - q_{i-1})^2 + \sum_{i=0}^L \left( \frac{r_i}{2} q_i^2 + \frac{g_i}{4} (q_i)^2 \right) \] (1.91)
for \( k, r, g > 0 \). The backward generator of the corresponding Markov process is
\[ L = \sum_{i=0}^L \left( m^{-1} p_i \cdot \nabla_{q_i} - (\nabla_{p_i} V)(q) \cdot \nabla_{p_i} - M_i^{-1} m^{-1} p_i \cdot \nabla_{p_i} + \beta_i^{-1} M_i^{-1} \nabla_{p_i}^2 \right). \] (1.92)

If \( \beta_i \equiv \beta \) then the dynamics (1.90) has the Gibbs state \( \nu(dx) = Z^{-1} e^{-\beta H(x)} \lambda(dx) \) as an invariant measure, where
\[ H(x) = \sum_{i=0}^L \frac{p_i^2}{2m} + V(q) \] (1.93)
is the Hamiltonian of the chain. We shall be mostly interested in the case when the dynamics inside the chain is Hamiltonian, that is when $M_i^{-1} = 0$ for $i \neq 0, L$. In that situation, the coefficients $\beta_i$ for $i \neq 0, L$ disappear from the stochastic equations (1.90). Dividing the drift $X_0$ into the pseudo-vector and vector parts so that

$$X_0^+(x) = (0, -M_i^{-1}m^{-1}p_i), \quad X_0^-(x) = (m^{-1}p_i, -\nabla q_iV(\underline{q}))$$

(1.94)

one infers from Eq. (1.81) that

$$\mathcal{J}_t = -\sum_{i=0}^{L-1} \beta_i m^{-1}p_i \cdot \left( \frac{d}{dt} \rho(x) + \nabla q_iV(\underline{q}) \right)$$

(1.95)

for such a choice. Note that the friction coefficients $M_i^{-1}$ do not enter into the latter expression which also does not depend on the choice of interpolating $\beta_i, i \neq 0, L$, if $M_i^{-1} = 0$ for $i \neq 9, L$. A simple calculation shows that

$$\mathcal{J}_t = \frac{d}{dt} \ln \rho_0(x_t) + \sum_{i=1}^{L} (\beta_i - \beta_{i-1}) j_{(i-1,i)}(x_t),$$

(1.96)

where

$$\mu_0(dx) = \rho_0(x) \lambda(dx) = 2_0^{-1} \exp \left[ \sum_{i=0}^{L} \beta_i \left( \frac{p_i^2}{2m} + \frac{r}{2} q_i^2 + \frac{g}{4N}(q_i^2)^2 \right) + \sum_{i=1}^{L} \beta_{i-1} + \beta_i, k \frac{1}{2}(q_{i-1} - q_i)^2 \right]$$

(1.97)

is the a local equilibrium measure and

$$j_{(i-1,i)}(x) = \frac{k}{2m} (p_{i-1} + p_i) \cdot (q_{i-1} - q_i)$$

(1.98)

is the energy (or heat) flux from site $i - 1$ to site $i$. Taking $\mu'_0 = \mu_0$, we obtain from Eq. (1.80) the quantity

$$W_r(x) = \sum_{i=1}^{L} (\beta_i - \beta_{i-1}) \int_0^r j_{(i-1,i)}(x_t) dt$$

(1.99)

for which the transient fluctuation relations (1.66) and (1.67) hold with all the implications, e.g. the inequality

$$E_{\mu_0} W_r(x) \geq 0.$$  

(1.100)

If the dynamics inside the chain is Hamiltonian, then different choices of $\beta_i$ for $i \neq 0, L$ lead to different fluctuation relations for the same system. E.g. for a linear interpolation between $\beta_0$ and $\beta_L$,

$$W_r(x) = (\beta_L - \beta_0) \frac{1}{L} \sum_{i=1}^{L} \int_0^r j_{(i-1,i)}(x_t) dt,$$

(1.101)

whether for the piecewise constant interpolation with a jump between sites $i - 1$ and $i$,

$$W_r(x) = (\beta_L - \beta_0) \int_0^r j_{(i-1,i)}(x_t) dt,$$

(1.102)

Different choices correspond to different local equilibrium measures $\mu_0(dx)$ none of which is a stationary state if $\beta_0 \neq \beta_L$ because

$$L^\dagger \rho_0 = \sum_{i=1}^{L} (\beta_i - \beta_{i-1}) j_{(i-1,i)} \neq 0.$$  

(1.103)

it was shown in refs. [JP99a, JP99b] that in the stationary state with an invariant measure $\nu(dx)$ (that in fact has not been proven to exist for the version of the model that we consider, see however [JP99a]),

$$E_{\nu} W_r(x) \geq 0$$

(1.104)

with the sharp inequality (for $\tau > 0$) if and only if $\beta_0 = \beta_N$ (the expectation on the right hand side of inequality (1.104), unlike the one in (1.100), is independent on the choice of interpolating $\beta_i$ - why?). Taking $W_r$ in the form (1.102), the latter result shows that, in the (putative) stationary state, the heat flows in average from the hot to the cold end of the chain, a reassuring result.
II. \textit{2nd Law of Stochastic Thermodynamics}

Although of tautological origin, the fluctuation relations considered in the previous section have important consequences that permit to make contact with the thermodynamical concepts in simple nonequilibrium situations relevant for modelisation of the dynamics of mesoscopic systems, like colloids, polymers, or bio-molecules, in contact with heat bath(s).

A. Work and heat in overdamped Langevin dynamics

We shall consider a system described by a an overdamped Langevin equation in \( \mathbb{R}^d \)

\[ dq = -M(\nabla U_t)(q) \, dt + (2D)^{1/2}dW(t), \tag{2.1} \]

with the positive mobility and diffusivity matrices \( M \) and \( D \) related by the Einstein relation

\[ M = \beta D. \tag{2.2} \]

Stochastic equation (2.1), together with the initial distribution \( \mu_0(dq) \), defines a nonstationary continuous-time Markov process with transition rates

\[ w_t(q,dq') = \left( -M(\nabla U_t) + D \nabla \right) \cdot \nabla \delta(q - q') \lambda(dq') \tag{2.3} \]

satisfying the detailed balance relations

\[ \lambda(dq) w_t(q,dq') = \lambda(dq') w_t(q', dq) e^{-\beta(U_t(q') - U_t(q))} \tag{2.4} \]

(\text{check it!}). Consequently, the Gibbs measures

\[ \nu_t(dq) = Z_t^{-1} e^{-\beta U_t(q)} \lambda(dq) \tag{2.5} \]

are instantaneously invariant, i.e. satisfy the relation \( L_t^* \nu_t = 0 \) for the backward generators

\[ L_t = \left( -M(\nabla U_t) + D \nabla \right) \cdot \nabla. \tag{2.6} \]

On the other hand, the time \( t \) probability distributions \( \mu_t(dq) = \rho_t(q) \lambda(dq) \) evolve according to the advection equation

\[ \partial_t \rho_t + \nabla(\rho_t v_t) = 0 \tag{2.7} \]

with the current velocity

\[ v_t(q) = -M(\nabla U_t)(q) - D \nabla \ln \rho_t(q), \tag{2.8} \]

see Eqs. (1.48) and (1.54). The quantity

\[ W_\tau[q] = \int_0^\tau (\partial_t U_t)(q_t) \, dt \tag{2.9} \]

has the interpretation of the work done over the system during time interval \([0, \tau]\). The white noise in the Langevin equation simulates the effect of thermal environment with which the system exchanges the heat. The heat dissipated in time interval \([0, \tau]\) is given by the formula

\[ Q_\tau[q] = -\int_0^\tau (\nabla U_t)(q_t) \circ dq(t). \tag{2.10} \]

Note that \( W_\tau \) and \( Q_\tau \) defined this way depend on the trajectory process. We shall call such quantities fluctuating. The assignment of the names may seem somewhat arbitrary and counterintuitive (it is the right-hand-side of (2.10) that looks as the work of the gradient force). For the rational behind the employed terminology, see [J07b]. Subtracting \( Q_\tau \) from \( W_\tau \), we obtain the relation

\[ W_\tau[q] - Q_\tau[q] = \int_0^\tau ((\partial_t U_t)(q_t) \, dt + (\nabla U_t)(q_t) \circ dq_t) = \int_0^\tau \frac{d}{dt} U(q_t) \, dt \]
\[ U_f(q_f) - U_0(q_0) \equiv \Delta_f U[q]. \] (2.11)

This is the **1st Law of Stochastic Thermodynamics** expressing the conservation of the energy for each realization of the process. The Jarzynski equality that we demonstrated in the previous lecture, takes now the form

\[ \mathbb{E}_{\nu_0} e^{-\beta W_f[q]} = e^{-\beta \Delta_f F}. \] (2.12)

where the initial measure is taken as the Gibbs one. By Jensen inequality, it implies that

\[ \mathbb{E}_{\nu_0} W_f[q] \geq \Delta_f F \] (2.13)

i.e. that the average work is bounded below by the change of the free energy between the initial and final times. The Jarzynski equality (2.12) contains, however, more information, implying the exponential suppression of the probability of the events for which \( W_f[q] \leq \Delta_f F - a \), as we discussed in Sec. IE.

**B. Entropy, entropy production, and the 2nd Law**

The Jarzynski equality (and the Crooks relation) extends to the case of the processes with arbitrary initial probability measures \( \mu_0(dq) = \rho_0(q) \lambda(dq) \) and \( \mu^*_0(dq) = \rho_0(q) \lambda(dq) \) and takes then the form

\[ \mathbb{E}_{\mu_0} e^{-W_f[q]} = 1, \] (2.14)

where

\[ W_f[q] = -\ln \rho_0(q_\tau) - \beta U_f(q_0) + \beta W_f[q] + \ln \rho_0(q_0) + \beta U_0(x_0) = -W[q^*] \] (2.15)

for \( (q^*)_t = q_{\tau t} \), see Eqs. (1.65) and (1.66), (1.67). The use of this freedom allows to obtain fluctuation relations for other physically important quantities. We shall take below \( \mu^*_0 = \mu_\tau \), where measure \( \mu_\tau(dq) = \rho_\tau(q) \lambda(dq) \) describes the time \( t \) distribution of the process. In other words, we assume that the reversed process starts from the time \( \tau \) distribution of the initial one. In this case,

\[ -\ln \rho_0(q_\tau) + \ln \rho_0(q_0) = -\ln \rho_\tau(q_\tau) + \ln \rho_0(q_0) \equiv -(\Delta_\tau \ln \rho)[q]. \] (2.16)

Let us define the time \( t \) fluctuating entropy of the system as

\[ S_t^{\text{fl}}(q) = -\ln \rho_t(q_t) \] (2.17)

so that

\[ \mathbb{E}_{\mu_0} S_t^{\text{fl}}(q) = -\int \ln \rho_t(q) \mu_\tau(dq) \equiv S[\mu_\tau] \] (2.18)

is the Gibbs-Shannon entropy of the measure \( \mu_\tau \), motivating the above definition (we set to 1 the Boltzmann constant \( k_B \)). Note that

\[ -(\Delta_\tau \ln \rho)[q] \equiv S_f[q] - S_0[q] \equiv \Delta_\tau S^{\text{fl}}[q] \] (2.19)

and that

\[ \mathbb{E}_{\mu_0} \Delta_\tau S_t^{\text{fl}}[q] = S[\mu_\tau] - S[\mu_0]. \] (2.20)

The change of the entropy of the system in interaction with the heat bath is accompanied by the change of the entropy of the thermal environment. Assuming that the relaxation times in the environment are much faster then that of the system (this assumption is inherent in the modelisation of the effect of the heat bath by the white noise), all processes in the environment may be considered quasi-stationary on the times scales relevant for the evolution of the systems and the change of the entropy of the environment may be related to the heat dissipation by the Clausius formula

\[ \Delta_\tau S^{\text{env}}[q] = \beta Q_f[q]. \] (2.21)
(extended to fluctuating quantities). Summing up the changes of the system and environment entropies, we obtain the total entropy production:

$$
\Delta_{\tau} S^{\text{tot}}[q] = \Delta_{\tau} S^{\text{sys}}[q] + \Delta_{\tau} S^{\text{env}}[q] = -(\Delta_{\tau} \ln \rho)[q] + \beta Q_{\tau}
$$

$$
= -(\Delta_{\tau} \ln \rho)[q] - \beta(\Delta_{\tau} U[q] - W_{\tau}[q]) = W_{\tau}[q],
$$

where $W_{\tau}$ is given by Eq. (2.15) and in the last line we used the 1st Law of Stochastic Thermodynamics (2.11). The last identity permits to rewrite Eq. (2.14) in the form

$$
E_{\mu_0} e^{-\Delta_{\tau} S^{\text{tot}}[q]} = 1
$$

first stated in this context in [S05]. The fluctuation relation (2.23) implies by the Jensen inequality the 2nd Law of Stochastic Thermodynamics:

$$
E_{\mu_0} \Delta_{\tau} S^{\text{tot}}[q] \geq 0
$$

Again, Eq. (2.23), or its Crooks’ extension

$$
e^{-\Delta S} p_{\tau}(\Delta S) = p'(\Delta S)
$$

contains more information than the 2nd Law, implying e.g. that the probability of fluctuations leading to total entropy production $\leq -s$ are exponentially suppressed:

$$
E_{\mu_0} 1_{\{\Delta_{\tau} S^{\text{tot}}[q] \leq -s\}} \leq e^{-s}.
$$

Note that if $\mu_0$ is equal to the Gibbs measure $\nu_0$ then inequality 2.24) reduces to

$$
E_{\nu_0} \left( \beta W[q] - \beta \Delta_{\tau} F - S[\mu_{\tau}||\nu_{\tau}] \right) \geq 0,
$$

where

$$
S[\mu_{\tau}||\nu_{\tau}] = \int \ln \left( \frac{\mu_{\tau}(dq)}{\nu_{\tau}(dq)} \right) \mu_{\tau}(dq) \geq 0
$$

is the entropy of the time $\tau$ distribution of the process relative to the time $\tau$ Gibbs measure. Hence the 2nd Law (2.24) implies a stronger version of inequality (2.13).

### C. Landauer Principle

The 2nd Law (2.24) may be also written as the inequality

$$
E_{\mu_0} Q_{\tau}[q] \geq -\beta^{-1} E_{\mu_0} \Delta_{\tau} S^{\text{sys}}[q] = -\beta^{-1} (S[\mu_{\tau}] - S[\mu_0])
$$

for the mean dissipated heat in terms of the change of the Gibbs-Shannon entropy between the initial and final statistical state of the system. In this form, the 2nd Law is closely related to the principle formulated by Landauer in 1961 [L61], see also [B82], stating that the erasure of one bit of information during a computation process conducted in thermal environment requires a release of heat equal to at least $\beta^{-1}\ln 2$ (in average). As an example, consider a bi-stable system that may be in two distinct states and undergoes a process that at final time leaves it always in, say, the second of those states, loosing the memory of the initial state. Such a device may be realized in the context of Stochastic Thermodynamics by an appropriately designed Langevin evolution that starts from the Gibbs state corresponding to a potential $R_0 = -\beta^{-1}\ln \rho_0$ with two symmetric wells separated by a high barrier and ends in a Gibbs state corresponding to a potential $R_{\tau} = -\beta^{-1}\ln \rho_{\tau}$ with only one of those wells, see FIG. 2. The change of system entropy in such a process is approximately

$$
S[\mu_{\tau}] - S[\mu_0] \approx -\ln 1 + 2(\ln \frac{1}{2})\frac{1}{2} = -\ln 2,
$$

with better and better approximation the deeper the wells or the lower the temperature, and Landauer’s lower bound for average heat release follows from inequality (2.29).
D. Means of work, heat and total entropy production

It is well known that the 2nd Law bound can be saturated for quasi-static processes that move infinitely slowly so that at intermediate times the instantaneous measures $\mu_t$ are (almost) equal to the Gibbs measures $\nu_t$. Suppose however, that we cannot afford to go too slowly. Indeed, in computational devices, we are interested in fast dynamics that arrives at the final state quickly but produces as little heat as possible. We are therefore naturally led to two questions:

- What is the lower bound for the total entropy production or the average heat release in the process that interpolates between given states in a time interval of fixed length?
- What is the dynamical protocol that leads to such a minimal total entropy production or heat release?

These questions make sense in a variety of setups. They are among the core ones of the so called Finite-Time Thermodynamics [A11] that was developed during last decades mostly with an eye on technological applications. Here we shall study them in the framework of Stochastic Thermodynamics of processes described by the overdamped Langevin equation (2.1) following [AM11, AG12].

First, note using the advection equation (2.7), that

$$
\mathbb{E}_{\mu_0} \Delta_\tau S^{sys}[\mathbf{q}] = S[\mu_\tau] - S[\mu_0] = -\int_0^\tau \frac{d}{dt} \int \ln \rho_t(q) \rho_t(q) \lambda(dq)
$$

$$
= \int_0^\tau dt \int [\nabla \cdot (\rho_t v_t)(q) + \ln \rho_t(q) \nabla \cdot (\rho_t v_t)(q)] \lambda(dq) = -\int_0^\tau dt \int \nabla (\ln \rho_t(q))(\rho_t v_t)(q) \lambda(dq)
$$

(2.31)

assuming that there are no boundary contributions from the integration by parts. On the other hand, using the 1st Law, we infer that

$$
\mathbb{E}_{\mu_0} \beta Q_\tau[\mathbf{q}] = -\beta \mathbb{E}_{\mu_0} (\Delta_\tau U[\mathbf{q}] - W_\tau[\mathbf{q}]) = -\beta \int_0^\tau dt \int \left[ \frac{d}{dt} \int U_t(q) \mu_t(dq) - \int (\partial_t U)(q) \mu_t(dq) \right]
$$

$$
= -\beta \int_0^\tau dt \int U_t(q) \partial_t \rho_t(q) \lambda(dq) = \beta \int_0^\tau dt \int U_t(q) \nabla (\rho_t v_t)(q) \lambda(dq)
$$

$$
= -\beta \int_0^\tau dt \int (\nabla U_t)(q)(\rho_t v_t)(q) \lambda(dq).
$$

(2.32)

Summing the expectations (2.31) and (2.32), we obtain, employing the expression for the current velocity $v_t$ from (2.8), the identity

$$
\mathbb{E}_{\mu_0} \Delta_\tau S^{tot}[\mathbf{q}] = \beta \int_0^\tau dt \int v_t(q) \cdot M^{-1} v_t(q) \rho_t(q) \lambda(dq)
$$

(2.33)
from which 2nd Law inequality follows directly.

We would like to find the minimum of the right hand side over all control potentials \( U_t \) that lead to the overdamped Langevin evolution (2.1) from the fixed initial density \( \rho_0(q) \) to the fixed final one \( \rho_T(q) \) in the fixed time interval \([0, \tau]\).

### E. Benamou-Brenier minimization and optimal mass transport

In [BB97, BB99], Benamou-Brenier solved a closely related problem of minimization of the right hand side of Eq. (2.33) over velocity fields \( v_t(q) \) constrained by the advection equation in (2.7), with the densities \( \rho_t \) fixed at the initial and final times but without imposing the special gradient form of \( v_t \). This was done as follows. Introducing the Lagrangian flow \( q_t(x) \) for velocities \( v_t \) satisfying the ODE

\[
\partial_t q_t(q_0) = v_t(q_t(q_0)), \quad q_t(q_0)|_{t=0} = q_0, \tag{2.34}
\]

and writing the solution of the advection equation in the form

\[
\rho_t(q) = \int \delta(q - q_t(q_0)) \, \rho_0(q_0) \, \lambda(dq_0), \tag{2.35}
\]

one may rewrite the functional on the right hand side of Eq. (2.33):

\[
\int_0^\tau dt \int v_t(q) \cdot M^{-1} v_t(q) \, \rho_t(q) \, \lambda(dq) = \int_0^\tau dt \int \lambda(dq) \int v_t(q) \cdot M^{-1} v_t(q) \, \delta(q - q_t(q_0)) \, \lambda(dq_0)
= \int_0^\tau dt \int (\partial_t q_t)(q_0) \cdot M^{-1} (\partial_t q_t)(q_0) \, \rho_0(q_0) \, \lambda(dq_0)
= \int_0^\tau \int (\partial_t q_t)(q_0) \cdot M^{-1} (\partial_t q_t)(q_0) \, dt \, \rho_0(q_0) \, \lambda(dq_0). \tag{2.36}
\]

In the first step, one minimizes the right hand side over the curves \( t \mapsto q_t(q_0) \) with fixed endpoint \( q_T(q_0) \). The minima are realized on straight lines leading to the functional

\[
\frac{1}{\tau} \int (q_T(q_0) - q_0) \cdot M^{-1} (q_T(q_0) - q_0) \, \rho_0(q_0) \, \lambda(dq_0)
\]

that is the quadratic cost function of the map \( q_0 \mapsto q_T(q_0) \). In the second step, one is left with the celebrated optimal mass transport problem of Monge-Kantorovich [V03] consisting of the minimization of the quadratic cost (2.37) over the diffeomorphisms \( q_0 \mapsto q_T(q_0) \) under the constraint that

\[
\int \delta(q - q_T(q_0)) \, \rho_0(q_0) \, \lambda(dq_0) = \rho_T(q), \tag{2.38}
\]

or, equivalently

\[
\rho_T(q_T(q_0)) \frac{\partial q_T(q_0)}{\partial q_0} = \rho_0(q_0), \tag{2.39}
\]

i.e. that the map \( q_0 \mapsto q_T(q_0) \) transports the density \( \rho_0 \) to \( \rho_T \). One of the results of the optimal transport theory states that if \( \rho_0 \) and \( \rho_T \) are smooth and have two moments then the minimal cost is attained on a unique diffeomorphism that is a gradient of a smooth convex function

\[
q_T(q_0) = M \nabla \psi(q_0). \tag{2.40}
\]

The corresponding minimizing velocity \( v_t \) with the linear Lagrangian flow \( q^{lin}_t(q_0) = t q_0 + (1-t) q_T(q_0) \) satisfies the inviscid Burgers equation

\[
\partial_t v + v_t \cdot \nabla v_t = 0 \tag{2.41}
\]

(which just says that the Lagrangian trajectories have no acceleration). Even more importantly for us, as a consequence of Eq. (2.40), the minimizing velocity \( v_t \) is also of a gradient type:

\[
v_t = M \nabla \psi_t = 0 \tag{2.42}
\]
Theorem (Finite-time refinement of the 2nd Law and of the Landauer bound). The mean total entropy production in an overdamped Langevin evolution during time \( \tau \) between the states with probability densities \( \rho_0 \) and \( \rho_\tau \) satisfies the bound

\[
\mathbb{E}_{\rho_0} \Delta S^{\text{tot}}[\mathbf{q}] \geq \frac{\beta}{\tau} K[\rho_0, \rho_\tau]
\]

where \( K[\rho_0, \rho_\tau] \) is the minimal quadratic cost (2.37) of transport of density \( \rho_0 \) to \( \rho_\tau \). The bound is saturated by the optimal protocol satisfying Eq. (2.46), where \( \psi_t \) and \( \rho_t \) are given by Eqs. (2.43) and (2.35), respectively, in terms of the linear interpolation \( q_t^{\text{lin}}(q_0) \) between \( q_0 \) and its image \( q_\tau(q_0) \) under the optimal transport map.

The above result has a geometric interpretation. The minimal quadratic cost \( K[\rho_0, \rho_\tau] \) is, by definition, the square of the Wasserstein distance between the measures \( \mu_0 \) and \( \mu_\tau \) that, formally, corresponds to the Riemannian metric on the space of probability densities [JKO98], with the square of the tangent vectors given by

\[
\|\partial_t\rho\|_W^2 = \int [(\partial_t\rho) (-\nabla \cdot \rho \nabla)^{-1} (\partial_t\rho)](q) \lambda(dq).
\]

The Fokker-Planck equation for the (1.44) corresponds to the gradient flow in metric (2.48) for the free energy functional

\[
\mathcal{F}_t(\rho) = \int [U + \beta^{-1} \ln \rho](q) \rho(q) \lambda(dq)
\]

and one has

\[
\mathbb{E}_{\rho_0} \Delta S^{\text{tot}}[\mathbf{q}] = \int_0^\tau \|\partial_t\rho\|_W^2 \, dt \geq dW(\mu_0, \mu_\tau)^2
\]

with the optimal protocol giving the (shortest) geodesic between \( \rho_0 \) and \( \rho_\tau \).

Corollary. Under the same assumptions, the mean heat release satisfies the bound

\[
\mathbb{E}_{\rho_0} Q_\tau[\mathbf{q}] \geq -\beta^{-1} (S[\mu_\tau] - S[\mu_0]) + \frac{1}{\tau} K[\rho_0, \rho_\tau],
\]

for \( q = q_t^{\text{lin}}(q_0) \) (2.43) satisfies the Hamilton-Jacobi equation

\[
\partial_t \psi_t + \frac{1}{\tau} (\nabla \psi_t) \cdot M(\nabla \psi_t) = 0.
\]
with the inequality saturated by the same optimal protocol.

The latter inequality providing a finite-time refinement of the bound (2.29), implies also a finite-time refinement of the Landauer bound in the situation where the change of the mean system entropy is given by Eq. (2.30). Such a refinement may be relevant in future computer designs [S11] (the present day computers still dissipate much more heat than the minimum allowed by the thermodynamical considerations). A recent experiment [B12] with a colloidal particle manipulated by laser tweezers measured the heat released in a process of memory erasure interpolating at room temperature between two states with Gibbs potential from FIG. 2, with the results plotted in FIG. 3.

![FIG. 3](image)

**FIG. 3:** Mean heat release as a function of time in the memory erasure experiment of [B12]

The control potential used in the experiment did not follow the optimal protocol and, as a result, the heat release in the 10s run exceeded 2.5 times the Landauer bound instead of the optimal 40%.

![FIG. 4](image)

**FIG. 4:** Gibbs and control potentials $R_t$ and $U_t$ for optimal 10s (left) and 1s (right) runs

![FIG. 5](image)

**FIG. 5:** Potentials $R_t$ and $U_t$ for times $\frac{3}{10}\tau$ and $tau$ (left) and for $\frac{15}{10}\tau$ and $\tau$ (right) for 1s run

The optimal protocols for 10s and 1s runs (the latter releasing almost 4 times more heat than the Landauer bound) are illustrated on FIG. 4 at initial, half- and final time. For 10s, the Gibbs potentials
\[ R_t = -\beta^{-1} \ln \rho_t \] are very close to control potentials \( U_t \) so that the optimal protocol is almost quasi-static. For \( t = 1 \), the \( R_t \) differ considerable from \( U_t \), also at the initial and final times, showing also more intricate structure in late times, with the persistence of the barrier separating the two wells, see FIG. 5. The initial and final jumps of the potential in the optimal protocol were first discovered in the case with quadratic potentials by an explicit calculation in [SS07]. The optimal transport map \( q_0 \mapsto q_\tau(q_0) \) leading to the optimal protocol has in the problem in question a form of a kink on the shifted identity map and the corresponding current velocities \( v_t \) build in time an (almost) shock, see FIG. 6.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6}
\caption{Optimal map \( q_0 \mapsto q_\tau(q_0) \) (left) and the corresponding current velocities (right)}
\end{figure}

In one dimensional problem as above, the optimal map \( q_0 \mapsto q_\tau(q_0) \) may be found numerically by sorting the points distributed with densities \( \rho_0 \) and \( \rho_\tau \) in the increasing order. In more than one dimension, finding such maps requires a more sophisticated Auction Algorithm, see [BF03].

Above, we dealt with nonstationary overdamped Langevin evolution without non-conservative forces. Adding such forces \( \phi_t \) as in Eq. (1.60) but keeping the Einstein relation (1.59) modifies the expression (2.10) for the dissipated heat to

\[ Q_\tau[q] = -\int_0^\tau (\nabla U_t + \phi_t)(q_t) \circ dq(t). \] (2.52)

The Fokker-Planck equation takes still the form of the advection equation (2.7) but the expression for the current velocity is modified by the addition of \( \phi_t \) to

\[ v_t(q) = M(-\nabla U_t + \phi_t)(q) - D\nabla \ln \rho_t(q) \] (2.53)

and is no more of the gradient type. The mean total entropy production is still given by Eq. (2.33). Since the Benamou-Brenier minimization held for arbitrary velocities, the bounds (2.47) and (2.51) hold in this case as well, but are saturated by the protocol discussed above without a nonconservative force.

More work on optimization in Stochastic Thermodynamics may be found in [GM08, EK10, AM12, PM12]. One of the problems still open is the extension of the above analysis to general underdamped Langevin evolutions whose relation to the overdamped limit contains some surprises [H80, SSD82, CB12].

III. FLUCTUATION-DISSIPATION RELATIONS

This is a lecture devoted to the relation between the fluctuation relation and laws of the linear response to perturbations around stationary states that, historically, were among the first results about nonequilibrium dynamics.

A. Hatano-Sasa fluctuation relation and general Fluctuation-Dissipation Theorem

Recall that the Hatano-Sasa transient fluctuation relation (1.69) holds for arbitrary continuous time nonstationary Markov process with backward generators \( L_t \) and a family of measures \( \nu_t(x) =\)
$e^{-\varphi(x)}\lambda(dx)$ such that $L^*_t\nu_0 = 0$. Now, following [PJP09], see also [H78], let us consider a family of stationary transition rates $w_\epsilon(x,dy)$ parametrized by $\epsilon = (\epsilon^s)$ in a neighborhood of $\epsilon = 0$, corresponding to backward generators $L_\epsilon$ with a family of invariant measures $\nu_\epsilon(dx) = e^{-\varphi_\epsilon(x)}\lambda(dx)$ such that

$$L^*_\epsilon\nu_\epsilon = 0. \quad (3.1)$$

For each protocol $t \mapsto \epsilon_t$ such that $\epsilon_t = 0$ for $t \leq 0$, we shall consider for $t \geq 0$ the nonstationary Markov process with backward generators $L_{\epsilon_t}$ that starts from the measure $\nu \equiv \nu_0$. For each of those protocols $\epsilon = (\epsilon_t)$, we have the Hatano-Sasa relation

$$E^{\epsilon}_t e^{-\int_0^t \frac{d\epsilon_s}{dt} \partial_{\epsilon_s} \varphi_\epsilon(x_s) \, dt} = 1. \quad (3.2)$$

Expanded up to the second order in $\epsilon$, this gives

$$-\int_0^\tau \frac{d\epsilon_t^a}{dt} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \, dt - \int_0^\tau \frac{d\epsilon_t^b}{dt} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \, dt$$

$$- \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{dt} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \, ds \bigg|_{\epsilon = 0} \frac{d\epsilon_t^a}{dt} \partial_{\epsilon_t} \varphi(x_t) \, ds = 0, \quad (3.3)$$

where $\varphi \equiv \varphi_0$ and $E_{\nu_t} \equiv E^{\epsilon}_t$. Note that by normalization,

$$\int e^{-\varphi(x)}\lambda(dx) = 1 \quad (3.4)$$

for all $\epsilon$. Expanding the latter relation to the second order, we obtain

$$\int \left( - e^a \partial_{\epsilon^a} \varphi(x) - \frac{1}{2} e^a e^b \partial_{\epsilon^a} \varphi(x) \partial_{\epsilon^b} \varphi(x) + \frac{1}{2} e^a e^b \partial_{\epsilon^a} \varphi(x) \partial_{\epsilon^b} \varphi(x) \right) e^{-\varphi(x)}\lambda(dx) = 0. \quad (3.5)$$

The last identity implies that

$$E_{\nu_t} \partial_{\epsilon^a} \varphi(x_t) = \int \partial_{\epsilon^a} \varphi(x_t) e^{-\varphi(x)} \lambda(dx) = 0 \quad (3.6)$$

and that

$$E_{\nu_t} \partial_{\epsilon^a} \partial_{\epsilon^b} \varphi(x_t) = \int \partial_{\epsilon^a} \partial_{\epsilon^b} \varphi(x_t) e^{-\varphi(x)} \lambda(dx) = \int \partial_{\epsilon^a} \varphi(x_t) \partial_{\epsilon^b} \varphi(x_t) e^{-\varphi(x)} \lambda(dx)$$

$$= E_{\nu_t} \partial_{\epsilon^a} \varphi(x_t) \partial_{\epsilon^b} \varphi(x_t), \quad (3.7)$$

where the right hand side is $t$-independent. Substituting these identities to Eq. (3.3), we obtain:

$$-\frac{1}{2} \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \partial_{\epsilon_s} \varphi(x_s)$$

$$- \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \partial_{\epsilon_s} \varphi(x_s) \bigg|_{\epsilon = 0} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \, ds$$

$$+ \frac{1}{2} \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_s) \partial_{\epsilon_s} \varphi(x_t) \, ds$$

$$= -\frac{1}{2} \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_s) \partial_{\epsilon_s} \varphi(x_t) \, ds$$

$$- \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \partial_{\epsilon_s} \varphi(x_t) \, ds$$

$$+ \frac{1}{2} \int_0^\tau \frac{d\epsilon_t^a}{dt} \int_0^\tau \frac{d\epsilon_s^b}{ds} E_{\nu_t} \partial_{\epsilon_t} \varphi(x_t) \partial_{\epsilon_s} \varphi(x_t) \, ds$$
Upon stripping of the arbitrary functions $\frac{d\nu}{dt}$, the last equation is equivalent to the identity

$$
\mathbb{E}_\nu \partial_s \varphi(x_s) \partial_s \varphi(x_t) - \mathbb{E}_\nu \partial_s \varphi(x_t) \partial_s \varphi(x_s)
= \int_0^\infty \frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s \varphi(x_s) \, ds + \int_0^t \frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s \varphi(x_s) \, d\sigma
$$

(3.9)

Because, by causality,

$$
\frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s \varphi(x_s) = 0
$$

for $s \leq \sigma$, Eq. (3.9) reduces upon taking $s \leq t$ to

$$
\mathbb{E}_\nu \partial_s \varphi(x_s) \partial_s \varphi(x_t) - \mathbb{E}_\nu \partial_s \varphi(x_t) \partial_s \varphi(x_s) = \int_0^t \frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s \varphi(x_s) \, d\sigma,
$$

(3.11)

or, in the differential form, to

$$
\partial_s \mathbb{E}_\nu \partial_s \varphi(x_s) \partial_s \varphi(x_t) = -\frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s \varphi(x_s) \bigg|_{\epsilon=0}.
$$

(3.12)

This is one of general forms of the **Fluctuation-Dissipation Theorem** (FDT) [H78, PJP09]. The left hand side is the time derivative of the **dynamical correlation function** of observables $\partial_s \varphi(x)$ in the stationary state, whereas the quantity on the right hand side is the **response function** measuring the change of the dynamical one point function of $\partial_s \varphi(x)$ under a small perturbation of $\epsilon$ concentrated around an earlier time that makes the dynamics nonstationary. The entry $\partial_s \varphi(x)$ plays a passive role in the identity (3.12) and could be replaced by an arbitrary function $O_\nu(x)$. On the other hand, for a general stationary dynamical correlation function with $s \leq t$ one has

$$
\partial_t \mathbb{E}_\nu O_b(s) O_a(t) = \int \nu(dx) O_b(x) \int \partial_t P_{s,t}(x, dy) O_a(y) = \mathbb{E}_\nu O_b(s) (L O_a)(t),
$$

$$
\partial_s \mathbb{E}_\nu O_b(s) O_a(t) = \int \nu(dx) O_b(x) \int \partial_s P_{s,t}(x, dy) O_a(y) = -\mathbb{E}_\nu (L' O_b)(s) (L O_a)(t),
$$

(3.13)

where

$$
L' = e^{\beta L} e^{-\beta}
$$

is the adjoint of $L$ with respect to the invariant measure $\nu(dx)$. We infer that the FDT (3.12) may be rewritten in the form

$$
\mathbb{E}_\nu L'(\partial_s \varphi(x_s)) O_a(x_t) = \frac{\delta}{\delta \epsilon_\nu} \mathbb{E}_\nu \partial_s O_a(x_t).
$$

(3.15)

**B. Other forms of the Fluctuation-Dissipation Theorem**

Let us consider a particular family of transition rates of the form

$$
w_{\epsilon}(x, dy) = e^{-\frac{\beta}{\epsilon} O_a(x)} w(x, dy) e^{\frac{\beta}{\epsilon} O_a(y)},
$$

(3.16)

for a family of functions $O_a(x)$, corresponding to the perturbed backward generators

$$
L_{\epsilon} = e^{-\frac{\beta}{\epsilon} O_a(x)} L e^{\frac{\beta}{\epsilon} O_a(x)} - e^{-\frac{\beta}{\epsilon} O_a(x)} \left( L e^{\frac{\beta}{\epsilon} O_a(x)} \right)
\approx \frac{\beta}{\epsilon} \left( [L, O_a] - (L O_a) \right) + o(\epsilon).
$$

(3.17)
Above, $\beta$ is introduced just for dimensional reason, see however below. The invariance condition (3.1) for the measures $\nu_\epsilon(dx)$ gives now to the 1st order in $\epsilon$ the condition

$$
e^\alpha L^\dagger(e^{-\varphi} \partial_\epsilon \varphi) + \frac{\beta}{\epsilon} e^\alpha \left([L, O_a] - (LO_a)\right)^\dagger \left(e^{-\varphi}\right) = 0,$$

i.e.

$$L' \left(\partial_\epsilon \varphi\right) = \frac{\beta}{\epsilon} (L'O_a + LO_a).$$

Plugging this expression into Eq. (3.15), we may rewrite it in the form

$$\partial_\epsilon \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) - \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) = \frac{2}{\beta} \frac{\delta}{\delta \epsilon_1} \bigg|_{\epsilon=0} \mathbb{E}_\nu \, O_a(x_t).$$

This is another form of the general FDT [CKP94, LCZ05, BMW09]. It does not require the knowledge of the invariant states, but involves explicitly the generator of the stationary process.

We may also rewrite the right hand side of Eq. (3.19) as $\beta L'O_a + \frac{\beta}{\epsilon} (L-L'O_a)$. This results in yet another form of the general FDT:

$$\partial_\epsilon \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) - \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) = \frac{1}{\beta} \frac{\delta}{\delta \epsilon_1} \bigg|_{\epsilon=0} \mathbb{E}_\nu \, O_a(x_t).$$

The latter form is useful for the Langevin process were

$$L = (-M(\nabla H) + M f + D \nabla) \cdot \nabla$$

and were

$$L' = e^\varphi L^\dagger e^{-\varphi} = (M(\nabla H) - M f) \cdot \nabla + 2 e^\varphi (\nabla \cdot e^{-\varphi}) D \nabla + \nabla \cdot D \nabla$$

since $L^\dagger(e^{-\varphi}) = 0$. Hence in this case,

$$\frac{1}{2}(L-L') = (-M(\nabla H) + M f - e^\varphi D(\nabla e^{-\varphi})) \cdot \nabla = v \cdot \nabla,$$

where $v(x)$ is the current velocity, see Eq. (1.54), in the stationary state. Using the the latter relation, we obtain the general FDT for the Langevin dynamics [CFG08]:

$$\partial_\epsilon \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) - \mathbb{E}_\nu \, (v \cdot \nabla O_b)(x_s) \, O_a(x_t) = \frac{1}{\beta} \frac{\delta}{\delta \epsilon_1} \bigg|_{\epsilon=0} \mathbb{E}_\nu \, O_a(x_t).$$

Note that for the stationary Langevin process with $M = M' = \beta D$, the perturbation (3.16) corresponds to the change of the Hamiltonian $H(x) \rightarrow H(x) - e^\alpha O_a(x)$.

In the situation with detailed balance for the stationary process with $\epsilon = 0$, see Eq. (1.71), one the generators $L$ and $L'$ coincide and $\varphi_\epsilon(x) = \beta (H(x) - e^\alpha O_a(x) - F_\epsilon)$, where $F_\epsilon$ is a constant. As a result, all three forms of the FDT reduce to the relation

$$\partial_\epsilon C^{ab}(t-s) \equiv \partial_\epsilon \mathbb{E}_\nu \, O_b(x_s) \, O_a(x_t) = \frac{1}{\beta} \frac{\delta}{\delta \epsilon_1} \bigg|_{\epsilon=0} \mathbb{E}_\nu \, O_a(x_t) \equiv \frac{1}{\beta} R^{ab}(t-s)$$

which is the classic equilibrium FDT [K57] relating the equilibrium dynamical correlation function $C^{ab}(t-s)$ to the response function $R^{ab}(t-s)$ of the equilibrium state to small perturbations.

**Example 6.** For the Einstein-Smoluchowski Brownian motion of Example 3 with scalar mass matrix $m$, $O_a(q, p) = p^a$ and $O_b(q, p) = q^b$, the stationary form of the dynamical correlation function is

$$C^{ab}(t-s) = m D \delta^{ab} e^{-\frac{m(t-s)}{D}}$$

(there is no complete stationary state in that case since the expectation value of $q^2_t$ diverges linearly in time), and the response function takes the form

$$R^{ab}(t-s) = \delta^{ab} e^{-\frac{t-s}{D}}.$$
Relation (3.26) reduces then to the Einstein relation (1.59), a prototype of the equilibrium FDT.

The possible usage of the FDT (3.26) is for extracting the response function, more difficult to measure, from the stationary dynamical correlation, more easily accessible, or for inferring the temperature $\beta^{-1}$ of a system in thermal equilibrium if both the response function and the dynamical correlation are accessible. Although near a nonequilibrium stationary states (NESS), the FDT does not have such a simple form, the ratio of the dynamical correlation to the response function is often used out of equilibrium to define effective temperatures, in particular in glassy systems [C11].

It was observed in [CFG08], see also [SS06], that the form (3.25) of the FDT for Langevin systems implies that one recovers the equilibrium form of the FDT in Lagrangian frame of the current velocity. It was shown subsequently in [CG09] that any nonequilibrium Langevin diffusion rewritten in the Lagrangian frame of its current velocity recovers the detailed balance property.

C. Relation of the response function to dissipation

The original name of the Fluctuation-Dissipation Theorem for the identity (3.26) comes from the fact that the dynamical 2-time function describes the correlation of fluctuations of the random variables $O(x_t)$ whereas the response function is related to dissipation of energy or heat. To understand the latter connection, let us consider the case of periodic a perturbation of the stationary equilibrium dynamics by taking $\epsilon_t = \epsilon_0 \cos(\omega t)$ so that $H_t(x) = H(x) - \epsilon_0 \sin(\omega t) O(x)$. For such a system, the average expectation of the work

$$W_\tau[x] = \int_0^\tau (\partial_t H_t)(x_t) \, dt = \epsilon_0 \omega \int_0^\tau \sin(\omega t) O(x_t) \, dt$$

considered per unit time:

$$\lim_{\tau \to \infty} \frac{1}{\tau} \mathbb{E}_\nu^{\epsilon_0} W_\tau[x] = \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \int_0^\tau \sin(\omega t) \mathbb{E}_\nu^{\epsilon_0} O(x_t) \, dt$$

$$= \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \int_0^\tau \sin(\omega t) \left( \mathbb{E}_\nu^{\epsilon_0} O(x_t) + \epsilon_0 \int_0^t \cos(\omega s) \frac{\delta}{\delta \epsilon_s} \mathbb{E}_\nu O(x_s) \, ds + o(\epsilon_0) \right) \, dt.$$  \hspace{1cm} (3.30)

The stationary contribution vanishes in the long time limit, so that denoting

$$\frac{\delta}{\delta \epsilon_s} \mathbb{E}_\nu O(x_t) \equiv R(t - s)$$

we obtain

$$\lim_{\tau \to \infty} \frac{1}{\tau} \mathbb{E}_\nu^{\epsilon_0} W_\tau[x] = \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \int_0^\tau dt \int_0^t \sin(\omega s) \cos(\omega \tilde{t} - s) \, ds + o(\epsilon_0^2)$$

$$= \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \int_0^\tau \cos(\omega s) ds \int_0^{\tau - s} \sin(\omega s + \sigma) R(\sigma) \, d\sigma + o(\epsilon_0^2)$$

$$= \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \left( \int_0^\tau \sin(\omega s) \cos(\omega \tilde{s}) ds \int_0^{\tau - s} \cos(\omega \sigma) R(\sigma) \, d\sigma \right)$$

$$+ \int_0^\tau \cos^2(\omega s) ds \int_0^{\tau - s} \sin(\omega \sigma) R(\sigma) \, d\sigma + o(\epsilon_0^2)$$

$$= \left( \lim_{\tau \to \infty} \frac{\epsilon_0 \omega}{\tau} \int_0^\tau \cos^2(\omega s) ds \right) \int_0^{\infty} \sin(\omega \sigma) R(\sigma) \, d\sigma + o(\epsilon_0^2)$$

$$= \frac{\epsilon_0 \omega}{2} \Im \tilde{R}(\omega) + o(\epsilon_0^2),$$  \hspace{1cm} (3.32)
assuming the integrable decay of \( R(\sigma) \) when \( \sigma \to \infty \), where

\[
\hat{R}(\omega) = \int_{0}^{\infty} e^{i\omega\sigma} R(\sigma) \, d\sigma
\]  

(3.33)

is the Fourier transform of the response function. In experiments, often the dissipation rate is directly measured, giving access to the imaginary part of the Fourier-space response function \( \hat{R}(\omega) \). The real part of \( \hat{R}(\omega) \) may then be obtained from the imaginary part by the Kramers-Kronig dispersion relation

\[
\text{Re} \hat{R}(\omega) = \frac{1}{\pi} \mathcal{P} \int \frac{\text{Im} \hat{R}(\omega')}{\omega' - \omega} \, d\omega'
\]  

(3.34)

holding for Fourier transforms of function vanishing at negative times.

D. A simple one-dimensional example

Let us consider an overdamped Langevin dynamics of a particle moving on a circle, given by the stochastic equation

\[
d\theta = M \left( -U'(\theta) + F \right) dt + \sqrt{2D} \, dW(t),
\]

(3.35)

where \( \theta \) is the angle modulo \( 2\pi \) parametrizing the position of the particle, \( M \) is the mobility and \( D = \beta^{-1} M \) is the diffusivity. Periodic function \( U(\theta) \) gives the potential and \( F \) is a constant part of the force (any nonconservative force may be separated into a constant plus a potential part in that situation). Eq. (3.35) models the dynamics of a colloidal particle of the radius 1 \( \mu m \) manipulated on a circle of radius 4.12 \( \mu m \) by a laser tweezer in an experiment performed at ENS Lyon [GP09]), in which case,

\[
MU(\theta) = H_0 \sin(\theta) \quad \text{for} \quad H_0 = 0.87 \text{ rad s}^{-1},
\]

(3.36)

\[
MF = 0.85 \text{ rad s}^{-1}, \quad D = 1.26 \times 10^{-2} \text{ rad}^2 \text{s}^{-1}.
\]

(3.37)

The diffusion (3.35) has the backward generator

\[
L = M \left( -U' + F \right) \partial_\theta + D \partial_\theta^2
\]

(3.38)

and a NESS with the invariant measure \( \nu(\theta) = \rho(\theta) d\theta \) for

\[
\rho(\theta) = Z^{-1} e^{-\beta(U(\theta) - F\theta)} \int_{\theta}^{\theta+2\pi} e^{\beta(U(\theta) - F\theta')} \, d\theta' \equiv e^{-\varphi(\theta)}
\]

(3.39)

that corresponds to the constant probability current

\[
j = \left( M(-U'(\theta) + F) - D \partial_\theta \right) e^{-\varphi(\theta)} = D Z^{-1} \left( 1 - e^{-2\pi \beta F} \right)
\]

(3.40)

and to the current velocity

\[
v(\theta) = j \rho(\theta)^{-1} = \frac{D (1 - e^{-2\pi \beta F})}{e^{-\beta(U(\theta) - F\theta)} \int_{\theta}^{\theta+2\pi} e^{\beta(U(\theta) - F\theta')} \, d\theta'}.
\]

(3.41)

The generator adjoint to \( L \) with respect to the invariant measure \( \nu \) is

\[
L' = e^{\varphi} L e^{-\varphi} = (M(-U' + F) - 2v) \partial_\theta + D \partial_\theta^2.
\]

(3.42)

It is the backward generator of the time-reversed process \( \{\theta_t^\ast\} \) defined with the rule of case (a) of Sec.1G (with \( \theta^\ast = \theta \)) which satisfies the overdamped Langevin equation

\[
d\theta' = (M(-U'(\theta') + F) - 2v(\theta')) \, dt + \sqrt{2D} \, dW(t).
\]

(3.43)

The three different forms of the FDT for this system (3.25), (3.20) and (3.12) were checked by comparing with the experimental measurements of the correlation and response functions, with similar
results confirming the theoretical predictions for times up to few seconds [GP09, GP11], see FIG. 7 for the case of Eq. (3.25). For the system in question, the anomalous term on the fluctuation side \((B(t)\) on FIG. 7) dominates the equilibrium term \((C(0) - C(t)\) on FIG. 7) so that the equilibrium form (3.26) of the FDT is grossly violated and has to be replaced by one of the nonequilibrium versions.

\[
\begin{align*}
\text{FIG. 7: Experimental verification of the time-integrated form of the FDT (3.25), different terms (left), error brackets (right), from [GP09]}
\end{align*}
\]

### E. Green-Kubo formula for diffusions

Let us consider the diffusion process given by the stochastic equation (1.41) with \(X_{0t} = X_0 + \epsilon_0 Y_0\) and \(X_0, Y_0\), and \(X_t\) time independent. Let \(\nu(dx) = e^{-\nu(x)}\lambda(dx)\) be the invariant measure when \(\epsilon = 0\). As was discussed in Lecture 1 in Sec. I G, a fluctuation relation

\[
E_{\rho_0}^x e^{-W_\epsilon}[x] = 1
\]

(3.44)

holds in this case for \(W_\epsilon\) given by Eqs. (1.80) and (1.87) with \(Y_t(x) = \epsilon_t Y_0\). In particular, the choice \(\rho_0(x) = e^{-\nu(x)} = \rho_0^*(x^*)\) results in the expression

\[
W_\epsilon[x] = \int_0^\tau \epsilon_t^a \left( (\partial_i \phi) Y_t^i - \partial_i Y_t^i(x_t) \right) dt \equiv \int_0^\tau \epsilon_t^a \mathcal{J}_a(x_t) dt.
\]

(3.45)

Expanding Eq. (3.44) to the second order in \(\epsilon\) around \(\epsilon = 0\), we obtain the relation

\[
\begin{align*}
- \int_0^\tau \epsilon_t^a E_\nu \mathcal{J}_a(x_t) dt & - \int_0^\tau \epsilon_t^a dt \int_0^t \epsilon_s^b \frac{\delta}{\delta \epsilon_s^a} \bigg|_{\epsilon = 0} E_{\rho_0}^x \mathcal{J}_a(x_s) ds \\
& \quad + \frac{1}{2} \int_0^\tau \epsilon_t^a \int_0^\tau \epsilon_s^a E_\nu \mathcal{J}_b(x_s) \mathcal{J}_a(x_t) ds = 0
\end{align*}
\]

(3.46)

or, stripping it from the arbitrary functions \(\epsilon\) and taking \(s \leq t\),

\[
E_\nu \mathcal{J}_a(x_t) = 0,
\]

(3.47)

\[
E_\nu \mathcal{J}_b(x_s) \mathcal{J}_a(x_t) = \frac{\delta}{\delta \epsilon_s^b} \bigg|_{\epsilon = 0} E_{\rho_0}^x \mathcal{J}_a(x_t).
\]

(3.48)

The first equation states that the stationary expectation of \(\mathcal{J}_a\) vanishes. The integration of the second equation over \(s\) from zero to \(t\) gives

\[
\int_0^t E_\nu \mathcal{J}_b(x_s) \mathcal{J}_a(x_t) ds = \partial_s \bigg|_{\epsilon = 0} E_{\rho_0}^x \mathcal{J}_a(x_t),
\]

(3.49)

where on the right hand side \(\epsilon = (\epsilon^a)\) is taken time independent. Assuming that for \(t \to \infty\) the expectation

\[
E_{\rho_0}^x \mathcal{J}_a(x_t) \xrightarrow{t \to \infty} \int \mathcal{J}_a(x) \nu_\epsilon(dx),
\]

(3.50)
and we may infer from the Eq. (3.51) the Onsager reciprocity relation with the same sign for all dynamics. For concreteness, let us consider the anharmonic chain but similar strategy may be applied to perturbations involving also the

\[
\int_{-\infty}^{\infty} \mathbb{E}_\nu J^a(x_s) J^b(x_t) = \partial_{\epsilon}|_{\epsilon=0} \int J^a(x) \nu_\epsilon(dx), \tag{3.51}
\]

This is the Green-Kubo formula [G54, K57] that permits to extract the linear-regime change of the stationary expectation value of observables \( J^a(x) \) under a perturbation of the dynamics, from their unperturbed dynamical correlation.

If the unperturbed process was time-reversible, i.e. if \( L = e^{\beta}L^*e^{-\beta} \), then

\[
\mathbb{E}_\nu J^b(x_s) J^a(x_t) = \mathbb{E}_\nu J^a(x_{s-1}) J^b(x_{s-1}) = \mathbb{E}_\nu J^a(x_s) J^b(x_t), \tag{3.52}
\]

and we may infer from the Eq. (3.51) the Onsager reciprocity relations

\[
\partial_{\epsilon}|_{\epsilon=0} \int J^a(x) \nu_\epsilon(dx) = \partial_{\epsilon}|_{\epsilon=0} \int J^b(x) \nu_\epsilon(dx). \tag{3.53}
\]

The Green-Kubo formula itself may be rewritten in this case in the symmetrized form

\[
\frac{1}{2} \int_{-\infty}^{\infty} J^b(x_s) J^a(x_t) ds = \partial_{\epsilon}|_{\epsilon=0} \int J^a(x) \nu_\epsilon(dx). \tag{3.54}
\]

The last three relations also hold if the unperturbed process is time-reversible only relative to an involution \( x \to x^* \) but the observables \( J^a(x) \) have are all either even or odd under it: \( J^a(x^*) = \pm J^a(x) \) with the same sign for all \( a \).

**Example 7.** Above, we considered for simplicity only perturbations of the drift term in the diffusions, but similar strategy may be applied to perturbations involving also the pure diffusion part of the dynamics. For concreteness, let us consider the anharmonic chain (1.90) of Example 5 in Sec.1G with \( M_i^{-1} = 0 \) for \( i \neq 0, L \) and with \( \beta_0 = \beta - \frac{1}{2} \epsilon \) and \( \beta_L = \beta + \frac{1}{2} \epsilon \) so that the functional (1.102) corresponding to the piecewise constant interpolation of \( \beta_i \) is equal to

\[
W_\tau(x) = \epsilon \int_0^\tau j_{(i-1,i)}(x_t) dt. \tag{3.55}
\]

Expanding the Jarzynski equality (1.66) to the second order in \( \epsilon \) and proceeding as before, one arrives at the identity

\[
\int j_{(i-1,i)}(x) \nu(dx) = 0, \tag{3.56}
\]

where \( \nu(dx) \) is the Gibbs measure at inverse temperature \( \beta \) for the chain, and at the Green-Kubo relation

\[
\frac{1}{2} \int_{-\infty}^{\infty} \mathbb{E}_\nu j_{(i-1,i)}(x_s) j_{(i-1,i)}(x_t) ds = \partial_{\epsilon}|_{\epsilon=0} \int j_{(i-1,i)}(x) \nu_\epsilon(dx), \tag{3.57}
\]

where \( \nu_\epsilon(dx) \) is the invariant nonequilibrium measure for the perturbed boundary temperatures (the equilibrium underdamped \( \epsilon = 0 \) dynamics is time-reversible under the involution that reverses the sign of momenta and the heat flux \( j_{(i-1,i)} \) is odd under it). One of the outstanding open problems of mathematical physics is the control of the large \( L \) behavior of the thermal conductivity

\[
\kappa(L, \beta) = L \beta^2 \partial_{\epsilon}|_{\epsilon=0} \int j_{(i-1,i)}(x) \nu_\epsilon(dx) \tag{3.58}
\]

giving the proportionality constant between the heat flux and the (infinitesimally small) temperature gradient imposed at the boundary. In particular, one would like to establish the conjectured Fourier law which (in a weak form) states that the limit \( \lim_{L \to \infty} \kappa(L, \beta) \) exists and is strictly positive [BL00, BK07].
IV. LARGE DEVIATIONS AND STATIONARY FLUCTUATION RELATIONS

In the presence of a small parameter \( \epsilon \), a family \( \mu_\epsilon(dX) \) of measures may exhibit a large deviations regime in which it takes an exponential form, with the inverse of the small parameter as the prefactor in front of the negative exponent. This is often formulated as the existence of a rate function \( I(X) \) such that

\[
-\inf_{X \in A^0} I(X) \leq \liminf_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(A) \leq \limsup_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(A) \geq -\inf_{X \in A} I(X),
\]

where \( A^0 \) is the interior and \( \bar{A} \) the closure of set \( A \). In less formal terms, this may be stated as the property

\[
\mu_\epsilon(dX) \sim e^{-\frac{1}{\epsilon} I(X)} \lambda(dX)
\]

or, if \( \mu_\epsilon(dX) = \rho_\epsilon(X) \lambda(dX) \), as the existence in a sufficiently weak form of the limit

\[
\lim_{\epsilon \to 0} \epsilon \ln \rho_\epsilon(X) = -I(X).
\]

The history of the large deviations theory is long and overlaps the works of the founding fathers of statistical mechanics. On the probability theory side, it goes back to contributions of the Swedish mathematician Harald Cramér from the thirties of the last century. In application to stochastic processes, small parameters may have different origin. One possibility is a small noise in the stochastic differential equations (e.g. low temperature in the Langevin equations). This is the domain of application of Freidlin-Wentzell theory of large deviations [FW84]. We shall encounter it below on a formal level of diffusions in a functional space. Another possibility, developed first by Donsker-Varadhan [DV75], is the long-time asymptotics of the solutions of stochastic equations, see also the textbooks [DS89, DZ98]. We shall need its version that, to my knowledge, was not explicitly considered in mathematical texts but appeared in the papers of physicists [CC07, MNW08].

A. Large deviations at long times

For a stationary diffusion Markov process \( \{x_t\} \) solving the stationary version of Eq. (1.41), define the empirical density and empirical current by the formulas

\[
\rho_\tau(x) = \tau^{-1} \int_0^\tau \delta(x - x_t) dt, \quad j_\tau(x) = \tau^{-1} \int_0^\tau \delta(x - x_t) \circ dx(t),
\]

where, as before, “\( \circ \)” signifies the Stratonovich convention. Assuming the ergodicity of the process, when \( \tau \to \infty \), \( \rho_\tau \) converges (in a weak sense) to the density \( \rho \) of the invariant measure \( \nu(dx) = \rho(x) \lambda(dx) \), and \( j_\tau \) converges to the probability current \( j \) given by

\[
j(x) = (\hat{X}_0(x) - \mathcal{D}(x) \nabla) \rho(x) \equiv j_\rho(x).
\]

see Eq. (1.45), which is conserved:

\[
\nabla \cdot j(x) = 0.
\]

We would like to inquire about the asymptotics of that convergence. The answer is provided by the large deviations form of the joint distribution function of the empirical density \( \rho_\tau \) and current \( j_\tau \):

\[
\mathbb{E}_\nu \delta[\rho - \rho_\tau] \delta[j - j_\tau] \sim_{\tau \to \infty} e^{-\tau I[\rho, j]},
\]

with the rate function

\[
I[\rho, j] = \begin{cases} 
\infty & \text{if } \nabla \cdot j \neq 0, \\
\frac{1}{2} \int \left[(j - j_\rho) \cdot (g \mathcal{D})^{-1} (j - j_\rho)\right] \lambda(dx) & \text{if } \nabla \cdot j = 0,
\end{cases}
\]

where \( j_\rho \) is given by Eqs. (4.5) with \( \rho \) replaced by \( \rho \). The large-deviations asymptotics for empirical densities or empirical currents only is then obtained by the "contraction principle":

\[
\mathbb{E}_\nu \delta[\rho - \rho_\tau] \sim_{\tau \to \infty} e^{-\tau I[\rho]}, \quad \mathbb{E}_\nu \delta[j - j_\tau] \sim_{\tau \to \infty} e^{-\tau I[j]}
\]
with
\[ I[q] = \min_j I[q, j], \quad I[j] = \min_q I[q, j] \] (4.10)
in a slightly abusive notations. The first minimum may be rewritten (why?) in terms of a maximum over Lagrange multipliers \( f(x) \):
\[ I[q] = -\min_f \int (\nabla f)(x) \cdot [qD(\nabla f) + \dot{j}_\theta](x) \lambda(dx) = -\frac{1}{4} \int [\nabla \cdot \dot{j}_\theta((\nabla \cdot qD)^{-1}(\nabla \cdot \dot{j}_\theta))(x) \lambda(dx). \] (4.11)

Note that \( I[q] \) is nonnegative and attains its vanishing minimum on the density \( \rho \) of the invariant measure such that \( \nabla \cdot \dot{j}_\rho = 0 \). The first line of Eq. (4.11) may be also rewritten as
\[ I[q] = -\min_{u>0} \int (u^{-1}L\rho)(x) \lambda(dx), \] (4.12)
where the last minimum is over positive functions \( u(x) = e^{f(x)} \). In the last form, the formula for the rate function \( I[q] \) holds for general continuous-time stationary Markov processes [DV75].

As for the rate function \( I[j] \), note that \( I[q, j] \) is a local non-linear functional of the density \( q > 0 \) constrained to be normalized, quadratic in its first derivatives, leading to the 2nd-order differential equation for the extrema. Even in one dimension where the latter equation reduces to an ODE and \( \nabla \cdot j = 0 \) means that \( j = \text{const.} \), the minimization over \( q \) cannot be explicitly solved. Nevertheless, for weak noise, one may resort to semiclassical instanton-gas type expansions which go back to ideas of Kramers from the 40’s of the last century and to Freidlin-Wentzell large deviations theory, and which are still being actively developed, see e.g. [E06, CC09].

**Example 8.** For the one-dimensional diffusion defined by Eqs. (3.35), (3.36) and (3.37), the drift has two zeros, one unstable at \( \theta_u \approx 0.214 \text{ rad} \) and one stable at \( \theta_s = 2\pi - \theta_u \). One obtains in this case [CC09]
\[ I(j) \approx -j \ln \frac{\sqrt{j^2 + 4\kappa_+\kappa_-} - j}{2\kappa_-} - \sqrt{j^2 + 4\kappa_+\kappa_-} + \kappa_+ + \kappa_- \] (4.13)
for \( \kappa_+ = A_0 \exp[D^{-1} A_-] \), \( \kappa_- = A_0 \exp[-D^{-1} A_+] \) where \( A_- \) is the (tiny negative) area under the negative part of the drift graph, \( A_+ \) is the one under the positive part of the graph, see the left plot in FIG. 8, and \( A_0 = M \sqrt{U''(\theta_u)U'(\theta_u)}/(2\pi) \). The most probable value of \( j \), for small \( D \) is \( j = \kappa_+ - \kappa_- \). The same large deviation function (4.13) may be obtained from a jump process with plus or minus jumps occurring with rates \( \kappa_\pm \) by looking at the statistics of the large sums of jumps [LS99]. Those jumps correspond in the diffusion to the tunneling to the right and to the left through the barriers separating the stable and unstable points of the drift.

![FIG. 8: Drift in Eq. (3.35) (left) and the comparison of numerical and theoretical rate functions for the empirical current (right)](image)

The large deviations statistics of the empirical current \( j_\tau \), that is \( \theta \)-independent in this regime, may be extracted from the one of its spatial mean
\[ \frac{1}{2\pi} \int_0^{2\pi} j_\tau(\theta) d\theta = \frac{1}{2\pi} \int_0^{\tau} d\theta = \] (4.14)
The numerical simulation of the minus logarithm of its distribution function of the latter divided by \( \tau = 4167 \text{ s} \) is shown on the right plot in FIG. 8. Note its oscillatory character (with the period \( 1/\tau \)). Its plot averaged over these oscillations compares well on the interval with sufficient number of events with the semiclassical formula (4.13) shifted by \( -(2\tau)^{-1} \ln \left( \frac{I(\ell)}{2\pi} \right) \) to include the one-loop correction. I do not know whether the convergence (4.3) (for \( X = j \) and \( \epsilon = \tau^{-1} \)) holds here pointwise or only after smearing with a test functions.

B. Gallavotti-Cohen type fluctuation relation

Defining the time-reversed diffusion process \( (x'_t) \) the way described in Sec. I G and using relation (1.79), we obtain the identity

\[
E'_\omega \ e^{-\mathcal{W}_\tau}(\rho_{\tau}, j_{\tau}) = E'_\omega \ F(\rho^*_{\tau}, -j^*_{\tau}), \tag{4.15}
\]

where, by definition,

\[
\rho^*_{\tau}[x] = \rho^*_t[x^*], \quad j^*_{\tau}[x] = -j^*_t[x^*] \tag{4.16}
\]

(the minus sign in the transformation of the current comes from the change of the sign of the time derivative of the process under the time reversal). An easy calculation shows that in their dependence on points in space, \( \rho^*_t \) and \( j^*_t \) are related to \( \rho_{\tau} \) and \( j_{\tau} \) by the geometric transformation rule for densities and currents:

\[
\varrho^*(x^*) \frac{\partial (x^*)}{\partial (x)} = \varrho(x), \quad j^*(x^*) \frac{\partial (x^*)}{\partial (x)} = \frac{\partial x^*}{\partial x} j(x), \tag{4.17}
\]

with \( \frac{\partial x^*}{\partial x} \) denoting the Jacobi matrix and \( \frac{\partial (x^*)}{\partial (x)} \) the Jacobian of the involution \( x \mapsto x^* \). On the other hand, by Eq. (1.80) and (1.81),

\[
\mathcal{W}_\tau = -\ln \rho'(x^*) + \tau \omega[\rho_{\tau}, j_{\tau}] + \ln \rho_0(x_0), \tag{4.18}
\]

where

\[
\omega[\varrho, j] = \int \left[ \hat{X}_0^+ \cdot \mathcal{D}^{-1} (j - X_0^- \varrho) - (\nabla \cdot X_0^- \varrho \right] (x) \lambda(dx) \tag{4.19}
\]

so that

\[
\int_0^\tau J_t \ dt = \tau \omega[\rho_{\tau}, j_{\tau}]. \tag{4.20}
\]

Comparing the large \( \tau \) asymptotics on both sides of identity (4.15) we infer the identity

\[
\mathcal{I}[\varrho, j] + \omega[\varrho, j] = \mathcal{I}^{*}[\varrho^*, -j^*] \tag{4.21}
\]

where \( \varrho^* \) and \( j^* \) are defined by the relations (4.17). This is the stationary fluctuation relation for rate functions describing the large deviations of empirical density and current. A simple exercise using Eqs. (4.8) and the identity

\[
j^*_{\varrho} = j_{\varrho} - 2X_0^- \varrho = -j_{\varrho} + 2\hat{X}_0^+ \varrho - 2D \nabla \varrho \tag{4.22}
\]

permits to verify Eq. (4.21) directly. The probability distribution of quantity (4.20) has also the large deviations regime with the rate function given by the contraction

\[
\mathcal{I}(\varpi) = \min_{\omega[\varrho, j] = \varpi} \mathcal{I}[\varrho, j]. \tag{4.23}
\]

From Eq. (4.21), using also the relation

\[
\omega'[\varrho^*, -j^*] = -\omega[\varrho, j], \tag{4.24}
\]

a consequence of the second equality in (1.80), we obtain immediately the fluctuation relation

\[
\mathcal{I}(\varpi) + \varpi = \mathcal{I}(-\varpi). \tag{4.25}
\]
The latter identity holds, in particular, for the time reversal with \( \hat{X}_0^+ = 0 \) considered in the case (c) of Sec.IG. In this instance,

\[
\omega[\rho_t, j_t] = - \int (\nabla \cdot X_0^+) (x) \rho_t(x) \lambda(dx) = - \tau^{-1} \int_0^\tau (\nabla \cdot X_0^-) (x_1) dt \equiv \omega[\rho_t], \tag{4.26}
\]

which reduces to the phase-space contraction rate in the deterministic case with \( X_0 \equiv 0 \) and \( X_0^- = X_0 \). The stationary fluctuation relation (4.25) for uniformly hyperbolic deterministic systems (in the case when the time-reversed dynamics coincides with the direct one) was proven in [GC95a, GC95b] as the Fluctuation Theorem. The existence of large deviations regime for the quantity \( \tau \omega[\rho_t] \) representing the phase-space contraction followed in that case from the thermodynamical formalism for such dynamical systems so that the Fluctuation Theorem of Gallavotti-Cohen is not a direct consequence of the relation (4.25) for the stochastic diffusions. In the latter case, we could obtain (4.25) from the transient fluctuation relation (4.15) holding for the stationary dynamics on any time interval, whereas there is no such relation for the general stationary deterministic systems that typically have singular invariant measures. The transient Evans-Searles relation for such systems employs the non-invariant smooth initial measures and a non-trivial work using the thermodynamical formalism would be needed to show that they lead for long times to the Gallavotti-Cohen relation.

The fluctuation relation (4.25) with \( \mathcal{T}' = \mathcal{T} \) should also hold for the large-deviations rate function of the cumulated heat flux \( W_\tau \) given by Eqs. (1.99), (1.101) or (1.102) in the non-equilibrium stationary state of the anharmonic chain with Hamiltonian dynamics in the interior that we discussed in Example 5 in Sec.IG), see [RBT02] for a proof of this fact for a closely related model.

### C. Large deviations for replicated diffusions

The final part of these lectures is based on a joint work in progress with F. Bouchet et C. Nardini [BGN]. The first half concerning the large deviations for independent replicated systems is rather well known, but we present it the spirit of the macroscopic fluctuation theory developed for the dynamics of boundary driven lattice gases in a series of papers of the Rome group, see e.g. [B06]. The second half that develops the macroscopic fluctuation theory for a non-equilibrium system of replicated diffusions with a mean-field interaction seems original.

Let us consider \( N \) independent copies \( (x^n_t, n = 1, \ldots, N) \), of identical diffusions satisfying stochastic equation (1.41). For such replicated system, we may define the dynamical empirical density and empirical current by the formulas

\[
\rho_N(t, x) = \frac{1}{N} \sum_{n=1}^N \delta(x - x^n_t), \quad j_N(t, x) = \frac{1}{N} \sum_{n=1}^N \delta(x - x^n_t) \circ \frac{dx^n_t}{dt}. \tag{4.27}
\]

Here and below, use bold letters for the quantities that depend on time and space. We shall also employ the notation \( \rho_{N1}(x) \) and \( j_{N1}(x) \) whenever we consider only the \( x \)-dependence for fixed \( t \). Note the continuity equation

\[
\frac{ \partial \rho_N }{ \partial t } + \nabla \cdot j_N = 0. \tag{4.28}
\]

Let \( \Phi[\alpha] \) be a functional of (possibly distributional) densities of the cylindrical form:

\[
\Phi[\alpha] = f \left( \int h_1(x_1) \rho(x_1) \lambda(dx_1), \cdots, \int h_k(x_k) \rho(x_k) \lambda(dx_k) \right). \tag{4.29}
\]

In its \( t \)-dependence, random variable \( \Phi(\rho_{N1}) \) satisfies the stochastic equation

\[
d\Phi[\rho_{N1}] = \int \frac{\delta \Phi[\rho_{N1}]}{\delta \rho(x)} \left( - \frac{1}{N} \sum_n (\nabla \delta)(x - x^n_t) (X_{0\alpha}(x^n_t) dt + X_{0\alpha}(x^n_t) \circ dW^n_{\alpha}(t)) \lambda(dx) \right)
\]

\[
= \int \left( \rho_{N1} X_{0\alpha} \frac{\delta \Phi[\rho_{N1}]}{\delta \rho} \right)(x) \lambda(dx) dt + \frac{1}{N} \sum_n \int X_{0\alpha}(x^n_t) (\nabla \delta \Phi[\rho_{N1}]) dx^n_t \circ dW^n_{\alpha}(t) \tag{4.30}
\]

which implies that

\[
\frac{d}{dt} \mathbb{E}_\rho \Phi[\rho_{N1}] = \mathbb{E}_\rho \int \left( j_{N1} \nabla \frac{\delta \Phi[\rho_{N1}]}{\delta \rho} \right)(x) \lambda(dx)
\]

30
\[ + \frac{1}{N} \int \rho_{Nt}(x) D_j^j(x) \delta(x-y) \nabla_x \cdot \nabla_y \left[ \frac{\delta^2 \Phi[\rho_{Nt}]}{\delta \rho(x) \delta \rho(y)} \lambda(dx) \lambda(dy) \right] = \mathbb{E}_{\rho_0} \left( L_{Nt} \Phi \right)[\rho_{Nt}], \]  

where \( \mathbb{E}_{\rho_0} \) denotes the expectation value over the replicated processes and \( j_{\rho_t} \) is given by Eq. (1.45). Note that the 2nd-order term in the functional operator \( L_{Nt} \) is proportional to \( \nabla^2 \). Formally, this is the same equation as the one for the expectation of the diffusion process in the space of densities solving the stochastic PDE

\[ \partial_t \rho + \nabla \cdot j_{\rho, \xi} = 0 \]  

(4.32)

where

\[ j_{\rho, \xi} = j_\rho + (2N^{-1} \rho D)^{1/2} \xi \]  

(4.33)

with the space-time white noise \( \xi \),

\[ \mathbb{E} \xi^i(t, x) = 0, \quad \mathbb{E} \xi^i(t, x) \xi^j(s, y) = \delta^{ij} \delta(t-s) \delta(x-y). \]  

(4.34)

Compare Eq. (4.32) to the continuity equation (4.28). In the limit \( N \to \infty \), Eq. (4.32) reduces to the standard Fokker-Plank equation

\[ \partial_t \rho + \nabla \cdot j_\rho = 0, \]  

(4.35)

for the instantaneous probability densities of a single copy of the process, see Eqs. (1.44) and (1.45).

D. Hamilton-Jacobi equation and Sanov Theorem

The probability distributions of the empirical densities \( \rho_{Nt} \) evolve by the adjoint operator \( L_{Nt}^* \) and their hypothetical densities by the formal adjoint \( L_{Nt}^\dagger \) built with the use of the rule

\[ \left( \frac{\delta}{\delta \rho} \right)^\dagger = - \frac{\delta}{\delta \rho}. \]  

(4.36)

In particular, assuming that those densities have the large-deviation form \( e^{-N \mathcal{F}_t[\rho]} \), we obtain in the leading order the Hamilton-Jacobi equation for \( \mathcal{F}_t \):

\[ \partial_t \mathcal{F}_t[\rho] + \int \left( \left( \nabla \delta \mathcal{F}_t[\rho] \right) \cdot \rho D \left( \nabla \delta \mathcal{F}_t[\rho] \right) + j_{\rho t} \cdot \nabla \delta \mathcal{F}_t[\rho] \right)(x) \lambda(dx) = 0 \]  

(4.37)

We shall call \( \mathcal{F}_t[\rho] \) the free energy of the replicated system. Eq. (4.37) is solved by the the relative entropy functional

\[ \mathcal{F}_t[\rho] = \int \left( \rho \ln \left( \frac{\rho}{\rho_t} \right) \right)(x) \lambda(dx) \equiv S[\rho||\rho_t] \]  

(4.38)

where \( \rho_t \) solves the Fokker-Planck equation (4.35).

Theorem (Dynamical version of the Sanov Theorem).

The solution (4.38) of the Hamilton-Jacobi equation (4.37) describes the time evolution of the rate function describing large deviations regime of the distribution of the empirical density \( \rho_{Nt} \) if the initial points \( x_0^N \) of the replicated processes are distributed (independently) with the probability density \( \rho_0 \).

Corollary. In the particular case of the replicated stationary diffusion process, the distribution of the empirical densities \( \rho_{Nt} \) stays time independent and for large \( N \) it takes the large deviations form with the rate function

\[ \mathcal{F}[\rho] = S[\rho||\rho] \]  

(4.39)

where \( \rho \) is the density of the invariant measure of the process. \( \mathcal{F}[\rho] \) solves the stationary Hamilton-Jacobi equation

\[ \int \left( \left( \nabla \delta \mathcal{F}[\rho] \right) \cdot \rho D \left( \nabla \delta \mathcal{F}[\rho] \right) + j_\rho \cdot \nabla \delta \mathcal{F}[\rho] \right)(x) \lambda(dx) = 0. \]  

(4.40)
Introducing the stationary current velocity in the space of densities by the formula

\[ \mathcal{V}[\varrho] = -\nabla \cdot \left( j_\varrho + \varrho \nabla \delta \mathcal{F}[\varrho] \frac{\delta \varrho}{\delta \varrho} \right), \quad (4.41) \]

the stationary Hamilton-Jacobi equation (4.40) may be rewritten as the orthogonality condition

\[ \int \frac{\delta \mathcal{F}[\varrho]}{\delta \varrho(x)} \mathcal{V}[\varrho](x) \lambda(dx) = 0. \quad (4.42) \]

For the time reversal corresponding of case (a) in Sec. I G, one obtains from Eq. (4.22) the relation

\[ j_\varrho^* = -j_\varrho - 2\varrho D \nabla \varphi - 2D \nabla \varrho \quad (4.43) \]

which may be rewritten in the form

\[ j_\varrho^* = -j_\varrho - 2\varrho D \nabla \delta \mathcal{F}[\varrho] \frac{\delta \varrho}{\delta \varrho} \quad (4.44) \]

A comparison with Eq. (4.41) yields the relations:

\[ \mathcal{V}[\varrho] = \nabla \cdot \left( j_\varrho^* + \varrho D \nabla \delta \mathcal{F}[\varrho] \frac{\delta \varrho}{\delta \varrho} \right) \quad (4.45) \]

and

\[ \mathcal{V}[\varrho] = \frac{1}{2} \left( \nabla \cdot j_\varrho^* - \nabla \cdot j_\varrho \right) \quad (4.46) \]

**Example 9.** For the diffusion on a circle (3.35),

\[ \mathcal{F}[\varrho] = \int_0^{2\pi} \varrho(\theta) \left( \ln \varrho(\theta) + \varphi(\theta) \right) d\theta \quad (4.47) \]

with \( \varphi(\theta) \) given by Eq. (3.39). In particular, in the equilibrium case with \( F = 0 \),

\[ \mathcal{F}[\varrho] = \int_0^{2\pi} \varrho(\theta) \left( \ln \varrho(\theta) + \beta U(\theta) \right) d\theta + \text{const.} \quad (4.48) \]

so that \( \mathcal{F} \) is equal in this instance to \( \beta \times \) the free energy of the gas of noninteracting particles in the thermal equilibrium at temperature \( \beta^{-1} \). Quantity \( \varrho \) is the density of the gas and \( U \) is the external potential. For the Langevin equation (3.35) (with any \( F \)),

\[ j_\varrho(\theta) = M \left( -U'(\theta) + F \right) \varrho - D \varrho'(\theta), \quad (4.49) \]

and for the time reversed one of Eq. (3.43),

\[ j_\varrho^*(\theta) = j_\varrho^*(\theta) = \left( M(-U'(\theta) + F) - 2v(\theta) \right) \varrho - D \varrho'(\theta), \quad (4.50) \]

where the current velocity \( v(\theta) \) is given by Eq. (3.41). In this case

\[ \mathcal{V}[\varrho](\theta) = -\partial_\varrho(\varrho v)(\theta). \quad (4.51) \]

**E. Dynamical large deviations for the replicated process**

One can show, at least formally, that the joint distribution of the dynamical empirical density \( \varrho_N \) and current \( j_N \) show for large \( N \) the large-deviations regime with the rate function

\[ \mathcal{I}[\varrho, j] = \begin{cases} \infty & \text{if } \partial_t \varrho + \nabla \cdot j \neq 0, \\ + \int \left[ (j - j_\varrho \cdot (\varrho D)^{-1}(j - j_\varrho)) \right](t,x) \ dt \ dx & \text{if } \partial_t \varrho + \nabla \cdot j = 0. \end{cases} \quad (4.52) \]
Note the similarities and the differences with the long-time rate function (4.8). In the formal argument using functional integrals, we shall replace \( \rho_N \) and \( j_N \) that satisfy Eq. (4.28) by \( \rho \) and \( j_{\rho, \xi} \) connected by Eq. (4.32). Thus

\[
E \psi[\rho_N, j_N] = E \int \psi[q, j_{\rho, \xi}] \delta[\partial_t q + \nabla \cdot j_{\rho, \xi}] \det \left( \frac{\delta(\partial_t q + \nabla \cdot j_{\rho, \xi})}{\delta q} \right) Dq
\]

\[
= E \int \psi[q, j] e^{i \int \lambda (j_{\rho, \xi} - j_{\rho} - j_{\rho, \xi})} \delta[\partial_t q + \nabla \cdot j] \det \left( \frac{\delta(\partial_t q + \nabla \cdot j)}{\delta q} \right) Da Dq Dj. \tag{4.53}
\]

Averaging over the white noise \( \xi \), we obtain:

\[
E \psi[\rho_N, j_N] = \int \psi[q, j] e^{i \int \lambda (j_{\rho, \xi} - j_{\rho} - j_{\rho, \xi})} \delta[\partial_t q + \nabla \cdot j] \det \left( \frac{\delta(\partial_t q + \nabla \cdot j)}{\delta q} \right) Da Dq Dj. \tag{4.54}
\]

From the last functional integral expression, we read off the large deviations rate function (4.52) to which the determinant does not contribute (a similar functional-integration argument may be used to obtain formula (4.8)).

The rate functions for the dynamical large deviations of solely the empirical density \( \rho_N \) or solely the empirical current \( j_N \) are given by the contraction:

\[
I[q] = \min_j I[q, j] = \frac{1}{\tau} \int \left[ (\partial_t q + \nabla \cdot j)(- \nabla \cdot D \nabla)^{-1}(\partial_t q + \nabla \cdot j) \right](t, x) dt \lambda(dx), \tag{4.55}
\]

\[
I[j] = \min_q I[q, j], \tag{4.56}
\]

with no closed expression in the latter case.

We shall denote by \( I_{\lambda}[q, j] \) the rate functions given by Eq. (4.52) with the time-integral restricted to the interval \( A \). The functionals \( I_{[0, \tau]}[q, j] \) and \( I_{[0, \tau]}[q^*, j^*] \) for the replicated direct and time reversed process, the latter obtained with the use of rules of Sec. 1G, satisfy the stationary fluctuation relation

\[
I_{[0, \tau]}[q, j] + \omega_{[0, \tau]}[q, j] = I_{[0, \tau]}[q^*, -j^*] \tag{4.57}
\]

where

\[
q^*(t, x) = q(t^*, x^*) \frac{\partial(x^*)}{\partial(x)}, \quad j^*(t, x) = \frac{\partial x}{\partial x^*} j(t^*, x^*) \frac{\partial(x^*)}{\partial(x)} \tag{4.58}
\]

and

\[
\omega_{[0, \tau]}[q, j] = \int_0^\tau dt \int \left[ \nabla_0^+ \cdot D_t^{-1}(j_t - X_0^- \delta t) - (\nabla \cdot X_0^-) \delta t \right](x) \lambda(dx)
\]

\[
- \left. \int_0^\tau \delta t \ln \delta t(x) \lambda(dx) \right|_0^\tau \tag{4.59}
\]

compare to relations (4.21), (4.17) and (4.19). The identities (4.58), (4.59) follow in a straightforward way from the relations

\[
j_{\rho, \xi} = j_{\rho} - 2X_0^- \rho = -j_{\rho} + 2 \nabla \cdot q_{\rho} - 2 D \nabla q_{\rho} \tag{4.60}
\]

that generalize Eqs. (4.22). In the particular case of the stationary process and the time reversal corresponding of case (a) in Sec. 1G,

\[
\omega_{[0, \tau]}[q, j] = F[\rho_0] - F[\rho_\tau]. \tag{4.61}
\]

By contraction, we infer then from relation (4.58) the identity

\[
I_{[0, \tau]}[q] + F[\rho_0] - F[\rho_\tau] = I_{[0, \tau]}[q^*] \tag{4.62}
\]

33
Let \( \rho_0 = \rho \), where \( \rho = e^{-\varphi} \) is the invariant density of the single process, so that \( F[\rho_0] = 0 \). Take \( \tau \to \infty \). Then the minimum of the right hand side over \( \varphi^* \) with \( \rho_0 \) and \( \varphi_{\infty} \) fixed is realized by the trajectory \( \varphi^* \) solving the reversed process Fokker-Planck equation

\[
\partial_t \varphi^* + \nabla \cdot J_{\varphi^*} = 0
\]  
(4.63)

that relaxes from \( \varphi_{\infty} \) to the invariant density \( \rho^* \) and \( I_{[0, \infty]}(\varphi^*) \) vanishes for such a trajectory. This is an expression of the (generalized) Onsager-Machlup principle: the most probable trajectory that describes the creation of the spontaneous fluctuation \( \varphi_{\infty} \) from the vacuum configuration \( \rho \) is the time reverse of the trajectory that describes the the most probable relaxation of the spontaneous fluctuation \( \varphi_{\infty} \) to the vacuum \( \rho^* \) in the time-reversed dynamics. Taking the minima on the both hand sides of Eq. (4.62), we obtain the identity

\[
F[\varphi] = \min_{\varphi_{\infty} = \varphi^*} I_{[0, \infty]}[\varphi] = I_{[0, \infty]}[\varphi^*]
\]  
(4.64)

that connects the rate functions for the large deviations of the invariant distribution and for the dynamical large deviations of the empirical density \( \varphi_N \).

F. Replicated diffusions with mean-field coupling

One may perturb the \( N \) replicated diffusions (1.41) by introducing a mean-field type coupling between the replicated processes by a pair potential \( \frac{1}{N} V_t(x_n - x_m) \), obtaining a coupled system of stochastic equations

\[
dx = X_{\text{int}}(x) dt - \frac{1}{N} \mathcal{M}_t \sum_{n=1}^{N} (\nabla V)_t(x_n - x_m) dt + X_{\text{int}}(x) \circ dW_t(t),
\]  
(4.65)

where \( \mathcal{M}_t(x) \) is a mobility matrix that we shall fix imposing the Einstein relation \( \beta \mathcal{M}_t = \mathcal{D}_t \) with \( \mathcal{D}_t \) defined by Eq. (1.46). For simplicity, we assumed that \( V_t(x) = V_t(-x) \) so that the \( m = n \) term does not contribute and may be included for free. One may still introduce the empirical dynamical densities \( \rho_N \) and currents \( j_N \) by Eqs. (4.27). The discussion concerning the large deviations of \( \rho_N \) and \( j_N \) for the replicated diffusions carries over to the interacting case after a modification of the formula Eq. (1.45) for the current \( j_p \), which becomes

\[
j_p(x) = (\rho_t X_{\text{int}} - \rho_t \mathcal{M}_t \nabla V_t(x) + \mathcal{D}_t \nabla \rho_t)(x),
\]  
(4.66)

picking up an additional term involving the effective mean field potential

\[
V_t \ast \rho_t(x) \equiv \int V_t(x - y) \rho_t(y) \lambda(dy).
\]  
(4.67)

After this modification, one still obtains the formal stochastic equation (4.32) reducing in the limit \( N \to \infty \) to Eq. (4.35). The latter becomes now a nonlinear Fokker-Planck equation due to the presence of a quadratic term in \( \rho_t \) in the expression for \( j_p \). The Hamilton-Jacobi equations (4.37) and (4.40) have still the same form but the Sanov solutions (4.38) and (4.39) are no more valid. Finding, in particular, the right solution of the free energy \( F[\varphi] \) in the stationary case is a mayor problem, see below, except of the special instance of equilibrium dynamics. The dynamical large deviations rate functions \( I[\varphi, j] \), \( I[\varphi] \) and \( I[j] \) are still given by Eqs. (4.52) and (4.55), (4.56). In the stationary case, if one defines the time reversed process as corresponding to the formal stochastic equation (4.32) with \( j_\rho \) in relation (4.33) replaced by \( j_\rho^* \) given by Eq. (4.44) and \( \mathcal{D}' \) defined as before then the stationary fluctuation relation (4.57) still holds for \( \omega[\varphi, j] \) given by Eq. (4.61) implying the Onsager-Machlup-type relations (4.62) and (4.64).

G. Perturbative solution for free energy \( F[\varphi] \)

In the stationary case, we may search for the solution the Hamilton-Jacobi equation (4.40) for the nonequilibrium free energy functional \( F[\varphi] \) in the form of a formal power series

\[
F[\varphi] = \sum_{n=0}^{\infty} F_n(\varphi),
\]  
(4.68)
in the interaction potential $V$ treated as a perturbation, where the term $F_n$ is of order $n$ in $V$ and where

$$F_0[\rho] = \int \left( \rho \ln \left( \frac{\rho}{\rho_0} \right) \right)(x) \lambda(dx)$$  \hspace{1cm} (4.69)$$
is the Sanov solution (4.39) for $V = 0$ with $\rho_0(x)$ standing for the density of the invariant measure of the diffusion (1.41). Inserting expansion (4.68) into the Hamilton-Jacobi equation (4.40) and gathering terms of the same order in $V$, we obtain the relations

$$\int \rho(x) \left[ \left( L + 2\left( \nabla F_0[\rho] \right) \cdot D\nabla \right) F_n[\rho] - (\beta(V + \rho) \cdot D\nabla \frac{F_{n-1}[\rho]}{\delta \rho}) + \sum_{m=1}^{n-1} \left( \nabla F_m[\rho] \right) \cdot D\left( \nabla \frac{F_{m-1}[\rho]}{\delta \rho} \right) \right] \lambda(dx) = 0,$$

where $L$ is the backward generator of the single unperturbed process. After a little algebra using Eq.(4.69) and left as an exercise, one may rewrite the above identities in the form

$$\int \rho(x) L' \frac{\delta F_n[\rho]}{\delta \rho(x)} \lambda(dx) = \int \rho(x) \left[ - (\beta V + \rho) \cdot D\nabla \frac{F_{n-1}[\rho]}{\delta \rho} + \sum_{m=1}^{n-1} \left( \nabla F_m[\rho] \right) \cdot D\left( \nabla \frac{F_{m-1}[\rho]}{\delta \rho} \right) \right] \lambda(dx),$$

where

$$L' = \rho_0^{-1} L^{\dagger} \rho_0 = -\hat{X}_0 \cdot \nabla + (\nabla + 2(\nabla \ln \rho_0)) \cdot D\nabla$$  \hspace{1cm} (4.72)$$
is the backward generator of the single process time-reversed according to the rules of case (a) in Sec.I G (with $x^* = x$). We shall search for the solution of this equations assuming that, for $n \geq 1$, $F_n[\rho]$ is a polynomial of degree $n + 1$ in $\rho$:

$$F_n[\rho] = \frac{1}{(n+1)!} \int \phi_n(y_0, \ldots, y_n) \rho(y_0) \cdots \rho(y_n) dy_0 \cdots dy_n$$  \hspace{1cm} (4.73)$$
with $\phi_n$ symmetric in its arguments. For the first order term, one obtains

$$\phi_1(x_0, x_1) = \beta V(x_1 - x_2) + L_{(01)}^{-1} \left( (v_0(0) - (v_0)(1)) \beta V(01) \right)(x_0, x_1)$$  \hspace{1cm} (4.74)$$
in the notation where

$$V_{(01)}(x_0, x_1) = V(x_0 - x_1), \hspace{1cm} (v_0)(0)(x_0, x_1) = v_0(x_0), \hspace{1cm} (v_0)(1)(x_0, x_1) = v_0(x_1),$$

$$L_{(01)} = L_{(0)} + L_{(1)}', \hspace{1cm} (4.75)$$
v_0(x) is the current velocity (1.49) of the stationary diffusion (1.41), and operator $L_{(m)}'$ acts on the $m^{th}$ variable. For the kernels $\phi_n$ with $n \geq 2$, Eqs.(4.71) imply the recursive relations

$$L'_{(0 \ldots n)} \phi_n(x_0, \ldots, x_n) = - \sum_{i,j=0 \atop i \neq j}^n \nabla_{x_i} \phi_{n-1}(x_i, (x_k)_{k \neq i, j}) \cdot D\beta(\nabla V)(x_i - x_j)$$

$$+ \sum_{p=1}^{n-1} \sum_{i=0}^{n-1} \nabla_{x_i} \phi_{n-p}(x_i, (x_j)_{j \in I_p}) \cdot D\nabla_{x_i} \phi_{n-p}(x_i, (x_j)_{j \in I_p}). \hspace{1cm} (4.76)$$

The solvability of the latter equations requires the orthogonality of the right hand side to the unique zero mode $\rho_0(x_0) \cdots \rho_0(x_n)$ of the operator adjoint to $L_{(0 \ldots n)}'$. This condition is equivalent to the vanishing of the right hand side of Eq.(4.71) for $\rho = \rho_0$. We have checked that this holds for the three first orders, but lack yet a general recursive proof. It is possible that one may obtain a closed and not only the recursive expression for the kernels $\phi_n$ or the corresponding functional $F_n[\rho]$ and that one may prove the convergence of the perturbative series in some neighborhood of $\rho_0$. Global convergence may be obstructed by phase transitions, see the next section. A perturbative solution for $F[\rho]$ necessarily has a perturbative stationary solution of the nonlinear Fokker-Planck equation

$$\rho(x) = \sum_{n=0}^{\infty} \rho_n(x)$$  \hspace{1cm} (4.77)$$
verifying \( \nabla \cdot j_{\rho_{N}} = 0 \) as an extremum since on such a configuration the Hamilton-Jacobi equation reduces to the identity

\[
\int \left[ \left( N \frac{\delta F[\rho]}{\delta \theta} \right) \cdot \rho D \left( \nabla \frac{\delta F[\rho]}{\delta \theta} \right) \right] (x) \lambda (dx) = 0 \quad (4.78)
\]

implying that

\[
\nabla \frac{\delta F[\rho]}{\delta \theta (x)} = 0. \quad (4.79)
\]

In any case, replicated diffusions coupled in a mean-field way seem to be among a few non-equilibrium systems, along with some special models of one-dimensional lattice systems, along with some special models of one-dimensional lattice glasses, see [B06], where the nonequilibrium free energy \( F[\rho] \) may be controlled analytically at least to some extent.

H. Diffusions on the circle with mean field coupling

Let us look more closely at the case when the original diffusion process is given by the stationary overdamped Langevin equation (3.35) on a circle and when we take the coupling potential \( V \) time-independent and \( M_t = M \), arriving at a system of stochastic equations

\[
d\theta^n = M \left( -U' (\theta^n) - \frac{1}{N} \sum_{m=1}^{N} V' (\theta^n - \theta^m) + F \right) dt + \sqrt{2D} dW^n (t). \quad (4.80)
\]

For \( F = 0 \), equations (4.80) describe an equilibrium dynamics with the invariant measure given by the Gibbs state

\[
\nu (d(\theta^n)) = Z_N^{-1} e^{-\beta \left( \sum_{n=1}^{N} U (\theta^n) + \frac{1}{N} \sum_{n,m=1}^{N} V (\theta^n - \theta^m) \right)} \prod_{n=1}^{N} d \theta^n. \quad (4.81)
\]

In this case, the large deviations rate function for the stationary distribution of empirical current \( \rho_{Nt} (\theta) \) is

\[
F[\rho] = \int_{0}^{2\pi} \rho (\theta) \left( \ln \rho (\theta) + \beta (U (\theta) + \frac{1}{2} V (\theta)) \right) d\theta + \text{const.} \quad (4.82)
\]

generalizing Eq. (4.48). It is a solution of the stationary Hamilton-Jacobi equation (4.40) with

\[
\dot{\rho}_{\theta} = \rho (U' + F - (V * \rho)' - D \rho') \quad (4.83)
\]

for \( F = 0 \) and it is of the form (4.73) with

\[
\phi_1 (\theta_0, \theta_1) = \beta V (\theta_0 - \theta_1) \quad (4.84)
\]

(which agrees with expression (4.74) because \( \nu_0 = 0 \) is this case) and with \( \phi_n = 0 \) for \( n \geq 2 \). It is easy that this, indeed, provides a solution of the recursion (4.76). Note that expression (4.82) is equal to \( \beta \times \) the free energy of the gas of interacting particles in the thermal equilibrium at temperature \( \beta^{-1} \). For \( F > 0 \), stochastic equation (4.80) describes a nonequilibrium \( N \)-particle dynamics, generalizing the single particle one (3.35).

**Example 10.** Consider the case when \( U (\theta) = -h \cos (\theta) \) and \( V (\theta) = J (1 - \cos (\theta)) \) with \( J > 0 \). In this instance, one may also view the angles \( \theta^n \) as describing planar spin vectors \( \vec{\sigma}_n = (\cos \theta^n, \sin \theta^n) \) in magnetic field \( b \) along the first axis, coupled with the ferromagnetic mean field coupling. For \( F = 0 \), the Gibbs state (4.81) exhibits in the limit \( N \to \infty \) a 2\textsuperscript{nd} order phase transition for \( h = 0 \) from the disordered stationary phase with \( \mathbb{E} \vec{\sigma}_n = 0 \) for \( \beta J \leq 2 \) to the mixture of ordered ones with \( \mathbb{E} \vec{\sigma}_n \neq 0 \) for \( \beta J > 2 \) [SFN72]. A pure ordered state is selected by turning on an infinitesimal magnetic field \( h = +0 \) and the other pure states are obtained by a simultaneous rotation of all spins. The sharp transition disappears for (non-infinitesimal) \( h \neq 0 \). The \( F \)-term describes the effect of a constant electric field perpendicular to the plane of spins, linearly growing with time. The model with \( F \neq 0 \) was studied in [SK86] and it is also closely related to the Kuramoto model of synchronization [K75] (that has drifts \( F \) dependent on \( n \), a subject of rich mathematical literature, see e.g. [GLP11]. At
$N = \infty$, the dynamics is described by the non-linear Fokker-Planck equation (4.35) whose stationary solutions have the form

$$\rho(\theta) = Z^{-1} e^{-\beta \left( (h+x_1) \cos \theta + x_2 \sin \theta + F \theta \right)} \int_0^{\theta + 2\pi} e^{\beta \left( (h+x_1) \cos \theta + x_2 \sin \theta + F \theta \right)} d\theta,$$

(4.85)

compare to Eq. (3.39), where the coefficients

$$\frac{x_1}{J} = \int_0^{2\pi} \cos \vartheta \rho(\vartheta) d\vartheta,$$

$$\frac{x_2}{J} = \int_0^{2\pi} \sin \vartheta \rho(\vartheta) d\vartheta.$$

(4.86)

have to be found self-consistently. The linearly stable solution solutions are then selected by the analysis of the spectrum of the linearization of the non-linear Fokker-Planck operator. For $F \neq 0$, integrals in Eqs. (4.86) are expressible by Bessel functions, leading to the equations

$$\frac{x_1}{J} = \frac{(x_1 + h) I_1(\beta x(h))}{x(h) I_0(\beta x(h))}, \quad \frac{x_2}{J} = \frac{x_2 I_1(\beta x(h))}{x(h) I_0(\beta x(h))},$$

(4.87)

where $x(h) = \sqrt{(x_1 + h)^2 + x_2^2}$. For $h = +0$, these equations are solved by $x_1 = 0 = x_2$ but the corresponding stationary solution of the nonlinear Fokker-Planck equation becomes unstable for $\beta J > 2$ where another stable solution with $x_1 > 0$ and $x_2 = 0$ (accompanied by the one reflected in the $x_2 = 0$ axes corresponding to $h = -0$) appears by a supercritical pitchfork bifurcation, see FIG. 9.

![FIG. 9: Ratio of the Bessel functions and the curves $\frac{I_1}{I_0}$ for $\beta J = \frac{3}{5} < 2$ and $\beta J = \frac{5}{2} > 2$](image)

For $h \neq 0$ there is a single solution of Eqs. (4.87) with $\text{sgn}(h) x_1 > 0$ and $x_2 = 0$. For $F > 0$ and $h = +0$, there is no stable stationary solution for $\beta J > 2$ but a stable periodic solution of the nonlinear Fokker-Planck equation

$$\rho(t, \theta) = \rho_{F=0}(\theta - MFt),$$

(4.88)

where $\rho_{F=0}$ is the stable solution for $F = 0$, appears by a Hopf bifurcation: the phase transition becomes dynamical. For $h \neq 0$ but small in absolute value, two dynamical phase transitions occur, with high and low temperature states stationary and the intermediate-temperature states periodic, and, finally, for $|\beta h|$ sufficiently large (in a $\beta F$- and $\beta J$-dependent way), there is a single stationary state at $N = \infty$ and the dynamical transitions disappear [SK86].


