Nonequilibrium statistical mechanics - II

Sourabh Lahiri

March 4, 2012

Contents

1 Langevin equation and thermodynamics

The Langevin equation of motion is given by

$$m\ddot{x} = -\frac{\partial U}{\partial x} - \gamma\dot{x} + \xi(t), \tag{1.1}$$

where $U(x,t)$ is the effective potential and $\xi(t)$ is Gaussian white noise:

$$\langle \xi(t) \rangle = 0, \qquad \langle \xi(t)\xi(t') \rangle = 2k_B T\gamma\delta(t-t'). \tag{1.2}$$

If the relaxation time m/γ is much less than any other time scale in the problem, then one neglects the inertial term to get the overdamped Langevin equation:

$$\gamma\frac{dx}{dt} = -\frac{\partial U}{\partial x} + \xi(t) \tag{1.3}$$

Multiplying above equation by dx and rearranging terms, one gets [1]

$$-\left[-\gamma\frac{dx}{dt} + \xi(t)\right]dx + \frac{\partial U}{\partial x}dx = 0. \tag{1.4}$$

Here, $-\left[-\gamma\frac{dx}{dt} + \xi(t)\right]$ is the *reaction* force to the heat bath exerted by the system. The work done by this reaction force will be the dissipated heat :

$$\boxed{dQ = -\left[-\gamma\frac{dx}{dt} + \xi(t)\right]dx.} \tag{1.5}$$

Thus, from (1.4) we have

$$dQ + dU - \frac{\partial U}{\partial t}dt = 0. \tag{1.6}$$

Since dQ is the heat dissipated and dU is the change in internal energy, one identifies the last term as work:

$$\boxed{dW = \frac{\partial U}{\partial t}dt} \tag{1.7}$$

Underdamped case :

The underdamped Langevin eqn is given by

$$m\frac{dv}{dt} = -\gamma v - \frac{\partial V(x,t)}{\partial x} + \xi(t). \tag{1.8}$$

$$\begin{aligned}
\therefore -(-\gamma v + \xi(t))dx &= -m\frac{dv}{dt}dx - \frac{\partial V}{\partial x}dx \\
\Rightarrow dQ &= -mvdv - dV + \frac{\partial V}{\partial t}dt \\
&= -d\left(\frac{1}{2}mv^2 + V\right) + \frac{\partial V}{\partial t}dt \\
\therefore Q &= -\Delta U + W.
\end{aligned} \tag{1.9}$$

2 The Jarzynski and Crooks relations

2.1 For Hamiltonian evolution

Let us consider a *thermally isolated* system at $t = 0$ in equilibrium [3]. Let the Hamiltonian depend on dynamical variables $x(t)$ and on a control parameter that varies according as the function $\lambda(t)$. The uncontrolled degrees of freedom, $x(t)$, evolve along a *deterministic* trajectory governed by Hamiltonian dynamics, from time $t = 0$ to $t = \tau$. Since the system is thermally isolated, the work done on it goes fully into changing its energy:

$$H(x_\tau) - H(x_0) = W[x(t)], \tag{2.1}$$

We have

$$\rho(x_0) = \frac{e^{-\beta H(x_0,\lambda_0)}}{Z(\lambda_0)} \tag{2.2}$$

Similarly for the reverse process,

$$\tilde{\rho}(\tilde{x}_0) = \frac{e^{-\beta H(x_\tau)}}{Z(\lambda_\tau)}. \tag{2.3}$$

Here, $\tilde{x}_0 = x_\tau^*$, with * operation implies inversion of velocities. The system now begins from equilibrium at the protocol value λ_τ, and then is allowed to evolve subject to the drive $\lambda(\tau - t)$.

$$\frac{\rho(x_0)}{\tilde{\rho}(\tilde{x}_0)} = \frac{Z(\lambda_\tau)}{Z(\lambda_0)}\ e^{\beta(H(x_\tau)-H(x_0))} = e^{\beta(W-\Delta F)}, \tag{2.4}$$

where we have used the statistical definition of free energy: $F = -k_B T \ln Z$.

Now, in Hamiltonian evolution, each initial state gives rise to a *unique* trajectory in phase space (one-to-one mapping), which in turn means that the probability for a given initial state x_0 must equal the probability of the entire trajectory generated from x_0, namely, the sequence of states $\{x_0, x_1, \cdots, x_\tau\}$:

$$\rho(x_0)\ dx_0 = P(x_0, x_1, \cdots, x_\tau)\ dx_0 dx_1 \cdots dx_\tau \equiv P[x(t)]\ \mathcal{D}[x_t]. \tag{2.5}$$

So the LHS of (2.4) can be equivalently written as ratios of the probabilities for the forward and reverse *trajectories*:

$$\frac{\rho(x_0)dx_0}{\tilde{\rho}(\tilde{x}_0)dx_0} = \frac{P[x(t)]\mathcal{D}[x_t]}{\tilde{P}[\tilde{x}(t)]\mathcal{D}[x_t]} = e^{\beta(W-\Delta F)}. \tag{2.6}$$

Now, suppose we find other trajectories, $x'(t)$, $x''(t), \cdots$ along which the *same* amount of work W has been done. Then (2.6) tells us

$$\frac{P[x(t)]\mathcal{D}[x_t]}{\tilde{P}[\tilde{x}(t)]\mathcal{D}[x_t]} = \frac{P[x'(t)]\mathcal{D}[x'_t]}{\tilde{P}[\tilde{x}'(t)]\mathcal{D}[x'_t]} = \frac{P[x''(t)]\mathcal{D}[x''_t]}{\tilde{P}[\tilde{x}''(t)]\mathcal{D}[x''_t]} = \cdots = e^{\beta(W-\Delta F)}. \tag{2.7}$$

Using the standard properties of ratios, we have

$$\frac{P[x(t)]\mathcal{D}[x_t] + P[x'(t)]\mathcal{D}[x'_t] + P[x''(t)]\mathcal{D}[x''_t] + \cdots}{\tilde{P}[\tilde{x}(t)]\mathcal{D}[x_t] + \tilde{P}[\tilde{x}'(t)]\mathcal{D}[x'_t] + \tilde{P}[\tilde{x}''(t)]\mathcal{D}[x''_t] + \cdots} = e^{\beta(W-\Delta F)}. \tag{2.8}$$

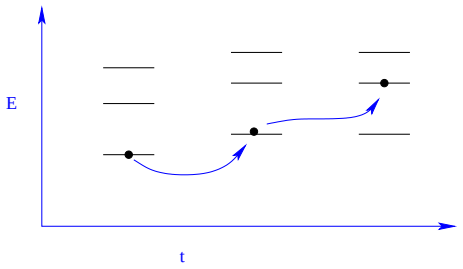
The numerator now is the sum of probabilities of trajectories along which work W has been performed on the system, while the denominator consists that of the corresponding reverse trajectories. Thus the numerator is the total probability of work W being done on the system, and the denominator is that of the work W done *by* the system along the reverse trajectory:

$$\frac{P_f(W)dW}{P_r(-W)dW} = e^{\beta(W-\Delta F)} \quad \Rightarrow \quad \boxed{\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta F)}}. \tag{2.9}$$

Now cross multiplication followed by integration of W on both sides gives the Jarzynski identity:

$$\boxed{\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta F}}. \tag{2.10}$$

2.2 For discrete time Markov process



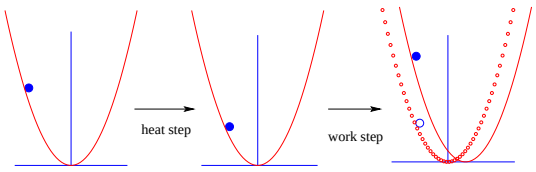
Consider the adjacent diagram. The Brownian particle is initially in state i_0 at the value of the external parameter λ_0. It performs the following steps in discrete time [4]:

$$(i_0, \lambda_0) \xrightarrow{\lambda_0} (i_1, \lambda_1) \xrightarrow{\lambda_1} (i_2, \lambda_2) \xrightarrow{\lambda_2} \cdots \xrightarrow{\lambda_{n-1}} (i_n, \lambda_n), \tag{2.11}$$

where the total time of observation has been broken up into n discrete time steps. Each of these steps can once again be broken up as follows:

$$(i_j, \lambda_j) \to (i_j, \lambda_{j+1}) \xrightarrow{\lambda_{j+1}} (i_{j+1}, \lambda_{j+1}), \tag{2.12}$$

where there is no loss of generality since the observable end result remains the same. The first step occurs when λ is changed instantaneously and the particle does not undergo change in its state. This step is caused solely due to change in λ, hence deriving its name *work step* . In the second step, the state of the particle changes at fixed λ. This can take place even in absence of external perturbation, so that it can be easily associated with the heat dissipated by the system, q_j. As a consequence, this step is termed as the *heat step* . These two together, as expected, gives rise to change in internal energy, $E(i_{j+1}, \lambda_{j+1}) - E(i_j, \lambda_j)$, during that step.



This is also easy to appreciate for a classical system. In the adjacent diagram, we have shown the division of an overdamped process of shifting a harmonic potential well (this is our protocol) into the heat and the work steps. During the heat step, the parameter value (position of potential minimum) remains the same but the particle moves to a different position. On the other hand, during the work step, the particle position is fixed but the minimum of the potential

has been shifted. The corresponding changes in energy of the particle are clearly visible. These are called "heat dissipated to bath" during heat step and "work done on the system" during the work step.

Thus the total work done by, heat *absorbed* by and change in energy of the system along the trajectory $\{i_0, i_1, \cdots, i_n\}$ are given by

$$\begin{aligned} W =& \sum_{j=0}^{n} [E(i_j, \lambda_{j+1}) - E(i_j, \lambda_j)] \\ Q =& \sum_{j=0}^{n} [E(i_{j+1}, \lambda_{j+1}) - E(i_j, \lambda_{j+1})] \\ \Delta U =& \sum_{j=0}^{n} [E(i_{j+1}, \lambda_{j+1}) - E(i_j, \lambda_j)]. \end{aligned} \tag{2.13a}$$

Now we make another assumption: in the second (heat) step, the path is microscopically reversible:

$$P(i_j, \lambda_{j+1})\ P[i_{j+1}, \lambda_{j+1} | i_j, \lambda_{j+1}] \ = \ P(i_{j+1}, \lambda_{j+1})\ P[i_j, \lambda_{j+1} | i_{j+1}, \lambda_{j+1}]. \tag{2.14}$$

This leads to the relation

$$\begin{aligned} \frac{P[i_{j+1}, \lambda_{j+1} | i_j, \lambda_{j+1}]}{P[i_j, \lambda_{j+1} | i_{j+1}, \lambda_{j+1}]} =& \frac{P(i_{j+1}, \lambda_{j+1})}{P(i_j, \lambda_{j+1})} \\ =& \exp\left[-\beta\{E(i_{j+1}, \lambda_{j+1}) - E(i_j, \lambda_{j+1})\}\right] \\ =& e^{\beta q_j}, \end{aligned} \tag{2.15}$$

where q_j is the heat *dissipated* by the system in the j^{Th} step. In a nutshell, we have

$$\left.\frac{P(i_j \to i_{j+1})}{P(i_{j+1} \to i_j)}\right|_{\lambda = \lambda_{j+1}} = e^{\beta q_j} \tag{2.16}$$

Now for a Markov process we have

$$\begin{aligned} \frac{P(i_0 \to i_n)}{P(i_0 \leftarrow i_n)} =& \frac{P(i_0 \xrightarrow{\lambda_0} i_1) P(i_1 \xrightarrow{\lambda_1} i_2) \cdots P(i_{n-1} \xrightarrow{\lambda_{n-1}} i_n)}{P(i_0 \xleftarrow{\lambda_0} i_1) P(i_1 \xleftarrow{\lambda_1} i_2) \cdots P(i_{n-1} \xleftarrow{\lambda_{n-1}} i_n)} \\ =& \exp\left[\beta \sum_{j=0}^{n} q_j\right] = e^{\beta Q}. \end{aligned} \tag{2.17}$$

If we represent the states by x_t in place of i_j, we can write

$$\boxed{\frac{p[x(t)|x_0]}{\tilde{p}[\tilde{x}(t)|\tilde{x}_0]} = e^{\beta Q}.} \tag{2.18}$$

where time runs from 0 to τ, while the *tilde* symbol implies time reversed path. This is the Crooks' Fluctuation Theorem.

However, all of this was for a given initial point x_0 for forward and $\tilde{x}_0$ for reverse process. If the initial distribution of both the forward and reverse processes are canonical, then we have

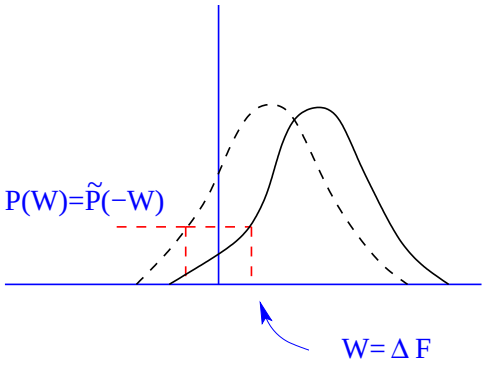
$$\frac{p[x(t)|x_0]\ p_0(x_0)}{\tilde{p}[\tilde{x}(t)|\tilde{x}_0]\ p_1(\tilde{x}_0)} = e^{\beta Q} e^{\beta(\Delta U - \Delta F)} = e^{\beta(W - \Delta F)}. \tag{2.19}$$

Here, $p[x(t)|x_0]$ is the path probability for the path $\{x_0, x_1, \cdots, x_\tau\}$ Now, we collect all the forward(reverse) trajectories along which W amount of work has been done on(by) the system. Using the ratio property as before (see the Hamiltonian case), we finally arrive at the relation

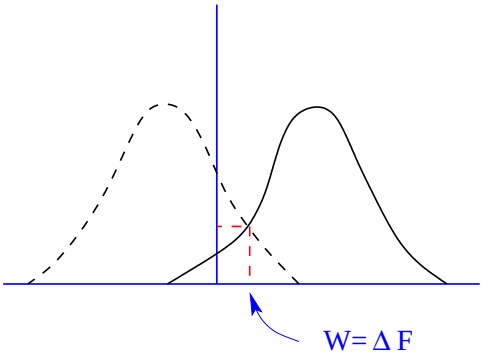
$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta F)}, \tag{2.20}$$

which is the Jarzynski nonequilibrium work theorem.

2.3 Some comments



1. Finding ΔF from work distribution: according to the Crooks' theorem, $P(W)/\tilde{P}(-W) = \exp(W - \Delta F)$, we find that the work probability values $P(W)$ and $\tilde{P}(-W)$ will be equal at $W = \Delta F$. Now if we "reflect" the $\tilde{P}(\tilde{W})$ plot about the $W = 0$ axis, we will find that they will actually intersect each other at $W = \Delta F$ (see adjacent figure. In first figure, the dotted curve represents $\tilde{P}(\tilde{W})$ distribution. In the lower plot, the dotted curve represents its reflection). The reflection may be achieved by simply plotting $\tilde{P}(-W)$ vs W instead of $\tilde{P}(W)$ vs W. This property is exploited to numerically extract information about the equilibrium free energy difference between the two states.



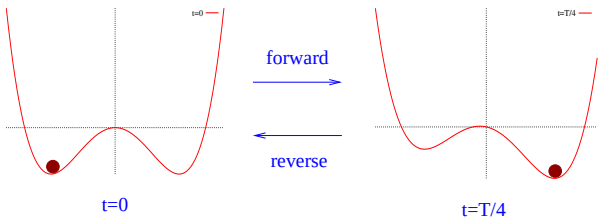
2. The Crooks' fluctuation theorem provides a partial answer to the Loschmidt paradox[1]: how does macroscopic irreversibility arise from microscopically time-reversible equations of motion? The Crooks fluctuation theorem says that for macroscopic systems, since the dissipated work will also be large (increases with the system size), the time-reversed paths will be exponentially suppressed, thus giving rise to irreversibility in macroscopic domain.
   However, we extract this information not in the sense Loschmidt would have liked, which is by time-reversing the entire universe. When the process is being carried out, we either disconnect the bath from the system, or keep it connected but analyze only the system states by means of stochastic dynamics.

   When we speak of time-reversed path, we time-reverse only the system states and are not concerned with the bath variables. At the end of the day, we find an answer that ought to be satisfied even if we would have time-reversed the entire universe! It seems that the source of the trouble is the enigmatic quantity called *dissipated heat*. As shown by Crooks, if Markovian dynamics is detailed balanced, then the difference between forward and corresponding reverse path is dependent entirely on this quantity.

[1] Loschmidt's paradox was a consequence of the H-theorem put forward by Boltzmann, which said that entropy of any isolated system always increases. But if the microscopic equations of motion are irreversible, why not the macroscopic phenomena? Forget about any external perturbation for the moment. We allow the isolated system to evolve for time τ, and at the end we reverse all the velocities, and allow the system to evolve for another interval τ. It ought to return to its initial state. Claim is, we will not be able to distinguish which of the two processes we are watching. However, the fluctuation theorems (in particular, the Seifert's integral fluctuation theorem) show that when the isolated system relaxes to equilibrium, the total entropy production of the system is always positive. This will hold even in the presence of an external time-dependent protocol.

It is sometimes suggested that we consider the time-reversed process by simply screening the video of the forward process backwards. In that case, it is often confusing to realize why it is said that the process is irreversible. For example, if a distribution $p(x_0)$ at time $t = 0$ goes to $p(x_\tau)$ at time $t = \tau$, it looks obvious that if we screen the video backwards, then a distribution $p(x_\tau)$ at $t = 0$ goes to $p(x_0)$ at $t = \tau$! Here it will be useful to keep in mind that we are not applying time-reversal in the exact sense of reverse screening of a video. If at all, we are doing this only with the video of the system, and are *not fidgeting in any way with the states of the bath.*

3. **Example: meaning of a time-reversed process:**



Suppose I have a particle performing over-damped motion inside a double-welled potential. As a typical (frequently occurring) trajectory of the forward process, I have shown the particle moving from the left to the right well when the potential is tilted towards right, say by using a sinusoidal force $f(t) = A \sin \omega t$, with t going from 0 to $T/4$, T being the time period.

Another typical trajectory for this process would be the particle being in the right well initially, and then staying throughout in the same well. Now let us come back to the first case. For the sake of simplicity, I assume that the particle hops into the right well instantaneously just before $t = T/4$. What shall be the corresponding reverse trajectory? It will be as follows: at $\tilde{t} = 0$, the particle is in the right well. As soon as the reverse protocol begins (this will bring the potential from the right to the left configuration), i.e. at $\tilde{t} = 0^+$, the particle hops into the left well (which is higher at this stage)! Obviously, this will be a rare phenomenon in the reverse process. On the other hand, the typical reverse trajectory will be as expected: if intially the particle is in the right well, then it stays in the right well throughout the process. Otherwise, it hops from the left to the right well. So the typical trajectories of forward process have reversed trajectories that become the rare ones along the reverse process. This is known as the phenomenon of "duality".

4. If some external force is applied, a particle that follows Newtonian dynamics will have an acceleration in the direction of force:

$$m\frac{d^2x}{dt^2} = F(x, t).$$

Now, one argues that there is no reason why the particle cannot move in the direction $\tau - t$, because the LHS remains same, and so will the RHS if t is replaced by $\tau - t$. It is very important to realize that we are not wondering whether a particle is actually going backward in time (i.e., from future to past). *We can never go backward in time.* If my Hamiltonian is $H(t)$, a particle can never experience a Hamiltonian $H(\tau - t)$, unless we explicitly apply the time-reversed Hamiltonian.

Let me provide a concrete example: suppose I compress a cylinder of gas by pushing a piston. Question is, can a few of the gas particles, during this compression process, observe that the piston is being pulled out? No! For doing that, we actually need to pull out the piston and carry on this process *forward* in time for the same time interval. Forward and time-reversed trajectories *cannot be observed in a single process.*

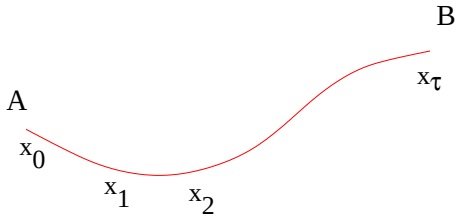
But who does that in nature? Who reverses the protocol? *Nature itself*! In fact, if the perturbation states (say, the positions and velocities of the piston in the cylinder) are also included as a part of the isolated system, it looks pretty obvious that since we can choose the initial state randomly, the exact time-reversed initial state along with time-reversed perturbation state can be chosen with equal probability.

3 The Entropy production fluctuation theorems

3.1 The Integral Fluctuation Theorem (IFT)

The Crooks Fluctuation Theorem reads

$$\boxed{\frac{p[x(t)|x_0]}{\tilde{p}[\tilde{x}(t)|\tilde{x}_0]} = e^{\Delta s_m}.} \tag{3.1}$$



where $p[x(t)|x_0]$ is the probability of forward path and $\tilde{p}[\tilde{x}(t)|\tilde{x}_0]$ is that of the reversed path, s_m being the medium entropy.

$$\begin{aligned} \text{One defines} \quad R \equiv & \ln\left\{\frac{p[x(t)|x_0]p_0(x_0)}{\tilde{p}[\tilde{x}(t)|\tilde{x}_0]p_1(\tilde{x}_0)}\right\} \\ = & \Delta s_m + \ln\frac{p_0(x_0)}{p_1(x_\tau)} \end{aligned} \tag{3.2}$$

The following general identity can be easily derived [5]:

$$\begin{aligned} \langle e^{-R}\rangle \equiv & \sum_{x(t),x_0} p[x(t)|x_0]p_0(x_0)e^{-R} \\ = & \sum_{\tilde{x}(t),\tilde{x}_0} \tilde{p}[\tilde{x}(t)|\tilde{x}_0]p_1(\tilde{x}_0) = 1. \end{aligned} \tag{3.3}$$

Note: There is *no* restriction on the choices of $p_0(x_0)$ and $p_1(x_\tau)$!!! However, if we *suppose* that $p_0(x_0)$ and $p_1(x_\tau)$ are respectively the initial and final probability densities of the path, then we say that the system entropy is given by

$$\boxed{s(t) = -\ln p(x,t) \quad \Rightarrow \quad \Delta s \equiv s(\tau) - s(0) = -\ln\frac{p_1(x_\tau)}{p_0(x_0)}.} \tag{3.4}$$

Note that the above definition is not unreasonable if we keep in mind that the change in the macroscopic entropy of an ensemble is given by

$$S = \langle \Delta s\rangle = -\int p(x,t)\ln p(x,t) = \langle -p(x,t)\rangle.$$

With the above definition, we have $R = \Delta s_{tot}$=change in *total entropy* of the system.

Let us see what the Jensen's inequality ($\langle e^x\rangle \geq e^{\langle x\rangle}$) has to say about the IFT: $\langle e^{-\Delta s_{tot}}\rangle = 1$ (see section 7.3 for a general proof of the Jensen's inequality, or the footnote here[1] for the proof in the present context). It says that

$$\langle e^{-\Delta s_{tot}}\rangle = 1 \geq e^{-\langle \Delta s_{tot}\rangle},$$

so that

$$\boxed{\langle \Delta s_{tot}\rangle \geq 0,} \tag{3.5}$$

[1] $e^x \geq 1 + x$ (always. Plot a graph and check!), so that [6] $\langle e^x\rangle = \langle e^{x-\langle x\rangle+\langle x\rangle}\rangle = e^{\langle x\rangle}\langle e^{x-\langle x\rangle}\rangle \geq e^{\langle x\rangle}\langle 1 + x - \langle x\rangle\rangle = e^{\langle x\rangle}$, Q.E.D.

which is the second law for mesoscopic systems. Note that the relation is valid for all times, irrespective of whether the system is in transient or steady state. In particular, we can consider an infinitesimal time of observation, dt, so that we get

$$\frac{d\langle s_{tot}\rangle}{dt} = \langle \dot{s}_{tot}\rangle \geq 0, \tag{3.6}$$

which is another form of the second law at mesoscopic scales.

Putting $p_0(x) = \frac{\exp[-V(x_0,\lambda_0)/T]}{Z_0}$ and $p_1(x) = \frac{\exp[-V(x_\tau,\lambda_\tau)/T]}{Z_\tau}$ in eqn. (3.2), one recovers the Jarzynski equality

$$\langle e^{-w_d/T}\rangle = 1. \tag{3.7}$$

3.2 The Detailed Fluctuation Theorem (DFT)

It is given by

$$\boxed{\frac{p(\Delta s_{tot})}{p(-\Delta s_{tot})} = e^{\Delta s_{tot}}.} \tag{3.8}$$

which holds in the steady state.

The symmetry function is defined as

$$f(\Delta s_{tot}) \equiv \ln\left[\frac{p(\Delta s_{tot})}{p(-\Delta s_{tot})}\right]. \tag{3.9}$$

which, when plotted as a function of Δs_{tot} should have a slope of 1 in the steady state.

Proof In steady state,

$$\begin{aligned} p(\Delta s_{tot}) =& p(R) = \langle \delta(R'-R)\rangle \\ =& \int_{x_0,x_\tau} p^s(x_0)p[x_\tau|x_0]\delta(R'-R)\mathcal{D}[x_t]dx_0dx_\tau, \end{aligned}$$

where $\mathcal{D}[x_t]dx_0dx_\tau$ is the integral over all trajectories[1], all initial points and all final points.

$$\therefore p(\Delta s_{tot}) = \int_{x_0,x_\tau} p^s(x_\tau)\tilde{p}[\tilde{x}_\tau|\tilde{x}_0]\left(e^R\right)\delta(R'-R)\mathcal{D}[x_t]dx_0dx_\tau.$$

(using the definition 3.2 of R).

Next, we use the fact that R changes sign in the time-reversed process (which once again follows from 3.2: the numerator and denominators get interchanged for the reverse process), and the fact that the Jacobian $\left|\frac{\partial x_t}{\partial \tilde{x}_{\tau-t}}\right|$ is 1 for any t (here we recall that $x(t) = \tilde{x}(\tau - t)$):

$$\begin{aligned} p(\Delta s_{tot}) =& e^R \int_{x_0,x_\tau} p^s(x_\tau)\tilde{p}[\tilde{x}_\tau|\tilde{x}_0]\delta(R'+R)\mathcal{D}[x_t]dx_0dx_\tau \\ =& e^R \int_{\tilde{x}_0,\tilde{x}_\tau} p^s(\tilde{x}_0)\tilde{p}[\tilde{x}_\tau|\tilde{x}_0]\delta(R'+R)\mathcal{D}[x_t]d\tilde{x}_0d\tilde{x}_\tau \end{aligned}$$

[1]See appendix A for the description of a functional integral

$$=e^R\langle\delta(R'+R)\rangle = e^R p(-R)$$
$$=e^R p(-\Delta s_{tot}). \qquad \text{Q.E.D.} \tag{3.10}$$

4 The Hatano-Sasa idenity

Now we embark upon a new identity, the Hatano-Sasa identity [7]. This time we consider transition between *nonequilibrium steady states* (NESS). Although the detailed fluctuation theorem given by Seifert for total entropy Δs_{tot} is valid under a particular steady state, this is not so for transition between two different steady states. For doing this, following Oono and Paniconi, the authors have divided the total dissipated heat into two parts:

$$Q_{tot} = Q_{hk} + Q_{ex}. \tag{4.1}$$

Here, Q_{hk} is called the *housekeeping heat* that is dissipated to keep the system in a given steady state, when the system is held in a single *macroscopic* state. On the other hand, Q_{ex} is the heat dissipated because of the change in the macroscopic state of the system. The steady state distribution is denoted by

$$\rho_{ss}(x,\lambda) = e^{-\phi(x,\lambda)}, \tag{4.2}$$

with $\lambda(t)$ being the external drive, and $\phi(x,\lambda)$ is an effective potential (dimensionless). The authors prove the following identity:

$$\boxed{\langle e^{-\beta Q_{ex}-\Delta\phi}\rangle = 1.} \tag{4.3}$$

Here, $\Delta\phi \equiv \phi(x_N,\lambda_N) - \phi(x_0,\lambda_0)$. First we prove the following identity:

$$\left\langle \prod_{i=0}^{N-1} \frac{\rho_{ss}(x_{i+1},\lambda_{i+1})}{\rho_{ss}(x_{i+1},\lambda_i)} \right\rangle = 1. \tag{4.4}$$

To prove this, we proceed as follows:

$$\begin{aligned}
\left\langle \prod_{i=0}^{N-1} \frac{\rho_{ss}(x_{i+1},\lambda_{i+1})}{\rho_{ss}(x_{i+1},\lambda_i)} \right\rangle &= \int \prod_{i=0}^{N} dx_i \ \rho_{ss}(x_0,\lambda_0) \prod_{i=0}^{N-1} P(x_{i+1}|x_i,\lambda_i) \ \frac{\rho_{ss}(x_{i+1},\lambda_{i+1})}{\rho_{ss}(x_{i+1},\lambda_i)} \\
&= \int \prod_{i=0}^{N} dx_i \ \prod_{i=0}^{N-1} \frac{P(x_{i+1}|x_i,\lambda_i)}{\rho_{ss}(x_{i+1},\lambda_i)} \prod_{i=0}^{N} \rho_{ss}(x_i,\lambda_i) \\
&= \int \prod_{i=0}^{N} dx_i \ \rho_{ss}(x_N,\lambda_N) \prod_{i=0}^{N-1} \frac{P(x_{i+1}|x_i,\lambda_i)\rho_{ss}(x_i,\lambda_i)}{\rho_{ss}(x_{i+1},\lambda_i)} \\
&= \int dx_N \ \rho_{ss}(x_N,\lambda_N) \prod_{i=1}^{N-1} \int dx_i \frac{P(x_{i+1}|x_i,\lambda_i)\rho_{ss}(x_i,\lambda_i)}{\rho_{ss}(x_{i+1},\lambda_i)} \\
&\qquad \times \frac{\int dx_0 P(x_1|x_0,\lambda_0)\rho_{ss}(x_0,\lambda_0)}{\rho_{ss}(x_1,\lambda_0)} \\
&= \int dx_N \ \rho_{ss}(x_N,\lambda_N) \prod_{i=1}^{N-1} \int dx_i \frac{P(x_{i+1}|x_i,\lambda_i)\rho_{ss}(x_i,\lambda_i)}{\rho_{ss}(x_{i+1},\lambda_i)} \times \frac{\rho_{ss}(x_1,\lambda_0)}{\rho_{ss}(x_1,\lambda_0)} \\
&= \int dx_N \ \rho_{ss}(x_N,\lambda_N) \prod_{i=1}^{N-1} \int dx_i \frac{P(x_{i+1}|x_i,\lambda_i)\rho_{ss}(x_i,\lambda_i)}{\rho_{ss}(x_{i+1},\lambda_i)} \\
&= \cdots = \int dx_N \ \rho_{ss}(x_N,\lambda_N) = 1.
\end{aligned} \tag{4.5}$$

which is same as (4.4). Here we have made use of the fact that the steady state probability distribution does not change with time (step 4 to step 5). This means,

$$\left\langle \exp\left[\sum_{i=0}^{N-1}\{-\phi(x_{i+1},\lambda_{i+1})+\phi(x_{i+1},\lambda_i)\}\right]\right\rangle = 1. \tag{4.6}$$

As $N\to\infty$, we can write the summation as an integral:

$$\boxed{\left\langle \exp\left[-\int_0^\tau dt\dot{\lambda}\frac{\partial\phi(x,\lambda)}{\partial\lambda}\right]\right\rangle = 1.} \tag{4.7}$$

Now, the *total* heat dissipated into bath is given by

$$Q_{tot} = \int_0^\tau dt[\gamma\dot{x}(t)-\xi(t)]\dot{x}(t), \tag{4.8}$$

where the system is described by the Langevin equation

$$\gamma\dot{x} = f - U'(x,\lambda)+\xi(t). \tag{4.9}$$

The symbols have their usual meanings. Now, we rewrite this equation as follows:

$$\gamma\dot{x} = b(x) - \frac{1}{\beta}\frac{\partial\phi}{\partial x}+\xi(t), \tag{4.10}$$

where

$$b(x) = f - U'(x,\lambda)+\frac{1}{\beta}\frac{\partial\phi}{\partial x}. \tag{4.11}$$

$$\begin{aligned}\therefore \beta Q_{tot} =& \beta\int_0^\tau[\gamma\dot{x}(t)-\xi(t)]\dot{x}(t)dt = \int_0^\tau\left[\beta b(x)-\frac{\partial\phi}{\partial x}\right]\dot{x}dt\\ =& \int_0^\tau\beta b(x)\dot{x}dt-\Delta\phi+\int_0^\tau\frac{\partial\phi}{\partial\lambda}\dot{\lambda}dt.\end{aligned} \tag{4.12}$$

We call the first term the *housekeeping heat* Q_{hk}:

$$Q_{hk}\equiv\int_0^\tau b(x)\dot{x}dt. \tag{4.13}$$

The *excess heat* Q_{ex} is now simply defined as

$$Q_{ex}\equiv Q_{tot}-Q_{hk}. \tag{4.14}$$

Thus, from (4.7), we get the desired identity (4.3).

5 The Harada-Sasa identity

Let us consider a Brownian particle following the underdamped Langevin equation [8]:

$$m\ddot{x}+\gamma\dot{x} = F_\lambda(x,t)+\epsilon f_p(t)+\xi(t), \tag{5.1}$$

where $\lambda(t)$ is the control parameter, ϵf_p is the perturbation force, while $F_{\lambda(t,y)}(x,t)$ is the force that is derived from the potential function, plus some external constant force (if present). Now, the fluctuation-dissipation theorem is given by,

$$\langle\dot{x}^2\rangle_0 = \frac{2}{\beta}R(t,t), \tag{5.2}$$

where $R(t,t)$ is the response function defined as

$$\langle \dot{x}(t)\rangle_\epsilon = \langle \dot{x}(t)\rangle_0 + \epsilon \int_0^t ds R(t,s) f_p(s). \tag{5.3}$$

Here, $\langle\cdots\rangle_\epsilon$ denotes the average in presence of external perturbation perturbation, while $\langle\cdots\rangle_0$ denotes that without perturbation. In a nonequilibrium state, the effective temperature of the system is defined through

$$T_{eff} \equiv \frac{\langle \dot{x}^2\rangle_0}{2k_B R(t,t)}. \tag{5.4}$$

From the underdamped Langevin equation, the probability of a trajectory is given by

$$P_{\lambda(t,y)}[x(t)|x_0] = N.\exp\left[-\frac{\beta}{4\gamma}\int_0^\tau dt(m\ddot{x}+\gamma\dot{x}-F_\lambda-\epsilon f_p)^2\right], \tag{5.5}$$

N being the normalization constant. The Seifert's theorem is then given by (using the above equation)

$$\frac{P_\lambda[x(t)|x_0]\rho_0(x_0)}{P_{\lambda^\dagger}[\tilde{x}(t)|\tilde{x}_0]\rho_\tau(\tilde{x}_0)} = e^{\Delta s_{tot}}, \tag{5.6}$$

where $\Delta s_{tot} = \beta Q + \langle\Delta\phi\rangle$, with

$$\begin{aligned} Q &\equiv \int_0^\tau dt\dot{x}(t)\left[F_{\lambda(t,y)}(x(t)) + \epsilon f_p(t) - m\ddot{x}(t)\right] = \int_0^\tau dt\dot{x}(t)[\gamma\dot{x}(t) - \xi(t)]; \\ \phi &\equiv -\ln\rho(t). \end{aligned} \tag{5.7}$$

To discuss the fluctuation-dissipation theorem, we first write the identity (which follows from direct differentiation):

$$\frac{\partial}{\partial\epsilon}\left\langle \dot{x}(t)\exp\left[-\epsilon\beta\int_0^t dt'\dot{x}(t')f_p(t')\right]\right\rangle\Big|_{\epsilon=0} = \frac{\partial\langle\dot{x}(t)\rangle_\epsilon}{\partial\epsilon}\Big|_{\epsilon=0} - \beta\int_0^t dt' f_p(t')\langle\dot{x}(t)\dot{x}(t')\rangle_0. \tag{5.8}$$

From (5.3) we readily obtain

$$\frac{\partial\langle\dot{x}(t)\rangle_\epsilon}{\partial\epsilon} = \int_0^t dt' R(t,t') f_p(t'). \tag{5.9}$$

The LHS of (5.8) can be evaluated directly:

$$\begin{aligned} &\frac{\partial}{\partial\epsilon}\left\langle \dot{x}(t)\exp\left[-\epsilon\beta\int_0^t dt'\dot{x}(t')f_p(t')\right]\right\rangle\Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial\epsilon}\int \mathcal{D}[x]P_\lambda[x(t)]\dot{x}(t)\exp\left[-\epsilon\beta\int_0^t dt'\dot{x}(t')f_p(t')\right]\Big|_{\epsilon=0} \\ &= N\frac{\partial}{\partial\epsilon}\int \mathcal{D}[x]\ \dot{x}(t)\ \rho_0(0) \\ &\quad\times\exp\left[-\frac{\beta}{4\gamma}\int_0^t dt'[m\ddot{x}(t')+\gamma\dot{x}(t')-F_\lambda(x(t'))-\epsilon f_p(t')]^2 - \epsilon\beta\int_0^t dt'\dot{x}(t')f_p(t')\right]\Big|_{\epsilon=0} \\ &= N\frac{\partial}{\partial\epsilon}\int \mathcal{D}[x]\ \dot{x}(t)\ \rho_{ss}(0) \\ &\quad\times\exp\left[\frac{\beta}{4\gamma}\int_0^t dt' 2\epsilon f_p(t')[m\ddot{x}(t')+\gamma\dot{x}(t')-F_\lambda(x(t'))] - \frac{\beta\epsilon^2}{4\gamma}\int_0^t dt' f_p^2(t')\right. \\ &\quad\left.-\ \epsilon\beta\int_0^t dt'\dot{x}(t')f_p(t') + \text{other terms}\right]\Big|_{\epsilon=0} \qquad \text{step(*)} \end{aligned}$$

$$
\begin{aligned}
&= N \int \mathcal{D}[x]\ \dot{x}(t)\ \rho_0(0) P[x(t)|x_0]_{\epsilon=0} \\
&\quad \times \left[\frac{\beta}{2\gamma} \int_0^t dt' f_p(t') [m\ddot{x}(t') + \gamma \dot{x}(t') - F_\lambda(x(t'))] - \beta \int_0^t dt' \dot{x}(t') f_p(t') \right] \\
&= \frac{\beta}{2\gamma} \int_0^t dt' f_p(t') \left\langle \dot{x}(t) [m\ddot{x}(t') + \gamma \dot{x}(t') - F_\lambda(x(t'))] \right\rangle_0 - \beta \int_0^t dt' f_p(t') \langle \dot{x}(t) \dot{x}(t') \rangle_0 \\
&= \boxed{\frac{\beta}{2\gamma} \int_0^t dt' f_p(t') \left\langle \dot{x}(t) [m\ddot{x}(t') - \gamma \dot{x}(t') - F_\lambda(x(t'))] \right\rangle_0 .}
\end{aligned}
\tag{5.10}
$$

In step (*), we have written down explicitly all the terms in the exponential that contain ϵ, because they will be acted upon by the derivative wrt ϵ. Finally, on taking $\epsilon = 0$, only the term that was quadratic in ϵ will disappear. The symbol $\langle \cdots \rangle_0$ implies averaging with respect to the unperturbed path probability $P[x(t)|x_0]_{\epsilon=0}$.

Now, suppose $f_p(t') = \delta(t' - (t-s))$, then for $s \neq t$, then using (5.8), (5.9) and (5.10), we get

$$
\begin{aligned}
\frac{\partial}{\partial \epsilon} \left\langle \dot{x}(t) e^{-\epsilon \beta \dot{x}(t-s)} \right\rangle &= R(t, t-s) - \beta \langle \dot{x}(t) \dot{x}(t-s) \rangle_0 \\
&= \frac{\beta}{2\gamma} \langle \dot{x}(t) [m\ddot{x}(t-s) - \gamma \dot{x}(t-s) - F_\lambda(x(t-s))] \rangle_0 \\
\Rightarrow R(t, t-s) - \frac{\beta}{2} \langle \dot{x}(t) \dot{x}(t-s) \rangle_0 &= \langle \dot{x}(t) [m\ddot{x}(t-s) - F_\lambda(x(t-s))] \rangle_0 \\
\Rightarrow \gamma \left[\langle \dot{x}(t) \dot{x}(t-s) \rangle_0 - \frac{2}{\beta} R(t, t-s) \right] &= \langle \dot{x}(t) [F_\lambda(x(t-s)) - m\ddot{x}(t-s)] \rangle_0
\end{aligned}
\tag{5.11}
$$

Now substituting $s = 0$ proveides the central result, the Harada-Sasa identity.

$$
\boxed{\gamma \left[\langle \dot{x}^2(t) \rangle_0 - \frac{2}{\beta} R(t,t) \right] = \langle \dot{x}(t) [F_\lambda(x(t)) - m\ddot{x}(t)] \rangle_0 .}
\tag{5.12}
$$

The LHS gives the extent of violation of the fluctuation dissipation theorem, while the RHS gives the rate of dissipation of heat into the bath.

6 Quantum Work Relations

6.1 Time reversal

In the nanoscopic (lower limit of mesoscopic) world, the quantum version of nonequilibrium work relations must be formulated. In this section, we will follow reference [10]. For this purpose, let us define the *time reversal* operator. This is defined as an antilinear operator that changes the sign of parameters that are *odd* in time, e.g., the momentum or magnetic field.

$$
\Theta H(t, B) = H(t, -B) \Theta \tag{6.1}
$$

Or, using the obvious relation $\Theta^2 = I$, one finds

$$
\Theta H(t, B) \Theta = H(t, -B). \tag{6.2}
$$

For the forward process, we have the initial state canonical density matrix given by

$$
\rho(0) = \frac{e^{-\beta H(0,B)}}{Z(0)} \tag{6.3}
$$

where $Z(0) = Tr\left[e^{-\beta H(0,B)}\right] \equiv e^{-\beta F(0)}$, which is basically $\sum_{\{E\}}\left[e^{-\beta E(0,B)}\right]$, a trace being the summation of eigenvalues. Now, starting from the equilibrium distribution at $t = 0$, the system is allowed to evolve till $t = T$ under Hamiltonian dynamics. The *forward* time evolution is given by

$$\boxed{i\hbar\frac{\partial}{\partial t}U_F(t,B) = H(t,B)U_F(t,B),} \tag{6.4}$$

which follows naturally from the Schrodinger's equation

$$\begin{aligned} i\hbar\frac{\partial\psi(x,t,B)}{\partial t} =& H(t,B)\psi(x,t,B) \\ \Rightarrow i\hbar\frac{\partial}{\partial t}[U_F(t,B)\psi(x,0,B)] =& H(t,B)[U_F(t,B)\psi(x,0,B)] \end{aligned} \tag{6.5}$$

which is true for arbitrary $\psi(x,0,B)$, and hence (6.4) immediately follows.

Next, we will move on to the Heisenberg representation in which some time -independent observable becomes time-dependent according to

$$A_F(t) = U_F^\dagger(t).A.U_F(t). \tag{6.6}$$

6.2 Digression: the density matrix

The average of an observable A will be given by (if $|\psi\rangle = \sum_i c_i|i\rangle$ be the state of the system)

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \sum_j\langle\psi|A|j\rangle\langle j|\psi\rangle = \sum_j\langle j|\psi\rangle\langle\psi|A|j\rangle = \mathrm{Tr}[\rho A]. \tag{6.7}$$

That was all about the system being in a *pure* state. However, in most cases we do not have the full information about the state of the system. In that case, we have what is known as the *statistical mixture of states*, in which we only know that the system has p_k probability of staying in the state ψ_k.

Thus, we will then have:

$$\begin{aligned} \langle A\rangle =& \sum_\alpha p_\alpha\langle\psi_\alpha|A|\psi_\alpha\rangle \\ =& \sum_\alpha p_\alpha\sum_j\langle\psi_\alpha|A|j\rangle\langle j|\psi_\alpha\rangle \\ =& \sum_{\alpha j} p_\alpha\langle j|\psi_\alpha\rangle\langle\psi_\alpha|A|j\rangle \\ =& \sum_j\langle j|\left(\sum_\alpha p_\alpha|\psi_\alpha\rangle\langle\psi_\alpha|\right)A|j\rangle \\ =& \mathrm{Tr}(\rho A), \end{aligned} \tag{6.8}$$

where the general density matrix is given by $\rho = \sum_\alpha p_\alpha|\psi_\alpha\rangle\langle\psi_\alpha|$. By the way, the $|\psi_\alpha\rangle$'s need not be orthogonal to each other. We can also express the density operator in terms of the orthogonal basis states:

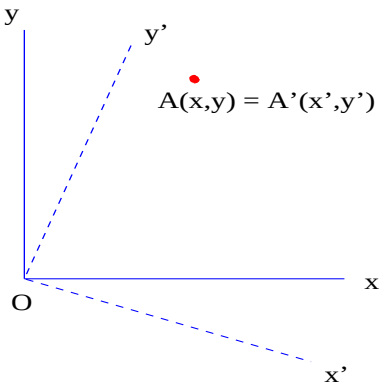
$$\rho = \sum_\alpha p_\alpha|\psi_\alpha\rangle\langle\psi_\alpha| = \sum_k p_k|k\rangle\langle k|, \tag{6.9}$$

where p_k is the probability of finding the system in the state $|k\rangle$.

[**Question**: If we know that the probability of being in state $|\psi_k\rangle$ is p_k, don't we have the full state as $|\Psi\rangle = \sum_k\sqrt{p_k}\,|\psi_k\rangle$?

Answer: Certainly not, because in general $|\Psi\rangle = \sum_k\sqrt{p_k}\,e^{i\alpha}|\psi_k\rangle$, and α, whose knowledge is essential to calculate interference effects, is unknown.]

6.3 Unitary transformation



Suppose we have been given a function A (say, the particle density) defined in the $x-y$ plane. Now the axes are rotated about the origin and the new axes are called x' and y' axes, respectively. Now consider any *given* point (x, y) in the $x-y$ axis. Tha value of A at this point is $A(x, y)$. This can be represented as a continuous matrix A_{xy} or $\langle x|A|y\rangle$. However, in the new basis, although the value of A at this point remains same (number of particles per unit area does not change), the functional form will look like $A'(x', y')$ or $\langle x'|A'|y'\rangle$. Let me denote the rotation operation by R, so that

$$|x'\rangle = R|x\rangle; \qquad |y'\rangle = R|y\rangle. \tag{6.10}$$

Then we have,

$$\langle x|A|y\rangle = \langle x'|A'|y'\rangle = \langle x|R^\dagger A' R|y\rangle. \tag{6.11}$$

Since the above is valid for any point on the plane, we must have the relation

$$A = R^\dagger A' R \quad \Rightarrow \quad \boxed{A' = RAR^\dagger.} \tag{6.12}$$

In words, an observer in the rotated coordinate system will see the given function as $RAR^\dagger$ instead of A. This was a basic example of unitary transformation. Let us now generalize this. We have an operator A represented in two bases, $\{|i\rangle\}$ and $\{|i'\rangle\}$, respectively. Each space may be multidimensional (our example above was for a two-dimensional case). Now, the matrix elements of some *operator* A will be given by $\langle i|A|j\rangle$. In the primed basis (formed by unitary transformation on the unprimed one), the matrix elements are $\langle i'|A'|j'\rangle$. By definition,

$$|i'\rangle = U|i\rangle, \tag{6.13}$$

where U is the unitary transform operator. Then we have,

$$\langle i|A|j\rangle = \langle i'|A'|j'\rangle = \langle i|U^\dagger A' U|j\rangle. \tag{6.14}$$

The above being valid for any i, j, we must have

$$A = U^\dagger A' U \quad \Rightarrow \quad \boxed{A' = UAU^\dagger.} \tag{6.15}$$

6.4 A universal work relation

So finally we have [10], in the case of our Heisenberg operator $A_F(t)$

$$\langle A_F(t)\rangle = Tr[\rho(0)A_F(t)]. \tag{6.16}$$

On the other hand, in the *backward* process, the Hamiltonian becomes $H(T-t, -B)$, so that now the initial density matrix (at $t = 0$) becomes

$$\rho(T) = \frac{e^{-\beta H(T,-B)}}{Z(T)} \tag{6.17}$$

where $Z(T) = \mathrm{Tr} e^{-\beta H(T,-B)} = e^{-\beta F(T)}$. The evolution operator for the *backward* process is defined as

$$\boxed{i\hbar \frac{\partial}{\partial t} U_R(t, B) = H(T-t, B)\ U_R(t, B).} \tag{6.18}$$

Next, we can proceed further after proving the following lemma:

$$\boxed{\Theta U_F(T-t,B)\ U_F^\dagger(T,B)\Theta = U_R(t,-B).} \tag{6.19}$$

For proving the above lemma, we first substitute $T-t$ in place of t in eqn. (6.4):

$$-i\hbar\frac{\partial}{\partial t}U_F(T-t,B) = H(T-t,B)\ U_F(T-t,B)$$

(note that $\frac{\partial}{\partial t}$ changes to $-\frac{\partial}{\partial t}$ as well). Now, multiplying by Θ from the left and by $U_F^\dagger(T,B)\Theta$ from the right, we get

$$+i\hbar\frac{\partial}{\partial t}[\Theta\ U_F(T-t,B)\ U_F^\dagger(T,B)\ \Theta] = H(T-t,-B)\ [\Theta\ U_F(T-t,B)U_F^\dagger(T,B)\Theta]$$

Comparing this with

$$i\hbar\frac{\partial}{\partial t}U_R(t,B) = H(T-t,B)\ U_R(t,B),$$

we immediately arrive at (6.19).

Putting $t\to T-t$ in (6.19) and thereafter multiplying from both left and right by Θ, one finds

$$U_F(t,B)\ U_F^\dagger(T,B) = \Theta\ U_R(T-t,-B)\ \Theta \tag{6.20}$$

Next, we have, for some operator A which has some *definite parity* under time-reversal.

$$\begin{aligned}
A_F(t) =& U_F^\dagger(t)\ A\ U_F(t)\\
=& U_F^\dagger(T)\ \underbrace{U_F(T)\ .U_F^\dagger(t)}_{[\Theta U_R(T-t,-B)\Theta]^\dagger}\ A\underbrace{U_F(t).\ U_F^\dagger(T)}_{\Theta U_R(T-t,-B)\Theta}U_F(T)\\
=& U_F^\dagger(T)\ .\Theta U_R^\dagger(T-t,-B)\underbrace{\Theta.\ A.\ \Theta}_{\epsilon_A A}U_R(T-t,-B)\Theta.\ U_F(T)\\
=& \epsilon_A U_F^\dagger(T)\ .\Theta\underbrace{U_R^\dagger(T-t,-B)\ .A.\ U_R(T-t,-B)}_{A_R(T-t)}\Theta.\ U_F(T)\\
=& \epsilon_A U_F^\dagger(T)\ .\Theta A_R(T-t)\Theta.\ U_F(T),
\end{aligned} \tag{6.21}$$

where $\epsilon_A = \pm 1$, and $A_R(T-t)\equiv U_R^\dagger(T-t,-B)\ .A.\ U_R(T-t,-B)$. Multiplying by $\lambda(t)$ on both sides and taking the exponents, we have

$$\begin{aligned}
\exp\left[\int_0^T dt\lambda(t)A_F(t)\right] =& \exp\left[\epsilon_A U_F^\dagger(T)\Theta\int_0^T dt\lambda(t)A_R(T-t)\Theta U_F(T)\right]\\
=& U_F^\dagger(T)\Theta\exp\left[\epsilon_A\int_0^T dt\lambda(t)A_R(T-t)\right]\Theta U_F(T)\\
=& U_F^\dagger(T)\Theta\exp\left[\epsilon_A\int_T^0(-dt)\lambda(T-t)A_R(t)\right]\Theta U_F(T)\\
=& U_F^\dagger(T)\Theta\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\Theta U_F(T).
\end{aligned} \tag{6.22}$$

In the third step, we have put $t\to T-t$. To get to the second step from the first, one simply needs to expand the integral in a Taylor's series and use $U_F^\dagger U_F = 1 = \Theta^2$:

$$\exp(\alpha^\dagger I\alpha) =1+\alpha^\dagger I\alpha+\frac{1}{2}\alpha^\dagger I\alpha.\alpha^\dagger I\alpha+\cdots=1+\alpha^\dagger I\alpha+\frac{1}{2}\alpha^\dagger I^2\alpha+\cdots$$
$$=\alpha^\dagger[1+I+I^2+\cdots]\alpha=\alpha^\dagger\exp(I)\alpha,$$

where $\alpha=\Theta U_F(T)$ in our case.

Now we have

$$\begin{aligned}
&\mathrm{Tr}\left\{\rho(0)\exp\left[\int_0^T dt\lambda(t)A_F(t)\right]\exp[-\beta H_F(T)]\exp[\beta H(0)]\right\}\\
=&\mathrm{Tr}\left\{\exp\left[\int_0^T dt\lambda(t)A_F(t)\right]e^{-\beta H_F(T)}e^{\beta H(0)}\frac{e^{-\beta H(0)}}{Z(0)}.\right\}\\
=&\mathrm{Tr}\left\{\frac{Z(T)}{Z(0)}\exp\left[\int_0^T dt\lambda(t)A_F(t)\right]\frac{e^{-\beta H_F(T)}}{Z(T)}\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{U_F^\dagger(T)\Theta\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\Theta U_F(T)\frac{e^{-\beta H_F(T)}}{Z(T)}\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\Theta U_F(T)\frac{e^{-\beta H_F(T)}}{Z(T)}U_F^\dagger(T)\Theta\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\Theta\frac{e^{-\beta H(T)}}{Z(T)}\Theta\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\frac{e^{-\beta H(T,-B)}}{Z(T)}\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\rho(T)\right\}\\
=&e^{-\beta\Delta F}\mathrm{Tr}\left\{\rho(T)\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\right\}.
\end{aligned}\tag{6.23}$$

In the first step, we have used the property of cyclic permutation within trace. Also, in going from third to fourth as well as from seventh (second-last) to last step, we have used the same property. We have used the definition $\rho(T)\equiv\frac{e^{-\beta H(T,-B)}}{Z(T)}$.

Thus, it follows that

$$\begin{aligned}
&\left\langle\rho(0)\exp\left[\int_0^T dt\lambda(t)A_F(t)\right]e^{-\beta H_F(T)}e^{\beta H(0)}\right\rangle_{F,B}\\
&=e^{-\beta\Delta F}\left\langle\exp\left[\epsilon_A\int_0^T dt\lambda(T-t)A_R(t)\right]\right\rangle_{R,-B}.
\end{aligned}\tag{6.24}$$

which is the central result of [10].

Now we can utilize the above powerful expression to generate several relations in non-equilibrium statistical mechanics or in response theory. In particular, if we pot $\lambda=0$ in (6.24), then we arrive

$$\left\langle e^{-\beta H_F(T)}e^{\beta H(0)}\right\rangle_{F,B}=e^{-\beta\Delta F}\tag{6.25}$$

which is the celebrated Jarzynski equality, as will be elaborated below.

6.5 Jarzynski equality: quantum version

We define the probability density for (Jarzynski) work done W [12] as

$$\begin{aligned}
P(W) =& \sum_{i,f} \delta[W - (\epsilon_f(T) - \epsilon_i(0))]\ P(\epsilon_f(T)|\epsilon_i(0))\ P(\epsilon_i(0)) \\
=& \sum_{i,f} \delta[W - (\epsilon_f(T) - \epsilon_i(0))]\ .|\langle\psi_f|U(T)|\psi_i\rangle|^2.\ \frac{e^{-\beta\epsilon_i(0)}}{Z(0)} \\
=& \sum_{i,f} \delta[W - (\epsilon_f(T) - \epsilon_i(0))]\ .\langle\psi_f|U(T)|\psi_i\rangle\ \langle\psi_i|U^\dagger(T)|\psi_f\rangle.\ \frac{e^{-\beta\epsilon_i(0)}}{Z(0)}.
\end{aligned}$$

The Fourier transform of $P(W)$ is defined through

$$\widehat{P}(u) \equiv \int P(W) e^{iuW} dW. \tag{6.26}$$

Integrating of W, which results in removal of the delta function, gives

$$\begin{aligned}
\widehat{P}(u) =& \sum_{i,f} \langle\psi_f|U(T)|\psi_i\rangle\ \langle\psi_i|U^\dagger(T)|\psi_f\rangle\ e^{iu[\epsilon_f(T)-\epsilon_i(0)]}\ \frac{e^{-\beta\epsilon_i(0)}}{Z(0)} \\
=& \frac{1}{Z(0)} \sum_{i,f} \langle\psi_f|U(T)|\psi_i\rangle\ \langle\psi_i|U^\dagger(T)|\psi_f\rangle e^{iu\epsilon_f(T)} e^{-(iu+\beta)\epsilon_i(0)} \\
=& \frac{1}{Z(0)} \sum_{i,f} \left\langle \psi_f \left| U(T) e^{-(iu+\beta)H} \right| \psi_i \right\rangle \left\langle \psi_i \left| U^\dagger(T) e^{iuH} \right| \psi_f \right\rangle \\
=& \frac{1}{Z(0)} \sum_{i,f} \left\langle \psi_i \left| U^\dagger(T) e^{iuH} \right| \psi_f \right\rangle \left\langle \psi_f \left| U(T) e^{-(iu+\beta)H} \right| \psi_i \right\rangle \\
=& \frac{1}{Z(0)} \sum_{i} \left\langle \psi_i \left| U^\dagger(T) e^{iuH} U(T) e^{-(iu+\beta)H} \right| \psi_i \right\rangle \\
=& \frac{1}{Z(0)} \sum_{i} \left\langle \psi_i \left| e^{iuH_F(T)} e^{-(iu+\beta)H} \right| \psi_i \right\rangle
\end{aligned} \tag{6.27}$$

Substituting $u = i\beta$, we get from (6.26) and (6.27) the following relation:

$$\widehat{P}(i\beta) = \left\langle e^{-\beta W} \right\rangle = \frac{1}{Z(0)} \sum_{i} \left\langle \psi_i \left| e^{-\beta H_F(T)} \right| \psi_i \right\rangle$$

By noting that $\rho(0) = \frac{e^{-\beta\epsilon_i(0)}}{Z(0)}$, i.e., $\frac{1}{Z(0)} = \rho(0) e^{-\beta\epsilon_i(0)}$, we get

$$\boxed{\left\langle e^{-\beta W} \right\rangle = \sum_{i} \left\langle \psi_i \left| \rho(0) e^{-\beta H_F(T)} e^{\beta H(0)} \right| \psi_i \right\rangle = \left\langle e^{-\beta H_F(T)} e^{\beta H(0)} \right\rangle.} \tag{6.28}$$

Thus, from (6.25) and (6.28), we arrive at the quantum Jarzynski equality:

$$\boxed{\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta F}.} \tag{6.29}$$

6.6 The Tasaki-Crooks Fluctuation Theorem

We consider a quantum system in *weak* thermal contact with the thermal bath until a time t_0, after which either the bath is switched off, or is retained at this weak level and can be neglected (the more general condition of strong system-bath coupling will be discussed in the next subsection)[13].

The protocol is described by a Hamiltonian $\{H(t)\}_{t_f,t_0}$, which govern the time evolution of the system in the interval (t_0, t_f). We have the initial density

$$\rho(t_0) = \frac{e^{-\beta H(t_0)}}{Z(t_0)}, \tag{6.30}$$

which forms continuous matrix depending on the initial state:

$$\rho(t_0) = \sum_i \frac{e^{-\beta E_i}}{Z(t_0)} |\psi_i(t_0)\rangle\langle\psi_i(t_0)|. \tag{6.31}$$

Here, $|\psi_i(t_0)\rangle$ are the eigenstates of the Hamiltonian $H(t_0)$. One first measures the energy $e_n(t_0)$ at time t_0 (and the system immediately collapses to the eigenstate $|\psi_n(t_0)\rangle$), then allows the system to evolve and finally measure the energy $e_m(t_f)$ at time t_f (at which the system collapses to eigenstate $|\psi_m(t_f)\rangle$). The work w is given by the difference between these two measurements (concept of heat does not appear): $w = e_m(t_f) - e_n(t_0)$. Thus the work probability is given by

$$p_{t_f,t_0}(w) \equiv \sum_{nm} \delta(w - [e_m(t_f) - e_n(t_0)])\ p(m, t_f|n)\ p(n), \tag{6.32}$$

where

$$p(m, t_f|n) \equiv |\langle\psi_m(t_f)|U(t_f, t_0)|\psi_n(t_0)\rangle|^2. \tag{6.33}$$

We now define the Fourier transform of $p(w)$:

$$\begin{aligned}
G_{t_f,t_0}(u) =& \int dw\ e^{iuw} \sum_{nm} \delta(w - [e_m(t_f) - e_n(t_0)])\ p(m, t_f|n)\ p(n) \\
=& \sum_{nm} e^{iu[e_m(t_f) - e_n(t_0)]}\ \langle\psi_m(t_f)|U(t_f, t_0)|\psi_n(t_0)\rangle\ \langle\psi_n(t_0)|U^\dagger(t_f, t_0)|\psi_m(t_f)\rangle\ \frac{e^{-\beta E_i}}{Z(t_0)} \\
=& \sum_{nm} \langle\psi_m(t_f)|U(t_f, t_0)e^{-iuH(t_0)}\rho(t_0)|\psi_n(t_0)\rangle\ \langle\psi_n(t_0)|U^\dagger U(t_f, t_0)e^{iuH(t_f)}|\psi_m(t_f)\rangle \\
=& \mathrm{Tr}\left[U(t_f, t_0)\ e^{-iuH(t_0)}\ U(t_f, t_0)\ e^{-iuH(t_0)}\ \rho(t_0)\right] \\
=& \mathrm{Tr}\left[U^\dagger(t_f, t_0)\ e^{iuH(t_f)}\ U(t_f, t_0)\ e^{-iuH(t_0)}\ \rho(t_0)\right] \\
=& \mathrm{Tr}\left[e^{iuH_H(t_f)}\ e^{-iuH(t_0)}\ \rho(t_0)\right],
\end{aligned} \tag{6.34}$$

where $H_H(t) = U^\dagger(t, t_0)\ e^{iuH(t)}\ U(t, t_0)$. In the above statements we have used the completeness of the initial states and cyclic property of trace. We can rewrite the above as

$$G_{t_f,t_0}(u) = \frac{1}{Z(t_0)} \mathrm{Tr}\left[U^\dagger(t_f, t_0)\ e^{iuH(t_f)}\ U(t_f, t_0)\ e^{-iuH(t_0)}\ e^{-\beta H(t_0)}\right]. \tag{6.35}$$

or,

$$Z(t_0)G_{t_f,t_0}(u) = \mathrm{Tr}\left[U^\dagger(t_f, t_0)\ e^{iuH(t_f)}\ U(t_f, t_0)\ e^{\overbrace{-(iu+\beta)}^{v} H(t_0)}\right]$$

$$
\begin{aligned}
&=\text{Tr}\left[U^\dagger(t_f,t_0)\ e^{i(-v+i\beta)H(t_f)}\ U(t_f,t_0)\ e^{ivH(t_0)}\right]\\
&=\text{Tr}\left[e^{-ivH(t_f)}e^{-\beta H(t_f)}\ U^\dagger(t_0,t_f)\ e^{ivH(t_0)}\ U(t_0,t_f)\right]\\
&=\text{Tr}\left[U^\dagger(t_0,t_f)\ e^{ivH(t_0)}\ U(t_0,t_f)\ e^{-ivH(t_f)}e^{-\beta H(t_f)}\right]\\
&=Z(t_f)G_{t_0,t_f}(v),
\end{aligned}
\tag{6.36}
$$

where we have written down the last line by comparing with the expression in the first line. Thus,

$$
\boxed{Z(t_0)G_{t_f,t_0}(u)=Z(t_f)G_{t_0,t_f}(v),}
\tag{6.37}
$$

which means

$$
G_{t_f,t_0}(u)=\frac{Z(t_f)}{Z(t_0)}G_{t_0,t_f}(v)=e^{-\beta\Delta F}G_{t_0,t_f}(v).
\tag{6.38}
$$

Taking the inverse Fourier transform on both sides, we get

$$
\begin{aligned}
p_{t_f,t_0}(w)=&e^{-\beta\Delta F}\frac{1}{2\pi}\int du\ e^{-iuw}G_{t_0,t_f}(-u+i\beta)\\
=&e^{-\beta\Delta F}\int du\ e^{(-u+i\beta)w}e^{\beta w}G_{t_0,t_f}(-u+i\beta)\\
=&e^{-\beta\Delta F}e^{\beta w}\int du\ e^{-i(-u+i\beta)(-w)}G_{t_0,t_f}(-u+i\beta)\\
=&e^{\beta(w-\Delta F)}\int dv\ e^{-iv(-w)}G_{t_0,t_f}(v) &(6.39)\\
=&e^{\beta(w-\Delta F)}p_{t_0,t_f}(-w). &(6.40)
\end{aligned}
$$

So we finally get

$$
\boxed{\frac{p_{t_f,t_0}(w)}{p_{t_0,t_f}(-w)}=e^{\beta(w-\Delta F)},}
\tag{6.41}
$$

which is the Tasaki-Crooks Fluctuation Theorem.

6.7 Fluctuation theorems for arbitrary open quantum systems

Let the total Hamiltonian of the (system+bath) be given by [20]

$$
H(t)=H_S(t)+H_B+H_{SB},
\tag{6.42}
$$

where the subscripts S and B stand for system and bath, respectively, and SB stands for the coupling between system and bath. This total Hamiltonian, being that of an isolated system (system+bath), must follow the Tasaki-Crooks fluctuation theorem:

$$
\frac{p_{t_f,t_0}(w)}{p_{t_0,t_f}(-w)}=\frac{Y(t_f)}{Y(t_0)}e^{\beta w},
\tag{6.43}
$$

where

$$
Y(t)=\text{Tr}\ e^{-\beta[H_S(t)+H_{SB}+H_B]}.
\tag{6.44}
$$

Note that we have written ΔF of (6.41) as ratio of the partition functions of the total Hamiltonian. Now, we need to find partition function only for the system, because we want to derive the TCFT for the system only. Since the bath is assumed to stay always at equilibrium, its free energy remains constant, and the

change in equilibrium free energy for the supersystem (consisting of system+bath) must be equal to the free energy change of the system of interest: $\Delta F = \Delta F_S$. We define the free energy of the system as

$$F_S(t) \equiv F(t) - F_B, \tag{6.45}$$

where t refers to $\lambda(t)$, i.e., the value of the external parameter at time t, and

$$F_B = \underset{B}{\mathrm{Tr}}\, e^{-\beta H_B}. \tag{6.46}$$

This satisfies the above mentioned condition:

$$\Delta F = \int_0^\tau dt \dot{\lambda} \frac{\partial F(t)}{\partial \lambda} = \int_0^\tau dt \dot{\lambda} \frac{\partial F_S(t)}{\partial \lambda} = \Delta F_S, \tag{6.47}$$

because F_B is time-independent. Multiplying by -1 and taking exponentials of both sides, we then get

$$Z_S(t) = Y(t) \cdot Z_B. \tag{6.48}$$

Therefore we have

$$Z_S(t) = \frac{Y(t)}{Z_B} = \frac{\underset{B,S}{\mathrm{Tr}}\; e^{-\beta[H_S(t)+H_{SB}+H_B]}}{\underset{B}{\mathrm{Tr}}\, e^{-\beta H_B}}. \tag{6.49}$$

Thus,

$$\frac{Y(t_f)}{Y(t_0)} = \frac{Z_S(t_f)}{Z_S(t_0)}, \tag{6.50}$$

since Z_B gets cancelled out. Then we know from (6.43) that

$$\boxed{\frac{p_{t_f,t_0}(w)}{p_{t_0,t_f}(-w)} = \frac{Z_S(t_f)}{Z_S(t_0)} e^{\beta w} = e^{\beta(W - \Delta F_S)}}. \tag{6.51}$$

7 Mathematical preliminaries for relative entropy

The relative entropy between two distributions, say $p(x)$ and $q(x)$, is defined as

$$D[p(x)||q(x)] \equiv \int dx\; p(x) \ln \frac{p(x)}{q(x)}. \tag{7.1}$$

It is also known as the *Kullback-Leibler* distance between the distributions $p(x)$ and $q(x)$. Note that it is not a proper distance function, in the sense that besides being asymmetric, it also violates the triangle inequality.

7.1 Multivariate Gaussian integration

The most general multivariate Gaussian distribution is

$$P(\mathbf{z}) = N \exp\left[-\frac{1}{2}\mathbf{z}^\dagger . \boldsymbol{\sigma}^{-1} . \mathbf{z}\right]. \tag{7.2}$$

where $\mathbf{z} \equiv (z_1, \cdots, z_n)$.

Little can we deduce about the normalization constant N by looking at the above equation. The situation, however, becomes vastly simplified if we keep in mind that if we can diagonalize $\boldsymbol{\sigma}^{-1}$, then we are through, because the components of $\mathbf{z}$ vectors become immediately decoupled to give rise to a product of one-variable Gaussian integrals. Accordingly, let $\mathbf{D}$ be the matrix that diagonalizes $\boldsymbol{\sigma}^{-1}$.

Thus we have

$$\begin{aligned} P(\mathbf{z}) =& N \exp\left[-\frac{1}{2}\mathbf{z}^\dagger \ .\mathbf{D}^\dagger\mathbf{D}.\boldsymbol{\sigma}^{-1}.\mathbf{D}^\dagger\mathbf{D}.\ \mathbf{z}\right] \\ =& N \exp\left[-\frac{1}{2}\mathbf{z}^\dagger\mathbf{D}^\dagger \ .\mathbf{A_D}.\ \mathbf{Dz}\right] \\ =& N \exp\left[-\frac{1}{2}(\mathbf{y}^\dagger.\mathbf{A_D}.\mathbf{y})\right] \\ =& N e^{-(\frac{y_1^2}{2\lambda_1}+\cdots+\frac{y_n^2}{2\lambda_n})} \end{aligned} \tag{7.3}$$

$\mathbf{A_D}$ being a diagonal matrix with eigenvalues $\frac{1}{\lambda_1}$, etc., $\lambda_1, \cdots, \lambda_n$ being themselves eigenvalues of $\boldsymbol{\sigma}$, and $\mathbf{y} \equiv \mathbf{Dz}$.

We have

$$\begin{aligned} \int P(\mathbf{z}) d\mathbf{z} =& 1 \\ \Rightarrow N \int e^{-(\frac{y_1^2}{2\lambda_1}+\cdots+\frac{y_n^2}{2\lambda_n})} d\mathbf{y} \left|\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right| =& 1 \end{aligned}$$

But, the Jacobian

$$\begin{aligned} \left|\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right| &= \left|\frac{\partial(z_1, \cdots, z_n)}{\partial(y_1, \cdots, y_n)}\right| = \left|\frac{\partial(z_1, \cdots, z_n)}{\partial(\sum D_{1j} z_j, \cdots, \sum D_{nj} z_j)}\right| \\ &= \left|\left(\frac{\partial(\sum D_{1j} z_j, \cdots, \sum D_{nj} z_j)}{\partial(z_1, \cdots, z_n)}\right)^{-1}\right| = |\mathbf{D}^{-1}| = 1, \end{aligned} \tag{7.4}$$

$\mathbf{D}$ being a unitary matrix. Therefore we finally have

$$N \int e^{-(\frac{y_1^2}{2\lambda_1}+\cdots+\frac{y_n^2}{2\lambda_n})} d\mathbf{y} = 1 = N\sqrt{(2\pi)^n \lambda_1 \cdots \lambda_n}. \tag{7.5}$$

which implies

$$N = \frac{1}{(2\pi)^{n/2}\sqrt{\det \sigma}}.$$

Therefore we finally have

$$\boxed{P(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \sigma}} e^{-\frac{1}{2}\mathbf{z}^\dagger.\boldsymbol{\sigma}^{-1}.\mathbf{z}}} \tag{7.6}$$

7.2 General expression for relative entropy for Gaussian distributions

We have the forward process

$$\rho_F(z) = \frac{1}{2\pi\sqrt{\det \sigma_F}} \exp\left[-\frac{1}{2}(z - \bar{z}_F)^T.\sigma_F^{-1}.(z - \bar{z}_F)\right] \tag{7.7a}$$

and the reverse process

$$\rho_R(z^*) = \frac{1}{2\pi\sqrt{\det\sigma_R}} \exp\left[-\frac{1}{2}(z^* - \bar{z}_R)^T.\sigma_R^{-1}.(z^* - \bar{z}_R)\right]. \tag{7.7b}$$

Here

$$\sigma_F = \begin{pmatrix} \sigma_{x^2} & \sigma_{xp} \\ \sigma_{xp} & \sigma_{p^2} \end{pmatrix}; \qquad \sigma_R = \begin{pmatrix} \sigma_{x^2} & -\sigma_{xp} \\ -\sigma_{xp} & \sigma_{p^2} \end{pmatrix},$$

which imply

$$\sigma_F^{-1} = \frac{1}{\det\sigma_F} \begin{pmatrix} \sigma_{p^2} & -\sigma_{xp} \\ -\sigma_{xp} & \sigma_{x^2} \end{pmatrix}; \qquad \sigma_R^{-1} = \frac{1}{\det\sigma_R} \begin{pmatrix} \sigma_{p^2} & \sigma_{xp} \\ \sigma_{xp} & \sigma_{x^2} \end{pmatrix}.$$

$$\begin{aligned} \therefore \ln\frac{\rho_F(z)}{\rho_R(z^*)} =& \frac{1}{2}\ln\frac{\det\sigma_R}{\det\sigma_F} \ - \ \frac{1}{2}(z^* - \bar{z}_R)^T.\sigma_R^{-1}.(z^* - \bar{z}_R) \ + \ \frac{1}{2}(z^* - \bar{z}_R)^T.\sigma_R^{-1}.(z^* - \bar{z}_R) \\ =& \frac{1}{2}\ln\frac{\det\sigma_R}{\det\sigma_F} - \frac{1}{2\det\sigma_F}[\Delta x^2\sigma_p^2 - 2\Delta x\Delta p\ \sigma_{xp} + \Delta p^2\sigma_x^2] \\ &+ \frac{1}{2}[(z^* - \bar{z}_F^*) + (\bar{z}_F^* - \bar{z}_R)]^T.\sigma_R^{-1}.[(z^* - \bar{z}_F^*) + (\bar{z}_F^* - \bar{z}_R)] \end{aligned} \tag{7.8}$$

Here, $\Delta x = x - \bar{x}$, $\Delta x^2 = (x - \bar{x})^2$, etc.

$$\begin{aligned} \int dz\ \rho_F(z)\ln\frac{\rho_F(z)}{\rho_R(z^*)} =& \frac{1}{2}\ln\frac{\det\sigma_R}{\det\sigma_F} - \frac{\sigma_x^2\sigma_p^2 - 2\sigma_{xp}^2 + \sigma_p^2\sigma_x^2}{2(\sigma_x^2\sigma_p^2 - \sigma_{xp}^2)} \\ &+ \frac{1}{2}\left\langle(z^* - \bar{z}_F^*)^T\sigma_R^{-1}(z^* - \bar{z}_F^*)\right\rangle_F + \frac{1}{2}(\bar{z}_F^* - \bar{z}_R)^T\sigma_R^{-1}(\bar{z}_F^* - \bar{z}_R) \\ &+ \frac{1}{2}\left\langle(\bar{z}_F^* - \bar{z}_F^*)^T\sigma_R^{-1}(z^* - \bar{z}_F^*)\right\rangle_F + \frac{1}{2}\left\langle(\bar{z}_F^* - \bar{z}_R)^T\sigma_R^{-1}(z^* - \bar{z}_F^*\right\rangle_F. \end{aligned} \tag{7.9}$$

The last to terms give zero, while second term is -1. The third term is

$$\begin{aligned} \frac{1}{2}\left\langle(z^* - \bar{z}_F^*)^T\sigma_R^{-1}(z^* - \bar{z}_F^*)\right\rangle_F =& \frac{1}{2\det\sigma_R}\left\langle\Delta x^2\sigma_p^2 - 2\Delta x\Delta p\ \sigma_{xp} + \Delta p^2\sigma_x^2\right\rangle_F \\ =& \frac{(\sigma_x^2\sigma_p^2 - \sigma_{xp}^2)}{\det\sigma_R} = \frac{1}{2}\mathrm{Tr}(\sigma_R^{-1}\sigma_F), \end{aligned} \tag{7.10}$$

which can be verified by direct evaluation. Thus, finally we have [14]

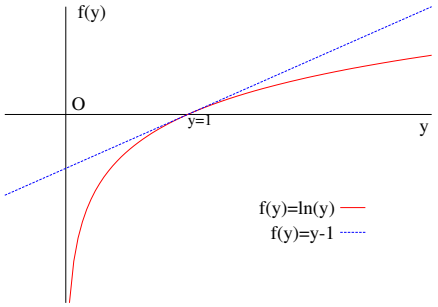
$$\boxed{D[\rho_F(z)||\rho_R(z^*)] = -1 + \frac{1}{2}\left[\ln\frac{\det\sigma_R}{\det\sigma_F} + \mathrm{Tr}(\sigma_R^{-1}\sigma_F^*)\right] + \frac{1}{2}(\bar{z}_F^* - \bar{z}_R)^T\sigma_R^{-1}(\bar{z}_F^* - \bar{z}_R).} \tag{7.11}$$

7.3 Some properties of relative entropy and Shannon entropy

1. **Positivity: simple method**

   The relative entropy is defined as

$$D[p(x)||q(x)] \equiv \int dx\ p(x)\ln\frac{p(x)}{q(x)}. \tag{7.12}$$



We will now use the following identity (see figure):

$$\ln y \leq y-1 \quad \Rightarrow \quad -\ln y \geq 1-y. \qquad \text{(always)}$$

$$\begin{aligned}
\therefore D[p(x)||q(x)] &= -\int dx\ p(x) \ln \frac{q(x)}{p(x)} \\
&\geq \int dx\ p(x) \left[1 - \frac{q(x)}{p(x)}\right] = \int dx\ p(x) - \int dx\ q(x) \\
&= 1 - 1 = 0. \qquad \text{QED}
\end{aligned} \tag{7.13}$$

2. **Positivity: longer but illuminating method**

For proving the positivity of the relative entropy, we simply need to use the following property of convex functions (jensen's inequality):

$$\langle f(x) \rangle \geq f(\langle x \rangle) \tag{7.14}$$

whose more general form is

$$\begin{aligned}
\langle f(g(x)) \rangle &\geq f(\langle g(x) \rangle) \\
\text{i.e.,} \quad \sum_x p(x) f(g(x)) &\geq f\left(\sum_x p(x) g(x)\right); \\
\text{or,} \quad \int dx\ p(x) f(g(x)) &\geq f\left(\int dx\ p(x) g(x)\right),
\end{aligned} \tag{7.15}$$

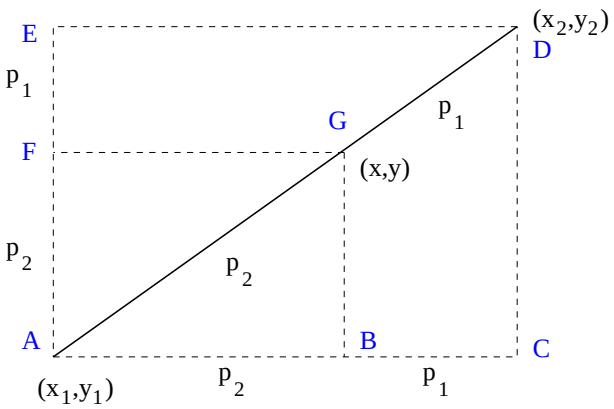
where $p(x)$ is a probability distribution function for the discrete variable case (second inequality) or a probability *density* function in the case of continuous x (last inequality). Noting that $-\ln x$ is a convex function of x, we have

$$\begin{aligned}
D[p(x)||q(x)] &= \sum_x p(x) \ln \frac{p(x)}{q(x)} = -\sum_x p(x) \ln \frac{q(x)}{p(x)} \\
&\geq -\ln\left(\sum_x p(x) \frac{q(x)}{p(x)}\right) = -\ln\left(\sum_x q(x)\right) \\
&= -\ln 1 = 0.
\end{aligned} \tag{7.16}$$

$$\therefore \boxed{D[p(x)||q(x)] \geq 0} \qquad Q.E.D. \tag{7.17}$$

Proof of the Jensen's inequality :

Let us first deal with a distribution function which takes up two values: p_1 and p_2, where $p_1 + p_2 = 1$. Now we consider the following geometry:



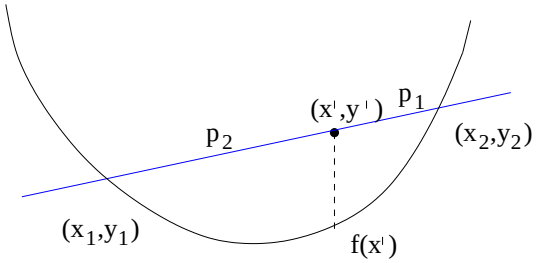
the adjacent diagram shows a straight line joining the points (x_1, y_1) and (x_2, y_2). The point (x, y) divides the above line segment in the ratio $p_1 : p_2$, with $p_1 + p_2 = 1$. We first note that using the similarity of triangles ($\triangle ABG \sim \triangle ACD$, and $\triangle AFG \sim \triangle AED$), we find that their corresponding projections on the x and the y axes also divide themselves in the same ratio.

Thus, we have

$$\frac{x_2 - x}{x - x_1} = \frac{y_2 - y}{y - y_1} = \frac{p_1}{p_2}. \tag{7.18}$$

From the above relations, we obtain

$$x = p_1 x_1 + p_2 x_2; \quad y = p_1 y_1 + p_2 y_2. \tag{7.19}$$



Next, we have a *convex* function $y \equiv f(x)$ (i.e., $y'' \geq 0$ at each x), say. The point (x', y') (see adjacent diagram) is given by $(p_1 x_1 + p_2 x_2,\ p_1 f(x_1) + p_2 f(x_2)) \equiv (\langle x \rangle, \langle f(x) \rangle)$. On the other hand, $f(x') = f(p_1 x_1 + p_2 x_2)$ is a point on the curve,as shown in the figure. Clearly, $f(\langle x \rangle)$ lies below $\langle f(x) \rangle$. That proves the definition (7.14).

Next, we proceed by method of induction [15]. Let $\sum_{i=1}^{k-1} p_i f(x_i) \geq f(\sum_{i=1}^{k-1} p_i x_i)$. Now,

$$\begin{aligned}
\sum_{i=1}^{k} p_i f(x_i) =& p_k f(x_k) + \sum_{i=1}^{k-1} p_i f(x_i) \\
=& p_k f(x_k) + (1 - p_k) \sum_{i=1}^{k-1} p'_i f(x_i), \quad \text{where } p'_i \equiv \tfrac{p_i}{1-p_k} \\
\geq& p_k f(x_k) + (1 - p_k) f\left(\sum_{i=1}^{k-1} p'_i x_i \right) \quad \text{(by assumption)} \\
\geq& f\left(p_k x_k + (1 - p_k) \sum_{i=1}^{k-1} p'_i x_i \right) \\
=& f\left(p_k x_k + \sum_{i=1}^{k-1} p_i x_i \right) = f\left(\sum_{i=1}^{k-1} p_i x_i \right) \qquad \text{Q.E.D.}
\end{aligned} \tag{7.20}$$

3. **Maximization of Shannon entropy:**

We have,

$$S = -\sum_i p_i \ln p_i; \qquad \sum_i p_i = 1. \tag{7.21}$$

Let us define a quantity

$$S' = -\sum_{i=1}^{N} p_i \ln p_i + \lambda \left(\sum_i p_i - 1 \right) = -\sum_{i-1}^{N} \left(p_i \ln p_i - \lambda p_i + \frac{\lambda}{N} \right), \tag{7.22}$$

where λ is a Lagrange multiplier that has been used to maximize S subject to the normalization constraint. We now maximize S' wrt the collection $\{p_i\}$ as follows:

$$\delta S' = 0 \;\Rightarrow\; \sum_i (1 + \ln p_i - \lambda)\delta p_i = 0. \tag{7.23}$$

Now, we know that all the p_i's are not independent, and so the coefficient of δp_i in the above equation cannot be immediately set to zero. However, we can always *choose* λ such that the coefficient of the dependent probability, say p_N, vanishes: $1 + \ln p_N - \lambda = 0$. Now all the coefficients are identically zero:

$$1 + \ln p_i - \lambda = 0 \;\Rightarrow\; p_i = e^{\lambda - 1} = const. \tag{7.24}$$

Although it is straightforward to see that if all probabilities are equal (some constant), then the normalization condition gives $p_i = 1/N$ for all i, let us derive it in a more formal way by maximizing S' wrt λ:

$$\frac{\partial S'}{\partial \lambda} = 0 \quad \Rightarrow \quad \sum_i \left(-p_i + \frac{1}{N} \right) = 0 \quad \Rightarrow \quad p_i = \frac{1}{N}. \tag{7.25}$$

We then readily obtain

$$S_{max} = -\sum_{i=1}^{N} \frac{1}{N} \ln \frac{1}{N} = \ln N. \tag{7.26}$$

A Functional integrals and derivatives

A.1 Functional derivative

I will follow the treatment of [17] in this entire discussion. A functional is a mapping from a functional *form* to a *number*.

Examples:

$$F[\phi] = \int_0^1 dx \; \left\{ \phi^2(x) + 3x \frac{\partial \phi}{\partial x} \right\}. \tag{A.1a}$$

$$F[\phi] = \phi(0). \tag{A.1b}$$

Note that a functional is different from the *function of a function*, e.g. $\phi^2(x) + x\phi(x)$, because in the latter case the value of the expression changes with change in x. On the other hand, the value of a *functional* will depend only on the specific *form* of the function, independent of the argument of the function.

Now let us define

$$F[\phi] = \int_a^b dx \; \Psi(x, \phi(x), \phi'(x), \phi''(x), \cdots, \phi^{(n)}(x)). \tag{A.2}$$

Here, of course, F is the *functional* of ϕ while Ψ is the *function* of ϕ and its dervatives. To calculate the *functional derivative* of $F[\phi]$ w.r.t. ϕ, we define

$$\begin{aligned} \delta F[\phi] \equiv & F[\phi + \delta\phi] - F[\phi] \\ = & \int_a^b dx \; \Psi(x, \phi + \delta\phi, \phi' + \delta\phi', \cdots, \phi^{(n)} + \delta\phi^{(n)}) - \int_a^b dx \; \Psi(x, \phi, \phi', \cdots, \phi^{(n)}), \end{aligned} \tag{A.3}$$

where $\delta\phi \equiv \epsilon\eta(x)$, and $\eta(x)$ is some function of x that vanishes at $x = a$ and at $x = b$. Here, ϵ is a small parameter that goes to zero. Now, keeping the first order terms in the Taylor's expansion of Ψ, we get

$$\delta F[\phi] = \int_a^b dx \left\{ \frac{\partial\Psi}{\partial\phi}\delta\phi + \frac{\partial\Psi}{\partial\phi'}\delta\phi' + \cdots + \frac{\partial\Psi}{\partial\phi^{(n)}}\delta\phi^{(n)} \right\}. \tag{A.4}$$

Now, we have

$$\int_a^b dx\ \frac{\partial\Psi}{\partial\phi'}\delta\phi' = \int_a^b dx\ \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi'}\delta\phi\right] - \int_a^b \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi'}\right]\delta\phi = -\int_a^b \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi'}\right]\delta\phi, \tag{A.5}$$

where in the last step we have set the first term equal to zero, because $\delta\phi(x) = \epsilon\eta(x)$ vanishes at the boundaries. Similarly,

$$\begin{aligned}
\int_a^b dx\ \frac{\partial\Psi}{\partial\phi^{(k)}}\delta\phi^{(k)} &= \int_a^b dx\ \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi^{(k)}}\delta\phi^{(k-1)}\right] - \int_a^b \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi^{(k)}}\right]\delta\phi^{(k-1)} \\
&= -\int_a^b \frac{d}{dx}\left[\frac{\partial\Psi}{\partial\phi^{(k)}}\right]\delta\phi^{(k-1)} \\
&= \cdots \ = \ (-1)^k \int_a^b dx\ \left(\frac{\partial}{\partial x}\right)^k \frac{\partial\Psi}{\partial\phi^{(k)}}\delta\phi,
\end{aligned} \tag{A.6}$$

where we have performed repeated integrations by parts. Therefore we get from (A.4),

$$\delta F[\phi] = \sum_{k=0}^{n} \int_a^b dx\ \left(-\frac{\partial}{\partial x}\right)^k \frac{\partial\Psi}{\partial\phi^{(k)}}\delta\phi. \tag{A.7}$$

We next define the functional derivative through

$$\delta F[\phi] \equiv \int_a^b \frac{\delta F[\phi]}{\delta\phi}\delta\phi(x). \tag{A.8}$$

Thus, equating (A.7) and (A.8), we get the functional derivative in terms of *ordinary derivatives* to be

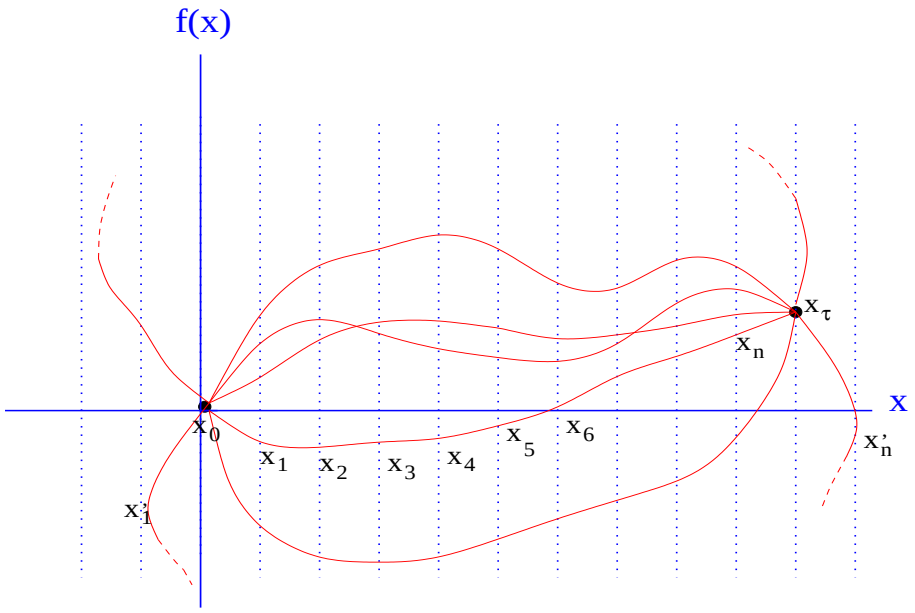
$$\boxed{\frac{\delta F[\phi]}{\phi(x)} = \sum_{k=0}^{n} \left(-\frac{\partial}{\partial x}\right)^k \frac{\partial\Psi}{\partial\phi^{(k)}}.} \tag{A.9}$$

A.2 Functional integral

This concept is more comprehensible compared to the functional derivative. To explain the concept of *functional* or *path* integral, let us first discretize the trajectories/paths (with different functions $\{\phi(x)\}$ describing each of them) that begin from some point x_0 and ends at some other point x_τ. Any such trajectory can be described as $\phi(x) \equiv \{x_0, x_1, x_2, \cdots, x_n, x_\tau\}$. Different such sets $\{x\}$ (with initial and final points same) will give different forms of $\phi(x)$. The functional integral is an integral over *each* of these points from x_1 to x_n, i.e.,

$$\int \mathcal{D}[\phi(x)]\ F[\phi] \equiv \int_{-\infty}^{\infty} dx_1 dx_2 \cdots dx_n\ F(x_0, x_1, x_2, \cdots, x_n, x_\tau). \tag{A.10}$$

Note that we have written the *functional* $F[\phi]$ as a *function* of *each* of the points constituting the function $\phi(x)$. This is okay, since any particular set $x_0, \cdots, x_\tau$ (which corresponds to a particular functional form $\phi(x)$) will correspond to some *number* $F(x_0, x_1, x_2, \cdots, x_n, x_\tau)$ unique to that functional form.



Concept of functional integral: discretization of the trajectories

Now, if there is a probability density provided for the initial and final pounts as well, then the value of $F[\phi]$ beginning from *any* initial point and ending in *any* final point will include an integral over x_0 and x_τ as well:

$$\int dx_0 dx_\tau \mathcal{D}[\phi(x)]\; p(x_0)p(x_\tau)F[\phi] \equiv \int_{-\infty}^{\infty} dx_0 dx_1 dx_2 \cdots dx_n dx_\tau\; p(x_0)p(x_\tau)F(x_0, x_1, x_2, \cdots, x_n, x_\tau). \quad \text{(A.11)}$$

Here, $p(x_0)$ and $p(x_\tau)$ are the initial and final probability densities, respectively. If the difference between the successive discrete points x_i and x_{i+1} tends to zero, we will get the value for a continuous trajectory.

Now, I would like to clarify one further point. Look at the above figure: the set $\{x_0, x_1, \cdots, x_n, x_\tau\}$ gives one of the paths while $\{x_0, x_1', \cdots, x_n', x_\tau\}$ gives another one. However, note that in the latter case x_1' is *less* than x_0. This is no surprise: conseder a particle moving in space. There is no reason why it cannot go backwards. Remember that if x_0 corresponds to $x(t_0)$ (where t denotes time) x_1 corresponds to $x(t_1)$, then since $t_1 > t_0 \not\Rightarrow x(t_0) > x(t_1)$, we need to be cautious while interpreting such diagrams.

B Path probability

We will consider a Markov process with the system dynamics given by the overdamped Langevin equation:

$$\dot{x}(t) = F(x,t) + \xi(t), \quad \text{(B.1)}$$

with $\langle \xi(t)\xi(t')\rangle = @D\delta(t-t')$ where $D = k_B T/\gamma$. For a Markov process, all transition probabilities get multiplied to give the transition probability between the initial and the final points, because each step is an independent stochastic variable:

$$P(x_0, x_1, \cdots, x_n) = p(x_1|x_0)\; p(x_2|x_1)\; \cdots\; p(x_n|x_{n-1}). \quad \text{(B.2)}$$

Now, the claim is the following:

$$P(x_0, x_1, \cdots, x_n) = P(\xi_0, \xi_1, \cdots, \xi_n) = p(\xi_0)\; p(\xi_1)\; p(\xi_2)\; \cdots\; p(\xi_n). \quad \text{(B.3)}$$

This readily follows from (B.1) by discretizing this equation:

$$\begin{aligned} \frac{\Delta x_n}{\Delta t} =& F(x_n, n\Delta t) + \xi_n \\ \Rightarrow \quad x_n =& x_{n-1} + F_n \Delta t + \xi_n \Delta t \end{aligned} \tag{B.4}$$

where the subscripts label the time. The recurrence is given by

$$x_1 = x_0 + \epsilon F(x_1, \epsilon) + \epsilon \xi_1, \tag{B.5a}$$
$$x_2 = x_1 + \epsilon F(x_2, 2\epsilon) + \epsilon \xi_2, \tag{B.5b}$$
$$\vdots$$
$$x_N = x_{N-1} + \epsilon F(x_N, N\epsilon) + \epsilon \xi_N. \tag{B.5c}$$

Now, eqn. (B.5a) can be inverted to get x_1 as a function of (x_0, ξ_1). In short, we have $x_1(x_0, \xi_1)$. Similarl, from (B.5b) we get $x_2(x_1, \xi_2) \to x_2(x_0, \xi_1, \xi_2)$. Reiterating this process, we can get for any i, $x_i(x_0, \xi_1, \xi_2, \cdots, \xi_i)$. In compact notation, we can write for the entirte trajectory, $\{x(t)\} = x(x_0, \{\xi(t)\})$. Thus, we have got a unique mapping: $\{x(t)\} \leftrightarrow \{\xi(t)\}$ for *any* given trajectory. Thus we have got two equivalent descriptions of a trajectory, and the probability of a given trajectory can be written as

$$P[x(t)|x_0]\ dx_0 dx_1 \cdots dx_N = P[\xi(t)]\ d\xi_0 d\xi_1 \cdots \xi_N. \tag{B.6}$$

Now, we usually prefer to work with the noise variable for the following reason. We can always write the probability of a trajectory as

$$P[\xi(t)] = P(\xi_1, \xi_2, \cdots, \xi_N) = p(\xi_1)\ p(\xi_2)\ \cdots\ p(\xi_N). \tag{B.7}$$

Obviously, the state variables do not enjoy the above luxury. Even if the process is assumed to be *Markovian*, when we can write

$$P[x(t)|x_0] = p(x_N|x_{N-1})\ p(x_{N-1}|p_{N-2})\ \cdots\ p(x_1|x_0), \tag{B.8}$$

even then the transition probabilities are hard to calculate and depend on the form of the external force as well as on the specific discretization scheme adopted (see, for example, Risken, sec. 4.4.2). Weiner, however, had found an elegant way to circumvent this problem. We discuss this in the next subsection.

C The Weiner integral

Let us write the Langevin equation (B.1) in discrete form [18, 19]:

$$\frac{x_j - x_{j-1}}{\varepsilon} = F_j + \xi_j, \tag{C.1}$$

where for any general discretization scheme, we write

$$F_j = \alpha F(x_j) + (1 - \alpha) F(x_{j-1}).$$

Also, we define a variable

$$y(t) \equiv \int_0^t dt'\ \xi(t'),$$

so that $\dot{y} = \xi(t)$. This ensures that we are dealing with a well-behaved variable rather than $\xi(t)$ which is highly fluctuating. There is an interesting property of y: it is *by definition* the same as x (other than an additive constant) *in the absence of any external drive*. This means that although the presence of external

drive changes the transition probability of x, it has no effect on y. Its transition probability still remains the solution of the simple diffusion equation, and is given by

$$p(y_2|y_1) = \frac{1}{\sqrt{4\pi D(t_2-t_1)}} \exp\left[-\frac{(y_2-y_1)^2}{4D(t_2-t_1)}\right]. \tag{C.2}$$

Thus, the path probability becomes

$$\begin{aligned} P[y(t)|y_0] =& p(y_N|y_{N-1})\; p(y_{N-1}|p_{N-2})\; \cdots\; p(y_1|y_0) \\ =& \frac{1}{(4\pi D\varepsilon)^{N/2}} \prod_{i=1}^{N} \exp\left[-\frac{(y_i-y_{i-1})^2}{4D\varepsilon}\right] \\ =& \frac{1}{(4\pi D\varepsilon)^{N/2}} \exp\left[-\sum_{i=1}^{N} \frac{(y_i-y_{i-1})^2}{4D\varepsilon^2}\varepsilon\right]. \end{aligned} \tag{C.3}$$

In the limit $\varepsilon \to 0$, we have

$$\begin{aligned} P[y(t)|y_0] =& \frac{1}{(4\pi D\; dt)^{N/2}} \exp\left[-\int_0^\tau dt\; \frac{\dot{y}^2(t)}{4D}\right] \\ =& \frac{1}{(4\pi D\; dt)^{N/2}} \exp\left[-\int_0^\tau dt\; \frac{\xi^2(t)}{4D}\right]. \end{aligned} \tag{C.4}$$

Note that if we use the original Langevin equation, $\gamma\dot{x} = F(x,t) + \xi(t)$ then we need to replace $D = k_BT/\gamma$ by $D = k_BT\gamma$ ($\because \xi \to \xi/\gamma$).

However, while employing the above trick, we also have to play a price: we need to find the *Jacobian*, $J \equiv \left|\frac{\partial y}{\partial x}\right|$, in order to integrate over the state variables $\{x(t)\}$ (or else write all the path functionals in terms of y). Then we rewrite the above Langevin equation as

$$\begin{aligned} y(t) =& x(t) - x(0) - \int_0^t F(x,t')dt' \\ \Rightarrow\quad y_j =& x_j - x_0 - \varepsilon \sum_{i=1}^{j} \{\alpha F(x_i) + (1-\alpha)F(x_{i-1})\} \\ \Rightarrow\quad y_j =& x_j - x_0 - \varepsilon\{[\alpha F(x_1) + (1-\alpha)F(x_0)] + [\alpha F(x_2) + (1-\alpha)F(x_1)] \\ & + \cdots + [\alpha F(x_j) + (1-\alpha)F(x_{j-1})]\} \\ =& x_j - x_0 - \varepsilon\{(1-\alpha)F(x_0) + \alpha F(x_j)\} - \varepsilon\{F(x_1) + \cdots + F(x_{j-1})\} \end{aligned} \tag{C.5}$$

Thus,

$$\frac{\partial y_j}{\partial x_j} = 1 - \varepsilon\alpha\frac{\partial F(x_j)}{\partial x_j}; \qquad \frac{\partial y_j}{\partial x_k} = -\varepsilon\frac{\partial F(x_k)}{\partial x_k} \qquad (0 < k < j). \tag{C.6}$$

We then find that the Jacobian is given by

$$J \equiv \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_N} \\ \vdots & & \ddots & \\ \frac{\partial y_N}{\partial x_1} & \frac{\partial y_N}{\partial x_2} & \cdots & \frac{\partial y_N}{\partial x_N} \end{vmatrix} = \begin{vmatrix} 1-\varepsilon\alpha F'(x_1) & 0 & \cdots & 0 \\ -\varepsilon F'(x_2) & 1-\varepsilon\alpha F'(x_2) & \cdots & 0 \\ \vdots & & \ddots & \\ -\varepsilon F'(x_N) & -\varepsilon F'(x_2) & \cdots & 1-\varepsilon\alpha F'(x_N) \end{vmatrix}, \tag{C.7}$$

whence we find that

$$J = \prod_{j=1}^{N} \{1 - \varepsilon\alpha F'(x_j)\}. \tag{C.8}$$

For $\varepsilon \to 0$, we get

$$J = \prod_{j=1}^{N} \exp[-\varepsilon\alpha F'(x_j)] = \exp\left[-\varepsilon\alpha \sum_{j=1}^{N} F'(x_j)\right] \quad \longrightarrow \quad \exp\left[-\alpha \int_0^\tau dt\ F'(x,t)\right]. \tag{C.9}$$

Now, combining (C.4) and (C.9) we get

$$\boxed{P[x(t)|x_0] = \exp\left[-\int_0^\tau \left\{\frac{\dot{y}^2(t)}{4D} + \alpha F'(x,t)\right\}\right]}. \tag{C.10}$$

In Stratonovich discretization, $\alpha = \frac{1}{2}$ while in Ito discretization, $\alpha = 1$. If the original Langevin equation (1.3) is used then, then $F \to F/\gamma$ in the Jacobian, while $D \to D\gamma^2 = k_B T\gamma$.

D The easily solvable Fokker Planck equation

As defined by Jarzynski [21], we define the following variable (we denote by $x(t)$ the full trajectory and by x_t the phase space point at time t):

$$g(x,t) \equiv \langle \delta(x - x_t) e^{-\beta W(t)} \rangle. \tag{D.1}$$

which means that the average of $e^{-\beta W(t)}$ has been taken with respect to the trajectories that pass through x at time t. Before embarking upon the FPE, it is helpful to know that a direct solution can be obtained via the Crooks' theorem. We proceed as follows:

$$\begin{aligned}
g(x,\tau) &\equiv \langle \delta(x - x_\tau) e^{-\beta W(\tau)} \rangle \\
&= \int dx_0 \cdots dx_\tau P[x(t)] \delta(x - x_\tau) e^{-\beta W(\tau)} \\
&= \int dx_0 \cdots dx_\tau P[x(t)|x_0] p(x_0) \delta(x - x_\tau) e^{-\beta W(\tau)} \\
&= \int dx_0 \cdots dx_\tau P[x(t)|x_0] \frac{e^{-\beta H(0)}}{Z_0} e^{-\beta W(\tau)} \delta(x - x_\tau) \\
&= \frac{1}{Z_0} \int dx_0 \cdots dx_\tau P[x(t)|x_0] e^{-\beta(H(\tau) - \Delta H(\tau))} e^{-\beta W(\tau)} \delta(x - x_\tau) \\
&= \frac{e^{-\beta H(\tau)}}{Z_0} \int dx_0 \cdots dx_\tau P[x(t)|x_0] e^{-\beta Q} \delta(x - x_\tau) \qquad \text{step (*)} \\
&= \frac{e^{-\beta H(\tau)}}{Z_0} \int d\tilde{x}_\tau \cdots d\tilde{x}_0 \tilde{P}[\tilde{x}(t)|\tilde{x}_0] \delta(\tilde{x} - \tilde{x}_0) \qquad \text{step (**)} \\
&= \frac{e^{-\beta H(\tau)}}{Z_0} \int d\tilde{x}_1 \cdots d\tilde{x}_\tau \tilde{P}[\tilde{x}(t)|\tilde{x}_0] \int d\tilde{x}_0 \delta(\tilde{x} - \tilde{x}_0) \\
\Rightarrow \quad & \boxed{g(x,\tau) = \frac{e^{-\beta H(\tau)}}{Z_0}.}
\end{aligned} \tag{D.2}$$

In step (*), we have used the first law, $Q = W - \Delta H$, while in step (**), we have used the Crooks' relation, $P[x(t)|x_0]e^{-\beta Q} = \tilde{P}[\tilde{x}(t)|\tilde{x}_0]$.

Now let us first see what happens for a Pure Hamiltonian evolution. We first write,

$$g(x,\tau) = \int dx_0 \cdots dx_\tau P[x(t)|x_\tau] p(x_\tau) \delta(x - x_\tau) e^{-\beta W(\tau)}$$

$$= p(x,\tau) \int dx_0 \cdots dx_{\tau-1} P[x(t)|x] e^{-\beta W[x(t)]}$$
$$= p(x,\tau) \langle e^{-\beta W} \rangle_{x,\tau}, \tag{D.3}$$

where the average is done with respect to the trajectories that pass through the point x at time τ.

Now, if we apply the $\frac{\partial}{\partial \tau}$ operator to g, then it will affect the two functions $p(x,\tau)$ and $W[x(t)]$, because $P[x(t)|x]$ inside the integral is independent of tau. Since we have $\dot{p}(x,t) = 0$, we get

$$\begin{aligned}
\frac{\partial g}{\partial \tau} &= \frac{\partial p(x,\tau)}{\partial \tau} \langle e^{-\beta W} \rangle_{x,\tau} - \beta p(x,\tau) \left\langle \frac{\partial W}{\partial \tau} e^{-\beta W} \right\rangle_{x,\tau} \\
&= \frac{\partial p(x,\tau)}{\partial \tau} \langle e^{-\beta W} \rangle_{x,\tau} - \beta p(x,t) \dot{W}(x,\tau) \left\langle e^{-\beta W} \right\rangle_{x,\tau} \\
&\Rightarrow \boxed{\frac{\partial g}{\partial \tau} = \left[\hat{R} - \beta \dot{W}(x,\tau) \right] g,}
\end{aligned} \tag{D.4}$$

where

$$\begin{aligned}
\hat{R}\, g &\equiv \{H, p\}_{x,p}\, g \qquad \text{(Hamiltonian evolution)} \\
&\equiv \frac{\partial p}{\partial \tau}\, g \qquad \text{(Stochastic evolution)}
\end{aligned} \tag{D.5}$$

Also, $\dot{W}(x,\tau)$ is a constant, owing to the fact that (x,τ) are fixed, so that it can be taken out of the average.

Of course, since in the derivation of (D.2) we have not assumed any particular dynamics, this solution will be valid for both Hamiltonian and stochastic evolutions.

E The time-reversal operator

The time-reversal operator is *anti*unitary, which means that is a unitary and antilinear operator. Antilinearity demands that

$$\Theta \lambda |\psi(t)\rangle = \lambda^* \Theta |\psi(t)\rangle. \tag{E.1}$$

The antiunitarity is required to avoid the unphysical situations of having energy state vectors having negative eigenvalues. Now, we know that the forward time evolution of a quantum state is given by

$$|\psi(t)\rangle = e^{-iHt/\hbar} |\psi(0)\rangle. \tag{E.2}$$

By definition of time reversal, if we begin with a time-reversed state $|\psi(t)\rangle$ and evolve it *forward* in time, then it must finally reach the state that is time reversed of $|\psi(0)\rangle$, namely, $\Theta|\psi(0)\rangle$:

$$\Theta |\psi(0)\rangle = e^{-iHt/\hbar} \Theta |\psi(t)\rangle. \tag{E.3}$$

This implies

$$\Theta |\psi(0)\rangle = e^{-iHt/\hbar} \Theta\, e^{-iHt/\hbar} |\psi(0)\rangle \quad \Rightarrow \quad e^{+iHt/\hbar} \Theta = \Theta\, e^{-iHt/\hbar} \tag{E.4}$$

Taking infinitesimal evolution for time interval δt, we get

$$\left(1 + \frac{iH\delta t}{\hbar}\right) \Theta = \Theta \left(1 - \frac{iH\delta t}{\hbar}\right) \quad \Rightarrow \quad iH\Theta = -\Theta iH. \tag{E.5}$$

Using the definition (E.1) for anilinearity , we arrive at

$$\boxed{\Theta H = H \Theta.} \tag{E.6}$$

Next, let us define two state vectors in the basis $\{|\phi_i\rangle\}$:

$$|\psi_1\rangle \equiv \sum_i \lambda_i |\phi_i\rangle; \qquad |\psi_2\rangle \equiv \sum_i \gamma_i |\phi_i\rangle. \tag{E.7}$$

$$\begin{aligned} \therefore |\Theta\psi_1\rangle \equiv \Theta|\psi_1\rangle &= \sum_i \lambda_i^* |\Theta\phi_i\rangle; \\ |\Theta\psi_2\rangle \equiv \Theta|\psi_2\rangle &= \sum_i \gamma_i^* |\Theta\phi_i\rangle. \end{aligned} \tag{E.8}$$

Thus,

$$\begin{aligned} \Rightarrow \langle\Theta\psi_2|\Theta\psi_1\rangle &= \sum_i \langle\Theta\psi_2|\Theta\phi_i\rangle\langle\Theta\phi_i|\Theta\psi_1\rangle \\ &= \lambda_i^*\gamma_i = \langle\psi_1|\psi_2\rangle = \langle\psi_2|\psi_1\rangle^*. \end{aligned} \tag{E.9}$$

This is another property that is unique to antiunitary operators (for a unitary uperator U, we will have $\langle U\psi_2|U\psi_1\rangle = \langle\psi_2|\psi_1\rangle$).

References

[1] K. Sekimoto, Prog. Theor. Phys. Supp. **130**, 17 (1998).

[2] M. L. Bellac, *Non equilibrium statistical mechanics*, Preliminary Version, website: *http://cel.archives-ouvertes.fr/cel-00176063/en/.*

[3] B. Cleuren, C. Van den Broeck, and R. Kawai, Phys. Rev. Lett. **96**, 050601 (2006).

[4] G. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[5] Udo Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[6] David Chandler, *Introduction to Modern Statistical Mechanics,* Oxford.

[7] T. Hatano and S. Sasa, Phys. Rev. Lett. **86**, 3463 (2001).

[8] S. Ito and M. Sano, Phys. Rev. E **84**, 021123 (2011).

[9] Y. Oono and M. Paniconi, Prog. Theor. Phys. Suppl. **130**, 17 (1998).

[10] David Andrieux and Pierre Gaspard, Phys. Rev. Lett. **100**, 230404 (2008)

[11] G. E. Crooks, Phys. Rev. E **60**, 2721 - 2726 (1999).

[12] P. Talkner, E. Lutz and P. Hanggi, Phys. Rev. E **75**, 050102 (2007).

[13] P. Talkner and P.Hanggi, J. Phys. A: Math. Theo. **40**, F569-F571 (2007).

[14] Jordan M. Horowitz, Christopher Jarzynski, Phys. Rev. E **79**, 021106 (2009).

[15] T. M. Cover and J. A .Thomas, *Elements of information theory*, Wiley interscience.

[16] James P. Sethna, *Entropy, Order Parameters, and Complexity*, Clarendon Press - Oxford.

[17] Tulsi Dass and S. K. Sharma, *Mathematical Methods In Classical And Quantum Physics*, Universities Press.

[18] H. S. Wio, *An Introduction to Stochastic Processes and Nonequilibrium Statistical Physics.*

[19] M. chaichian, A. Tureanu and A. Zahabi, Phys. Rev. E **81**, 066309 (2010).

[20] M. Campisi, P. Talkner and P. Hanggi, Phys. Rev. Lett. **102**, 210401 (2009).

[21] C. Jarzynski, Phys. Rev. Lett. **56**, 5018 (1997).

Index