A Fluctuation Relation for a general feedback protocol

Alberto Imparato

We consider a setup where a demon observes a system with microsopic coordinates x (for example $x = (q, p)$ for a 1D underdamped Brownian particle) in a time interval $0 \le t \le t_\mathrm{f}$, and can change some parameters in the Hamiltonian, if a given event or a set of events occurs.

We assume that the system is initially at equilibrium with a probability distribution $p(x, t = 0) = p^{eq}(x, \lambda_0)$. We also choose as initial probability distribution for the *backward* trajectories $p(x^\dagger, t^\dagger = 0) = p^{eq}(x^\dagger, \lambda_f)$. In no way we are assuming that the system is at equilibrium at the end of the forward trajectory, on the contrary the system can be arbitrarily far from equilibrium at final time, see discussion in section 4.6 in PP. Thus, in absence of feedback control, and using an arbitrary protocol $\boldsymbol{\lambda}(t)$, the following fluctuation relation for the work holds

$$\frac{P_{no}[\boldsymbol{x}|\boldsymbol{\lambda}]}{P^R_{no}[\boldsymbol{x}^\dagger|\boldsymbol{\lambda}^\dagger]} = \mathrm{e}^{\beta(w[\boldsymbol{x},\boldsymbol{\lambda}]-\Delta F)}, \tag{1}$$

with $\Delta F = -k_B T\,(\ln Z(\lambda_f) - \ln Z(\lambda_0))$.

A bit of notation: in the following $\boldsymbol{x}(t)$ is a trajectory in the system phase space, $\boldsymbol{\lambda}(t)$ is a trajectory in the parameter phase space (aka the protocol), $\lambda_{FB}[\boldsymbol{x}(t)]$ is the specific protocol generated by the demon while observing the trajectory $\boldsymbol{x}(t)$. Strictly speaking $\lambda_{FB}[\boldsymbol{x}(t)]$ is an operator mapping functions to functions $\bar{\boldsymbol{\lambda}}(t) = \lambda_{FB}[\boldsymbol{x}(t)]$ according to the given demon rules. I will use the bar to indicate that the function $\bar{\boldsymbol{\lambda}}(t)$ is one of the possible outcomes for the protocol, i.e., there is at least one trajectory $\boldsymbol{x}(t)$ that generates the protocol $\bar{\boldsymbol{\lambda}}(t)$.

Notice that the work is a functional of both the trajectory $\boldsymbol{x}(t)$ and the protocol $\boldsymbol{\lambda}(t)$: $w[\boldsymbol{x}(t), \boldsymbol{\lambda}(t)]$.

Given the feedback rules, of all the possible functions $\boldsymbol{\lambda}(t)$, $0 \le t \le t_\mathrm{f}$, only a subset can be generated by the demon while observing all the possible trajectories $\boldsymbol{x}$. Let

$$\Omega = \left\{\bar{\boldsymbol{\lambda}}(t) |\, \exists\, \boldsymbol{x}(t) : \bar{\boldsymbol{\lambda}}(t) = \Lambda_{FB}[\boldsymbol{x}(t)]\right\} \tag{2}$$

be the set of all possible protocols $\bar{\boldsymbol{\lambda}}$ that can be generated by the demon, i.e. the set of all the functions $\boldsymbol{\lambda}(t)$ for which there exists at least one microscopic trajectories resulting in that output protocol.

In the following we drop the explicit dependence on time of the trajectory and the protocol in order to lighten the notation. We introduce the ensemble of trajectories generating the same protocol $\Omega_{\bar{\boldsymbol{\lambda}}} = \left\{\boldsymbol{x} : \bar{\boldsymbol{\lambda}} = \Lambda_{FB}[\boldsymbol{x}]\right\}$. This is useful as different trajectories in the phase space can generate the same protocol. This also emphasizes that if the experimentally accessible outcome is the protocol rather than the microscopic trajectory, the information that can be gathered (or not) must pertain the protocol rather than the single trajectory $\boldsymbol{x}$.

We now introduce the joint probability $P_{FB}[\boldsymbol{x}, \boldsymbol{\lambda}]$ which is non-vanishing only if $\boldsymbol{\lambda}$ is a protocol associated to at least one trajectory $\boldsymbol{x}$: $\boldsymbol{\lambda} \in \Omega$, and (trivially) if $\boldsymbol{x}$ generates the specific $\boldsymbol{\lambda}$, thus

$$P_{FB}[\boldsymbol{x}, \boldsymbol{\lambda}] \neq 0 \qquad \text{if } \boldsymbol{\lambda} \in \Omega \text{ and } \boldsymbol{x} \in \Omega_{\bar{\boldsymbol{\lambda}}}. \tag{3}$$

Given that $\bar{\boldsymbol{\lambda}}$ is one of the possible output protocols with FB, we also notice that

$$P_{FB}[\boldsymbol{x}, \bar{\boldsymbol{\lambda}}] = P_{no}[\boldsymbol{x}|\bar{\boldsymbol{\lambda}}] \qquad \text{if } \boldsymbol{x} \in \Omega_{\bar{\boldsymbol{\lambda}}}, \tag{4}$$

and thus the probability of observing a given protocol with feedback is

$$P_{FB}[\bar{\boldsymbol{\lambda}}] = \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{FB}[\boldsymbol{x}, \bar{\boldsymbol{\lambda}}] = \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{no}[\boldsymbol{x}|\bar{\boldsymbol{\lambda}}] \equiv Q[\bar{\boldsymbol{\lambda}}]. \tag{5}$$

Fixing $\bar{\boldsymbol{\lambda}}$ and without using the FB control, $Q[\bar{\boldsymbol{\lambda}}]$ is the cumulative probability of all the trajectories in $\Omega_{\bar{\boldsymbol{\lambda}}}$ out of all the possible trajectories. It is also the cumulative probability of all those trajectories that would result in a protocol $\bar{\boldsymbol{\lambda}}$ when the demon is active.

Similarly we define

$$Q^R[\bar{\boldsymbol{\lambda}}^\dagger] = \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P^R_{no}[\boldsymbol{x}^\dagger|\bar{\boldsymbol{\lambda}}^\dagger] = \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{no}[\boldsymbol{x}|\bar{\boldsymbol{\lambda}}]\mathrm{e}^{-\beta(w[\boldsymbol{x},\bar{\boldsymbol{\lambda}}]-\Delta F)}, \tag{6}$$

which is the cumulative probability of observing the reverse trajectories belonging to $\Omega_{\bar{\boldsymbol{\lambda}}}$ with the time-reversed protocol $\bar{\boldsymbol{\lambda}}^\dagger$ and without FB. We notice that the identification between P_{FB} and P_{no}, eq. (5) only holds for the forward trajectories with the forward protocol.

We continue by writing the exponential average, constrained by a given output protocol

$$\left\langle \mathrm{e}^{-\beta(w-\Delta F)} \right\rangle_{FB,\bar{\boldsymbol{\lambda}}} = \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{FB}[\boldsymbol{x}|\bar{\boldsymbol{\lambda}}]\mathrm{e}^{-\beta(w[\boldsymbol{x},\bar{\boldsymbol{\lambda}}]-\Delta F)} = \frac{1}{P_{FB}[\bar{\boldsymbol{\lambda}}]} \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{FB}[\boldsymbol{x},\bar{\boldsymbol{\lambda}}]\mathrm{e}^{-\beta(w[\boldsymbol{x},\bar{\boldsymbol{\lambda}}]-\Delta F)}$$
$$= \frac{1}{P_{FB}[\bar{\boldsymbol{\lambda}}]} \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P^R_{no}[\boldsymbol{x}^\dagger|\bar{\boldsymbol{\lambda}}^\dagger] = \frac{Q^R[\bar{\boldsymbol{\lambda}}^\dagger]}{Q[\bar{\boldsymbol{\lambda}}]} = \mathrm{e}^{\Delta\mathbf{I}[\bar{\boldsymbol{\lambda}}]} \tag{7}$$

which defines the information $\Delta\mathbf{I}[\bar{\boldsymbol{\lambda}}] = -\ln\left(Q[\bar{\boldsymbol{\lambda}}]/Q^R[\bar{\boldsymbol{\lambda}}^\dagger]\right)$ which is a functional of $\bar{\boldsymbol{\lambda}}$ and of its time reverse.

We finally have the Sagawa-Ueda fluctuation theorem by summing over all the possible protocols

$$\left\langle \mathrm{e}^{-\beta(w-\Delta F+k_BT\Delta\mathbf{I})} \right\rangle_{FB} = \sum_{\bar{\boldsymbol{\lambda}}\in\Omega} P_{FB}[\bar{\boldsymbol{\lambda}}] \int_{\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}} \mathcal{D}\boldsymbol{x}\; P_{FB}[\boldsymbol{x}|\bar{\boldsymbol{\lambda}}]\mathrm{e}^{-\beta(w-\Delta F+k_BT\Delta\mathbf{I}[\bar{\boldsymbol{\lambda}}])} = 1. \tag{8}$$

We see that, by defining the information in terms of the cumulative probability of the trajectories that results in a single protocol (trajectories $\boldsymbol{x}\in\Omega_{\bar{\boldsymbol{\lambda}}}$), i.e. the probability of observing a given protocol with FB, eq. (5), we obtain a constrained FT, holding for the specific occurrence of the protocol, eq. (7)