

Lectures on Fluctuation Relations

Ian Ford, University College London, UK

1 Preliminaries

Prediction is very difficult, especially about the future - Niels Bohr

Ours is a harsh and unforgiving universe, and not just in the little matters that seem to conspire against us, like the weather on our holidays and the lateness of trains. What I mean is that the fundamental rules of the evolution of the world seem to be unfairly biased against those who seek to predict the future. Do not believe claims that the laws of physics are simple and beautiful in their entirety! Elegant symmetries in the fundamental building blocks are all very well, but the rules of the game are fiendishly complex if we try to work out what will happen next, given what we know about the present. And that question is central to any claims that we might understand the world, and thereby control events around us.

If the rules of change were simple, of course, then quite possibly there would be no complex universe of which to make sense. Games with simple rules, and easily predictable outcomes, like noughts and crosses (tic-tac-toe), are of limited interest and potential. Chess is famously more rich in possible developments, but even greater richness of outcome is to be found from a machine consisting of many intricately connected moving parts, turned by a handle perhaps, transforming one state of affairs into another. The physical world on most length and time scales works like this; its richness is evident, the existence within it of sentient, self-reproducing creatures such as ourselves being one of the more astonishing examples.

In such an interesting universe we are obliged to try to engage with and perhaps control events around us, and to do so we need to develop ideas about how the world evolves. But each moving part in a system of a high degree of complexity interacts with many others, particularly as time goes on, and in ways that are very sensitive to details. That is the harsh and unforgiving nature I mentioned earlier. In order to predict the future, it is necessary to take into account all the disparate interactions between the components of a system because they might all be, or later become, crucial to its gross evolution. Technically, this is due to so-called non-linearity in the equations of motion describing the system. A second difficulty is that it is virtually impossible to gather sufficient knowledge about the present state of a system to enable an exact prediction about the future to be made. A good guess about the position of all the moving parts of a complex machine before the handle starts turning is not enough: the sensitivity to the details makes the system unforgiving: very soon the predictions will go awry. A good example is the workings of a weather system. We would like to predict whether it will rain in central London next Saturday, but in order to do so with confidence the meteorological situation on the previous Sunday throughout western Europe would need to be measured in considerable detail. The guesses and approximations that weather services need to make in order to fill in the gaps lead inescapably to predictions that eventually go wrong. We find that the weather forecaster talks of a '30% chance of showers' or similar. What a shame it is that we cannot do better!

The chance of rain just mentioned sums up our position in predicting the future of all but the most simple systems like the pendulum of a clock or the orbits of the planets. We must accept limitations and must be prepared to embrace the uncertain future that arises from uncertainty in the present. We must expect that uncertainty will be an element in our description of the future development of a system. If we are lucky, or attentive to detail in constructing a mathematical model of the system, this will grow only slowly. If we are unlucky, or sloppy in observing the present state of the world, it will grow quickly. We might expect there to be a state of general ignorance towards which the predictions of our model might ultimately tend, whereby we cannot be sure about anything: rain, snow, heatwave or hurricane. We expect fluctuations in outcomes. In complex systems, future events will eventually be obscured behind a haze of uncertainty, and we are left to guess what will happen.

These lectures are about a set of so-called *fluctuation relations* that offer some insight into the range of the predicted behaviour of complex systems. They can quantify the unexpected outcome in terms of the expected, and they offer an interpretation of the second law of thermodynamics and the peculiar concept of entropy, which we discuss in the next section. The universe is a harsh and unforgiving place, but there are ways to make some sense of it, and the fluctuation relations are such tools that promise a better understanding of entropy, uncertainty, and the limits of predictability.

2 Entropy and the second law

Ignorance and uncertainty has never been an unusual state of affairs in human perception, and the desire for clarity has been a constant feature. In science, Newton's laws of motion provided tools that seemed to dispel some of the haze. Here were mathematical models that enabled the future to be foretold! Never mind the problems with determinism and the role of human free will that ensued in some quarters: for practical matters such as predicting the paths of celestial bodies, Newton's work and that which followed was a considerable step forward. It inspired thinking about predicting future behaviour in other fields, such as in economics, and in particular it led to developments in thermodynamics. Proponents in these fields wished to establish the rules of the game, in mathematical form if possible, and thereby to understand and control events.

Our field is thermodynamics, the study of the behaviour of systems through which matter and energy can flow. The particular focus in the early days of the subject was the heat engine, a device whereby fuel, and the heat it can generate, can be converted into mechanical work, or motive power. In the nineteenth century the steam engine, an example of such a device, was the basis of leading edge technologies, and an understanding of its operation was developed. It is fascinating that such practical matters required consideration of fundamental questions and that the principles worked out then are still relevant today, and to systems well beyond the steam engine.

The operation of an engine was discovered to produce a quantity called entropy, and that it characterised the efficiency with which energy in the fuel could be converted into motion. Indeed entropy seemed to be generated whenever heat or matter flowed. The second law of thermodynamics (one of several established in a clear attempt to emulate Newton's laws of motion) states that the total entropy of the evolving universe is always increasing. The first law of thermodynamics concerns the conservation of energy and there are a couple others that need not concern us. It is the second law that has attracted the most discussion, even now, over 150 years since its introduction. We do not debate the meaning of Newton's second law. Why is the second law of thermodynamics so controversial?

One reason is that it is hard to understand how there can be a physical quantity that never decreases. Such a statement demands the breakage of the principle of time reversal symmetry. This is a slightly obscure requirement: how can time be reversed such that we can appreciate such a symmetry? What we mean is that Newton's fundamental equations of motion do not specify a preferred direction in which time evolves. Time is a coordinate in a description of the world and it is just convention that real world events take place in the direction where this coordinate increases. Given that we cannot actually run time backwards, we can demonstrate this symmetry in the following way. A sequence of events that takes place according to time reversal symmetric equations can be inverted by instantaneously reversing all the velocities of all the participating components and then proceeding forward in time once again. If the moon's velocity were suddenly reversed, it would retrace its orbit around the Earth but in the opposite direction, if all other influences except the Earth's gravity were ignored. If all the velocities and rotations of all major objects in the solar system were reversed, the inversion of the sequence of events would proceed to an even better approximation, maybe for millennia. And if every particle in the universe had its velocity reversed, from a perspective of classical mechanics we could imagine events proceeding that took us back to the Big Bang, or at least until quantum mechanical effects caused a deviation. And furthermore, none of it would appear to be physically strange: it would be business as normal according to the underlying Newtonian dynamics.

Reversing the velocities is an unfeasible task, but the point is that any evolution can be imagined in reverse. We therefore don't expect to observe any quantity that only ever increases with time, a point made first by Loschmidt. In principle we ought to be able to initiate a steam engine such that it sucks exhaust gases back into its furnace and combines them into coal, thereby consuming energy from the slowing down of the locomotive. But we have never seen such an event, and the denial of such a spectacle is taken as evidence of the operation of the second law of thermodynamics. The reconstruction of coal at the expense of macroscopic motion would require the reversal of the generation of entropy that characterises the forward process, since it would in every respect be an inversion of the normal sequence of events. Such a destruction of entropy is disallowed by the second law and that is why the process is impossible. Time reversal symmetry is broken by the apparent illegality of certain processes, and that seems odd.

A second reason is that entropy is a rather ethereal concept. In thermodynamics it is defined in terms of differences in energy flows that occur during the operation of a engine, for example, and the amount generated can depend on how quickly the process is conducted: the faster it goes, the more entropy is produced. This is far away from the more familiar and solid ideas about physical objects and events. And the second law simply states that entropy is generated, not how much is generated. A law that only specifies the sign of a change sounds rather incomplete.

A third reason is that the debate about the nature of entropy has persisted for so long, and has brought

about such widely differing views, that positions have become entrenched. Resolution is thereby made more difficult.

But what has emerged in the last 15 years or so is the realisation that Newton's laws of motion, when supplemented by the acceptance of uncertainty in the way models are built and calculations are initiated and developed, can lead quite naturally to a quantity that grows with time, namely uncertainty itself. This is in the nature of the harsh and unforgiving laws of the universe described in the previous section. Any model of the evolution of a physical system that recognises its own incompleteness will generate additional uncertainty as it is evolved. And if the velocities are all instantaneously reversed, in the hope that the sequence of events might thereafter be reversed, uncertainty will continue to be generated. Newton's laws are time-reversal symmetric, but the inclusion and evolution of uncertainty is not. Such a basis for predicting the future will naturally break time reversal symmetry. Entropy generation with time might therefore be just the leakage of our confidence in the predictions of the model that we might employ. It is a consequence of the incompleteness of the model and in the initial conditions that are assumed.

Immediately objections to this can be raised. Entropy is uncertainty? Then do the energy flows I measure when I operate my machine depend on how well I measure the positions of its parts before I start? How can sloppy initial measurements affect the efficiency of my machine? How might better measurements improve the efficiency? The machine would perform in the same way even if you had managed to list in detail the positions of all its cogs and pistons to begin with, and I had neglected to do so. How would my estimate of entropy change differ from yours?¹ And how can entropy arise from a mere model? It is a function of heat flows. It is a real quantity that can be measured with precision instruments, not an mental illusion that depends on which model is employed to describe the process. This is utter nonsense! Or is it so? These objections need to be addressed if the connection between entropy and uncertainty, or to use a related concept, information, is to be understood.

3 Basic thermodynamics

Before we proceed further, let us remind ourselves of the basic rules of thermodynamic processes. A system possesses energy E and can receive additional incremental contributions in the form of heat dQ from a so-called heat reservoir or bath at temperature T , and work dW from an external mechanical device that might squeeze or stretch the system. Immediately we might ponder what exactly is the difference between heat and work: it helps perhaps to view them roughly as an increment in kinetic and in potential energy, respectively.

We write the first law of thermodynamics (energy conservation) in the form

$$dE = dQ + dW, \quad (1)$$

and the second law as

$$\oint \frac{dQ}{T} \leq 0, \quad (2)$$

which requires further explanation. The integration symbol means that the system is taken around a cycle of heat and work transfers, ending with the same macroscopic system parameters, such as temperature and volume, that it started with. The temperature $T(t)$ of the heat bath, and hence the system, might change around the cycle, and so might the system volume and shape. However, we often write the second law in a more powerful form:

$$dS_i = dS + dS_r, \quad (3)$$

where each term is an increment in a thermodynamic property called entropy during an incremental segment of the cycle. The change in system entropy is denoted dS , and dS_r denotes the entropy gained by the heat reservoir, defined as

$$dS_r = -\frac{dQ}{T}, \quad (4)$$

and dS_i is the total entropy change of the two combined (the 'universe'). We see that equation (2) corresponds to the condition $\oint dS_i \geq 0$, since $\oint dS = 0$, but the more powerful reading of the second law is that

$$dS_i \geq 0, \quad (5)$$

for any segment of a thermodynamic process. We might write $dS = dS_i - dS_r$ and regard dS_i as a contribution to the change in entropy of a system brought about by a change in reservoir temperature or by the exertion

¹Hint at a resolution: the prior measurements needed to affect the outcome of a thermodynamic process go well beyond the position of the cogs, and would concern atomic detail.

of external mechanical force, but which is not associated with a flow of entropy from the heat bath, which is the dQ/T term. For a thermally isolated system, where $dQ = 0$, equation (3) tells us that $dS = dS_i$ and the second law says that the system entropy must increase during any adiabatic process that transforms a system from one equilibrium state to another.

Let us now consider a familiar process of heat transfer through a system that is in contact with two heat reservoirs, one at temperature T_1 and the other at T_2 . In a steady state, we presume that the system entropy remains constant, even though heat dQ is flowing through the system, per unit time say, from reservoir 1 to reservoir 2. The entropy lost by reservoir 1 is dQ/T_1 and the amount gained by reservoir 2 is dQ/T_2 . The overall change in entropy of system and both reservoirs is $dQ(1/T_2 - 1/T_1)$. Now, we expect dQ to take the same sign as $T_1 - T_2$, based on common experience about the flow of heat from hot to cold bodies. Without damaging the argument, we might assume it is proportional to that temperature difference. Hence the change in overall entropy is proportional to $(T_1 - T_2)(1/T_2 - 1/T_1)$ which is positive, irrespective of the values of the two temperatures. The total entropy increases.

Boltzmann tried to explain what this ever-increasing quantity might represent at a microscopic level. He considered a thermally isolated gas of particles interacting through pairwise collisions. The quantity

$$H(t) = \int f(\mathbf{v}, t) \ln f(\mathbf{v}, t) d\mathbf{v}, \quad (6)$$

where $f(\mathbf{v}, t)d\mathbf{v}$ is the population of particles with a velocity within the range $d\mathbf{v}$ about \mathbf{v} , can be proved to decrease with time, or remain constant if the population is in a Maxwell-Boltzmann distribution that is characteristic of equilibrium. He obtained this result using the assumption that collisions between particles at velocities \mathbf{v}_1 and \mathbf{v}_2 take place at a rate proportional to the product of populations at those velocities, namely $f(\mathbf{v}_1, t)f(\mathbf{v}_2, t)$. Thus Boltzmann claimed that H was proportional to the negative of system entropy and that his so-called H -theorem provided a microscopic justification for the second law. Unfortunately, his proposal was flawed, or rather limited to a very particular case. It was noted that Newton's laws of motion could not lead to a quantity that always increased with time. It would be incompatible with the principle of time reversal symmetry. For every entropy producing microscopic trajectory, there would be another, obtained by a reversal of velocities, that would destroy entropy. Boltzmann's H , defined in terms of microscopic velocities that evolved according to Newton's equations, could not solely decrease. So Boltzmann revised his ideas, stating that H would decrease on most but not all occasions. The H -theorem has a statistical meaning that has its origin in the assumption that the rate of collisions depends on a product of population distributions. Such a neglect of correlation in velocities of colliding particles, both in the past and in the future, is incompatible with Newtonian dynamics, and we thereby depart from a description that respects time-reversal symmetry. However, a more telling difficulty emerged when Gibbs proposed that the entropy of a system, modelled statistically as an ensemble of many systems sampled from a probability density $P(\{\mathbf{x}, \mathbf{v}\})$, where the $\{\mathbf{x}, \mathbf{v}\}$ denote all the positions and velocities of particles in a system, could be written

$$S_{\text{Gibbs}} = -k \int P(\{\mathbf{x}, \mathbf{v}\}) \ln P(\{\mathbf{x}, \mathbf{v}\}) \prod d\mathbf{x} d\mathbf{v}, \quad (7)$$

where k is Boltzmann's constant and the integration is over all possible positions and velocities of all the particles. The use of P defined over the complete so-called phase space instead of the particle velocity distribution f is a more general statistical representation of the system state, and the Gibbs representation of entropy is a very successful model, compatible with all of classical equilibrium thermodynamics. But the probability density P for an isolated system evolves according to Liouville's equation of motion, and it may be shown as a consequence that S_{Gibbs} is a *constant* of the motion. How, then, can the entropy of an isolated system increase? Either equation (7) is valid only for equilibrium situations, or something is wrong.

4 Stochastic dynamics

4.1 Master equations

Let us pursue the assertion made earlier that only when Newton's laws are supplemented by the explicit inclusion of uncertainty can sense be made of the second law. The deterministic rules of evolution of a property of a system need to be replaced by rules for the evolution of the probability that the property should take a particular value. We must first establish what we mean by *probability*, since it can cause confusion. We prefer it not to mean the limiting frequency that an event might occur amongst a large number of trials. We follow those for whom probability represents a distillation, in numerical form, of best judgement or belief about the state of a system. We regard it as a tool for the evaluation of *expectation* values of system properties, which quite literally

represent what we expect to observe based on our beliefs or information about a system. In this way, we have no need to justify the imagined infinite number of repeated trials required to establish limiting frequencies, or the imagined infinite number of copies of a system in a literal ensemble. We only want to make predictions about a single system. On the other hand, we have invoked something that seems to many people to be over-reliant on a subjective viewpoint, namely upon the modelling assumptions/beliefs of the model-builder. This comes to a particular focus when making an initial assessment of the probability distribution. But in fact the two interpretations lead to laws for the evolution of probability that are of the same form. The proposed rules for a belief or information-based interpretation must obviously be compatible with frequency-based data. Thus to a certain extent we can use either interpretation, according to our personal taste, although to remain noncommittal can perpetuate confusion.

We need to derive equations that describe the evolution of probability. Consider the case of a symmetric Markov random walk in one dimension, where a step is taken at regular time intervals, but with no preference between taking a left or right step. This means that the probability that the particle is at position x_m at timestep $n + 1$ is one half the probability that it was positioned at x_{m-1} at timestep n , plus one half the probability that it was then at x_{m+1} . In short

$$P_{n+1}(x_m) = \frac{1}{2}P_n(x_{m-1}) + \frac{1}{2}P_n(x_{m+1}), \quad (8)$$

which is a simple rule for the evolution of probability as time progresses. It is nothing more than a statement of logic: a propagation of belief (or population if you prefer) according to specific rules.

In more general cases, where there are many transitions that might bring a particle to the position x_m , we write

$$P_{n+1}(x_m) = \sum_{m'=-\infty}^{\infty} T(x_m - x_{m'}|x_{m'})P_n(x_{m'}), \quad (9)$$

where $T(\Delta x|x)$ is the transition probability for making a step of size Δx given a starting position of x . The Markov property we have invoked is that the transition probability should not depend on the previous history of the walker; only the position x prior to making the step. In principle, T could be time dependent, but we ignore this possibility here. The transition probability is normalised such that

$$\sum_{m=-\infty}^{\infty} T(x_m - x_{m'}|x_{m'}) = 1, \quad (10)$$

since the total probability that *any* transition is made, starting from $x_{m'}$, is unity.

The probability of being at position m at time $n + 1$ is a sum of probabilities of all possible previous histories that lead to this point. In the Markov case these probabilities of histories may be written as the probability $P_n(x_{m'})$ of being at position m' at time n , times the probability of making the jump from m' to m . Equation (9) is called a *master equation*.

For the random walk,

$$T(x_m - x_{m'}|x_{m'}) = \frac{1}{2}(\delta_{m-1\ m'} + \delta_{m+1\ m'}), \quad (11)$$

written in terms of the Kronecker delta δ_{ij} , which is unity if the two indices are equal, but zero otherwise. The two terms in the brackets represent, respectively, steps to the right ($m = m' + 1$) and left ($m = m' - 1$). There are only two transitions possible if one starts at $x_{m'}$. Hence, for this example,

$$P_{n+1}(x_m) = \sum_{m'=-\infty}^{\infty} \frac{1}{2}(\delta_{m-1\ m'} + \delta_{m+1\ m'})P_n(x_{m'}) = \frac{1}{2}P_n(x_{m-1}) + \frac{1}{2}P_n(x_{m+1}), \quad (12)$$

using the $\sum_j \delta_{ij}F_j = F_i$ property of the Kronecker delta. This corresponds to the earlier equation (8), as it should.

4.2 Kramers-Moyal and Fokker-Planck equations

The Kramers-Moyal and Fokker-Planck equations describe the evolution of probability distributions (more properly probability density functions) which are continuous in space (K-M) and additionally in time (F-P). We start with an integral form of the master equation for the evolution of a probability density function which is continuous in space

$$P(x, t + \tau) = \int T(\Delta x|x - \Delta x, t)P(x - \Delta x, t)d\Delta x, \quad (13)$$

which is often called the Chapman-Kolmogorov equation. We have swapped the discrete time label n for a parameter t . The Markovian transition probability $T(\Delta x|x, t)$ describes a jump from x through distance Δx in a period τ starting from time t . Note that T now has dimensions of inverse length (it is really a transition probability *density* and later we shall use the notation \mathcal{P} to denote such densities instead of T), and is normalised according to $\int T(\Delta x|x, t)d\Delta x = 1$.

We can turn this integral equation into a differential equation by Taylor expanding the integrand in Δx :

$$T(\Delta x|x - \Delta x, t)P(x - \Delta x, t) = T(\Delta x|x, t)P(x, t) + \sum_{n=1}^{\infty} \frac{1}{n!} (-\Delta x)^n \frac{\partial^n (T(\Delta x|x, t)P(x, t))}{\partial x^n}, \quad (14)$$

so

$$P(x, t + \tau) = P(x, t) + \int d\Delta x \sum_{n=1}^{\infty} \frac{1}{n!} (-\Delta x)^n \frac{\partial^n (T(\Delta x|x, t)P(x, t))}{\partial x^n}. \quad (15)$$

Now we define the Kramers-Moyal (K-M) coefficients, proportional to moments of T :

$$M_n(x, t) = \frac{1}{\tau} \int d\Delta x (\Delta x)^n T(\Delta x|x, t), \quad (16)$$

and then we obtain the Kramers-Moyal equation:

$$\frac{1}{\tau} (P(x, t + \tau) - P(x, t)) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n (M_n(x, t)P(x, t))}{\partial x^n}. \quad (17)$$

This is an infinite order differential-difference equation for evolving the probability density function P , given an underlying transition probability density T . (Sometimes the Kramers-Moyal equation is defined with a time derivative of P on the left hand side instead of the difference).

Let us calculate some example K-M coefficients for the uneven Markov random walk in one dimension, where probabilities of right and left steps are equal, but the right step is of length $a + u$, while the left step is of length $a - u$, (with $a > u$). The transition probability density is independent of starting position and time and we can write

$$T(\Delta x|x, t) = T(\Delta x) = \frac{1}{2}(\delta(\Delta x - (a + u)) + \delta(\Delta x + (a - u))), \quad (18)$$

using the Dirac delta function, defined to be infinite where the argument vanishes, and zero everywhere else, and satisfying the important conditions $\int_{-\infty}^{\infty} \delta(y)dy = 1$ and $\int_{-\infty}^{\infty} f(y)\delta(y - y_0)dy = f(y_0)$. Probability *densities* can be greater than unity, so we need not worry about infinite values of T ! We get the first four K-M coefficients:

$$\begin{aligned} M_1 &= \frac{1}{\tau} \int d\Delta x \Delta x \frac{1}{2}(\delta(\Delta x - a - u) + \delta(\Delta x + a - u)) = \frac{1}{2\tau} (a + u - a + u) = \frac{u}{\tau} \\ M_2 &= \frac{1}{\tau} \int d\Delta x (\Delta x)^2 \frac{1}{2}(\delta(\Delta x - a - u) + \delta(\Delta x + a - u)) = \frac{a^2 + u^2}{\tau} \\ M_3 &= \frac{1}{\tau} \int d\Delta x (\Delta x)^3 \frac{1}{2}(\delta(\Delta x - a - u) + \delta(\Delta x + a - u)) = \frac{3a^2u + u^3}{\tau} \\ M_4 &= \frac{1}{\tau} \int d\Delta x (\Delta x)^4 \frac{1}{2}(\delta(\Delta x - a - u) + \delta(\Delta x + a - u)) = \frac{a^4 + 6a^2u^2 + u^4}{\tau} \quad \text{etc.} \end{aligned} \quad (19)$$

The K-M equation is rather intractable, due to the infinite number of higher derivatives on the right hand side. However, let us confine attention to evolution in continuous time, and consider only stochastic processes which are continuous in space in this limit. This excludes processes which involve discontinuous jumps: the allowed step lengths must go to zero as the timestep goes to zero.

For the random walk, for example, we allow a , u and τ to go to zero such that $\lim(u/\tau) \rightarrow v_D$ and $\lim(a^2/\tau) \rightarrow 2D$, both constants, in which case every K-M coefficient vanishes except the first and second. Furthermore, the difference on the left hand side of the K-M equation now does become a time derivative and we end up with a diffusion equation with a so-called drift term characterised by a so-called drift velocity v_D :

$$\frac{\partial P(x, t)}{\partial t} = -v_D \frac{\partial P(x, t)}{\partial x} + D \frac{\partial^2 P(x, t)}{\partial x^2}. \quad (20)$$

This is an example of a so-called Fokker-Planck equation (FPE), that describes a continuous stochastic process in space and time. If the drift velocity were zero, it would reduce to the ordinary diffusion equation. Note that it is probability that is diffusing, not a physical property like gas concentration.

Why do we call the first term on the right hand side a drift term? Consider equation (20) without the diffusive term. The solution can be written

$$P(x, t) = P(x - v_D t, 0) = \psi(x - v_D t), \quad (21)$$

which states that the probability density at position x at time t is the same as that which lay at position $x - v_D t$ at $t = 0$: it has therefore drifted through space at velocity v_D . Let's check this solution:

$$\left(\frac{\partial P(x, t)}{\partial x} \right)_t = \left(\frac{\partial \psi(x - v_D t)}{\partial x} \right)_t = \frac{d\psi(x - v_D t)}{d(x - v_D t)} \left(\frac{\partial(x - v_D t)}{\partial x} \right)_t = \psi', \quad (22)$$

where the prime indicates a first derivative, and

$$\left(\frac{\partial P(x, t)}{\partial t} \right)_x = \left(\frac{\partial \psi(x - v_D t)}{\partial t} \right)_x = \frac{d\psi(x - v_D t)}{d(x - v_D t)} \left(\frac{\partial(x - v_D t)}{\partial t} \right)_x = -v_D \psi' = -v_D \left(\frac{\partial P(x, t)}{\partial x} \right)_t, \quad (23)$$

as required. The FPE is a Liouville equation supplemented by diffusion. A general version will have first and second K-M coefficients that have spatial and temporal dependence:

$$\frac{\partial P(x, t)}{\partial t} = - \frac{\partial (M_1(x, t) P(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2 (M_2(x, t) P(x, t))}{\partial x^2}. \quad (24)$$

This equation describes stochastic processes with a probability current J , defined by

$$J = M_1(x, t) P(x, t) - \frac{1}{2} \frac{\partial (M_2(x, t) P(x, t))}{\partial x}. \quad (25)$$

This is apparent when we rewrite the FPE in the form

$$\frac{\partial P(x, t)}{\partial t} = - \frac{\partial}{\partial x} \left(M_1(x, t) P(x, t) - \frac{1}{2} \frac{\partial (M_2(x, t) P(x, t))}{\partial x} \right) = - \frac{\partial J}{\partial x}, \quad (26)$$

which is a continuity equation for probability density. If we write

$$J = \left(M_1(x, t) - \frac{1}{2} \frac{\partial M_2(x, t)}{\partial x} \right) P(x, t) - \frac{1}{2} M_2(x, t) \frac{\partial (P(x, t))}{\partial x}, \quad (27)$$

we see a drift term proportional to probability density, and a diffusive term proportional to the gradient of probability density.

The FPE is a Kramers-Moyal equation (in continuous time) with terms beyond the second derivative neglected. Arguments for the neglect of terms beyond the first two for a more general transition probability density are rather technical. Often the dropping of higher terms is justified non-rigorously by simply noting that the higher moments of the transition probability density are small compared with the first two. There are exceptions, for example transition probability densities with infinite moments, that require special treatments, but we shall not consider such things here.

4.3 Wiener process

The Wiener process is a particular stochastic process in continuous time and space: the limit of the symmetric Markov random walk in one dimension as timestep and spatial step go to zero. The probability density $P(x, t)$ evolves according to the zero-drift diffusion equation

$$\frac{\partial P(x, t)}{\partial t} = D \frac{\partial^2 P(x, t)}{\partial x^2}, \quad (28)$$

with an initial condition $P(x, 0) = \delta(x)$.

The statistical properties of the Wiener process may be obtained by solving the diffusion equation. We define the characteristic function

$$G(k, t) = \int_{-\infty}^{\infty} P(x, t) e^{ikx} dx, \quad (29)$$

such that

$$P(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(k, t) e^{-ikx} dk. \quad (30)$$

We Fourier transform both sides of the diffusion equation to get

$$\int_{-\infty}^{\infty} \frac{\partial P}{\partial t} e^{ikx} dx = D \int_{-\infty}^{\infty} \frac{\partial^2 P(x, t)}{\partial x^2} e^{ikx} dx, \quad (31)$$

which after two integrations by parts, and assuming that P and $\partial P/\partial x$ go to zero as $x \rightarrow \pm\infty$ for all t , becomes

$$\frac{\partial G}{\partial t} = -k^2 DG. \quad (32)$$

Since $P(x, 0) = \delta(x)$, we have $G(k, 0) = 1$, and a full solution to the dynamics of the characteristic function is

$$G(k, t) = \exp(-k^2 Dt). \quad (33)$$

The inverse Fourier transform could now be performed, but it is easier to proceed by differentiating both sides of equation (30) with respect to x and use equation (33):

$$\frac{\partial P(x, t)}{\partial x} = -\frac{i}{2\pi} \int_{-\infty}^{\infty} k G(k, t) e^{-ikx} dk = -\frac{i}{2\pi} \int_{-\infty}^{\infty} k \exp(-k^2 Dt) e^{-ikx} dk, \quad (34)$$

and then integrate by parts:

$$\begin{aligned} \frac{\partial P(x, t)}{\partial x} &= -\frac{i}{2\pi} \left(\left[-\frac{1}{2Dt} \exp(-k^2 Dt) e^{-ikx} \right]_{-\infty}^{\infty} + \frac{1}{2Dt} \int_{-\infty}^{\infty} \exp(-k^2 Dt) (-ix) e^{-ikx} dk \right) \\ &= -\frac{x}{2Dt} P(x, t), \end{aligned} \quad (35)$$

which leads to

$$\int \frac{dP}{P} = -\frac{1}{2Dt} \int x dx, \quad (36)$$

and upon integrating and normalising we get the solution

$$P(x, t) = \frac{1}{(4\pi Dt)^{1/2}} \exp\left(-\frac{x^2}{4Dt}\right). \quad (37)$$

The development of a gaussian probability density function for position, starting from a delta function at the origin at $t = 0$, with a variance proportional to time t , is a characteristic of the Wiener process. It is ubiquitous in stochastic process modelling, because of the Central Limit Theorem: almost all (symmetrical) elementary transition probability densities T will give rise to the Wiener process in the continuum limit.

4.4 Ornstein-Uhlenbeck process

A second important stochastic process describes the random evolution of the velocity of a particle. We shall approach this from a different basis: the Langevin equation. This is a phenomenological treatment of the dynamics. We imagine that the velocity evolves due to a damping term together with a random noise:

$$\dot{v} = -\gamma v + b\xi(t), \quad (38)$$

where γ is the so-called friction coefficient, and the noise ξ has statistical properties

$$\langle \xi(t) \rangle = 0, \quad (39)$$

where the brackets represents an expectation over the probability distribution of the noise, and

$$\langle \xi(t)\xi(t') \rangle = \delta(t - t'), \quad (40)$$

which states that the noise is sampled from a distribution with zero mean and with no autocorrelation in time. For simplicity, the distribution of noise is assumed to be gaussian. The singular variance implied by equation (40) at $t = t'$ seems to present a problem but this can be handled.

The Langevin description of uncertain dynamics is equivalent to a stochastic process described by a Fokker-Planck equation. We can demonstrate this by evaluating the K-M coefficients, rewriting equation (16) in the form

$$M_n(v, t) = \frac{1}{\tau} \int d\Delta v (\Delta v)^n T(\Delta v|v, t) = \frac{1}{\tau} \langle (v(t + \tau) - v(t))^n \rangle. \quad (41)$$

This proposes the equivalence between an average of an increment Δv over a transition probability density T and over the statistics of the noise ξ . We then employ the integral of equation (38), for small τ :

$$v(t + \tau) - v(t) = -\gamma \int_t^{t+\tau} v dt + b \int_t^{t+\tau} \xi(t') dt' \approx -\gamma v(t) \tau + b \int_t^{t+\tau} \xi(t') dt', \quad (42)$$

and according to the assumed properties of the noise and in the limit $\tau \rightarrow dt$ we get $v(t + \tau) - v(t) = dv$ and $\langle dv \rangle = -\gamma v dt$ such that $M_1(v) = \langle \dot{v} \rangle = -\gamma v$. We also construct

$$(v(t + \tau) - v(t))^2 = (\gamma v(t) \tau)^2 - 2\gamma v(t) \tau b \int_t^{t+\tau} \xi(t') dt' + b^2 \int_t^{t+\tau} dt' \int_t^{t+\tau} dt'' \xi(t') \xi(t''), \quad (43)$$

and with the appropriate use of statistical properties and the continuum limit, we get $\langle (dv)^2 \rangle = b^2 dt$ and hence $M_2 = b^2$. We have therefore established that the equivalent FPE to the Langevin equation (38) is

$$\frac{\partial P(v, t)}{\partial t} = \frac{\partial (\gamma v P(v, t))}{\partial v} + \frac{b^2}{2} \frac{\partial^2 P(v, t)}{\partial v^2}. \quad (44)$$

It may be demonstrated that $P(v) \propto \exp(-\gamma v^2/b^2)$ is a stationary solution to this equation. If this is to represent the Maxwell-Boltzmann solution expected for the velocity distribution of a particle of mass m in thermal equilibrium with a bath at temperature T , then b must be related to T and γ in the form $b^2 = 2kT\gamma/m$, where k is Boltzmann's constant. This is a connection known as a fluctuation dissipation relation: b characterises the fluctuations and γ the dissipation in the Langevin equation.

Furthermore, it may be shown that the time-dependent solution to equation (44), with initial condition $P(v, t_0) = \delta(v - v_0)$ is

$$P(v, t) = \mathcal{P}(v, t | v_0, t_0) = \sqrt{\frac{m}{2\pi kT(1 - e^{-2\gamma(t-t_0)})}} \exp\left(-\frac{m(v - v_0 e^{-\gamma(t-t_0)})^2}{2kT(1 - e^{-2\gamma(t-t_0)})}\right). \quad (45)$$

This is a gaussian with time-dependent mean and variance. The notation \mathcal{P} is used now to denote the transition probability density for a continuous stochastic process, defined by equation (38) or (44), starting from initial value v_0 at initial time t_0 , and ending at the final value v at time t . This behaviour is known as an Ornstein-Uhlenbeck process.

The same procedure can be used to describe the motion of a particle in a harmonic potential $\phi(x) = \kappa x^2/2$, in the limit that the frictional damping coefficient γ is very large. The Langevin equations that describe the dynamics are now

$$\begin{aligned} \dot{v} &= -\gamma v - \frac{\kappa}{m} x + b \xi(t) \\ \dot{x} &= v, \end{aligned} \quad (46)$$

which reduce in the so-called overdamped limit to

$$\dot{x} = -\frac{\kappa}{m\gamma} x + \frac{b}{\gamma} \xi(t), \quad (47)$$

which then has the same form as the evolution equation (38) (but for position instead of velocity). The solutions for the stochastic evolution obtained in this section apply equally well to this case, and we shall use these later on.

4.5 Summary

Let us take stock of our position. We have developed tools to determine the evolution of probability distributions and densities, or equivalently the stochastic evolution of particle velocities or positions. The evolution equations are not time reversal symmetric since the principle of causality is assumed: the probability of an event depends only upon events that precede it in time, and not by events in the future.

The evolution of a probability density of system properties is designed to capture the growth in uncertainty during the motion. The dynamics takes the form of random transitions occurring during the course of an evolution: a continuous injection of uncertainty, it might be thought. This injection might be traced back to uncertainty in the *initial* conditions for the system: indefiniteness in some of those conditions will take time to have an effect upon the system property of interest. For example, as time evolves the uncertainty in the velocity

of one particle in a gas increases as particles that were initially far away have the opportunity to get closer and interact. Uncertainty in the exact initial velocity of those once-distant particles gives rise to uncertainty in the Langevin ξ term at later times. So the picture is that initial uncertainty affects subsequent uncertainty and that it is typical that uncertainty grows. We now wish to see how the rather vague concept of uncertainty might be commodified and related to entropy and irreversibility.

5 Entropy generation and stochastic irreversibility

5.1 The reversibility of a stochastic trajectory

The usual statement of the second law in thermodynamics is that it is impossible to observe the reverse of an entropy producing process. Let us immediately reject this version of the law and recognise that nothing is impossible. A bird's egg might fall off a cliff and shatter on the rocks below. But there is never stillness at the microscopic level, and without breaking any law of mechanics, the molecular motion of the rocks and fragments of egg might conspire to reverse their macroscopic motion, bringing the pieces of egg back together again and projecting it back safely into the nest. This is not ridiculous: it is an inevitable consequence of the time reversal symmetry of Newton's laws. All we need for this event to occur is to create the right initial conditions. Of course, that is where the problem lies: it is virtually impossible to engineer such a situation, but virtually impossible is not absolutely impossible. So, better put, the second law of thermodynamics states that it is *highly unlikely* to observe the reverse of an entropy producing process. We might get somewhere if we calculate just how unlikely.

Let us be specific and consider a thermodynamic process, operating upon a single particle, that consists of a time-dependent change in the potential $\phi(x, t)$ within which it moves. We could also consider time dependence in the reservoir temperature, but for the moment, we assume this to be constant. The overdamped Langevin equation of motion of the particle is written

$$\dot{x} = -\frac{1}{m\gamma} \frac{\partial \phi(x, t)}{\partial x} + \frac{b}{\gamma} \xi(t), \quad (48)$$

and we employ a harmonic potential with a time-dependent spring constant: $\phi(x, t) = \kappa(t)x^2/2$. The equivalent FPE would then be

$$\frac{\partial P(x, t)}{\partial t} = \frac{\kappa(t)}{m\gamma} \frac{\partial (xP(x, t))}{\partial x} + \frac{kT}{m\gamma} \frac{\partial^2 P(x, t)}{\partial x^2}. \quad (49)$$

We consider a thermodynamic process where the spring constant undergoes a sequence of shifts from κ_i to κ_{i+1} at times t_i . So starting with κ_0 , a change is made to κ_1 at time t_0 , to κ_2 at time t_1 , etc. The time intervals between shifts are all equal to τ . Thus after time $t_n = n\tau$, the spring constant has changed to κ_n , and the probability density has evolved to $P(x, n\tau) = P_{\text{end}}(x)$, having started from $P(x, 0) = P_{\text{start}}(x)$. The probability density that a trajectory of the particle consists of the positions (x_0, x_1, \dots, x_n) at times $(0, \tau, \dots, n\tau)$ is given by

$$p(\{x_i\}) \propto \mathcal{P}_{\kappa_n}(x_n, t_n | x_{n-1}, t_{n-1}) \mathcal{P}_{\kappa_{n-1}}(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \cdots \mathcal{P}_{\kappa_1}(x_1, t_1 | x_0, t_0) P_{\text{start}}(x_0), \quad (50)$$

where each transition probability density is labelled by the prevailing value of the spring constant for the period $t_{i-1} \rightarrow t_i$. The probability density to be at position x_n at t_n is obtained by integrating over all contributing trajectories:

$$P_{\text{end}}(x_n) \propto \int \prod_{i=0}^{n-1} dx_i \mathcal{P}_{\kappa_n}(x_n, t_n | x_{n-1}, t_{n-1}) \mathcal{P}_{\kappa_{n-1}}(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \cdots \mathcal{P}_{\kappa_1}(x_1, t_1 | x_0, t_0) P_{\text{start}}(x_0) \quad (51)$$

Now what do we mean by reversibility? Intuitively this must relate to the behaviour of the system when it is driven by a sequence of changes in spring constant at specified instants in time, and then the reversed sequence. Specifically, how well can the effect on the system of the first sequence be annulled by the reversed sequence? There are several subtle points to bear in mind here, but for our purposes, let us make it a requirement of reversibility that there is equality between the probability density that the trajectory (x_0, x_1, \dots, x_n) is followed during the forward sequence, or protocol, and the probability density that the corresponding antitrajectory $(x_n, x_{n-1}, \dots, x_0)$ is followed during the reversed sequence, or antiprotocol. An extension of this requirement is then to make it apply to all possible trajectories. Then we know that any observed response of the system to the protocol might be undone with equal probability. This is what we mean by a reversible change. For large systems, fluctuations are slight and there is a dominant trajectory that is followed during the protocol, and for

a reversible phenomenon the dominant antitrajectory followed during the antiprotocol is the exact reverse of the trajectory.

The measure of irreversibility, for a particular trajectory, is therefore the degree to which the trajectory and antitrajectory probability densities differ. Let us define, with suggestive notation,

$$\Delta S_{\text{tot}} = k \ln \left(\frac{\mathcal{P}_{\kappa_n}(x_n, t_n | x_{n-1}, t_{n-1}) \mathcal{P}_{\kappa_{n-1}}(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \cdots \mathcal{P}_{\kappa_1}(x_1, t_1 | x_0, t_0) P_{\text{start}}(x_0)}{\mathcal{P}_{\kappa_1}(x_0, t_n | x_1, t_{n-1}) \cdots \mathcal{P}_{\kappa_{n-1}}(x_{n-2}, t_2 | x_{n-1}, t_1) \mathcal{P}_{\kappa_n}(x_{n-1}, t_1 | x_n, t_0) P_{\text{end}}(x_n)} \right), \quad (52)$$

so the trajectory is reversible if $\Delta S_{\text{tot}} = 0$, under the prevailing dynamics, for all trajectories, and otherwise irreversible. We shall make the claim that ΔS_{tot} is the *total production of entropy in the universe* associated with the trajectory (x_0, x_1, \dots, x_n) and the specified protocol. We shall see that the probability distribution for ΔS_{tot} has properties that ensures that its expectation value is non-negative. But this does not mean that negative values of ΔS_{tot} are impossible; just that they are rare. These results are consequences of the structure of equation (52): this structure gives rise to fluctuation relations, the objective of these notes.

We see in equation (52) the promised connection between entropy production and the evolution of uncertainty, embodied in the probability density P and the transition probability densities \mathcal{P} . It is fundamental to these ideas that without stochasticity in the dynamics, there is no change in the entropy of the universe.

5.2 Integral fluctuation relation

The first of the results we want to establish is called an *integral fluctuation theorem*. Let us simplify the notation in the definition of ΔS_{tot} :

$$\Delta S_{\text{tot}}[x_n \leftarrow x_0] = k \ln \left(\frac{\mathcal{P}_{\text{for}}[x_n | x_0] P_{\text{start}}(x_0)}{\mathcal{P}_{\text{rev}}[x_0 | x_n] P_{\text{end}}(x_n)} \right), \quad (53)$$

where $\mathcal{P}_{\text{for}}[x_n | x_0]$ is the trajectory probability density for the forward protocol, and $\mathcal{P}_{\text{rev}}[x_0 | x_n]$ is the antitrajectory probability density for the reversed protocol. Now evaluate the expectation value of $\exp(-\Delta S_{\text{tot}}/k)$ over all trajectories:

$$\begin{aligned} \langle \exp(-\Delta S_{\text{tot}}/k) \rangle &= \int \mathcal{P}_{\text{for}}[x_n | x_0] P_{\text{start}}(x_0) \exp(-\Delta S_{\text{tot}}/k) dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{for}}[x_n | x_0] P_{\text{start}}(x_0) \left[\frac{\mathcal{P}_{\text{rev}}[x_0 | x_n] P_{\text{end}}(x_n)}{\mathcal{P}_{\text{for}}[x_n | x_0] P_{\text{start}}(x_0)} \right] dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{rev}}[x_0 | x_n] P_{\text{end}}(x_n) dx_0 \dots dx_n = 1, \end{aligned} \quad (54)$$

since the last line is the normalisation condition for the probability density over all antitrajectories. This is the integral fluctuation relation. It is assumed that the allowed ranges of all the x_i is the same. We now employ Jensen's inequality

$$\langle \exp(z) \rangle \geq \exp\langle z \rangle, \quad (55)$$

valid for any distribution of the z , to conclude that

$$\langle \exp(-\Delta S_{\text{tot}}/k) \rangle = 1 \geq \exp(-\langle \Delta S_{\text{tot}} \rangle / k), \quad (56)$$

such that

$$\langle \Delta S_{\text{tot}} \rangle \geq 0. \quad (57)$$

This deceptively simple construction and analysis has yielded a quantity that is expected to increase under the defined stochastic dynamics. Might it indeed be related to entropy increase? Let us see.

5.3 Total entropy change arising from a trajectory

It is revealing to look further into the structure of ΔS_{tot} by inserting the transition probability density for the Ornstein-Uhlenbeck process of overdamped spatial motion in a harmonic potential:

$$\mathcal{P}_{\kappa}(x', t' | x, t) = \left(\frac{\kappa}{2\pi kT(1 - e^{-2\kappa(t'-t)/(m\gamma)})} \right)^{1/2} \exp \left(-\frac{\kappa(x' - xe^{-\kappa(t'-t)/(m\gamma)})^2}{2kT(1 - e^{-2\kappa(t'-t)/(m\gamma)})} \right), \quad (58)$$

into a regrouped version of equation (52), with $t_i - t_{i-1} = \tau$:

$$\begin{aligned}
\Delta S_{\text{tot}} &= k \ln \left(\frac{\mathcal{P}_{\kappa_n}(x_n, t_n | x_{n-1}, t_{n-1}) \cdots \mathcal{P}_{\kappa_1}(x_1, t_1 | x_0, t_0) P_{\text{start}}(x_0)}{\mathcal{P}_{\kappa_n}(x_{n-1}, t_1 | x_n, t_0) \cdots \mathcal{P}_{\kappa_1}(x_0, t_n | x_1, t_{n-1}) P_{\text{end}}(x_n)} \right) \\
&= \sum_{i=1}^n \left(\frac{\kappa_i (x_{i-1} - x_i e^{-\kappa_i \tau / (m\gamma)})^2}{2T (1 - e^{-2\kappa_i \tau / (m\gamma)})} - \frac{\kappa_i (x_i - x_{i-1} e^{-\kappa_i \tau / (m\gamma)})^2}{2T (1 - e^{-2\kappa_i \tau / (m\gamma)})} \right) + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_n)} \right) \\
&= - \sum_{i=1}^n \frac{\kappa_i}{2T} (x_i^2 - x_{i-1}^2) + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_n)} \right) \\
&= \sum_{i=1}^n \frac{1}{2T} (\kappa_i x_{i-1}^2 - \kappa_{i-1} x_{i-1}^2) - \sum_{i=1}^n \frac{1}{2T} (\kappa_i x_i^2 - \kappa_{i-1} x_{i-1}^2) + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_n)} \right) \\
&= \sum_{i=1}^n \frac{1}{2T} \frac{(\kappa_i - \kappa_{i-1})}{\tau} x_{i-1}^2 \tau - \sum_{i=1}^n \frac{1}{T} \frac{(\phi(x_i) - \phi(x_{i-1}))}{\tau} \tau + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_n)} \right). \tag{59}
\end{aligned}$$

If the sequence of shifts in spring constant were replaced by a continuous time-dependent $\kappa(t)$, this would reduce to

$$\Delta S_{\text{tot}} = \frac{1}{T} \int_0^t dt' \left(\frac{\partial \phi(x(t'), t')}{\partial t'} - \frac{d\phi(x(t'), t')}{dt'} \right) - k \ln \left(\frac{P_{\text{end}}(x(t))}{P_{\text{start}}(x(0))} \right), \tag{60}$$

where we have reinserted the potential $\phi(x, t) = \kappa(t)x^2/2$, and replaced the x_i with the continuous function $x(t')$.

It is a crucial observation that the term in brackets inside the integrand is the negative of the rate of heat transfer to the system at time t' . In this example, work is the energy transfer to the system brought about by changes to the potential. So the work increment at time t_i is $(\kappa_{i+1} - \kappa_i)(x(t_i))^2/2$ and the rate of working is $\dot{W} = \partial\phi/\partial t'$. Since the only energy E in an overdamped isothermal system is potential energy ϕ (the kinetic energy being always constant), the term in brackets is $(\dot{W} - \dot{E})$. According to the first law, this difference is $-\dot{Q}$, the rate of heat transfer from the system into the heat bath.

Note that according to the normal rules of calculus

$$d\phi = \frac{\partial \phi(x, t)}{\partial t} dt + \frac{\partial \phi(x, t)}{\partial x} dx, \tag{61}$$

and so we might write

$$\frac{\partial \phi}{\partial t} dt - \frac{d\phi}{dt} dt = - \frac{\partial \phi}{\partial x} dx = f(x, t) dx, \tag{62}$$

where we have introduced the force $f(x, t)$ acting on the particle. Accordingly, we might be tempted to regard the term in brackets in equation (60) as the work done on the system, not the loss of heat to the environment, but there is danger of misinterpretation here associated with the fact that the system is overdamped. Furthermore, dx is a random variable and the rules of calculus for such variables can differ from those of ordinary variables, a subject we shall not go into. It is safer to work with the integrand in the form given in equation (60) and its interpretation as heat transfer, which is sound.

So we have now derived the result

$$\Delta S_{\text{tot}} [x(t) \leftarrow x(0)] = - \frac{1}{T} \int_0^t \dot{Q} dt' - k \ln \left(\frac{P_{\text{end}}(x(t))}{P_{\text{start}}(x(0))} \right) = - \frac{\Delta Q}{T} - k \ln \left(\frac{P(x(t), t)}{P(x(0), 0)} \right), \tag{63}$$

for this particular isothermal process. We claim again that it is the total entropy production associated with a particular trajectory from $x(0)$ to $x(t)$ driven by the underlying stochastic dynamics of a Langevin equation. It can also be written

$$\Delta S_{\text{tot}} = \Delta S_{\text{res}} + \Delta S, \tag{64}$$

where $-\Delta Q/T$ has been identified as the change in entropy of the reservoir ΔS_{res} and ΔS is the change in the system entropy S , defined here as

$$S(x, t) = -k \ln(P(x, t) dx), \tag{65}$$

where an increment dx has been included to take care of dimensions. Notice the difference between this and the Gibbs entropy of equation (7). The inclusion of dx is not the crucial difference. More important is the recognition that the Gibbs entropy is an average of the system entropy over all possible configurations of the system, namely

$$S_{\text{Gibbs}} = \int S(x, t) P(x, t) dx = \langle S \rangle. \tag{66}$$

We have defined a system entropy (65) that appears to be a property of a distinct ‘configuration’ x of the system at time t , and not of a whole ensemble of systems. Some careful interpretation is required here, since S is defined in terms of the probability density over the configurations of the system, and this suggests a role for an ensemble and not a single state. But if we regard $P(x, t)$ as a statement of belief about a single system, instead of a representation of the outcomes of many trials conducted on an ensemble of systems, then S is indeed a property that we can ascribe to a single system. It depends not only upon its current configuration x , but also on the stochastic dynamics that brought it to that configuration at time t , and the state of belief, or of information, we had about its configuration at the initial time. It is not necessary to maintain the mental device of the ensemble and we need only consider a single system, which to my mind is preferable.

The earlier result $\langle \Delta S_{\text{tot}} \rangle \geq 0$ allows us to show that

$$\Delta S_{\text{Gibbs}} - \frac{\langle \Delta Q \rangle}{T} \geq 0, \quad (67)$$

which is an echo of equations (3) and (5). We are now saying that the thermodynamic internal entropy production ΔS_i is equivalent to the mean total entropy production $\langle \Delta S_{\text{tot}} \rangle$.

However, we have in $S(x, t)$ a more powerful concept than the Gibbs entropy, and in ΔS_{tot} a more specific measure of irreversibility than ΔS_i . The total entropy change can be negative, in accord with our acceptance of the remote possibility of eggs returning to nests, etc. And considering the likelihood of such events leads us to further fluctuation relations. Furthermore, whilst system entropy is clearly linked to uncertainty about the state of a system, it is clearly affected by our choice about how well to categorise the system: the category width dx appears in equation (65). If we are sure that a particle is to be found between a specified x and $x + dx$, then its entropy according to equation (65) would be zero. If we had chosen dx to be smaller, we might not be so sure where the particle was. The entropy is dependent upon the scale of discretisation of state space. It *does* depend on the level of our perception or measurements, as we suggested would be the case at the end of section 2.

Finally, although the derivation has been confined to a particular example, it is possible to be much more general and consider an overdamped system subject to a general potential $\phi(x, t)$ and to an additional non-conservative force. The temperature of the heat bath can be made a function of time and even space. The behaviour under finite damping can be considered. We would proceed by determining the transition probability densities of the relevant Fokker-Planck equation, and combining them in a suitable fashion to construct ΔS_{tot} . Equation (64) would still hold, with

$$\Delta S_{\text{res}} = - \int \frac{dQ}{T}, \quad (68)$$

but with a more elaborate definition of the heat transfer ΔQ . Unfortunately, it is beyond the scope of these notes to be able to demonstrate this powerful result in its full generality.

6 Further results

6.1 Incomplete integral fluctuation relation

When the initial probability distribution extends over a smaller range of possible positions than the final distribution, we must modify the integral fluctuation relation. Whilst we can still write, following equation (54):

$$\begin{aligned} \langle \exp(-\Delta S_{\text{tot}}/k) \rangle &= \int \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \exp(-\Delta S_{\text{tot}}/k) dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \left[\frac{\mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n)}{\mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0)} \right] dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) dx_0 \dots dx_n, \end{aligned} \quad (69)$$

the final integral is not unity. It is an integral of the probability density of all (reversed) trajectories starting in the range allowed at the end of the process, but finishing in the range allowed initially. This excludes some available possibilities, so the integral is less than unity, and hence

$$\langle \exp(-\Delta S_{\text{tot}}/k) \rangle \leq 1, \quad (70)$$

for a process that is characterised by the lifting of a constraint upon system position. We may still employ the result

$$1 \geq \langle \exp(-\Delta S_{\text{tot}}/k) \rangle \geq \exp(-\langle \Delta S_{\text{tot}} \rangle / k), \quad (71)$$

and deduce that $\langle \Delta S_{\text{tot}} \rangle \geq 0$ still holds, but some of the further implications of the integral fluctuation relation, that we now explore, will need modification. But for the rest of this section, we assume that the process involves no release of constraint on the range of positions that a system might explore.

6.2 Jarzynski equality

Let us consider

$$\Delta S_{\text{tot}} = k \ln \left(\frac{\mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0)}{\mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n)} \right), \quad (72)$$

for an isothermal process that starts and ends in canonical equilibrium, i.e. $P_{\text{start}}(x_0) \propto \exp(-(F_{\text{start}} - E_{\text{start}}(x_0))/kT)$ and $P_{\text{end}}(x_n) \propto \exp(-(F_{\text{end}} - E_{\text{end}}(x_n))/kT)$. Then

$$\Delta S_{\text{tot}} = -\frac{\Delta Q}{T} - \frac{1}{T} (\Delta F - \Delta E) = \frac{1}{T} (\Delta W - \Delta F) = \frac{W_d}{T}, \quad (73)$$

having defined a free energy difference $\Delta F = F_{\text{end}} - F_{\text{start}}$, $\Delta E = E_{\text{end}} - E_{\text{start}}$ and $\Delta W = \Delta E - \Delta Q$, and introduced the so-called dissipative work $W_d = \Delta W - \Delta F$. Then the integral fluctuation relation tells us that

$$\langle \exp(-W_d/kT) \rangle = 1, \quad (74)$$

for such a process. Equivalently

$$\langle \exp(-\Delta W/kT) \rangle = \exp(-\Delta F/kT), \quad (75)$$

which is known as the Jarzynski equality. A time-dependent change to the potential acting upon a system is associated with the performance of work, and the Jarzynski equality tells us that the average of an exponential function of this mechanical work is related to the change in the free energy corresponding to the change in the potential. In classical thermodynamics, a free energy change is sometimes called the ‘reversible work’ since it is equal to a mechanical work performed quasistatically, meaning slowly, without entropy generation. In other words the dissipative work in a quasistatic process is zero. The Jarzynski equality tells us how this result is modified for a non-quasistatic process, one where mechanical work is performed at a finite rate such that the system is driven away from thermal equilibrium and the process is irreversible. The Jarzynski equality implies that $\langle W_d \rangle \geq 0$, or

$$\langle \Delta W \rangle \geq \Delta F. \quad (76)$$

The equals sign applies if the probability distribution of ΔW is a delta function: $P(\Delta W) = \delta(\Delta W - \Delta F)$. For a quasistatic process all possible measurements of ΔW are the same, and equal to ΔF . Fluctuations away from a unique possible outcome are associated with irreversibility and with $\langle \Delta W \rangle > \Delta F$.

Note that the Jarzynski equality holds for a processes in which work is performed but where the system does not necessarily end in equilibrium. The reason is that such a process could be continued with zero further performance of work until a true thermal equilibrium is reached. For such an extended process, the Jarzynski equality would of course hold. However, the statistics of the work performed would not change between the ending of the work and the achievement of equilibration. Thus $\langle \exp(-W_d/kT) \rangle = 1$ holds for any isothermal process that begins in equilibrium.

6.3 Crooks relation

Let us investigate further properties of the probability distribution of isothermal work performed during a process. Once again we consider one where the system starts and ends in equilibrium, such that $\Delta S_{\text{tot}}^{\text{for}} = (\Delta W - \Delta F)/T$. The probability density specifying that the work done is equal to $\Delta W'$ may be expressed as

$$P_{\text{for}}(\Delta W') = \int [dx] \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \delta(\Delta W_{\text{for}}[x] - \Delta W'), \quad (77)$$

where the integration is over all trajectories $x_0 \rightarrow x_n$ possible under the dynamics and $P_{\text{start}}(x_0)$ takes canonical form. The suffix ‘for’ is added to the distribution and work to distinguish it from new quantities that we shall introduce shortly. We proceed by writing

$$P_{\text{for}}(\Delta W') = \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \exp\left(\frac{\Delta W_{\text{for}}[x] - \Delta F}{kT}\right) \delta(\Delta W_{\text{for}}[x] - \Delta W'), \quad (78)$$

having inserted the particular form for ΔS_{tot} in terms of $\Delta W_{\text{for}} = \int_0^t \partial\phi/\partial t' dt'$. This reduces to

$$P_{\text{for}}(\Delta W') = \exp\left(\frac{\Delta W' - \Delta F}{kT}\right) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \delta(\Delta W_{\text{for}}[x] - \Delta W'). \quad (79)$$

Now we consider a process driven by a sequence of changes to system potential that is opposite to the sequence that drives the forward process. The total entropy production for the reverse trajectory generated by this reverse process, starting at equilibrium in the final state, is defined as

$$\Delta S_{\text{tot}}^{\text{rev}} = \frac{1}{T} (\Delta W_{\text{rev}} - (-\Delta F)) = k \ln \left(\frac{\mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n)}{\mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0)} \right), \quad (80)$$

where start and end distributions have been swapped, and the labelling of the transition probability densities have been changed to reflect the nature of the inverted processes under consideration. Note that the change in free energy associated with this reverse process is equal and opposite to the change associated with the forward process. ΔW_{rev} remains a functional of the coordinates x_i traversed during the reverse trajectory, except that they are encountered along the reverse trajectory in the opposite order. Clearly $\Delta W_{\text{rev}} + \Delta W_{\text{for}} = 0$ and so we write

$$P_{\text{for}}(\Delta W') = \exp \left(\frac{\Delta W' - \Delta F}{kT} \right) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \delta(-\Delta W_{\text{rev}}[x] - \Delta W'). \quad (81)$$

Then by recognising that

$$P_{\text{rev}}(-\Delta W') = \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \delta(\Delta W_{\text{rev}}[x] + \Delta W'), \quad (82)$$

we establish that

$$P_{\text{for}}(\Delta W) = \exp \left(\frac{\Delta W - \Delta F}{kT} \right) P_{\text{rev}}(-\Delta W), \quad (83)$$

where we have removed the primes for clarity of presentation. Thus the probability distributions of work done in a forward and a reversed process, starting and ending in canonical equilibrium (such that $\Delta S_{\text{tot}} = W_d/T$), are related to one another. As before, the statistics do not change after the termination of the performance of work, so equation (83) holds even for processes that do not end in equilibrium.

The analysis might be extended to consider the distribution of the dissipative work $W_d = \Delta W - \Delta F$ in each of the forward and reverse processes, with the result that

$$P_{\text{for}}(W_d) = \exp \left(\frac{W_d}{kT} \right) P_{\text{rev}}(-W_d). \quad (84)$$

This result is known as the Crooks relation. It has the direct implication that negative values of dissipative work are possible. The probability of dissipative work $-W_d$ during a reverse process is equal to the probability of dissipative work W_d during a forward process, multiplied by an exponential factor $\exp(-W_d/kT)$. For large magnitudes of dissipative work, on a scale of kT , such probabilities become very small, but nevertheless they are non-zero. We begin to be able to quantify the likelihood of fluctuations away from the expected behaviour in thermodynamic processes.

6.4 Detailed fluctuation relation

The Crooks relation connects probability distributions of work and dissipative work for the forward and reverse processes, as long as they begin in equilibrium. A similar relation is satisfied by the probability distribution for ΔS_{tot} generated during time-symmetric (and therefore identical) forward and backward protocols, starting with general but *identical* initial and final distributions P_{start} and P_{end} denoted by \hat{P} . The procedure is very similar to that just followed. We write

$$\begin{aligned} P_{\text{for}}(\Delta S'_{\text{tot}}) &= \int [dx] \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \delta(\Delta S_{\text{tot}}^{\text{for}}[x] - \Delta S'_{\text{tot}}) \\ &= \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \exp \left(\frac{\Delta S_{\text{tot}}^{\text{for}}[x]}{k} \right) \delta(\Delta S_{\text{tot}}^{\text{for}}[x] - \Delta S'_{\text{tot}}) \\ &= \exp \left(\frac{\Delta S'_{\text{tot}}}{k} \right) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \delta(\Delta S_{\text{tot}}^{\text{for}}[x] - \Delta S'_{\text{tot}}) \\ &= \exp \left(\frac{\Delta S'_{\text{tot}}}{k} \right) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{end}}(x_n) \delta(-\Delta S_{\text{tot}}^{\text{rev}}[x] - \Delta S'_{\text{tot}}) \\ &= \exp \left(\frac{\Delta S'_{\text{tot}}}{k} \right) P_{\text{rev}}(-\Delta S'_{\text{tot}}), \end{aligned} \quad (85)$$

where we have employed the result

$$\Delta S_{\text{tot}}^{\text{for}}[x] + \Delta S_{\text{tot}}^{\text{rev}}[x] = -\frac{\Delta Q}{T} + k \ln \left(\frac{\hat{P}(x_0)}{\hat{P}(x_n)} \right) + \frac{\Delta Q}{T} + k \ln \left(\frac{\hat{P}(x_n)}{\hat{P}(x_0)} \right) = 0 \quad (86)$$

so the likelihood of negative entropy production in a reversed process is related to the probability of positive entropy production in the forward process, but exponentially suppressed. This holds for any initial conditions. The relation is called a detailed fluctuation relation in recognition of this specific claim to the probability of negative entropy changes.

For the protocol we have specified, and the requirement of identical initial and final probability densities, the distinction between the forward and backward processes vanishes. We can drop the for and rev suffices and hence the detailed fluctuation relation becomes

$$P(\Delta S_{\text{tot}}) = \exp \left(\frac{\Delta S_{\text{tot}}}{kT} \right) P(-\Delta S_{\text{tot}}). \quad (87)$$

We would expect this to apply to the entropy generated over an integer number of periods of a periodically driven stochastic system, with a driving protocol that is time-symmetric during each period. An even simpler case would be a steady state where the conservative forces are time-independent and the system is maintained out of equilibrium by a constant non-conservative force.

6.5 Dissipation function

In a later section, we shall see there is good reason to consider now a quantity called the dissipation function, denoted $\Omega(x)$, defined on a trajectory driven by a protocol whilst keeping the temperature constant. Its integral over the trajectory has a slightly different structure compared with ΔS_{tot} :

$$\int_0^t \Omega_{\text{for}}(x(t')) dt' = \bar{\Omega}_{\text{for}}[x_n, \dots, x_0] t = \ln \left(\frac{\mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0)}{\mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n)} \right), \quad (88)$$

where it is to be noted that the distribution in the denominator is that which applies at the start of the process and not its end. We have introduced $\bar{\Omega}$ to represent the mean value of Ω over the trajectory, and the suffix ‘for’ to make clear that it is associated with a forward process. Clearly

$$\bar{\Omega}_{\text{for}} t = \frac{\Delta S_{\text{tot}}^{\text{for}}}{k} + \ln \left(\frac{P_{\text{end}}(x_n)}{P_{\text{start}}(x_n)} \right), \quad (89)$$

and

$$\bar{\Omega}_{\text{for}} t = - \int_0^t \frac{dQ}{kT} + \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{start}}(x_n)} \right). \quad (90)$$

If the system starts out in canonical equilibrium with $P_{\text{start}}(x) \propto \exp(-E_{\text{start}}(x)/kT)$ then

$$\bar{\Omega}_{\text{for}} t = - \int_0^t \frac{dQ}{kT} + \frac{1}{kT} (E_{\text{start}}(x_n) - E_{\text{start}}(x_0)). \quad (91)$$

If the time dependence of the protocol is *cyclic*, such that the functional form of the system energy is the same at the end as it was at the beginning, then $E_{\text{start}}(x_n) = E_{\text{end}}(x_n)$ and

$$\bar{\Omega}_{\text{for}} t = -\frac{\Delta Q}{kT} + \frac{\Delta E}{kT} = \frac{\Delta W}{kT}, \quad (92)$$

and the mean dissipation function is proportional to the work done, under these special conditions. Notice that the process need not end in equilibrium, only that the system should have the same energy function $E(x)$ at beginning and end.

But what can we say about the statistical properties of $\bar{\Omega}t$ in the general case, and hence about work done in this specific case? It turns out that $\bar{\Omega}t$ satisfies an integral fluctuation relation in general and a detailed fluctuation relation in certain circumstances. First, we note that

$$\begin{aligned} \langle \exp(-\bar{\Omega}_{\text{for}} t) \rangle &= \int \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \exp(-\bar{\Omega}_{\text{for}} t) dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \left[\frac{\mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n)}{\mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0)} \right] dx_0 \dots dx_n \\ &= \int \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n) dx_0 \dots dx_n = 1, \end{aligned} \quad (93)$$

which establishes the integral fluctuation relation and the implication that $\langle \bar{\Omega}_{\text{for}} t \rangle \geq 0$ or $\langle \bar{\Omega}_{\text{for}} \rangle \geq 0$ for any process. It is assumed that the process does not involve the release of a constraint so that the range of positions at the end is equal to the range at the start. Next we note that in general $\bar{\Omega}_{\text{for}}[x_n \leftarrow x_0]$ and $\bar{\Omega}_{\text{rev}}[x_0 \leftarrow x_n]$ are equal and opposite (they refer to protocol and reversed protocol starting from the same initial probability distribution), and construct

$$\begin{aligned}
P_{\text{for}}(\bar{\Omega}') &= \int [dx] \mathcal{P}_{\text{for}}[x_n|x_0] P_{\text{start}}(x_0) \delta(\bar{\Omega}_{\text{for}}[x] - \bar{\Omega}') \\
&= \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n) \exp(\bar{\Omega}_{\text{for}}[x]t) \delta(\bar{\Omega}_{\text{for}}[x] - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n) \delta(\bar{\Omega}_{\text{for}}[x] - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) \int [dx] \mathcal{P}_{\text{rev}}[x_0|x_n] P_{\text{start}}(x_n) \delta(-\bar{\Omega}_{\text{rev}}[x] - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) P_{\text{rev}}(-\bar{\Omega}'),
\end{aligned} \tag{94}$$

If the protocol is time-symmetric about its mid point, the ‘for’ and ‘rev’ suffices may be dropped. This establishes what is called the transient, or Evans-Searles fluctuation relation for the probability distribution of the mean dissipation function. For example, this result holds for a relaxation towards equilibrium, in the absence of any external work (and hence for a time-symmetric protocol) having started in an arbitrary initial distribution $P_{\text{start}}(x_0)$. If the protocol consists of a time-independent non-conservative field, once again it is time-symmetric and if the system starts out in canonical equilibrium, then not only will the statistics of $\bar{\Omega}$ satisfy equation (94), but also $\bar{\Omega}t$ will correspond to the work done.

This latter situation is of interest when considering a steady dissipative state of the system. We begin in canonical equilibrium and switch on a non-conservative force for a period t . A statistical steady state is reached after a time of order t_{rel} . Under these conditions $\bar{\Omega}$ is the mean rate of performance of work over the trajectory and equation (94) is satisfied. If t is much longer than t_{rel} , then the contribution in the interval $0 \rightarrow t_{\text{rel}}$ to the mean work $kT \int_0^t \Omega(x(t')) dt'$ becomes negligible. In fact we can replace this integral to good accuracy by the expression $kT \bar{\Omega}_{\text{ss}} t$ where $\bar{\Omega}_{\text{ss}}$ is the mean of the dissipation function when the system is in the steady state. The corresponding symmetry relation satisfied by the distribution of rate of performance of work in the steady state is

$$\left(\frac{1}{t} \ln \left(\frac{P(\dot{W}_{\text{ss}}^t)}{P(-\dot{W}_{\text{ss}}^t)} \right) \right) \approx \frac{\dot{W}_{\text{ss}}^t}{kT}, \tag{95}$$

where \dot{W}_{ss}^t is the mean rate of performance of work on the system over an interval t when the system is in a steady state: there will be deviations from the expected constant rate. This result has been designated the steady state fluctuation relation.

The development of results starting from the dissipation function appears more restrictive and less physically motivated than the development starting from ΔS_{tot} . Only in certain circumstances is the dissipation function, or its time integral, related to a thermodynamic property such as the work done. Nevertheless, it has been introduced here in order to prepare for a description in section 8 of the dynamics of systems in thermal contact with a heat bath that uses deterministic instead of stochastic equations of motion. Historically, this approach provided the initial insights that lead to the development of fluctuation relations.

Another point to be made when introducing the dissipation function is that it demonstrates that there might be system properties that on average grow with time but nevertheless do not correspond to entropy production in all circumstances. For example, in order to satisfy ourselves that ΔS_{tot} represented entropy production, the integral fluctuation relation and the consequence $\langle \Delta S_{\text{tot}} \rangle \geq 0$ was not sufficient: it was necessary to demonstrate that it entered into the appropriate relations with thermodynamic process variables such as equation (64).

7 Examples of the fluctuation relations in action

The development of theoretical results of the kind we have seen so far is all very well, but their meaning is best appreciated by considering examples, which we do in this section. In the first two cases, we shall consider overdamped stochastic dynamics, such that the velocities are always in an equilibrium Maxwell-Boltzmann distribution and never enter into consideration for entropy production. And we shall focus on the harmonic oscillator, since we understand its properties well. The only drawback of the harmonic oscillator is that it is a rather special case and some of its properties are not general. In the third case we allow the velocity distribution

of a particle to be driven out of equilibrium by a heat bath with a time-dependent temperature. In the final case we consider an adiabatic work process.

7.1 Harmonic oscillator subject to a step change in spring constant

Let us start by considering a 1-d classical harmonic oscillator that is subject to an instantaneous step change in spring constant from κ_0 to κ_1 at $t = 0$. Until this time, the system is in thermal equilibrium at temperature T , and it remains in contact with a thermal bath at this temperature after the shift in spring constant. We are interested in the statistics of the entropy change associated with the process.

Starting from equation (59) we write

$$\Delta S_{\text{tot}} = -\frac{\kappa_1}{2T} (x_1^2 - x_0^2) + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_1)} \right), \quad (96)$$

where we use notation $x_1 = x(t)$ and $x_0 = x(0)$. We employ the initial canonical distribution

$$P_{\text{start}}(x_0) = \left(\frac{\kappa_0}{2\pi kT} \right)^{1/2} \exp \left(-\frac{\kappa_0 x_0^2}{2kT} \right), \quad (97)$$

and the final distribution is

$$P_{\text{end}}(x_1) = \int_{-\infty}^{\infty} dx_0 \mathcal{P}_{\kappa_1}(x_1, t | x_0, 0) P_{\text{start}}(x_0), \quad (98)$$

with the transition probability density given by equation (58). Hence

$$\begin{aligned} P_{\text{end}}(x_1) &= \int_{-\infty}^{\infty} dx_0 \left(\frac{\kappa_1}{2\pi kT(1 - e^{-2\kappa_1 t/(m\gamma)})} \right)^{1/2} \exp \left(-\frac{\kappa_1 (x_1 - x_0 e^{-\kappa_1 t/(m\gamma)})^2}{2kT(1 - e^{-2\kappa_1 t/(m\gamma)})} \right) \left(\frac{\kappa_0}{2\pi kT} \right)^{1/2} \exp \left(-\frac{\kappa_0 x_0^2}{2kT} \right) \\ &= \left(\frac{\tilde{\kappa}(t)}{2\pi kT} \right)^{1/2} \exp \left(-\frac{\tilde{\kappa}(t) x_1^2}{2kT} \right), \end{aligned} \quad (99)$$

with

$$\tilde{\kappa}(t) = \frac{\kappa_0 \kappa_1}{\kappa_0 + e^{-2\kappa_1 t/(m\gamma)} (\kappa_1 - \kappa_0)}, \quad (100)$$

such that P_{end} is always gaussian. The coefficient $\tilde{\kappa}(t)$ evolves monotonically from κ_0 at $t = 0$ to κ_1 as $t \rightarrow \infty$. Then

$$\Delta S_{\text{tot}}(x_1, x_0, t) = -\frac{\kappa_1}{2T} (x_1^2 - x_0^2) + \frac{k}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}(t)} \right) - \frac{\kappa_0 x_0^2}{2T} + \frac{\tilde{\kappa}(t) x_1^2}{2T}, \quad (101)$$

is the entropy production associated with a trajectory that begins at x_0 at $t = 0$ and ends at x_1 at time t , and is not specified in between. We can average this over the probability distribution for such a trajectory to get

$$\begin{aligned} \langle \Delta S_{\text{tot}} \rangle_t &= \int dx_0 dx_1 \mathcal{P}_{\kappa_1}(x_1, t | x_0, 0) P_{\text{start}}(x_0) \Delta S_{\text{tot}}(x_1, x_0, t) \\ &= k \left(-\frac{1}{2} \frac{\kappa_1}{\tilde{\kappa}(t)} + \frac{1}{2} \frac{\kappa_1}{\kappa_0} + \frac{1}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}(t)} \right) - \frac{1}{2} + \frac{1}{2} \right) \\ &= \frac{k}{2} \left(\frac{\kappa_1}{\kappa_0} - \frac{\kappa_1}{\tilde{\kappa}(t)} + \ln \left(\frac{\kappa_0}{\tilde{\kappa}(t)} \right) \right), \end{aligned} \quad (102)$$

making full use of the separation of ΔS_{tot} into quadratic terms, and the gaussian character of each distribution. At $t = 0$ this is zero, and as $t \rightarrow \infty$ the limit is

$$\langle \Delta S_{\text{tot}} \rangle_{\infty} = \frac{k}{2} \left(\frac{\kappa_1}{\kappa_0} - 1 - \ln \left(\frac{\kappa_1}{\kappa_0} \right) \right), \quad (103)$$

which is positive since $\ln z \leq z - 1$ for all z . Furthermore,

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \frac{k}{2\tilde{\kappa}^2} \frac{d\tilde{\kappa}}{dt} (\kappa_1 - \tilde{\kappa}), \quad (104)$$

and it is easy to convince oneself that this is positive at all times during the evolution, irrespective of the values of κ_1 and κ_0 . If $\kappa_1 > \kappa_0$ then $\tilde{\kappa}$ increases and all factors on the right hand side of equation (104) are positive. If

$\kappa_1 < \kappa_0$ then $\tilde{\kappa}$ decreases but always remains greater than κ_1 and the total entropy production is still positive as the relaxation process proceeds.

The work done on the system is simply the input of potential energy at the shift in spring constant:

$$\Delta W(x_1, x_0, t) = \frac{1}{2} (\kappa_1 - \kappa_0) x_0^2, \quad (105)$$

and so the mean work performed up until time t (indeed any $t > 0$) is

$$\langle W \rangle_t = \frac{kT}{2} \left(\frac{\kappa_1}{\kappa_0} - 1 \right), \quad (106)$$

which is greater than $\Delta F = (kT/2) \ln(\kappa_1/\kappa_0)$. The mean dissipative work is

$$\langle W_d \rangle = \langle W \rangle_t - \Delta F = \frac{kT}{2} \left(\frac{\kappa_1}{\kappa_0} - 1 - \ln \left(\frac{\kappa_1}{\kappa_0} \right) \right), \quad (107)$$

and this equals the mean entropy generated as $t \rightarrow \infty$ derived in equation (103), since the system started in equilibrium. More specifically, let us verify the Jarzynski equality, which is expected to apply since the system started the process in equilibrium:

$$\begin{aligned} \langle \exp(-\Delta W/kT) \rangle &= \int dx_0 P_{\text{start}}(x_0) \exp(-(\kappa_1 - \kappa_0) x_0^2 / 2kT) \\ &= (\kappa_0/\kappa_1)^{1/2} = \exp(-\Delta F/kT), \end{aligned} \quad (108)$$

since $F = \frac{1}{2} \ln \kappa + \text{constants}$ for the harmonic oscillator.

We have already demonstrated the result $\langle \Delta S_{\text{tot}} \rangle \geq 0$ that emerges from the integral fluctuation relation. It is important to demonstrate that the latter is satisfied. We consider

$$\begin{aligned} \langle \exp(-\Delta S_{\text{tot}}/k) \rangle &= \langle \exp \left(\frac{\kappa_1}{2kT} (x_1^2 - x_0^2) - \frac{1}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}} \right) + \frac{\kappa_0 x_0^2}{2kT} - \frac{\tilde{\kappa} x_1^2}{2kT} \right) \rangle \\ &= \int dx_1 dx_0 \mathcal{P}_{\kappa_1}(x_1, t | x_0, 0) P_{\text{start}}(x_0) \exp \left(\frac{\kappa_1}{2kT} (x_1^2 - x_0^2) - \frac{1}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}} \right) + \frac{\kappa_0 x_0^2}{2kT} - \frac{\tilde{\kappa} x_1^2}{2kT} \right) \\ &= \left(\frac{\tilde{\kappa}}{\kappa_0} \right)^{1/2} \int dx_1 dx_0 \left(\frac{\kappa_1}{2\pi kT (1 - e^{-2\kappa_1 t/(m\gamma)})} \right)^{1/2} \exp \left(-\frac{\kappa_1 (x_1 - x_0 e^{-\kappa_1 t/(m\gamma)})^2}{2kT (1 - e^{-2\kappa_1 t/(m\gamma)})} \right) \\ &\quad \times \left(\frac{\kappa_0}{2\pi kT} \right)^{1/2} \exp \left(-\frac{\kappa_0 x_0^2}{2kT} \right) \exp \left(\frac{\kappa_1}{2kT} (x_1^2 - x_0^2) + \frac{\kappa_0 x_0^2}{2kT} - \frac{\tilde{\kappa} x_1^2}{2kT} \right) \\ &= 1, \end{aligned} \quad (109)$$

which is an integral best checked using Mathematica.

It is even more challenging to calculate the probability distribution of values taken by ΔS_{tot} in this process and in its reverse (where a system starts at equilibrium with spring constant κ_1 and it changes at $t = 0$ to κ_0), and to show that it satisfies a detailed fluctuation relation or Crooks relation. This can be done, but actually it is easier to demonstrate numerically by sampling the form of ΔS_{tot} or of ΔW by generating x_1 and x_0 from the joint probability distribution.

7.2 Smoothly squeezed harmonic oscillator

Now let us consider a process where work is performed isothermally on a particle by the continuous variation of the spring constant. We employ equation (60) and write

$$\begin{aligned} \Delta S_{\text{tot}} &= \frac{1}{T} \int_0^t dt' \left(\frac{\partial \phi(x_{t'}, t')}{\partial t'} - \frac{d\phi(x_{t'}, t')}{dt'} \right) + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_t)} \right) \\ &= \frac{1}{T} \int_0^t \frac{1}{2} \dot{\kappa} x_{t'}^2 dt' - \frac{1}{2T} \kappa(t) x_t^2 + \frac{1}{2T} \kappa_0 x_0^2 + k \ln \left(\frac{P_{\text{start}}(x_0)}{P_{\text{end}}(x_t)} \right). \end{aligned} \quad (110)$$

Once again, for convenience, we assume the initial state to be in canonical equilibrium. The evolving distribution P satisfies the Fokker-Planck equation for an Ornstein-Uhlenbeck process, namely

$$\frac{\partial P}{\partial t} = \frac{\kappa(t)}{m\gamma} \frac{\partial (xP)}{\partial x} + \frac{kT}{m\gamma} \frac{\partial^2 P}{\partial x^2}. \quad (111)$$

If P is initially canonical then it retains its gaussian form. This can be established by inserting a solution of the form

$$P_{\text{end}}(x_t) = P(x_t, t) = \left(\frac{\tilde{\kappa}(t)}{2\pi kT} \right)^{1/2} \exp \left(-\frac{\tilde{\kappa}(t)x_t^2}{2kT} \right), \quad (112)$$

with a ‘retarded spring constant’ $\tilde{\kappa}(t)$ that satisfies

$$\frac{d\tilde{\kappa}}{dt} = -\frac{2}{m\gamma} \tilde{\kappa} (\tilde{\kappa} - \kappa), \quad (113)$$

and with initial condition $\tilde{\kappa}(0) = \kappa_0$. We can solve for $\tilde{\kappa}$: write $z = \tilde{\kappa}^{-1}$ such that

$$\frac{dz}{dt} = \frac{2}{m\gamma} (1 - \kappa z). \quad (114)$$

This has integrating factor solution

$$z(t) \exp \left(\frac{2}{m\gamma} \int_0^t \kappa(t') dt' \right) = z(0) + \int_0^t \exp \left(\frac{2}{m\gamma} \int_0^{t'} \kappa(t'') dt'' \right) \frac{2}{m\gamma} dt', \quad (115)$$

or equivalently

$$\frac{1}{\tilde{\kappa}(t)} = \frac{1}{\kappa(0)} \exp \left(-\frac{2}{m\gamma} \int_0^t \kappa(t') dt' \right) + \frac{2}{m\gamma} \int_0^t \exp \left(-\frac{2}{m\gamma} \int_{t'}^t \kappa(t'') dt'' \right) dt'. \quad (116)$$

Returning to the entropy production, we now write

$$\Delta S_{\text{tot}} = \frac{1}{T} \int_0^t \frac{1}{2} \dot{\kappa} x_{t'}^2 dt' - \frac{1}{2T} \kappa(t) x_t^2 + \frac{1}{2T} \kappa_0 x_0^2 + \frac{k}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}(t)} \right) - \frac{\kappa_0 x_0^2}{2T} + \frac{\tilde{\kappa}(t) x_t^2}{2T}, \quad (117)$$

and we also have

$$\Delta W = \int_0^t \frac{1}{2} \dot{\kappa} x_{t'}^2 dt'. \quad (118)$$

We can investigate the statistics of these quantities:

$$\langle \Delta W \rangle = \int_0^t \frac{1}{2} \dot{\kappa} \langle x_{t'}^2 \rangle dt' = \int_0^t \frac{1}{2} \dot{\kappa} \frac{kT}{\tilde{\kappa}} dt', \quad (119)$$

and using $W_d = \Delta W - \Delta F$ the rate of performance of dissipative work is

$$\frac{d\langle W_d \rangle}{dt} = \frac{\dot{\kappa} kT}{2\tilde{\kappa}} - \frac{dF(\kappa(t))}{dt} = \frac{\dot{\kappa} kT}{2\tilde{\kappa}} - \frac{\dot{\kappa} kT}{2\kappa} = \frac{kT}{2} \dot{\kappa} \left(\frac{1}{\tilde{\kappa}} - \frac{1}{\kappa} \right). \quad (120)$$

Whilst the positivity of $\langle W_d \rangle$ is assured for this process, as a consequence of the Jarzynski equation and the initial equilibrium condition, the rate of its evolution can be both positive and negative, according to this result.

The expectation value for total entropy production in equation (117), on the other hand, is

$$\langle \Delta S_{\text{tot}} \rangle = \frac{1}{T} \int_0^t \frac{1}{2} \dot{\kappa} \frac{kT}{\tilde{\kappa}} dt' - \frac{1}{2T} \kappa(t) \frac{kT}{\tilde{\kappa}} + \frac{1}{2T} kT + \frac{k}{2} \ln \left(\frac{\kappa_0}{\tilde{\kappa}(t)} \right) - \frac{kT}{2T} + \frac{kT}{2T}, \quad (121)$$

and the rate of change of this quantity is

$$\begin{aligned} \frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} &= \frac{\dot{\kappa} k}{2\tilde{\kappa}} - \frac{k}{2} \left(\frac{\dot{\kappa}}{\tilde{\kappa}} - \frac{\kappa}{\tilde{\kappa}^2} \dot{\kappa} \right) - \frac{k}{2} \frac{\dot{\kappa}}{\tilde{\kappa}} \\ &= \frac{k}{2} \frac{\dot{\kappa}}{\tilde{\kappa}} \frac{(\kappa - \tilde{\kappa})}{\tilde{\kappa}^2} \\ &= \frac{k}{m\gamma} \frac{(\kappa - \tilde{\kappa})^2}{\tilde{\kappa}}. \end{aligned} \quad (122)$$

The monotonic increase in entropy with time is explicit.

It is of more interest to verify that detailed fluctuation relations hold. Analytic demonstration based upon equation (117) and the probability density for a particular trajectory throughout the entire period is challenging, but a numerical approach based upon generating sample trajectories is feasible.

7.3 Heated particle

Our studies of the harmonic oscillator enable us to consider the process of heat transfer to a free particle and to illustrate entropy production without dissipative work. Consider the Langevin equation for the velocity of a particle moving in one dimension in contact with a time-dependent heat bath:

$$\dot{v} = -\gamma v + \left(\frac{2kT(t)\gamma}{m} \right)^{1/2} \xi(t), \quad (123)$$

where $\xi(t)$ is white noise. The equivalent Fokker-Planck equation is

$$\frac{\partial P(v, t)}{\partial t} = \frac{\partial (\gamma v P(v, t))}{\partial v} + \frac{kT(t)\gamma}{m} \frac{\partial^2 P(v, t)}{\partial v^2}, \quad (124)$$

and the transition probability density for an interval $t_{i-1} \rightarrow t_i$, during which the bath temperature is a constant $T(t_i) = T_i$, is

$$\mathcal{P}_{T_i}(v_i, t_i | v_{i-1}, t_{i-1}) = \left(\frac{m}{2\pi k T_i (1 - e^{-2\gamma(t_i - t_{i-1})})} \right)^{1/2} \exp \left(-\frac{m(v_i - v_{i-1} e^{-\gamma(t_i - t_{i-1})})^2}{2k T_i (1 - e^{-2\gamma(t_i - t_{i-1})})} \right), \quad (125)$$

using the notation $v_i = v(t_i)$. We wish to calculate the entropy generated:

$$\Delta S_{\text{tot}} = k \ln \left(\frac{\mathcal{P}_{T_n}(v_n, t_n | v_{n-1}, t_{n-1}) \mathcal{P}_{T_{n-1}}(v_{n-1}, t_{n-1} | v_{n-2}, t_{n-2}) \cdots \mathcal{P}_{T_1}(v_1, t_1 | v_0, t_0) P_{\text{start}}(v_0)}{\mathcal{P}_{T_1}(v_0, t_n | v_1, t_{n-1}) \cdots \mathcal{P}_{T_{n-1}}(v_{n-2}, t_2 | v_{n-1}, t_1) \mathcal{P}_{T_n}(v_{n-1}, t_1 | v_n, t_0) P_{\text{end}}(v_n)} \right), \quad (126)$$

associated with a time-series (v_0, v_1, \dots, v_n) in velocity. The analogy with equation (52) for the isothermally squeezed harmonic oscillator suggests that most of our analysis is already done.

We can use those earlier results to show that

$$\begin{aligned} \Delta S_{\text{tot}} &= \sum_{i=1}^n \left(\frac{m(v_{i-1} - v_i e^{-\gamma\tau})^2}{2T_i(1 - e^{-2\gamma\tau})} - \frac{m(v_i - v_{i-1} e^{-\gamma\tau})^2}{2T_i(1 - e^{-2\gamma\tau})} \right) + k \ln \left(\frac{P_{\text{start}}(v_0)}{P_{\text{end}}(v_n)} \right) \\ &= - \sum_{i=1}^n \frac{m}{2T_i} (v_i^2 - v_{i-1}^2) + k \ln \left(\frac{P_{\text{start}}(v_0)}{P_{\text{end}}(v_n)} \right) \\ &= - \sum_{i=1}^n \frac{\Delta Q}{T_i} + k \ln \left(\frac{P_{\text{start}}(v_0)}{P_{\text{end}}(v_n)} \right) \\ &= - \int_0^t \frac{dQ}{T(t')} + k \ln \left(\frac{P_{\text{start}}(v_0)}{P_{\text{end}}(v_t)} \right), \end{aligned} \quad (127)$$

where the change in kinetic energy of the particle over time period $t_{i-1} \rightarrow t_i$, given by $\Delta E = m(v_i^2 - v_{i-1}^2)/2$, is provided by heat transfer from the bath, since no mechanical work has been performed, in which case we can use the notation ΔQ instead of ΔE . In the last line, we have gone over to a continuum description of temperature variation.

Now we assume that $P_{\text{start}}(v_0)$ is a canonical distribution and solve the Fokker-Planck equation (124) subject to this initial condition. We find

$$P_{\text{end}}(v_t) = P(v_t, t) = \left(\frac{m}{2\pi k \tilde{T}(t)} \right)^{1/2} \exp \left(-\frac{mv_t^2}{2k \tilde{T}(t)} \right), \quad (128)$$

with the retarded temperature $\tilde{T}(t)$ given by the solution to

$$\frac{d\tilde{T}}{dt} = -2\gamma (\tilde{T} - T), \quad (129)$$

with initial condition $\tilde{T}(0) = T_0$. Hence

$$\Delta S_{\text{tot}} = - \int_0^t \frac{dQ}{T(t')} + \frac{k}{2} \ln \left(\frac{\tilde{T}(t)}{T_0} \right) + \frac{mv_t^2}{2\tilde{T}(t)} - \frac{mv_0^2}{2T_0}. \quad (130)$$

The counterpart to the dissipative work for this thermal process is what we might call the Clausius entropy production, namely

$$S_c = - \int_0^t \frac{dQ}{T(t')} + \Delta S_{\text{Gibbs}}, \quad (131)$$

where ΔS_{Gibbs} is the difference in Gibbs thermodynamic entropy of the system with reference to the initial and final temperatures defined by the heating process. For the single particle, we use the result

$$S_{\text{Gibbs}}(T) = -k \int P(v) \ln P(v) dv = -k \int P(v) \left(\frac{1}{2} \ln \left(\frac{m}{2\pi kT} \right) - \frac{mv^2}{2kT} \right) dv = -\frac{k}{2} \ln \left(\frac{m}{2\pi kT} \right) + \frac{k}{2}, \quad (132)$$

so

$$\Delta S_{\text{Gibbs}} = \frac{k}{2} \ln \left(\frac{T(t)}{T_0} \right), \quad (133)$$

and

$$\Delta S_{\text{tot}} = S_c - \frac{k}{2} \ln \left(\frac{T(t)}{T_0} \right) + \frac{k}{2} \ln \left(\frac{\tilde{T}(t)}{T_0} \right) + \frac{mv_t^2}{2\tilde{T}(t)} - \frac{mv_0^2}{2T_0}. \quad (134)$$

The average of both sides over the initial and final velocities gives

$$\langle \Delta S_{\text{tot}} \rangle = \langle S_c \rangle - \frac{k}{2} \ln \left(\frac{T(t)}{T_0} \right) + \frac{k}{2} \ln \left(\frac{\tilde{T}(t)}{T_0} \right) = \langle S_c \rangle - \frac{k}{2} \ln \left(\frac{T(t)}{\tilde{T}(t)} \right). \quad (135)$$

Clearly, if the process ends in equilibrium, with $\tilde{T}(t) = T(t)$, then the expectation value of the Clausius entropy production equals that of the total entropy production. If not, then the two quantities differ, just as we saw the mean dissipative work differed from the mean total entropy production, divided by temperature, for an isothermal work process that did not end in equilibrium.

Clearly

$$\frac{d\langle \Delta S_{\text{tot}} \rangle}{dt} = \frac{d\langle S_c \rangle}{dt} - \frac{k}{2} \left(\frac{1}{T} \frac{dT}{dt} - \frac{1}{\tilde{T}} \frac{d\tilde{T}}{dt} \right) = \frac{d\langle S_c \rangle}{dt} - \frac{k}{2} \left(\frac{1}{T} \frac{dT}{dt} + \frac{2\gamma(\tilde{T} - T)}{T} \right), \quad (136)$$

and since the left hand side is intrinsically positive, whilst dT/dt could be as negative as we like, the rate of change of Clausius entropy production could be negative, even though the integrated mean change in Clausius entropy production must be positive. Once again this is an analogue of the behaviour of the mean dissipative work. In addition, one can investigate the integral fluctuation relation numerically.

7.4 Oscillator subject to adiabatic work

Verifying the detailed fluctuation relation proves to be a challenging task, as we have found. However, one example where a relatively simple demonstration can be performed is the case of an oscillator, starting in canonical equilibrium, that undergoes a process consisting of adiabatic work caused by the change in spring constant, followed by heat exchange with a heat bath to bring the system once again into equilibrium. If the work process were time-symmetric about its mid point, then the probability distribution of work done should satisfy a detailed fluctuation relation. And we can calculate this distribution by considering the adiabatic work segment of the process alone.

The simplest suitable work process is a step change in spring constant from κ_0 to κ_1 at $t = 0$, and a return shift from κ_1 to κ_0 at $t = \tau$. The work done is

$$\Delta W = \frac{1}{2}(\kappa_1 - \kappa_0)x_0^2 + \frac{1}{2}(\kappa_0 - \kappa_1)x_\tau^2, \quad (137)$$

and after some analysis one can prove that the distribution of work done takes the form

$$P(\Delta W) = \exp \left(\frac{\Delta W}{2kT} \right) K_0 \left(\frac{|\Delta W|}{2kT} \left(1 + \frac{2kT}{|\langle \Delta W \rangle|} \right) \right), \quad (138)$$

where the mean work done is

$$\langle \Delta W \rangle = kT \left(\frac{1}{2} \left(\frac{\kappa_0}{\kappa_1} + \frac{\kappa_1}{\kappa_0} \right) - 1 \right) \sin^2(\omega_1 \tau). \quad (139)$$

The factor involving the modified Bessel function of the second kind K_0 in equation (138) is clearly symmetric in ΔW . The exponential prefactor then ensures that the detailed fluctuation theorem

$$P(\Delta W) = \exp(\Delta W/kT)P(-\Delta W) \quad (140)$$

is satisfied.

8 Deterministic thermal systems

Historically, it was the consideration of deterministic dynamical systems that provided the insights for the development of fluctuation relations. Such systems can be evolved in time using equations of motion without noise designed to maintain a constraint over part of the system. Such a constraint might be to keep fixed the total kinetic energy of some or all of the particles in the system. This makes the equations non-Hamiltonian in form and non-conservative of energy. The aim of seminal studies by Evans and Searles was to develop a system function that would display a tendency to grow with time, and which could be used to represent entropy production. The H -theorem of Boltzmann described earlier was a similar attempt in this direction. One difficulty that emerged at that time is that a quantity that increases, on average, as time progresses starting from a time t_0 , must have decreased on average in the period prior to t_0 . Boltzmann struggled to make sense of how this could be connected with thermodynamic entropy. We shall confront this same problem.

We shall also have to confront a difficulty that will be obvious in view of our repeated statements to the effect that the expected increase in total entropy is a consequence of the increasing uncertainty in system properties as time progresses. By their very nature, deterministic equations do not generate additional uncertainty. The configuration of the system at a time t is precisely specified if the configuration at t_0 is known. If the latter is only specified through a probability distribution, and is therefore uncertain, it remains the case that all future and past configurations arising from each starting configuration are determined, and uncertainty is therefore not changed by the passage of time. It is for that reason that the Gibbs entropy, the commodification of the uncertainty associated with a system, is a constant of the motion under Hamilton's equations. Something other than the increase in uncertainty will have to emerge in a deterministic framework if it is to represent entropy.

We shall not attempt to discuss the approach in full detail. Our starting point is that the system is to be described using a probability density over its dynamical variables Γ , which include the positions and velocities of all particles. In our discussions of stochastic dynamics in the overdamped limit, we were able to focus solely on particle position, but now we need to give attention to the velocity coordinate as well. It will introduce some extra considerations.

An initial probability density $P(\Gamma, 0)$ evolves into a density $P(\Gamma, t)$. Furthermore, the starting point of a trajectory Γ_0 is linked uniquely to a terminating point Γ_t . It may then be shown that

$$P(\Gamma_t, t) = P(\Gamma_0, 0) \exp\left(-\frac{\Delta Q}{kT}\right), \quad (141)$$

where we consider the system to be thermalised to a time-independent temperature T , and ΔQ is the heat transferred to the system during the process. For a system without constraint, and hence isolated, the heat transfer is zero, and the resulting $P(\Gamma_t, t) = P(\Gamma_0, 0)$ is an expression of Liouville's theorem: the probability density is conserved along any trajectory followed by the system.

We now reconsider the dissipation function previously described in section 6.5. We have

$$\bar{\Omega}(\Gamma_0)t = -\frac{\Delta Q}{kT} + \ln\left(\frac{P(\Gamma_0, 0)}{P(\Gamma_t, 0)}\right) = \ln\left(\frac{P(\Gamma_t, t)}{P(\Gamma_t, 0)}\right), \quad (142)$$

where the mean dissipation function $\bar{\Omega}$ for the trajectory of duration t starting from point Γ_0 is recognised to be a function of Γ_0 , and not the entire trajectory, as would be the case for stochastic dynamics. As before, we can investigate the statistical properties of $\bar{\Omega}t$. We have

$$\begin{aligned} \langle \exp(-\bar{\Omega}t) \rangle &= \int d\Gamma_0 P(\Gamma_0, 0) \exp(-\bar{\Omega}t) \\ &= \int d\Gamma_t P(\Gamma_t, t) \exp(-\bar{\Omega}t) \\ &= \int d\Gamma_t P(\Gamma_t, 0) = 1, \end{aligned} \quad (143)$$

having imposed a conservation law of probability $d\Gamma_0 P(\Gamma_0, 0) = d\Gamma_t P(\Gamma_t, t)$ so $\langle \bar{\Omega} \rangle \geq 0$. This is encouraging since it reproduces one of the results obtained using stochastic dynamics. Notice, however, that it applies for

negative time as well as positive. The time dependence of this proposed measure of irreversibility respects the time reversal symmetry of the evolution equations, as it should.

Furthermore, for a protocol that is time-symmetric about its mid point, it must be the case that $\bar{\Omega}(\Gamma_0) = -\bar{\Omega}(\Gamma_t^*)$, where Γ_t^* is related to the point Γ_t by an inversion of the velocity coordinates. The point Γ_t^* evolves to the point Γ_0^* (the velocity reversed counterpart to the point Γ_0) over the course of the time-symmetric protocol of duration t . Hence, in the usual fashion,

$$\begin{aligned}
P(\bar{\Omega}') &= \int d\Gamma_0 P(\Gamma_0, 0) \delta(\bar{\Omega}(\Gamma_0) - \bar{\Omega}') \\
&= \int d\Gamma_t P(\Gamma_t, t) \delta(\bar{\Omega}(\Gamma_0) - \bar{\Omega}') \\
&= \int d\Gamma_t P(\Gamma_t, t) \exp(\bar{\Omega}(\Gamma_0)t) \frac{P(\Gamma_t, 0)}{P(\Gamma_t, t)} \delta(\bar{\Omega}(\Gamma_0) - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) \int d\Gamma_t P(\Gamma_t, 0) \delta(\bar{\Omega}(\Gamma_0) - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) \int d\Gamma_t P(\Gamma_t, 0) \delta(-\bar{\Omega}(\Gamma_t^*) - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) \int d\Gamma_t^* P(\Gamma_t^*, 0) \delta(-\bar{\Omega}(\Gamma_t^*) - \bar{\Omega}') \\
&= \exp(\bar{\Omega}'t) P(-\bar{\Omega}'),
\end{aligned} \tag{144}$$

where we note that the Jacobian for the transformation of the integration variable from Γ_t to Γ_t^* is unity and assume that the initial distribution is symmetric in the velocity coordinates: $P(\Gamma_t, 0) = P(\Gamma_t^*, 0)$. Apart from this last condition, these are the very criteria that were used to prove the transient fluctuation relation within the framework of stochastic dynamics.

If we assume that the initial distribution is in canonical equilibrium, such that $P(\Gamma, 0) \propto \exp(-H(\Gamma, 0)/kT)$, where $H(\Gamma, 0)$ is the Hamiltonian at $t = 0$, then we find from equation (142) that

$$\bar{\Omega}(\Gamma_0)t = -\frac{\Delta Q}{kT} + \frac{1}{kT} (H(\Gamma_t, 0) - H(\Gamma_0, 0)), \tag{145}$$

and if the Hamiltonian at time t returns to its functional form at $t = 0$, then $H(\Gamma_t, 0) = H(\Gamma_t, t)$ and we get

$$\bar{\Omega}(\Gamma_0)t = -\frac{\Delta Q}{kT} + \frac{1}{kT} (H(\Gamma_t, t) - H(\Gamma_0, 0)) = -\frac{\Delta Q}{kT} + \frac{\Delta E(\Gamma_0)}{kT} = \frac{\Delta W(\Gamma_0)}{kT}, \tag{146}$$

and the integrated dissipation function is proportional to the work performed on the system during the process, as found in equation (92). We deduce that the expectation value of the work is positive, and that its probability distribution satisfies a detailed fluctuation relation.

Deterministic methods are capable of deriving other statistical results involving the work performed on a system, including the Jarzynski equation and the Crooks relation. However, it seems that a parallel development of the statistics of ΔS_{tot} is not possible. The fundamental problem is revealed if we try to construct the deterministic counterpart to equation (63):

$$\Delta S_{\text{tot}} = -\frac{\Delta Q}{T} - k \ln \left(\frac{P(\Gamma_t, t)}{P(\Gamma_0, 0)} \right), \tag{147}$$

but using equation (141) this is identically zero. As might have been expected, uncertainty is not changing under deterministic dynamics, and entropy, in the form that we have chosen to define it, is not produced. What we do obtain from a treatment using deterministic dynamics, however, are relationships involving the statistics of work performed, and heat transferred, that correspond to those obtained from a treatment using stochastic dynamics. The conditions for the validity of the results are similar. Obviously this ought to be the case if both treatments capture the essence of the dynamics of a system under some kind of thermal constraint.

9 Final remarks

The aim of these notes was to explore the origin, application and limitations of fluctuation relations. We have done this largely within the framework of stochastic dynamics; in fact a specific kind of stochastic dynamics with white noise and often employing the overdamped limit. There is good reason to look at such special cases,

since the derivations are easier and the physical interpretation of the phenomena is more straightforward. It is in the analysis of explicit examples where learning is often to be found. Nevertheless, the results can be extended to other situations, though we have not proved this.

The fluctuation relations provide information about the statistics of measurable quantities associated with thermodynamic processes, in particular the mechanical work done upon, or the heat transferred to a system in contact with a heat bath. In the thermodynamic limit, the statistics are simple: there are negligible deviations from the mean, and work and heat transfers appear to be deterministic. But for finite size systems, there are fluctuations, and the statistics of these will satisfy one or more fluctuation relation. These can be very specific requirements, for example that the probability of a fluctuation with negative dissipative work is related to the probability of a fluctuation with positive dissipative work in the reversed process. Or the outcome can take the form of an inequality that merely demonstrates that the mean dissipative work is positive.

The core concept in the analysis is entropy production, defined not as the difference in the properties of an initial and a final ensemble of systems, but as a quantity associated with a particular evolution of a single system under the influence of the stochastic dynamics. It is not restricted to changes that are positive or zero: it can go negative, and the probability of such an excursion might be described by a fluctuation relation. As a consequence, however, its expectation value over all possible realisations of the dynamics is rigorously non-negative. Once again, in the thermodynamic limit the entropy production appears to behave deterministically since all fluctuations disappear.

Entropy production in the framework we have employed is associated with the increase in uncertainty with time brought about by the stochastic dynamics, which are themselves a consequence, and representation, of the lack of a detailed initial specification of the system, and of its environment. Intuition suggests that initial uncertainty usually grows with time, and entropy production is a commodification of this concept. But it turns out that it can be related, in certain circumstances, to heat and work transfers, and so the growth of uncertainty can sometimes be measured in terms of thermodynamic process variables.

It is ironic that entropy evolves quasi-deterministically in the thermodynamic limit, since we might say that it is fundamentally reliant upon stochastic dynamical behaviour. But for large systems it evolves according to equations that lack noise terms. This raises the fundamental concern, noticed well over a century ago, that entropy satisfies evolution equations that break time reversal symmetry, whilst every fundamental physical quantity of a system must evolve according to time reversal symmetric equations. If only thermodynamics had initially been developed to describe small systems! Fluctuations would have been obvious and the quantities in question, such as heat and work, would have been recognised as stochastic variables, not deterministic. Then, perhaps, much of the controversy over the strange violation of time reversal symmetry might have been avoided.

Fluctuations in predicted future behaviour are inevitable if we prepare a system using macroscopic constraints, without taking care to specify the smaller details. Uncertainty, present at the start of the process, propagates into the future and usually grows. The total entropy, a representation of uncertainty, usually goes up as time progresses, but occasionally goes down. In contrast, when we investigate a system at the microscopic level, the usual strategy is to try to constrain all details in the initial state and specify the dynamics unambiguously. Then uncertainty does not change as a consequence of the evolution and successful model equations of motion of system properties ought to satisfy time reversal symmetry. There is no entropy production if there is no prospect for an increase in uncertainty.

The nagging question remains how it is that the entropy production seems to depend upon how carefully one measures the initial state of a system, or the detailed interactions that it experiences during its processing. How can the power of my microscope, and the use I make of it before compressing a gas, affect the entropy change and hence dissipative work I perform upon compressing it? The key point, I think, is that any restrictions upon the initial uncertainty of configuration, however small, will affect the possible values of work done during the process. Furthermore, if initial information restricts the number of configurations available to a system, this alters the partition function and hence the free energy. The gathering of information about the initial state and the dynamics that operate can ultimately reduce the possible evolution to a case where a uniquely determined initial configuration evolves into a uniquely determined final state. Such an evolution involves no heat transfer, which we associate with uncertain dynamics. Then the change in free energy reduces to a change in energy, which is equal to the unique value of work performed, and the dissipative work then vanishes. Initial information does affect outcomes: it does alter the apparent change in total entropy. The two are intimately linked, perhaps in a way that might surprise the pioneering thermodynamicists who developed the concept of entropy so long ago.

Further reading

The following is a brief selection from the literature in this field:

G. Crooks, J. Stat. Phys. 90 (1998) 1481
 D.J. Evans and D.J. Searles, Adv. Phys. 51 (2002) 1529
 G. Gallavotti and E.G.D. Cohen, J. Stat. Phys. 80 (1995) 931.
 R.J. Harris and G.M. Schütz, J. Stat. Mech. (2007) P07020
 C. Jarzynski, Phys. Rev. Lett. 78 (1997) 2690
 J.L. Lebowitz and H. Spohn, J. Stat. Phys. 95 (1999) 333.
 J. Kurchan, J. Phys. A Math. Gen. 31 (1998) 3719.
 U. Seifert, Eur. Phys. J. B64 (2008) 423.
 K. Sekimoto, Stochastic Energetics, Lecture Notes in Physics, 799, Springer 2010.