3. Stochastic Processes

We learn in kindergarten about the phenomenon of *Brownian motion*, the random jittery movement that a particle suffers when it is placed in a liquid. Famously, it is caused by the constant bombardment due to molecules in the surrounding the liquid. Our goal in this section is to introduce the mathematical formalism that allows us to model such random behaviour.

3.1 The Langevin Equation

In contrast to the previous section, we will here focus on just a single particle. However, this particle will be sitting in a background medium. If we know the force F acting on the particle, its motion is entirely deterministic, governed by

$$m\ddot{\vec{x}} = -\gamma\dot{\vec{x}} + \vec{F} \tag{3.1}$$

In contrast to the previous section, this is not a Hamiltonian system. This is because we have included a friction term with a coefficient γ. This arises due to the viscosity, η, of the surrounding liquid that we met in the previous section. If we model the particle as a sphere of radius a then there is a formula due to Stokes which says $\gamma = 6\pi\eta a$. However, in what follows we shall simply treat γ as a fixed parameter. In the presence of a time independent force, the steady-state solution with $\ddot{\vec{x}} = 0$ is

$$\dot{\vec{x}} = \frac{1}{\gamma}\vec{F}$$

For this reason, the quantity $1/\gamma$ is sometimes referred to as the *mobility*.

Returning to (3.1), for any specified force $\vec{F}$, the path of the particle is fully determined. This is seemingly at odds with the random behaviour observed in Brownian motion. The way in which we reconcile these two points is, hopefully, obvious: in Brownian motion the force $\vec{F}$ that the particle feels is itself random. In fact, we will split the force into two pieces,

$$\vec{F} = -\nabla V + \vec{f}(t)$$

Here V is a fixed background potential in which the particle is moving. Perhaps V arises because the particle is moving in gravity; perhaps because it is attached to a spring. But, either way, there is nothing random about V. In contrast, $\vec{f}(t)$ is the random force that the particle experiences due to all the other atoms in the liquid. It is sometimes referred to as *noise*. The resulting equation is called the *Langevin equation*

$$m\ddot{\vec{x}} = -\gamma\dot{\vec{x}} - \nabla V + \vec{f}(t) \tag{3.2}$$

Although it looks just like an ordinary differential equation, it is, in fact, a different beast known as a *stochastic differential equation*. The reason that it's different is that we don't actually know what $\vec{f}(t)$ is. Yet, somehow, we must solve this equation anyway!

Let's clarify what is meant by this. Suppose that you did know the microscopic force $\vec{f}(t)$ that is experienced by a given particle. Then you could, in principle, go ahead and solve the Langevin equation (3.2). But the next particle that you look at will experience a different force $\vec{f}(t)$ so you'll have to solve (3.2) again. And for the third particle, you'll have to solve it yet again. Clearly, this is going to become tedious. What's more, it's unrealistic to think that we will actually know $\vec{f}(t)$ in any specific case. Instead, we admit that we only know certain crude features of the force $\vec{f}(t)$ such as, for example, its average value. Then we might hope that this is sufficient information to figure out, say, the average value of $\vec{x}(t)$. That is the goal when solving the Langevin equation.

3.1.1 Diffusion in a Very Viscous Fluid

We start by solving the Langevin equation in the case of vanishing potential, $V = 0$. (For an arbitrary potential, the Langevin equation is an unpleasant non-linear stochastic differential equation and is beyond our ambition in this course. However, we will discuss some properties of the case with potential is the following section when we introduce the Fokker-Planck equation). We can simplify the problem even further by considering Brownian motion in a very viscous liquid. In this case, motion is entirely dominated by the friction term in the Langevin equation and we ignore the inertial term, which is tantamount to setting $m = 0$.

When $m = 0$, we're left with a first order equation,

$$\dot{\vec{x}}(t) = \frac{1}{\gamma}\,\vec{f}(t)$$

For any $\vec{f}(t)$, this can be trivially integrated to give

$$\vec{x}(t) = \vec{x}(0) + \frac{1}{\gamma}\int_0^t dt'\ \vec{f}(t') \tag{3.3}$$

At this point, we can't go any further until we specify some of the properties of the noise $\vec{f}(t)$. Our first assumption is that, on average, the noise vanishes at any given time. We will denote averages by $\langle\,\cdot\,\rangle$, so this assumption reads

$$\langle \vec{f}(t)\rangle = 0 \tag{3.4}$$

Taking the average of (3.3) then gives us the result:

$$\langle \vec{x}(t) \rangle = \vec{x}(0)$$

The is deeply unsurprising: if the average noise vanishes, the average position of the particle is simply where we left it to begin with. Nonetheless, it's worth stressing that this doesn't mean that all particles sit where you leave them. It means that, if you drop many identical particles at the origin, $\vec{x}(0) = \vec{0}$, they will all move but their average position — or their centre of mass — will remain at the origin.

We can get more information by looking at the variance of the position,

$$\langle\, (\vec{x}(t) - \vec{x}(0))^2 \,\rangle$$

This will tell us the average spread of the particles. We can derive an expression for the variance by first squaring (3.3) and then taking the average,

$$\langle\, (\vec{x}(t) - \vec{x}(0))^2 \,\rangle = \frac{1}{\gamma^2} \int_0^t dt_1' \int_0^t dt_2' \; \langle\, \vec{f}(t_1') \cdot \vec{f}(t_2') \,\rangle \tag{3.5}$$

In order to compute this, we need to specify more information about the noise, namely its correlation function $\langle\, f_i(t_1) f_j(t_2) \,\rangle$ where we have resorted to index notation, $i, j = 1, 2, 3$ to denote the direction of the force. This is specifying how likely it is that the particle will receive a given kick f_j at time t_2 given that it received a kick f_i at time t_1.

In many cases of interest, including that of Brownian motion, the kicks imparted by the noise are both fast and uncorrelated. Let me explain what this means. Suppose that a given collision between our particle and an atom takes time $\tau_{\rm coll}$. Then if we focus on time scales less than $\tau_{\rm coll}$ then there will clearly be a correlation between the forces imparted on our particle because these forces are due to the same process that's already taking place. (If an atom is coming in from the left, then it's still coming in from the left at a time $t \ll \tau_{\rm coll}$ later). However if we look on time scales $t \gg \tau_{\rm coll}$, the force will be due to a different collision with a different atom. The statement that the noise is uncorrelated means that the force imparted by later collisions knows nothing about earlier collisions. Mathematically, this means

$$\langle\, f_i(t_1) f_j(t_2) \,\rangle = 0 \quad \text{when} \quad t_2 - t_1 \gg \tau_{\rm coll}$$

The statement that the collisions are fast means that we *only* care about time scales $t_2 - t_1 \gg \tau_{\rm coll}$ and so can effectively take the limit $\tau_{\rm coll} \to 0$. However, that doesn't quite mean that we can just ignore this correlation function. Instead, when we take the limit $\tau_{\rm coll} \to 0$, we're left with a delta-function contribution,

$$\langle\, f_i(t_1) f_j(t_2)\,\rangle = 2D\gamma^2\,\delta_{ij}\,\delta(t_2 - t_1) \tag{3.6}$$

Here the factor of γ^2 has been put in for convenience. We will shortly see the interpretation of the coefficient D, which governs the strength of the correlations. Noise which obeys (3.4) and (3.6) is often referred to as *white noise*. It is valid whenever the environment relaxes back down to equilibrium much faster than the system of interest. This guarantees that, although the system is still reeling from the previous kick, the environment remembers nothing of what went before and kicks again, as fresh and random as the first time.

Using this expression for white noise, the variance (3.5) in the position of the particles is

$$\langle\, (\vec{x}(t) - \vec{x}(0))^2\,\rangle = 6D\,t \tag{3.7}$$

This is an important result: the root-mean square of the distance increases as $\sqrt{t}$ with time. This is characteristic behaviour of diffusion. The coefficient D is called the *diffusion constant*. (We put the factor of γ^2 in the correlation function (3.6) so that this equation would come out nicely).

3.1.2 Diffusion in a Less Viscous Liquid

Let's now return to the Langevin equation (3.2) and repeat our analysis, this time retaining the inertia term, so $m \neq 0$. We will still set $V = 0$.

As before, computing average quantities — this time both velocity $\langle\, \dot{\vec{x}}(t)\,\rangle$ and position $\langle\, \vec{x}(t)\,\rangle$ is straightforward and relatively uninteresting. For a given $\vec{f}(t)$, it is not difficult to solve (3.2). After multiplying by an integrating factor $e^{\gamma t/m}$, the equation becomes

$$\frac{d}{dt}\left(\dot{\vec{x}} e^{\gamma t/m}\right) = \frac{1}{m}\vec{f}(t) e^{\gamma t/m}$$

which can be happily integrated to give

$$\dot{\vec{x}}(t) = \dot{\vec{x}}(0) e^{-\gamma t/m} + \frac{1}{m}\int_0^t dt'\ \vec{f}(t')\, e^{\gamma(t'-t)/m} \tag{3.8}$$

We now use the fact that the average of noise vanishes (3.4) to find that the average velocity is simply that of a damped particle in the absence of any noise,

$$\langle \dot{\vec{x}}(t)\rangle = \dot{\vec{x}}(0) e^{-\gamma t/m}$$

Similarly, to determine the average position we have

$$\vec{x}(t) = \vec{x}(0) + \int_0^t dt' \; \dot{\vec{x}}(t') \tag{3.9}$$

From which we get

$$\begin{aligned}\langle \vec{x}(t) \rangle &= \vec{x}(0) + \int_0^t dt' \; \langle \dot{\vec{x}}(t') \rangle \\ &= \vec{x}(0) + \frac{m}{\gamma} \dot{\vec{x}}(0) \left(1 - e^{-\gamma t/m}\right)\end{aligned}$$

Again, this is unsurprising: when the average noise vanishes, the average position of the particle coincides with that of a particle that didn't experience any noise.

Things get more interesting when we look at the expectation values of quadratic quantities. This includes the variance in position $\langle \vec{x}(t) \cdot \vec{x}(t) \rangle$ and velocity $\langle \dot{\vec{x}}(t) \cdot \dot{\vec{x}}(t) \rangle$, but also more general correlation functions in which the two quantities are evaluated at different times. For example, the correlation function $\langle \dot{x}_i(t_1) \dot{x}_j(t_2) \rangle$ tells us information about the velocity of the particle at time t_2 given that we know where its velocity at time t_1. From (3.8), we have the expression,

$$\langle \dot{x}_i(t_1)\dot{x}_j(t_2)\rangle = \langle \dot{x}_i(t_1) \rangle \langle \dot{x}_j(t_2) \rangle + \frac{1}{m^2} \int_0^{t_1} dt_1' \int_0^{t_2} dt_2' \; \langle f_i(t_1') f_j(t_2') \rangle \, e^{\gamma(t_1' + t_2' - t_1 - t_2)/m}$$

where we made use of the fact that $\langle \vec{f}(t) \rangle = 0$ to drop the terms linear in the noise $\vec{f}$. If we use the white noise correlation function (3.6), and assume $t_2 \geq t_1 > 0$, the integral in the second term becomes,

$$\begin{aligned}\langle \dot{x}_i(t_1)\dot{x}_j(t_2)\rangle &= \langle \dot{x}_i(t_1) \rangle \langle \dot{x}_j(t_2) \rangle + \frac{2D\gamma^2}{m^2} \delta_{ij} \, e^{-\gamma(t_1+t_2)/m} \int_0^{t_1} dt' \; e^{2\gamma t'/m} \\ &= \langle \dot{x}_i(t_1) \rangle \langle \dot{x}_j(t_2) \rangle + \frac{D\gamma}{m} \delta_{ij} \left(e^{-\gamma(t_2 - t_1)/m} - e^{-\gamma(t_1+t_2)/m}\right)\end{aligned}$$

For very large times, $t_1, t_2 \to \infty$, we can drop the last term as well as the average velocities since $\langle \dot{\vec{x}}(t) \rangle \to 0$. We learn that the correlation between velocities decays exponentially as

$$\langle \dot{x}_i(t_1)\dot{x}_j(t_2)\rangle \to \frac{D\gamma}{m} \delta_{ij} \, e^{-\gamma(t_2 - t_1)/m}$$

This means that if you know the velocity of the particle at some time t_1, then you can be fairly confident that it will have a similar velocity at a time $t_2 < t_1 + m/\gamma$ later. But if you wait longer than time m/γ then you would be a fool to make any bets on the velocity based only on your knowledge at time t_1.

Finally, we can also use this result to compute the average velocity-squared (which, of course, is the kinetic energy of the system). At late times, the any initial velocity has died away and the resulting kinetic energy is due entirely to the bombardment by the environment. It is independent of time and given by

$$\langle \dot{\vec{x}}(t) \cdot \dot{\vec{x}}(t) \rangle = \frac{3D\gamma}{m} \tag{3.10}$$

One can compute similar correlation functions for position $\langle\, x_i(t_1) x_j(t_2) \,\rangle$. The expressions are a little more tricky but still quite manageable. (Combining equations (3.9) and (3.8), you can see that you will a quadruple integral to perform and figuring out the limits is a little fiddly). At late times, it turns out that the variance of the position is given by the same expression that we saw for the viscous liquid (3.7),

$$\langle\, (\vec{x}(t) - \vec{x}(0))^2 \,\rangle = 6D\, t \tag{3.11}$$

again exhibiting the now-familiar $\sqrt{t}$ behaviour for the root-mean-square distance.

3.1.3 The Einstein Relation

We brushed over something important and lovely in the previous discussion. We computed the average kinetic energy of a particle in (3.10). It is

$$E = \frac{1}{2} m \langle \dot{\vec{x}} \cdot \dot{\vec{x}} \rangle = \frac{3}{2} D\gamma$$

But we already know what the average energy of a particle is when it's bombarded by its environment: it is given by the equipartition theorem and, crucially, depends only on the temperature of the surroundings

$$E = \frac{3}{2} k_B T$$

It must be therefore that the diffusion constant D is related to the mobility $1/\gamma$ by

$$D = \frac{k_B T}{\gamma} \tag{3.12}$$

That's rather surprising! The diffusion constant captures the amount a particle is kicked around due to the background medium; the mobility expresses the how hard it is for a particle to plough through the background medium. And yet they are related. This equation is telling us that diffusion and viscosity both have their microscopic origin in the random bombardment of molecules. Notice that D is inversely proportional to γ. Yet you might have thought that the amount the particle is kicked increases as the viscosity increases. Indeed, looking back at (3.6), you can see that the amount the particle is kicked is actually proportional to $D\gamma^2 \sim T\gamma$. Which is more in line with our intuition.

Equation (3.12) is known as the Einstein relation. It is an important example of the fluctuation-dissipation theorem. The fluctuations of the particle as it undergoes its random walk are related to the drag force (or dissipation of momentum) that the particle feels as it moves through the fluid.

The Einstein relation gives an excellent way to determine Boltzmann's constant experimentally. Watch a particle perform a Brownian jitter. After time t, the distance travelled by the particle (3.7) should be

$$\langle \vec{x}^{\,2} \rangle = \frac{k_B T}{\pi \eta a}\, t$$

where we have used the Stokes formula $\gamma = 6\pi\eta a$ to relate the mobility to the viscosity μ and radius a of the particle. This experiment was done in 1909 by the French physicist Jean Baptiste Perrin and won him the 1926 Nobel prize.

3.1.4 Noise Probability Distributions

So far, we've only needed to use the two pieces of information about the noise, namely,

$$\langle\, \vec{f}(t)\, \rangle = 0 \tag{3.13}$$

$$\langle\, f_i(t_1) f_j(t_2)\, \rangle = 2D\gamma^2 \delta_{ij} \delta(t_1 - t_2) \tag{3.14}$$

However, if we wanted to compute correlation functions involving more than two velocities or positions, it should be clear from the calculation that we would need to know higher moments of the probability distribution for $\vec{f}(t)$. In fact, the definition of *white noise* is that there are no non-trivial correlations other than $\langle\, f_i(t_1) f_j(t_2)\, \rangle$. This doesn't mean that the higher correlation functions are vanishing, just that they can be reduced to the two-time correlators. This means that for N even,

$$\langle\, f_{i_1}(t_1) \ldots f_{i_N}(t_N)\, \rangle = \langle f_{i_1}(t) f_{i_2}(t_2)\, \rangle \ldots \langle f_{i_{N-1}}(t_{N-1}) f_{i_N}(t_N)\, \rangle + \text{ permutations}$$

while, for N odd

$$\langle\, f_{i_1}(t_1) \ldots f_{i_N}(t_N)\, \rangle = 0$$

Another way of saying this is that all but the second cumulant of the probability distribution vanish.

Instead of specifying all these moments of the distribution, it is often much more useful to specify the probability distribution for $\vec{f}(t)$ directly. However, this is a slightly subtle object because we want to specify the probability for an entire function $\vec{f}(t)$, rather than a single random variable. This means that the probability distribution must be a functional: you give it a function $\vec{f}(t)$ and it spits back a number which, in this case, should be between zero and one.

The good news is that, among the class of probability distributions over functions, the white noise distribution is by far the easiest! If we were dealing with a single random variable, the distribution that has only two-point correlators but no higher is the Gaussian. And, suitably generalised, this also works for our functional probability distribution. The probability distribution that gives white noise is

$$\text{Prob}[f(t)] = \mathcal{N} \exp\left(- \int_{-\infty}^{+\infty} dt \, \frac{\vec{f}(t) \cdot \vec{f}(t)}{4D\gamma^2} \right)$$

where $\mathcal{N}$ is a normalization factor which is needed to ensure that the sum over all probabilities gives unity. This "sum" is really a sum over all functions $\vec{f}(t)$ or, in other words, a functional integral. The normalization condition which fixes $\mathcal{N}$ is then

$$\int Df(t) \; \text{Prob}[f(t)] = 1 \tag{3.15}$$

With this probability distribution, all averaging over the noise can now be computed as a functional integral. If you have any function $g(x)$, then its average is

$$\langle \, g(x) \, \rangle = \mathcal{N} \int Df(t) \; g(x_f) \, e^{-\int dt \; \vec{f}^2/4D\gamma^2}$$

where the notation x_f means the solution to the Langevin equation in the presence of a fixed source f.

Let's now show that the Gaussian probability distribution indeed reproduces the white noise correlations as claimed. To do this, we first introduce an object $Z[\vec{J}(t)]$ known as a *generating function*. We can introduce a generating function for any probability distribution, so let's keep things general for now and later specialise to the Gaussian distribution.

$$Z[\vec{J}(t)] = \int Df(t) \; \text{Prob}[f(t)] \exp\left(\int_{-\infty}^{+\infty} dt \; \vec{J}(t) \cdot \vec{f}(t) \right)$$

This generating function is a functional: it is a function of any function $\vec{J}(t)$ that we care to feed it. By construction, $Z[0] = 1$, courtesy of (3.15).

As the notation Z suggests, the generating function has much in common with the partition function that we work with in a first course of statistical mechanics. This is most apparent in the context of statistical field theories where the generating function is reminiscent of the partition function. Both are functional, or path, integrals. These objects are also important in quantum field theory where the names partition function and generating function are often used synonymously.

The function $\vec{J}$ that we have introduced is, in this context, really little more than a trick that allows us to encode all the correlation functions in $Z[\vec{J}]$. To see how this works. Suppose that we differentiate Z with respect to $\vec{J}$ evaluated at some time $t = t_1$ and then set $\vec{J} = 0$. We have

$$\left.\frac{\delta Z}{\delta J_i(t_1)}\right|_{\vec{J}=0} = \int Df(t)\ f_i(t_1)\,\mathrm{Prob}[f(t)] = \langle\, f_i(t_1)\,\rangle$$

Playing the same game, first taking n derivatives, gives

$$\begin{aligned}\left.\frac{\delta^n Z}{\delta J_{i_1}(t_1)\delta J_{i_2}(t_2)\ldots J_{i_n}(t_n)}\right|_{\vec{J}=0} &= \int Df(t)\ f_{i_1}(t_1)f_{i_2}(t_2)\ldots f_{i_n}(t_n)\,\mathrm{prob}[f(t)] \\ &= \langle\, f_{i_1}(t_1)f_{i_2}(_2)\ldots f_{i_n}(t_n)\,\rangle\end{aligned}$$

So we see that if we can compute $Z[\vec{J}]$, then successive correlation functions are simply the coefficients of a Taylor expansion in $\vec{J}$. This is particularly useful for the Gaussian distribution where the generating function is,

$$Z[\vec{J}(t)] = \mathcal{N}\int Df(t)\ \exp\left(-\int_{-\infty}^{+\infty} dt\ \frac{\vec{f}(t)\cdot\vec{f}(t)}{4D\gamma^2} - \vec{J}(t)\cdot\vec{f}(t)\right)$$

But this is nothing more than a Gaussian integral. (Ok, it's an infinite number of Gaussian integrals because it's a functional integral. But we shouldn't let that phase us). We can easily compute it by completing the square

$$Z[\vec{J}(t)] = \mathcal{N}\int Df(t)\ \exp\left(-\frac{1}{4D\gamma^2}\int_{-\infty}^{+\infty} dt\ \left[\vec{f}(t) - 2D\gamma^2\vec{J}(t)\right]^2 - 4D^2\gamma^4\vec{J}(t)\cdot\vec{J}(t)\right)$$

After the shift of variable, $\vec{f} \to \vec{f} - 2D\gamma^2\vec{J}$, the integral reduces to (3.15), leaving behind

$$Z[\vec{J}(t)] = \exp\left(D\gamma^2\int_{-\infty}^{+\infty} dt\ \vec{J}(t)\cdot\vec{J}(t)\right)$$

Now it is an easy matter to compute correlation functions. Taking one derivative, we have

$$\frac{\delta Z}{\delta J_i(t_1)} = 2D\gamma^2\, J_i(t_1)\, Z[\vec{J}]$$

But this vanishes when we set $J = 0$, in agreement with our requirement (3.13) that the average noise vanishes. Taking a second derivative gives,

$$\frac{\delta^2 Z}{\delta J_i(t_1)\delta J_j(t_2)} = 2D\gamma^2\delta_{ij}\delta(t_1 - t_2)Z[\vec{J}] + 4D^2\gamma^4 J_i(t_1)J_j(t_2)Z[\vec{J}]$$

Now setting $\vec{J} = 0$, only the first term survives and reproduces the white noise correlation (3.14). One can continue the process to see that all higher correlation functions are entirely determined by $\langle f_i\, f_j\,\rangle$.

3.1.5 Stochastic Processes for Fields

Finally, it's worth mentioning that Langevin-type equations are not restricted to particle positions. It is also of interest to write down stochastic processes for fields. For example, we may want to consider a time dependent process for some order parameter $m(\vec{r}, t)$, influenced by noise

$$\frac{\partial m}{\partial t} = c\nabla^2 m - am - 2bm^2 + f$$

where $f(\vec{r}, t)$ is a random field with correlations $\langle f \rangle = 0$ and

$$\langle\, f(\vec{r}_1, t_1) f(\vec{r}_2, t_2)\,\rangle \sim \delta^d(\vec{r}_1 - \vec{r}_2)\delta(t_1 - t_2)$$

A famous example of a stochastic process is provided by the fluctuating boundary between, say, a gas and a liquid. Denoting the height of the boundary as $h(\vec{r}, t)$, the simplest description of the boundary fluctuations is given by the Edwards-Wilkinson equation,

$$\frac{\partial h}{\partial t} = \nabla^2 h + f$$

A somewhat more accurate model is given by the Kardar-Parisi-Zhang equation,

$$\frac{\partial h}{\partial t} = \nabla^2 h + \lambda(\nabla h)^2 + f$$

We won't have anything to say about the properties of these equations in this course. An introduction can be found in the second book by Kardar.

3.2 The Fokker-Planck Equation

Drop a particle at some position, say $\vec{x}_0$ at time t_0. If the subsequent evolution is noisy, so that it is governed by a stochastic Langevin equation, then we've got no way to know for sure where the particle will be. The best that we can do is talk about probabilities. We will denote the probability that the particle sits at $\vec{x}$ at time t as $P(\vec{x}, t; \vec{x}_0, t_0)$.

In the previous section we expressed our uncertainty in the position of the particle in terms of correlation functions. Here we shift perspective a little. We would like to ask: what probability distribution $P(\vec{x}, t; \vec{x}_0, t_0)$ would give rise to the same correlation functions that arose from the Langevin equation?

We should stress that we care nothing about the particular path $\vec{x}(t)$ that the particle took. The probability distribution over paths would be a rather complicated functional (rather like those we saw in Section 3.1.4). Instead we will ask the much simpler question of the probability that the particle sits at $\vec{x}$ at time t, regardless of how it got there.

It is simple to write down a formal expression for the probability distribution. Let's denote the solution to the Langevin equation for a given noise function $\vec{f}$ as $\vec{x}_f$. Of course, if we know the noise, then there is no uncertainty in the probability distribution for $\vec{x}$. It is simply $P(\vec{x},t) = \delta(\vec{x}-\vec{x}_f)$. Now averaging over all possible noise, we can write the probability distribution as

$$P(\vec{x},t) = \langle\, \delta(\vec{x}-\vec{x}_f)\,\rangle \tag{3.16}$$

In this section, we shall show that the $P(\vec{x},t)$ obeys a simple partial differential equation known as the Fokker-Planck equation.

3.2.1 The Diffusion Equation

The simplest stochastic process we studied was a particle subject to random forces in a very viscous fluid. The Langevin equation is

$$\dot{\vec{x}}(t) = \frac{1}{\gamma}\,\vec{f}(t)$$

In Section 3.1.1 we showed that the average position of the particle remains unchanged: if $\vec{x}(t=0)=\vec{0}$ then $\langle \vec{x}(t)\rangle = \vec{0}$ for all t. But the variance of the particle undergoes a random walk (3.7),

$$\langle\, \vec{x}(t)^2\,\rangle = 6Dt \tag{3.17}$$

For this simple case, we won't derive the probability distribution: we'll just write it down. The probability distribution that reproduces this variance: it is just a Gaussian

$$P(\vec{x},t) = \left(\frac{1}{4\pi Dt}\right)^{3/2} e^{-\vec{x}^{\,2}/4Dt} \tag{3.18}$$

where the factor out front is determined by the normalization requirement that

$$\int d^3x\, P(x,t) = 1$$

for all time t. Note that there is more information contained in this probability distribution that the just the variance (3.17). Specifically, all higher cumulants vanish. (This means, for example, that $\langle \vec{x}^{\,3}\,\rangle = 0$ and $\langle\, \vec{x}^{\,4}\,\rangle = 3\langle\, \vec{x}^{\,2}\,\rangle$ and so on). But it simple to check that this is indeed what arises from the Langevin equation with white noise described in Section 3.1.4.

The probability distribution (3.18) obeys the *diffusion equation*,

$$\frac{\partial P}{\partial t} = D\nabla^2 P$$

This is the simplest example of a Fokker-Planck equation. However, for more complicated versions of the Langevin equation, we will have to work harder to derive the analogous equation governing the probability distribution P.

3.2.2 Meet the Fokker-Planck Equation

Let's now consider the a more general stochastic process. We'll still work in the viscous limit for now, setting $m = 0$ so that we have a first order Langevin equation,

$$\gamma\dot{\vec{x}} = -\nabla V + \vec{f} \tag{3.19}$$

A quadratic V corresponds to a harmonic oscillator potential and the Langevin equation is not difficult to solve. (In fact, mathematically it is the same problem that we solved in Section 3.1.2. You just have to replace $\dot{\vec{x}} = \vec{v} \to \vec{x}$). Any other V gives rise to a non-linear stochastic equation (confusingly sometimes called "quasi-linear" in this context) and no general solution is available. Nonetheless, we will still be able to massage this into the form of a Fokker-Planck equation.

We begin by extracting some information from the Langevin equation. Consider a particle sitting at some point x at time t. If we look again a short time δt later, the particle will have moved a small amount

$$\delta\vec{x} = \dot{\vec{x}}\,\delta t = -\frac{1}{\gamma}\,\nabla V\,\delta t + \frac{1}{\gamma}\int_t^{t+\delta t} dt'\,\vec{f}(t') \tag{3.20}$$

Here we've taken the average value of the noise function, f, over the small time interval. However, we've assumed that the displacement of the particle $\delta\vec{x}$ is small enough so that we can evaluate the force ∇V at the original position $\vec{x}$. (It turns out that this is ok in the present context but there are often pitfalls in making such assumptions in the theory of stochastic processes. We'll comment on one such pitfall at the end of this Section). We can now compute the average. Because $\langle\vec{f}(t)\rangle = 0$, we have

$$\langle\,\delta\vec{x}\,\rangle = -\frac{1}{\gamma}\,\nabla V\,\delta t \tag{3.21}$$

The computation $\langle\,\delta x_i\,\delta x_j\rangle$ is also straightforward,

$$\begin{aligned}\gamma^2\langle\,\delta x_i\delta x_j\rangle = \langle\partial_i V\partial_j V\rangle\delta t^2 - \delta t\int_t^{t+\delta t} dt'\,\langle\partial_i V\,f_j(t') + \partial_j\,V f_i(t')\rangle \\ + \int_t^{t+\delta t} dt'\int_t^{t+\delta t} dt''\langle\,f_i(t')\,f_j(t'')\,\rangle\end{aligned}$$

Both the first two terms are order δt^2. However, using (3.6), one of the integrals in the third term is killed by the delta function, leaving just one integral standing. This ensures that the third term is actually proportional to δt,

$$\langle\, \delta x_i \delta x_j \,\rangle = 2\delta_{ij} D\, \delta t + \mathcal{O}(\delta t^2) \tag{3.22}$$

We will ignore the terms of order δt^2. Moreover, It is simple to see that all higher correlation functions are higher order in δt. For example, $\langle \vec{x}^{\,4} \rangle \sim \delta t^2$. These too will be ignored.

Our strategy now is to construct a probability distribution that reproduces (3.21) and (3.22). We start by considering the conditional probability $P(\vec{x}, t+\delta t; \vec{x}^{\,\prime}, t)$ that the particle sits at $\vec{x}$ at time $t+\delta t$ given that, a moment earlier, it was sitting at $\vec{x}^{\,\prime}$. From the definition (3.16) we can write this as

$$P(\vec{x}, t+\delta t; \vec{x}^{\,\prime}, t) = \langle\, \delta(\vec{x} - \vec{x}^{\,\prime} - \delta\vec{x})\,\rangle$$

where δx is the random variable here; it is the distance moved in time δt. Next, we do something that may look fishy: we Taylor expand the delta-function. If you're nervous about expanding a distribution in this way, you could always regulate the delta function in your favourite manner to turn it into a well behaved function. However, more pertinently, we will see that the resulting expression sits inside an integral where any offending terms make perfect sense. For now, we just proceed naively

$$P(\vec{x}, t+\delta t; \vec{x}^{\,\prime}, t) = \left(1 + \langle\, \delta x_i \,\rangle \frac{\partial}{\partial x_i'} + \frac{1}{2} \langle\, \delta x_i\, \delta x_j \,\rangle \frac{\partial^2}{\partial x_i' \partial x_j'} + \ldots \right) \delta(\vec{x} - \vec{x}^{\,\prime}) \tag{3.23}$$

We have truncated at second order because we want to compare this to (3.27) and, as we saw above, $\langle\, \delta\vec{x} \rangle$ and $\langle\, \delta\vec{x}^{\,2} \,\rangle$ are the only terms that are order δt.

We now have all the information that we need. We just have to compare (3.27) and (3.23) and figure out how to deal with those delta functions. To do this, we need one more trick. Firstly, recall that our real interest is in the evolution of the probability $P(\vec{x}, t; \vec{x}_0, t_0)$, given some initial, arbitrary starting position $\vec{x}(t = t_0) = \vec{x}_0$. There is an obvious property that this probability must satisfy: if you look at some intermediate time $t_0 < t' < t$, then the particle has to be somewhere. Written as an equation, this "has to be somewhere" property is called the Chapman-Kolmogorov equation

$$P(\vec{x}, t; \vec{x}_0, t_0) = \int_{-\infty}^{+\infty} d^3\vec{x}^{\,\prime}\; P(\vec{x}, t; \vec{x}^{\,\prime}, t')\, P(\vec{x}^{\,\prime}, t'; \vec{x}_0, t_0) \tag{3.24}$$

Replacing t by $t + \delta t$, we can substitute our expression (3.23) into the Chapman-Kolmogorov equation, and then integrate by parts so that the derivatives on the delta function turn and hit $P(\vec{x}', t'; \vec{x}_0, t_0)$. The delta-function, now unattended by derivatives, kills the integral, leaving

$$P(\vec{x}, t+\delta t; \vec{x}_0, t_0) = P(\vec{x}, t; \vec{x}_0, t_0) - \frac{\partial}{\partial x_i}\Big(\langle\, \delta x_i \,\rangle P(\vec{x}, t; \vec{x}_0, t_0)\Big)$$
$$+\frac{1}{2}\langle\, \partial x_i \, \partial x_j \,\rangle \frac{\partial^2}{\partial x_i \partial x_j} P(\vec{x}, t; \vec{x}_0, t_0) + \ldots \tag{3.25}$$

Using our expressions for $\langle \delta x \rangle$ and $\langle \delta x \delta x \rangle$ given in (3.21) and (3.22), this becomes

$$P(\vec{x}, t+\delta t; \vec{x}_0, t_0) = P(\vec{x}, t; \vec{x}_0, t_0) + \frac{1}{\gamma}\frac{\partial}{\partial x_i}\left(\frac{\partial V}{\partial x_i} P(\vec{x}, t; \vec{x}_0, t_0)\right) \delta t$$
$$+D\frac{\partial^2}{\partial x^2} P(\vec{x}, t; \vec{x}_0, t_0)\, \delta t + \ldots \tag{3.26}$$

But we can also get a much simpler expression for the left-hand side simply by Taylor expanding with respect to time,

$$P(\vec{x}, t+\delta t; \vec{x}_0, t_0) = P(\vec{x}, t; \vec{x}_0, t_0) + \frac{\partial}{\partial t} P(\vec{x}, t; \vec{x}_0, t_0)\, \delta t + \ldots \tag{3.27}$$

Equating (3.27) with (3.26) gives us our final result,

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma}\nabla \cdot (P \nabla V) + D\nabla^2 P \tag{3.28}$$

This is the *Fokker-Planck* equation. This form also goes by the name of the Smoluchowski equation or, for probabilists, Kolomogorov's forward equation.

Properties of the Fokker-Planck Equation

It is useful to write the Fokker-Planck equation as a continuity equation

$$\frac{\partial P}{\partial t} = \nabla \cdot \vec{J} \tag{3.29}$$

where the *probability current* is

$$\vec{J} = \frac{1}{\gamma} P \nabla V + D \nabla P \tag{3.30}$$

The second term is clearly due to diffusion (because there's a big capital D in front of it); the first term is due to the potential and is often referred to as the *drift*, meaning the overall motion of the particle due to background forces that we understand.

One advantage of writing the Fokker-Planck equation in terms of a current is that we see immediately that probability is conserved, meaning that if $\int d^3x\, P = 1$ at some point in time then it will remain so for all later times. This follows by a standard argument,

$$\frac{\partial}{\partial t}\int d^3x\, P = \int d^3x\, \frac{\partial P}{\partial t} = \int d^3x\ \nabla\cdot\vec{J} = 0$$

where the last equality follows because we have a total derivative (and we are implicitly assuming that there's no chance that the particle escapes to infinity so we can drop the boundary term).

The Fokker-Planck equation tells us how systems evolve. For some systems, such as those described by the diffusion equation, there is no end point to this evolution: the system just spreads out more and more. However, for generic potentials V there are time-independent solutions to the Fokker-Planck equation obeying $\nabla\cdot\vec{J} = 0$. These are the equilibrium configurations. The solution is given by

$$P(\vec{x}) \sim e^{-V(\vec{x})/\gamma D}$$

Using the Einstein relation (3.12), this becomes something very familiar. It is simply the Boltzmann distribution for a particle with energy $V(\vec{x})$ in thermal equilibrium

$$P(\vec{x}) \sim e^{-V(\vec{x})/k_BT} \tag{3.31}$$

Isn't that nice! (Note that there's no kinetic energy in the exponent as we set $m = 0$ as our starting point).

An Application: Escape over a Barrier

As an application of the Fokker-Planck equation, consider thermal escape from the one-dimensional potential shown in Figure 7. We'll assume that all the particles start off sitting close to the local minimum at $x_{\rm min}$. We model the potential close to this point as

$$V(x) \approx \frac{1}{2}\omega_{\rm min}^2(x - x_{\rm min})^2$$

and we start our particles in a distribution that is effectively in local equilbrium (3.31), with

$$P(x, t=0) = \sqrt{\frac{\omega_{\rm min}^2}{2\pi k_BT}}\, e^{-\omega_{\rm min}^2(x-x_{\rm min})^2/2k_BT} \tag{3.32}$$

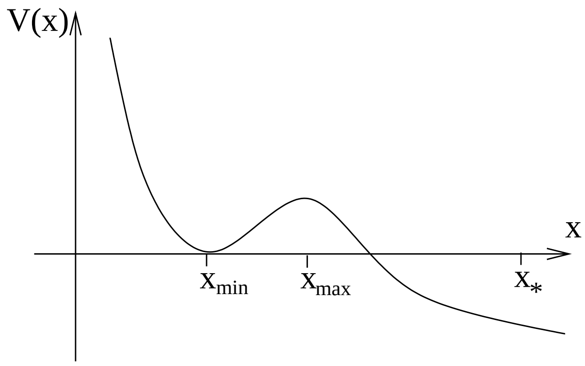

Figure 7: Escape over a Barrier.

But, globally, $x_{\rm min}$ is not the lowest energy configuration and this probability distribution is not the equilibrium configuration. In fact, as drawn, the potential has no global minimum and there is no equilibrium distribution. So this isn't what we'll set out to find. Instead, we would like to calculate the rate at which particles leak out of the trap and over the barrier.

Although we're clearly interested in a time dependent process, the way we proceed is to assume that the leakage is small and so can be effectively treated as a steady state process. This means that we think of the original probability distribution of particles (3.32) as a bath which, at least on the time scales of interest, is unchanging. The steady state leakage is modelled by a constant probability current $J = J_0$, with J given by (3.30). Using the Einstein relation $D = k_BT/\gamma$, we can rewrite this as

$$J = \frac{k_BT}{\gamma} e^{-V(x)/k_BT} \frac{\partial}{\partial x}\left(e^{+V(x)/k_BT} P\right)$$

The first step is to integrate $J_0\, e^{+V(x)/k_BT}$ between the minimum $x_{\rm min}$ and some distance far from all the action, $x \gg x_{\rm max}$, which we may as we call $x = x_\star$,

$$\int_{x_{\rm min}}^{x_\star} J_0\, e^{V(x)/k_BT} = \frac{k_BT}{\gamma}\left[e^{V(x_{\rm min})/k_BT} P(x_{\rm min}) - e^{V(x_\star)/k_BT} P(x_\star)\right]$$

But we can take the probability $P(x_\star)$ to be vanishingly small compared to $P(x_{\rm min})$ given in (3.32), leaving us with

$$\int_{x_{\rm min}}^{x_\star} J_0\, e^{V(x)/k_BT} \approx \frac{k_BT}{\gamma}\sqrt{\frac{\omega_{\rm min}^2}{2\pi k_BT}} \tag{3.33}$$

Meanwhile, the integral on the left-hand-side is dominated by the maximum of the potential. Let's suppose that close to the maximum, the potential looks like

$$V(x) \approx V_{\rm max} - \frac{1}{2}\omega_{\rm max}^2 (x - x_{\rm max})^2$$

Then we'll write the integral as

$$J_0 \int_{x_{\rm min}}^{x_\star} e^{V(x)/k_BT} \approx J_0\, e^{V_{\rm max}/k_BT} \sqrt{\frac{2\pi k_BT}{\omega_{\rm max}^2}} \tag{3.34}$$

Combining the two expressions (3.33) and (3.34), we get our final result for the rate of escape over the barrier

$$J_0 \approx \frac{\omega_{\rm min}\omega_{\rm max}}{2\pi\gamma}\, e^{-V_{\rm max}/k_BT}$$

3.2.3 Velocity Diffusion

So far we've ignored the inertia term, setting $m = 0$. Let's now put it back in. We can start by setting the potential to zero, so that the Langevin equation is

$$m\ddot{\vec{x}} = -\gamma\dot{\vec{x}} + \vec{f}(t)$$

But, we can trivially rewrite this as a first order equation involving $\vec{v} = \dot{\vec{x}}$,

$$m\dot{\vec{v}} = -\gamma\vec{v} + \vec{f}(t)$$

This means that if we're only interested in the distribution over velocities, $P(\vec{v}, t)$, then we have exactly the same problem that we've just looked at, simply replacing $\vec{x} \to \vec{v}$ and $\gamma \to m$. (Actually, you need to be a little more careful. The diffusion constant D that appears in (3.28) was really $D\gamma^2/\gamma^2$ where the numerator arose from the noise correlator and the denominator from the $\gamma\dot{\vec{x}}$ term in the Langevin equation. Only the latter changes, meaning that this combination gets replaced by $D\gamma^2/m^2$). The resulting Fokker-Planck equation is

$$\frac{\partial P}{\partial t} = \frac{1}{m}\frac{\partial}{\partial \vec{v}} \cdot \left(\gamma P \vec{v} + \frac{D\gamma^2}{m}\frac{\partial P}{\partial \vec{v}} \right) \tag{3.35}$$

The equilibrium distribution that follows from this obeys $\partial P/\partial t = 0$, meaning

$$\frac{\partial P}{\partial \vec{v}} = -\frac{m}{D\gamma} P\vec{v} \quad\quad \Rightarrow \quad\quad P = \left(\frac{m}{2\pi k_BT}\right)^{3/2} e^{-m\vec{v}^2/2k_BT}$$

where we've again used the Einstein relation $D\gamma = k_BT$. This, of course, is the Maxwell-Boltzmann distribution.

In fact, we can do better than this. Suppose that we start all the particles off at $t = 0$ with some fixed velocity, $\vec{v} = \vec{v}_0$. This mean that the probability distribution is a delta-function, $P(\vec{v}, t = 0) = \delta^3(\vec{v} - \vec{v}_0)$. We can write down a full time-dependent solution to the Fokker-Planck equation (3.35) with this initial condition.

$$P(\vec{v}, t) = \left(\frac{m}{2\pi k_B T(1 - e^{-2\gamma t/m})}\right)^{3/2} \exp\left(-\frac{m(\vec{v} - \vec{v}_0 e^{-\gamma t/m})^2}{2k_B T(1 - e^{-2\gamma t/m})}\right)$$

As $t \to \infty$, we return to the Maxwell-Boltzmann distribution. But now this tells us how we approach equilibrium.

The Kramers-Chandrasekhar Fokker-Planck Equation

As our final example of a Fokker-Planck equation, we can consider the Langevin equation with both acceleration term and potential term,

$$m\ddot{\vec{x}} = -\gamma\dot{\vec{x}} - \nabla V + \vec{f}(t)$$

Now we are looking for a probability distribution over phase space, $P(\vec{x}, \dot{\vec{x}}, t)$. The right way to proceed is to write this as two first-order equations. The first of these is simply the definition of velocity $\vec{v} = \dot{x}$. The second is the Langevin equation

$$m\dot{\vec{v}} = -\gamma\vec{v} - \nabla V + \vec{f}(t)$$

These can now be combined into a single Langevin equation for six variables. Once armed with this, we need only follow the method that we saw above to arrive at a Fokker-Planck equation for $P(\vec{x}, \vec{v}, t)$,

$$\left(\frac{\partial}{\partial t} + v^i \frac{\partial}{\partial x^i}\right) P = \frac{1}{m}\frac{\partial}{\partial v^i}\left(\gamma v^i P + P\frac{\partial V}{\partial x^i}\right) + \frac{D\gamma^2}{m^2}\frac{\partial^2 P}{\partial v^i \partial v^i} \tag{3.36}$$

This form of the Fokker-Planck equations is sometimes called the Kramers equation and sometimes called the Chandrasekhar equation.

Note that this equation is now capturing the same physics that we saw in the Boltzmann equation: the probability distribution $P(\vec{x}, \vec{v}, t)$ is the same object that we called $f_1(\vec{r}, \vec{p}; t)$ in Section 2. Moreover, it is possible to derive this form of the Fokker-Planck equation, starting from the Boltzmann equation describing a heavy particle in a surrounding bath of light particles. The key approximation is that in small time intervals δt, the momentum of the heavy particle only changes by a small amount. Looking back, you can see that this was indeed an assumption in the derivation of the Fokker-Planck equation in Section 3.2.2, but not in the derivation of the Boltzmann equation.

Integrating over Velocity

The equation (3.36) governing the probability distribution over phase space $P(\vec{x}, \vec{v}, t)$ looks very different from the Fokker-Planck equation governing the probability distribution over configuration space (3.28). Yet the related Langevin equations are simply related by setting $m = 0$ or, equivalently, looking at systems with large γ. How can we derive (3.28) from (3.36)?

The computation involves a careful expansion of (3.36) in powers of $1/\gamma$. Let's see how this works. Firstly, we use the Einstein relation to write $D\gamma = k_B T$, and the rearrange the terms in (3.36) to become

$$\frac{\partial}{\partial v^i}\left(\frac{k_B T}{m^2}\frac{\partial}{\partial v^i} + \frac{v^i}{m}\right) P = \frac{1}{\gamma}\left(\frac{\partial}{\partial t} + v^i \frac{\partial}{\partial x^i} - \frac{1}{m}\frac{\partial V}{\partial x^i}\frac{\partial}{\partial v^i}\right) P \tag{3.37}$$

We're going to solve this perturbatively in $1/\gamma$, writing

$$P = P^{(0)} + \frac{1}{\gamma}P^{(1)} + \frac{1}{\gamma^2}P^{(2)} + \ldots$$

As our first pass at this, we drop anything that has a $1/\gamma$, which mean that $P^{(0)}$ must be annihilated by the left-hand-side of (3.37). and This is a simple differential equation, with solution

$$P^{(0)}(v, x, t) = e^{-mv^2/2k_B T}\,\phi^{(0)}(x, t)$$

for any function $\phi^{(0)}(x, t)$. Of course, the velocity dependence here is simply the Maxwell-Boltzmann distribution. To figure out what restrictions we have on $\phi^{(0)}$, we need to go to the next order in perturbation theory. Keeping terms of $\mathcal{O}(1/\gamma)$, the differential equation (3.37) becomes

$$\frac{\partial}{\partial v^i}\left(\frac{k_B T}{m^2}\frac{\partial}{\partial v^i} + \frac{v^i}{m}\right) P^{(1)} = \left(\frac{\partial}{\partial t} + v^i \frac{\partial}{\partial x^i} + \frac{v^i}{k_B T}\frac{\partial V}{\partial x^i}\right) \phi^{(0)} e^{-mv^2/2k_B T} \tag{3.38}$$

But this equation cannot be solved for arbitrary $\phi^{(0)}$. This is simplest to see if we just integrate both sides over $\int d^3v$: the left-hand-side is a total derivative and so vanishes. On the right-hand-side, only one term remains standing and this must vanish. It is

$$\frac{\partial \phi^{(0)}}{\partial t} = 0$$

So $\phi^{(0)} = \phi^{(0)}(x)$. With this constraint, the solution to (3.38) is, again, straightforward to write down. It is

$$P^{(1)}(x, v, t) = \left(-mv^i \frac{\partial \phi^{(0)}}{\partial x^i} - \frac{m}{k_B T} v^i \frac{\partial V}{\partial x^i}\,\phi^{(0)} + \phi^{(1)}(x, t)\right) e^{-mv^2/2k_B T}$$

At this point, it doesn't look like we're making much progress. We still don't know what $\phi^{(0)}(x)$ is and we've now had to introduce yet another arbitrary function, $\phi^{(1)}(x,t)$ which carries all the time dependence. Let's plug this once more into (3.37), now working to order $\mathcal{O}(1/\gamma^2)$. After a little bit of algebra, the equation becomes

$$\frac{\partial}{\partial v^i}\left(\frac{k_BT}{m^2}\frac{\partial}{\partial v^i}+\frac{v^i}{m}\right)P^{(2)} = \left[-mv^iv^j\frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial x^j}+\frac{1}{k_BT}\frac{\partial V}{\partial x^j}\right)\phi^{(0)}\right.$$
$$+\frac{\partial V}{\partial x^i}\left(\delta_{ij}-\frac{m}{k_BT}v^iv^j\right)\left(\frac{\partial}{\partial x^j}+\frac{1}{k_BT}\frac{\partial V}{\partial x^j}\right)\phi^{(0)}$$
$$\left.+\left(\frac{\partial}{\partial t}+v^i\frac{\partial}{\partial x^i}+\frac{v^i}{k_BT}\frac{\partial V}{\partial x^i}\right)\phi^{(1)}\right]e^{-mv^2/2k_BT}$$

Once again, there's a consistency condition that must be realised if this equation is to have a solution. Integrating over $\int d^3v$, the left-hand-side is a total derivative and therefore vanishes. Any term linear in v on the right-hand-side also vanishes. But so too do the terms on the second line: you can check that the Gaussian integral over the δ_{ij} term exactly cancels the v^iv^j term. The resulting consistency condition is

$$\frac{\partial\phi^{(1)}}{\partial t} = k_BT\frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial x^i}-\frac{1}{k_BT}\frac{\partial V}{\partial x^i}\right)\phi^{(0)} \tag{3.39}$$

where the overall factor of k_BT on the right-hand-side comes only arises when you do the Gaussian integral over $\int d^3v$.

Now we're almost there. (Although it probably doesn't feel like it!). Collecting what we've learned, to order $\mathcal{O}(1/\gamma)$, the probability distribution over phase space takes the form

$$P(x,v,t) = \left(\phi^{(0)}-\frac{mv^i}{\gamma}\frac{\partial\phi^{(0)}}{\partial x^i}-\frac{mv^i}{\gamma k_BT}\frac{\partial V}{\partial x^i}\phi^{(0)}+\frac{\phi^{(1)}}{\gamma}\right)e^{-mv^2/2k_BT}$$

But to make contact with the earlier form of the Fokker-Planck equation (3.28), we want a distribution over configuration space. We get this by simply integrating over velocities. We'll also denote the resulting probability distribution as $P(x,t)$, with only the arguments to tell us that it's a different object:

$$P(x,t) = \int d^3v\,P(x,v,t) = \sqrt{\frac{2\pi k_BT}{m}}\left(\phi^{(0)}(x)+\frac{1}{\gamma}\phi^{(1)}(x,t)\right)$$

But now we can use the consistency condition (3.39) to compute $\partial P/\partial t$. Working only to order $\mathcal{O}(1/\gamma)$, this reads

$$\frac{\partial P}{\partial t} = \frac{k_BT}{\gamma}\frac{\partial}{\partial x^i}\left(\frac{\partial}{\partial x^i}+\frac{1}{k_BT}\frac{\partial V}{\partial x^i}\right)P$$

Which is precisely the Fokker-Planck equation (3.28) that we saw previously.

3.2.4 Path Integrals: Schrödinger, Feynman, Fokker and Planck

There is a close similarity between the Fokker-Planck equation and the Schrödinger equation in quantum mechanics. To see this, let's return to the first order Langevin equation

$$\dot{\vec{x}} = \frac{1}{\gamma}\left(-\nabla V + \vec{f}\right) \tag{3.40}$$

and the corresponding Fokker-Planck equation (3.28). We can change variables to

$$P(x,t) = e^{-V(x)/2\gamma D}\tilde{P}(x,t) \tag{3.41}$$

Substituting into the Fokker-Planck equation, we see that the rescaled probability $\tilde{P}$ obeys

$$\frac{\partial \tilde{P}}{\partial t} = D\nabla^2\tilde{P} + \left(\frac{1}{2\gamma}\nabla^2 V - \frac{1}{4\gamma^2 D}(\nabla V)^2\right)\tilde{P} \tag{3.42}$$

There are no first order gradients $\nabla\tilde{P}$; only $\nabla^2\tilde{P}$. This form of the Fokker-Planck equation looks very similar to the Schrödinger equation.

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + U(\vec{x})\psi$$

All that's missing is a factor of i on the left-hand-side. Otherwise, with a few trivial substitutions, the two equations look more or less the same. Note, however, that the relationship between the potentials is not obvious: if we want to relate the two equations, we should identify

$$U = -\frac{1}{2\gamma}\nabla^2 V + \frac{1}{4D\gamma^2}(\nabla V)^2 \tag{3.43}$$

The relationship between the evolution of quantum and classical probabilities is also highlighted in the path integral formulation. Recall that the Schrödinger equation can be reformulated in terms of function integrals, with the quantum amplitude for a particle to travel from $\vec{x} = \vec{x}_i$ at time $t = t_i$ to $\vec{x}_f$ at time t_f given by[8].

$$\langle \vec{x}_f, t_f | \vec{x}_i, t_i\rangle = \mathcal{N}\int \mathcal{D}x(t)\ \exp\left(\frac{i}{\hbar}\int dt\ \frac{\dot{\vec{x}}^2}{2m} - U(\vec{x})\right)$$

[8] A derivation of the path integral from the Schrödinger equation can be found in many places, including Section 4 of the notes, http://www.damtp.cam.ac.uk/user/tong/dynamics.html

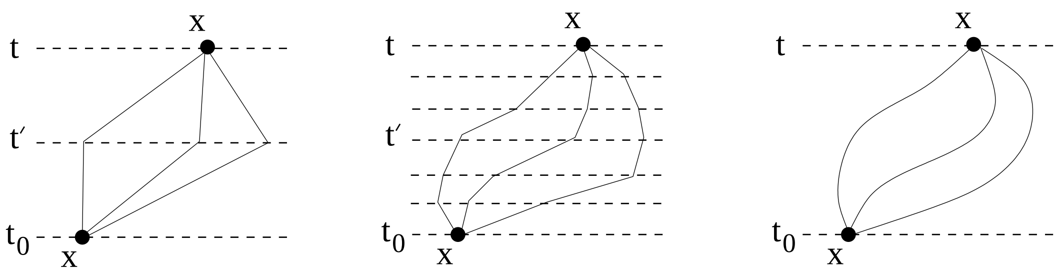

Figure 8: From Chapman-Kolmogorov to Feynman.

where $\mathcal{N}$ is a normalization factor. Here the integral is over all paths which start at $(\vec{x}_i, t_i)$ and end at $\vec{x}_f, t_f)$. By analogy, we expect there to be a similar path integral formulation of the classical probability for a particle in the Langevin environment (3.40) to travel from $\vec{x}_i$ to $\vec{x}_f$. Indeed, the existence of a path integral formulation for this problem is very natural. The essence of this can already be seen in the Chapman-Kolmogorov equation (3.24)

$$P(\vec{x}, t; \vec{x}_0, t_0) = \int_{-\infty}^{+\infty} d^3\vec{x}' \; P(\vec{x}, t; \vec{x}', t')\, P(\vec{x}', t'; \vec{x}_0, t_0)$$

This simply says that to get from point A to point B, a particle has to pass through some position in between. And we sum up the probabilities for each position. Adding many more intervening time steps, as shown in Figure 8, naturally suggests that we should be summing over all possible paths.

Deriving the Path Integral

Here we will sketch the derivation of the path integral formula for the Fokker-Planck equation. We've already met function integrals in Section 3.1.4 where we introduced the probability distribution for a given noise function $\vec{f}(t)$

$$\text{Prob}[f(t)] = \mathcal{N} \exp\left(-\int dt\; \frac{\vec{f}(t)\cdot\vec{f}(t)}{4D\gamma^2}\right) \tag{3.44}$$

subject to the normalization condition

$$\int \mathcal{D}f(t)\, \text{Prob}[f(t)] = 1 \tag{3.45}$$

But given a fixed noise profile $\vec{f}(t)$ and an initial condition, the path of the particle is fully determined by the Langevin equation (3.40). Let's call this solution $\vec{x}_f$. Then the probability that the particle takes the path $\vec{x}_f$ is the same as the probability that the force is $\vec{f}$,

$$\begin{aligned}\text{Prob}[\vec{x}_f(t)] = \text{Prob}[\vec{f}(t)] &= \mathcal{N} \exp\left(-\int dt\ \frac{\vec{f}(t)\cdot\vec{f}(t)}{4D\gamma^2}\right)\\ &= \mathcal{N}\exp\left(-\frac{1}{4D\gamma^2}\int dt\ (\gamma\dot{\vec{x}}_f + \nabla V(\vec{x}_f))^2\right)\end{aligned}$$

where, in the last line, we've used the Langevin equation (3.40) to relate the force to the path taken. But since this equation holds for any path $\vec{x}_f$, we can simply drop the f label. We have the probability that the particle takes a specific path $\vec{x}(t)$ given by

$$\text{Prob}[\vec{x}(t)] = \mathcal{N}\exp\left(-\frac{1}{4D\gamma^2}\int dt\ (\gamma\dot{\vec{x}} + \nabla V)^2\right)$$

The total probability to go from $\vec{x}_i$ to $\vec{x}_f$ should therefore just be the sum over all these paths. With one, slightly fiddly, subtlety: the probability is normalized in (3.45) with respect to the integration measure over noise variable $\vec{f}$. And we want to integrate over paths. This means that we have to change integration variables and pick up a Jacobian factor for our troubles. We have

$$\begin{aligned}\text{Prob}[\vec{x}_f, t_f; \vec{x}_i, t_i] &= \mathcal{N}\int \mathcal{D}f(t)\ \exp\left(-\frac{1}{4D\gamma^2}\int dt\ (\gamma\dot{\vec{x}}_f + \nabla V(\vec{x}_f)^2\right)\\ &= \mathcal{N}\int \mathcal{D}x(t)\ \det\mathcal{M}\ \exp\left(-\frac{1}{4D\gamma^2}\int dt\ (\gamma\dot{\vec{x}} + \nabla V)^2\right) \quad (3.46)\end{aligned}$$

Here the operator $\mathcal{M}(t,t')$ that appears in the Jacobian be thought of as $\delta f(t)/\delta x(t')$. It can be written down by returning to the Langevin equation (3.40) which relates $f$and x,

$$\mathcal{M}(t,t') = \gamma\frac{\partial}{\partial t}\delta(t-t') + \nabla^2 V\delta(t-t')$$

If we want to think in a simple minded way, we can consider this as a (very large) matrix $\mathcal{M}_{tt'}$, with columns labelled by the index t and rows labelled by t'. We'll write the two terms in this matrix as $\mathcal{M} = A + B$ so the determinant becomes

$$\det(A+B) = \det A\ \det(1 + A^{-1}B) \quad (3.47)$$

The first operator $A = \gamma\partial/\partial t\,\delta(t-t')$ doesn't depend on the path and its determinant just gives a constant factor which can be absorbed into the normalization $\mathcal{N}$. The operator A^{-1} in the second term is defined in the usual way as

$$\int dt'\ A(t,t')A^{-1}(t,' t'') = \delta(t-t'')$$

where the integral over $\int dt'$ is simply summing over the rows of A and the columns of A^{-1} as in usual matrix multiplication. It is simple to check that the inverse is simply the step function

$$A^{-1}(t', t'') = \frac{1}{\gamma}\theta(t' - t'') \tag{3.48}$$

Now we write the second factor in (3.47) and expand,

$$\det(1 + A^{-1}B) = \exp\Big(\mathrm{Tr}\log(1 + A^{-1}B)\Big) = \exp\Big(\sum_n \mathrm{Tr}(A^{-1}B)^n/n\Big) \tag{3.49}$$

Here we should look in more detail at what this compact notation means. The term $\mathrm{Tr}A^{-1}B$ is really short-hand for

$$\mathrm{Tr}A^{-1}B = \int dt\, dt'\ A^{-t}(t, t')B(t', t)$$

where the integral over $\int dt'$ is multiplying the matrices together while the integral over $\int dt$ comes from taking the trace. Using (3.48) we have

$$\mathrm{Tr}A^{-1}B = \frac{1}{\gamma}\int dt\, dt'\ \theta(t - t')\nabla^2 V\,\delta(t - t') = \frac{\theta(0)}{\gamma}\int dt\ \nabla^2 V$$

The appearance of $\theta(0)$ may look a little odd. This function is defined to be $\theta(x) = +1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$. The only really sensible value at the origin is $\theta(0) = 1/2$. Indeed, this follows from the standard regularizations of the step function, for example

$$\theta(x) = \lim_{\mu\to 0}\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\mu}\right)\right) \quad \Rightarrow \quad \theta(0) = \frac{1}{2}$$

What happens to the higher powers of $(A^{-1}B)^n$? Writing them out, we have

$$\mathrm{Tr}(A^{-1}B)^n = \int dt \int dt_1 \ldots dt_{2n-1}\ \theta(t - t_1)\delta(t_1 - t_2)\theta(t_2 - t_3)\delta(t_3 - t_4)\ldots$$
$$\ldots\theta(t_{2n-2} - t_{2n-1})\delta(t_{2n-1} - t)\,\frac{(\nabla^2 V)^n}{\gamma^n}$$

where we have been a little sloppy in writing $(\nabla^2 V)^n$ because each of these is actually computed at a different time. We can use the delta-functions to do half of the integrals, say all the t_n for n odd. We get

$$\mathrm{Tr}(A^{-1}B)^n = \int dt \int dt_2 dt_4 \ldots\ \theta(t - t_2)\theta(t_2 - t_4)\theta(t_4 - t_6)\ldots\theta(t_{2n-2} - t)\,\frac{(\nabla^2 V)^n}{\gamma^n}$$

But this integral is only non-vanishing only if $t > t_2 > t_4 > t_6 > \ldots > t_n > t$. In other words, the integral vanishes. (Note that you might think we could again get contributions from $\theta(0) = 1/2$, but the integrals now mean that the integrand has support on a set of zero measure. And with no more delta-functions to rescue us, gives zero. The upshot of this is that the determinant (3.49) can be expressed as a single exponential

$$\det(1 + A^{-1}B) = \exp\left(\frac{1}{2\gamma}\int dt\ \nabla^2 V\right)$$

We now have an expression for the measure factor in (3.46). Using this, the path integral for the probability becomes,

$$\begin{aligned}\text{Prob}[\vec{x}_f, t_f; \vec{x}_i, t_i] &= \mathcal{N}' \int \mathcal{D}x(t)\ \exp\left(-\frac{1}{4D\gamma^2}\int dt\ (\gamma\dot{\vec{x}} + \nabla V)^2 + \frac{1}{2\gamma}\int dt\ \nabla^2 V\right) \\ &= \mathcal{N}' e^{[V(x_f) - V(x_i)]/2\gamma D} \int \mathcal{D}x(t)\ \exp\left(-\int dt\ \frac{\dot{\vec{x}}^2}{4D} + U\right)\end{aligned}$$

where U is given in (3.43). Notice that the prefactor $e^{[V(x_f)-V(x_i)]/2\gamma D}$ takes the same form as the map from probabilities P to the rescaled $\tilde{P}$ in (3.41). This completes our derivation of the path integral formulation of probabilities.

3.2.5 Stochastic Calculus

There is one final generalization of the Langevin equation that we will mention but won't pursue in detail. Let's return to the case $m = 0$, but generalise the noise term in the Langevin equation so that it is now spatially dependent. We write

$$\gamma\dot{\vec{x}} = -\nabla V + b(\vec{x})\,\vec{f}(t) \tag{3.50}$$

This is usually called the non-linear Langevin equation. The addition of the $b(\vec{x})$ multiplying the noise looks like a fairly innocuous change. But it's not. In fact, annoyingly, this equation is not even well defined!

The problem is that the system gets a random kick at time t, the strength of which depends on its position at time t. But if the system is getting a delta-function impulse at time t then its position is not well defined. Mathematically, this problem arises when we look at the position after some small time δt. Our equation (3.20) now becomes

$$\delta\vec{x} = \dot{\vec{x}}\,\delta t = -\frac{1}{\gamma}\nabla V\ \delta t + \frac{1}{\gamma}\int_t^{t+\delta t} dt'\, b(\vec{x}(t'))\,\vec{f}(t')$$

and our trouble is in making sense of the last term. There are a couple of obvious ways we could move forward:

- Ito: We could insist that the strength of the kick is related to the position of the particle immediately before the kick took place. Mathematically, we replace the integral with

$$\int_t^{t+\delta t} dt'\, b(\vec{x}(t'))\, \vec{f}(t') \;\longrightarrow\; b(\vec{x}(t)) \int_t^{t+\delta t} dt'\, \vec{f}(t')$$

  This choice is known as *Ito stochastic calculus.*

- Stratonovich: Alternatively, we might argue that the kick isn't really a delta function. It is really a process that takes place over a small, but finite, time. To model this, the strength of the kick should be determined by the average position over which this process takes place. Mathematically, we replace the integral with,

$$\int_t^{t+\delta t} dt'\, b(\vec{x}(t'))\, \vec{f}(t') \;\longrightarrow\; \frac{1}{2}\left[b(\vec{x}(t+\delta t)) + b(\vec{x}(t))\right] \int_t^{t+\delta t} dt'\, \vec{f}(t')$$

  This choice is known as *Stratonovich stochastic calculus.*

Usually in physics, issues of this kind don't matter too much. Typically, any way of regulating microscopic infinitesimals leads to the same macroscopic answers. However, this is not the case here and the Ito and Stratonovich methods give different answers in the continuum. In most applications of physics, including Brownian motion, the Stratonovich calculus is the right way to proceed because, as we argued when we first introduced noise, the delta-function arising in the correlation function $\langle f(t)f(t')\rangle$ is just a convenient approximation to something more smooth. However, in other applications such as financial modelling, Ito calculus is correct.

The subject of stochastic calculus is a long one and won't be described in this course[9]. For the Stratonovich choice, the Fokker-Planck equation turns out to be

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma}\nabla\cdot\left[P(\nabla V - D\gamma^2 b\nabla b)\right] + D\nabla^2(b^2 P)$$

This is also the form of the Fokker-Planck equation that you get by naively dividing (3.50) by $b(\vec{x})$ and the defining a new variable $\dot{\vec{y}} = \dot{\vec{x}}/b$ which reduces the problem to our previous Langevin equation (3.19). In contrast, if we use Ito stochastic calculus, the $b\nabla b$ term is absent in the resulting Fokker-Planck equation.

[9] If you'd like to look at this topic further, there are a number of links to good lecture notes on the course webpage.