Information Geometry of Spatially Periodic Stochastic Systems

Rainer Hollerbach [1],*[ID] **and Eun-jin Kim** [2][ID]

[1] Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK
[2] School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK
* Correspondence: rh@maths.leeds.ac.uk

Abstract: We explore the effect of different spatially periodic, deterministic forces on the information geometry of stochastic processes. The three forces considered are $\mathbf{f_0} = \sin(\pi x)/\pi$ and $\mathbf{f_\pm} = \sin(\pi x)/\pi \pm \sin(2\pi x)/2\pi$, with $\mathbf{f_-}$ chosen to be particularly flat (locally cubic) at the equilibrium point $x = 0$, and $\mathbf{f_+}$ particularly flat at the unstable fixed point $x = 1$. We numerically solve the Fokker–Planck equation with an initial condition consisting of a periodically repeated Gaussian peak centred at $x = \mu$, with μ in the range $[0, 1]$. The strength D of the stochastic noise is in the range 10^{-4}–10^{-6}. We study the details of how these initial conditions evolve toward the final equilibrium solutions and elucidate the important consequences of the interplay between an initial PDF and a force. For initial positions close to the equilibrium point $x = 0$, the peaks largely maintain their shape while moving. In contrast, for initial positions sufficiently close to the unstable point $x = 1$, there is a tendency for the peak to slump in place and broaden considerably before reconstituting itself at the equilibrium point. A consequence of this is that the information length $\mathcal{L}_\infty$, the total number of statistically distinguishable states that the system evolves through, is smaller for initial positions closer to the unstable point than for more intermediate values. We find that $\mathcal{L}_\infty$ as a function of initial position μ is qualitatively similar to the force, including the differences between $\mathbf{f_0} = \sin(\pi x)/\pi$ and $\mathbf{f_\pm} = \sin(\pi x)/\pi \pm \sin(2\pi x)/2\pi$, illustrating the value of information length as a useful diagnostic of the underlying force in the system.

Keywords: stochastic processes; Fokker–Planck equation; information length

1. Introduction

It is of interest to apply the idea of a metric to problems involving stochastic processes, e.g., [1–6]. Given a metric, the differences between different Probability Density Functions (PDFs) can be quantified, with different metrics focusing on a range of aspects, and hence most suitable for various applications. Fisher information [7] yields a metric where distance is measured in units of the PDF's width. The distance in the Fisher metric is thus dimensionless, and represents the number of statistically different states [8].

By extending the statistical distance in [8] to time-dependent situations, we recently introduced a way of quantifying information changes associated with time-varying PDFs [9–16]. We first compare two PDFs separated by an infinitesimal increment in time, and consider the corresponding infinitesimal distance. Integrating in time gives the total number of statistically distinguishable states that the system passes through, called the *information length* $\mathcal{L}$, e.g., [6–8,14]. Another interpretation of $\mathcal{L}$ that can be useful is as a measure of the total elapsed time in units of an 'information-change' dynamical timescale.

We start by defining the dynamical time $\tau(t)$ as

$$\mathcal{E} \equiv \frac{1}{[\tau(t)]^2} = \int \frac{1}{p(x,t)} \left[\frac{\partial p(x,t)}{\partial t} \right]^2 dx. \tag{1}$$

That is, $\tau(t)$ is the characteristic timescale over which the information changes, and quantifies the PDF's correlation time. Alternatively, $1/\tau$ quantifies the (average) rate of change of information in time. A PDF that evolves such that $\mathcal{E}$ is constant in time is referred to as a geodesic, along which the information propagates at a uniform rate [6]. The information length $\mathcal{L}(t)$ is then defined by

$$\mathcal{L}(t) = \int_0^t \frac{dt_1}{\tau(t_1)} = \int_0^t \sqrt{\int dx \frac{1}{p(x,t_1)} \left[\frac{\partial p(x,t_1)}{\partial t_1} \right]^2} \, dt_1. \tag{2}$$

which can be interpreted as measuring time in units of τ. It is important to note that $\mathcal{L}$ has no dimension (unlike entropy) and represents the total number of statistically different states that a system passes through in time between 0 and t. If we know the parameters that determine the PDF $p(x,t)$, $\mathcal{E}$ and $\mathcal{L}$ in Equations (1) and (2) can be written in terms of the Fisher metric tensor defined in the statistical space spanned by those parameters. However, it is not always possible to have access to the parameters that govern PDFs, for instance, in the case of PDFs calculated from data. The merit of Equations (1) and (2) is thus that $\mathcal{E}$ and $\mathcal{L}$ can be directly calculated from PDFs even without knowing the parameters governing the PDFs, nor the Fisher metric. For instance, $\mathcal{L}$ was calculated from PDFs of music data in [10]. In the work here, we first compute time-dependent PDFs by solving the Fokker–Planck equation numerically, and then calculate $\mathcal{E}$ and $\mathcal{L}$ from these PDFs as additional diagnostics.

Unlike quantities such as entropy, relative entropy, Kullback–Leibler divergence, or Jensen divergence, information length is a Lagrangian measure, that is, it includes the full details of the PDF's evolution, and not just the initial and final states. $\mathcal{L}_\infty$, the total information length over the entire evolution, is then particularly useful to quantify the proximity of any initial PDF to a final attractor of a dynamical system. In previous work [12,15] we explored these aspects of $\mathcal{L}$ for restoring forces that were power-laws in the distance to the attractor. For instance, for the Ornstein–Uhlenbeck process, which is a linear relaxation process, we showed that $\mathcal{L}$ consists of two parts: the first is due to the movement of the mean position measured in units of the width of the PDF, and the second is due to the entropy change. Thus, the total entropy change that is often discussed in previous works (e.g., [17]) contributes only partially to $\mathcal{L}$. Importantly, for the Ornstein–Uhlenbeck process, $\mathcal{L}_\infty$ increases linearly from the stable equilibrium point (with its minimum value at the stable equilibrium point) with the mean position of the initial PDFs regardless of the strength of the stochastic noise and the width of the initial PDFs. The linear relation indicates that a linear process preserves a linearity of the underlying process. Heseltine & Kim [18] shows that this linear relation is lost for other metrics (e.g., Kullback–Leibler divergence, Jensen divergence). Note that $\mathcal{L}$ is related to the integral of the square root of the infinitesimal relative entropy (see Appendix A). In comparison, for a chaotic attractor, $\mathcal{L}_\infty$ varies sensitively with the mean position of a narrow initial PDF, taking its minimum value at the most unstable point [9]. This sensitive dependence of $\mathcal{L}_\infty$ on the initial PDF is similar to a Lyapunov exponent.

These results highlight $\mathcal{L}_\infty$ as an alternative diagnostic to understand attractor structures of dynamical systems. It is this attractor structure that we are interested in in this paper. We thus focus on the relaxation problem as in [9,12,15,18] by considering periodic deterministic forces and elucidate the importance of the initial condition and its interplay with the deterministic forces in the relaxation and thus attractor structure.

2. Model

We consider the following nonlinear Langevin equation:

$$\frac{dx}{dt} = -f(x) + \xi. \tag{3}$$

Here x is a random variable; $f(x)$ is a deterministic force; ξ is a stochastic forcing, which for simplicity can be taken as a short-correlated Gaussian random forcing as follows:

$$\langle \xi(t)\xi(t') \rangle = 2D\delta(t - t'), \tag{4}$$

where the angular brackets represent the average over ξ, $\langle \xi \rangle = 0$, and D is the strength of the forcing.

In [15] we considered the choice $f(x) = x^n$ and investigated how varying the degree of nonlinearity $n = 3, 5, 7$ affects the system. In this work we take $f(x)$ to be periodic in x, and explore some of the new effects this can create. The three choices of $f(x)$ we consider are

$$\mathbf{f_0} = \sin(\pi x)/\pi, \qquad \mathbf{f_\pm} = \sin(\pi x)/\pi \pm \sin(2\pi x)/2\pi. \tag{5}$$

Figure 1 shows these profiles, which are all anti-symmetric in x, and periodic on the interval $x \in [-1, 1]$. All three choices have $x = 0$ as an attractor, and $x = 1$ as an unstable fixed point. The particular combinations of harmonics for $\mathbf{f_\pm}$ were chosen so that they are locally cubic rather than linear at either $x = 0$ (for $\mathbf{f_-}$) or $x = 1$ (for $\mathbf{f_+}$). In applications such a Brownian motors many specific choices of $f(x)$ are considered to model particular physics. However, as noted in the introduction, we are here more interested in attractor structures in the relaxation problem, in particular, how initial conditions and stochastic noise interact with deterministic forces and the role of the asymmetry of the deterministic force and the stable and unstable fixed points on the local dynamics.
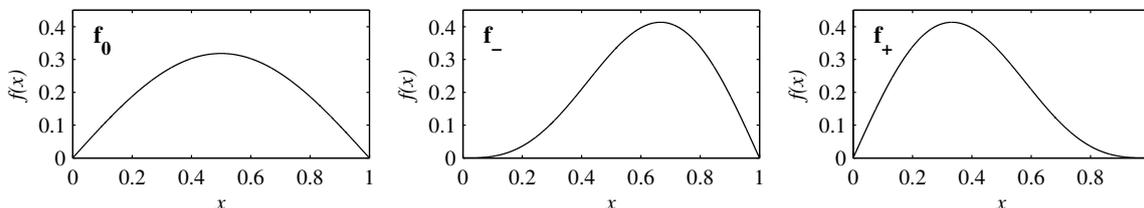
Figure 1. The three profiles $\mathbf{f_0}$, $\mathbf{f_-}$ and $\mathbf{f_+}$. Note how $\mathbf{f_-}$ is chosen to be flat at the attractor $x = 0$, and $\mathbf{f_+}$ at the unstable fixed point $x = 1$. All three choices are anti-symmetric in x, and periodic with period 2.

Comparing these three periodic functions with the previous choices, two significant differences stand out. First, for $f(x) = x^n$ with $n = 3, 5, 7$, all initial conditions are pushed directly toward the origin, and there are no unstable fixed points. It is therefore of particular interest to see how the choices here behave for initial conditions near $x = 1$. Second, $f(x) = x^{3,5,7}$ all curve upward (that is, have $f'' > 0$ for all $x > 0$), whereas the choices here have different combinations of curvatures, which will turn out to have clearly identifiable effects.

The Fokker–Planck equation [19,20] corresponding to Equation (3) is

$$\partial_t p(x, t) = \partial_x(f(x)p) + D\partial_{xx}p. \tag{6}$$

In [15] we solved the corresponding equation by finite-differencing in x. For the periodic systems considered here, it is more convenient to start with the Fourier expansion

$$p(x, t) = a_0(t) + \sum_{k=1}^{K} \Big(a_k(t)\cos(k\pi x) + b_k(t)\sin(k\pi x) \Big). \tag{7}$$

The coefficients a_k and b_k are then time-stepped using second-order Runge-Kutta. The term $\partial_x(f(x)p)$ is separated out into the relevant Fourier components using a fast Fourier transform. (For the very simple choices of $f(x)$ considered here, consisting of at most two Fourier modes, it would be straightforward to do this separation analytically, and thereby do the entire calculation purely in Fourier space, but the code was developed with more general choices for $f(x)$ in mind, where this approach becomes increasingly cumbersome as the number of harmonics in $f(x)$ increases. For such more general choices of $f(x)$ the FFT approach is most convenient).

Resolutions in the range $K = 2^{11} - 2^{14}$ are used, and carefully checked to ensure fully resolved solutions. Time-steps were in the range 10^{-4}–10^{-5}, and were again varied to ensure proper accuracy. Another useful test of the numerical implementation is to monitor the coefficient a_0: this is time-stepped along with the others, but must in fact remain constant if the total probability $\int p\,dx = 2a_0$ is to remain constant. It was found that if the initial condition is correctly set to have $a_0 = 0.5$, then this was maintained throughout the entire subsequent evolution.

The initial conditions are of the form

$$p(x,0) = \frac{1}{\sqrt{2\pi D_0}} \exp\left[-\frac{(x-\mu)^2}{2D_0}\right], \tag{8}$$

that is, Gaussians centred at $x = \mu$ and having half-width scaling as $\sqrt{D_0}$. We are interested in the range $\mu \in [0,1]$; by symmetry the range $\mu \in [-1,0]$ would behave the same, simply approaching $x = 0$ from the other direction.

This initial condition is also periodic, on the same $x \in [-1,1]$ interval as the entire problem. For the purposes of actually implementing Equation (8), it was most convenient to consider the range as being $x \in [-0.5, 1.5]$. In particular, for $\mu \in [0,1]$ and the values of D_0 considered here, Equation (8) yields results at $x = -0.5$ and $x = 1.5$ that are different, but both are so vanishingly small that the discrepancy does not need to be smoothed out in defining the initial condition. If instead Equation (8) were implemented on either $x \in [0,2]$ or $x \in [-1,1]$, then μ near either 0 or 1 would be more awkward to handle correctly.

In [15] we also used a Gaussian initial condition, with $D_0 = 10^{-6}$, and then explored the regime $D = 10^{-6}$ to 10^{-9}. Here we are again interested in the regime $D_0 \geq D$, which allows at least the initial parts of the evolution to be nondiffusive. Having the initial peak be so narrow that $D_0 < D$ can also be interesting in other contexts (e.g., [21]), but diffusive effects are then necessarily important from the outset, which would obscure some of the dynamics of interest here. We therefore focus on the range $D_0 = 10^{-2}$ to 10^{-4}, and $D = 10^{-4}$ to 10^{-6}.

3. Results

Figure 2 shows how the peak amplitudes evolve in time for the three choices $\mathbf{f_0}$, $\mathbf{f_-}$ and $\mathbf{f_+}$. Starting with the initial position $\mu = 0.5$ in the top row, we see that the solutions for $\mathbf{f_0}$ and $\mathbf{f_+}$ equilibrate to their final values on very rapid timescales, involving relatively little variation with D. In contrast, the timescales for $\mathbf{f_-}$ are much longer, and vary substantially with D. Comparing the $\mathbf{f_-}$ results here with Figure 1 in [15], we see that $\mathbf{f_-}$ is exactly analogous to the previous $f(x) = x^3$. This is because for $x \leq 0.5$ the shape of $\mathbf{f_-}$ is very close to a cubic. Similarly, for $x \leq 0.5$ the shape of $\mathbf{f_0}$ is still reasonably close to linear, and the evolution is therefore essentially like the linear Ornstein–Uhlenbeck process $f(x) = x$, for which an exact analytic solution exists [21].
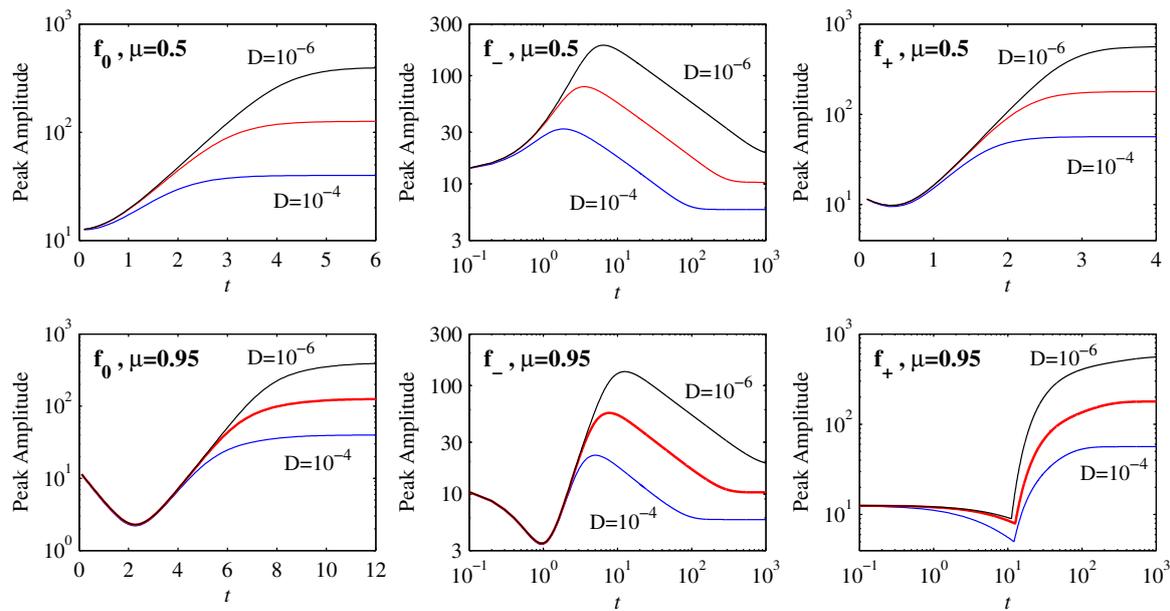
Figure 2. Peak amplitudes as functions of time, for the three choices $\mathbf{f}_*$ as labelled. The initial positions are at $\mu = 0.5$ in the top row, and $\mu = 0.95$ in the bottom row, with initial widths $D_0 = 10^{-3}$ in all cases. $D = 10^{-4}$ to 10^{-6} as labelled, also colour-coded as blue, red, black. Note also that some panels have t on a linear scale, indicating a very rapid adjustment process, whereas others have t on a logarithmic scale, corresponding to much slower dynamics.

It is only $\mathbf{f}_+$ whose shape is already substantially different from either linear or cubic even on the interval $x \le 0.5$, being close to linear for $x \le 0.2$ but strongly curved for $0.2 < x \le 0.5$. Correspondingly $\mathbf{f}_+$ also shows a new effect, namely an initial reduction in the peak amplitudes. This effect becomes even more pronounced for $\mathbf{f_0}$ and $\mathbf{f}_-$ and the initial position $\mu = 0.95$, in the bottom row of Figure 2. This reduction in the peak amplitudes is *not* caused by diffusive spreading but is a consequence of the non-diffusive ($\xi = D = 0$) evolution resulting from the interplay between an initial PDF and the deterministic force. We note in particular how $D = 10^{-4}$ to 10^{-6} yield identical reductions in amplitudes here. It is worth comparing this with the non-diffusive evolution in [15] where the opposite behavior—an initial increase in peak amplitudes (the same effect as seen here for $\mathbf{f}_-$, $\mu = 0.5$)—was observed. The interplay between the initial PDF and the deterministic force is elaborated below.

If $f(x)$ is such that it increases more rapidly than linearly, i.e., curves upward, then those parts of any initial condition furthest from the origin are pushed toward it fastest, whereas those parts closest move more slowly. The result is that an initial Gaussian peak bunches up on itself, causing the amplitude to increase. In contrast, if $f(x)$ curves downward the opposite effect occurs, and an initial Gaussian peak is spread out, even before diffusion starts to play a role. Eventually of course the peak moves sufficiently close to the origin that the behaviour is as before, explaining why the behaviour at later times is similar to the previous $\mu = 0.5$ results.

Finally, the behaviour for $\mathbf{f}_+$ with $\mu = 0.95$ is yet again different, namely an initial reduction in amplitude up to $t \approx 10$, followed by an abrupt increase. This is caused by a fundamentally new peak forming at the origin, rather than the initial peak moving toward it. Note also that time here is on a logarithmic scale, corresponding to a very slow equilibration process, unlike the previous case $\mathbf{f}_+$ with $\mu = 0.5$.

Figures 3–5 illustrate these various behaviours in more detail, showing the actual PDFs at different times for $\mathbf{f_0}$, $\mathbf{f}_-$ and $\mathbf{f}_+$, respectively. Starting with $\mathbf{f_0}$, we see how the peak initially located at $\mu = 0.95$ becomes broader as it moves toward the origin, an effect again not caused by diffusion, but rather by the curvature of $\mathbf{f_0}$ at these values of x. Note for example how the solutions at $t = 3$ or 4 have much

steeper leading edges (nearer to the origin) than trailing edges, caused by the trailing edges moving so much slower. Another feature to note is how parts of the solution reach the origin coming from the 'other' direction. That is, if the initial condition is a peak centred at $\mu = 0.95$, and having half-width 0.07 (corresponding to $D_0 = 10^{-3}$), then a small but non-negligible portion of the initial condition is in the range $x \geq 1$, as seen also in Figure 3. For this part of the initial condition the nearest attractor is $x = 2$ rather than $x = 0$. Viewed on the interval $x \in [-0.2, 1.2]$, this part therefore approaches from negative x values, as seen at times $t = 5$ and 6. (The interval $x \in [1.2, 1.8]$ or equivalently $x \in [-0.8, -0.2]$ is not shown in these figures because the amplitudes are rather small there, due to the PDFs being very spread out as they traverse this range). Finally, between $t = 6$ and 9 we see how the two peaks coming from negative and positive x values combine to form the single final equilibrium consisting of a Gaussian centred at the origin.
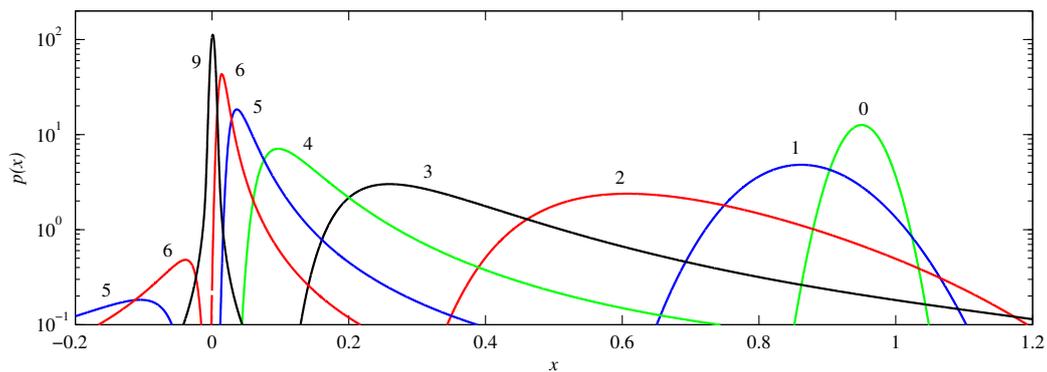
Figure 3. The solutions $p(x, t)$ for $\mathbf{f_0}$, $D = 10^{-5}$, and initial condition $\mu = 0.95$ and $D_0 = 10^{-3}$. The numbers beside individual curves indicate the times, from $t = 0$ to 9. The different colours are for clarity only, but do not indicate a specific colour $\leftrightarrow$ time relationship.
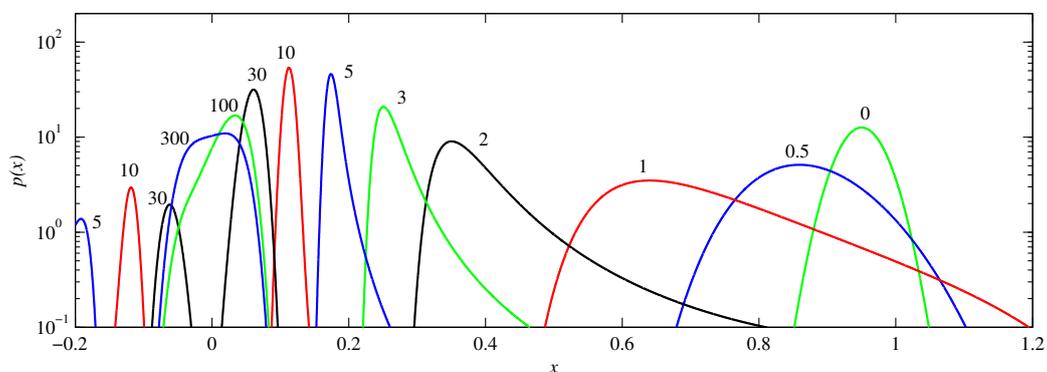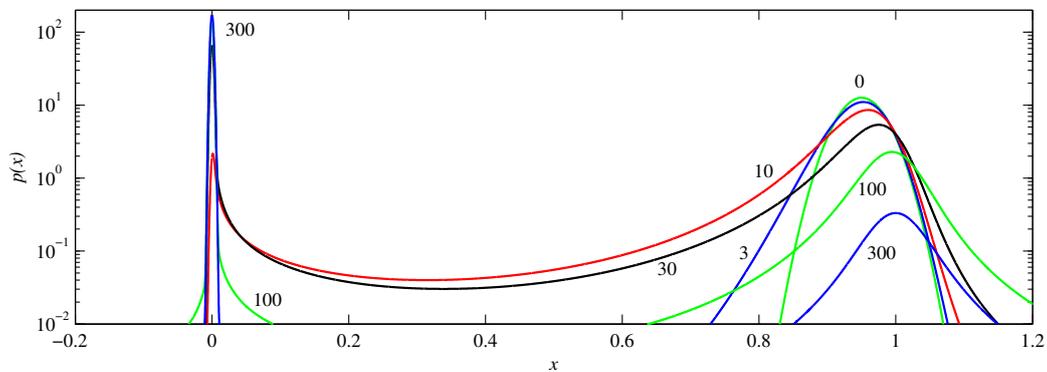

Figure 4. The solutions $p(x, t)$ for $\mathbf{f_-}$, $D = 10^{-5}$, and initial condition $\mu = 0.95$ and $D_0 = 10^{-3}$. The numbers beside individual curves indicate the times, from $t = 0$ to 300. Different colours are again only to help distinguish between the different lines.

Figure 4 shows the corresponding solutions for $\mathbf{f_-}$. For small durations the behaviour is very similar to that seen in Figure 3, except that it happens roughly twice faster (e.g., compare $t = 0.5$ in Figure 4 with $t = 1$ in Figure 3). This is readily understandable by noting that the slope of $\mathbf{f_-}$ near $x = 0.95$ is roughly twice that of $\mathbf{f_0}$, yielding faster evolution. The later evolution is much slower though, with the merging of the two peaks only occurring between $t = 30$ and 100, and even $t = 300$ still displaying some asymmetry, and hence not yet the final quartic profile. This is the same very slow final adjustment process previously analysed in detail in [15], and is caused by $\mathbf{f_-}$ being cubic rather than linear near the origin.

Figure 5 shows the solutions for $\mathbf{f_+}$. We see the behaviour alluded to above, of an abrupt transition from one peak to another. Because $\mathbf{f_+}$ is so flat near $x = 1$, there is hardly any tendency to push the

initial peak away. Instead, it simply broadens out, slumping as it spreads. A new peak then forms at the origin, overtaking the original one in amplitude around $t \approx 10$, as previously noted in Figure 2. Note though that long after this time a significant portion of the original peak still remains near $x = 1$, and this portion only fades away on very long timescales; $x = 1$ is an unstable fixed point, but $\mathbf{f}_+$ is so small everywhere near $x = 1$ that there is very little tendency to push the solutions away from there.


Figure 5. The solutions $p(x, t)$ for $\mathbf{f}_+$, $D = 10^{-5}$, and initial condition $\mu = 0.95$ and $D_0 = 10^{-3}$. The numbers beside individual curves indicate the times, from $t = 0$ to 300.

As noted in the introduction, we are particularly interested in the effects that these various different types of behaviour have on the information length quantities $\mathcal{E}(t)$ and $\mathcal{L}_\infty$. Figure 6 shows $\mathcal{E}(t)$ for the same solutions as before in Figure 2. We see that $\mathcal{E}$ is initially uniform, and independent of D (provided D is sufficiently small in comparison with D_0), corresponding to the 'geodesic' behaviour first identified by [6]. For some configurations, $\mathcal{E}$ then immediately transitions to an exponential decay, whereas for others it first has a power-law decay before ultimately decaying exponentially. Correspondingly, the timescales to achieve $\mathcal{E} \leq 10^{-8}$ also vary dramatically, as seen by the various linear and logarithmic scales for t. Different scaling regimes signify fundamentally different dynamics.
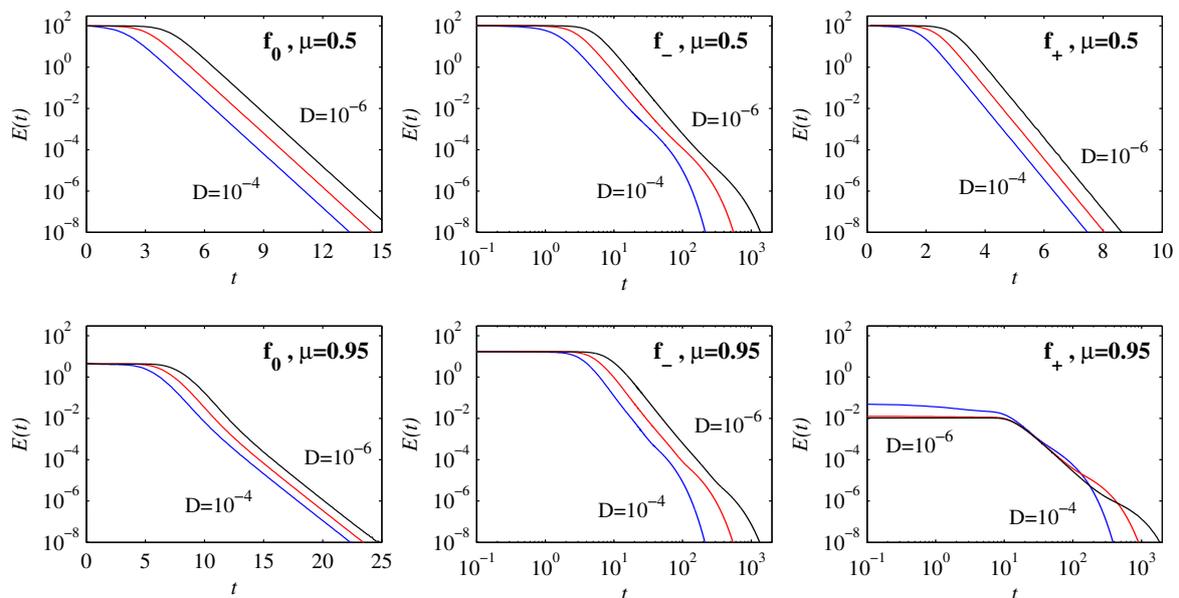
Figure 6. $\mathcal{E}$ as a function of time, for the six configurations as in Figure 2. Note again how the scale for t is sometimes linear and sometimes logarithmic.
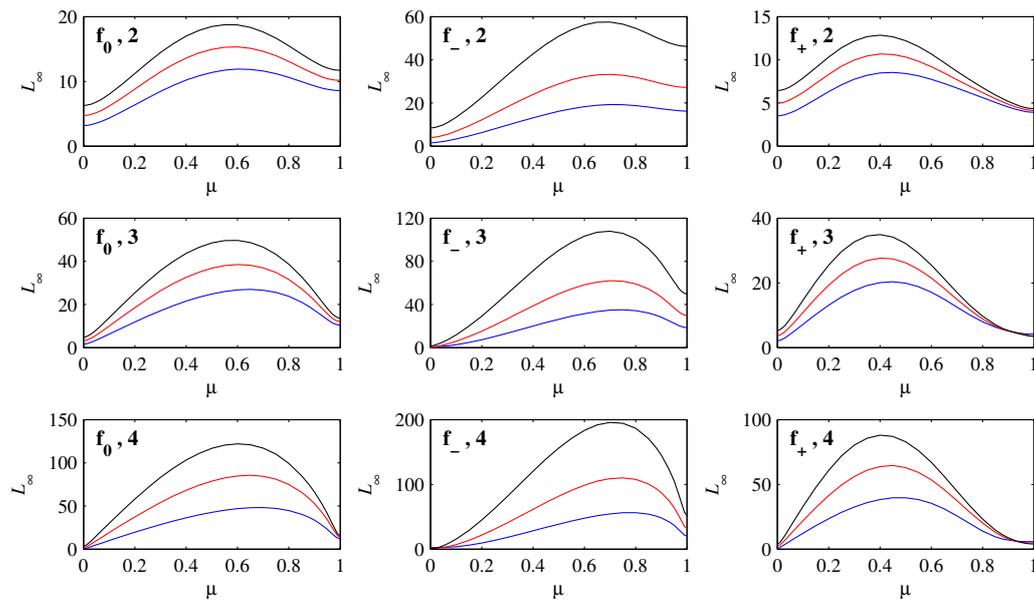
Figure 7. The total information length $\mathcal{L}_\infty$ as a function of initial position μ. The labels $\mathbf{f}_*$, $\mathbf{n}$ indicate the particular profile $\mathbf{f}_*$, and the initial width $D_0 = 10^{-n}$. Within each panel the three curves are $D = 10^{-4}$ (lowest, blue), $D = 10^{-5}$ (middle, red), and $D = 10^{-6}$ (top, black).

Figure 7 shows how $\mathcal{L}_\infty$ varies with μ, for $D_0 = 10^{-2}$ to 10^{-4}, and $D = 10^{-4}$ to 10^{-6} within each panel. It is interesting to note how the shapes generally mimic the corresponding functions $\mathbf{f}_0$, $\mathbf{f}_-$ and $\mathbf{f}_+$. The largest values always occur for intermediate values of μ, even though larger values correspond to initial conditions that have farther to travel to reach the origin. Such initial conditions also spread out much more though, as seen above, and according to the interpretation of information length, this should indeed reduce $\mathcal{L}$. Very close to $x = 1$ the $\mathcal{L}_\infty$ values are particularly small, because having peaks collapse in place and reform at the new location is an informationally very efficient way to move, as seen also in other contexts [13,22,23].

Finally, Figure 8 shows the time, call it T_∞, needed for $\mathcal{E}$ to drop to 10^{-8}. The precise cutoff $\mathcal{E} = 10^{-8}$ is of course somewhat arbitrary, but as seen in Figure 6 is sufficiently small to be in the exponential decay regime in all cases. This is therefore a convenient measure of the time taken to reach $\mathcal{L}_\infty$, and any even smaller cutoff would only add small increments to T_∞ (and essentially nothing to $\mathcal{L}_\infty$).

Starting with $\mathbf{f}_0$, we note first that T_∞ is on a linear scale, meaning that each reduction of D by a factor of 10 only adds a constant amount to T_∞. This is the same effect already seen in Figure 2, where smaller D requires slightly longer to settle in to the final states. Equivalently, smaller D in Figure 6 remains in the flat, geodesic regime for slightly longer times. The other feature to note for $\mathbf{f}_0$ is the behaviour near $\mu = 1$, where T_∞ increases strongly, and increasingly abruptly for smaller D_0. This can be understood by noting that if $\sqrt{D_0} \ll 1 - \mu$, the initial condition Equation (8) is essentially zero at $x = 1$, whereas if μ is within $\sqrt{D_0}$ of 1, Equation (8) does have a non-negligible component at $x = 1$. Therefore, if $\sqrt{D_0} \ll 1 - \mu$ the initial peak will simply move monotonically toward the origin, which occurs on a rapid timescale, whereas if $1 - \mu \le \sqrt{D_0}$ the evolution will include a significant component of the slumping-in-place behaviour, which we saw only happens on slower timescales.
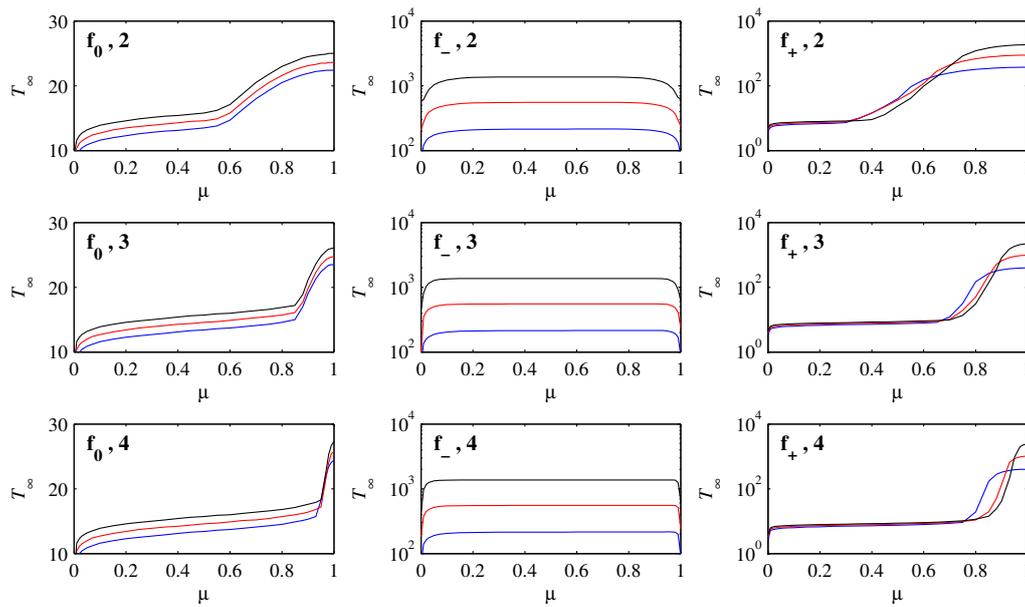
Figure 8. As in Figure 7, but now showing the time T_∞ that it takes to achieve $\mathcal{E} = 10^{-8}$, that is, a measure of the time it takes to reach $\mathcal{L}_\infty$. Note how T_∞ has a linear scale for $\mathbf{f_0}$, but logarithmic scales for $\mathbf{f_-}$ and $\mathbf{f_+}$. The colour-coding is again $D = 10^{-4}$ blue, $D = 10^{-5}$ red, $D = 10^{-6}$ black.

For $\mathbf{f_-}$, the scale for T_∞ is logarithmic, so that each reduction of D by a factor of 10 increases T_∞ by a factor of $\sqrt{10}$. For intermediate values of μ, T_∞ is also essentially independent of μ. The equilibration time is completely dominated by the final settling-in time, just as in the cubic case in [15], and the initial motion of the peak toward the origin is negligible in comparison. For very small values of μ the behaviour is different, with much smaller values of T_∞. If $\mu \leq \sqrt{D_0}$, the peak is essentially at the origin already, making the adjustment quicker. Finally, there is a similar end-effect for μ sufficiently close to 1; if $1 - \mu \leq \sqrt{D_0}$, the initial peak is essentially at the unstable fixed point, and the evolution is the slumping-in-place behaviour, which has a faster final adjustment than if the peak moves toward the origin and then adjusts its shape there (but still scaling as $D^{-1/2}$).

Finally, $\mathbf{f_+}$ is qualitatively similar to $\mathbf{f_0}$, in the sense that T_∞ is a monotonically increasing function of μ. Indeed, for intermediate values of μ the behaviour is virtually identical to $\mathbf{f_0}$, with T_∞ increasing by a constant amount every time D is decreased by a constant factor. (This is simply not visible because T_∞ is on a logarithmic rather than linear scale here). Because $\mathbf{f_0}$ and $\mathbf{f_+}$ are both linear near the origin, the extremely slow final adjustment that happens for $\mathbf{f_-}$ does not apply to either of them, leaving only this much weaker dependence on D. The behaviour near $\mu = 1$, with the very strong increase in T_∞, and again more abruptly for smaller D_0, is again because this is the regime where the slumping-in-place behaviour occurs. Also, because $\mathbf{f_+}$ is so much flatter near $x = 1$ than either of $\mathbf{f_0}$ of $\mathbf{f_-}$, this slumping-in-place behaviour is much slower for $\mathbf{f_+}$ than for the other choices (recall again how long the peak at $x = 1$ lasts in Figure 5). This explains why T_∞ is on a logarithmic scale for $\mathbf{f_+}$ but on a linear scale for $\mathbf{f_0}$, even though for intermediate values of μ they exhibit the same (weak) scaling with D.

4. Conclusions

The results presented here extend our previous work [12,15] to the deterministic forces that are periodic in space. This naturally allows for forces $f(x)$ that curve in opposite directions in different regions, as well as unstable fixed points. The deterministic force can also be adjusted to be particularly flat at either the stable equilibrium or the unstable fixed points, which both turn out to have important consequences, with either choice yielding particularly long timescales, scaling as $D^{-1/2}$. The interesting consequences of the interplay between an initial PDF and a deterministic force on the PDF evolution

and scalings of different quantities were discussed in detail by considering three types of periodic deterministic forces and comparing the results.

In particular, we computed how the rate of information change and the resulting total information length $\mathcal{L}_\infty$ depend on the position of an initial Gaussian peak. We found that for all choices of $f(x)$, the unstable fixed points yield comparatively small $\mathcal{L}_\infty$, even though they are farthest away from the final equilibrium points. It is particularly interesting that $\mathcal{L}_\infty$ as a function of initial position qualitatively follows $f(x)$, indicating the close connection between the information geometry and the underlying forcing.

Finally, we note that this work can be extended in many different directions, including: (1) If the initial condition is not one Gaussian peak for every period of $f(x)$, but only one peak for some much larger (tending to infinity) number of periods of $f(x)$, then one can study how this initial peak gradually spreads out, how that compares with pure diffusion, and what effect the precise shape of $f(x)$ might have [24–26]. This is also related to so-called anomalous diffusion [27–29], which can be considerably more general though, involving fractional derivatives. (2) If the force also includes a constant component tending to push the initial condition in a particular direction, it is of interest to study situations where the constant force is comparable to the periodic component of the force [30–32]. (3) Allowing the force to vary in time as well as space is relevant to so-called Brownian motors [33–36]. Work in some of these areas is currently ongoing.

Author Contributions: The underlying ideas were developed by E.K. and R.H. Numerical calculations were done by R.H. The paper was written by R.H. and E.K.; both authors have read and approved the final manuscript.

Appendix A

We now recall how $\tau(t)$ and $\mathcal{L}(t)$ in Equations (1) and (2) are related to the relative entropy (Kullback–Leibler divergence) [6,11]. We consider two nearby PDFs $p_1 = p(x, t_1)$ and $p_2 = p(x, t_2)$ at time $t = t_1$ and t_2, and the limit of a very small $\delta t = t_2 - t_1$ to do Taylor expansion of $D[p_1, p_2] = \int dx\, p_2 \ln(p_2/p_1)$ by using

$$\frac{\partial}{\partial t_1} D[p_1, p_2] = -\int dx\, p_2 \frac{\partial_{t_1} p_1}{p_1}, \tag{A1}$$

$$\frac{\partial^2}{\partial t_1^2} D[p_1, p_2] = \int dx\, p_2 \left\{ \frac{(\partial_{t_1} p_1)^2}{p_1^2} - \frac{\partial_{t_1}^2 p_1}{p_1} \right\}, \tag{A2}$$

$$\frac{\partial}{\partial t_2} D[p_1, p_2] = \int dx \left\{ \partial_{t_2} p_2 + \partial_{t_2} p_2 \left[\ln p_2 - \ln p_1 \right] \right\}, \tag{A3}$$

$$\frac{\partial^2}{\partial t_2^2} D[p_1, p_2] = \int dx \left\{ \partial_{t_2}^2 p_2 + \frac{(\partial_{t_2} p_2)^2}{p_2} + \partial_{t_2}^2 p_2 \left[\ln p_2 - \ln p_1 \right] \right\}. \tag{A4}$$

In the limit $t_2 \to t_1 = t$ $(p_2 \to p_1 = p)$, Equations (A1)–(A4) give us

$$\lim_{t_2 \to t_1} \frac{\partial}{\partial t_1} D[p_1, p_2] = \lim_{t_2 \to t_1} \frac{\partial}{\partial t_2} D[p_1, p_2] = \int dx \partial_t p = 0,$$

$$\lim_{t_2 \to t_1} \frac{\partial^2}{\partial t_1^2} D[p_1, p_2] = \lim_{t_2 \to t_1} \frac{\partial^2}{\partial t_2^2} D[p_1, p_2] = \int dx \frac{(\partial_t p)^2}{p} = \frac{1}{\tau^2}. \tag{A5}$$

Up to $O((dt)^2)$ $(dt = t_2 - t_1)$, Equation (A5) and $D(p_1, p_1) = 0$ lead to

$$D[p_1, p_2] = \frac{1}{2} \left[\int dx \frac{(\partial_t p(x, t))^2}{p(x, t)} \right] (dt)^2, \tag{A6}$$

and thus the infinitesimal distance $dl(t_1)$ between t_1 and $t_1 + dt$ as

$$dl(t_1) = \sqrt{D[p_1, p_2]} = \frac{1}{\sqrt{2}} \sqrt{\int dx \frac{(\partial_{t_1} p(x, t_1))^2}{p(x, t_1)} dt}. \tag{A7}$$

By summing $dt(t_i)$ for $i = 0, 1, 2, ..., n-1$ (where $n = t/dt$) in the limit $dt \to 0$, we have

$$\lim_{dt \to 0} \sum_{i=0}^{n-1} dl(idt) = \lim_{dt \to 0} \sum_{i=0}^{n-1} \sqrt{D[p(x, idt), p(x, (i+1)]} \, dt \propto \int_0^t dt_1 \sqrt{\int dx \frac{(\partial_{t_1} p(x, t_1))^2}{p(x, t_1)}} = \mathcal{L}(t), \tag{A8}$$

where $\mathcal{L}(t)$ is the information length. Thus, $\mathcal{L}$ is related to the sum of infinitesimal relative entropy. Note that $\mathcal{L}$ is a Lagrangian distance between PDFs at time 0 and t and sensitively depends on the particular path that a system passed through reaching the final state. In contrast, the relative entropy $D[p(x, 0), p(x, t)]$ depends only on PDFs at time 0 and t and thus does not tell us about intermediate states between initial and final states.

References

1. Jordan, R.; Kinderlehrer, D.; Otto, F. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **1998**, *29*, 1–17. [CrossRef]
2. Gibbs, A.L.; Su, F.E. On choosing and bounding probability metrics. *Int. Stat. Rev.* **2002**, *70*, 419–435. [CrossRef]
3. Lott, J. Some geometric calculations on Wasserstein space. *Commun. Math. Phys.* **2008**, *277*, 423–437. [CrossRef]
4. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
5. Costa, S.; Santos, S.; Strapasson, J. Fisher information distance. *Discrete Appl. Math.* **2015**, *197*, 59–69. [CrossRef]
6. Kim, E.; Lee, U.; Heseltine, J.; Hollerbach, R. Geometric structure and geodesic in a solvable model of nonequilibrium process. *Phys. Rev. E* **2016**, *93*, 062127. [CrossRef] [PubMed]
7. Frieden, B.R. *Science from Fisher Information*; Cambridge University Press: Cambridge, UK, 2004.
8. Wootters, W.K. Statistical distance and Hilbert space. *Phys. Rev. D* **1981**, *23*, 357–362. [CrossRef]
9. Nicholson, S.B.; Kim, E. Investigation of the statistical distance to reach stationary distributions. *Phys. Lett. A* **2015**, *379*, 83–88. [CrossRef]
10. Nicholson, S.B.; Kim, E. Structures in sound: Analysis of classical music using the information length. *Entropy* **2016**, *18*, 258. [CrossRef]
11. Heseltine, J.; Kim, E. Novel mapping in non-equilibrium stochastic processes. *J. Phys. A* **2016**, *49*, 175002. [CrossRef]
12. Kim, E.; Hollerbach, R. Signature of nonlinear damping in geometric structure of a nonequilibrium process. *Phys. Rev. E* **2017**, *95*, 022137. [CrossRef] [PubMed]
13. Hollerbach, R.; Kim, E. Information geometry of non-equilibrium processes in a bistable system with a cubic damping. *Entropy* **2017**, *19*, 268. [CrossRef]
14. Kim, E.; Lewis, P. Information length in quantum systems. *J. Stat. Mech.* **2018**, 043106. [CrossRef]
15. Hollerbach, R.; Dimanche, D.; Kim, E. Information geometry of nonlinear stochastic systems. *Entropy* **2018**, *20*, 550. [CrossRef]
16. Hollerbach, R.; Kim, E.; Mahi, Y. Information length as a new diagnostic in the periodically modulated double-well model of stochastic resonance. *Phys. A* **2019**, *525*, 1313–1322. [CrossRef]
17. Van den Broeck, C.; Esposito, M. Three faces of the second law. II. Fokker–Planck formulation. *Phys. Rev. E* **2010**, *82*, 011144. [CrossRef] [PubMed]
18. Heseltine, J; Kim, E. Comparing information metrics for a coupled Ornstein–Uhlenbeck process. **2019**, in preparation.
19. Risken, H. *The Fokker–Planck Equation: Methods of Solution and Applications*; Springer: Berlin, Germany, 1996.
20. Klebaner, F. *Introduction to Stochastic Calculus with Applications*; Imperial College Press: London, UK, 2012.

21. Kim, E.; Hollerbach, R. Time-dependent probability density function in cubic stochastic processes. *Phys. Rev. E* **2016**, *94*, 052118. [CrossRef]

22. Kim, E.; Tenkès, L.-M.; Hollerbach, R.; Radulescu, O. Far-from-equilibrium time evolution between two gamma distributions. *Entropy* **2017**, *19*, 511. [CrossRef]

23. Kim, E.; Hollerbach, R. Geometric structure and information change in phase transitions. *Phys. Rev. E* **2017**, *95*, 062107. [CrossRef]

24. Dean, D.S.; Gupta, S.; Oshanin, G.; Rosso, A.; Schehr, G. Diffusion in periodic, correlated random forcing landscapes. *J. Phys. A* **2014**, *47*, 372001. [CrossRef]

25. Dean, D.S.; Oshanin, G. Approach to asymptotically diffusive behavior for Brownian particles in periodic potentials: Extracting information from transients. *Phys. Rev. E* **2014**, *90*, 022112. [CrossRef] [PubMed]

26. Sivan, M.; Farago, O. Probability distribution of Brownian motion in periodic potentials. *Phys. Rev. E* **2018**, *98*, 052117. [CrossRef]

27. Metzler, R.; Jeon, J.-H.; Cherstvy, A.G.; Barkai, E. Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **2014**, *16*, 24128–24164. [CrossRef] [PubMed]

28. Zaburdaev, V.; Denisov, S.; Klafter, J. Lévy walks. *Rev. Mod. Phys.* **2015**, *87*, 483–530. [CrossRef]

29. Evangelista, L.R.; Lenzi, E.K. *Fractional Diffusion Equations and Anomalous Diffusion*; Cambridge University Press: Cambridge, UK, 2018.

30. Risken, H.; Vollmer, H.D. Brownian motion in periodic potentials; nonlinear response to an external force. *Z. Phys. B* **1979**, *33*, 297–305. [CrossRef]

31. Lindner, B.; Schimansky-Geier, L.; Reimann, P.; Hänggi, P.; Nagaoka, M. Inertia ratchets: A numerical study versus theory. *Phys. Rev. E* **1999**, *59*, 1417–1424. [CrossRef]

32. Guérin, T.; Dean, D.S. Universal time-dependent dispersion properties for diffusion in a one-dimensional critically tilted potential. *Phys. Rev. E* **2017**, *95*, 012109. [CrossRef]

33. Reimann, P. Brownian motors: Noisy transport far from equilibrium. *Phys. Rep.* **2002**, *361*, 57–265. [CrossRef]

34. Vorotnikov, D. Analytical aspects of the Brownian motor effect in randomly flashing ratchets. *J. Math. Biol.* **2014**, *68*, 1677–1705. [CrossRef]

35. Frezzato, D. Dissipation, lag, and drift in driven fluctuating systems. *Phys. Rev. E* **2017**, *96*, 062113. [CrossRef] [PubMed]

36. Ethier, S.N.; Lee, J. The tilted flashing Brownian ratchet. *Fluct. Noise Lett.* **2019**, *18*, 1950005. [CrossRef]