IOP PUBLISHING NONLINEARITY

Nonlinearity **24** (2011) R19–R49 doi:10.1088/0951-7715/24/6/R01

INVITED ARTICLE

Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems—an analytical theory

Hong Qian

Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA

E-mail: hqian@u.washington.edu

Received 16 November 2010, in final form 22 March 2011
Published 21 April 2011
Online at stacks.iop.org/Non/24/R19

Recommended by J A Glazier

Abstract

The nonlinear dynamics of biochemical reactions in a small-sized system on the order of a cell are stochastic. Assuming spatial homogeneity, the populations of n molecular species follow a multi-dimensional birth-and-death process on $\mathbb{Z}^n$. We introduce the Delbrück–Gillespie process, a continuous-time Markov jump process, whose Kolmogorov forward equation has been known as the chemical master equation, and whose stochastic trajectories can be computed via the Gillespie algorithm. Using simple models, we illustrate that a system of nonlinear ordinary differential equations on $\mathbb{R}^n$ emerges in the infinite system size limit. For finite system size, transitions among multiple attractors of the nonlinear dynamical system are rare events with exponentially long transit times. There is a separation of time scales between the deterministic ODEs and the stochastic Markov jumps between attractors. No diffusion process can provide a global representation that is accurate on both short and long time scales for the nonlinear, stochastic population dynamics. On the short time scale and near deterministic stable fixed points, Ornstein–Uhlenbeck Gaussian processes give linear stochastic dynamics that exhibit time-irreversible circular motion for open, driven chemical systems. Extending this individual stochastic behaviour-based nonlinear population theory of molecular species to other biological systems is discussed.

Mathematics Subject Classification: 82C31, 37N25, 92C40

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Recent studies of biochemical reaction systems in a mesoscopic volume such as a cell have firmly established the chemical master equation (CME) as the basis of an analytical theory for cellular dynamics [1–7]. A system's volume V is a natural parameter in the theory: an ideal elementary chemical reaction $A + B \to C$ has a rate constant k with the dimension of $[\text{Time}]^{-1} \times [\text{Concentration}]^{-1}$. In the CME, the probability of this reaction occurring in an infinitesimal Δt is the dimensionless $k(n_A n_B / V)(\Delta t)$, when there are n_A and n_B molecules of types A and B, respectively. With increasing system size, V, the stochastic dynamics predicted by a CME has been mathematically shown to approach the deterministic solution of the kinetic differential equations based on the law of mass action for homogeneous chemical reactions [8]. The CME, therefore, 'is not an alternative to the deterministic kinetics, it is a more complete kinetic description which is capable of modelling reactions with and without fluctuations', for systems with small and large V [9].

In this review, I would like to take this new perspective a step further. A great many nonlinear systems of ordinary differential equations (ODEs) one studies describe dynamics of populations of one type or another. Examples include molecular species in biochemical reactions, cell and virus populations in immunology, human populations in demography and biological species in ecology. At the mechanistic level, all these dynamics are concerned with *birth* and *death* of individuals whose basic unit is 1. Therefore, every such dynamic model based on a nonlinear deterministic ODE system has a corresponding stochastic counterpart based on a birth-and-death process (BDP). If a nonlinear ODE system is defined on $\mathbb{R}^n$, the corresponding BDP is defined on $\mathbb{Z}^n$. First-order ODEs correspond to Markov jump processes with continuous time [10, 11].

Ever since the work of Einstein, Smoluchowski, Langevin, and Kramers, stochastic differential equations (SDEs), also known as diffusion processes by probabilists [12], have always been considered as the stochastic counterpart of ODEs [13, 14]. However, as anyone who has developed an SDE model for an applied problem knows, the choice for the coefficient $\boldsymbol{\Gamma}(\boldsymbol{x})$ in an SDE $\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{b}(\boldsymbol{x})\,\mathrm{d}t + \boldsymbol{\Gamma}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{W}(t)$ is almost always rather arbitrary. (The only exception is the guiding principle for fluctuating equilibrium dynamics based on the *fluctuation–dissipation theory* which we shall discuss later.) A BDP, however as we shall see, provides a rather complete stochastic description for the dynamics from mechanisms based on statistics of an individual's behaviour. There is no artificial separation of the deterministic $\boldsymbol{b}(\boldsymbol{x})$ and stochastic $\boldsymbol{\Gamma}(\boldsymbol{x})\boldsymbol{W}(t)$ as in a SDE. (See section 7.3. This is called 'intrinsic noise' in cellular biochemistry.) Even more important, as we shall discuss in section 4.4, is the 'diffusion theory's dilemma' that invalidates the diffusion-model approach to nonlinear stochastic *population* dynamics.

The BDP theory provides further insights into the theory of nonlinear population dynamics extensively studied since the 1970s. There is a fundamental concept that does not exist in the theory of deterministic nonlinear dynamics, the concept of 'rare events': something that occurs with a very small probability, but on an evolutionarily long time scale, it will occur with probability one! We have recently argued that [15] this emergent stochastic transition among different attractors, on the time scale beyond the infinity of the deterministic dynamics, is one of the origins of 'complexity' [5]. It is these dynamics that exhibit 'dynamic symmetry breaking' [16] and 'singular points' at which the dynamics are truly unpredictable [17], giving rise to complex dynamics with high information content [16].

It is safe to say that statistical inference is currently one of the key approaches to complex systems and their dynamics. Bioinformatics and statistical genomics are dominant applied mathematics in cellular molecular biology. The above nonlinear stochastic dynamic

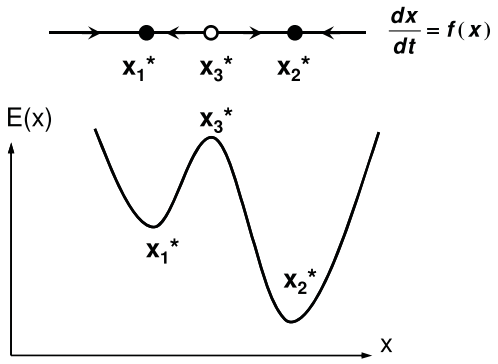

Figure 1. A pictorial introduction of the Kramers' barrier crossing problem in nonlinear, stochastic dynamics. Deterministic dynamics always go 'downhill' toward lower values of $E(x)$. Therefore, any dynamics with initial positive $x(0) < x_3^*$ will end at x_1^*, and with $x(0) > x_3^*$ will end at x_2^*. After reaching a stable fixed point, x_1^* or x_2^*, there will be no possibility of leaving. However, with stochastic elements in the dynamics, there are possibilities to go 'uphill.' With exceptional luck, continuous uphill movement leads to a transition between the two domains of attraction. Exceptional luck means the barrier crossings occur only on an extremely long time scale. x_3^* has been called the 'singular point' by James Clerk Maxwell since 'influences whose physical magnitude is too small to be taken account of by a finite being, may produce results of the greatest importance' [17].

perspective, however, clearly suggests that statistical approaches, while they can be powerful in representing data with statistical significance, cannot be useful in understanding the rare events. In fact, the very existence of multistabilities, i.e. alternative attractors, cannot be inferred from 'normal' statistical data. Mechanistic *deterministic* models can predict the existence of alternative attractors. Mechanistic *stochastic* models can further estimate the *lifetime* of an attractor. The actual time of the rare transition, however, is a random variable with exponential distribution, which is memoryless in defiance of causality.

Proving global asymptotic stability of a dynamical system, of course, has always been the ultimate goal of engineering. However, with increasing complexities, this becomes a less and less feasible task even in traditional engineering. On the other hand, one of the best understood 'rare events' is discrete chemical reactions in terms of Kramers' theory [18]. Recently in [19] we have shown that the nonlinear bifurcation theory of Thom–Zeeman's catastrophe, the phase transition theory from statistical mechanics, and Kramers' theory of barrier crossing (also known as decay of metastable states [20, 21]) are three different aspects (e.g. deterministic, steady state and kinetic) of a rare event. All these classical theories are called for in BDP dynamics.

Figure 1 shows the canonical pictorial introduction of the problem of 'barrier crossing' as a rare event. From a deterministic nonlinear dynamics standpoint, this system has three fixed points, two stable (x_1^* and x_2^* represented by the filled circles) and one unstable (x_3^* represented by the open circle). Barrier crossing requires movement against the deterministic force (shown by the arrows) which are low probability events. However, it is the cumulation of these unlikely events that leads to 'spectacular' or 'disastrous' phenomena in complex, stochastic nonlinear dynamical systems.

Our discussion of the nonlinear stochastic dynamics of biochemical reactions is organized in this paper as follows.

In section 2, using a simple example from mesoscopic chemical reaction systems, we introduce the 'bottom-up' approach to stochastic population dynamics based on mechanisms at the individual's level. The example illustrates how nonlinear, bistable behaviour emerges from such a dynamical model. The analysis of stochastic dynamics gives rise to the concept of multiple time scales.

Section 3 provides a systematic treatment of the Ornstein–Uhlenbeck Gaussian process as the linear stochastic process near a fixed point of a dynamical system. In a nutshell, the stability, i.e. hyperbolicity, of a fixed point is determined by the stationary probability distribution $f^{\rm st}(\boldsymbol{x})$, and the type of a fixed point, i.e. node *versus* focus, etc, is determined by the stationary, divergence-free circulation $\boldsymbol{j}^{\rm st}(\boldsymbol{x})$.

Section 4 presents the widely practiced 'top-down' approach based on SDEs and related diffusion processes. We suggest, however, that when approximating the large, but not infinite, population limit of a BDP with bistability, diffusion theory encounters a dilemma. It can provide a faithful representation for either the stationary behaviour or the fluctuating 'downhill' dynamics, but not both. We further illustrate the intimate relation of this problem to several other issues: the Keizer's paradox [22], Kurtz's convergence theorem with finite time [8], and van Kampen's conditional diffusion equation [23, 24].

Section 5 gives a brief discussion of two types of bistability in a mesoscopic chemical reaction system: that with a macroscopic, deterministic nonlinear counterpart, and that without. It is shown that their difference can be understood from the volume dependence of the transition rates between the two attractors.

In section 6, we show how insights from studying stochastic, nonlinear chemical reaction systems can be useful to the studies of other population dynamics. We try to establish some kinetic isomorphism between chemical dynamics and ecological dynamics.

Section 7 concludes the paper with some discussions and outlooks.

In the appendices, we have given some details of the mathematical results used in the main text. Much of this material is not found in the literature.

2. Nonlinear stochastic population dynamics: the individual-based approach

In this section, we present the theoretical development of stochastic models for nonlinear chemical reaction dynamics. The approach here is 'bottom-up' since we use an individual's stochastic behaviour as the starting point, considering one individual molecule at a time. As we shall see, this approach is in sharp contrast to the 'top-down' approach of section 4.

The approach we advance is general for any chemical and biochemical reaction system. However, we shall not present the theory in its most general form that often obscures the insights. Rather, we shall use a simple example to illustrate the theoretical approach. Let us consider the biochemical reaction system given by

$$A + 2X \underset{\alpha_2}{\overset{\alpha_1}{\rightleftharpoons}} 3X \qquad \text{and} \qquad X \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B. \tag{1}$$

This nonlinear chemical reaction system is known as the Schlögl model [25, 26]. The autocatalytic step in fact is widely observed in cellular biochemistry such as Src family kinase signalling, Rabaptin-5 mediated Rab5 GTPase activation in endocytosis, Xenopus oocyte maturation via a mitogen-activated protein kinase (MAPK) pathway, and self-regulating gene networks [4, 5, 27, 28].

2.1. Analysis of the deterministic dynamics

According to the law of mass action [29, 30], the nonlinear differential equation for $x(t)$, the concentration of the molecular species X in (1), is [19, 26, 31, 32]

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \beta_2 b - \beta_1 x + \alpha_1 a x^2 - \alpha_2 x^3, \tag{2}$$

where a and b represent the concentrations of chemical species A and B, which are assumed to be sustained at constant values. Any biochemical system in living organisms has to have at least one 'source' and one 'sink' species. This feature has been called an *open driven chemical system*. A system in contact with only a single material reservoir is called *grand canonical system* according to Gibbs [33, 34]. The latter necessarily approaches a 'dead' chemical equilibrium.

Non-dimensionalizing equation (2) with new variables and parameters

$$z = \frac{\beta_1 x}{\beta_2 b}, \qquad \tau = \beta_1 t, \qquad \sigma = \frac{\alpha_2 (\beta_2 b)^2}{\beta_1^3}, \qquad \gamma = \frac{a\alpha_1 \beta_1}{\alpha_2 \beta_2 b}, \tag{3}$$

we have

$$\frac{\mathrm{d}z}{\mathrm{d}\tau} = 1 - z + \sigma\left(\gamma z^2 - z^3\right) = f(z). \tag{4}$$

It is easy to show that for a wide range of parameter values, $f(z) = 0$ has three positive roots, corresponding to $\frac{\mathrm{d}z}{\mathrm{d}\tau} = f(z)$ having two stable fixed points and one unstable fixed point. See figure 2(a).

Equation (4) exhibits bistability when the parameter pair (σ, γ) is in the region bound by the curve in parametric form with

$$\sigma = \frac{z-2}{z^3} \qquad \text{and} \qquad \gamma = \frac{(2z-3)z}{z-2}. \tag{5}$$

Figure 2(b) shows that the region in which bistability exists has a cusp at $\sigma = \frac{1}{27}, \gamma = 9$. The nonlinear dynamics exhibits the canonical saddle-node bifurcation and Thom–Zeeman's catastrophe [29].

2.2. The CME and stochastic models

There are four elementary reactions in system (1). In an aqueous solution, the occurrence of a reaction is a random event with exponentially distributed waiting time. The stochastic dynamics of the number of molecule X, $n(t)$, therefore, is a one-dimensional BDP. As a continuous-time Markov process, the BDP has its Kolmogorov forward equation, the CME, in the form [7, 26]

$$\frac{\mathrm{d}}{\mathrm{d}t} p(n,t) = p(n-1,t)\mu_{n-1} - p(n,t)\left(\mu_n + \lambda_n\right) + p(n+1,t)\lambda_{n+1}, \tag{6}$$

in which

$$\mu_n(V) = \frac{\alpha_1 a n(n-1)}{V} + \beta_2 b V \qquad \text{and} \qquad \lambda_n(V) = \frac{\alpha_2 n(n-1)(n-2)}{V^2} + \beta_1 n, \tag{7}$$

are the birth and death rates of the process. Note both are functions of the system's size V.

For this simple system, it is not difficult to show heuristically that in the limit of $V \to \infty$ and $n \to \infty$, but $n/V \to x$, the stochastic dynamics following the BDP becomes the solution to the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu(x) - \lambda(x), \tag{8}$$

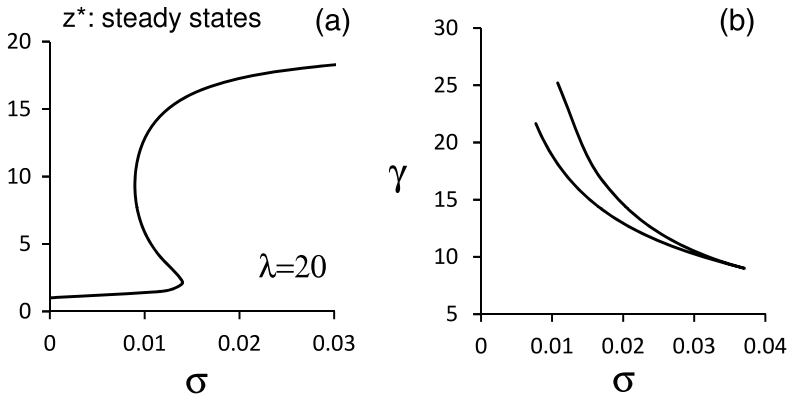

Figure 2. Nonlinear chemical reaction system (1) can exhibit bistability. (*a*) Fixed points of the ODE in equation (4), z^*, are obtained from $f(z) = 0$, as a function of σ with $\lambda = 20$: $\sigma = (z^* - 1)/(\gamma - z^*)/z^{*2}$. (*b*) The region of parameter space (σ, λ) in which the ODE is bistable has a cusp at $\sigma = \frac{1}{27}, \gamma = 9$. In the statistical physics theory of phase transition, the cusp is also known as a critical point [19].

with

$$\mu(x) = \lim_{n\to\infty} \frac{\mu_n(n/x)}{n/x} = \alpha_1 a x^2 + \beta_2 b \qquad \text{and} \qquad \lambda(x) = \lim_{n\to\infty} \frac{\lambda_n(n/x)}{n/x} = \alpha_2 x^3 + \beta_1 x.$$

Equation (8) is exactly equation (2). See [8] for a rigorous, general proof of the important limit theorem that connects the BDP following the CME and the ODE according to the law of mass action. Also see [35] for an alternative proof.

Because the problem is one-dimensional, the stationary probability distribution to equation (6) is readily obtained:

$$p^{\mathrm{st}}(n; V) = p^{\mathrm{st}}(0) \prod_{k=1}^{n} \frac{\mu_{k-1}(V)}{\lambda_k(V)}, \tag{9}$$

where $p^{\mathrm{st}}(0)$ is determined by normalization of the probability distribution.

The distribution (9) has several important properties:

(i) Its extrema are located at n^* where $\mu_{n^*-1} = \lambda_{n^*}$. An extreme corresponds to, therefore, a fixed point of the ODE, where $\mu(x) = \lambda(x)$.

(ii) One can obtain an asymptotic expansion when $V, n \to \infty$ and $n/V \to x$:

$$p^{\mathrm{st}}(n; V) \longrightarrow f^{\mathrm{st}}(x; V) = \mathrm{e}^{-V\phi(x)}, \qquad \text{where} \quad \phi(x) = \int_0^x \ln\left[\frac{\lambda(v)}{\mu(v)}\right] \mathrm{d}v. \tag{10}$$

Note that when $V \to \infty$, the $f^{\mathrm{st}}(x, V)$ converges to the global minimum of $\phi(x)$.

(iii) The function $\phi(x)$ is a Lyapunov function for the ODE (8):

$$\frac{\mathrm{d}\phi(x(t))}{\mathrm{d}t} = \frac{\mathrm{d}\phi(x)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} = \{\mu(x) - \lambda(x)\} \ln\left(\frac{\lambda(x)}{\mu(x)}\right) \leqslant 0. \tag{11}$$

The rhs is equal to zero iff $\mu(x) = \lambda(x)$. While this result might not be too surprising in the case of one-dimensional dynamics, differentiable $\phi(x)$ can be obtained for many higher dimensional chemical reaction systems without detailed balance. $\phi(x)$, known as the large deviation rate function in the theory of probability, is a Lyapunov function for the ODEs from the mass-action law! [36, 37]

For the non-dimensionalized Schlögl model, one has

$$\phi(z) = \int_0^z \ln \frac{(\sigma u^2 + 1)u}{\sigma\gamma u^2 + 1}\,\mathrm{d}u = z \ln \frac{(\sigma z^2 + 1)z}{\sigma\gamma z^2 + 1} + \frac{2}{\sqrt{\sigma}} \arctan\left(\sqrt{\sigma} z\right) - \frac{2}{\sqrt{\sigma\gamma}} \arctan\left(\sqrt{\sigma\gamma} z\right) - z. \tag{12}$$

(iv) The two basins of attraction should be understood as two states of the chemical reaction system (1). They are the 'emergent properties' of the stochastic nonlinear population dynamics. For a given system, one could be in one of the states for a very long time. Still, the elementary operations at the individual level dictate the existence of the other state. This is a perspective that one cannot gain from pure statistical inference.

While the ODE predicts the existence of bistability, it cannot provide an estimation for the stability of the states. The stability can be obtained from the BDP by computing the mean first passage times (MFPTs). For the simple one-dimensional system, let n_1^* and n_2^* be the two peak positions and n_3^* the trough position of $p^{\mathrm{st}}(n)$, then [24]

$$T_{n_1^* \to n_2^*} = \sum_{n=0}^{n_1^*} \sum_{m=n_1^*+1}^{n} \frac{p^{\mathrm{st}}(n)}{\lambda_m p^{\mathrm{st}}(m)} + \sum_{n=n_1^*+1}^{n_2^*-1} \sum_{m=n+1}^{n_2^*} \frac{p^{\mathrm{st}}(n)}{\lambda_m p^{\mathrm{st}}(m)}. \tag{13}$$

In the case of $V \to \infty$, equation (13) becomes (appendix B.2)

$$T_{x_1^* \to x_2^*} \approx \frac{2\pi \mathrm{e}^{V(\phi(x_3^*) - \phi(x_1^*))}}{\lambda(x_3^*)\sqrt{-\phi''(x_1^*)\phi''(x_3^*)}}. \tag{14}$$

This time grows exponentially with system size V. Transitions between the two domains of attraction (DoA) are rare events.

(v) A comprehensive theory emerges from analysing this simple model. There are three different time scales in the nonlinear, stochastic population dynamics: (1) the time scale of individual reactions, the α's and β's, which we call the molecular signalling time scale t_{ms} in the context of cellular biochemistry, (2) the time scale of nonlinear network dynamics t_{nd} and (3) the time scale on which the transitions between the DoA occurs, i.e. $T_{x_1^* \to x_2^*}$ and $T_{x_2^* \to x_1^*}$, which we call cellular evolution t_{ce}. In nonlinear deterministic dynamics, a long time means $t \gg t_{\mathrm{nd}}$ but it is still $\ll t_{\mathrm{ce}}$. On this time scale, a system settles into one attractor depending on the initial state. However, on the time scale $t \gg t_{\mathrm{ce}}$, the system will establish a probability distribution between the two DoA.

(vi) There is a great separation of time scales between t_{nd} and t_{ce} for a system with large populations. In this case, on the time scale $\gg t_{\mathrm{nd}}$ but $\ll t_{\mathrm{ce}}$, the system's behaviour is captured by a bifurcation diagram, such as that in figure 2(*a*). However, on the time scale $\gg t_{\mathrm{ce}}$, the stationary probability distribution given in equation (10) shows that the global minimum of $\phi(x)$ will have probability of almost 1, while other minimum will have only probability $\propto \mathrm{e}^{-cV}$ where c is a positive constant. Therefore, for large systems with $t \gg t_{\mathrm{ce}}$, the bifurcation diagram in figure 2(*a*) has to be modified by the Maxwell construction [19, 32]. This is the subject of phase transition theory in statistical mechanics.

3. Ornstein–Uhlenbeck processes: linear analysis of stochastic dynamics

One very useful method for analysing nonlinear dynamical systems is the local, linear analysis of fixed points. For nonlinear stochastic dynamics, the corresponding linear analysis is the theory of Gaussian–Markov processes, also known as Ornstein–Uhlenbeck processes. The

subject has been extensively studied by physicists such as Einstein, Chandrasekhar, Ornstein–Uhlenbeck–Wang, Onsager–Machlup, Lax, Keizer, Fox, and many others [13, 38, 40, 41]. Recent work has paid particular attention to the issue of time irreversibility and the breakdown of detailed balance in Gaussian processes [42, 43].

We consider linear SDE in the vector form

$$\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{B}\boldsymbol{x}(t)\,\mathrm{d}t + \boldsymbol{\Gamma}\,\mathrm{d}\boldsymbol{W}(t), \tag{15}$$

in which $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are $n \times n$ constant matrices and $\boldsymbol{x}$ and $\boldsymbol{W}$ are n-dimensional column vectors. $\boldsymbol{W}(t)$ contains n independent standard Brownian motions.

SDE (15) can be analysed by using several different methods, including the direct method

$$\boldsymbol{x}(t) = \mathrm{e}^{Bt}\left(\boldsymbol{x}(0) + \int_0^t \mathrm{e}^{-Bs}\boldsymbol{\Gamma}\,\mathrm{d}\boldsymbol{W}(s)\right), \tag{16}$$

and the Fourier transform method

$$-\mathrm{i}\omega\widetilde{\boldsymbol{x}}(\omega) = \boldsymbol{B}\widetilde{\boldsymbol{x}}(\omega) + \boldsymbol{\Gamma}\widetilde{\boldsymbol{\xi}}(\omega), \tag{17}$$

in which

$$\boldsymbol{x}(t) = \int_{-\infty}^{\infty}\widetilde{\boldsymbol{x}}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega \qquad \text{and} \qquad \frac{\mathrm{d}\boldsymbol{W}(t)}{\mathrm{d}t} = \int_{-\infty}^{\infty}\widetilde{\boldsymbol{\xi}}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega. \tag{18}$$

The Fourier transform of independent white noise satisfies

$$\left\langle\widetilde{\boldsymbol{\xi}}^{*}(\omega)\widetilde{\boldsymbol{\xi}}^{\mathrm{T}}(\omega)\right\rangle = \boldsymbol{I}, \tag{19}$$

the identity matrix, where $\langle\cdots\rangle$ is the ensemble average.

The stationary $\boldsymbol{x}(t)$ has a multivariate Gaussian distribution

$$f^{\mathrm{st}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\det{}^{\frac{1}{2}}(\boldsymbol{\Xi})}\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\boldsymbol{x}\right) \tag{20}$$

in which the symmetric matrix $\boldsymbol{\Xi}$ is the covariant matrix satisfying the Lyapunov matrix equation [44]

$$\boldsymbol{B}\boldsymbol{\Xi} + \boldsymbol{\Xi}\boldsymbol{B}^{\mathrm{T}} = -\boldsymbol{A} \quad \text{with} \quad \boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}. \tag{21}$$

3.1. Power spectrum of a stationary OU process with circulation

A stationary $\boldsymbol{x}(t)$ has not only a distribution, given in equation (20), but also temporal correlation. One way to characterize the temporal dynamics is by the power spectrum. From equation (17) we have [45]

$$\widetilde{\boldsymbol{x}}^{*}(\omega) = [\mathrm{i}\omega\boldsymbol{I} - \boldsymbol{B}]^{-1}\,\boldsymbol{\Gamma}\widetilde{\boldsymbol{\xi}}^{*}(\omega) \qquad \text{and} \qquad \widetilde{\boldsymbol{x}}^{\mathrm{T}}(\omega) = -\widetilde{\boldsymbol{\xi}}^{\mathrm{T}}(\omega)\boldsymbol{\Gamma}^{\mathrm{T}}\,[\mathrm{i}\omega\boldsymbol{I} + \boldsymbol{B}]^{-\mathrm{T}}\,. \tag{22}$$

Thus, we have the power spectra for a multi-dimensional, stationary OU process

$$\boldsymbol{\Theta}(\omega) \triangleq \left\langle\widetilde{\boldsymbol{\xi}}^{*}(\omega)\widetilde{\boldsymbol{\xi}}^{\mathrm{T}}(\omega)\right\rangle = -\,[\mathrm{i}\omega\boldsymbol{I} - \boldsymbol{B}]^{-1}\,\boldsymbol{A}\,[\mathrm{i}\omega\boldsymbol{I} + \boldsymbol{B}]^{-\mathrm{T}}\,. \tag{23}$$

In other words,

$$\boldsymbol{\Theta}^{-1}(\omega) = \underbrace{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B} + \omega^{2}\boldsymbol{A}^{-1}}_{\text{symmetric}} + \mathrm{i}\omega\underbrace{\left(\boldsymbol{A}^{-1}\boldsymbol{B} - \boldsymbol{B}^{T}\boldsymbol{A}^{-1}\right)}_{\text{anti-symmetric}}. \tag{24}$$

In [42], it was shown that a stationary OU process is time-reversible iff $\boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$, and furthermore $\boldsymbol{\Xi} = \frac{1}{2}\boldsymbol{B}^{-1}\boldsymbol{A}$. Therefore, a time-reversible OU process has

$$\boldsymbol{\Theta}(\omega) = \left[\boldsymbol{B}^{2} + \omega^{2}\boldsymbol{I}\right]^{-1}\boldsymbol{A}. \tag{25}$$

We see that, in this case, the power spectra $\boldsymbol{\Theta}(\omega)$ is a real symmetric matrix and has a Lorentzian form with its peak located at $\omega = 0$ [46, 47].

When $\boldsymbol{A}^{-1}\boldsymbol{B} \neq \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$, the anti-symmetric term in equation (24) indicates that the stationary $\boldsymbol{x}(t)$ has a circular motion [48]. This circular motion can be best illustrated by considering the stationary Fokker–Planck equation (FPE) for the SDE (15):

$$\frac{\partial f(\boldsymbol{x})}{\partial t} = \nabla \cdot \left(\frac{1}{2}\boldsymbol{A}\nabla f(\boldsymbol{x}) - \boldsymbol{B}\boldsymbol{x} f(\boldsymbol{x}) \right) = 0. \tag{26}$$

Integrating equation (26) once,

$$\tfrac{1}{2}\boldsymbol{A}\nabla f^{\mathrm{st}}(\boldsymbol{x}) - \boldsymbol{B}\boldsymbol{x} f^{\mathrm{st}}(\boldsymbol{x}) = -\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x}) \qquad \text{with} \qquad \nabla \cdot \boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x}) = 0. \tag{27}$$

This can be rewritten as

$$\tfrac{1}{2}\nabla \ln f^{\mathrm{st}}(\boldsymbol{x}) - \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x} = -\boldsymbol{A}^{-1}\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x}) f^{-1}(\boldsymbol{x}). \tag{28}$$

Symmetry, $\boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$, means $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}$ is a gradient force. Then by the uniqueness of the solution to the linear elliptic equation, $\boldsymbol{j}^{\mathrm{st}} = 0$ and $\ln f^{\mathrm{st}}(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x} + \mathrm{const}$. That is $2\boldsymbol{\Xi} = \boldsymbol{B}^{-1}\boldsymbol{A}$.

If $\boldsymbol{A}^{-1}\boldsymbol{B} \neq \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$, then $\boldsymbol{j}^{\mathrm{st}} \neq 0$. In fact,

$$\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x}) = \left(\boldsymbol{B} + \tfrac{1}{2}\boldsymbol{A}\boldsymbol{\Xi}^{-1}\right)\boldsymbol{x} f^{\mathrm{st}}(\boldsymbol{x}). \tag{29}$$

The divergence-free $\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x})$ represents certain circular motion, which occurs only when a vector field is non-gradient.

It is easy to verify that the $\boldsymbol{j}^{\mathrm{st}}(x)$ and $\nabla \ln f^{\mathrm{st}}(\boldsymbol{x})$ are orthogonal to each other [43]:

$$\begin{aligned} \nabla \ln f^{\mathrm{st}}(\boldsymbol{x}) \cdot \boldsymbol{j}^{\mathrm{st}}(x) &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\left(\boldsymbol{B} + \tfrac{1}{2}\boldsymbol{A}\boldsymbol{\Xi}^{-1}\right)\boldsymbol{x} f^{\mathrm{st}}(\boldsymbol{x}) \\ &= \tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{\Xi}^{-1}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\right)\boldsymbol{x} f^{\mathrm{st}}(\boldsymbol{x}) = 0. \end{aligned} \tag{30}$$

3.2. The Green–Kubo–Zwanzig relation

Because the SDE in (15) is linear, it is also easy to obtain

$$E^{x_0}\left[\boldsymbol{x}(t)|\boldsymbol{x}(0) = \boldsymbol{x}_0\right] = \mathrm{e}^{\boldsymbol{B}t}\boldsymbol{x}_0, \tag{31}$$

where $E^{x_0}[\cdots|\boldsymbol{x}(0) = \boldsymbol{x}_0]$ is the conditional ensemble average with given initial $\boldsymbol{x}(0) = \boldsymbol{x}_0$. Then the stationary time correlation function matrix is, for $t \geqslant 0$,

$$E^{\mathrm{st}}\left[E^{x_0}\left[\boldsymbol{x}(t)|\boldsymbol{x}(0) = \boldsymbol{x}\right]\boldsymbol{x}^{\mathrm{T}}\right] = E^{\mathrm{st}}\left[\mathrm{e}^{\boldsymbol{B}t}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right] = \mathrm{e}^{\boldsymbol{B}t}\boldsymbol{\Xi}. \tag{32}$$

Therefore,

$$\boldsymbol{G}_{xx}(t) = \langle \boldsymbol{x}(\tau)\boldsymbol{x}^{\mathrm{T}}(\tau + t)\rangle = \begin{cases} \boldsymbol{\Xi}\mathrm{e}^{\boldsymbol{B}^{\mathrm{T}}t} & t \geqslant 0, \\ \mathrm{e}^{-\boldsymbol{B}t}\boldsymbol{\Xi} & t \leqslant 0. \end{cases} \tag{33}$$

We note that

$$\boldsymbol{G}_{xx}(-t) = \boldsymbol{G}_{xx}^{\mathrm{T}}(t). \tag{34}$$

For time-reversible processes, $\boldsymbol{B}\boldsymbol{\Xi} = \boldsymbol{\Xi}\boldsymbol{B}^{\mathrm{T}}$. Hence, $\boldsymbol{G}_{xx}(-t) = \boldsymbol{G}_{xx}^{\mathrm{T}}(t) = \boldsymbol{G}_{xx}(t)$ [49].

The Green–Kubo–Zwanzig formula concerns the mathematical relation between the transport coefficients and the integrals of the time-correlation function of the velocity [46, 50]. In the case of multi-dimensional Gaussian processes, the velocity is simply $\boldsymbol{v}(t) = \boldsymbol{B}\boldsymbol{x}(t)$ and $\langle \boldsymbol{v}(\tau)\boldsymbol{v}^{\mathrm{T}}(\tau+t)\rangle = \boldsymbol{B}\boldsymbol{G}_{xx}(t)\boldsymbol{B}^{\mathrm{T}}$, where the correlation function matrix $\boldsymbol{G}_{xx}(t) = \langle \boldsymbol{x}(\tau)\boldsymbol{x}(\tau+t)\rangle$ is the Fourier transform of $\boldsymbol{\Theta}(\omega)$. Assuming all the eigenvalues of $\boldsymbol{B}$ have negative real parts,

the integral of the time-correlation function is the spectrum value of the process at $\omega = 0$. Therefore, from equation (23) we have

$$\int_{-\infty}^{\infty} \langle \boldsymbol{v}(\tau)\boldsymbol{v}^{\mathrm{T}}(\tau+t)\rangle \,\mathrm{d}t = \boldsymbol{B}\left(\int_{-\infty}^{\infty} \boldsymbol{G}_{xx}(t)\,\mathrm{d}t\right)\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{B}\boldsymbol{\Theta}(0)\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{A}. \quad (35)$$

Note that in this version, the Green–Kubo–Zwanzig formula requires no time-reversibility. It is a consequence of a linear stochastic dynamical system.

However, if one considers only $t \geqslant 0$, then

$$\int_{0}^{\infty} \langle \boldsymbol{v}(\tau)\boldsymbol{v}^{\mathrm{T}}(\tau+t)\rangle \,\mathrm{d}t = -\boldsymbol{B}\boldsymbol{\Xi}\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{B}^{\mathrm{T}} = -\boldsymbol{B}\boldsymbol{\Xi}. \quad (36)$$

In this case, the rhs is $\frac{1}{2}\boldsymbol{A}$ if and only if $\boldsymbol{B}^{-1}\boldsymbol{A}$ is symmetric, i.e. the Gaussian process is time reversible.

3.3. Gaussian processes in a plane

In nonlinear dynamical systems, local linear stability analysis near a fixed point in a plane has provided great insights into the nature of fixed points. In this section, we shall carry out a similar analysis for Gaussian OU processes in a plane.

We consider the two matrices

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \quad (37)$$

where $\boldsymbol{A}$ is positive definite. Solving the Lyapunov matrix equation (21) we have the covariance matrix

$$\boldsymbol{\Xi} = -\frac{1}{2(b_{11}+b_{22})\det(\boldsymbol{B})} \times \begin{pmatrix} (b_{11}b_{22} - b_{12}b_{21} + b_{22}^2)a_{11} & -b_{21}b_{22}a_{11} + 2b_{11}b_{22}a_{12} \\ -2b_{12}b_{22}a_{12} + b_{12}^2 a_{22} & -b_{11}b_{12}a_{22} \\ -b_{21}b_{22}a_{11} + 2b_{11}b_{22}a_{12} & b_{21}^2 a_{11} - 2b_{11}b_{21}a_{12} \\ -b_{11}b_{12}a_{22} & +(b_{11}^2 + b_{11}b_{22} - b_{12}b_{21})a_{22} \end{pmatrix}. \quad (38)$$

Thus,

$$\boldsymbol{\Xi}^{-1} = -\frac{2}{(b_{11}+b_{22})(\det(\boldsymbol{A})+\delta^2)} \times \begin{pmatrix} b_{21}^2 a_{11} - 2b_{11}b_{21}a_{12} & b_{21}b_{22}a_{11} - 2b_{11}b_{22}a_{12} \\ +(b_{11}^2 + b_{11}b_{22} - b_{12}b_{21})a_{22} & +b_{11}b_{12}a_{22} \\ b_{21}b_{22}a_{11} - 2b_{11}b_{22}a_{12} & (b_{11}b_{22} - b_{12}b_{21} + b_{22}^2)a_{11} \\ +b_{11}b_{12}a_{22} & -2b_{12}b_{22}a_{12} + b_{12}^2 a_{22} \end{pmatrix}, \quad (39)$$

in which

$$\delta = \frac{b_{11}a_{12} + b_{12}a_{22} - b_{21}a_{11} - b_{22}a_{12}}{b_{11}+b_{22}}.$$

Note that $-(b_{11}+b_{22}) > 0$ for a stable fixed point.

What is the relationship between the $\boldsymbol{B}$, the linear stability matrix, and the $\boldsymbol{\Xi}^{-1}$, the inverse of covariance matrix, that constitutes the quadratic form $\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\boldsymbol{x}$? We make the following observations:

(a) The determinants of $\boldsymbol{B}$ and $\boldsymbol{\Xi}^{-1}$ have same sign:

$$\det\left(\boldsymbol{\Xi}^{-1}\right) = \frac{4\det(\boldsymbol{B})}{\det(\boldsymbol{A})+\delta^2}. \quad (40)$$

(b) The hyperbolicity of the fixed point of $\dot{\boldsymbol{x}} = \boldsymbol{B}\boldsymbol{x}$ is in agreement with the quadratic function $\phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Xi}\boldsymbol{x} = -\ln f^{\mathrm{st}}(\boldsymbol{x}) + \text{const}$. This can be shown from

$$\dot{\boldsymbol{x}}^{\mathrm{T}} \cdot \nabla\phi(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\boldsymbol{x} = -\tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Xi}^{-1}\boldsymbol{A}\boldsymbol{\Xi}^{-1}\boldsymbol{x} \leqslant 0. \tag{41}$$

However, the nature of a stable fixed point, i.e. being a node or a focus, is determined by the sign of the *discriminant*

$$\mathrm{Tr}^2(\boldsymbol{B}) - 4\det(\boldsymbol{B}). \tag{42}$$

This information is not contained in the symmetric $\boldsymbol{\Xi}^{-1}$ whose discriminant is always greater than zero. To see whether the eigenvalues of $\boldsymbol{B}$ are complex, we turn to the divergence-free $\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x})$.

(c) $\boldsymbol{j}^{\mathrm{st}}(\boldsymbol{x}) = 0 \Longleftrightarrow \boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$.

(d) Since $\boldsymbol{A}$ is positive definite, we have a real, symmetric matrix $\boldsymbol{A}^{\frac{1}{2}}$. If $\boldsymbol{B}$ has a pair of complex eigenvalues, then $\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{\frac{1}{2}}$ has a pair of complex eigenvalues, and therefore $\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{\frac{1}{2}} \neq \left(\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{A}^{\frac{1}{2}}\right)^{\mathrm{T}} = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-\frac{1}{2}}$. That is, $\boldsymbol{A}^{-1}\boldsymbol{B} \neq \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$. If the fixed point is a focus, then the Gaussian process is time-irreversible for any $\boldsymbol{A}$.

(e) If $\boldsymbol{B}$ has all real eigenvalues and is diagonalizable, then $\exists\boldsymbol{Q}$ such that $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}$ is diagonal: $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}$. One can choose $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$ and have $\boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}$. If the fixed point is a node, then $\exists\boldsymbol{A}$ such that the Gaussian process is time-reversible.

(f) For an irreversible stationary process, its power spectrum can exhibit a peak at $\omega > 0$, indicating inherent frequency [51]. Spectral peaking, however, is only a sufficient condition for irreversibility, but a not necessary condition. For planar Gaussian processes, we have from equation (23)

$$\begin{aligned}\boldsymbol{\Theta}(\omega) = {} & \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-\mathrm{T}} + \mathrm{i}\omega\boldsymbol{B}^{-1}\left(\boldsymbol{B}^{-1}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B}^{-\mathrm{T}}\right)\boldsymbol{B}^{-\mathrm{T}} \\ & - \omega^2\left(\boldsymbol{B}^{-3}\boldsymbol{A}\boldsymbol{B}^{-\mathrm{T}} - \boldsymbol{B}^{-2}\boldsymbol{A}\boldsymbol{B}^{-2\mathrm{T}} + \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-3\mathrm{T}}\right) + O\left(\omega^3\right).\end{aligned} \tag{43}$$

Therefore, a condition for $\Theta_{11}(\omega)$ having an off-zero peak is its curvature at $\omega = 0$ being positive

$$\begin{aligned}\frac{\mathrm{d}^2\Theta_{11}(0)}{\mathrm{d}\omega^2} = {} & \left(\left(b_{12}b_{21} - 2b_{22}^2 - 2b_{11}b_{22}\right)b_{12}b_{21} - b_{22}^4\right)a_{11} \\ & + \left(b_{11}^2 + b_{22}^2 + 2b_{21}b_{21}\right)\left(2b_{12}b_{21}a_{12} - b_{12}^2a_{22}\right) \geqslant 0.\end{aligned} \tag{44}$$

The rhs can be rewritten as

$$\begin{aligned}& -\underbrace{(b_{21},\ -b_{12})\,\boldsymbol{A}\begin{pmatrix} b_{21} \\ -b_{12}\end{pmatrix}}_{\text{positive}}\left(b_{11}^2 + b_{22}^2 + 2b_{12}b_{21}\right) \\ & + a_{11}\left(b_{11}b_{22} - b_{12}b_{21}\right)^2 + a_{11}\left(b_{21}^2 - b_{22}^2\right)\left(b_{11}^2 + b_{22}^2 + 2b_{12}b_{21}\right).\end{aligned} \tag{45}$$

We see that the second term in (45) is positive since $a_{11} > 0$. If b_{12} and b_{21} have opposite signs, and $\left(b_{11}^2 + b_{22}^2\right) + b_{12}b_{21} < 0$, then the first term in (45) also becomes positive. But under this condition,

$$\begin{aligned}\mathrm{Tr}^2(\boldsymbol{B}) - 4\det(\boldsymbol{B}) < {} & -2b_{11}b_{22} + 3b_{12}b_{21} \leqslant 2|b_{11}b_{22}| + 3b_{12}b_{21} \\ \leqslant {} & \left(b_{11}^2 + b_{22}^2\right) + 3b_{12}b_{21} < 0.\end{aligned}$$

We shall show next, however, that it is actually possible to find an $\boldsymbol{A}$ such that for a $\boldsymbol{B}$ with a node, the corresponding OU stationary process has a power spectrum with its peak at $\omega > 0$.

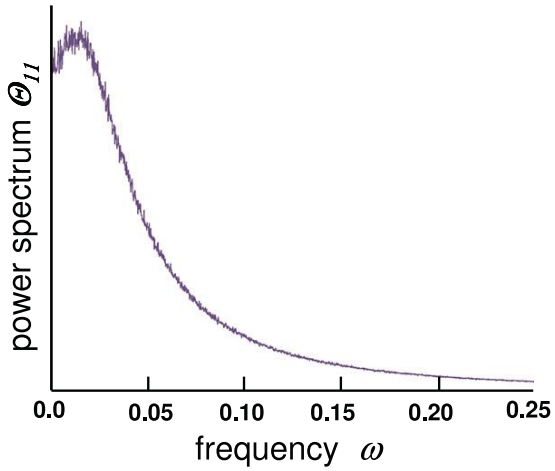

Figure 3. A planar linear dynamical system with a stable node (eigenvalues -2 and -3), when coupled to a white noise, becomes a time-irreversible OU Gaussian process exhibiting strong rotational motion. The power spectrum $\Theta_{11}(\omega)$, adopted from [45], shows an off-zero peak. (Figure provided by Dr Jia-zeng Wang.)

3.4. Noise-induced strong circular motion in a plane: an example

A Gaussian process is time irreversible iff $\boldsymbol{j}^{\rm st}(\boldsymbol{x}) \neq 0$. A non-zero $\boldsymbol{j}^{\rm st}$ indicates a certain kind of circulation in the dynamics of $\boldsymbol{x}(t)$. Power spectral peaking is only a sufficient but not necessary condition for the circulation. Only when the circular motion is sufficiently 'strong' will its power spectrum exhibit an off-zero peak [51].

The presence of noise can induce strong circulation in an ODE system with only a node and no hint of any circular motion in the deterministic dynamics. We borrow the example given in [45], considering

$$\boldsymbol{B} = \begin{pmatrix} -4 & 1 \\ -2 & -1 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{46}$$

The two eigenvalues of $\boldsymbol{B}$, -2 and -3, are both real. However, the curvature of $\Theta_{11}(\omega)$ at $\omega = 0$ according to equation (44) is 10. Figure 3 shows the power spectrum.

Noise-induced circulations, or oscillations, have been extensively studied in the past in connection with the phenomena called *stochastic resonance* and *coherence resonance*. See [51–53] for studies from the perspective of irreversible stationary stochastic processes. The example in (46) and figure 3 is perhaps the most elementary version of this interesting phenomenon.

4. Diffusion theory of nonlinear population dynamics with fluctuations

The SDE,

$$\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{b}(\boldsymbol{x})\,\mathrm{d}t + \boldsymbol{\Gamma}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{W}(t), \tag{47}$$

and the diffusion process it defines, is widely used to represent stochastic dynamics of natural and engineering systems [10, 13, 14, 24]. This is a much researched mathematical subject in both pure and applied mathematics.

Dynamics with fluctuations have been extensively studied in statistical physics, ever since the work of Einstein, Smoluchowski and Langevin, respectively, in 1905, 1906 and 1908. But it was mathematician K Itô who finally unified the mathematics of Einstein and Smoluchowski in terms of partial differential equations, and Langevin's approach in terms of SDEs (47). The description of the stochastic dynamics by diffusion processes is universal for Markov processes with continuous paths [14].

Still, one should note that the diffusion theory was conceived, in physics, as a 'correction' to deterministic dynamics. It is a phenomenological approach to stochastic fluctuations with a clear 'top-down' characteristic. When modelling with a diffusion process, it is almost always the case that one is uncertain about how to choose the $\mathbf{\Gamma}$. It is worth noting that when modelling an equilibrium process in physics and chemistry, this problem was solved by the so called *fluctuation–dissipation theory* [45]. The $\mathbf{\Gamma}$ is not determined from the physical mechanism of the problem, but rather from requiring the stationary process to agree with known physics—Boltzmann's law, Onsager's regression hypothesis and time reversibility.

With such a rich history, it is natural for one to be interested in representing the dynamics of large, but not infinite, populations with diffusion processes [54, 55]. In particular, one may ask if the diffusion process description, which has enjoyed great successes in physics, can be the appropriate model for the CME in macroscopic volume with fluctuations? The answer turns out to be a surprising 'no' [56, 57]. The diffusion processes can fail to provide a global approximation for the nonlinear stochastic dynamics of a bistable population with birth and death [22, 26, 56, 57]. The crux of the matter turns out to rest precisely with the rare events. A rare event occurs with exponentially small probability e^{-cV} and exponentially long time e^{+cV}, where $c > 0$ and V is the system's size. The diffusive stochastic processes with continuous trajectories are not accurate enough global representations [58].

As a mathematical result, this has been known for a long time. Kurtz's 1971 limit theorem is only valid for finite time due to precisely this problem [8]. van Kampen has repeatedly emphasized that a diffusion approximation can only be obtained for master equations with *small individual jumps* [23, 24]. He actually developed a sophisticated treatment of diffusion approximations for the master equation, order-by-order, called system-size expansion [24, chapter 10]. This theory provides a satisfying approximation for the stochastic relaxation in the limit of large V. It is shown that the only mathematically valid diffusion process one can derive from a CME is a Gaussian process *conditioned* on a given deterministic solution to the corresponding ODE (see appendix A). Both excluded the rare jumps between multiple nonlinear attractors. This approach, thus, does not address how to obtain a stationary distribution with multistability.

In the present review, we shall revisit this problem from a different perspective. From Kurtz's theorem and van Kampen's system-size expansion, we know that in the limit of large system size, one can obtain a diffusion approximation near a fixed point, as we have done in section 3. This approximation, however, underestimates the time for barrier crossing (see equation (63)). On the other hand, it is also possible to obtain a different diffusion approximation which gives the correct, global stationary distribution. But this one misrepresents the downhill dynamics. We call this 'diffusion theory's dilemma' [58].

4.1. A simple example

We again use the example of a one-dimensional BDP in equation (6), which resembles a difference scheme of a diffusion equation of Fokker–Planck type:

$$\frac{\partial f(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(A(x)f(x,t)\right) - \frac{\partial}{\partial x}\left(b(x)f(x,t)\right). \tag{48}$$

If we identify $x = n/V$, and $\mathrm{d}x = 1/V$, then the Kramers–Moyal expansion [10, 24, 40], truncated at the second-order, yields

$$A(x) = \frac{\mu(x) + \lambda(x)}{V} \qquad \text{and} \qquad b(x) = \mu(x) - \lambda(x). \tag{49}$$

Since the $A(x)$ term is on the order of $\frac{1}{V}$, equation (48) can also be written as

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\mu(x) + \lambda(x)}{2V}\frac{\partial f}{\partial x} - (\mu(x) - \lambda(x))f(x,t)\right), \tag{50}$$

with the divergence form for the diffusion. The difference is in the $b(x)$ term on the order of $O(1/V)$, which is negligible.

The stationary distribution to equation (50) is readily obtained as

$$\widetilde{f}^{\mathrm{st}}(x) = \mathrm{e}^{-V\widetilde{\phi}(x)}, \qquad \text{where} \qquad \widetilde{\phi}(x) = 2\int_0^x \frac{\lambda(v) - \mu(v)}{\lambda(v) + \mu(v)}\,\mathrm{d}v. \tag{51}$$

Now comparing the $\widetilde{\phi}(x)$ with the $\phi(x)$ in equation (10), we see that they are not identical. However, both have the same extrema x^*

$$\frac{\mathrm{d}}{\mathrm{d}x}\widetilde{\phi}(x^*) = \frac{\mathrm{d}}{\mathrm{d}x}\phi(x^*) = 0 \quad \text{at} \quad \mu(x^*) = \lambda(x^*). \tag{52}$$

In fact, both have identical curvature near an extreme

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\widetilde{\phi}(x^*) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\phi(x^*) = \frac{1}{\mu(x^*)}\left(\frac{\mathrm{d}\lambda(x^*)}{\mathrm{d}x} - \frac{\mathrm{d}\mu(x^*)}{\mathrm{d}x}\right). \tag{53}$$

This means both have identical linear Gaussian dynamics near a fixed point. Equation (50) is a good approximation for the local dynamics.

$\widetilde{\phi}(x)$ and $\phi(x)$ can have very different global behaviour [26, 56]. To illustrate this, let us consider the particular example where

$$\mu(x) = \alpha_1 a x^2 + \beta_2 b \qquad \text{and} \qquad \lambda(x) = \alpha_2 x^3 + \beta_1 x \tag{54}$$

with $\alpha_1 = 6$, $\alpha_2 = 1.2$, $\beta_1 = 5.37$, $\beta_2 = 0.25$, $a = 1$ and $b = 1.4$. Figure 4 shows $\phi(x)$ as a solid blue line and $\widetilde{\phi}(x)$ as a dashed orange line. The two functions are indeed very similar (figure 4(*a*)); however, a careful inspection shows that the $\phi(x_1^*) < \phi(x_2^*)$ but $\widetilde{\phi}(x_1^*) > \widetilde{\phi}(x_2^*)$ (figure 4(*b*)). Therefore, when $V \to \infty$, the $f^{\mathrm{st}}(x) \to \delta(x - x_1^*)$ but $\widetilde{f}^{\mathrm{st}}(x) \to \delta(x - x_2^*)$.

4.2. Keizer's paradox

The disagreement between $\widetilde{\phi}(x)$ and $\phi(x)$ in figure 4 illustrates that a naive, truncated Kramers–Moyal expansion of the BDP (6) in the form of FPE (50) yields good local approximations near every fixed point, but cannot provide a globally satisfying approximation for sufficiently long times with uniform convergence of $V \to \infty$ with respect to $\forall t$. This failure is intimately related to the rare events that connect the bistability of the corresponding ODE.

The issue can be further elucidated by an even simpler model. Keizer [40] discussed the autocatalytic reaction system with

$$A + X \underset{\alpha_2}{\overset{\alpha_1}{\rightleftharpoons}} 2X \qquad \text{and} \qquad X \xrightarrow{\beta} B. \tag{55}$$

The ODE following the law of mass action is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha_1 a x - \alpha_2 x^2 - \beta x. \tag{56}$$

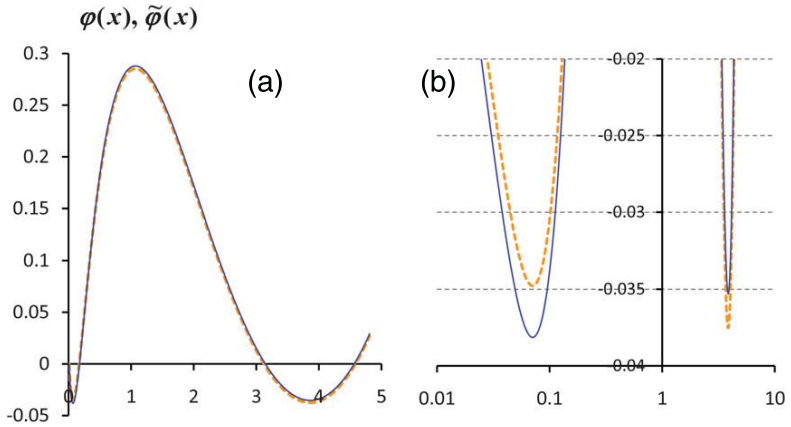

Figure 4. A comparison between the $\phi(x) = -(1/V)\ln f^{\mathrm{st}}(x)$, in solid blue, obtained from equation (6) in the limit of large volume, given by equation (10), and the $\widetilde{\phi}(x) = -(1/V)\ln \widetilde{f}^{\mathrm{st}}(x)$, in dashed orange, obtained from the diffusion approximation of the CME (6). For the Schlögl model, even though these two functions can be quite similar as shown in (*a*), a careful inspection in (*b*) shows that their global minima are different. It is at the left well for $\phi(x)$ and at the right well for $\widetilde{\phi}(x)$. This difference in the infinite-volume limit becomes very significant: the probability approaches to 1 at the global minimum.

For parameters $\alpha_1 a - \beta > 0$, this is in fact the celebrated logistic equation in population dynamics, with growth rate $\alpha_1 a - \beta$ and carrying capacity $x^* = \frac{\alpha_1 a - \beta}{\alpha_2}$. The ODE has two fixed points: unstable $x = 0$ and stable x^*.

In the chemical reaction context, Keizer observed that the ODE's stable steady state is inconsistent with the stationary distribution of the CME model for the reaction system (55). The CME is again a BDP with

$$\mu_n = \alpha_1 a n \qquad \text{and} \qquad \lambda_n = \frac{\alpha_2 n(n-1)}{V} + \beta n. \tag{57}$$

Because $\mu_0 = 0$, the $n = 0$ is an absorbing state of the BDP, and its stationary distribution has probability 1 for $n = 0$, i.e. extinction. But the $x = 0$ is an *unstable* fixed point of the ODE!

The resolution to these seemingly paradoxical results is simple [22, 40]. As indicated in section 2.2, system (55) again has a separation of the nonlinear network dynamics time scale t_{nd} and the cellular evolution time scale t_{ce}. The fixed point of the ODE is for $t \gg t_{\mathrm{nd}}$, but it is still $t \ll t_{\mathrm{ce}}$. For $t \gg t_{\mathrm{ce}}$, the system will be $n = 0$ with probability 1. However, for $t_{\mathrm{nd}} \ll t \ll t_{\mathrm{ce}}$, the system has a quasi-stationary distribution centred around the non-zero x^*. One can obtain this distribution as the eigenfunction associated with the largest non-zero eigenvalue of the CME. The eigenvalue, which $\propto \mathrm{e}^{-cV}$ $(c > 0)$, gives the time scale for reaching extinction.

Noting that from Kramers' theory, all barrier crossing rare events involve an exponentially slow 'climbing' to a saddle point and then a rapid 'descending' afterward (see appendix B). Keizer's paradox, therefore, is also at the root of the failure in section 4.1. The ODE predicts an infinitely long time for the climbing, but the FPE (50) predicts a time that is too short for the climbing.

4.3. The tale of two diffusion equations

Hänggi *et al* [56] proposed a different FPE that gives the correct stationary distribution

$$\frac{\partial f(x,t)}{\partial t} = \frac{1}{V}\frac{\partial^2}{\partial x^2}\left(\frac{\mu(x)-\lambda(x)}{\ln\mu(x)-\ln\lambda(x)}f\right) - \frac{\partial}{\partial x}\left((\mu(x)-\lambda(x))\,f\right). \tag{58}$$

It is not difficult to show that the stationary solution to equation (58) is the same as that in equation (10).

The 'physics rationale' for equation (58) is based on Onsager-type transport law as follows. First, the stationary distribution for the CME (6), in the limit of large V, is

$$p^{\mathrm{st}}(x) = \exp\left\{-V\int^x \ln\left(\frac{\lambda(z)}{\mu(z)}\right)\mathrm{d}z\right\}. \tag{59}$$

The stochastic potential for the system is $\phi(x) = -(1/V)\ln p^{\mathrm{st}}(x)$ and the thermodynamic force is $F(x) = -\mathrm{d}\phi(x)/\mathrm{d}x = \ln(\mu(x)/\lambda(x))$. The macroscopic ODE should be velocity $\times$ frictional coefficient = force,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu(x)-\lambda(x) = \eta^{-1}(x)F(x). \tag{60}$$

Therefore, this yields

$$\eta^{-1}(x) = \frac{\mu(x)-\lambda(x)}{\ln\mu(x)-\ln\lambda(x)}. \tag{61}$$

And the diffusion coefficient $\propto \eta^{-1}(x)$. This relation ensures the logarithm of the stationary distribution being $\propto -\phi(x)$. This approach, therefore, amounts to enforcing the deterministic kinetics and the stationary distribution. In a one-dimensional system, these two constraints essentially determine a FPE.

The two diffusion equations in (50) and (58) have the same drift $b(x)$ given in equation (49), but different diffusion coefficients,

$$A_{\mathrm{KM}}(x) = \mu(x)+\lambda(x) \qquad \text{and} \qquad A_{\mathrm{HGTT}}(x) = \frac{2(\mu(x)-\lambda(x))}{\ln\mu(x)-\ln\lambda(x)}, \tag{62}$$

where subscripts KM and HGTT stand for Kramers–Moyal and the authors of [56], respectively.

The two diffusion coefficients are the same near x^* where $\mu(x) = \lambda(x)$, i.e. the fixed point of the $b(x)$:

$$\frac{\mu-\lambda}{\ln(\mu/\lambda)} \approx \lambda\left[1+\frac{(\mu/\lambda)-1}{2}-\frac{(\mu/\lambda-1)^2}{12}+\cdots\right] \approx \frac{\mu+\lambda}{2}. \tag{63}$$

However, away from the fixed point of $b(x)$, HGTT's diffusion coefficient is always smaller than that of Kramers–Moyal's.

The HGTT diffusion, unfortunately, is not the full solution to the problem. First, it is not clear how to generalize this approach to higher dimensional problems. More importantly, it actually poses a dilemma. As an approximation to the CME, the Kramers–Moyal's diffusion gives the same finite time dynamics as the CME with large V, but a wrong stationary distribution. On the other hand, the HGTT diffusion gives the correct stationary distribution, but a wrong conditional diffusion equation for the finite time dynamics, as shown in appendix A. The HGTT diffusion does give the correct mean time for downhill dynamics, as does the KM diffusion, since the mean time for downhill is independent of diffusion (see equation (B.3)). However, the variances and distributions of the downhill times are different from the correct KM diffusion.

Therefore, no diffusion processes, with any possible $A(x)$ and $b(x)$, will give a satisfying representation of the dynamics predicted by a CME with bistability in the limit of $V \to \infty$.

The origin of this difficulty is the tremendous separation of time scales in the 'uphill' and 'downhill' motion. For a unistable system, uphill motion will only lead to an exponentially small probability which can be safely neglected. However, for systems with multi-stability, the exponentially small probability is responsible for establishing the correct probability between the two stochastic attractors. This difficulty renders the diffusion theory not capable of reasonably representing nonlinear stochastic population fluctuations. Rather, a hybrid model that combines continuous diffusion with discrete Markov jump processes is required.

4.4. Diffusion theory's dilemma

The problem can be even more tellingly stated as follows. According to the standard derivation of the diffusion equation from a discrete state and discrete time Markov chain with forward probability p, backward probability $q = 1 - p$, spacing δ and time step τ [59],

$$D = \frac{p+q}{2}\frac{\delta^2}{\tau} \qquad \text{and} \qquad b = (p-q)\frac{\delta}{\tau}. \tag{64}$$

Or

$$p = \frac{D\tau}{\delta^2} + \frac{b\tau}{2\delta} \qquad \text{and} \qquad q = \frac{D\tau}{\delta^2} - \frac{b\tau}{2\delta}. \tag{65}$$

We see that if both the diffusion coefficient D and the drift rate (bias) b exist, then in the limit of τ and $\delta \to 0$, $p/q \to 1$. In other words, diffusion theory assumes that in the very small spatial and temporal scale, the motion is purely random without bias. This feature, as we shall see, is inconsistent with the CME in the limit of $V \to \infty$.

If the p and q ($\neq p$) exist, then in the limit of τ and $\delta \to 0$, we have $D/b \to 0$. The diffusion is negligible. This is Kurtz's theorem [8].

It seems to us that the stochastic trajectory of the CME in the thermodynamic limit, depends on one's perspective. It is either a smooth, deterministic function of time, or a discontinuous stochastic function of time. There is really no diffusion process like behaviour. Thus we coined the term 'diffusion theory's dilemma.'

4.5. Diffusion theory's dilemma and exponentially small asymptotics

The discussion so far has explained why the CME in general does not converge to a proper diffusion. As in the law of large numbers, the proper limit is a system of deterministic ODEs [8]. To have a proper diffusion, one has to 'eliminate' the 'mean value.' The situation is completely analogous to the sum of N identical, independently distributed random variables, $Y_N = X_1 + X_2 + \cdots X_N$. There is simply no way to capture both the mean value and the variance of Y_N with a single scaling. The scaling for the law of large numbers is N^{-1}, while for the Central Limit Theorem it is $N^{-1/2}$. This is precisely the idea behind the van Kampen's system-size expansion, order-by-order [23, 24].

However, we still need to explain why the asymptotic form of the FPE with $1/V$ diffusion coefficient gives an erroneous stationary distribution. The insights from the present work point to this. The asymptotic order in the equation is singular. The stationary solution contains *exponentially small asymptotics* e^{-cV} ($c > 0$). This is the well-understood singularly perturbed linear two-point boundary value problem [60].

To give a better feel for the exponentially small asymptotics, let us consider the master equation with $\mu_n = \mu$ and $\lambda_n = \lambda$. Then the MFPT from n to zero, with reflecting boundary at n, is [10, 24]

$$T_n = \frac{1}{\mu - \lambda}\left(\frac{1 - (\lambda/\mu)^n}{(\lambda/\mu)^n - (\lambda/\mu)^{n+1}}\right) + \frac{n}{\lambda - \mu}. \tag{66}$$

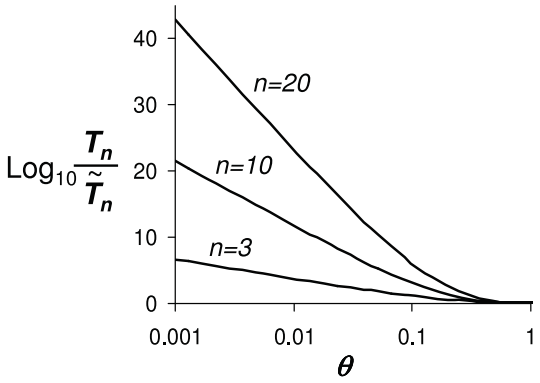

Figure 5. Comparison between the MFPTs in a BDP with constant birth and death rates, μ and λ, and corresponding Brownian motion with a constant drift $D = (\mu+\lambda)\delta^2/2$ and $V = (\mu-\lambda)\delta$. T_n is obtained from the discrete model and $\widetilde{T}_n$ is obtained from the corresponding diffusion process with $x = n\delta$. $\theta = \lambda/\mu < 1$ indicates the process to the absorbing state 0 is 'uphill.' With longer and longer 'climbing', i.e. larger n, the two times diverge exponentially.

Now if we consider the distance between n to $n+1$ being δ, and let $\delta \to 0$ and $n \to \infty$, but $n\delta \to x$, then we have a FPE,

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f}{\partial x^2} - V\frac{\partial f}{\partial x}, \tag{67}$$

where $D = (\mu+\lambda)\delta^2/2$ and $V = (\mu-\lambda)\delta$. The corresponding MFPT for this problem is [10, 24]

$$T_x = \frac{1}{V}\left[\frac{D}{V}\left(\mathrm{e}^{\frac{V}{D}x} - 1\right) - x\right]. \tag{68}$$

To compare the two results in equations (68) and (66), we rewrite equation (68) using $x = n\delta$ to obtain

$$\widetilde{T}_n = \frac{(\mu+\lambda)}{2(\mu-\lambda)^2}\left(\mathrm{e}^{\frac{2(\mu-\lambda)n}{(\mu+\lambda)}} - 1\right) + \frac{n}{\lambda-\mu}. \tag{69}$$

Comparing $\widetilde{T}_n$ and T_n in equations (69) and (66) we have

$$\frac{T_n}{\widetilde{T}_n} = \frac{\left(\frac{1-\theta^n}{\theta^n-\theta^{n+1}}\right) - n}{\frac{(1+\theta)}{2(1-\theta)}\left(\mathrm{e}^{\frac{2(1-\theta)n}{(1+\theta)}} - 1\right) - n}, \tag{70}$$

where $\theta = \lambda/\mu$. We then have

$$\lim_{n\to\infty}\frac{T_n}{\widetilde{T}_n} = \begin{cases}\infty & \theta < 1,\\ 1 & \theta \geqslant 1.\end{cases} \tag{71}$$

We note that $\theta > 1$ means the motion from positive x to zero is downhill. $\theta < 1$ means the motion from positive x to zero is uphill. The diffusion approximation for the master equation breaks down for the uphill dynamics! Figure 5 shows the ratio $T_n/\widetilde{T}_n$ as a function of θ and finite n.

We see that for the case of downhill dynamics ($\theta > 1$), both T_n and $\widetilde{T}_n$ approach $n/(\lambda-\mu)$, where $\lambda - \mu$ is the velocity. For the case of uphill dynamics ($\theta < 1$), both $\sim \mathrm{e}^{cn}$, but with

different, positive c's. In fact we have

$$\lim_{n\to\infty}\frac{1}{n}\ln\frac{T_n}{\widetilde{T}_n}=-\ln\theta-\frac{2(1-\theta)}{1+\theta}>0. \qquad (\theta<1). \tag{72}$$

We see the two expressions in equation (63) appear again here.

When $\theta\geqslant 1$, we indeed have that $T_n/\widetilde{T}_n$ converges with order $1/n$:

$$\lim_{n\to\infty} n\ln\frac{T_n}{\widetilde{T}_n}=\begin{cases}1 & \theta=1,\\ \frac{1}{2} & \theta>1.\end{cases} \tag{73}$$

This is at the heart of both examples. The stationary probabilities between two peaks are determined by the ratio of two exponentially long times to transition (back and forth), or two rare events with exponentially small probabilities. In other words, if an approximation can not give the 'exponent' correctly, then it is not a meaningful 'approximation"! In probability theory, this is the domain of the large deviation theory [19, 61].

4.6. The Delbrück–Gillespie process versus the Wright-Fisher model

The dynamic model for a mesoscopic, homogeneous chemical or biochemical reaction system is a stochastic process with birth and death. Any stochastic Markov process has two different mathematical representations: its stochastic trajectories and its time-dependent probability distribution following a Kolmogorov forward equation. In the context of the present review, they are the Gillepie algorithm [2] and the CME [62], respectively. While these two views of a stochastic process are mathematically equivalent, each only gives a partial understanding. For this reason, we would like to introduce the term *Delbrück–Gillespie* to refer to the stochastic process itself.

Tan [54, p 271] extensively discussed the conditions for a valid diffusion approximation (48) of finite BDPs like (6). A similar analysis is presented in appendix C. It showed that the necessary condition for a master equation to converge to a non-degenerate diffusion is

$$\ln\left(\frac{\mu_{n-1}(V)}{\lambda_n(V)}\right)\sim O\left(\frac{1}{V}\right)=\mathrm{d}x. \tag{74}$$

This form yields a stationary probability density $f^{\mathrm{st}}(x)$ which is properly supported on all x. When this condition is not met, as in the CME, one has

$$\ln\left(\frac{\mu_{n-1}(V)}{\lambda_n(V)}\right)=\left(V\ln\frac{\mu(x)}{\lambda(x)}\right)\mathrm{d}x.$$

The V on the rhs gives rise to the form $f^{\mathrm{st}}(x)=\mathrm{e}^{-V\phi(x)}$, which in the limit of $V=\infty$ will have x only supported at the global minimum of $\phi(x)$.

The population genetic models, on the other hand, have long enjoyed their fruitful relationship with the diffusion processes [54, 55, 63]. There is indeed a significant difference between the chemical reaction system and the genetic system. The random sampling in discrete genetic models is equivalent to a *long-range diffusion*, not just among the nearest neighbours. Hence it has a valid diffusion equation in the limit of large sample size, with both diffusion and drift terms being finite. This suggests in the infinite population size, deterministic limit, its nonlinear dynamics for the number density has the form $\mathrm{d}x/\mathrm{d}t=b_1(x)V+b_0(x)$, where the term $b_1(x)V$ means the individual reaction is aware of the size of the entire system. This is a *volume-dependent rate*.

This is indeed the case for the discrete population genetic drift model of Wright and Fisher [54, 55, 63]. But it does not arise from chemical kinetics in an ideal solution. The issue is as follows. In population genetic models, the conditional variance of an individual step is much greater than that in a Delbrück–Gillespie process. It is on the same order as the conditional expectation. In a chemical reaction, the former is a higher order infinitesimal.

5. Nonlinear and stochastic bistabilities

A basin of attraction around a stable fixed point in a deterministic nonlinear dynamical system corresponds to a peak in its stochastic counterpart. The converse is not necessarily true. A stochastic CME can have peaks that do not correspond to fixed points in the deterministic system of ODEs. We call the latter stochastic stability and the former nonlinear stability.

Consider the following autocatalytic reaction system [4, 5]

$$E + \chi E^* \xrightarrow{k_1} (\chi + 1)E^* \qquad \text{and} \qquad E^* \underset{k_3}{\overset{k_2}{\rightleftharpoons}} E, \tag{75}$$

in which χ can be either 1 or 2. This model resembles Keizer's logistic system in equation (55) and the Schlögl system in equation (1). Assuming the system's volume is V and there are N total number of E and E^* molecules, the stationary probability distribution for the number of E^*, p_n, satisfies the steady-state CME [4, 19, 27]

$$\begin{aligned}(N-n+1)\left(\frac{k_1(n-1)\cdots(n-\chi)}{V^\chi}+k_3\right)p_{n-1}\\ -\left[(N-n)\left(\frac{k_1 n\cdots(n-\chi+1)}{V^\chi}+k_3\right)+k_2 n\right]p_n + k_2(n+1)p_{n+1}=0.\end{aligned} \tag{76}$$

Solving the equation yields

$$p_n = p_0 \prod_{\ell=0}^{n-1} \frac{(N-\ell)}{(\ell+1)}\left(\frac{k_1\ell\cdots(\ell-\chi+1)}{k_2 V^\chi}+\frac{k_3}{k_2}\right), \tag{77}$$

where p_0 is a normalization factor. Let $x = n/N$ be the fraction of E^* among E and E^*. The probability distribution can be written as

$$\ln p(x) = \ln p_0 + \sum_{z=0,\delta}^{x-\delta} \ln\left[\frac{(1-z)(\theta z\cdots(z-\chi\delta+\delta)+\eta)}{(z+\delta)}\right], \tag{78}$$

where $\eta = k_3/k_2$, $\delta = 1/N$, and $\theta = (k_1/k_2)(N/V)^\chi$.

When the system size tends to infinity, V and $N \to \infty$, N/V tends to a finite concentration E_t, and $\delta \to 0$, we have an integral expression of the probability distribution

$$\ln f(x) = \text{const} + N\int_0^x \ln\left[\frac{(1-z)\,(\theta z^\chi+\eta)}{z}\right]\mathrm{d}z, \tag{79}$$

with continuous $x \in [0, 1]$, The distribution $f(x)$ has its extrema at the roots of the equation

$$\frac{(1-x)\,(\theta x^\chi+\eta)}{x} = 1. \tag{80}$$

The extrema of $f(x)$ match precisely with the macroscopic steady states from the law of mass action

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (k_1(E_t x)^\chi + k_3)\,(1-x) - k_2 x = 0. \tag{81}$$

Equation (80) gives only monostability for $\chi = 1$ and the possibility of bistability for $\chi = 2$. This is the macroscopic behaviour of the chemical reaction system in (75). However, for smaller system sizes, the distribution in equation (77) can in fact have two peaks even for $\chi = 1$, if $\delta > \eta$. In this case, the peak locations of the distribution p_n are at $n_1^* = 0$ and at n_2^*, the larger root of the quadratic equation

$$\theta\left(n_2^*\right)^2 - N(\theta - 1 - \eta)n_2^* + (N - N^2\eta) = 0. \tag{82}$$

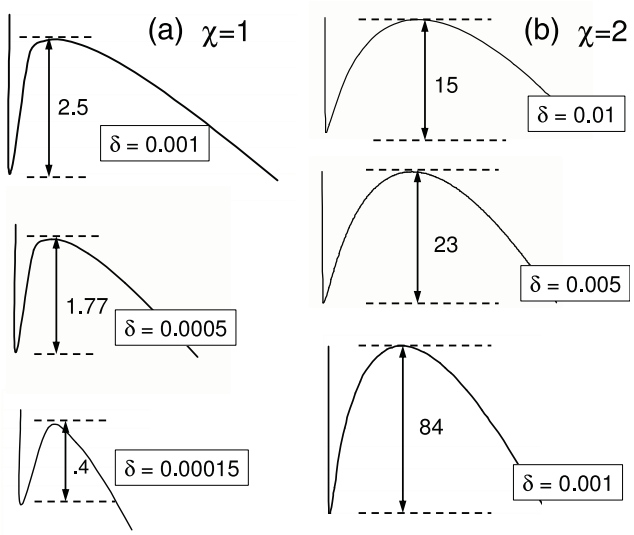

Figure 6. $-\ln p_n$ (the ordinate) as a function of n (the abscissa) and δ. (a) $\chi = 1$, $\theta = 1.5$, $\eta = 0.0001$. The peak locations, the smaller root of equation (82), are at $z = 0.002, 0.0008$, and 0.0001 for $\delta = 0.001, 0.0005, 0.00015$, respectively. ($b$) $\chi = 2$, $\theta = 10$, $\eta = 0.001$. The peak locations are at $z = 0.14, 0.125$, and 0.114, respectively. With increasing system's size, i.e., $\delta \to 0$, the lifetime of the state in (a) decreases while in (b) it increases. (a) is called *stochastic bistability* and (b) is called *nonlinear bistability*.

Figure 6(a) shows the $-\ln p_n$ for three different values of δ. We see that the stability of the 'energy well' at $n = 0$ decreases when δ tends to zero. The well disappears when $\delta < \eta$. In contrast, for $\chi = 2$, figure 6(b) shows the stability of the energy well at $n = 0$ increases when δ tends to zero.

The distinction between nonlinear and stochastic bistabilities is related to the concept of 'enthalpic barriers' in the Arrhenius theory of the chemical reaction rate [64], $k = e^{-\Delta H^1/k_B T + \Delta S^1/k_B}$ in which ΔH^1 and ΔS^1 are called activation enthalpy and entropy, respectively. With decreasing temperature, i.e. decrease the thermal randomness, the rate of crossing an enthalpic barrier decreases exponentially if $\Delta H^1 > 0$ but increases if $\Delta H^1 < 0$.

6. Kinetic isomorphism and general population dynamics

While we have so far focused on biochemical reaction kinetics, the theory of nonlinear, stochastic multi-dimensional BDPs we developed in the present paper could and should be applied to many other population dynamics [65]. In this section, we shall establish a kinetic isomorphism between chemical reaction systems and general population dynamics such as predator-and-prey, competition, and cannibalism. By doing so, our understanding and development of the stochastic, nonlinear biochemical dynamics can be easily transferred to the studies of many other population systems in ecology, infection epidemics, and sociology[1].

[1] It seems to us that a distinct feature of sociological dynamics is the possibility of 'volume-dependent rate' discussed in section 4.6, due to the rapid information exchange and government control in modern society.

6.1. The three types of predation functional responses

In mathematical ecology [66], the predation functional response characterizes the rate of prey consumed as a function of the density of the prey population under a constant environment including the predators. There are three widely used types of functional responses. Let r be the rate and x be the prey population density, then the three types are

$$r_1(x) \propto x, \qquad r_2(x) \propto \frac{x}{a+x}, \qquad \text{and} \qquad r_3(x) \propto \frac{x^n}{a^n + x^n}\ (n > 1). \tag{83}$$

The most important distinction between type I and types II and III is that the latter have a saturation effect. When there is a sufficiently large population of prey, more than enough for all the predators, the rate of consumption of the prey levels off.

These three different types of functional responses in equation (83) can be precisely represented by the following three types of chemical reactions, respectively:

$$X + A \xrightarrow{\alpha} B, \qquad X + A \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} XA \xrightarrow{\delta} B, \qquad \text{and} \qquad nX + A \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} X_nA \xrightarrow{\delta} B, \tag{84}$$

with the corresponding ODEs according to the law of mass action

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha a x, \qquad \frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\delta a x}{\left(\frac{\beta+\delta}{\alpha}\right) + x}, \qquad \text{and} \qquad \frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\delta a x^n}{\left(\frac{\beta+\delta}{\alpha}\right) + x^n}, \tag{85}$$

where a is the total concentration of molecular species A, which is kept constant in the reaction. The derivation for the expressions in equation (85) from chemical kinetic schemes in equation (84) involves singular perturbations and can be found in many enzyme kinetic textbooks [7].

It is interesting to note the correspondence between type III response and the molecular cooperativity. In ecological systems, type III response is associated with learning, that is, the natural improvement of a predator's searching and attacking efficiency as prey density increases.

6.2. Birth and death rates in multi-prey predation

Let us now consider the case of one predator population Y who has n different possible prey species X_i, $1 \leqslant i \leqslant n$. Let the consumption rate of X_i, per Y, in the absence of all the other X_j's ($j \neq i$, $1 \leqslant j \leqslant n$) be type II functional response,

$$\frac{a_i x_i}{1 + a_i \tau_i x_i},$$

where a_i is the attack rate and τ_i is called handling time. In the presence of all the X_i, one then has ([67], section 7.2)

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i \left(\frac{a_i y}{1 + \sum_{j=1}^{n} a_j \tau_j x_j} \right) \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = y \left(\frac{\sum_{i=1}^{n} e_i a_i x_i}{1 + \sum_{i=1}^{n} a_i \tau_i x_i} \right), \tag{86}$$

where e_i is known as consumer efficiency. The first equation is the death rate of the prey population X_i caused by the predator, and the second equation is the birth rate of the predator with multiple preys.

In biochemical reaction terms, the predator Y is an autocatalytic enzyme which transforms the various X_i into Y, where

$$X_i + Y \underset{\beta_i}{\overset{\alpha_i}{\rightleftharpoons}} X_i Y \xrightarrow{\kappa_i} A + Y \qquad \text{and} \qquad X_i + Y + A \underset{\psi_i}{\overset{\phi_i}{\rightleftharpoons}} X_i Y A \xrightarrow{\delta_i} X_i + 2Y. \tag{87}$$

If we set the concentration of A, $a = \frac{\alpha_i}{\beta_i} \times \frac{\psi_i}{\phi_i}$, and assume the binding steps are in rapid equilibrium, then the chemical kinetic equations for the concentrations of X_i and Y are

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -\frac{\kappa_i \left(\frac{\alpha_i}{\beta_i}\right) x_i y}{1 + \sum_{i=1}^{n} \frac{\alpha_i}{\beta_i} x_i} \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = y \left(\frac{\sum_{i=1}^{n} \delta_i \left(\frac{\alpha_i}{\beta_i}\right) x_i}{1 + \sum_{i=1}^{n} \frac{\alpha_i}{\beta_i} x_i} \right). \tag{88}$$

Comparing equation (88) with equation (86), we have

$$\frac{\alpha_i}{\beta_i} = a_i \tau_i, \qquad \kappa_i = \frac{1}{\tau_i}, \qquad \delta_i = e_i. \tag{89}$$

6.3. Four planar population systems

Here we consider the dynamics of two interacting populations is a planar nonlinear system and give the chemical kinetic equivalences of several well-known examples. Almost all textbooks on differential equations and in mathematical biology discuss such systems [29, 66].

Lotka–Volterra's predator–prey model. The widely studied model for predator and prey dynamics was originally developed as a system of chemical reactions containing autocatalysis,

$$A \xrightarrow{k_1} X, \qquad Y \xrightarrow{k_2} B \qquad \text{and} \qquad X + Y \xrightarrow{k_3} 2Y. \tag{90}$$

The corresponding ODEs from the law of mass action are given in equation (92) with $\widetilde{k}_3 c = k_4 = k_3$. In ecological terms, X and Y are the prey and predator, respectively. The prey is the sole food of the predator, and it has a linear grow rate of $k_1 a$ in the absence of the predator. In the absence of the prey, the predator has a death rate of k_2.

In an ecological context, there is no fundamental reason for the two xy terms in equation (92) to be equal. Hence, the faithful chemical reaction representation for the predator–prey model is

$$A \xrightarrow{k_1} X, \qquad Y \xrightarrow{k_2} B, \qquad X + Y + C \xrightarrow{\widetilde{k}_3} X + 2Y \qquad \text{and} \qquad X + Y \xrightarrow{k_4} D + Y. \tag{91}$$

With the concentrations for chemical species X, Y, A, B, C being x, y, a, b, c, we have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (k_1 a)\, x - k_4 x y \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = -k_2 y + \left(\widetilde{k}_3 c\right) x y. \tag{92}$$

See [68, 69] for a recent study of a generalization of the Lotka–Volterra system with a chemical perspective, which yields new insights to the classic problem.

Competition model. The second widely studied type of planar population dynamics involves a competition between two species. In chemical kinetic terms,

$$\begin{aligned} &A + X \xrightarrow{k_1} 2X, \qquad X + X \xrightarrow{k_2} C, \qquad X + Y \xrightarrow{k_3} Y + E, \\ &B + Y \xrightarrow{k_4} 2Y, \qquad Y + Y \xrightarrow{k_5} D, \qquad X + Y \xrightarrow{k_6} X + F. \end{aligned} \tag{93}$$

The mass-action kinetic equations for dynamical X and Y, with constant populations of A and B, are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (k_1 a)\, x - k_2 x^2 - k_3 x y \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = (k_4 b)\, y - k_5 y^2 - k_6 x y. \tag{94}$$

Both X and Y, in the absence of the competition, have logistic growth. Species X has a linear growth rate of $k_1 a$ and carrying capacity of $\frac{k_1 a}{k_2}$, and species Y has $k_4 b$ and $\frac{k_4 b}{k_5}$. Equation (94) is precisely the equations in section 3.5 of [29].

Mutualism or symbiosis. Murray [29] also presented a model for symbiosis in which species X and Y are in cooperation. In this case, the signs of the xy terms in equation (94) are positive rather than negative. The chemical reaction system that yields such dynamics is

$$A + X \xrightarrow{k_1} 2X, \qquad X + X \xrightarrow{k_2} E, \qquad X + Y + C \xrightarrow{k_3} 2X + Y,$$
$$B + Y \xrightarrow{k_4} 2Y, \qquad Y + Y \xrightarrow{k_5} F, \qquad X + Y + D \xrightarrow{k_6} X + 2Y. \tag{95}$$

The corresponding ODEs are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (k_1 a)\, x - k_2 x^2 + (k_3 c)\, xy \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = (k_4 b)\, y - k_5 y^2 + (k_6 d)\, xy. \tag{96}$$

Models of cannibalistic demography. A single population with cannibalism can be 'modelled' by the following chemical reaction system, known as energy relay [70],

$$2X + E \xrightarrow{k_1} X + E^*, \qquad A + E^* \xrightarrow{k_2} X + E^*, \qquad E^* \xrightarrow{k_3} E. \tag{97}$$

Let the concentrations for X and E^* be x and e^*, and the total E and E^* together is a constant e_t. Then, the law of mass action gives us

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -k_1(e_t - e^*)x^2 + k_2 a e^* \qquad \text{and} \qquad \frac{\mathrm{d}e^*}{\mathrm{d}t} = k_1(e_t - e^*)x^2 - k_3 e^*. \tag{98}$$

Treating E and E^* as an 'enzyme' following the Michaelis–Menten kinetics, we have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{(k_1 e_t)\,(k_2 - k_3)}{k_1 x^2 + k_3} x^2. \tag{99}$$

A juveniles and adults two-age model with cannibalism of juveniles by adults [71] can be understood as a predator–prey system (adult as predator, juvenile as prey) with a population transfer from juveniles to adults, such that

$$A \xrightarrow{k_1} X, \qquad Y \xrightarrow{k_2} B, \qquad X + Y + C \xrightarrow{k_3} X + 2Y,$$
$$X + Y \xrightarrow{k_4} D + Y \qquad \text{and} \qquad X \xrightarrow{k_5} Y, \tag{100}$$

in which X and Y are the juveniles and adults. Therefore, we have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (k_1 a)\, x - k_4 xy - k_5 x \qquad \text{and} \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = k_5 x - k_2 y + \left(\widetilde{k_3 c}\right) xy. \tag{101}$$

7. Discussion and outlook

7.1. Nonlinear, stochastic biochemical dynamics as a new paradigm

Currently, there are mainly two mathematical approaches to biological systems and phenomena. One is based on *principles and mechanisms* and the other is based on *data*. Research on protein molecular dynamics and on the Hodgkin–Huxley equation of excitable cells belongs to the the first kind, while statistical research on bioinformatics, ecology, economics, etc belong to the second kind. In between, there are *modellers* who struggle to develop mathematical principles from the data. Mathematical biologists are a large group of modellers. Accordingly, modelling of biological dynamics has been roughly divided into *deterministic* and *statistical* approaches. The study of *cellular biochemical dynamics*, as a paradigm, offers a new perspective on biological dynamics.

The *nonlinear, stochastic cellular biochemical dynamics* offers a new mathematical framework for dynamics that encompasses both deterministic and statistical aspects of

modelling. The benefits go further than this. Perhaps one of the most important insights is the emergence of *rare events* which has infinitesimal probability to occur in a regular time scale, but it will occur with probability 1 on an evolutionary time scale. Rare events can be understood by neither classical deterministic mathematics nor normal statistics. The only tool we know of is mechanistic stochastic modelling.

Cancers, ecological catastrophes, stock market crashes, and sociopolitical revolutions are all rare events. It is these rare events that are truly *unpredictable* in the classical sense, giving the appearance of *free will* [16]. John Hopfield called it *dynamic symmetry breaking*. James Clerk Maxwell has said, 'It is manifest that the existence of unstable conditions renders impossible the prediction of future events, if our knowledge of the present state is only approximate, and not accurate. At these (unstable) points, influences whose physical magnitude is too small to be taken account of by a finite being, may produce results of the greatest importance. All great results produced by human endeavour depend on taking advantage of these singular states when they occur.' [17]

7.2. Stochastic dynamics in terms of multi-dimensional BDPs and diffusion processes

Stochastic processes have gradually become an indispensable and powerful mathematical description of biological dynamics, from cellular biochemical to ecological systems. Yet, compared with the understanding of nonlinear deterministic dynamical systems, our current in-depth knowledge of applied stochastic processes are still rather limited. This is particularly true for stochastic processes with time-irreversibility. Time-reversible processes are appropriate models for equilibrium dynamics. Stochastic dynamics of living systems have to be time-irreversible [33, 34, 48].

The interaction between the stochastic aspect and the nonlinear aspect of dynamics creates complex behaviour. This is a subject that is yet to be fully explored. Markov processes, the stochastic counterpart of first-order ordinary differential equations, have two equally valid mathematical representations, the trajectories and the Kolmogorov forward equations. For diffusion processes that have been widely employed in physics and chemistry, these correspond to SDEs and the FPE [72]. For multi-dimensional BDPs, they correspond to the Lebowitz–Gillespie algorithm [2, 73] and master equations, respectively.

The Delbrück–Gillespie process is the stochastic counterpart of the deterministic mass-action kinetics. It is a full range analytical theory of dynamics of homogeneous chemical and biochemical reaction systems. It is more than either the wildly popular Gillespie algorithm or the CME alone. It has an emergent nonlinear differential equation system as well as the emerging stochastic jump dynamics on an evolutionary time scale [5].

Since the pioneering work of Einstein, Smoluchowski, Langevin and Kramers, the diffusion process, with its continuous but everywhere non-differentiable trajectory, has become the dominant mathematical theory for stochastic processes. The physicists' approach to stochastic dynamics, however, is markedly 'deterministic centric.' The entire stochastic enterprise of statistical physics is to understand macroscopic, deterministic behaviour from the atomic nature of matters.

The stochasticity has always been considered merely as 'fluctuations.' In fact, physicists have long believed that there is no stochasticity in a macroscopic world. This view, of course, has been justified by the law of large number in the theory of probability. This perspective, as we have shown in the present paper, needs to be modified to embrace a macroscopic complexity with variations and stochastic jumps [16]. The law of large numbers, it turns out, requires an infinitely long time and large system. For mesoscopic systems [74], there are stochastic

dynamics beyond the deterministic limit [15, 19], and for macroscopic systems, stochasticity occurs on an evolutionary time scale.

Finally, in biological and many other non-mechanical systems, a BDP is a more fundamental approach to stochastic dynamics than the continuous diffusion. Certainly, it is more consistent with all the deterministic differential equation models based on number density, be it individuals, organisms, cells or molecules. The diffusion approach, however, is only phenomenological. It has to rely on additional information to specify the diffusion coefficient. This is why the fluctuation–dissipation relations are so essential in equilibrium statistical physics.

7.3. Intrinsic and extrinsic fluctuations: stochastic processes versus random dynamical systems (RDSs)

According to van Kampen [24], 'internal or intrinsic fluctuations are caused by the fact that the system itself consists of discrete particles; it is inherent in the very mechanism by which the system evolves.' External or extrinsic noises, on the other hand, often reside in systems' parameters or environments. The distinction between these two types of randomness in a dynamic process can be best illustrated if one considers two trajectories with different initial conditions, and asks if the two dynamics utilize two different, independent sequences of realizations of random events, or the same sequence. In the theory of RDSs [75], it is the latter. A SDE can be interpreted as both. However, a Delbrück–Gillespie (DG) process does not fit the RDS perspective. The random process underlying a DG process is a time-changed Poisson process. To consider two DG processes with same realization of a Poisson process can only come from a globally synchronized clock [5]. Therefore, the fluctuations in the DG process are due to intrinsic noise. There is no separation between the deterministic nonlinear dynamics and the stochastic fluctuations.

Acknowledgments

I thank Professors R E O'Malley and W J Heuett for carefully reading the manuscript, and Professor O'Malley for numerous discussions. Collaborations with Drs Jia-zeng Wang and Da Zhou of Peking University are highly appreciated.

Appendix A. Conditional diffusion equation

Let us consider the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \epsilon\sqrt{A(X_t)}\,\mathrm{d}B_t, \qquad X_0 = x_0. \tag{A.1}$$

Let $\xi(t)$ be the deterministic solution to the ODE

$$\frac{\mathrm{d}\xi(t)}{\mathrm{d}t} = b(\xi(t)), \qquad \xi(0) = x_0. \tag{A.2}$$

Then we can consider

$$Y_t = \frac{X_t - \xi(t)}{\epsilon}, \qquad Y_0 = 0. \tag{A.3}$$

which satisfies the SDE

$$\mathrm{d}Y_t = \frac{b(\epsilon Y_t + \xi(t)) - b(\xi(t))}{\epsilon}\mathrm{d}t + \sqrt{A(\epsilon Y_t + \xi(t))}\,\mathrm{d}B_t. \tag{A.4}$$

The probability density function for Y_t satisfies the time-inhomogeneous FPE

$$\frac{\partial}{\partial t} f_Y(y,t) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(A(\epsilon y + \xi(t)) f_Y\right) - \frac{\partial}{\partial y}\left(\frac{b(\epsilon y + \xi(t)) - b(\xi(t))}{\epsilon} f_Y\right). \tag{A.5}$$

For small ϵ, equation (A.5) can be approximated as

$$\frac{\partial}{\partial t} f_Y(y,t) = \frac{A(\xi(t))}{2}\frac{\partial^2 f_Y}{\partial y^2} - b'(\xi(t))\frac{\partial\,(y f_Y)}{\partial y}. \tag{A.6}$$

In particular, if the deterministic solution to (A.2) asymptotically approaches a fixed point x^*, and if we choose $X_0 = x^*$, then equation (A.6) is simplified to the FPE for an Ornstein–Uhlenbeck Gaussian process,

$$\frac{\partial}{\partial t} f_Y(y,t) = \frac{A(x^*)}{2}\frac{\partial^2 f_Y}{\partial y^2} - b'(x^*)\frac{\partial\,(y f_Y)}{\partial y}. \tag{A.7}$$

Appendix A.1. Gaussian solution to the conditional diffusion

To solve equation (A.6), we introduce

$$\langle Y_t\rangle = \int_{-\infty}^{\infty} y f_Y(y,t)\,\mathrm{d}y, \qquad \langle Y_t^2\rangle = \int_{-\infty}^{\infty} y^2 f_Y(y,t)\,\mathrm{d}y, \tag{A.8}$$

then we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle Y_t\rangle = b'(\xi(t))\langle Y_t\rangle, \qquad \langle Y_0\rangle = 0,\\
&\frac{\mathrm{d}}{\mathrm{d}t}\langle Y_t^2\rangle = A(\xi(t)) + 2b'(\xi(t))\langle Y_t^2\rangle, \qquad \langle Y_0^2\rangle = 0,\\
&\langle Y_t\rangle = 0,\\
&\langle Y_t^2\rangle = \begin{cases} b^2(\xi(t))\displaystyle\int_0^t \frac{a(\xi(s))}{b^2(\xi(s))}\,\mathrm{d}s, & \xi'(t) \neq 0\\ \dfrac{A(x^*)}{2|b'(x^*)|}\left(1 - \mathrm{e}^{-2|b'(x^*)|t}\right), & \xi(t) \equiv x^*, b'(x^*) < 0. \end{cases}
\end{aligned} \tag{A.9}$$

It is easy to verify that the solution to equation (A.6) is

$$f_Y(y,t) = \frac{1}{\sqrt{2\pi\langle Y_t^2\rangle}}\exp\left(-\frac{y^2}{2\langle Y_t^2\rangle}\right). \tag{A.10}$$

Appendix A.2. Linear fluctuation theory according to conditional diffusion

If the $\xi(t)$ is near a stable fixed point of the $b(x)$, then $\xi(t|x_0) = x^* + (x_0 - x^*)\mathrm{e}^{-\beta t}$ where $\beta = |b'(x^*)|$. Furthermore, $b(\xi(t)) \approx -\beta(x_0 - x^*)\mathrm{e}^{-\beta t}$. Then equation (A.9) gives

$$\langle Y_t^2\rangle = \frac{A(x^*)}{2\beta}\left(1 - \mathrm{e}^{-2\beta t}\right). \tag{A.11}$$

And the autocorrelation function for the stationary process is

$$\begin{aligned}
\langle X_\tau X_0\rangle^{\mathrm{st}} &= \int_{\infty}^{\infty} \mathrm{d}x_0 f^{\mathrm{st}}(x_0) x_0 \xi(\tau|x_0)\\
&= \langle X\rangle^{\mathrm{st}} x^*\left(1 - \mathrm{e}^{-\beta\tau}\right) + \langle X^2\rangle^{\mathrm{st}}\mathrm{e}^{-\beta\tau}.
\end{aligned} \tag{A.12}$$

Therefore,

$$\langle \Delta X_\tau \Delta X_0\rangle^{\mathrm{st}} = \langle (\Delta X)^2\rangle^{\mathrm{st}}\mathrm{e}^{-\beta\tau}. \tag{A.13}$$

These results have been obtained many times in the theory of stochastic linear relaxation, in the work of L Onsager, M Lax, T L Hill and J Keizer. We see that the conditional diffusion shares the same principle. The stochastic dynamics is a 'correction term' to the deterministic behaviour.

Appendix B. The mean first passage time (MFPT)

Appendix B.1. MFPT for a 1D diffusion process

The MFPT T for a diffusion process with diffusion coefficient $A(x)/2$ and drift $b(x)$, from x_1 to x_2, satisfies [10, 14, 24]

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{A(x)}{2}\frac{\mathrm{d}T(x)}{\mathrm{d}x}+b(x)\frac{\mathrm{d}T(x)}{\mathrm{d}x}=-1,\qquad \frac{\mathrm{d}T(x_1)}{\mathrm{d}x}=0,\qquad T(x_2)=0, \tag{B.1}$$

$$T_{x_1\to x_2}=\int_{x_1}^{x_2}\mathrm{e}^{-\phi(x)}\,\mathrm{d}x\int_{x}^{x_2}\mathrm{e}^{\phi(y)}\frac{2dy}{A(y)},\qquad \phi(x)=-\int^{x}\frac{2b(x)}{A(x)}\,\mathrm{d}x. \tag{B.2}$$

If interval (x_1, x_2) contains an energy well at x_1^*, $b(x_1^*)=0$, $b'(x_1^*)<0$, and an energy barrier at x_3^*, $b(x_3^*)=0$, $b'(x_3^*)>0$, then one can use Laplace's method to simplify equation (B.2) to

$$T_{x_1\to x_2}=\frac{4\pi\mathrm{e}^{\phi(x_3^*)-\phi(x_1^*)}}{A(x_3^*)\sqrt{|\phi''(x_1^*)\phi''(x_3^*)|}}+\int_{x_3^*}^{x_2}\frac{2\,\mathrm{d}x}{A(x)|\phi'(x)|}. \tag{B.3}$$

The second term is for downhill relaxation. In fact, $A(x)|\phi'(x)|/2=|b(x)|$. Thus the downhill time is essentially determined by the drift $b(x)$, independent of $A(x)$. The first term is the time for barrier crossing. The inverse of this expression is known as Kramers formulae for reaction rate. For a barrier crossing problem, the second term can be neglected.

Appendix B.2. MFPT for a 1D BDP

A same calculation can be carried out for a BDP [10]. For a one-dimensional CME with birth rate μ_n and death rate λ_n, in the limit of large V, $\mu_{xV}\to\mu(x)V$, $\lambda_{xV}\to\lambda(x)V$, and one obtains [64]

$$T_{x_1\to x_2}=\frac{2\pi\mathrm{e}^{V(\phi(x_3^*)-\phi(x_1^*))}}{\lambda(x_3^*)\sqrt{|\phi''(x_1^*)\phi''(x_3^*)|}}+\int_{x_3^*}^{x_2^*}\frac{1}{\lambda(y)|\phi'(y)|}\,\mathrm{d}y, \tag{B.4}$$

in which

$$\phi(x)=\int^{x}\ln\left[\frac{\lambda(z)}{\mu(z)}\right]\mathrm{d}z. \tag{B.5}$$

Note that at fixed point x_3^*, $\lambda(x_3^*)=\mu(x_3^*)$, which corresponds to $A(x_3^*)/2$. Hence the first terms in equations (B.3) and (B.4) are completely identical. The detailed $\phi(x)$ and $A(x)$ between x_1^* and x_3^* do not matter to the barrier crossing time.

The downhill time, however, is significantly different from that of diffusion theory. Using either $A(x)=\mu(x)+\lambda(x)$, as in the Kramers–Moyal expansion, or $A(x)=2(\mu(x)-\lambda(x))/(\ln\mu(x)-\ln\lambda(x))$, as in Onsager's theory, gives the same result in equation (B.7),

$$\text{diffusion}:\int\frac{\mathrm{d}x}{\mu(x)-\lambda(x)}, \tag{B.6}$$

$$\text{CME}:\qquad\int\frac{\mathrm{d}x}{\lambda(x)\,(\ln\mu(x)-\ln\lambda(x))}. \tag{B.7}$$

These are the different predictions based on the diffusion theory and on the CME.

Appendix C. Master equations with and without diffusion limit

Let us consider the canonical FPE for diffusion processes,

$$\frac{\partial f}{\partial t} = \frac{\epsilon}{2}\frac{\partial^2}{\partial x^2}\left(A(x)f\right) - \frac{\partial}{\partial x}\left(b(x)f\right). \tag{C.1}$$

If we discretize the x in terms of a uniform interval δ, we have

$$\begin{aligned}\frac{\mathrm{d}f(x,t)}{\mathrm{d}t} &= \left[\frac{\epsilon A(x-\delta)}{2\delta^2} + \frac{b(x-\delta)}{2\delta}\right] f(x-\delta) \\ &\quad - \left[\left(\frac{\epsilon A(x)}{2\delta^2} + \frac{b(x)}{2\delta}\right) + \left(\frac{\epsilon A(x)}{2\delta^2} - \frac{b(x)}{2\delta}\right)\right] f(x) \\ &\quad + \left[\frac{\epsilon A(x+\delta)}{2\delta^2} - \frac{b(x+\delta)}{2\delta}\right] f(x+\delta).\end{aligned} \tag{C.2}$$

Therefore, if the birth and death rates of a master equation, $\mu_n(V)$ and $\lambda_n(V)$, are in the forms of

$$\mu_n = V\left[\frac{\epsilon A(n/V)}{2\delta} + \frac{b(n/V)}{2}\right] \qquad \text{and} \qquad \lambda_n = V\left[\frac{\epsilon A(n/V)}{2\delta} - \frac{b(n/V)}{2}\right], \tag{C.3}$$

then we have the master equation in (6). Note that both μ_n and λ_n have to be non-negative. This is guaranteed by the δ being sufficiently small. We see that if ϵ is smaller, the δ has to be smaller as well.

The stationary distribution of the master equation is readily obtained, such that

$$p_n^{\mathrm{st}} = p_0 \exp\left[\sum_{k=1}^{n} \ln\frac{\mu_{k-1}}{\lambda_k}\right] \tag{C.4}$$

and

$$\frac{f^{\mathrm{st}}(x)}{V} = p_0 \lim_{V\to\infty} \exp\left[\sum_{k=1}^{xV} \ln\frac{A\left(\frac{k-1}{V}\right)\frac{\epsilon}{\delta} + b\left(\frac{k-1}{V}\right)}{A\left(\frac{k}{V}\right)\frac{\epsilon}{\delta} - b\left(\frac{k}{V}\right)}\right], \tag{C.5}$$

which yields

$$f^{\mathrm{st}}(x)\,\mathrm{d}x = A\exp\left[V\int_0^x \ln\frac{(\epsilon/\delta)A(x)+b(x)}{(\epsilon/\delta)A(x)-b(x)}\,\mathrm{d}z\right]. \tag{C.6}$$

Note that if we consider ϵ and $\delta \to 0$, but $\epsilon/\delta \to \nu$, then we have

$$f^{\mathrm{st}}(x)\,\mathrm{d}x = A\exp\left[-\frac{\nu}{\epsilon}\int_0^x \ln\frac{\nu A(x)-b(x)}{\nu A(x)+b(x)}\,\mathrm{d}z\right]. \tag{C.7}$$

The $\ln f^{\mathrm{st}}(x)$ has its extrema at x^* with $b(x^*) = 0$. Furthermore, the curvature at x^* is $2b'(x^*)/(\epsilon A(x^*))$, which is independent of ν.

More interestingly, if $\nu = 1$, we have

$$f^{\mathrm{st}}(x)\,\mathrm{d}x = A\exp\left[-\frac{1}{\epsilon}\int_0^x \ln\frac{A(x)-b(x)}{A(x)+b(x)}\,\mathrm{d}z\right]. \tag{C.8}$$

This is the case of the CME in which μ_n and λ_n have the form of $\mu_n = \mu(x)V$ and $\lambda_n = \lambda(x)V$ as in equation (C.3).

If, however, $\nu = \infty$,

$$f^{\mathrm{st}}(x)\,\mathrm{d}x = A\exp\left[\frac{1}{\epsilon}\int_0^x \frac{2b(z)}{A(z)}\,\mathrm{d}z\right]. \tag{C.9}$$

Only equation (C.9) recovers the correct stationary distribution to the FPE in (C.1). One can not have $\nu \to 0$ because of the discussion following equation (C.3).

Equations (C.9) and (C.8) are precisely the correct and wrong stationary distributions according to [56] and Kramers–Moyal's diffusion equations, respectively. Their difference is exactly the two expressions on the lhs and rhs of equation (63).

References

[1] Samoilov M S and Arkin A P 2006 *Nature Biotechnol.* **24** 1235
[2] Gillespie D T 2007 *Annu. Rev. Phys. Chem.* **58** 35
[3] Liang J and Qian H 2010 *J. Comput. Sci. Technol.* **25** 154
[4] Qian H and Bishop L M 2010 *Int. J. Mol. Sci.* **11** 3472
[5] Qian H 2010 *J. Stat. Phys.* **141** 990
[6] Wilkinson D J 2006 *Stochastic Modeling for Systems Biology* (New York: Chapman & Hall/CRC)
[7] Beard D A and Qian H 2008 *Chemical Biophysics: Quantitative Analysis of Cellular Systems* (Cambridge: Cambridge University Press)
[8] Kurtz T G 1971 *J. Appl. Probab.* **8** 344
[9] Qian H, Saffarian S and Elson E L 2002 *Proc. Natl Acad. Sci. USA* **99** 10376
[10] Gardiner C W 1983 *A Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences* (Berlin: Springer)
[11] Kurtz T G 1981 *Approximation of Population Processes* (Cambridge: Cambridge University Press)
[12] Stroock D W and Varadhan S R S 1979 *Multidimensional Diffusion Processes* (Berlin: Springer)
[13] Wax N 1954 *Selected Papers on Noise and Stochastic Processes* (New York: Dover)
[14] Schuss Z 2010 *Theory and Applications of Stochastic Processes: An Analytical Approach* (New York: Springer)
[15] Qian H, Shi P-Z and and Xing J 2009 *Phys. Chem. Chem. Phys.* **11** 4861
[16] Hopfield J J 1994 *J. Theor. Biol.* **171** 53.
[17] Prigogine I and Stengers I 1984 *Order Out of Chaos* (Boulder, CO: New Sci. Lib. Shambhala Pub.)
[18] Hänggi P, Talkner P and Borkovec M 1990 *Rev. Mod. Phys.* **62** 251
[19] Ge H and Qian H 2011 *J. R. Soc. Interface* **8** 107
[20] Landauer R and Swanson J A 1961 *Phys. Rev.* **121** 1668
[21] Langer J S 1969 *Ann. Phys.* **54** 258
[22] Vellela M and Qian H 2007 *Bull. Math. Biol.* **69** 1727
[23] van Kampen N G 1961 *Can. J. Phys.* **39** 551
[24] van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland/Elsevier Science)
[25] Schlögl F 1972 *Z. Physik.* **253** 147
[26] Vellela M and Qian H 2009 *J. R. Soc. Interf.* **6** 925
[27] Bishop L M and Qian H 2010 *Biophys. J.* **98** 1
[28] Shi P-Z and Qian H 2011 *J. Chem. Phys.* **134** 065104
[29] Murray J D 2004 *Mathematical Biology I: An Introduction* (New York: Springer)
[30] Epstein I R and Pojman J A 1998 *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos* (Oxford: Oxford University Press)
[31] Qian H and Reluga T C 2005 *Phys. Rev. Lett.* **94** 028101
[32] Ge H and Qian H 2009 *Phys. Rev. Lett.* **103** 148103
[33] Qian H 2006 *J. Phy. Chem. B* **110** 15063 (feature article)
[34] Qian H 2007 *Annu. Rev. Phys. Chem.* **58** 113
[35] Keizer J 1986 *Probability, Statistical Mechanics, and Number Theory: A Volume Dedicated to Mark Kac* (*Advances in Mathematics Supplementary Studies* vol 9) ed G-C Rota (New York: Academic) p 1
[36] Hu G 1986 *Zeit. Phys.* B **65** 103
[37] Ge H and Qian H 2010 arxiv.1011.4049
[38] Onsager L and Machlup S 1953 *Phys. Rev.* **91** 1505
[39] Lax M 1960 *Rev. Mod. Phys.* **32** 25
[40] Keizer J 1979 *Acc. Chem. Res.* **12** 243
[41] Fox R F 1978 *Phys. Rep.* **48** 180
[42] Qian H 2001 *Proc. R. Soc.* A **457** 1645
[43] Kwon C, Ao P and Thouless D J 2005 *Proc. Natl Acad. Sci. USA* **102** 13029
[44] Gajic Z and Qureshi M T J 2008 *Lyapunov Matrix Equation in System Stability and Control* (New York: Dover)

[45] Jiang D-Q, Qian H and Wang J-Z 2011 in preparation
[46] Jiang D-Q and Zhang F-X 2003 *J. Math. Phys.* **44** 4681
[47] Chen Y, Qian M-P and Xie J-S 2007 *J. Math. Phys.* **48** 073301
[48] Jiang D-Q, Qian M and Qian M-P 2004 *Mathematical Theory of Nonequilibrium Steady States: On the Frontier of Probability and Dynamical Systems* (New York: Springer)
[49] Qian H and Elson E L 2004 *Proc. Natl Acad. Sci. USA* **101** 2828
[50] Zwanzig R 1965 *Annu. Rev. Phys. Chem.* **16** 67
[51] Qian H and Qian M 2000 *Phys. Rev. Lett.* **84** 2271
[52] Qian M, Wang G-X and Zhang X-J 2000 *Phys. Rev.* E **62** 6469
[53] Qian M and Zhang X-J 2001 *Phys. Rev.* E **65** 011101
[54] Tan W Y 2000 *Stochastic Modeling of AIDS Epidemiology and HIV Pathogenesis* (Singapore: World Scientific)
[55] Fleming W H 1975 *Suppl. Adv. Appl. Probab.* **7** 100
[56] Hänggi P, Grabert H, Talkner P and Thomas H 1984 *Phys. Rev.* A **29** 371
[57] Baras F, Malek-Mansour M and Pearson J E 1996 *J. Chem. Phys.* **105** 8257
[58] Zhou D and Qian H 2011 arxiv.1101.2945
[59] Feller W 1957 *An Introduction to Probability Theory and Its Applications* vol 1, 2nd edn (New York: Wiley)
[60] O'Malley R E 2008 *SIAM Rev.* **50** 459
[61] Touchette H 2009 *Phys. Rep.* **478** 1
[62] McQuarrie D A 1967 *J. Appl. Probab.* **4** 413
[63] Ewens W J 2004 *Mathematical Population Genetics: I. Theoretical Introduction* 2nd edn (New York: Springer)
[64] Zhang Y, Ge H and Qian H 2010 arxiv.1011.2554
[65] Wang Z-J, Qian M and Qian H 2011 in preparation
[66] Kot M 2001 *Elements of Mathematical Ecology* (Cambridge: Cambridge University Press)
[67] Dercole F and Rinaldi S 2008 *Analysis of Evolutionary Processes: The adaptive dynamics approach and its applications* (Princeton, NJ: Princeton University Press)
[68] Li Y, Qian H and Yi Y 2008 *J. Chem. Phys.* **129** 154505
[69] Li Y, Qian H and Yi Y 2010 *J. Dyn. Diff. Eqns* **22** 491
[70] Qian H 2010 *J. Phys. Chem.* B **114** 16105
[71] Cushing J M 1991 *Math. Biosci.* **107** 47
[72] Risken H 1989 *The Fokker–Planck Equation: Methods of Solutions and Applications* 2nd edn (New York: Springer)
[73] Bortz A B, Kalos M H and Lebowitz J L 1975 *J. Comput. Phys.* **17** 10
[74] Laughlin R B, Pines D, Schmalian J, Stojković B P and Wolynes P G 2000 *Proc. Natl Acad. Sci. USA* **97** 32
[75] Arnold L 1998 *Random Dynamical Systems* (New York: Springer)