Introduction to dynamical large deviations of Markov processes

Hugo Touchette *

National Institute for Theoretical Physics (NITheP), Stellenbosch 7600, South Africa
Institute of Theoretical Physics, Department of Physics, University of Stellenbosch, Stellenbosch 7600, South Africa

HIGHLIGHTS

- Nonequilibrium systems driven in steady states are modelled by Markov processes.
- Observables of these systems are defined as time-integrated functionals of their states.
- Techniques based on spectral methods are presented to obtain the large deviation functions of these observables characterizing their fluctuations in the long-time limit.
- In many cases the large deviation spectral problem reduces to solving a Schrödinger-type equation.

ARTICLE INFO

ABSTRACT

These notes give a summary of techniques used in large deviation theory to study the fluctuations of time-additive quantities, called dynamical observables, defined in the context of Langevin-type equations, which model equilibrium and nonequilibrium processes driven by external forces and noise sources. These fluctuations are described by large deviation functions, obtained by solving a dominant eigenvalue problem similar to the problem of finding the ground state energy of quantum systems. This analogy is used to explain the differences that exist between the fluctuations of equilibrium and nonequilibrium processes. An example involving the Ornstein–Uhlenbeck process is worked out in detail to illustrate these methods. Exercises, at the end of the notes, also complement the theory.

1. Introduction

My aim in these notes is to give an introduction to large deviation techniques used to calculate the probability distribution of physical quantities, called *dynamical observables*, which are time-integrated functionals of Markov processes. These observables have come to play in the last years an important role in the context of nonequilibrium systems, and the recent discovery of fluctuation symmetries (also called fluctuation relations) generally satisfied by these systems. They also appear naturally when defining energy-like quantities, such as work and heat, in the context of noisy systems modelled by Markov processes, a field of research commonly referred to now as *stochastic thermodynamics* (see Seifert's contribution to this volume).

The results covered are known in large deviation theory, but are not widespread, certainly not in statistical physics, and cannot be found all put together in a single reference. The invitation to lecture at the FPSP School (to which I participated as

---

* Correspondence to: National Institute for Theoretical Physics (NITheP), Stellenbosch 7600, South Africa.
  E-mail address: htouchet@alum.mit.edu.

a student in 2005) offers a good opportunity to summarize them by appealing to what physics students know best: quantum mechanics.

The link with quantum mechanics is a natural one because all large deviation calculations related to time-additive functionals of Markov processes reduce, when one does not consider the low-noise limit, to an eigenvalue calculation, which is very close in spirit to finding the energy levels of a quantum system and, particularly, its ground state energy. For this reason, it is not surprising to see many techniques of quantum mechanics (such as the Bethe ansatz and the density matrix renormalization group) being applied to study the fluctuations of nonequilibrium processes. In a way, if you know quantum mechanics, you also know Markov processes.

Compared to quantum systems, however, there is a fundamental difference that arises when dealing with Markov processes, namely, that the operator or matrix that we must consider to study their steady state and fluctuations is, in general, not Hermitian. This raises many technical but important questions, which are precisely the questions that are hard to find in the literature, such as:

- What can we say in general about the spectrum of non-Hermitian operators?
- Under what conditions is that spectrum real?
- Which function space must we use to solve the eigenvalue problem?
- What are the boundary conditions for the eigenfunctions?
- What is their normalization?
- What distinguishes equilibrium from nonequilibrium processes in terms of fluctuations, dynamical or otherwise?

The fact that we have to deal with non-Hermitian operators is the main reason why studying nonequilibrium processes is more challenging than studying equilibrium processes, but at the same time so much more interesting. We are far from completely understanding nonequilibrium systems, and I do not believe, in fact, that there is or will be a general theory of these systems (not at least in the way many physicists picture it). But I do believe that having a good mathematical grounding in the subject — at least as good as the one we expect in quantum mechanics or any modern physics subject — is necessary for making progress.

Obviously, such a pedagogical goal cannot be achieved in 15 pages of notes (a book is in progress), so I decided to focus here on Langevin-type equations modelling noisy systems driven by external forces and baths, sacrificing mathematical rigor, as always, for clarity. With this, one should keep in mind that all the results discussed can be applied more generally (sometimes with minor modifications) to a large class of Markov processes, including Markov chains and Markov jump processes. Some references and exercises on those are given in the text.

2. Markov processes

2.1. Stochastic differential equations

We consider throughout these notes Markov processes defined by the following **stochastic differential equation (SDE)**:

$$dX_t = F(X_t)dt + \sigma \, dW_t, \tag{1}$$

where

- X_t is a vector in $\mathbb{R}^n$ representing the **state** of the system at the time t. Note that we do not use bold symbols to represent vectors.
- $F : \mathbb{R}^n \to \mathbb{R}^n$ is a vector field that drives the deterministic evolution of X_t when there is no noise ($\sigma = 0$). We call this function the **force** or the **drift** of the system. It can explicitly depends on time, for example, by having $F(X_t, t)$, but we do not consider this possibility here.
- W_t is a vector of independent **Brownian** or **Wiener motions**, whose increments dW_t are Gaussian-distributed with zero mean and variance dt. In many applications, W_t is taken to have as many components as X_t, so that $W_t \in \mathbb{R}^n$, but it is also possible to have $W_t \in \mathbb{R}^m$ with $m > n$ or $m < n$. For simplicity, we consider here $m = n$.
- σ is the **noise matrix**, assumed here not to depend on X_t to simplify the discussion. That matrix has dimensions $n \times m$ to match the dimensions of X_t and W_t. For $W_t \in \mathbb{R}^n$, it is an $n \times n$ matrix.

The SDE above governing the evolution of X_t (as an infinitesimal difference equation) is called in mathematics an **Itô SDE**. For the purpose of these notes, we can consider X_t to be defined equivalently by a noisy **ordinary differential equation (ODE)** having the more common form

$$\dot{X}_t = F(X_t) + \sigma \xi_t, \tag{2}$$

where ξ_t is an n-dimensional **Gaussian white noise** corresponding formally to the time-derivative of W_t and defined by the properties

$$\langle \xi_t \rangle = 0, \qquad \langle \xi_t^i \xi_{t'}^j \rangle = \delta_{ij}\delta(t - t'), \tag{3}$$

where i and j denote specific components of $\xi(t)$ and $\langle \cdot \rangle$ denotes the expectation or mean (with respect to whatever random variable or process found in the brackets).

I assume in these notes that the reader knows that Brownian motion is nowhere differentiable, which explains why it is preferable to express a diffusion process in the Itô rather than in the noisy ODE form. I also assume some familiarity with stochastic calculus, though this is not essential, as long as one is aware that the rules for calculating derivatives and integrals of stochastic processes differ slightly from those of normal calculus. For background references on SDEs and stochastic calculus, see the Further reading section. Some important differences between stochastic and normal calculus will be pointed out along the notes.

2.2. State distribution and generator

The first natural problem to consider when studying a Markov process is to determine its probability density $P(X_t = x) = p(x, t)$, starting from an initial density $p(x, 0)$ that represents the distribution with which we sample the initial state X_0. For the general SDE shown in (1), that density is known to be given by the **Fokker–Planck equation**,

$$\partial_t p(x, t) = -\nabla \cdot (F(x)p(x, t)) + \frac{1}{2} \nabla \cdot D\nabla p(x, t), \tag{4}$$

which involves the symmetric matrix $D = \sigma\sigma^T$, called the **covariance matrix**. To emphasize that this partial differential equation (PDE) is linear, we can express it as

$$\partial_t p(x, t) = L^\dagger p(x, t), \tag{5}$$

where

$$L^\dagger = -\nabla \cdot F + \frac{1}{2} \nabla \cdot D\nabla \tag{6}$$

is a linear differential operator called the **Fokker–Planck generator**. The analogy with the Schrödinger equation of quantum mechanics should be obvious.

The Fokker–Planck equation can also be rewritten, as is well known, as a conservative equation involving the (vector) probability current

$$J_t(x) = F(x)p(x, t) - \frac{D}{2}\nabla p(x, t), \tag{7}$$

called the **Fokker–Planck current**, as

$$\partial_t p(x, t) + \nabla \cdot J_t(x) = 0. \tag{8}$$

This form is useful for interpreting the nature of the **stationary probability density** $p_s(x)$ satisfying

$$L^\dagger p_s = 0 \tag{9}$$

or equivalently $\nabla \cdot J_s = 0$, where J_s is the **stationary current** associated with p_s. This shows that $p_s(x)$ is the eigenfunction of $L^\dagger$ with eigenvalue 0. When the process X_t is **ergodic**, $p(x, t) \rightarrow p_s(x)$ from any initial density as $t \rightarrow \infty$.

For the next sections, it is useful to note that the Fokker–Planck equation also determines the evolution of expectations of X_t having the general form

$$\langle f(X_t) \rangle = \int p(x, t)f(x)dx, \tag{10}$$

where f is any smooth function of the process. Indeed, it can be proved using the natural inner product defined by the expectation that

$$\partial_t \langle f(X_t) \rangle = \langle (Lf)(X_t) \rangle, \tag{11}$$

where L is the adjoint of $L^\dagger$ (see Appendix A). This operator, which is simply called the **generator** of X_t, acts on the function f rather than the density p, as is explicit from the notation above, and is equal here to

$$L = F \cdot \nabla + \frac{1}{2} \nabla \cdot D\nabla. \tag{12}$$

This result is much less known in physics than the Fokker–Planck equation, though it is as important. It corresponds, in a way, to the Heisenberg picture of quantum mechanics which describes the evolution of observables (after taking their expectation), as opposed to the Schrödinger picture which describes the evolution of probabilities. The generator will become useful for treating large deviations.

Table 1

Comparison of equilibrium and nonequilibrium Markov processes in steady states.

Equilibrium	Nonequilibrium
Reversible	Non-reversible
$J_s(x) = 0$	$J_s(x) \neq 0$ although $\nabla \cdot J_s = 0$
Spectrum of L real	Spectrum of L generally complex

2.3. Examples

There are many standard SDEs used in physics to model noisy systems driven by forces, external reservoirs, heat baths and noise sources in general. The following is a representative list — for more examples, see the reading list in Section 2.6.

- **Kramers or underdamped Langevin equation:**

$$dq_t = \frac{p_t}{m} dt$$
$$dp_t = \left(-\nabla V(q_t) + \phi_t - \Gamma \frac{p_t}{m}\right) dt + \sqrt{2\Gamma/\beta}\, dW_t, \tag{13}$$

where q_t is the position, p_t the momentum, $V(q)$ is a potential, Γ is the friction, ϕ_t is an external force, and β is the inverse temperature of the thermal noise. The noise acts as a force in Newton's equation and so only affects the momentum, not the position.

- **Overdamped Langevin equation:**

$$dq_t = \Gamma^{-1}(-\nabla V + \phi_t)dt + \sqrt{2\Gamma^{-1}/\beta}\, dW_t. \tag{14}$$

This is an SDE for the position obtained by taking the overdamped ($m \to 0$) limit of Kramers equation.

- **Gradient SDEs:**

$$dX_t = -\nabla U(X_t)dt + \sigma\, dW_t \tag{15}$$

with σ proportional to the identity matrix, that is, $\sigma = \varepsilon \mathbb{1}$. The stationary density of this SDE is the Gibbs distribution

$$p_s(x) = c\, e^{-2U(x)/\varepsilon^2}, \tag{16}$$

where c is a normalization constant (see Exercise 3).

- **Linear diffusions:**

$$dX_t = -MX_t dt + \sigma\, dW_t, \tag{17}$$

where M is an $n \times n$ matrix assumed to be positive definite (positive eigenvalues) in order for X_t to have a stationary density (see Exercise 4).

- **Ornstein–Uhlenbeck process:**

$$dX_t = -\gamma X_t dt + \sigma\, dW_t \tag{18}$$

with $X_t \in \mathbb{R}$ and $W_t \in \mathbb{R}$. This is obviously a gradient SDE with quadratic potential $U(x) = \gamma x^2/2$ having a Gibbs stationary distribution with $\sigma = \varepsilon$.

2.4. Equilibrium versus nonequilibrium processes

The distinction between equilibrium and nonequilibrium systems in the context of stochastic processes is based on the notion of time reversibility or, equivalently, detailed balance. It would take too much space to fully explain these notions, so we only summarize them. The idea, essentially, is that a process is an **equilibrium process** if the probability of any given trajectory is the same as the probability of that trajectory reversed in time. If that is not the case, then the process is **nonequilibrium**.

For Markov processes, it can be proved that this definition of equilibrium in terms of "forward" and "backward" trajectories is equivalent to the notion of detailed balance, which is itself related (in most cases) to having a vanishing probability current in the Fokker–Planck equation (see Exercise 2). Moreover, all of these notions are related in general to the eigenvalues of the generator L.

We summarize these connections in Table 1, assuming that the process X_t is stationary, that is to say, it has a stationary distribution and its initial condition X_0 is drawn according to that distribution.

Table 2
Comparison between Markov processes and quantum mechanics.

	Markov	Quantum
State	X_t	$|\psi(t)\rangle$ or $\psi(x, t)$
Distribution	$p(x, t)$	$|\psi(x, t)|^2$
Evolution	Fokker–Planck	Schrödinger
Generator	L	H (Hamiltonian)
Propagator	$U(t) = e^{L^\dagger t}$	$U(t) = e^{-iHt/\hbar}$
Inner product	$\langle p, f \rangle$	$\langle \psi, \psi \rangle = \langle \psi | \psi \rangle$
Duality	$\langle p, Af \rangle = \langle A^\dagger p, f \rangle$	$\langle \psi, A\psi \rangle = \langle A\psi, \psi \rangle = \langle \psi | A | \psi \rangle$
Self-adjoint?	Not necessarily	Always

2.5. Comparison with quantum mechanics

It is useful at this point to reflect on the structure of Markov diffusions, especially the linear form of the Fokker–Planck equation and its generator, by comparing it to what we know about the evolution of quantum systems — see Table 2.

From this table, you should see that the theory of Markov processes has much in common with quantum mechanics, as announced in the introduction. Both are linear theories of evolution for a "vector" corresponding to the probability density $p(x, t)$ for Markov processes and the wavefunction $\psi(x, t)$ for quantum systems. As a result, the central object in both theories is the generator of that evolution, which is the Fokker–Planck operator $L^\dagger$ for Markov processes and the Hamiltonian H for quantum systems.

The main difference between the two sides is that, unlike the Hamiltonian of closed quantum systems, the generator of Markov processes is not always self-adjoint, a property deeply related to whether we are dealing with an equilibrium or nonequilibrium process. This means, in practice, that we have to be careful when diagonalizing matrices or operators and dealing, in general, with eigenvectors or eigenfunctions (see Appendix B).

2.6. Further reading

- SDEs and Langevin equations: [1]. For more technical yet readable presentations, see [2] and especially [3].
- Stochastic calculus: [1,4] for the theory and [5] for simulations.
- Theory of Markov processes focusing more on Markov chains and jump processes (also called continuous-time Markov chains): A very good reference, for its scope, clarity and number of exercises, is [6]. Though a maths textbook, it should be compulsory reading for any serious students in statistical physics. For a leaner textbook, see [2].
- Fokker–Planck equation: [7].
- Stochastic processes with applications in physics: [8,9]. In my opinion, statistical physics is in real need of a modern reference on that front.

3. Dynamical observables

The study of Markov processes in physics has focused a lot in the past on the statistical properties of the state X_t — its distribution, its average, in addition to its variance and covariance, which can be related to diffusion and transport coefficients [10]. In the last 20 years or so, researchers have also become interested in the statistics of *time-integrated quantities involving* X_t. An example is the **mechanical work** done by the force F on X_t over a time interval $[0, T]$, as calculated by

$$W_T = \int_0^T F(X_t) \circ dX_t, \tag{19}$$

where the circle $\circ$ indicates that the integral is to be calculated using the midpoint Riemann integral rule, also called the **Stratonovich convention**.[1] This random variable depends obviously not only on the state X_t at time t, but on the whole trajectory of this process between $t = 0$ and $t = T$. Thus, for different (random) trajectories, one typically gets different (random) work values. For this reason, W_T is often called an **additive functional** of the process or, more physically, a **dynamical observable**.

We list next other physical examples:

- **Potential energy:** The change in time of potential energy can be written as

$$\Delta U_T = U(X_T) - U(X_0) = \int_0^T \nabla U(X_t) \circ dX_t. \tag{20}$$

---

[1] The work is a scalar quantity, so the product $F(X_t) \circ dX_t$ is also a scalar product.

This holds whether the SDE of X_t is gradient or not, but only if the integral is interpreted with the Stratonovich convention, which follows the standard rules of calculus. In the Itô convention, corresponding to the left-point Riemann rule, there would be additional terms in the integral, coming from Itô's calculus [1], which are unphysical. This explains why the work W_T must be defined in the Stratonovich convention: for a potential force, the work is the change of potential.

- **Empirical distribution:**

$$\rho_T(x) = \frac{1}{T} \int_0^T \delta(X_t - x)\, dt. \tag{21}$$

This random function represents the fraction of time in $[0, T]$ that the process X_t "takes" the value x. For an ergodic process, it converges to the stationary density p_s for almost all trajectories. Mathematically, we say that ρ_T converges in probability to p_s as $T \to \infty$.

- **Empirical current:**

$$J_T(x) = \frac{1}{T} \int_0^T \delta(X_t - x) \circ dX_t. \tag{22}$$

This random field represents intuitively the local mean "velocity" of X_t at x, if we view dX_t formally as $\dot{X}_t dt$. It is more aptly called the empirical current because it represents a flow at each point x that converges in probability to the stationary Fokker–Planck current $J_s(x)$ as $T \to \infty$ (see Exercise 8). The Stratonovich rule is also important for this convergence.

- **Entropy production:**

$$\Sigma_T = 2 \int_0^T (D^{-1} F(X_t)) \circ dX_t. \tag{23}$$

This is an important quantity in the theory of nonequilibrium systems, related to the non-reversibility (or nonequilibrium nature) of X_t and, as is clear from its definition, also to the work W_T [11]. The factor 2 is there to obtain the normal thermodynamic relation $\Sigma_T = \beta \Delta V_T$ for the overdamped Langevin equation (14) (check this).

From these examples, it is natural to consider a general class of dynamical observables having the form

$$A_T = \frac{1}{T} \int_0^T f(X_t)\, dt + \frac{1}{T} \int_0^T g(X_t) \circ dX_t, \tag{24}$$

where f (scalar) and g (vector) are two arbitrary functions that depend on the system and physical quantity considered, and $\circ$ denotes, as before, the Stratonovich (scalar) product. This choice of convention is actually not important − we could use the Itô convention instead with a slight modification of the results that will come next. What is more important is the factor $1/T$ which is there for two reasons: first, to make A_T intensive in time and, second, to guarantee, following the examples of the empirical density and empirical current, that A_T converges in probability to a constant (its mean) in the long-time limit, $T \to \infty$.

4. Large deviations

4.1. Large deviation principle

Our goal now is to study the probability distribution $P(A_T = a)$ of a given dynamical observable A_T and Markov process X_t. In general, it is very difficult to obtain that distribution exactly. In many cases, however, it is known that it has the following general asymptotic form:

$$P(A_T = a) \approx e^{-TI(a)} \tag{25}$$

in the limit of large observation time T. The meaning of this approximation, which is called the **large deviation principle (LDP)**, is that the dominant contribution of $P(A_T = a)$ is a decaying exponential, and so that any corrections to that contribution is sub-exponential in T.[2] The exponent or **rate function** $I(a)$ is therefore the essential information that we need to find in order to characterize the fluctuations of A_T. In particular,

- $I(a) \geq 0$ for all a, so $P(A_T = a)$ decays exponentially fast with T, except for values a such that $I(a) = 0$.
- For Markov processes, there is usually only one point a^* where $I(a^*) = 0$, so if $P(A_T = a)$ decays where $I(a) > 0$, it must concentrate on a^* by conservation of probability.

---

[2] See Appendix B of [12] or Sect. 1.2 of [13] for the mathematical definition of the LDP. The loose definition given here assumes that $P(A_T = a)$ is a probability density.

- The zero of $I(a)$ is a reflection of the Law of Large Numbers: it gives the most probable or **typical value** of A_T in the long-time limit and coincides with the mean of A_T. In an experiment where the work W_T per unit time would be measured, for example, one would see that most trajectories do work close to its mean value $\langle W_T \rangle$. Only rarely would we see trajectories that require more or less work than this average.
- The rate function $I(a)$ characterizes the rare fluctuations of A_T around this typical value. In general, it is not a parabola, so fluctuations of dynamical observables are in general not Gaussian. This is the most important point about large deviation theory — the fact that it characterizes fluctuations beyond Gaussian fluctuations and so beyond the Central Limit Theorem.

These properties will become clear once we start calculating rate functions for specific processes and observables.

We describe next the most common way of obtaining rate functions using a result of large deviation theory known as the Gärtner–Ellis Theorem, which essentially boils down, for additive functionals of Markov processes, to calculating a dominant eigenvalue of a linear operator. This is not an easy task to carry out in many applications, but it is definitely simpler than calculating the exact distribution of A_T.

4.2. Spectral problem

The **Gärtner–Ellis Theorem** is based on the following function:

$$\lambda(k) = \lim_{T \to \infty} \frac{1}{T} \ln \langle e^{TkA_T} \rangle, \tag{26}$$

known as the **scaled cumulant generating function (SCGF)**. The theorem says, in its simplified version, that if $\lambda(k)$ exists for $k \in \mathbb{R}$ and is differentiable in k, then

(1) A_T satisfies a large deviation principle, so its distribution has the scaling form (25);
(2) Its rate function $I(a)$ is the **Legendre–Fenchel transform** of $\lambda(k)$:

$$I(a) = \max_{k \in \mathbb{R}} \{ka - \lambda(k)\}. \tag{27}$$

The question now is, how do we calculate $\lambda(k)$?

For this, we can use another result of probability theory, known as the Feynman–Kac formula, to cast the evolution of the generating function of A_T as a linear PDE similar to the Fokker–Planck equation, which involves some linear operator $\mathcal{L}_k$, called the **tilted generator**, and then study the asymptotic evolution of that PDE to realize that it is dominated by the dominated eigenvalue of $\mathcal{L}_k$. These steps are presented in Appendix C and lead to the main result of these notes, namely,

$$\lambda(k) = \zeta_{\max}(\mathcal{L}_k), \tag{28}$$

where $\zeta_{\max}(\mathcal{L}_k)$ denotes the dominant eigenvalue of $\mathcal{L}_k$. To be more precise, $\lambda(k)$ *is equal to* $\zeta_{\max}(\mathcal{L}_k)$ *whenever* $\lambda(k)$ *exists as a SCGF*. This is an important precision (see Exercise 17). For the SDE (1) and the observable A_T defined in (24), the tilted generator is explicitly given by

$$\mathcal{L}_k = F \cdot (\nabla + kg) + \frac{1}{2}(\nabla + kg) \cdot D(\nabla + kg) + kf, \tag{29}$$

where f and g are the functions entering in A_T (see Appendix C and Exercise 9). Note that $\mathcal{L}_{k=0} = L$ and so $\lambda(0) = \zeta_{\max}(L) = 0$.

This result is a PDE generalization of the Perron–Frobenius Theorem about positive matrices, guaranteeing that $\lambda(k)$ is real, and applies essentially whenever the spectrum $\mathcal{L}_k$ is gapped. The resulting dominant eigenvalue problem is similar to a quantum eigenvalue problem, except that $\mathcal{L}_k$ is not in general a Hermitian operator. This makes the calculation of $\lambda(k)$ more complicated.

In fact, compared to quantum mechanics, we must now not only consider the eigenvalue problem

$$\mathcal{L}_k r_k(x) = \lambda(k) r_k(x), \tag{30}$$

where $r_k(x)$ is the "right" eigenfunction associated with the dominant eigenvalue $\lambda(k)$. We must also solve in parallel the dual eigenvalue problem

$$\mathcal{L}_k^\dagger l_k(x) = \lambda(k) l_k(x), \tag{31}$$

where $l_k(x)$ is the corresponding "left" eigenfunction, by requiring overall the following boundary condition:

$$r_k(x) l_k(x) \xrightarrow{|x| \to \infty} 0. \tag{32}$$

The reason for this, in short, is that the duality between $\mathcal{L}_k$ and $\mathcal{L}_k^\dagger$ is equivalent to performing integration by parts (Appendix A) and $r_k(x) l_k(x)$ is the boundary term in that integration that must vanish at infinity. Therefore, we see that, contrary to quantum mechanics, we cannot just solve the direct eigenvalue problem for $\mathcal{L}_k$ by requiring that $r_k(x)$ alone decays to 0 at

infinity — we must consider the direct and dual eigenvalue problems at the same time with (32) to find the correct dominant eigenvalue.

This duality also determines the correct normalization to use for the eigenfunctions, which turns out to be

$$\int r_k(x)\,l_k(x)\,dx = 1 \tag{33}$$

(see Appendix B). For convenience, we also impose

$$\int l_k(x)\,dx = 1. \tag{34}$$

Note that $l_{k=0} = p_s$ and $r_{k=0} = 1$ (see Exercise 10), so both normalization conditions reduce to $\int p_s(x)dx = 1$ for $k = 0$.

4.3. Symmetrization

The tilted generator $\mathcal{L}_k$ is known, in many cases, to have a real spectrum even though it is not Hermitian.[3] This happens, for example, when dealing with gradient SDEs and observables such that $g = 0$. In this case, it should be possible to transform $\mathcal{L}_k$ to a Hermitian operator $\mathcal{H}_k$ in an unitary way (so as to preserve the spectrum) and then work only with $\mathcal{H}_k$ using techniques from quantum mechanics to find $\lambda(k)$.

For the case of gradient SDEs mentioned, this is indeed possible and the unitary transformation or **symmetrization** that we must use is given by

$$\mathcal{H}_k = p_s^{1/2}\mathcal{L}_k p_s^{-1/2}, \tag{35}$$

where $p_s(x)$ is the Gibbs stationary distribution of the SDE shown in Eq. (16) (see also Exercise 3). Note that this is an operator transformation: when applied to a function ϕ, $\mathcal{H}_k$ acts as follows:

$$\mathcal{H}_k\phi = p_s^{1/2}(\mathcal{L}_k p_s^{-1/2}\phi), \tag{36}$$

so that $\mathcal{L}_k$ is applied to the product $p_s^{-1/2}\phi$. The resulting function is then multiplied by $p_s^{1/2}$.

For a gradient SDE defined in (15) and an observable A_T such that $f \neq 0$ but $g = 0$, it is not difficult to see by direct calculation (see Exercise 12) that $\mathcal{H}_k$ has the form

$$\mathcal{H}_k = \frac{\varepsilon^2}{2}\Delta - V_k, \tag{37}$$

where $\Delta = \nabla^2$ is the Laplacian, ε is the noise amplitude of the SDE, and

$$V_k(x) = \frac{|\nabla U(x)|^2}{2\varepsilon^2} - \frac{\Delta U(x)}{2} - kf(x) \tag{38}$$

is an effective quantum-like potential. Our eigenvalue problem thus reduces to

$$\mathcal{H}_k\psi_k = \lambda(k)\psi_k, \tag{39}$$

where ψ_k is the eigenfunction of $\mathcal{H}_k$ associated with $\lambda(k)$.

We recognize in this equation the time-independent Schrödinger equation, up to a minus sign, with $\varepsilon^2 = \hbar^2/m$ and potential V_k. The sign difference means that what we find as the dominant (largest) eigenvalue of $\mathcal{H}_k$ corresponds to the ground state (smallest) energy of $-\mathcal{H}_k$.

The eigenvalue $\lambda(k)$ is the same for $\mathcal{L}_k$ and $\mathcal{H}_k$ since these two operators are unitarily related. The eigenfunctions ψ_k of $\mathcal{H}_k$, however, are different from the eigenfunctions r_k of $\mathcal{L}_k$. In general, the two are related together (see Exercise 13) by

$$\psi_k(x) = p_s(x)^{1/2}r_k(x). \tag{40}$$

Moreover, it can be verified that

$$\psi_k(x) = p_s(x)^{-1/2}l_k(x). \tag{41}$$

This is interesting because it implies that the boundary condition (32) that we have for the full eigenvalue problem now reduces to

$$\psi(x)^2 \xrightarrow{|x|\to\infty} 0, \tag{42}$$

---

[3] Being Hermitian is only a sufficient condition for having a real spectrum, not a necessary condition.

Table 3
Spectral problem for equilibrium and nonequilibrium processes.

X_t	L symmetrizable?	g	$\mathcal{L}_k$ symmetrizable?
Reversible	Yes	0	Yes
		Gradient $g = \nabla\varphi$	Yes
		Non-gradient	No
Non-reversible	No (generally)	Any	No (generally)

which is the normal quantum boundary condition (modulo a complex conjugate), while the normalization condition in (33) reduces to

$$\int \psi(x)^2 \, dx = 1. \tag{43}$$

In a more practical way, it also implies that we can now focus on only one eigenvalue problem with natural (quantum) boundary conditions imposed only on ψ_k.

This symmetrization method greatly simplifies, obviously, the calculation of large deviation functions. A natural question is, to which class of Markov processes can it be applied to besides gradient SDEs? Table 3 gives some answers. The basic idea is that, if we consider an equilibrium process and an "equilibrium-type" observable characterized by $g = 0$ or g gradient, then $\mathcal{L}_k$ can be symmetrized. If X_t is a nonequilibrium process, then $\mathcal{L}_k$ cannot be symmetrized in general because L itself cannot be symmetrized. Finally, if X_t is reversible but g is not gradient, then $\mathcal{L}_k$ generally cannot be symmetrized because the fluctuations of the observable that we consider are essentially created in a nonequilibrium way (see Exercise 21).

4.4. Example: Ornstein–Uhlenbeck process

We close these notes by applying the material of the previous sections to a specific example involving the Ornstein–Uhlenbeck process defined in (18). Our goal is to find the rate function $I(a)$ characterizing the fluctuations of the following observable:

$$A_T = \frac{1}{T} \int_0^T X_t \, dt, \tag{44}$$

which represents mathematically the area under the paths of the process, and which can be related physically to the mechanical work performed by a laser tweezer on a Brownian particle immersed in water [14].

To obtain the rate function, we first write down the tilted generator associated with this process and observable, noting in this case that $f(x) = x$ and $g = 0$:

$$\mathcal{L}_k = L + kf = -\gamma x \frac{d}{dx} + \frac{\varepsilon^2}{2} \frac{d^2}{dx^2} + kx, \qquad k \in \mathbb{R}. \tag{45}$$

This operator is not Hermitian, but since X_t is gradient (any SDE on $\mathbb{R}$ is gradient) and $g = 0$, it can be symmetrized with its (Gaussian) Gibbs distribution

$$p_s(x) = \sqrt{\frac{\gamma}{\pi \varepsilon^2}} \, e^{-\gamma x^2/\varepsilon^2} \propto e^{-2U(x)/\varepsilon^2} \tag{46}$$

to

$$\mathcal{H}_k = \frac{\varepsilon^2}{2} \frac{d^2}{dx^2} - \frac{\gamma^2 x^2}{2\varepsilon^2} + \frac{\gamma}{2} + kx. \tag{47}$$

We recognize the Hamiltonian of the 1D quantum harmonic oscillator if we multiply by -1 and shift the space to $x \to x + \varepsilon^2 k/\gamma^2$. The SCGF, corresponding to minus the known ground state of the quantum oscillator which does not depend on the shift (see any textbook on quantum mechanics), is therefore found to be

$$\lambda(k) = \frac{\varepsilon^2 k^2}{2\gamma^2}. \tag{48}$$

This exists and is differentiable for all $k \in \mathbb{R}$, so we can the apply the Legendre–Fenchel transform (27), which in this case reduces to a simple Legendre transform (why?), to finally obtain

$$I(a) = \frac{\gamma^2 a^2}{2\varepsilon^2}. \tag{49}$$

This shows that the fluctuations of A_T are Gaussian around the typical value $A_T = 0$, a result consistent with the fact that linear integrals of Gaussian processes are Gaussian.

Although we do not need the eigenvectors $r_k(x)$ and $l_k(x)$ in the calculation of the SCGF and rate function, it is instructive to compute them. For the quantum oscillator, it is well known that the eigenfunctions are given in terms of Hermite polynomials. Considering only the ground state, we find here

$$\psi_k(x) = \left(\frac{\gamma}{\pi \varepsilon^2}\right)^{1/4} \exp\left(-\frac{\gamma\left(x - \varepsilon^2 k/\gamma^2\right)^2}{2\varepsilon^2}\right). \tag{50}$$

Using the transformations (40) and (41), and applying the normalization conditions (33) and (34), we then find

$$r_k(x) = \exp\left(\frac{kx}{\gamma} - \frac{3\varepsilon^2 k^2}{4\gamma^3}\right) \tag{51}$$

and

$$l_k(x) = \sqrt{\frac{\gamma}{\pi \varepsilon^2}} \exp\left(-\frac{\gamma\left(2x - \varepsilon^2 k/\gamma^2\right)^2}{4\varepsilon^2}\right). \tag{52}$$

This clearly shows that $r_k(x)$ or $l_k(x)$ do not decay to zero as $x \to \pm\infty$ — it is their product again that does so, in agreement with (32). Finally, note that $r_{k=0}(x) = 1$ while $l_{k=0}(x) = p_s(x)$, as pointed out before (see also Exercise 10).

The exercises found at the end of these notes extend this simple calculation to other equilibrium-type observables, before slowly moving into the realm of nonequilibrium processes. For more information about nonequilibrium large deviations, and the current research done on this topic, see the pointer references next.

4.5. Further reading

- Mathematical theory of large deviations: [13,15].
- Applications of large deviations in statistical physics: [12,16].
- Quantum approach to large deviations: [17,18].
- Related approach based on the thermodynamics of trajectories: [19–21].
- Stochastic thermodynamics: [22].
- Fluctuations of interacting particle systems modelling particle and energy transport: [23–25].
- Fluctuation relations and symmetries: [11,26].
- Theory of fluctuation processes explaining, with modified SDEs, how large deviations are created in time: [27–29].
- Entropy production: [11].
- Fluctuations of empirical density and current (level 2.5 of large deviations): [30].
- Large deviations for open quantum systems: [31].
- Low-noise large deviations: [12].
- Large deviation simulations: [32,33].

Acknowledgments

Appendix A. Dual spaces for Markov processes

The expectation

$$\langle f(X_t) \rangle = \int p(x, t) f(x) \, dx \tag{A.1}$$

of a function f of X_t defines the following natural **scalar** or **inner product** in the theory of Markov processes:

$$\langle p, f \rangle = \int p(x) f(x) \, dx, \tag{A.2}$$

which connects the space of normalized probability densities and the space of functions of X_t, also called **test functions** or **observables**. (See Table 2 for a comparison of this inner product and the one used in quantum mechanics.)

Applying an operator either on f or on p leads us to define the notion of **dual** or **adjoint operator**: if L acts on f, then its adjoint $L^\dagger$ acts on p according to

$$\langle p, Lf \rangle = \langle L^\dagger p, f \rangle. \tag{A.3}$$

Because the integral defining the inner product is performed with dx, the Lebesgue measure, we say that $L^\dagger$ is the adjoint of L with respect to the Lebesgue measure.

For differential operators, the duality between L and $L^\dagger$ simply corresponds to performing integration by parts:

$$\langle p, \frac{df}{dx} \rangle = \int p\,df = pf|_{\text{boundary}} - \int f\,dp, \tag{A.4}$$

which leads to $(\nabla)^\dagger = -\nabla\cdot$ for the gradient and $(\Delta)^\dagger = \Delta$ for the Laplacian, if we choose p and f such that the boundary term (usually at infinity) vanishes. The Fokker–Planck equation also follows from this duality by noting from (11) that

$$\partial_t \langle f(X_t) \rangle = \langle p_t, Lf \rangle = \langle L^\dagger p_t, f \rangle. \tag{A.5}$$

Since this holds for any test functions, we must then have (4).

Appendix B. Non-Hermitian operators

Non-Hermitian or non-self-adjoint operators (we do not make a difference between the two here) can be diagonalized in the same way as Hermitian operators in quantum mechanics — just think about symmetric versus non-symmetric matrices. The only difference is that, being non-Hermitian (essentially, non-symmetric), we have to distinguish between "left" and "right" eigenfunctions (or eigenvectors) to build their spectral decomposition.

To be precise, let us consider a linear differential operator L and consider the eigenvalue problem

$$Lv(x) = \lambda v(x), \tag{B.1}$$

which can be solved for a set of eigenvalues λ, called the **spectrum** of L, and their corresponding eigenfunctions $v(x)$. We do not put indices to these objects as is commonly done — their numbering or labelling is implicit. If L is not self-adjoint, the dual eigenvalue problem

$$L^\dagger u(x) = \beta u(x) \tag{B.2}$$

has a spectrum that is the complex conjugate of the spectrum of L, that is, $\lambda = \beta^*$, but will in general have a completely different set of eigenfunctions $u(x)$.

As in quantum mechanics, eigenfunctions u or v associated with distinct eigenvalues are orthogonal, so that

$$\langle u_i, v_j \rangle = \int u_i^*(x) v_j(x)\,dx = \delta_{ij} \tag{B.3}$$

by properly normalizing them. Moreover, it can be shown that

$$\delta(x - x') = \sum_i u_i^*(x) v_i(x'), \tag{B.4}$$

a property known in quantum mechanics as the **completeness relation**. As a result, we see that the set of eigenfunctions u and v form a complete basis onto which any function can be decomposed. For an application of this decomposition for solving the Fokker–Planck equation, see Sec. 5.4 of [7]; for its application to our problem of finding large deviation functions, see the next appendix on the Feynman–Kac formula.

If L is self-adjoint, then $u = v$ and we recover the usual complete basis of quantum mechanics. Moreover, if L is a matrix, then the dual eigenvalue problem (B.2) is equivalent to

$$u^\dagger L = \beta^\dagger u^\dagger = \lambda u^\dagger, \tag{B.5}$$

where $u^\dagger$ is now seen as a row vector multiplying L. In this sense, it is common to call u the **left eigenvector** of L and v the **right eigenvector** of L.

Appendix C. Feynman–Kac formula

The linear structure of the Fokker–Planck equation (5) means that we can write down the time-dependent density $p(x, t)$ as

$$p(x, t) = U(t)p(x, 0), \tag{C.1}$$

where $U(t) = e^{tL^\dagger}$ is an operator acting on the initial density, called the **propagator**.[4] This operator is well known in quantum mechanics and leads to what we call a **semi-group** structure for the evolution of $p(x, t)$ (or the wavefunction) as a result of

---

[4] This applies to homogeneous SDEs with time-independent drift. For a time-dependent drift, the generator is formally the time-ordered exponential of the time-dependent generator.

the fact that $U(t) = U(s)U(s')$ for $t = s + s'$. For stochastic processes, this semi-group property *is* the Markov property and $L^\dagger$ is simply the generator of that semi-group.

The probability density of X_t or the wavefunction of a quantum system are not the only objects whose evolution forms a semi-group. The expectation $\langle f(X_t) \rangle$ also does, since its evolution is linear, its generator being L.

In the 1940s, Mark Kac (pronounced *khats*) showed that the general exponential functional of a Markov process X_t defined as

$$G(x, t) = \left\langle e^{\int_0^t c(X_s)ds} \right\rangle_x \tag{C.2}$$

also has, amazingly, a semi-group structure, provided that the function c is smooth enough and such that the expectation exists. Here the notation $\langle \cdot \rangle_x$ means that X_t is started at x, that is, $X_0 = x$ with probability 1. Specifically, he showed that

$$\partial_t G(x, t) = \mathcal{L}_c G(x, t), \tag{C.3}$$

where $\mathcal{L}_c = L + c$, L being the generator of X_t. This linear PDE, coupled with the initial condition $G(x, 0) = \langle e^0 \rangle_x = 1$, is what is called the **Feynman–Kac formula**.

There are many derivations of that formula, some of which based on analogies with quantum mechanics [18]. The simplest I could find is very similar to what we call in physics the Kramers–Moyal expansion (see [10]) and proceeds by considering $G(x, t + dt)$ to write

$$G(x, t + dt) = e^{c(x)dt} \left\langle e^{\int_{dt}^{t+dt} c(X_s)ds} \right\rangle_x. \tag{C.4}$$

The initial condition for the expectation in this formula does not match the start of the integral, so we should propagate X_0 to X_{dt} with the SDE (1) to obtain

$$G(x, t + dt) = e^{c(x)dt} \int p(\xi) G(x + \xi, t) d\xi, \tag{C.5}$$

where $p(\xi)$ is the probability density of the Gaussian increment ξ that takes us from $X_0 = x$ to $X_{dt} = x + \xi$. By Taylor-expanding $G(x + \xi, t)$ to second order around x, and taking the expectation with respect to ξ, we finally arrive at the correct PDE.

To connect this result to our goal of calculating large deviations, we only need to notice that the generating function $\langle e^{TkA_T} \rangle$ entering in the definition of the SCGF $\lambda(k)$ in Eq. (26) is a particular case of Kac's functional, at least for $g = 0$, corresponding to $c(x) = kf(x)$ so that $\mathcal{L}_c = \mathcal{L}_k$ with a slight abuse of notations. Therefore, this generating function satisfies the Feynman–Kac equation, which enables us to write

$$G(x, t) = \langle e^{tkA_t} \rangle_x = (e^{t\mathcal{L}_k}1)(x), \tag{C.6}$$

where the propagator $e^{t\mathcal{L}_k}$ acts on the initial condition $G(x, 0) = 1$ to yield some non-trivial function of x at time t. From there, we can use the results of Appendix B about the spectral decomposition of non-Hermitian operators to expand the initial unit function in the eigenbasis of $\mathcal{L}_k$. This yields, with the completeness relation (B.4) and the normalization (34) adopted,

$$1 = \sum_i r_k^{(i)}(x), \tag{C.7}$$

where $r_k^{(i)}$ denotes an eigenfunction of $\mathcal{L}_k$ (not necessarily all real), so that

$$G(x, t) = \sum_i e^{\zeta_i t} r_k^{(i)}(x), \tag{C.8}$$

where ζ_i denotes an eigenvalue of $\mathcal{L}_k$. In the long-time limit, the eigenvalue with largest real part will dominate the sum (if there is a gap) and, thus, we arrive at the result mentioned in Section 4.2 that the SCGF is the largest eigenvalue of $\mathcal{L}_k$.

The same result holds for $g \neq 0$, since there is also a Feynman–Kac formula for this case, one in fact that was never considered by Kac (see Exercise 9).

Exercises

All exercises are rated, rather subjectivity, according to Knuth's logarithmic rating system whereby 00 = immediate, 10 = simple, 20 = medium, 30 = moderately hard, 40 = term project, 50 = research problem. Feel free to contact me for questions or comments.

1. [10] The conservation of probability in quantum mechanics is expressed by the requirement that H is Hermitian. What is the corresponding property of L for Markov processes?

2. [15] Show that the stationary density of Kramers equation (13), in the absence of external forces ($\phi_t = 0$), is a Gibbs distribution involving the total energy $H(q, p) = p^2/(2m) + V(q)$. Calculate its associated stationary Fokker–Planck current. Does it vanish? Discuss the consequence of this result in view of the fact that equilibrium systems are supposed to have vanishing currents. Source: Chap. 10 of [7].

3. [15] Prove that the stationary density of the gradient SDE (15) is the Gibbs distribution (16). Do we still have that stationary distribution when σ is not proportional to the identity matrix? Calculate the associated stationary current.

4. [20] Show that the stationary density of the linear SDE (17), with attractor at $x = 0$, is a Gaussian distribution of the form

$$p_s(x) = \sqrt{\frac{\det C}{2\pi}} \exp\left(-\frac{1}{2}x \cdot Cx\right),\tag{C.9}$$

where C is a symmetric, positive matrix. What is the equation satisfied by C involving M and D? What is the stationary current $J_s(x)$? When is this density Gibbsian?

5. [20] Prove that the generator L of gradient SDEs is self-adjoint with respect to the following inner product

$$\langle f, g \rangle_{p_s} = \int f(x)g(x) p_s(x)\, dx,\tag{C.10}$$

where p_s is the Gibbs stationary distribution. What do you conclude for the spectrum of L?

6. [20] For gradient SDE (15), $\langle Lf, f \rangle_{p_s}$ is a so-called Dirichlet form:

$$\langle Lf, f \rangle_{p_s} = \langle f, Lf \rangle_{p_s} = -\frac{\varepsilon^2}{2} \|\nabla f\|_{p_s}^2,\tag{C.11}$$

where $\|\cdot\|_p$ is the norm weighted by p. Prove this result.

7. [15] Show that the expectation of the empirical density $\rho_T(x)$ defined in (21) is the stationary density $p_s(x)$ when X_t is ergodic. Explain why this must also be the most probable value of ρ_T in the long-time limit. [Hint: What is the distribution of ρ_T? Does it satisfy an LDP?]

8. [15] Show that the expectation of the empirical current $J_T(x)$ defined in (22) is the stationary Fokker–Planck current.

9. [20] The form of the tilted generator $\mathcal{L}_k$ is derived from the Feynman–Kac equation in Appendix C for the case $g = 0$. Adapt the proof for the case $g \neq 0$ leading to the full expression of $\mathcal{L}_k$ shown in (29).

10. [10] Show that $l_{k=0}(x) = p_s(x)$ and $r_{k=0}(x) = 1$ for all x.

11. [15] Show for gradient SDEs that $Lp_s = p_s L^\dagger$. Then use this result to show that $\mathcal{H}_k$, as defined by the symmetrization (35), is Hermitian.

12. [10] Derive for gradient SDEs the expression of the quantum generator $\mathcal{H}_k$ shown in (37), together with the corresponding potential $V_k(x)$ shown in (38).

13. [10] Derive the relations (40) and (41) between ψ_k, r_k, and l_k using the relation (35) between $\mathcal{H}_k$ and $\mathcal{L}_k$.

14. [20] Repeat the calculations of Section 4.4 for

$$V_T = \frac{1}{T}\int_0^T X_t^2\, dt,\tag{C.12}$$

which represents the empirical variance of X_t. What is the related quantum problem? Find $\lambda(k)$, $r_k(x)$ and $l_k(x)$ (correctly normalized). Source: [17].

15. [45] Repeat the previous exercise by replacing X_t^2 by X_t^k, $k > 2$.

16. [20] Study the large deviations of the entropy production Σ_T, as defined in Eq. (23), of the Ornstein–Uhlenbeck process. Assume, first, that X_0 and X_T have fixed values, e.g., $X_0 = a$ and $X_T = b$. Then repeat the calculation for $X_0 = a$, but by integrating X_T over $\mathbb{R}$. Finally, assume that X_0 is distributed according to $p_s(x)$. You should see in each case a very different effect of the "boundary terms" X_0 and X_T.

17. [25] Combine the calculation of Section 4.4 with the previous exercise to find the rate function of

$$Q_T = \frac{X_0^2}{T} - \frac{X_T^2}{T} + \frac{1}{T}\int_0^T X_t\, dt\tag{C.13}$$

for the Ornstein–Uhlenbeck process. This observable can be related to the heat exchanged by a Brownian particle with its environment when manipulated by laser tweezers [34]. Obtain the rate function for the three cases considered before: (i) $X_0 = a$, $X_T = b$; (ii) $X_0 = a$ and $X_T \in \mathbb{R}$; (iii) $X_0 \sim p_s(x)$ and $X_T \in \mathbb{R}$. You should see that the region where $\lambda(k) < \infty$ is different in each case, although $\zeta_{\max}(\mathcal{L}_k)$ is defined for all $k \in \mathbb{R}$. [Hint: You will need the generating function of the χ^2 distribution.]

18. [30] Calculate for the Ornstein–Uhlenbeck process the rate function $I(r)$ associated with

$$r_T = \frac{1}{T}\int_0^T \chi_{[-1,1]}(X_t)\, dt.\tag{C.14}$$

where $\chi_S(x)$ is the indicator function equal to 1 if $x \in S$ and 0 otherwise. This observable represents the fraction of time X_t spends in the interval $[-1, 1]$, so it is a random variable taking values in $[0, 1]$. What happens when $\gamma \to 0$? Sources: [35,36].

19. [40] Show that the occupation fraction r_T defined in the previous exercise can be obtained by integrating the empirical distribution $\rho_T(x)$ defined in (21) over $x \in [-1, 1]$. Can you use this result to derive the rate function $I(r)$ from the known rate function(al) $I(\rho)$ of $\rho_T(x)$? Source: [35].

20. [25] Repeat the exercise about the occupation time for $\chi_{[0,\infty)}(x)$ so as to study the fraction of time X_t stays positive. Source: [18].

21. [20] Consider the SDE

$$d\theta_t = \gamma \, dt + \sigma \, dW_t, \qquad \theta_t \in [0, 2\pi), \tag{C.15}$$

representing the motion of a Brownian particle on the unit circle with a drive or torque γ. Is this a gradient (equilibrium) system? What is the stationary Fokker–Planck current $J_s(x)$? Obtain the rate function $I(j)$ characterizing the fluctuations of the total empirical current

$$J_T = \frac{1}{T} \int_0^T d\theta_t = \int J_T(x) dx. \tag{C.16}$$

What are the natural boundary conditions for $r_k(x)$ and $l_k(x)$ in this case? Can $\mathcal{L}_k$ be symmetrized when $\gamma = 0$? Source: [37] and references therein.

22. [25] Calculate the rate function of the entropy production Σ_T for the 2D linear transversal SDE defined by

$$F(x, y) = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{C.17}$$

and $\sigma = \varepsilon \mathbb{1}$. Is this SDE gradient? Why is it called "transversal"? What do you obtain if you try to symmetrize it? Source: [38].

23. [25] Show that the tilted generator associated with the entropy production Σ_T satisfies the symmetry

$$\mathcal{L}_k^\dagger = \mathcal{L}_{-k-c}, \tag{C.18}$$

where c is some constant. Derive from this a symmetry satisfied by the SCGF and rate function. Such a symmetry is known as a **fluctuation relation**. Source: [11].

24. [30] Derive the expression of the tilted generator $\mathcal{L}_k$ for SDEs in which the noise is multiplicative, that is, in which the noise matrix σ depends on x. In this case, you must specify the stochastic convention used for interpreting the product $\sigma(x) dW_t$.

25. [25] Another approach for deriving the SCGF is to discretize in time an SDE to obtain a Markov chain for which A_T is then a sum. Use this approach to confirm that the SCGF $\lambda(k)$ is given by the dominant eigenvalue of $\mathcal{L}_k$. Source: Chap. V of [39] and the lecture notes of my large deviation course (see my website).

26. [40] Rewrite all the large deviations of these notes for Markov chains. Then do the same for Markov jump processes or continuous-time Markov chains. Source: [28].

27. [45] Do the eigenvectors r_k and l_k have any probabilistic or physical interpretation? Sources: [27,28].

28. [25] We have seen that, in the mapping to the quantum Hamiltonian, the noise parameter is essentially $\hbar$. With this, comment on the following: The low-noise limit of SDEs is equivalent to the semi-classical limit of quantum mechanics. Source: [12].

In memoriam

I dedicate this paper to the memory of E. G. D. Cohen (1923–2017): friend, collaborator, physicist and raconteur extraordinaire, who wisely told me once that the most important thing in science is to follow your heart.

References

[1] K. Jacobs, Stochastic Processes for Physicists: Understanding Noisy Systems, Cambridge University Press, Cambridge, 2010.
[2] Z. Brzeź, T. Zastawniak, Basic Stochastic Processes, in: Springer Undergraduate Mathematics Series, Springer, New York, 1999.
[3] G.A. Pavliotis, Stochastic Processes and Applications, Springer, New York, 2014.
[4] B. Øksendal, Stochastic Differential Equations, Springer, 2000.
[5] D.J. Higham, An algorithmic introduction to numerical simulation of stochastic differential equations, SIAM Rev. 43 (2001) 525–546.
[6] G. Grimmett, D. Stirzaker, Probability and Random Processes, Oxford University Press, Oxford, 2001.
[7] H. Risken, The Fokker–Planck Equation: Methods of Solution and Applications, third ed., Springer, Berlin, 1996.
[8] C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences, second ed., in: Springer Series in Synergetics, 13, Springer, New York, 1985.
[9] N.G. van Kampen, Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, 1992.
[10] L.E. Reichl, A Modern Course in Statistical Physics, fourth ed., Wiley, New York, 2016.
[11] J.L. Lebowitz, H. Spohn, A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics, J. Stat. Phys. 95 (1999) 333–365.
[12] H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. 478 (2009) 1–69.
[13] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Springer, New York, 1998.
[14] R. van Zon, E.G.D. Cohen, Stationary and transient work-fluctuation theorems for a dragged Brownian particle, Phys. Rev. E 67 (2003) 046102.

[15] F. den Hollander, Large Deviations, in: Fields Institute Monograph, AMS, Providence, 2000.
[16] R.J. Harris, H. Touchette, Large deviation approach to nonequilibrium systems, in: R. Klages, W. Just, C. Jarzynski (Eds.), Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond, in: Reviews of Nonlinear Dynamics and Complexity, vol. 6, Wiley-VCH, Weinheim, 2013, pp. 335–360.
[17] S.N. Majumdar, A.J. Bray, Large-deviation functions for nonlinear functionals of a Gaussian stationary Markov process, Phys. Rev. E 65 (2002) 051112.
[18] S.N. Majumdar, Brownian functionals in physics and computer science, Curr. Sci. 89 (2005) 2076–2092.
[19] V. Lecomte, C. Appert-Rolland, F. van Wijland, Chaotic properties of systems with Markov dynamics, Phys. Rev. Lett. 95 (2005) 010601.
[20] V. Lecomte, C. Appert-Rolland, F. van Wijland, Thermodynamic formalism for systems with Markov dynamics, J. Stat. Phys. 127 (2007) 51–106.
[21] V. Lecomte, C. Appert-Rolland, F. van Wijland, Thermodynamic formalism and large deviation functions in continuous time Markov dynamics, C. R. Phys. 8 (2007) 609–619.
[22] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Progr. Phys. 75 (2012) 126001.
[23] K. Mallick, The exclusion process: A paradigm for non-equilibrium behaviour, Physica A 418 (2015) 17–48.
[24] B. Derrida, Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current, J. Stat. Mech. 2007 (2007) P07023.
[25] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, C. Landim, Stochastic interacting particle systems out of equilibrium, J. Stat. Mech. 2007 (2007) P07014.
[26] R.J. Harris, G.M. Schütz, Fluctuation theorems for stochastic dynamics, J. Stat. Mech. 2007 (2007) P07020.
[27] R. Chetrite, H. Touchette, Nonequilibrium microcanonical and canonical ensembles and their equivalence, Phys. Rev. Lett. 111 (2013) 120601.
[28] R. Chetrite, H. Touchette, Nonequilibrium markov processes conditioned on large deviations, Ann. Henri Poincaré 16 (2015) 2005–2057.
[29] R. Chetrite, H. Touchette, Variational and optimal control representations of conditioned and driven processes, J. Stat. Mech. 2015 (2015) P12001.
[30] A. Barato, R. Chetrite, A formal view on 2.5 large deviations and fluctuation relations, J. Stat. Phys. 160 (2015) 1154–1172.
[31] J.P. Garrahan, I. Lesanovsky, Thermodynamics of quantum jump trajectories, Phys. Rev. Lett. 104 (2010) 160601.
[32] J.A. Bucklew, Introduction to Rare Event Simulation, Springer, New York, 2004.
[33] H. Touchette, A basic introduction to large deviations: Theory, applications, simulations, in: R. Leidl, A.K. Hartmann (Eds.), Modern Computational Science 11: Lecture Notes from the 3rd International Oldenburg Summer School, BIS-Verlag der Carl von Ossietzky Universität Oldenburg, 2011.
[34] R. van Zon, E.G.D. Cohen, Extended heat-fluctuation theorems for a system with deterministic and stochastic forces, Phys. Rev. E 69 (2004) 056121.
[35] F. Angeletti, H. Touchette, Diffusions conditioned on occupation measures, J. Math. Phys. 53 (2016) 023303.
[36] P. Tsobgni Nyawo, H. Touchette, A minimal model of dynamical phase transition, Europhys. Lett. 116 (2016) 50009.
[37] P. Tsobgni Nyawo, H. Touchette, Large deviations of the current for driven periodic diffusions, Phys. Rev. E 94 (2016) 032101.
[38] J.D. Noh, Fluctuations and correlations in nonequilibrium systems, J. Stat. Mech. 2014 (2014) P01013.
[39] J.A. Bucklew, Large Deviation Techniques in Decision, Simulation and Estimation, in: Wiley Series in Probability and Mathematical Statistics, Wiley Interscience, New York, 1990.