Diffusions conditioned on occupation measures
Florian Angeletti and Hugo Touchette

Citation: Journal of Mathematical Physics 57, 023303 (2016); doi: 10.1063/1.4941384
View online: http://dx.doi.org/10.1063/1.4941384
View Table of Contents: http://aip.scitation.org/toc/jmp/57/2
Published by the American Institute of Physics
Diffusions conditioned on occupation measures

Florian Angeletti and Hugo Touchette

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

(Received 16 October 2015; accepted 24 January 2016; published online 16 February 2016)

A Markov process fluctuating away from its typical behavior can be represented in the long-time limit by another Markov process, called the effective or driven process, having the same stationary states as the original process conditioned on the fluctuation observed. We construct here this driven process for diffusions spending an atypical fraction of their evolution in some region of state space, corresponding mathematically to stochastic differential equations conditioned on occupation measures. As an illustration, we consider the Langevin equation conditioned on staying for a fraction of time in different intervals of the real line, including the positive half-line which leads to a generalization of the Brownian meander problem. Other applications related to quasi-stationary distributions, metastable states, noisy chemical reactions, queues, and random walks are discussed. © 2016 AIP Publishing LLC.

I. INTRODUCTION

The stationary distribution of a Markov process gives, when it is unique, the average fraction of time the process spends in any given state in the long-time limit. When finite-time trajectories are considered, fluctuations around this average occupation occur, with a probability that depends on the forces and noise acting on the process. The position of a Brownian particle, for example, is positive in one dimension on average half of the time, yet sample trajectories have a strong tendency to stay positive or negative for any finite time, pushing the positive occupation above or below $\frac{1}{2}$. Similarly, Brownian particles evolving in complex potentials tend to spend most of their time around the stable equilibria of the acting potential but are also likely to “climb” it in finite time to reach possible unstable or metastable states.

Similar fluctuations of the occupation that persist in time are observed in almost all random systems, including jump processes describing noisy chemical reactions and particle transport, phase ordering and coarsening dynamics in magnetic systems, financial time series, queueing systems, as well as random walks on graphs. In these and many other applications, it is of interest not only to determine the probability that a process ventures in an atypical region of the state space, for example, around a metastable or unstable state, but also to describe with a modified process the effective dynamics of the process in that region.

We show in this paper how to formulate this problem as a Markov occupation conditioning problem which can be solved using the general framework proposed recently in Refs. 14–16. The general idea is illustrated in Fig. 1. We consider a general Markov process $X_t$ and condition probabilistically that process on spending a fraction $R_T$ of the time interval $[0,T]$ in some subset $S$ of its state space. Following Refs. 14–16, we then derive a new Markov process $\hat{X}_t$, called the driven process, which is equivalent to the conditioned process at the level of stationary states. In particular, the mean occupation of $\hat{X}_t$ in $S$ is $R_T$, so it realizes what is a fluctuation for $X_t$ in a typical way.

htouchette@sun.ac.za
FIG. 1. Illustration of the driven process for the occupation region $S = [0, 1]$. Left: Sample trajectory of a process $X_t$ (black curve) spending about 40% of its time in $S$ (gray region) compared to a sample trajectory of the driven process $\hat{X}_t$ (blue curve) representing the process $X_t$ conditioned on spending 80% of its time in $S$. Right: Fraction $R_T$ of the time interval $[0, T]$ spent in $S$ as a function of $T$ for $X_t$ (black) and $\hat{X}_t$ (blue).

The driven process is in this sense the modified process mentioned before: it represents the effective (stochastic) dynamics of $X_t$ as this process is seen to “fluctuate” in $S$ for a fraction of time given by the occupation measure $R_T$. When applied to noisy chemical reactions, for example, the driven process gives an effective chemical reaction with modified rates accounting for concentration fluctuations. For a random walk visiting a “rare” graph component, it gives a new random walk that concentrates on that component.

This effective representation of fluctuations can be constructed for any ergodic Markov processes, including Markov chains and jump processes. Here, we focus on diffusions described by stochastic differential equations in order to provide a new application of the framework developed in Refs. 14–16 and to set a template for applications based on continuous-space and continuous-time Markov models. As an example, we study in Sec. III the Langevin equation conditioned on staying in the interval $[a, b]$, the half-line $[a, \infty)$, and the point $\{a\}$ for a fraction of the time interval $[0, T]$. The second conditioning is a variant of the so-called Brownian meander, corresponding to Brownian motion restricted to stay positive for $t \in [0, T]$. Other physical and manmade applications of the driven process related to more complex diffusions, jump processes, and random walks are mentioned in the Conclusion of the paper.

II. OCCUPATION CONDITIONING

We explain in this section how the conditioned and the driven processes are constructed for a conditioning involving an occupation measure. This is a special case of the framework presented in Refs. 14–16 dealing with general, time-integrated random variables for the conditioning.

A. Model

We consider a pure diffusion process $X_t \in \mathbb{R}^m$ described by the following (Itô) stochastic differential equation (SDE),

$$dX_t = F(X_t)dt + \sigma dW_t,$$

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is the drift, $W_t$ is an $n$-dimensional Brownian motion, and $\sigma$ is the $m \times n$ noise matrix, assumed for simplicity to be constant in space and non-singular (invertible). The probability density $p(x, t)$ of this process evolves according to the Fokker-Planck equation

$$\partial_t p(x, t) = L^* p(x, t),$$

expressed here in terms of the Fokker-Planck operator

$$L^* = -\nabla \cdot \nabla + \frac{1}{2} \nabla \cdot D \nabla,$$
with $D = \sigma \sigma^T$ the diffusion matrix. For the remaining, we also need the adjoint of the Fokker-Planck operator,

$$L = F \cdot \nabla + \frac{1}{2} \nabla \cdot D \nabla,$$

which generates the evolution of expectations of $X_t$.

Given the evolution of $X_t$, we now consider a subset $S \subset \mathbb{R}^m$ and look at the fraction of time that $X_t$ spends in $S$ in the time interval $[0, T]$ using

$$R_T = \frac{1}{T} \int_0^T \mathbb{1}_S(X_t) \, dt,$$

where $\mathbb{1}_S(x)$ denotes the indicator function equal to 1 if $x \in S$ and 0 otherwise. This random variable, which explicitly depends on both $S$ and $T$, is called the occupation measure of $S$ or the (normalized) local time when $S$ is a single point. Assuming that $X_t$ has a unique stationary distribution $\rho^*$ satisfying $L^* \rho^* = 0$, we have by the ergodic theorem that

$$\lim_{T \to \infty} R_T = E_{\rho^*}[\mathbb{1}_S] = \int_S \rho^*(x) \, dx,$$

so that $R_T$ converges in probability to the mean occupation in $S$ given by $\rho^*(S)$.

For finite integration times, $R_T$ fluctuates around this concentration point according to its probability density $P(R_T = r)$, which can be expressed for large times as

$$P(R_T = r) = e^{-TI(r)/\sigma(T)}$$

or, equivalently,

$$\lim_{T \to \infty} \frac{1}{T} \ln P(R_T = r) = I(r).$$

This scaling of the distribution is known as a large deviation principle (LDP). The rate of decay $I(r)$ is called the rate function and can be obtained from the contraction principle of large deviation theory by the following minimization:

$$I(r) = \min_{\rho : C(\rho) = r} J(\rho),$$

which involves the Donsker-Varadhan or level-2 rate function,

$$J(\rho) = \min_{h > 0} \int \rho(x)(h^{-1}Lh)(x) \, dx,$$

and the contraction linking $\rho$ to $R_T$,

$$C(\rho) = \int_{\mathbb{R}^m} \rho(x)\mathbb{1}_S(x) \, dx = \int_S \rho(x) \, dx.$$

This result is derived in the Appendix.

The LDP (7) shows that $X_t$ is exponentially unlikely for long times $T$ to enter the region $S$ for a fraction $R_T$ of time, except when $R_T$ is the stationary fraction $r^*$ of time spent in $S$. The ergodic theorem indeed states that $P(R_T = r^*) \to 1$ as $T \to \infty$, which implies $I(r^*) = 0$, corresponding to the typical occupation of $X_t$. Any other fraction $R_T \neq r^*$ represents an atypical occupation of $X_t$ in $S$ characterized by $I(r) > 0$ and so $P(R_T = r) \to 0$ as $T \to \infty$. For more information about the large deviations and applications of occupation times, see Refs. 25–28.

**B. Conditioned and driven processes**

We now consider a fixed occupation $R_T = r$ of $X_t$ in $S$ and derive the effective driven process $\hat{X}_t$ that describes $X_t$ conditioned on (or restricted to) this occupation. The construction of $\hat{X}_t$ is explained in Refs. 14–16 and requires that we find the dominant eigenvalue $\lambda(k)$ and corresponding eigenfunction $r_k$ of the tilted generator, defined by

$$\mathcal{L}_k = L + k \mathbb{1}_S,$$

where $\mathcal{L}_k$ is called the tilted generator.
where \( k \in \mathbb{R} \) and \( L \) is the generator (4) of \( X_t \). With these elements, the driven process is defined as the Markov process with modified generator

\[
L_k = r_k^{-1} L_k r_k - r_k^{-1} (L_k r_k)
\]  

(13)

acting on functions \( f \) according to

\[
(L_k f)(x) = \frac{1}{r_k(x)} (L_k r_k f)(x) - \frac{1}{r_k(x)} (L_k r_k)(x) f(x)
\]

\[= \frac{1}{r_k(x)} (L_k r_k f)(x) - \lambda k f(x).\]  

(14)

As shown in Ref. 15, the effect of this transform on SDE (1) is to change the drift \( F \) to the modified or driven drift

\[
F_k(x) = F(x) + D \nabla \ln r_k(x)
\]  

(15)

while keeping the diffusion matrix \( D \) constant. The evolution of the driven process \( \hat{X}_t \) is thus given by the modified SDE

\[
d\hat{X}_t = F_k(\hat{X}_t)dt + \sigma dW_t
\]  

(16)

perturbed by the same Gaussian noise as \( X_t \) but involving the new driven drift \( F_k \).

The connection between the driven process and the conditioning of \( X_t \) on \( R_T = r \) is illustrated again in Fig. 1 and is fully explained in Ref. 15. The idea briefly is that the driven process \( \hat{X}_t \) and the conditioned process \( X_t | R_T = r \) have the same stationary properties, in addition to having similar probabilities for their trajectories as \( T \rightarrow \infty \), if the rate function \( I(r) \) of \( R_T \) is convex and \( k \) is chosen so that

\[
k = I'(r).
\]  

(17)

In this sense, we then say that \( \hat{X}_t \) is equivalent to \( X_t | R_T = r \) or realizes that conditioned process in the long-time limit.

This equivalence is similar to the equivalence of the microcanonical and canonical ensembles in equilibrium statistical mechanics:15 the conditioned process \( X_t | R_T = r \) is essentially a process generalization of the microcanonical ensemble in which the “energy” \( R_T \) is constant and equal to \( r \), whereas the driven process is a generalization of the canonical ensemble in which \( R_T \) fluctuates but concentrates in the “thermodynamic limit” \( T \rightarrow \infty \) to \( r \), the constant of the microcanonical ensemble. This is achieved by matching the “temperature” \( k \) to the constraint \( R_T = r \) according to (17), which is an analog of the thermodynamic temperature-energy relation.

Another way to understand the driven process is as an optimal change of measure or process.16 Recall that the event \( R_T = r \) is a rare fluctuation in the original process \( X_t \) having an exponentially small probability for long times \( T \). The driven process, by contrast, is such that \( R_T = r \) happens with certainty as \( T \rightarrow \infty \), so that the transformation (13) modifies the process \( X_t \) to make a rare occupation typical. In general, many transformed processes can be used to achieve this reweighting of rare events. The driven process is special in that it the process closest to \( X_t \), with respect to a metric defined by the relative entropy, that makes the occupation \( R_T = r \) typical; see Ref. 16 for more details.

C. Spectral problem and effective potential

The difficulty of constructing the driven process comes from solving the spectral problem

\[
L_k r_k = \lambda(k) r_k
\]  

(18)

for the dominant eigenvalue and its corresponding eigenfunction. Depending on the form of generator \( L \) considered and, more precisely, its self-adjointness, three cases arise for obtaining \( \lambda(k) \) and \( r_k \).

**Case 1:** \( L = L^\dagger \). This is the simplest case determining a reversible process with respect to the Lebesgue (uniform) measure. In this case, the techniques of quantum mechanics apply:
the spectrum of $L$ or $L_k$ is real and the eigenfunction $r_k$ must be found by solving (18) with vanishing boundary condition for $r_k^2(x)$ as $|x| \to \infty$.

**Case 2:** $L \neq L^\dagger$ but the spectrum of $L$ is real. This arises, for example, when $X_t$ is a reversible or equilibrium diffusion having a gradient drift

$$F = -\frac{D}{2} \nabla U$$

and, therefore, a Gibbs stationary distribution

$$p^*(x) = e^{-U(x)}.$$  

In this case, it is known that $L$ is self-adjoint with respect to an inner product defined with $p^*$ and that this can be used to “symmetrize” $L$ into a self-adjoint operator $H$, playing the role of a quantum Hamiltonian.  

This symmetrization is simply defined as

$$H = e^{-U/2} L e^{U/2}$$

and leads, when applied to $L_k$, to the tilted Hamiltonian

$$\mathcal{H}_k = e^{-U/2} L_k e^{U/2} = \frac{D}{2} \left\{ -\Delta + \frac{\Delta U}{2} - \left(\frac{\nabla U}{2}\right)^2 \right\} + k \mathbb{I}_S.$$  

This operator has of course the same real spectrum as $L_k$, so that

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

but its dominant eigenfunction $\psi_k$, obtained with the natural vanishing boundary condition $\psi_k(x)^2 = 0$ at infinity, is related to $r_k$ by $r_k = e^{U/2} \psi_k$.

**Case 3:** $L \neq L^\dagger$ and the spectrum is complex. This happens when $F$ is not gradient, $\sigma$ depends on $X_t$, or external reservoirs are included in this process as boundary conditions. In this case, $X_t$ represents a genuine nonequilibrium process supporting non-vanishing probability currents, for which $L$ or $L_k$ cannot be symmetrized. Moreover, the spectral problem (18) on its own is incomplete: it must be solved in tandem with the dual problem

$$L_k^\dagger l_k = \lambda(k) l_k$$

with the boundary condition $l_k(x) r_k(x) = 0$ at infinity. This is more difficult to solve in general than the case of self-adjoint operators (cases 1 and 2).

We focus in the rest of the paper mostly on case 2, which is equivalent to a quantum ground state problem with effective Schrödinger Hamiltonian $\mathcal{H}_k$. Assuming that the drift is conservative, as in (19), we can express the driven drift (15) in this case also in gradient form,

$$F_k = -\frac{D}{2} \nabla U_k,$$

by introducing the effective or driven potential

$$U_k(x) = U(x) - 2 \ln r_k(x) = -2 \ln \psi_k(x),$$

which realizes the occupation conditioning.

Non-reversible diffusions falling in case 3 cannot be represented by such an effective potential, even though the modified drift $F_k$, as given by (15), is always a gradient perturbation of the original drift $F$ when $D$ is constant. This property of $F_k$ comes from the time-additive form of $R_T$. For other conditionings, based, for example, on currents or the entropy production, the perturbation $F_k - F$ can have a non-conservative and, therefore, nonequilibrium component; see Sec. 5.5 of Ref. 15 for more detail.
III. APPLICATION

We illustrate in this section our results for an exactly solvable model based on the linear Langevin equation or one-dimensional Ornstein-Uhlenbeck process defined by

\[ dX_t = -\gamma X_t dt + \sigma dW_t, \]

where \( X_t \in \mathbb{R} \), \( W_t \in \mathbb{R} \), with \( \gamma \) and \( \sigma \) positive constants. This process obviously falls in case 2 of Sec. II, as do all processes defined on \( \mathbb{R} \) without sinks or sources. The linear drift \( F(x) = -\gamma x \) is associated with the parabolic potential

\[ U(x) = \frac{\alpha x^2}{2}, \]

where \( \alpha = 2\gamma/D \) and \( D = \sigma^2 \). The quantum problem that we need to solve therefore is

\[ \left[ \frac{d^2}{dx^2} + \frac{\alpha}{2} - \frac{\alpha^2}{4} x^2 + \frac{2k}{D} \mathbb{I}_S(x) \right] \psi(x) = \frac{2\lambda}{D} \psi(x), \]

which we convert to

\[ \Psi''(x) - \frac{x^2}{4} \Psi(x) + \left[ \frac{1}{2} + \frac{2}{Da} \left( k \mathbb{I}_S(x) - \lambda \right) \right] \Psi(x) = 0 \]

with the rescaling \( \Psi(x) = \psi(x/\sqrt{\alpha}) \). The same quantum problem can be obtained using path integral methods as applied to Brownian functionals and the Feynman-Kac equation, see Refs. 27 and 29.

Equation (30) is essentially the Weber equation with piecewise-constant coefficients, representing a quantum harmonic oscillator with piecewise-shifted potential. It can be solved exactly in and out of the conditioning interval \( S \) and then pieced together by requiring continuity at the boundaries of \( S \). This is done next for three occupations, namely, \( S = [a, b] \) (finite interval), \( S = [a, \infty) \) (half-line), and \( S = \{a\} \) (point conditioning).

A. Finite interval

For \( S = [a, b] \), we must solve (30) on the three regions \( (-\infty, a) \), \( [a, b] \), and \( (b, \infty) \), and piece the three solutions, as mentioned, continuously at \( x = a \) and \( x = b \). Over each region, the Weber equation has the form

\[ \Psi''(x) - \left( \frac{x^2}{4} + \nu(x) \right) \Psi(x) = 0, \]

where

\[ \nu(x) = \frac{2}{aD} \left( \lambda - \frac{\alpha D}{4} - k \mathbb{I}_{S\setminus\{S\}}(x) \right). \]

This function takes only two values, denoted from now on by \( \nu = \nu(S) \) and \( \nu' = \nu(\mathbb{R}\setminus S) \).

The solution space of the Weber equation is spanned by

\[ s_1(\nu, x) = e^{-\nu x^2/4} \gamma_{11} \left( \frac{\nu}{2} + \frac{1}{4} \right), \]

\[ s_2(\nu, x) = xe^{-\nu x^2/4} \gamma_{11} \left( \frac{\nu}{2} + \frac{3}{4} \right), \]

where \( \gamma_{11} \) is the confluent hypergeometric function of the first kind. From these two particular solutions, it is possible to construct a solution outside \([a, b]\) that decays to 0 at infinity,

\[ W(\nu, x) = \frac{1}{2^{\nu + \frac{1}{4} \sqrt{\nu}}} \left[ \cos \left( \frac{\sqrt{\nu}}{2} \right) \Gamma \left( \frac{1}{4} - \frac{\sqrt{\nu}}{2} \right) s_1(\nu, x) - \sqrt{2} \sin \left( \frac{\sqrt{\nu}}{2} \right) \Gamma \left( \frac{1}{4} - \frac{\sqrt{\nu}}{2} \right) s_2(\nu, x) \right]. \]
Combining these solutions, we then construct the complete eigenfunction as

$$\Psi(x) = \begin{cases} K_1 W(v', -x) & x < a \\ K_2 s_1(v, x) + K_3 t_2(v, x) & a < x < b \\ K_4 W(v', x) & b < x \end{cases},$$  \hspace{0.1cm} (35)

where the $K_i$'s are constants to be adjusted by imposing continuity.

To this end, we define the vector $K = (K_1, K_2, K_3, K_4)^T$ and the matrix

$$C(\lambda, k) = \begin{pmatrix} -W(v', -a) & s_1(v, a) & s_2(v, a) & 0 \\ \partial_x W(v', -a) & \partial_s s_1(v, a) & \partial_s s_2(v, a) & 0 \\ 0 & s_1(v, b) & s_2(v, b) & -W(v', b) \\ 0 & \partial_s s_1(v, b) & \partial_s s_2(v, b) & -\partial_x W(v', b) \end{pmatrix},$$  \hspace{0.1cm} (36)

where $\partial_x$ denotes the first derivative with respect to the second coordinate (noted $x$ above). The continuity of $\Psi(x)$ at $a$ and $b$ is equivalent to the following linear equation:

$$C(\lambda, k) K = 0.$$  \hspace{0.1cm} (37)

Non-trivial solutions therefore exist if, and only if,

$$\det C(\lambda, k) = 0.$$  \hspace{0.1cm} (38)

This defines a transcendental equation in $\lambda$ involving hypergeometric functions, which can easily be solved numerically to obtain $\lambda(k)$ with an arbitrary precision. To find the associated rate function $I(r)$, we then use the fact that $\lambda(k)$ and $I(r)$ are related by Legendre transform when the former is differentiable,$^{22-24}$

$$I(r) = \sup_{k \in \mathbb{R}} \{ kr - \lambda(k) \}.$$  \hspace{0.1cm} (39)

In parametric form, we therefore have

$$I(\lambda'(k)) = k \lambda'(k) - \lambda(k), \hspace{0.5cm} k \in \mathbb{R}.$$  \hspace{0.1cm} (40)

Figure 2 shows the result of these expressions for $S = [0, 1]$. As can be seen, the tails of $\lambda(k)$ are asymptotically linear with slopes 0 and 1 as $|k| \to \infty$, reflecting the fact that $I(r)$ is defined for $r \in [0, 1]$ and is steep at $r = 0$ and $r = 1$. This is important for what follows as it means, following (17), that the effective potential associated with no occupation in $[a, b]$ is obtained by taking the limit $k \to -\infty$, whereas full occupation in $[a, b]$ is obtained with $k \to \infty$. In between, $k$ is related to the occupation fraction $r$ via (17) or equivalently $\lambda'(k) = r$.

To find the effective potential $U_k$, we compute the kernel of the matrix $C(k, \lambda)$ to obtain $\Psi$ via (35), and then rescale $\Psi$ back to $\psi$. Figure 3 shows the result of these calculations for $S = [0, 1]$ and different values of $k$. For $k > 0$, we see that $U_k(x)$ becomes steeper around $[0, 1]$ compared to the “natural” potential $U(x)$ obtained for $k = 0$. This confines the process inside $[0, 1]$ and so increases

![Figure 2](image-url)  \hspace{0.1cm} FIG. 2. SCGF (left) and rate function (right) for $S = [0, 1]$. Parameters: $\alpha = D/2 = 1.$
FIG. 3. Effective potential $U_k(x)$ for $S = [0, 1]$. Left: $k = 0 : 2 : 10$ (from bottom to top curves) using the notation $k = k_{\text{min}} : dk : k_{\text{max}}$. Right: $k = 0 : -2 : -10$ (from bottom to top curves). Parameters: $\alpha = D/2 = 1$.

naturally the time spent inside this interval. In the limit $k \to \infty$, the process is completely confined inside that interval by an infinitely steep potential $U_\infty(x)$ shown in Fig. 4. In this case, it is easy to see by analogy with the confined harmonic oscillator that $U_\infty(x)$ must diverge logarithmically near $x = 0$ and $x = 1$, since $\psi_k(x)$ vanishes at these points. This yields a diffusive version of the so-called $Q$-process, arising in the context of quasi-stationary distributions, which corresponds here to the Ornstein-Uhlenbeck process conditioned on not leaving $[0, 1]$.

The effective potential is more interesting for $k < 0$. In this case, a non-trivial barrier develops inside $[0, 1]$ so as to “deconfine” the process from $[0, 1]$, leading to a reduced occupation in that interval. As $k \to -\infty$, $U_k(x)$ becomes steep near $x = 0^+$, as shown in Fig. 4, preventing the process to reach $[0, 1]$ from negative initial conditions. It also becomes steep near $x = 1^+$ while being raised, as shown in the left plot of Fig. 3. However, because the height of the potential obtained for $x > 1$ does not play any role when it becomes disconnected from the one obtained for $x < 0$, we can shift the former down to zero, yielding the limiting potential shown in Fig. 4. This leads effectively to a breaking of ergodicity for the process conditioned on not entering $[0, 1]$: the process started in the region $x < 0$ stays in that region and cannot visit the region $x > 1$ because of the infinite barrier at $x = 0$. Conversely, when $X_t$ is started in the region $x > 1$, it stays in that region and cannot cross to $x < 0$. For initial conditions in $[0, 1]$, the process is not defined, at least not in the formal limit $k = -\infty$. For any finite $k$, however, the driven process is ergodic.

B. Halfline

We now consider $S = [a, \infty)$ as the occupation set to show how our results can be used to study variants of the Brownian meander process corresponding to Brownian motion conditioned on

FIG. 4. Black: Effective potential $U_\infty(x)$ preventing any occupation in $S = [0, 1]$. Blue: Effective potential $U_\infty(x)$ forcing a total occupation in $S = [0, 1]$. Parameters: $\alpha = D/2 = 1$. 
staying positive. The Weber solution in this case has two branches,

$$\Psi(x) = \begin{cases} 
K_1W(\nu', -x) & x < a \\
K_2W(\nu, x) & x > a 
\end{cases}$$  \hspace{1cm} (41)

linked continuously at $x = a$ by solving (38) using the matrix

$$C(\lambda, k) = \begin{pmatrix}
-W(\nu', -a) & W(\nu, a) \\
-\partial_x W(\nu', -a) & \partial_x W(\nu, a)
\end{pmatrix}.$$  \hspace{1cm} (42)

Figure 5 shows the results of the numerical calculation of $\lambda(k)$ and $I(r)$ from this matrix for $a = 1$, which are overall qualitatively similar to those of Fig. 2 because of the restriction $r \in [0,1]$. In Fig. 6, we show the effective potential $U_k(x)$ for positive and negative values of $k$ related to a confinement in the region $x > 1$ and $x < 1$, respectively. The shape of $U_k(x)$ is also qualitatively similar to the previous results obtained for $[0,1]$, except that it develops only one steep barrier instead of two. As before, the divergence of $U_k(x)$ near $x = 1$ appearing in the limit $k \to \pm\infty$ is logarithmic, since the wavefunction $\psi_k(x)$ of the quantum harmonic oscillator with an infinite wall at $x = 1$, the equivalent quantum problem, vanishes at $x = 1.31,32$

Another interesting feature to observe in Fig. 6 is that the “natural” potential $U = U_0$ is not modified much by the conditioning on the side of occupation; that is to say, more occupation for $x > 1$ (respectively, $x < 1$) does not change the right (respectively, left) branch of $U$ significantly. This can be understood from the quantum perspective by noting that the introduction of a wall or well in the parabolic potential does not affect the tails of $\psi_k$ far away from this wall or well. The same phenomenon can also be explained using recent results showing that the modified force $F_k$ minimizes a cost function involving a weighted integral of $(F - F_k)^2$ and the occupation $R_T = r$. As a result, the drift $F$ is modified only minimally whenever it contributes “naturally” to the occupation

![FIG. 6. Effective potential $U_k(x)$ for $S = [1, \infty)$](image)
targeted. This cost, importantly, is a function of the drift and not the potential, so that large differences between \( U_k \) and \( U \), such as those seen in the right plot of Fig. 3, do not necessarily translate into large differences between \( F_k \) and \( F \) and, therefore, large costs.

Considering \( a = 0 \) instead of \( a = 1 \) does not change these results much. The only difference is that the rate function shown in Fig. 5 is symmetric about \( r = 0.5 \), which leads to an effective potential \( U_k(x) \) for \( k < 0 \) that is the mirror image of \( U_k(x) \) for \( k > 0 \), that is, \( U_k(x) = U_{-k}(-x) \). A logarithmic singularity near \( x = 0 \) also appears for \( a = 0 \) in the limits \( k \to -\infty \) and \( k \to \infty \), which restrict the occupation in the negative and positive regions, respectively. The positive case is interesting as it is related to the so-called arc sine law\(^{27,38,39} \) and leads to a generalization of the Brownian meander. Indeed, solving the Weber equation for \( k \to \infty \), which is equivalent to the quantum harmonic oscillator with a wall,\(^{31,32} \) we find that the asymptotic Ornstein-Uhlenbeck meander defined as the SDE (27) conditioned on staying positive at all times, is a nonlinear diffusion with potential \( U_{\text{OUm}} = U_\infty \) having the following tails:

\[
U_{\text{OUm}}(x) \sim \begin{cases} 
-c \ln x & x \to 0^+ \\
\beta x^2 / 2 & x \to \infty
\end{cases},
\]

where \( c \) and \( \beta \) are constants that can be determined numerically from \( C(\lambda, k) \). The drift of this meander is thus given asymptotically by

\[
F_{\text{OUm}}(x) \sim \begin{cases} 
c / x & x \to 0^+ \\
-\beta x & x \to \infty
\end{cases}.
\]

For pure Brownian motion (\( \gamma = \alpha = 0 \)), we find \( \beta = 0 \), which is consistent with the exact drift of the Brownian meander; see Eq. (21) of Ref. 19.

**C. Point occupation**

The third and last application that we consider is the point occupation at \( x = a \), obtained by replacing \( \mathbb{I}_S(x) \) by \( \delta(x - a) \) in the definition of \( R_T \) to obtain the local time at \( a \). This case can also be considered as the limit \( \epsilon \to 0 \) of \( S = [a - \epsilon/2, a + \epsilon/2] \), with \( \mathbb{I}_S \) replaced by \( \mathbb{I}_S(x) / \epsilon \), and leads to the following Weber solution:

\[
\Psi(x) = \begin{cases} 
K_1 W(\nu', -x) & x < a \\
K_2 W(\nu', x) & x > a
\end{cases}
\]

with continuity conditions

\[
K_1 W(\nu, -a) - K_2 W(\nu, a) = 0 \\
K_1 \partial_x W(\nu, -a) + K_2 \partial_x W(\nu, a) = k \Psi(a).
\]

In the particular case \( a = 0 \), these conditions reduce to the following relation:

\[
k = -\frac{\cot(\pi \left(\frac{\nu}{2} + \frac{1}{2}\right)) \Gamma\left(\frac{3}{4} - \frac{\nu}{2}\right)}{\sqrt{2} \Gamma\left(\frac{3}{4} - \frac{\nu}{2}\right)},
\]

which can be used as an implicit equation to find \( \lambda(k) \) via expression (32) of \( \nu \).

The solution for \( \psi_k \) that we find in this case is similar to the one found for the half line, except that the derivative of \( \psi_k \) is now discontinuous because of the delta source at \( x = a \), and jumps according to the second line in (46). This introduces a kink at \( x = a \) in the effective potential \( U_k(x) \), illustrated in Fig. 7 for \( a = 0 \), which is reminiscent of the kink seen in the potential of Brownian motion with dry or solid friction.\(^{40-42} \) This makes sense intuitively: for the process \( X_t \) to have a larger local time at \( x = a \), it has to “stick” more onto that point, similarly to what is observed with solid friction. Conditioning on having a smaller local time at \( x = a \) forces, on the other hand, \( X_t \) to “avoid” that point as if there was a “negative” solid friction force (see left plot of Fig. 7).

In the limit \( k \to \infty \), the potential \( U_k \) becomes degenerate and concentrates the process on \( x = a \), whereas for \( k \to -\infty \), it develops an infinite barrier at \( x = a \) with two logarithmic branches.
that prevents occupation onto that point. The latter limit yields a $Q$-process version of the Ornstein-Uhlenbeck process conditioned on not reaching $x = a$, which also breaks ergodicity.

IV. PERTURBATION THEORY

We complement the exact results of Sec. III by developing a perturbation theory in the parameter $k$ for obtaining $\lambda(k)$, $I(r)$, and $U_k(x)$. In principle, this perturbation can be applied around any value $k$ for which the spectrum of $\mathcal{H}_k$ or $L_k$ is known, even if $L_k$ is not symmetrizable. For simplicity we consider reversible processes with effective (self-adjoint) Hamiltonian $\mathcal{H}_k$ and develop a perturbation in the form

$$\mathcal{H}_{k+\Delta k} = \mathcal{H}_k + \Delta k \ 1_\mathcal{S}. \tag{48}$$

A natural starting point is $k = 0$, since $\mathcal{H}_0 = H$ is simply the Hamiltonian (obtained by symmetrization of $L$) of the quantum harmonic oscillator with shifted energy levels, so that $\lambda(0) = 0$ and $\psi_0 = e^{-U/2} = \sqrt{\beta}^2$.

The application of standard perturbation theory for self-adjoint operators with non-degenerate spectrum gives directly

$$\partial_k \lambda_n(k) = \langle \Psi_n(k) | 1_\mathcal{S} | \Psi_n(k) \rangle \tag{49}$$

and

$$\partial_k \Psi_n(k) = \sum_{m \neq n} \frac{\langle \Psi_m(k) | 1_\mathcal{S} | \Psi_n(k) \rangle}{\lambda_m(k) - \lambda_n(k)} \Psi_m(k). \tag{50}$$

Here, we use the quantum bracket notation for the inner product and now denote by $\lambda_n(k)$ and $\Psi_n(k)$ the $n$th eigenvalue of $\mathcal{H}_k$ and its corresponding eigenfunction, respectively. The matrix elements

$$N_{i,j}(k) = \langle \Psi_i(k) | 1_\mathcal{S} | \Psi_j(k) \rangle \tag{51}$$

driving the “evolution” of $\lambda_n(k)$ and $\Psi_n(k)$ as a function of $k$ have a natural geometric interpretation: they represent an orthogonality defect of the basis $\{ \Psi_n(k) \}$ with respect to the modified inner product,

$$\langle \Psi_m(k) | 1_\mathcal{S} | \Psi_n(k) \rangle = \int_{\mathcal{S}} \Psi_m^*(k,x) \Psi_n(k,x) dx, \tag{52}$$

which defines mathematically a semi-positive sesquilinear form. To complete these equations, we can calculate the evolution of the orthogonality defect matrix $N$ itself with the perturbation,

$$\partial_k N_{i,j}(k) = (\Psi_i(k) | 1_\mathcal{S} | \partial_k \Psi_j(k)) + (\partial_k \Psi_i(k) | 1_\mathcal{S} | \Psi_j(k))$$

$$= \sum_{m \neq i} \frac{N_{i,m}(k) N_{m,j}(k)}{\lambda_i(k) - \lambda_m(k)} + \sum_{m \neq j} \frac{N_{j,m}(k) N_{m,i}(k)}{\lambda_j(k) - \lambda_m(k)}. \tag{53}$$
FIG. 8. Left: SCGF computed through perturbation for $S = [0, 1]$ with $M = 2, 5, 10, 20$ modes (from bottom to top curves). The black curve shows the exact $\lambda(k)$. Right: Corresponding rate function. The black curve shows the exact $I(r)$. Parameters: $\alpha = D/2 = 1$.

which simplifies on the diagonal to

$$\partial_k N_{i,j}(k) = \sum_{m \neq i} \frac{2|N_{i,m}(k)|^2}{\lambda_m(k) - \lambda_i(k)}.$$  \hspace{1cm} (54)

It can be checked that these differential equations for $N(k)$ admit the set of diagonal matrices as fixed points. Knowing that $k \to -\infty$ is equivalent to total deconfinement in $S$, we find however that $N(-\infty) = 0$ is the only stable fixed point reached in that limit. Similarly, since $k \to \infty$ corresponds to total confinement in $S$, we must have $N(+\infty) = I$.

Equations (49), (50), and (53) define a set of coupled nonlinear differential equations that can be used to find the SCGF $\lambda(k)$, which corresponds to the dominant eigenvalue $\lambda_0(k)$, and its associated eigenfunction $\Psi$, which corresponds to $\Psi_0(k)$. From the dominant eigenfunction, we then find the driven potential $U_k$ as in Sec. III. Moreover, using the 0th component of (49), we can obtain the rate function $I(r)$ by rewriting the parametric expression (40) as

$$I(N_{0,0}(k)) = k N_{0,0}(k) - \lambda_0(k).$$  \hspace{1cm} (55)

Figure 8 shows the perturbation results for $\lambda(k)$ and $I(r)$ obtained by integrating Eqs. (49), (50), and (53) starting from the known eigenvalues $\lambda_n(0)$ and eigenstates $\Psi_n(0)$ of the quantum harmonic oscillator. The results are for the unit interval occupation, $S = [0, 1]$, and are also obtained by truncating $N(k)$ to a finite size $M$. As can be seen, the difference between the perturbative and exact results decreases by increasing $M$, as expected, and becomes negligible for $M = 10$. The perturbation also converges quickly for $S = [1, \infty)$ (results not shown), but not for the point occupation case, as shown in Fig. 9. For the latter, the SCGF $\lambda(k)$ obtained by perturbation strongly differs from the exact result obtained from the methods of Sec. III for $k$ beyond some positive value $k_c$, which is only slightly shifted by increasing $M$. This arises because the introduction of
a Dirac well in a potential (the quantum problem for \( k > 0 \)) strongly modifies the eigenfunctions and eigenvalues of that potential. By comparison, a Dirac wall (the quantum problem for \( k < 0 \)) modifies these eigenfunctions essentially only in the way that they are joined at \( x = a \), which leads to a small perturbation of the eigenvalues, including the dominant eigenvalue \( \lambda(k) \) which converges to a constant as \( k \to -\infty \).

To complete the perturbation analysis, we show in Fig. 10 the evolution of the matrix \( N(k) \) for increasing values of \( k \), obtained by integrating the same differential equations truncated to order \( M = 10 \). For \( S = [1, \infty) \) (top panel), we see that this evolution essentially involves three phases: a first for \( k < 0 \) in which \( N \) has the approximate block form

\[
N \approx \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{I} \end{pmatrix},
\]

a second around \( k = 0 \) in which \( N \approx \mathbb{I} \), and then a third phase obtained for \( k > 0 \) in which

\[
N \approx \begin{pmatrix} \mathbb{I} & 0 \\ 0 & 0 \end{pmatrix}.
\]

The first and last block phases are approximations of the extreme solutions \( N(-\infty) = 0 \) and \( N(\infty) = \mathbb{I} \), respectively, containing errors in the lower block coming from the truncation. In each case, the upper corner of \( N \) follows the extreme solutions, which confirms that the largest eigenvalues—in particular, the dominant eigenvalue—are minimally affected by truncation. A similar evolution of \( N(k) \) is observed for \( S = [0, 1] \) (lower panel), although convergence to \( N(-\infty) = 0 \) and \( N(\infty) = \mathbb{I} \) in this case is slower and involves more truncation errors outside the upper corner of \( N \). These results are obtained with a direct truncation of \( N(k) \) in the eigenbasis defined by \( \Psi_n(k) \); a more efficient truncation leading to smaller errors could be constructed in principle by choosing a different function basis.

V. CONCLUSION

We have shown how a Markov process which is observed to spend a long time in some region of its state space can be represented by a modified Markov process, called the driven process, representing physically the dynamics of the original process restricted to that region. We have constructed this driven process for the Ornstein-Uhlenbeck process and have shown how it can be used to obtain two important probabilistic constructions, namely, stochastic meanders which are confined in a certain region of space, and \( Q \)-processes which avoid a region of space.

The application of these results to higher-dimensional diffusions that are reversible should follow the example of the Ornstein-Uhlenbeck process. In this case, the driven process is obtained by solving a corresponding quantum ground state problem, as we have seen, which means that it
can be solved using many powerful techniques of quantum mechanics (e.g., discretization, mesh, or base function methods).\textsuperscript{45} For nonreversible diffusions, the problem is more complicated: there is no mapping to the quantum problem and the full spectral problem that must be solved involves, as mentioned, the tilted generator and its dual, with non-trivial boundary conditions imposed on the product of their respective eigenfunctions. An alternative method is to construct the driven process using optimal control representations detailed in Ref. 16 or to discretize the underlying space to obtain a jump process which can then be studied using exact diagonalization or the density-matrix renormalization techniques developed in Refs. 46–48.

For jump processes, the tilted generator becomes the tilted matrix

\[
W_k(x, y) = W(x, y) + k \, \mathbb{1}_S(x) \delta_{x,y},
\]

where \(W(x, y)\) is the transition rate (probability per unit time) for the transition \(x \rightarrow y\), and \(\delta_{x,y}\) is the Kronecker symbol. Moreover, the driven process is then the jump process with modified transition rates given by

\[
W_k(x, y) = r_k(x)^{-1} W(x, y) r_k(y),
\]

where \(r_k\) is the eigenvector associated with the dominant eigenvalue \(\lambda(k)\) of \(W_k\).\textsuperscript{15}

This result suggests many possible applications of the occupation conditioning problem beyond diffusions, including, for example,

- Chemical reactions producing abnormally high or low concentrations of chemical species because of thermal noise. In this case, the state \(X_t\) is the vector \((n_1^t, n_2^t, \ldots, n_m^t)\) of concentrations in time for \(m\) chemical species so that \(X_t \in \mathbb{N}^m_+\) or \(X_t \in \mathbb{R}^m_+\).\textsuperscript{1,2}
- Queues in which the number \(X_t \in \mathbb{N}_+\) of waiting “customers” goes beyond a certain threshold such as the queue capacity, see, e.g., Refs. 9, 49, and 50.
- Random walks on regular or random graphs that visit “rare” or “metastable” nodes or graph components (e.g., nodes with low pagerank).\textsuperscript{10–13} In this case, \(X_t\) is simply the node visited at time \(t\) while the state space is the set of nodes.
- Interacting particle systems on lattice, such as the zero-range process, showing condensation transitions where a macroscopic number of particles get to occupy one lattice site.\textsuperscript{51–54} The dynamics leading to this condensation and metastable phases related to it have been studied using occupation conditioning in Refs. 54–56.
- Other general Markov processes having metastable states, see, e.g., Refs. 57–59 and references therein. The occupation set \(S\) defining the conditioning can be chosen to include one or more metastable states, or a set of states connecting stable and metastable states so as to study transition pathways, also called reactive paths.

In all cases, the driven process provides a way to understand the dynamics of a stochastic process as it evolves in atypical states (concentrations, nodes, regions, etc.). This can take the form of a chemical reaction with modified rates, as already mentioned, or a queue with modified arrival and serving rates leading to a specific mean occupation. Similar interpretations apply to the other applications listed above and should yield new insights in understanding in general how large fluctuations arise in time and how they can be simulated efficiently.

ACKNOWLEDGMENTS

We thank Satya Majumdar for useful comments. Financial support for this work was received from the National Institute for Theoretical Physics (Postdoctoral Fellowship), Stellenbosch University (Project Funding for New Appointee), and the National Research Foundation of South Africa (Grant Nos. 90322 and 96199).
APPENDIX: LARGE DEVIATION PRINCIPLE FOR THE OCCUPATION

The contraction principle of large deviation theory states the following.\textsuperscript{22–24} Let $A_T$ be a random variable satisfying the LDP

$$P(A_T = a) = e^{-T I_A(a) + o(T)}$$

with rate function $I_A(a)$ and let $B_T$ be another random variable such that $B_T = C(A_T)$. Then $B_T$ also satisfies a LDP,

$$P(B_T = b) = e^{-T I_B(b) + o(T)},$$

with rate function

$$I_B(b) = \min_{a: C(a) = b} I_A(a).$$

This is called the contraction principle because the function $C(a)$ can be many-to-one, in which case the fluctuations of $A_T$ are “contracted” to the fluctuations of $B_T$.

To use this result for the occupation $R_T$, as defined in \textsuperscript{(5)}, we consider the empirical density

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

which represents the fraction of time (in the density sense) that $X_t = x$ over the time interval $[0, T]$. It is known from the Donsker-Varadhan theory (see Refs. \textsuperscript{22–24}) that the random function $\rho_T$ satisfies the LDP

$$P(\rho_T = \rho) = e^{-T I_\rho(\rho) + o(T)},$$

with the rate function given indirectly by the minimization shown in \textsuperscript{(10)}. Given that

$$R_T = \int_{\mathbb{R}^m} \rho_T(x) \, \mathbb{1}_S(x) \, dx = C(\rho_T),$$

we then obtain from the contraction principle \textsuperscript{(A3)} the result shown in \textsuperscript{(9)}.

---
