Linear response theory and transient fluctuation relations for diffusion processes: a backward point of view

Fei Liu\textsuperscript{1}, Huan Tong\textsuperscript{2}, Rui Ma\textsuperscript{1} and Zhong-can Ou-Yang\textsuperscript{1,2}

\textsuperscript{1} Center for Advanced Study, Tsinghua University, Beijing 100084, People’s Republic of China
\textsuperscript{2} Institute of Theoretical Physics, The Chinese Academy of Sciences, PO Box 2735, Beijing 100080, People’s Republic of China

E-mail: liufei@mail.tsinghua.edu.cn

Received 27 July 2010, in final form 29 September 2010
Published 18 November 2010
Online at stacks.iop.org/JPhysA/43/495003

Abstract

A formal apparatus is developed to unify derivations of the linear response theory and a variety of transient fluctuation relations for continuous diffusion processes from a backward point of view. The basis is a perturbed Kolmogorov backward equation and the path integral representation of its solution. We find that these exact transient relations could be interpreted as a consequence of a generalized Chapman–Kolmogorov equation, which intrinsically arises from the Markovian characteristic of diffusion processes.

PACS numbers: 05.70.Ln, 02.50.Ey, 87.10.Mn

1. Introduction

One of the important developments in nonequilibrium statistical physics in the past two decades is the discovery of a variety of fluctuation theorems or fluctuation relations [1–12]. These relations were usually expressed as exact equalities about statistics of entropy production or dissipated work in dissipated systems. In near-equilibrium region, these fluctuation relations reduce to the fluctuation–dissipation theorems (FDTs) [13, 14]. Hence they are also regarded as a nonperturbative extension of the FDTs in far-from-equilibrium regions [1, 4, 15]. Analogous to many new findings in physics, the mathematical techniques for proving these fluctuation relations have been available for many decades. For instance, thanks to the work of Lebowitz and Sphon [4] and Hummer and Szabo [16], we know that in Markovian stochastic dynamics these theorems or relations have an intimate connection with the Kolmogorov backward equation [17] (1931) and applications of the famous Feynman–Kac [18, 19] (1948) and Girsanov formulas [20, 21] (1960). The involvement of the backward equation or more precisely, its perturbed versions, is not occasional. Previously, many studies have proved...
Figure 1. Schematic diagram of the backward and forward points of view. The solid and dashed lines represent the original process and its reversal, respectively.

that various fluctuation theorems or relations arise from the differences between the original stochastic system and its time reversal [1–12, 22] (an excellent synthesis from this point of view is given in the article by Chetrite and Gawedzki [23]). The standard backward equation concerns how the system evolves in a given state or subset at future time \( t \) from the past time \( t' \). Intriguingly, this past time \( t' \) also plays the role of time in the reversed process of the original one between times 0 and \( t \), e.g., defining \( s = t - t' \) in figure 1. Hence, the backward rather than forward equation or the Fokker–Planck equation naturally accounts for the time-reversed characteristics of a stochastic system. Actually, this idea was implied in looking for conditions for the detailed balance principle of homogeneous Markov stochastic systems [24, 25]. Throughout this work, we call a discussion using the past time \( t' \) ‘backward’ to distinguish from a conventional ‘forward’ discussion using the future time \( t \).

Although the fluctuation relations are of importance and extensive attention was paid to them in past two decades, there was little work concerning this connection for a long time. The reasons may be twofold. On the one hand, physicists are not very familiar with the backward equation compared to the Fokker–Plank equation. The introduction of the backward equation in classic textbooks [24, 25] usually concerned its equivalence with the forward equation. Its application is solely limited to the first passage time or exit problems. On the other hand, as mentioned previously, time reversal is very relevant to the fluctuation relations. Almost all of them could be evaluated by a ratio of the probability densities of observing a stochastic trajectory and its reverse in a stochastic system and its time reversal, respectively [3, 8, 9, 12, 26]. Hence physicists familiar with quantum physics may favor the direct path integral approach [27, 28]. Recently, some studies began to investigate and exploit the connection between the fluctuation relations and the backward equation [23, 29–31]. For instance, Ge and Jiang [29] employed a perturbed backward equation and the Feynman–Kac formula to reinvestigate Hummer and Szabo’s earlier derivation [16] of the Jarzynski equality [6, 7] from mathematical rigors. They obtained a generalized multidimensional version of this equality. Using an abstract time reversal argument, Chetrite and Gawedzki [23] established an exact fluctuation relation between the perturbed Markovian generator of a forward Markovian process and the generator of its time-reversed process, though they did not explicitly claim a perturbed backward equation. Inspired by Ge and Jiang’s idea, we obtained two time-invariable integral identities for very general diffusion and discrete jump processes, respectively [30, 31]. Considering that several transient integral fluctuation relations [6, 7, 10–12] are their path integral representations in specific cases, we called the two identities the generalized integral fluctuation relations (GIFRs)\(^3\). Our further analysis showed that these GIFRs had well-defined time reversal explanations that are consistent with those achieved by Chetrite and Gawedzki [23]. Hence, their detailed versions or the transient detailed

\(^3\) We use ‘relation’ instead of ‘theorem’ used previously by the authors [30, 31] to avoid confusion with the fluctuation theorems specifically named for the nonequilibrium steady state [1–4].
fluctuation relations should be easily established. In addition to simplicity in evaluation, the most impressive point of using perturbed backward equations is that a specific time reversal is defined naturally and explicitly given a specific integral fluctuation relation. In particular, the latter could be designed ‘freely’ from these GIFRs. This contrasts with the conventional direct path integral approach (including [23]), which usually first requires a specific time reversal and then obtains a specific integral fluctuation relation. Previous work showed that the definition of time reversal may be nontrivial, e.g., in the Hatano–Sasa equality [10].

The aims of this work are twofold. First, we attempt to give a comprehensive version of our previous work about the transient fluctuation relations in continuous diffusion processes [30]. In addition, many details that were missed or only very briefly reported previously will be addressed, which include the classification of the existing integral fluctuation relations and time reversals, and derivation of the generalized detailed balance condition from the point of view of the GIFR. We also present several new theoretical results. The most significant progress is finding that the time-invariable integral identity we obtained previously is a generalized Chapman–Kolmogorov equation in diffusions; the path integral representation of the well-known Chapman–Kolmogorov equation may be regarded as the first integral fluctuation relation. Additionally, we uniformly obtain the GIFR for degenerate- and nondegenerate-type diffusions by employing a limited Girsanov formula (see appendix A). In previous work [4, 23] the fluctuation relation for the latter type was considered individually. Our second aim is to show that there is an alternative way of using the backward equation to derive the classical linear response theory [13, 14]. Furthermore, a simple extension of this ‘lost’ approach results in the transient fluctuation relations found almost 40 years later. Although it is widely accepted that the fluctuation relations reduce to the linear response theory when they are approximated linearly near equilibrium [1, 4, 15, 23], one may see a significant difference between their derivations: in textbooks [24] the linear response theory always starts from an evaluation of a probability distribution function using the time-dependent perturbation theory, whereas the fluctuation relations did not use this function at all. This discrepancy could be obviously diminished if one employs the backward equation to evaluate the linear response of a perturbed system at the very beginning. Moreover, this reevaluation directed our attention to the importance of the Chapman–Kolmogorov equation. We are tempted to wonder whether the dominated forward idea using the forward equation postpones the findings of the transient fluctuation relations.

The organization of this work is as follows. In section 2, we first introduce essential notions about the continuous diffusion process, three important formulas, and a backward interaction representation. Inspired by the characteristic of the novel representation, we then propose a generalized Chapman–Kolmogorov equation. On the basis of this equation, in section 3, we derive the linear response theory from a backward point of view and obtain a formally nonperturbative relation. To overcome a drawback of the relation, we investigate a general version of the GIFR which is mainly investigated in section 4. This section includes a time reversal explanation of the GIFR, and a classification of the transient fluctuation relations and time reversals in the literature from that point of view. In section 5, we present another kind of integral fluctuation relation, the Girsanov equality. On the basis of the time reversal interpretation of the GIFR, in section 6 we derive a generalized detailed balance condition. In section 7, we illustrate a specific GIFR in two typical physical systems by analytical and numerical methods, respectively. In section 8 the summary is given. Because our work mainly focuses on the unification of linear response theory and a variety of transient fluctuation relations, little physics is discussed here. We expect that this apparent shortage could be partially remedied by our physical examples and extensive physical discussion in the literature [32–34]. In addition, we must also admit that this work involves many formal
evaluations, the validity of which is not explored rigorously from a mathematical point of view. Such effort is of course interesting and valuable. However, we do not think it is a serious obstacle to justify our current results in terms of their generality and physical relevance.

2. Elements of stochastic diffusion processes

We begin with some notions about stochastic processes. We consider a general $N$-dimensional stochastic system $x = \{x_i\}, i = 1, \ldots, N$, described by a stochastic differential equation (SDE) [25]

$$\text{d}x(t) = A(x, t) \text{d}t + B^{1/2}(x, t) \text{d}W(t), \quad (1)$$

where $\text{d}W$ is an $N$-dimensional Wiener process, $A = \{A_i\}$ denotes an $N$-dimensional drift vector and $B^{1/2}$ is the square root of a $N \times N$ semipositive definite and symmetric diffusion matrix $B$:

$$B = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}, \quad (2)$$

where $B$ is a $M \times M$ ($M \leq N$) positive definite submatrix. We call a stochastic process nondegenerate type for $M = N$, and degenerate type otherwise; the Smoluchowski [35] and Kramers equations [36] are their typical representatives. One usually converts the SDE into two equivalent partial differential equations of the transition probability density $G(x, t|x', t')$ $(t > t')$: the Kolmogorov forward equation of the Fokker–Planck equation

$$\partial_t G = \mathcal{L}(x, t) G = \left[ -\partial_{x_i} A_i(x, t) + (1/2) \partial_{x_i} \partial_{x_l} B_{il}(x, t) \right] G \quad (3)$$

and the Kolmogorov backward equation

$$\partial_{t'} G = -\mathcal{L}^+(x', t') G = \left[ A_i(x', t') \partial_{x_i} + (1/2) B_{il}(x', t') \partial_{x_i} \partial_{x_l} \right] G. \quad (4)$$

Both the initial condition (3) and final condition (4) are $\delta(x - x')$, respectively. We follow Ito’s convention for the SDE and use Einstein’s summation convention throughout this work unless explicitly stated. Different from the forward equation, (4) is about the past time $t'$, and generally $G(x, t|x', t')$ does not have a probability interpretation with respect to the variable $x'$. The connection between the forward and backward equations may be seen from the famous Chapman–Kolmogorov equation [25]

$$G(x_2, t_2|x_1, t_1) = \int \text{d}x G(x_2, t_2|x, t') G(x, t'|x_1, t_1), \quad (5)$$

with $t_1 \leq t' \leq t_2$. An alternative expression is its derivative with respect to time $t$,

$$0 = \partial_{t'} \left[ \int \text{d}x G(x_2, t_2|x, t') G(x, t'|x_1, t_1) \right]. \quad (6)$$

Equation (6) implies that the operators $\mathcal{L}$ and $\mathcal{L}^+$ are adjoint to each other if one substitutes the time derivatives on the right-hand side with the forward and backward equations. Conversely, through the same equation we can also obtain the backward (forward) equation using the adjoint characteristic of the operators if the forward (backward) equation is known first. Intriguingly, the trivial (6) gives an alternative understanding of the Chapman–Kolmogorov equation (5), namely it is a consequence of the constant $\int \text{d}x G(x_2, t_2|x, t') G(x, t'|x_1, t_1)$ by choosing $t' = t_1$ and noting the initial condition of the transition probability density.
2.1. Three important formulas

There are three important formulas which are very useful in this work. The first is the Feynman–Kac formula, which was originally found by Feynman in quantum mechanics [18] and extended by Kac [19] to the stochastic process. The second formula is a limited version of the standard Girsanov formula [20]. In appendix A, we briefly introduce the Feynman–Kac formula and explain the limited Girsanov formula in more detail as it is not the standard content of textbooks [37]. The third important formula we used concerns the decomposition of the action of the operator $L$ on a multiplied function $Ef$.

Proposition 1.

$$L(Ef) = L'(Ef) - 2\partial_i [J_i(f)E],$$

(7)

where $E$ and $f$ are arbitrary functions, $J_i$ is a probability current components on the function $f$ defined by

$$J_i(f) = A_i(x, t) f - (1/2)\partial_x [B_i(x, t) f].$$

(8)

The proof is straightforward. We show that this simple formula plays a key role in constructing time reversal interpretation of the fluctuation relations. To the best of our knowledge, (7) is not present explicitly in the literature.

2.2. Backward interaction representation

Here, we present an uncommon representation to evaluate the average of an observable $O(x)$ in a stochastic system. Given the system’s probability density function $\rho(x, t)$ at time $t$, the average of the observable at time $t$ can be obtained by

$$\langle O \rangle(t) = \int dx' O(t|x', t') \rho(x', t'),$$

(9)

namely by integrating the dynamic observable weighted by the probability distribution at some time $t'$, where the dynamic observable is

$$O(t|x', t') = \int dx O(x|x', t)G(x, t|x', t'),$$

(10)

with $t' \leq t$. Obviously, choosing time $t' = t$, (9) is just the standard definition of the average. In order to make (9) available at general time $t'$, one must provide the evolution equations about $O(t|x', t')$ and $\rho(x, t')$ with respect to $t'$ simultaneously. It is easily seen that the former equation is just the backward equation (4) with a final condition $O(t|x', t) = O(x')$, while the latter is the forward equation (3) with initial condition $\rho(x, 0)$. Because this representation possesses the same sprint as that of interaction representation in quantum theory, except that here is about past time $t'$, we simply call it backward interaction representation. At a first glance, this representation is not very useful due to the undetermined dynamic observable and a probability density function. However, it is indeed very convenient and insightful when deriving linear response theory and unifying the various fluctuation relations.

2.3. Generalized Chapman–Kolmogorov equation

Our backward interaction representation immediately reminds us that differentiating the integral on the right-hand side of (9) with respect to time $t'$ is zero since the dynamic observable and the probability density function, respectively, follow the backward and forward equations as mentioned above. This point is essentially consistent with the Chapman–Kolmogorov
equation (6). Importantly, these past time-invariable integrals, revealed in (6) and (9), inspire us to propose a simple universal observation for any stochastic system, which is the central result of this work.

**Proposition 2.** Given function $D(t|x, t')$ satisfying a perturbed backward Kolmogorov equation

$$
\partial_t D(t|x, t') = -\mathcal{L}_a^x(x, t')D(t|x, t') - f^{-1}(x, t') \left[ \partial_x f - \mathcal{L}(f) \right](x, t')D(t|x, t') + f^{-1}(x, t') \left[ \mathcal{L}_o^x(g) - g \mathcal{L}_a^x \right](x, t')D(t|x, t'),
$$

with a final condition $D(t|x, t) = D(x)$, where $g(x, t')$ is the arbitrary smooth positive function, and the arbitrary operators $\mathcal{L}_o$ and $\mathcal{L}_a$ are adjoint to each other, the integral of the function weighted by a normalized positive $f(x, t)$ is past time invariable, namely

$$
\partial_x \int \text{d}x D(t|x, t') f(x, t') = 0.
$$

The proof is straightforward by substituting (11) into the above equation. Note that the first perturbation term in the first line is indispensable for (12) except for the function $f = \rho$, while the second perturbation term in the second line is not.

We immediately see that the Chapman–Kolmogorov equations (6) and (9) are the specific cases of the general (12). For instance, by choosing the function $f$ to be the system's probability density function $\rho$ and the final conditions $D(x) = O(x)$ and $\mathcal{L}_o = 0$, one obtains the latter equation. As mentioned above, $\mathcal{L}_o$ may be indeed nonzero. This point will be exploited in deriving the linear response theory. Because (12) is analogous to (6) and is important in the reminder of this work, we call it the generalized Chapman–Kolmogorov equation specifically.

We see shortly that both the linear response theory and a variety of fluctuation relations are the consequences of choosing different $f$, $g$ and the operator $\mathcal{L}_a$.

3. **Linear response theory from a backward point of view**

Evaluating the linear response of a system to an external perturbation is an essential ingredient of the FDTs [13, 14]. For the stochastic diffusion system, the conventional approach was based on the forward Fokker–Plank equation and applied the time-dependent perturbation theory [24, 38]. Here we show that the same result can be achieved using the perturbed Kolmogorov backward equation (11) and the Feynman–Kac formula. In particular, this revisiting inspires us to find a nonperturbative formal solution to the perturbation problem, which is very relevant to the later transient fluctuation relations.

We begin by setting some notations for a usual perturbation problem in a stochastic system [24]. Assume that the perturbed system (denoted by the subscript ‘p’) has a Fokker–Planck operator $\mathcal{L}_p(x, t) = \mathcal{L}_o + \mathcal{L}_c$, where $\mathcal{L}_o$ and $\mathcal{L}_c$ are the unperturbed (denoted by the subscript ‘o’) and perturbed (denoted by the subscript ‘c’) parts, respectively, and that the perturbation is applied at time $0$. For the sake of generality, the unperturbed system may be stationary or nonstationary, and the perturbation is arbitrary. Further assuming the probability distribution functions of the unperturbed and perturbed systems to be $\rho_o(x, t)$ and $\rho_p(x, t)$, respectively, they are consistent at initial time $0$. The aim of the linear response theory is to evaluate the average of an observable $O(x)$ in the perturbed system in terms of the unperturbed system’s knowledge, which can be realized using the backward interaction representation that we introduced in section 2.2. To do so, we introduce the dynamic observable $O_p(t|x, t')$ for the perturbed system, which follows the Kolmogorov backward equation (4) with an adjoint operator $\mathcal{L}_p^+(x, t') = \mathcal{L}_o^+ + \mathcal{L}_c^+$.  

J. Phys. A: Math. Theor. 43 (2010) 495003  F Liu et al
Instead of the conventional perturbation technique to solve the perturbed $O_p(t|x, t')$ (appendix B), here we make use of the previous observation that there are other functions having the same average while following different evolution equations, e.g., the function $O'_p(t|x, t')$ satisfying the perturbed backward equation (11) with $L_o = L_e, f = g = \rho_p$ and the final condition $O'_p(t|x, t) = O(x)$. Under this circumstance, its evolution equation may be rewritten as

$$\partial_t O'_p(t|x, t') = -L'_o(x, t') O'_p(t|x, t') - \rho^{-1}_p L_e(\rho_p)(x, t') O'_p(t|x, t').$$

(13)

According to the Feynman–Kac formula and generalized Chapman–Kolmogorov equation (12), we immediately obtain its path integral representation.

**Proposition 3.**

$$\langle O \rangle_p(t) = \left\{ \exp \left[ \int_0^t \rho^{-1}_p L_e(\rho_p)(\tau, x(\tau)) d\tau \right] O(x(t)) \right\}_o,$$

(14)

where the average on the right-hand side is over the trajectories starting from the initial distribution function $\rho_0(x, 0)$ and determined by the unperturbed stochastic process.

**Remark 1.** There is an analogous identity if we interchange the subscripts ‘$p$’ and ‘$o$’ and change the sign before $L_e$ into minus in (14) since the identification between the perturbed and unperturbed systems is arbitrary in physics.

Although the ‘nonperturbative’ (14) always holds and is independent of the strength of the perturbation, it in fact provides little knowledge about the perturbed nonequilibrium process because of the unknown probability density function $\rho_p$. However, there is an exception when the external perturbation is very small. Then one may approximate $\rho_p$ with the unperturbed one and expand the exponential to the first order obtaining

$$\langle O \rangle_p(t) \approx \langle O \rangle_o(t) + \int_0^t d\tau \left[ \rho^{-1}_o L_e(\rho_o)(\tau) O(t) \right]_o,$$

(15)

which is the result of the linear response theory [24]. To make it more available in practice, one still needs further effort to eliminate $\rho_o$, which results into a variety of FDTs in equilibrium [13, 14] or nonequilibrium regions [39–44]. The interested reader may refer to appendix C, where we present a relatively concise derivation for them and obtain more general formulas. (We are not very clear whether the above results are still available for very interesting nonergodic stochastic systems, e.g. [45–48].)

4. Generalized integral fluctuation relation

To avoid the fatal drawback of the unknown $\rho_p$ above, a strategy is to apply some functions that are known or well understood. Whether such effort is valuable depends on what we can obtain from it, e.g., achieving new evaluation approaches or physical understanding about stochastic processes. To do so, we return to the general perturbed backward equation (11) with the simplest but nontrivial case: $L_o = \partial_{x_i}$ (analogous to the momentum operator in quantum theory) and $g = S_i$. Then (11) becomes

$$\cdots + 2 f^{-1}(x, t') \left[ \partial_{x_i} S_i(x, t') + S_i(x, t') \partial_{x_i} \right] D(t|x, t'),$$

(16)

where ‘...’ represents the first line in the original equation, $N$-dimensional vector $S = \{S_i\}$ is required to be zero at the boundaries of the system or to be periodic if the system is periodic. On the basis of the generalized Chapman–Kolmogorov equation (12), Feynman–Kac and limited
Girsanov formulas (appendix A), for a certain vector $S$ whose last $(N - M)$ components vanish we obtain

**Proposition 4.**

$$
\left[ e^{-\int_0^t \mathcal{J}[f, S][x(\tau), \tau) d\tau} \right] D(x(t)) = \langle D(t) \rangle,
$$

(17)

where the integrand is

$$
\mathcal{J}[f, S] = f^{-1} \left[ (\mathcal{L} - \partial_x) f + 2\partial_x S \right] + R[-2 f^{-1} S]
\quad = f^{-1} \left[ (\mathcal{L} - \partial_x) f + 2\partial_x S + 2 f^{-1} S_i (B^{-1})_{ji} (S_j) \right] + 2 f^{-1} S_i (B^{-1})_{ji} (v_j - A_j),

(18)

the inverse of $B$ is formally defined by

$$
B^{-1} = \begin{bmatrix}
B^{-1} & 0 \\
0 & 0
\end{bmatrix},

(19)
$$

the average on the left-hand side is over the trajectories starting from the initial distribution function $f(x, 0)$ and determined by the stochastic process (1), and the average on the right-hand side is performed over the distribution $f(x, t)$.

We call (17) as the GIFR that is obviously more general than the previous one limited in the nondegenerate-type diffusions [30]. Note that time 0 in this relation may be indeed replaced by any time $t' (\leq t)$.

### 4.1. Time reversal interpretation

As mentioned at the very beginning, the backward equation has a natural connection with time reversal. In this section, we will show how to realize it for (16) concretely.

We start from the following observation. Multiplying both sides of (16) by $f(x, t')$ and doing a rearrangement, we obtain

$$
\partial_t \left[ D(t|x, t') f(x, t') \right] = -f(x, t') \mathcal{L}^+ D(t|x, t') + \mathcal{L}(f)(x, t') D(t|x, t')
\quad + 2 \partial_x [S_i (x, t') D(t|x, t')].

(20)
$$

Compared with (7), we see that if $S_i$ is the probability current $J_i(f)$, the right-hand side of the above equation is just $-\mathcal{L}[D(t|x, t') f(x, t')]$. Defining a new time parameter $s = t - t'$ further, we immediately obtain a Fokker–Planck equation about the multiplied function $D(t|x, t') f(x, t')$ with a Fokker–Planck operator $\mathcal{L}(x, t - s)$. Namely, we find the time reversal origin of (12) with the specific $S_i$. For a general case, however, this argument is yet inadequate since $J_i(f)$ is uniquely determined by the drift vector $A$, diffusion matrix $B$ and function $f$, which prompts us to propose a general time reversal given below.

Assume coordinates $x$ of the stochastic system to be even or odd, according to their rules under a time reversal: if $x_i \to +x_i$ is even and $x_i \to -x_i$ is odd; in abbreviation $x_i \to \tilde{x}_i = \varepsilon_i x_i$ and $\varepsilon_i = \pm 1$. The drift vector splits into ‘irreversible’ and ‘reversible’ parts:

$$
A = A^{\text{irr}} + A^{\text{rev}}. \quad \text{Under a time reversal, these vectors and the diffusion matrix are assumed to be transformed into } \tilde{A} = \tilde{A}^{\text{irr}} + \tilde{A}^{\text{rev}} \text{ and } \tilde{B}, \text{ respectively, where}
$$

$$
\tilde{A}_i^{\text{irr}}(x, t') = \varepsilon_i A_i^{\text{irr}}(x, s),
\tilde{A}_i^{\text{rev}}(x, t') = -\varepsilon_i A_i^{\text{rev}}(x, s),
\tilde{B}_{ij}(x, t') = \varepsilon_i \varepsilon_j B_{ij}(x, s).

(21)
$$

No summation over repeated indices here.
Proposition 5. Equation (16) is equivalent to a time-reversed forward Fokker–Planck equation

\[ \partial_s p(\tilde{x}, s) = L_R(\tilde{x}, s)p(\tilde{x}, s) \]

\[ = \left[ -\partial_{\tilde{x}_i} \tilde{A}_i(\tilde{x}, s) + (1/2)\partial_{\tilde{x}_i} \partial_{\tilde{x}_l} \tilde{B}_{il}(\tilde{x}, s) \right] p(\tilde{x}, s), \]

(22)
given a splitting

\[ A^{irr}_i(x, t'| f, S) = f^{-1}(x, t') \left[ S_i(x, t') + (1/2)\partial_{\tilde{x}_i} (B_{il} f)(x, t') \right], \]

(23)
\[ A^{rev}_i(x, t'| f, S) = A_i(x, t') - A^{irr}_i(x, t'| f, S), \]

and

\[ p(\tilde{x}, s) = \left[ \int D(t|\tilde{x}', t)f(\tilde{x}', t) d\tilde{x}' \right]^{-1} \]

\[ D(t|x, t') f(x, t'). \]

(24)
The proof is straightforward by substituting (24) into the time-reversed Fokker–Planck equation (22).

We give several remarks about this proposition.

Remark 2. The time-reversed transformations of the drift vector and diffusion matrix (21) are actually an inhomogeneous extension of homogeneous diffusion case [24]. Given a time-reversed drift vector \( \tilde{A}_i \), the fact can be seen by rewriting the splitting (23) as

\[ A^{irr}_i(x, t') = [A_i(x, t') + \varepsilon_i \tilde{A}_i(\tilde{x}, s)]/2, \]

(25)
\[ A^{rev}_i(x, t') = [A_i(x, t') - \varepsilon_i \tilde{A}_i(\tilde{x}, s)]/2. \]

Remark 3. Equation (23) states that the arbitrary vector \( S \) may be regarded as an irreversible probability current on the function \( f \) with the irreversible drift (23) or (25). For the former we denote \( S_l = J^{irr}_l(f|S, f) \) to indicate that the irreversible drift may depend on \( S \) and \( f \).

Remark 4. Equation (24) gives an alternative understanding about the generalized Chapman–Kolmogorov equation (12): the integral on its left-hand side is proportional to the total probability of \( p(\tilde{x}, s) \), which is time invariable with respect to time \( s \) (or \( t' \)).

Remark 5. Given \( S = \mathbf{J}^{irr}(f) \), the conditions on the diffusion matrix and drift vector for a time-reversible homogeneous Fokker–Planck equation (\( \tilde{A}^{irr} = A^{irr}, \tilde{A}^{rev} = A^{rev}, \) and \( \tilde{B} = B \)) to have a stationary equilibrium solution \( \rho^{eq}(x) \) that satisfies the detailed balance principle [24, 25] are identical to the requirement that \( f = \rho^{eq}(x) \) and the other terms except for \( L^* \) on the right-hand side of (16) vanish, respectively.

4.2. The GIFR and existing integral fluctuation relations

To demonstrate previous formal evaluation valuable, in this section we show several specific cases of the GIFR (17) that are relevant with existing integral fluctuation relations [6, 7, 10–12]. One may see that this general relation provides a simple and clear way to understand and classify these fluctuation relations.

4.2.1. \( S = \mathbf{J}^{irr}(f) \) with natural splitting. This may be the most natural consideration if we already know a splitting of the drift vector independent of \( S \) and \( f \). Because the function \( f \) is indeed arbitrary in (24), one may specify a decomposition \( p(\tilde{x}, s) \propto 1 \times D_{(1)}(t|x, t') \) and the
new function $D_{1(1)}(t|x, t')$ still satisfies (16) except for $S = J^{\text{irr}}(1)$ therein. Due to the same $p(\dot{x}, s)$, these two decompositions have a simple connection
\[
D(t|x, t') = \frac{D_{1(1)}(t|x, t') \int D(t|x', t) f(x', t) \, dx'}{f(x, t') \int D_{1(1)}(t|x', t) \, dx'}.
\] (26)

This equation immediately results in a relationship between the functionals (18) of the path integral representations of $D(t|x, 0)$ and $D_{1}(t|x, 0)$.

**Proposition 6.**
\[
\int_{0}^{t} \mathcal{J}[f, J^{\text{irr}}(f)](x(\tau), \tau) \, d\tau = -\ln f(x(t), t) + \int_{0}^{t} \mathcal{J}(x(\tau), \tau) \, d\tau
\] (27)

and
\[
\mathcal{J}(1) = \partial_{x} \dot{A}_{x}^{\text{irr}} - \partial_{x} A_{x}^{\text{rev}} + 2 \dot{A}_{x}^{\text{irr}}(\mathbf{B}^{-1})_{i j} \dot{A}_{x}^{\text{irr}} + 2 \dot{A}_{x}^{\text{irr}}(\mathbf{B}^{-1})_{i j} (v_{j} - A_{j}^{\text{rev}}) = 2 \dot{A}_{x}^{\text{irr}}(\mathbf{B}^{-1})_{i j} (v_{j} - A_{j}^{\text{rev}}) - \partial_{x} A_{x}^{\text{rev}} \ (S),
\] (28)

where $\dot{A}_{x}^{\text{irr}} = A_{x}^{\text{irr}} - \partial_{x} B_{i}/2$ and letter ‘S’ in the second line denotes that time integral of this equation is Stratonovich integral.

The proof is simple by noting that $\mathcal{J}(1) = \mathcal{J}[1, J^{\text{irr}}(1)]$ and the term $\ln f(x(t), t)$ is from the final condition of $D_{1}(t|x, t)$.

**Remark 6.** For another function $f$ having the same initial distribution with the stochastic system itself, namely $f(x, 0) = \rho(x, 0)$, (27) implies that
\[
\int_{0}^{t} \mathcal{J}[f, J^{\text{irr}}(f)](x(\tau), \tau) \, d\tau = \ln \rho(x(t), t) + \int_{0}^{t} \mathcal{J}[\rho, J^{\text{irr}}(\rho)](x(\tau), \tau) \, d\tau.
\] (29)

**Remark 7.** For any pair of very different smooth positive functions having only the same expressions at times 0 and $t$, the GIFRs are completely identical.

Compared with the original one, the function $D_{1}(t|x, t')$ is distinctive because its functional is completely determined by intrinsic characteristics of the system and environment, including the drift vector and diffusion matrix. An analogous expression was obtained earlier in [23] ((7.5) therein) by using an abstract time reversal argument. We must emphasize that (28) is more general than the previous one, because it also accounts for degenerate-type diffusions; see the example below.

**Example 1.** Consider a one-dimensional underdamped Brownian particle
\[
dr = \Pi \cdot \nabla H_{0} \, dt + \mathbf{F} \cdot \mathbf{d}r - \Gamma \cdot \mathbf{P} \, dt + \sqrt{2m/\beta} \mathbf{d}W,
\] (30)

where these vectors are $\mathbf{r}^{T} = [p, x]$, $\mathbf{P}^{T} = [p, 0]$, $\mathbf{F}^{T} = [F, 0]$, $\nabla^{T} = [\partial_{p}, \partial_{x}]$, the matrices are
\[
\Pi = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \gamma_{0} & 0 \\ 0 & 0 \end{bmatrix},
\] (31)

$x$ is a spatial position coordinate, $p$ is a momentum coordinate, $F$ is a nonconservative additive force, $\gamma_{0}$ is a constant friction coefficient unit mass, and $\beta^{-1} = k_{B} T$ with Boltzmann constant $k_{B}$ and coordinate-independent environmental temperature $T$. This model has an

---

4 Equation (16) in our previous work [30] is a specific case of (28) from this work, where the reversible drift was assumed to vanish. But we must point out that additional term $(1/2)(\mathbf{B}^{-1})_{i j} \partial_{x} (\mathbf{B}^{-1})_{i j}$ in the previous equation is a flaw we made in transforming from Ito integral to Stratonovich integral.
even spatial coordinate and an odd momentum coordinate. For a simple Hamiltonian
\( H_0 = p^2 / 2m + U(x, t) \), a canonical splitting is \([24, 25]\)
\[
A^{irr}(r, t) = -\gamma_0 \cdot P, \quad A^{rev}(r, t) = \Pi \cdot \nabla H_0 + F.
\] (32)
Then the \( p \)-component of the irreversible current on the function \( f \) is
\[
J_{irr}^p(f) = -\gamma_0 pf(p, x) - \partial_p [\beta^{-1} m \gamma_0 f(p, x)], \quad (33)
\]
and the \( x \)-component \( J_{irr}^x(f) \) vanishes. Therefore, the condition for the GIFR with degenerate
diffusion matrix is satisfied, and
\[
J(1) = -\beta d d\tau \left( \frac{-\partial_x U + F}{} \right).
\] (34)
If the temperature is a function with respect to the spatial coordinate, one may easily check
that the time integral of \( J(1) \) is (6.12) in \([4]\) which was called entropy flow from the system to
environment along a trajectory.

The physical meaning of functional (27) for \( f = \rho \) has been well acknowledged
\([11, 12, 23]\); the first and second terms are the Gibbs entropy production of the system
and the entropy production in the environment along a stochastic trajectory between times 0
and \( t \), respectively. Hence the GIFR (17) under this case is the integral fluctuation relation
of the overall entropy production. This relation also implies that for a diffusion process, the
mean overall entropy production is always nonnegative (the second law of thermodynamics).
We can see it by directly using the Jensen inequality to the GIFR with \( D(x) = 1 \) or evaluating
the mean instantaneous rate of overall entropy production, the latter of which is
\[
\langle J[f, J_{irr}(f)] \rangle = 2 \int d\rho \rho^{-1} [J_{irr}(\rho)] \beta^{-1} m \gamma_0 f(\rho, x) \geq 0 \quad (35)
\]
since all the other terms in (18) vanish after the ensemble average (the last term due to the
definition of Ito integral \([25]\)). In addition, considering that the ensemble average of the first
term over \( \rho \) in (29) is the relative entropy between the two distributions \( \rho \) and \( f \) that is always
nonnegative, we have

**Proposition 7.** The nonnegative \( \langle J[f, J_{irr}(f)] \rangle \) is minimum with respect to the function \( f \) if
and only if the function is the probability density function of the stochastic system itself.

### 4.2.2. \( S \) with a posterior splitting.
For an arbitrary vector \( S \), the results obtained in the previous case are still correct except that the irreversible and reversible drift components
are the functions of \( f \) and \( S \) through the splitting (23). Because such a splitting is defined
under these given functions, we roughly call it posterior. In the following we focus on the
simplest case: \( S = 0 \). Under this circumstance, functional (18) is simply
\[
J[f, 0] = f^{-1} (L - \partial_x) f, \quad (36)
\]
and the splitting is
\[
A^{irr}(x, t[f]) = \frac{1}{2 f(x, t)} \partial_x (B_{il} f)(x, t),
\]
\[
A^{rev}(x, t[f]) = A_i(x, t) - A^{irr}(x, t[f]). \quad (37)
\]
There are two integral fluctuation relations in the literature very relevant to (36).

**Example 2.** \( f \)-independent splitting. The splitting (37) usually depends on the given
function \( f \). Nevertheless, there is a very intriguing exception if we choose \( \rho^{eq}(x, t) \), which
is the transient stationary equilibrium solution of the stochastic system with time parameter fixed, namely
\[ \mathcal{L}(x, t) \rho_{eq}(x, t) = 0, \quad J^{\text{irr}}(\rho^{\text{eq}}) = 0. \] (38)

Obviously, the splitting (37) is now \(f\)-independent. In this case (36) is simply \(J[\rho^{\text{eq}}, 0] = -\partial_t \ln \rho^{\text{eq}},\) the time integral which was interpreted as dissipated work [6, 7]. We can easily see that the GIFR (17) with the final conditions \(D(x) = 1\) and \(\delta(x - z)\) is the celebrated Jarzynski equality [6, 7] and the key (4) in Hummer and Szabó’s work [16], respectively.

**Example 3.** \(f\)-dependent splitting. A famous example of virtually \(f\)-dependent splitting in the literature is for the stochastic system having a transient nonequilibrium steady state [10]
\[ \mathcal{L}(x, t) \rho_{ss}(x, t) = 0, \quad J(\rho_{ss}) \neq 0. \] (39)
The time reversal corresponding the splitting (37) with \(f = \rho_{ss}\) was also called current reversal [23]. In this case (36) is the same as the previous case in formality. The time integral of the functional was called the excess heat production and the GIFR with \(D(x) = 1\) is the Hatano–Sasa equality [10].

In addition to the above two well-known integral fluctuation relations, (36) also reveals several simpler relations.

**Example 4.** The most obvious case is to choose \(f = \rho\), the probability density function of the stochastic system itself and \(J[\rho, 0] = 0\) simply. Correspondingly, (16) reduces to the standard Kolmogorov backward equation (4) and the GIFR (17) is a trivial path integral representation of the standard Chapman–Kolmogorov equation (5) or (9) according to different final conditions. The corresponding time reversal or splitting (37) was called the complete reversal [23].

**Example 5.** In the perturbation problem in section 3, we may choose the stochastic system to be the unperturbed one \(\mathcal{L} = \mathcal{L}_o\) and \(f = \rho_p\) or the perturbed one \(\mathcal{L} = \mathcal{L}_p\) and \(f = \rho_o\) as discussed previously, (36) then becomes
\[
\begin{align*}
J[\rho_p, 0] &= \rho_p^{-1}(\mathcal{L}_o - \partial_t \rho_p) = -\rho_p^{-1} \mathcal{L}_e(\rho_p), \\
J[\rho_o, 0] &= \rho_o^{-1}(\mathcal{L}_p - \partial_t \rho_o) = -\rho_o^{-1} \mathcal{L}_e(\rho_o).
\end{align*}
\] (40)

respectively. We immediately see that the corresponding GIFRs are (14) and those mentioned in remark 1, respectively. Namely, we give an alternative approach to obtain the linear response theory. Although these functionals or identities look very similar, their time reversals are significantly distinct. Let us consider a simple situation in which the unperturbed system is in equilibrium \(\rho_0^\text{eq}(x)\) and a perturbation \(A_e(x, t)\) is imposed on the original drift \(A_o(x) = A_0^\text{rev}(x) + A_0^\text{irr}(x)\) as usual. For the first choice the posterior splitting (37) is apparently \(\rho_p\)-dependent and we usually do not know their concrete expressions because of the unknown probability density function. In contrast, for the second choice, because \(\rho_0^\text{eq}(x)\) satisfies the detailed balance condition, the splitting (37) is simply
\[
A^{\text{irr}}(x, t | \rho_o) = A_p^\text{irr}(x), \quad A^{\text{rev}}(x, t | \rho_o) = A_o^\text{rev}(x) + A_c(x, t).
\] (42)

This is a new example with vanishing \(S\) and \(f\)-independent time reversal. Interestingly, whatever the perturbation is, reversible or irreversible in physics, it is always classified into the reversible drift in the time-reversed system \(\mathcal{L}_R\) (22).
5. Girsanov equality

Recalling (18), one may note that any ensemble average of the term \( f^{-1} S_i (B^{-1})_{ij} S_j \) is always non-negative due to the semi-positive definite diffusion matrix \( B \). In fact, this observation has an alternative indirect explanation. Considering a perturbed forward Fokker–Plank equation

\[
\partial_t \rho' = L'(x, t) \rho' + 2 \partial_i [ f^{-1}(x, t) S_j S_i(x, t) \rho'(x, t)].
\]  

(43)

Employing the limited Girsanov formula, we may obtain an integral fluctuation relation.

**Proposition 8.**

\[
\langle e^{-\int_0^t R[-2f^{-1}S](x(\tau), \tau) \, d\tau} \rangle = \langle D' \rangle(t).
\]  

(44)

We simply call it the Girsanov equality. The Jensen inequality indicates that the ensemble average of the functional of the equality is non-negative. It is worth emphasizing that the averages of both sides of (44) are respectively over \( \rho'(x, 0) \) and \( \rho'(x, t) \) rather than the arbitrary function \( f \) as in the GIFR (17). Compared the Girsanov equality with the GIFR, we find

**Proposition 9.** Given vector \( S \) divergenceless, there exits a relation

\[
\int_0^t J[f, S](x(\tau), \tau) \, d\tau = \int_0^t J[f, 0](x(\tau), \tau) \, d\tau + \int_0^t R[-2f^{-1}S](x(\tau), \tau) \, d\tau.
\]  

(45)

Particularly, these three functionals satisfy their respective integral fluctuation relations, simultaneously.

This proposition is a general version of a specific case found by Speck and Seifert when they investigated nonequilibrium steady-state system [51]. Choosing the divergenceless vector to be \( S = J(\rho^{ss}) \) and \( f = \rho^{ss} \) the transient steady-state solution in (39), we immediately see the left-hand side of (45) is the overall entropy production functional (27); the first term on the right-hand side is the excess heat production functional in the Hatano–Sasa equality, and the last term is the housekeeping heat functional [52]. Therefore the integral fluctuation relation (44) is an identity about the housekeeping heat [51].

6. Generalized detailed balance condition

The path integral representation of the solution of (16) presents a relationship between \( D(t|x, t') \) with a general final condition and the one \( D(x_2, t_2|x_1, t_1) \) with a specific final condition \( \delta(x_1 - x_2) \), which is simply

\[
D(t|x_1, t_1) = \int dx_2 [e^{-\int_0^{t_2} \mathcal{J} d\tau} \delta(x(t_2) - x_2)] \times e^{\int_0^{t_1} \mathcal{J} d\tau} D[x(t)]
\]  

\[
= \int dx_2 D(t|x_2, t_2) D(x_2, t_2|x_1, t_1).
\]  

(46)

In the first line we inserted a \( \delta \)-function at time \( t_2 \) between times \( t_1 \) and \( t \), and the second line is a consequence of Markovian property. On the other hand, the probability density function of the time-reversed system (22) at time \( s_1 = t - t_1 \) can be constructed by the density function at earlier time \( s_2 = t - t_2 \), given the transition probability \( G_R \),

\[
p(\tilde{x}_1, s_1) = \int G_R(\tilde{x}_1, s_1|\tilde{x}_2, s_2) p(\tilde{x}_2, s_2) \, d\tilde{x}_2.
\]  

(47)

With (46), (47) and (24), we obtain
Proposition 10.
\[ G_R(\tilde{x}_1, s_1|\tilde{x}_2, s_2) f(x_2, t_2) = D(x_2, t_2|x_1, t_1) f(x_1, t_1). \] (48)

Remark 8. For a time-reversible homogeneous stochastic system, choosing \( S = J^{irr}(\rho) \) and \( f = \rho^{eq}(x) \), both the transition probability \( G_R(x, t|\tilde{x}', t') \) (48) then reduces to the principle of detail balance written in terms of conditional probabilities [25].

Chetrite and Gawedzki [23] have obtained an analogous expression and they called it ((7.15) therein) the generalized detailed balance relation. Starting from (48), one may obtain the generalized version of Crooks’ relation [8, 9]. Because this discussion has been given in great detail in [23], we turn to two concrete physical examples to illustrate a specific GIFR.

7. Analytical and numerical examples of the GIFR

In this section we examine a specific GIFR (17), with
\[ f = \rho, \quad S = \alpha J(\rho) \quad \text{and} \quad D(x) = 1, \] (49)
in two typical physical systems that have been performed experimentally [49, 53], where \( \alpha \) is an arbitrary number, \( \rho \) is the real probability density function of the systems and \( J \) is the corresponding probability current. Our consideration of the given GIFR is mainly for simplicity. In addition, this relation obviously includes the fluctuation relation of overall entropy production for \( \alpha = 1 \) and the standard Chapman–Kolmogorov equation for \( \alpha = 0 \) (example 4).

We begin by investigating the GIFR (17) for a one-dimensional overdamped Brownian particle that is in equilibrium in a harmonic potential with the spring constant \( k \) at time \( t = 0 \) and is dragged by a time-dependent external harmonic potential \( k(x - v^* t)^2/2 \), where \( v^* \) is the moving velocity of the potential [49]. The motion equation of the particle’s position is described by a SDE
\[ dx = -\mu_0 k(x - v^* t) \, dt + \sqrt{2k_B T \mu_0} \, dW. \] (50)
where \( \mu_0 \) is the particle’s constant mobility. In order to simplify the mathematics we resort to a dimensionless form for the SDE by measuring the position \( x \) and time \( t \) in units of \( \sqrt{k_B T/k} \) and \( (\mu_0 k)^{-1} \), respectively. Equation (50) then becomes
\[ dx = -(x - t) \, dt + \sqrt{2} \, dW. \] (51)
Here we specially set the dimensionless velocity \( v^* \) as 1, which should not apparently change the following evaluation and conclusion. The solution of this equation is simply
\[ x(t) = x(t') e^{-u(t-t')} + t - t' e^{-u(t-t')} - 1 + e^{-u(t-t')} + \sqrt{2} \int_{t'}^t e^u \, dW(\tau). \] (52)
Namely, the dragged Brownian particle is a simple Gaussian stochastic process and its probability density function is Gaussian [50]:
\[ \rho(x, t) = (\sqrt{2\pi})^{-1} \exp[-(x - t + 1 - e^{-t})^2/2]. \] (53)
Under this circumstance, the functional of the given GIFR is
\[ \int_{t'}^t \left\{ -2\alpha [x(\tau) - \tau + 1 - e^{-\tau}] (1 - e^{-\tau}) + \alpha^2 (1 - e^{-\tau})^2 \right\} d\tau + \sqrt{2} \alpha (1 - e^{-t}) \, dW(\tau). \] (54)
Figure 2. The specific GIFR \((49)\) in a nonequilibrium steady-state system that is realized by a one-dimensional overdamped Brownian particle driven by a constant external force \(F\) in a periodic potential \(V(x) = V_0 \cos(x)/2\) with a periodicity \(L\). In simulation the number of stochastic trajectories is 20,000, \(V_0 = 60k_BT\) and \(F = 60\pi k_BT/L\). (a) The GIFR as a function of the number \(\alpha\). We show the simulation results at times 10 and 200 in units \(L^2/4\pi^2 k_BT\mu\), respectively. The dashed line is the expected theoretical value. (b) The probability density functions \(P(h)\) of functional \((27)\) having value of \(h\). We choose \(t = 10\) and \(\alpha = -1, 1, 2\), respectively.

Because \(x(t)\) is a Gaussian process, the path integral representation of \(D(t|x, t')\) can be explicitly obtained by evaluating the expectation and variance of \((54)\). After a simple calculation, we find

\[
\ln D(t|x, t') = -\alpha (x - t')(2e^{-|t-t'|} - 2e^{-2|t-t'|} + e^{-|t-t'|}) \\
- \left(\alpha^2/2\right)(2e^{-|t-t'|} - 2e^{-2|t-t'|} + (\alpha - 2\alpha^2)(e^{-|t-t'|} - 2)e^{-|t-t'|}) \\
+ (2\alpha^2 - 3\alpha)e^{-|t-t'|} + (\alpha - \alpha^2/2)e^{-2|t-t'|} + (\alpha^2 - \alpha)(e^{-2|t-t'|} - 2e^{-|t-t'|} - 2),
\]

and particularly,

\[
D(t|x, t')\rho(x, t) = \left(\sqrt{2}\pi\right)^{-1} \exp\left[-\left[x - (t' + \alpha(e^{-t} - 2) e^{-|t-t'|}) \\
- (\alpha - 1)e^{-|t-t'|} + 2\alpha - 1\right]^2/2\right].
\]

Obviously, the Gaussian function in the above equations indicates that the integration of \((56)\) with respect to \(x\) is indeed equal to 1, i.e., we prove the specific GIFR \((17)\). We must emphasize that the presence of a Gaussian function on the right-hand side of the above equations is not occasional. In fact, as we claimed about the time reversal interpretation of the GIFR, \(D(t|x, t')\rho(x, t)\) is proportional to the probability density function \(\rho(x,s)\) of the time-reversed diffusion process \((22)\) with an initial distribution \(\rho(x, t)\), which follows a time-reversed SDE given by

\[
dx = -(x - s + 2(\alpha - 1)(1 - e^{-t+s})) ds + \sqrt{2}dW.
\]

One may easily see it in terms of the splitting \((23)\) and prove that \(\rho(x, t - t')\) is identical to \((56)\). A similar conclusion also holds for the Jarzynski equality investigated by Ge and Jiang \([29]\) in the same physical system.

The dragged Brownian particle is somewhat distinctive since its distribution is Gaussian even if the particle is in a far-from-equilibrium regime. We consider the second system whose probability density function is not Gaussian \([24]\). In this system, a one-dimensional
overdamped Brownian particle is driven by a constant external force $F$ in a periodic potential $V(x) = V_0 \cos(x)/2$ with a periodicity $L$ [53], the SDE of which is

$$d x = \mu_0 \left( V_0 \sin(x)/2 + F \right) dt + \sqrt{2k_B T \mu_0} dW.$$ \hspace{1cm} (58)

We are interested in the specific GIFR when the particle is in a nonequilibrium steady state. Differing from the former, the GIFR here cannot be evaluated analytically and we have to resort to a Langevian simulation. Figure 2(a) shows the results of the GIFR as a function of the number $\alpha$ at two times. We see that the agreement between the theoretical prediction and simulation is satisfactory, especially in the case of a shorter time. The reason for the larger deviation for $\alpha = 2$ is given in figure 2(b); compared with the other two cases, its probability density function of functional (27) is significantly skewed toward larger values. Note that there is a slight increase in the probability of $h$ with negative values.

8. Conclusion

In this work, we formally unify the derivations of the linear response theory and the transient fluctuation relations for Markov diffusion stochastic systems from a backward point of view using the perturbed Kolmogorov backward equations (11) and its path integral representation. The motivation of this reinvestigation of the linear response theory is that the conventional approach for the theory is based on the forward Fokker–Planck equation and time-dependent perturbation, which is not used in the evaluations of the transient fluctuation relations. Our results show that a derivation using the backward equation could be very simple and flexible even if the unperturbed Markovian system is nonstationary. Importantly, this study reminds us that the time-invariable integral identity we found previously is the generalization of the well-known Chapman–Kolmogorov equation. In addition, one may note that our evaluations heavily depend on the path integral representation of the partial differential equation. Only in this representation does the physics of the perturbed Kolmogorov backward equation emerge explicitly. Such a situation is analogous to the relationship between the Schrödinger equation and the Feynman path integral in quantum physics. Hence one might criticize that the partial differential equation is unnecessary because all of the above results could be evaluated by the direct path integral approach. This point is of course correct in principle. However, as mentioned at the very beginning, such a ‘bottom-up’ idea needs the known time reversal or splitting of the drift vector. Except for very simple or intuitive cases, finding a meaningful or novel time reversal or splitting is not a trivial task. It would be desirable if there were some rules or guides for this task. We think that the perturbed backward equation meets this requirement. This is also natural. After all, the fluctuation relations are the identities of ensemble statistic properties of a stochastic process. In a word, the roles played by these perturbed backward equations and their path integral representations are complementary. Considering that the generalized Chapman–Kolmogorov equation is the cornerstone of this work, which intrinsically arises from the Markovian characteristic of diffusion processes, we believe that the evaluations and results developed here should also be available to other Markovian stochastic processes, e.g., general discrete jump processes with continuous or discrete time.

Acknowledgments

This work was supported in part by Singular Basic Research Foundation and the National Science Foundation of China under grant nos 10547002 and 10704045.
Appendix A. The Feynman–Kac formula and limited Girsanov formula

The Feynman–Kac formula is a path integral representation of a partial differential equation \[18, 19\]. Assuming a partial differential equation

\[
\frac{\partial}{\partial t} u(x, t') = -L u(x, t') - g(x, t') u(x, t'),
\]

(A.1)

with a final condition \(u(x, t) = q(x)\), its solution has a path integral representation

\[
u(x, t') = \mathcal{F} \left( \exp \left[ \int_{t'}^t g(x(\tau), \tau) d\tau \right] q(x(t)) \right).
\]

(A.2)

where the expectation \(\mathcal{F}\) is an average over all trajectories \(\{x(\tau)\}\) determined by SDE (1) taken conditioned on \(x(t') = x\). Letting \(g = 0\) and \(q(x)\) be a \(\delta\)-function, the Feynman–Kac formula also gives a path integral representation of the backward equation (4).

Roughly speaking, the standard Girsanov formula is about the probability densities of observing the same trajectory \(\{x(\tau)\}\) between time \(t'\) and \(t\) in two different stochastic systems with the same diffusion matrix. This formula particularly requires the diffusion matrix to be nondegenerate. However, degenerate diffusion processes are generic in physics, e.g., the Kramers equation \[36\]. Therefore, one needs a revised Girsanov formula that is still valid for the degenerate diffusion case.

We start by writing out the probability density of a stochastic trajectory \(\{x(\tau)\}\) in the stochastic system (1) \[28\]:

\[
\mathcal{P}[\{x(\tau)\}] = \left[ \prod_{k>M} \delta(x_k - A_k) \right] \int \prod_{i=M}^N \mathcal{P} \left[ \eta_i \right] \exp \left[ -\frac{1}{2} \int_{t'}^t \eta_i \eta_i dx \right] \prod_{i=1}^M \delta(x_i - A_i - (B^{-1})_{ij} \eta_j),
\]

(A.3)

where \(\eta_i = dW_i/ds\) are standard white noises, and \(\delta\)-functions should be understood as a product of a sequence of terms on all times between \(t'\) and \(t\). The expression in the first square brackets on the right-hand side indicates that noises directly act on the first \(M\) coordinates only. Assuming that another stochastic system (denoted by prime) has a different drift vector \(A' = A + a\). Obviously, if there is any nonzero component \(a_k\) with \(k > M\), the \(\delta\) functions in the first square brackets make the ratio of the probability densities of the same trajectory in these two systems meaningless. In physics this means that we never observe the same trajectory in these two systems. In contrast, if nonzero components of \(a\) are restricted to the first \(M\) component, namely \(a_k = 0\) with \(k > M\), the ratio or the Radon–Nikodym derivative of these two probability densities is finite, which is given by

\[
\mathcal{P}'[\{x(\tau)\}] = \mathcal{P}[\{x(\tau)\}] e^{-\int_{t'}^t R[a](\tau, x(\tau)) d\tau},
\]

(A.4)

where the integrand is

\[
R[a] = \frac{1}{2} a_i (B^{-1})_{ij} a_j - a_i (B^{-1})_{ij} v_j - A_i,
\]

(A.5)

the summations are restricted to the first \(M\) component, \(v_i = dx_i/d\tau\), and the integral is defined by Ito stochastic integral \[25\]. We call (A.4) the limited Girsanov formula to distinguish it from the standard one. We may also conveniently rewrite this limited formula into the standard one by formally defining the inverse of the diffusion matrix in (19) and bear in mind the application condition.
Appendix B. Other backward approaches deriving the linear response theory

B.1. Standard perturbation method

The standard perturbation technique is to regard the last term in the adjoint operator \( L^+_o(x, t') = L^+_o + L^+_e \) as a small perturbation. We expand the dynamic observable to first order

\[
O_p(t|x, t') = O_o(t|x, t') + O_1(t|x, t') + \cdots, \tag{B.1}
\]

and impose the final conditions \( O_o(t|x, t) = O(x) \) and \( O_1(t|x, t) = 0 \). Substituting it into the backward equation (4), we obtain the zero- and first-order terms satisfying

\[
\begin{align*}
\partial_t' O_o(t|x, t') &= -L^+_o(x, t') O_o(t|x, t'), \\
\partial_t' O_1(t|x, t') &= -L^+_o(x, t') O_1(t|x, t') - L^+_e(x, t') O_o(t|x, t'),
\end{align*} \tag{B.2}
\]

respectively, and their solutions have path integral representations (e.g. theorem 7.6 in [54])

\[
\begin{align*}
O_o(t|x, t') &= \langle x(t) \rangle_o, \\
O_1(t|x, t') &= \langle \int_t^{t'} d\tau L^+_e(x, \tau) O_o(t|x(\tau), \tau) \rangle_o,
\end{align*} \tag{B.3, B.4}
\]

respectively. \( O_o(t|x, t') \) is obviously the dynamic observable in the unperturbed system. Then the linear approximation of the average of the observable is

\[
\langle O \rangle_p(t) = \langle O \rangle_o(t) + \int_0^t d\tau \int dx \rho_o(x, \tau) L^+_e(x, \tau) O_o(t|x, \tau) + \cdots
\]

\[
= \langle O \rangle_o(t) + \int_0^t d\tau \langle [\rho_0^{-1} L^+_e(\rho_0)](\tau) O(t) \rangle_o + \cdots, \tag{B.5}
\]

where we have used the adjoint characteristic of \( L^+_e \) in the second line.

B.2. A twisted Chapman–Kolmogorov equation

The second approach is more direct and interesting. Let us consider a ‘twisted’ Chapman–Kolmogorov equation

\[
\partial_t' \left[ \int dx O_p(t|x, t') \rho_o(x, t') \right] = - \int dx \rho_o(x, t') L^+_e O_p(t|x, t'). \tag{B.6}
\]

We must emphasize this is exact. Integrating both sides with respect to time \( t' \) from 0 to \( t \), we immediately see that the left-hand side is just the minus of the difference between the averages of the observable in the perturbed and unperturbed systems. If the first-order approximation was concerned about, namely the subscript ‘\( p \)’ is replaced with ‘\( o \)’ on the right-hand side of (B.6), we reobtain (15).

Appendix C. Several nonequilibrium fluctuation–dissipation theorems

The classical FDTs state that the linear response function of an equilibrium system to a small perturbation is proportional to the two-point time-correlation function of the unperturbed system [13, 14]. This topic is attracting considerable interest due to continuous efforts of extending the standard one to the nonequilibrium region [39–43]. Here, we briefly discuss two intriguing FDTs [42, 43] in two typical physical models in terms of derivation. We will
show that although these two theorems are nontrivial in physical interpretation, they may be regarded as simple applications of two general identities
\[
\partial_o [B_i(\partial_o E) f] = 2[\mathcal{L}(E f) - E \mathcal{L}(f) + (\partial_o E) J_i(f)] \tag{C.1}
\]
\[
= \mathcal{L}(E f) - E \mathcal{L}(f) + \mathcal{L}'(E) f, \tag{C.2}
\]
where \( E \) and \( f \) are the arbitrary functions. They should be implicitly applied in the previous studies. Intriguingly, these two equations result in the important (7) as well.

Overdamped Brownian motion. The SDE of a multidimensional overdamped Brownian particle is
\[
\mathbf{dx} = \mathbf{M}(x, t) [-\nabla U(x, t) + \mathbf{F}(x, t)] \, dt + \mathbf{B} \, d\mathbf{W}(t), \tag{C.3}
\]
where \( \mathbf{F} \) is a nonconservative additive force, the nonnegative mobility and diffusion matrices are related by \( 2\mathbf{M} = \beta \mathbf{B} \). We assume that perturbation is realized as usually by adding a time-dependent potential \(-h(t) V(x)\) to the original one \( U(x, t) \). Then the perturbed component \( \mathcal{L}_e \) is \(-h(t') \partial_o M_i(\partial_o V)\). Substituting it into (15), we obtain the response function
\[
R_O(t, \tau) = \delta(O)|_{p(t) \delta h(t')}|_{h=0} = \langle [\rho_o^{-1}(t) \partial_o [\rho_o M_i(\partial_o V)(t) O(t)]] \rangle_o. \tag{C.4}
\]
This expression seems very different from the standard FDT [14], even if the unperturbed system is in equilibrium. However, this difference is not intrinsic. Choosing \( \mathcal{L} = \mathcal{L}_o \) as the Fokker–Planck operator of (C.3) and \( E = V(x, t') \) and noticing that the left-hand side of (C.1) is just \( 2\mathcal{L}_e(\rho_o)/h(t')\beta \), we obtain two new expressions of (C.4) given by [42, 43]
\[
R_O(t, \tau) = \beta \frac{d}{dt} \langle V(\tau) O(t) \rangle_o - \beta \langle [\rho_o^{-1}(t) \partial_o [\rho_o V]](\tau) O(t) \rangle_o \tag{C.5}
\]
\[
= \frac{\beta}{2} \frac{d}{dt} \langle V(\tau) O(t) \rangle_o - \frac{\beta}{2} \langle \mathcal{L}_o^* V(\tau) O(t) \rangle_o. \tag{C.6}
\]
Although the FDT (C.5) still faces the difficulty of unknown probability distribution, \( \rho_o \) as (C.4), it intuitively indicates that the responses are different for the unperturbed systems prepared in equilibrium and nonequilibrium states; the latter has nonvanishing probability current. In contrast, the FDT (C.6) does not need this distribution and is more useful in practical simulation or experiment. The second term on the right-hand side was interpreted as a correlation with a dynamical activity [43].

Underdamped Brownian motion. The second model is one-dimensional underdamped Brownian motion (30) (no apparent differences in discussion for the multidimensional case). According to the types of the perturbations, several different FDTs under specific conditions may be obtained. The simplest case is that the perturbation is still through a potential \(-h(t) V(x)\) and \( \mathcal{L}_o = -h(t)(\partial_o V)\partial_o \). We can of course obtain a FDT as (C.4) by directly substituting \( \mathcal{L}_o \) into (15) (not shown here). In addition, one may expect that the left-hand side of (C.1) is still proportional to \( \mathcal{L}_e(\rho_o) \) as that in the overdamped case. This is indeed true if choosing \( E \equiv p \partial_o V(x) \). We then obtain
\[
R_O(t, \tau) = \frac{\beta}{\gamma_o m} \frac{d}{d\tau} \langle (p \partial_o V)(\tau) O(t) \rangle_o - \frac{\beta}{\gamma_o m} \langle [\rho_o^{-1} J_i(\rho_o) \partial_o (p \partial_o V)](\tau) O(t) \rangle_o \tag{C.7}
\]
\[
= \frac{\beta}{2\gamma_o m} \frac{d}{d\tau} \langle (p \partial_o V)(\tau) O(t) \rangle_o - \frac{\beta}{2\gamma_o m} \langle \mathcal{L}_o^* (p \partial_o V)(\tau) O(t) \rangle_o. \tag{C.8}
\]
These new FDTs seem to be very different from (C.5) and (C.6). For instance (C.7) is not as good as (C.5) in concept because the current \( J_i \) is not zero even if the unperturbed
system has canonical distribution (in equilibrium and $\mathbf{F} = 0$). Particularly, these FDTs cannot automatically reduce to the standard FDT \cite{14} in the deterministic Hamiltonian system by simply choosing $\gamma_0 = 0$. These problems could be avoided if one notes that the left-hand side of (C.1) vanishes for $E = V(x)$ (the same consequence given $\gamma_0 = 0$) and introduces a modified current

$$\mathbf{J}(\rho_0) = \mathbf{J}(\rho_0) + \beta^{-1} \Pi \nabla \rho_0.$$  

(C.9)

Then we obtain another FDT

$$R_O(t, \tau) = \beta \frac{d}{d\tau} \langle V(\tau) O(t) \rangle_0 - \beta \langle \left[ \rho^{-1}_0 \mathbf{J}_r(x_0) \right] (\tau) O(t) \rangle_0.$$  

(C.10)

This expression is the same as (C.5) and the last term vanishes for an unperturbed system having canonical distribution. We must emphasize that (C.10) also holds even if $\gamma_0$ is a function of the coordinate $x$.

For general perturbations that depend on spatial and momentum coordinates simultaneously, e.g. $-h(t)V(x, p)$ \cite{38}, the above FDTs usually do not hold. For instance, (C.10) is modified by an additional term

$$+ \beta m \gamma_0 \left[ \left[ \rho^{-1}_0 \mathbf{J}_x(\rho_0) - p/m \right] \partial_p V + \beta^{-1} \partial^2_p V \right] (\tau) O(t) \rangle_0.$$  

(C.11)

Finally, if one temporarily forgets the time derivative in these FDTs, we can obtain a more concise FDT

$$R_O(t, \tau) = \beta \left[ A_r \Omega_i \partial_i V \right] (\tau) O(t) \rangle_0 - \beta \langle \left[ \rho^{-1}_0 \mathbf{J}_r(\rho_0) \Omega_i \partial_i V \right] (\tau) O(t) \rangle_0.$$  

(C.12)

where the matrix $\Omega$ is $-\Pi (m \Gamma - \Pi)^{-1}$. One may easily prove that (C.12) and (C.10) are identical if the potential $V$ is a function of the spatial coordinate $x$ only or given $\gamma_0 = 0$.

References

[1] Evans D J, Cohen E G D and Morriss G P 1993 Phys. Rev. Lett. 71 2401
[2] Gallavotti G and Cohen E G D 1995 Phys. Rev. Lett. 74 2694
[3] Kurchan J 1998 J. Phys. A: Math. Gen. 31 3719
[4] Lebowitz J L and Spohn H 1999 J. Stat. Phys. 95 333
[5] Bochkov G N and Kuzovlev Yu E 1977 Sov. Phys.—JETP 45 125
[6] Jarzynski C 1997 Phys. Rev. Lett. 78 2690
[7] Jarzynski C 1997 Phys. Rev. E 56 5018
[8] Crooks G E 1999 Phys. Rev. E 60 2721
[9] Crooks G E 1999 Phys. Rev. E 61 2361
[10] Hatano T and Sasa S I 2001 Phys. Rev. Lett. 86 3463
[11] Maes C 2003 Sem. Poincare 2 29
[12] Seifert U 2005 Phys. Rev. Lett. 95 040602
[13] Callen H B and Welton T A 1951 Phys. Rev. 83 34
[14] Kubo R 1966 Rep. Prog. Phys. 29 255
[15] Gallavotti G 1996 Phys. Rev. Lett. 77 4334
[16] Hummer G and Szabo A 2003 Proc. Natl Acad. Sci. USA 98 3658
[17] Kolmogorov A 1931 Math. Ann. 104 415
[18] Feynman R P 1948 Rev. Mod. Phys. 20 367
[19] Kac M 1949 Trans. Am. Math. Soc. 65 1
[20] Girsanov I V 1960 Theory Probabil. Appl. 5 285
[21] Cameron R H and Martin W T 1945 Trans. Am. Math. Soc. 58 184
[22] Kurchan J 2007 J. Stat. Mech. P07005
[23] Chetrite R and Gawedzki K 2008 Commun. Math. Phys. 282 469
[24] Risken H 1984 The Fokker–Planck Equation (Berlin: Springer)
[25] Gardiner C W 1983 Handbook of Stochastic Methods (Berlin: Springer)
[26] Chernyak V, Chertkov M and Jarzynski C 2006 J. Stat. Mech P08001
[27] Onsager L and Machlup S 1953 Phys. Rev. 91 1505
[28] Zinn-Justin J 1996 Quantum Field Theory and Critical Phenomena (New York: Oxford University Press)
[29] Ge H and Jiang D Q 2008 J. Stat. Phys. 131 67
[30] Liu F and Ou-Yang Z C 2009 Phys. Rev. E 79 060107
[31] Liu F, Luo Y P, Huang M C and Ou-Yang Z C 2009 J. Phys. A: Math. Theor. 42 332003
[32] Sevick E M, Prabhakar R, Williams S R and Searles D J 2008 Annu. Rev. Phys. Chem. 59 603
[33] Harris R J and Schutz G M 2007 J. Stat. Mech P07020
[34] Ritort F 2008 Adv. Chem. Phys. 137 31
[35] Smoluchowski M V 1915 Ann. Phys. 48 1103
[36] Kramers H A 1940 Physica 7 284
[37] Stroock D W and Varadhan S R S 1979 Multidimensional Diffusion Processes (Berlin: Springer)
[38] Marini Bettolo Marconi U, Puglisi A, Rondoni L and Vulpiani A 2008 Phys. Rep. 461 111
[39] Diezemann G 2005 Phys. Rev. E 72 011104
[40] Hanggi P and Thomas H 1982 Phys. Rep. 88 207
[41] Speck T and Seifert U 2006 Europhys. Lett. 74 391
[42] Chetrite R, Falkovich G and Gawedzki K 2008 J. Stat. Mech P08005
[43] Baiesi M, Maes C and Wynants B 2009 Phys. Rev. Lett. 103 010602
[44] Baiesi M, Maes C and Wynants B 2009 J. Stat. Phys. 137 1094
[45] Barbi F, Bologna M and Grigolini P 2005 Phys. Rev. Lett. 95 220601
[46] Allegrini P, Bologna M and Grigolini P 2007 Phys. Rev. Lett. 99 010603
[47] Aquino G, Bologna M, Grigolini P and West B J 2010 Phys. Rev. Lett. 105 040601
[48] Sokolov I M and Klafter J 2006 Phys. Rev. Lett. 97 140602
[49] Wang G M, Sevick E M, Mittag E, Searles D J and Evans D J 2002 Phys. Rev. Lett. 89 050601
[50] van Zon R and Cohen E G D 2003 Phys. Rev. E 67 046102
[51] Speck T and Seifert U 2005 J. Phys. A: Math. Gen. 38 L581
[52] Oono Y and Paniconi M 1998 Prog. Theor. Phys. Suppl. 130 29
[53] Speck T, Blickle V, Bechinger C and Seifert U 2007 Europhys. Lett. 79 30002
[54] Karatzas I and Shreve S E 1988 Brownian Motion and Stochastic Calculus (New York: Springer)