Jarzynski's equality, fluctuation theorems, and variance reduction: Mathematical analysis and numerical algorithms

Carsten Hartmann [1] Christof Schütte [2,3] Wei Zhang [3]

Abstract

In this paper, we study Jarzynski's equality and fluctuation theorems for diffusion processes. While some of the results considered in the current work are known in the (mainly physics) literature, we review and generalize these nonequilibrium theorems using mathematical arguments, therefore enabling further investigations in the mathematical community. On the numerical side, variance reduction approaches such as importance sampling method are studied in order to compute free energy differences based on Jarzynski's equality.

Keywords Jarzynski's equality, fluctuation theorem, nonequilibrium dynamics, free energy difference, variance reduction, reaction coordinate

Contents

[1] Institut für Mathematik, Brandenburgische Technische Universität Cottbus-Senftenberg, D-03046 Cottbus, Germany; carsten.hartmann@b-tu.de

[2] Institut für Mathematik, Freie Universität Berlin, D-14195 Berlin, Germany; christof.schuette@fu-berlin.de

[3] Zuse Institute Berlin, D-14195 Berlin, Germany; wei.zhang@fu-berlin.de

1 Introduction

Nonequilibrium work relations concern the behavior of dynamical systems which are out of equilibrium under nonequilibrium driving forces. Different from linear response theory [42, 49] where systems are required to be close to equilibrium, nonequilibrium work relations refer to a set of equalities which hold for general systems far away from equilibrium. And the most remarkable ones include Jarzynski's equality [37, 38] and Crooks's fluctuation theorem [15]. In particular, Jarzynski's equality relates free energy differences to the work that is applied to the system in order to drive the system from one state to another within a finite period of time. Since its first report in 1997 [37, 38], considerable amount of research work has been done both numerically and experimentally to study the computation of free energy differences, by driving the system out of equilibrium using nonequilibrium forces [27, 51, 50, 68, 67]. In recent years, inspired by the work [57], there has also been growing research interest to generalize both Jarzynski's equality and fluctuation theorems to nonequilibrium systems under discrete feedback controls [58, 54, 34, 59].

Although Jarzynski's equality ensures that free energy differences can be calculated by pulling the system using any control forces (protocols) and the transition can be done within any finite time, the efficiency of Monte Carlo estimators for free energy computation based on Jarzynski's equality crucially depends on the control protocols and therefore careful design is needed. Various techniques, such as importance sampling in trajectory space [68, 51], the use of both forward and reversed trajectories [16, 67, 50, 64], the interacting particle system techniques [55], and the escorted free energy simulation method [63, 64], have been proposed in order to improve the efficiency of Monte Carlo estimators. Meanwhile, we note that several recent works have considered optimal control protocols which minimize either average work or average heat [62, 60, 2, 4]. However, it is important to point out that, although these protocols are optimal in certain sense and are physically interesting, they do not necessarily provide the optimal Monte Carlo estimators in the sense of smallest variance. Readers are referred to [27, 52, 40, 19, 67] for detailed discussions on related issues.

In the aforementioned literature, the concept of free energy is often defined as a function of physical parameters, e.g., temperature, volume or pressure, which characterize the macroscopic status of physical system. This is termed as the alchemical transition case in [45]. Free energy also plays an important role in the study of model reduction of complex (molecular) systems along a given reaction coordinate or collective variables. In this context, free energy is often defined as a function of reaction coordinate which in turn depends on the state of the system. And calculating free energy differences along a given reaction coordinate has attracted considerable attentions in the study of molecular systems [35, 1, 65, 11, 45]. Similar to the alchemical transition case, Jarzynski-like equalities and their applications in free energy calculation have been considered in [44, 46].

Motivated by the development of nonequilibrium work relations and their potential applications, the goal of the current work is to understand these results from a mathematical point of view, and to study variance reduction approaches, such as importance sampling, in Monte Carlo methods for free energy calculation based on Jarzynski's equality. In the alchemical transition case, we provide mathematical proofs of both Jarzynski's equality and fluctuation theorems in a general setting based on the theory of stochastic differential equations, making them more accessible for readers in mathematical community (we refer to the previous study [25] for a mathematical proof of Jarzynski's equality). It is worth emphasizing that the nonequilibrium diffusion processes in our setting are allowed to be irreversible and can have multiplicative noise. Furthermore, the Jarzynski's equality is generalized to allow noisy control protocols. This generalization may be useful to study systems in experiments [36], since the implementations of control protocols through physical devices are typically imprecise to some extent. As an advantage of our mathematical approach, it allows us to elucidate the connection between thermodynamic integration identity and Jarzynski's equality, which were usually considered as two distinct identities involving free energy differences. Such a connection is indeed known in physics community [14], but we believe it is helpful to present its mathematical derivation. In the reaction coordinate case, we prove a fluctuation theorem and derive a Jarzynski-like equality based on the fluctuation theorem. These results complement the previous mathematical studies in [44, 46]. In both the alchemical transition case and the reaction coordinate case, following our previous studies [72, 30, 31], we investigate variance reduction approaches in order to compute free energy differences using Monte Carlo method based on Jarzynski's equality.

The paper is organized as follows. In Section 2, we study the Jarzynski's equality and fluctuation theorem in the alchemical transition case. In particular, the cases when the control protocols are noisy will be considered. Information-theoretic formulation of Jarzynski's equality, the importance sampling method, as well as the cross-entropy method will be discussed in the context of free energy calculation. In Section 3, we study the Jarzynski-like equality and the fluctuation theorem in the reaction coordinate case. Information-theoretic formulations and variance reduction approaches will be discussed following a similar reasoning as in Section 2. Two simple numerical examples are studied in detail in Section 4 to illustrate the numerical issues of Monte Carlo estimators for free energy calculation as well as the variance reduction ideas proposed in this work. In Appendix A two asymptotic regimes of nonequilibrium processes(fast mixing and slow driving) and, in particular, connections between Jarzynski's equality and thermodynamic integration identity will be discussed. Appendix B records the thermodynamic integration identity in the reaction coordinate case. Appendix C contains an alternative proof of the fluctuation theorem (Theorem 2) in the alchemical transition case. The proof of the fluctuation theorem in the reaction coordinate case (Theorem 3) is given in Appendix D.

2 Jarzynski's equality and fluctuation theorem: alchemical transition case

In this section, we study the Jarzynski's equality and the fluctuation theorem in the alchemical transition case. In Subsection 2.1, we introduce the dynamical systems which will be studied in this section and fix notations. Jarzynski's equality and fluctuation theorem will be studied from Subsection 2.2 to Subsection 2.3. Finally, Information-theoretic formulation of Jarzynski's equality, as well as the cross-entropy method will be discussed in Subsection 2.4 and Subsection 2.5, respectively.

2.1 Mathematical setup

Consider the stochastic process $x(s) \in \mathbb{R}^n$ which satisfies the stochastic differential equation (SDE)

$$dx(s) = b(x(s), \lambda(s))\, ds + \sqrt{2\beta^{-1}}\sigma(x(s), \lambda(s))\, dw^{(1)}(s)\,, \quad s \geq 0\,, \tag{1}$$

where $\beta > 0$ is a constant, $w^{(1)}(s)$ is a d_1-dimensional Brownian motion with $d_1 \geq n$. Both the drift vector $b : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and the matrix $\sigma : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n\times d_1}$ are smooth functions depending on the *control protocol* $\lambda(s) \in \mathbb{R}^m$, which we assume is governed by

$$d\lambda(s) = f(\lambda(s), s)\, ds + \sqrt{2\epsilon}\, \alpha(\lambda(s), s)\, dw^{(2)}(s)\,. \tag{2}$$

In the above, $\epsilon \geq 0$ is related to the intensity of the noise, $\lambda(0) \in \mathbb{R}^m$ is fixed, $f : \mathbb{R}^m \times \mathbb{R}^+ \rightarrow \mathbb{R}^m$, $\alpha : \mathbb{R}^m \times \mathbb{R}^+ \rightarrow \mathbb{R}^{m\times d_2}$ are smooth functions, and $w^{(2)}(s)$ is a d_2-dimensional Brownian motion independent of $w^{(1)}(s)$. Notice that in equation (2), functions f, α are assumed to be independent of $x(s)$, and therefore the control protocol $\lambda(s)$ is of feedback form with respect to itself but does not depend on the system state $x(s)$. More generally, in Subsection 2.3, we will also consider the case when the control protocol is of feedback form with respect to both processes $x(s)$ and $\lambda(s)$, i.e.,

$$d\lambda(s) = f(x(s), \lambda(s), s)\, ds + \sqrt{2\epsilon}\, \alpha(x(s), \lambda(s), s)\, dw^{(2)}(s)\,. \tag{3}$$

In both cases (2) and (3), the infinitesimal generator of the dynamics $\lambda(s)$ for fixed $x(s)$ is given by

$$\mathcal{L}_2 = f \cdot \nabla_\lambda + \epsilon\, (\alpha\alpha^T) : \nabla^2_\lambda\,, \tag{4}$$

where ∇_λ denotes the gradient operator with respect to the variable $\lambda \in \mathbb{R}^m$ and

$$(\alpha\alpha^T) : \nabla^2_\lambda \phi := \sum_{1\leq i,j\leq m} (\alpha\alpha^T)_{ij} \frac{\partial^2 \phi}{\partial\lambda_i \partial\lambda_j}\,,$$

for a smooth function ϕ of variable $\lambda \in \mathbb{R}^m$.

For fixed parameter $\lambda \in \mathbb{R}^m$, the dynamics (1) reads

$$dx(s) = b(x(s), \lambda)\, ds + \sqrt{2\beta^{-1}}\sigma(x(s), \lambda)\, dw^{(1)}(s)\,, \quad s \geq 0\,, \tag{5}$$

and its infinitesimal generator is

$$\mathcal{L}_1 = b(\cdot, \lambda) \cdot \nabla + \frac{1}{\beta} a(\cdot, \lambda) : \nabla^2 \,, \tag{6}$$

where the matrix $a = \sigma\sigma^T$ and ∇ denotes the gradient operator with respect to $x \in \mathbb{R}^n$. Correspondingly, the infinitesimal generator of the joint process $(x(s), \lambda(s))$ is

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \,, \tag{7}$$

since the two Brownian motions $w^{(1)}(s)$, $w^{(2)}(s)$ are independent. Throughout this article, we assume that the drift and noise coefficients satisfy appropriate Lipschitz and growth conditions, such that equations (1)-(3) have unique strong solutions [53]. For each fixed parameter $\lambda \in \mathbb{R}^m$, we further assume that the process $x(s)$ in (5) is ergodic and has a unique invariant measure μ_λ satisfying

$$\mu_\lambda(dx) = \rho(x, \lambda) dx \,, \quad \int_{\mathbb{R}^n} \rho(x, \lambda) dx = 1 \,. \tag{8}$$

Furthermore, we introduce the potential

$$V(x, \lambda) = -\beta^{-1} \ln \rho(x, \lambda) + \text{constant} \,, \tag{9}$$

where the constant only depends on the parameter λ. Equivalently, we have $\rho(x, \lambda) = \frac{1}{Z(\lambda)} e^{-\beta V(x,\lambda)}$, and the normalization constant $Z(\lambda)$ is given by

$$Z(\lambda) = \int_{\mathbb{R}^n} e^{-\beta V(x,\lambda)} dx \,. \tag{10}$$

The free energy of the system (5) for a fixed parameter $\lambda \in \mathbb{R}^m$ is defined as

$$F(\lambda) = -\beta^{-1} \ln Z(\lambda) \,. \tag{11}$$

To proceed, we follow the previous study [70] and introduce the quantity

$$J_i(x, \lambda) = b_i - \frac{1}{\beta\rho} \sum_{j=1}^{n} \frac{\partial(a_{ij}\rho)}{\partial x_j} \,, \quad 1 \le i \le n \,. \tag{12}$$

Note that both here and in the following, J_i, b_i denote the ith component of the vectors J, b, respectively. Also, the dependence of the functions on the variables x and λ will be omitted when no ambiguities arise. Since the probability measure μ_λ in (8) is the invariant measure of the dynamics (5), we can verify that

$$\operatorname{div}\Big(J(x, \lambda) e^{-\beta V(x,\lambda)}\Big) \equiv 0 \,, \quad \rho - \text{a.e. } x \in \mathbb{R}^n \,, \tag{13}$$

for every $\lambda \in \mathbb{R}^m$. Thus, (1) can be written as

$$dx_i(s) = J_i ds + \frac{1}{\beta\rho} \sum_{j=1}^{n} \frac{\partial(a_{ij}\rho)}{\partial x_j} ds + \sqrt{2\beta^{-1}} \sum_{j=1}^{d_1} \sigma_{ij}\, dw_j^{(1)}(s) \,, \quad 1 \le i \le n \,, \tag{14}$$

or, in vector form,

$$dx(s) = \Big(J - a\nabla V + \frac{1}{\beta} \nabla \cdot a\Big) ds + \sqrt{2\beta^{-1}} \sigma\, dw^{(1)}(s) \,, \tag{15}$$

where $\nabla \cdot a$ denotes the vector in $\mathbb{R}^n$ with components

$$(\nabla \cdot a)_i = \sum_{j=1}^{n} \frac{\partial a_{ij}}{\partial x_j}\,, \quad 1 \le i \le n\,. \tag{16}$$

Finally, we introduce two physical quantities which are associated to the trajectories of the stochastic processes $x(s), \lambda(s)$ and will be relevant for our subsequent study. For each trajectory $x(s)$, $\lambda(s)$ of the dynamics (1), (3) on the time interval $[t_1, t_2] \subseteq [0, T]$, the *change of internal energy* and the *work* done to the system are defined as

$$\begin{aligned} \Delta\mathcal{U}_{(t_1,\,t_2)} =& V\big(x(t_2), \lambda(t_2)\big) - V\big(x(t_1), \lambda(t_1)\big) \\ W_{(t_1,\,t_2)} =& \int_{t_1}^{t_2} \nabla_\lambda V(x(s), \lambda(s)) \circ d\lambda(s)\,, \end{aligned} \tag{17}$$

respectively. Note that, in (17), the notation '$\circ$' indicates that Stratonovich integration has been used. Using the relation between Stratonovich integration and Ito integration, we can verify the alternative expression

$$\begin{aligned} W_{(t_1,\,t_2)} =& \int_{t_1}^{t_2} \Big(\nabla_\lambda V \cdot f + \epsilon\, \alpha\alpha^T : \nabla^2_\lambda V\Big)\big(x(s), \lambda(s), s\big)\, ds \\ &+ \sqrt{2\epsilon} \int_{t_1}^{t_2} \big(\alpha^T \nabla_\lambda V\big)\big(x(s), \lambda(s), s\big) \cdot dw^{(2)}(s)\,, \end{aligned} \tag{18}$$

where Ito integration has been used.

In the following, we will omit the subscripts and adopt the notation $W = W_{(t_1,t_2)}$ when we consider the time interval $[t_1, t_2] = [0, T]$. Similarly, $W(t)$ will be used to denote the work $W_{(0,t)}$ for $t \in [0, T]$.

2.2 Jarzynski's equality under noisy control protocol

Jarzynski's equality can be derived using different approaches [40]. In this subsection, we will provide a simple argument to obtain the (generalized) Jarzynski's equality, where the nonequilibrium processes $x(s)$ can be irreversible for fixed parameter λ, the diffusion coefficient σ in the equation (1) of $x(s)$ can be position dependent (multiplicative noise), and the control protocol $\lambda(s)$ can be stochastic ($\epsilon > 0$). The proof has some similarities with the one in [36] using the Feynman-Kac formula. As an advantage of our method, it allows us to figure out the connections between thermodynamic integration and Jarzynski's equality by analyzing the related PDEs. See Remark 1 and Appendix A for more details.

Before starting, we first introduce the quantity

$$\begin{aligned} g(x, \lambda, t) =& \mathbf{E}_{x,\lambda,t}\Big(\varphi(x(T), \lambda(T))\, e^{-\beta W_{(t,T)}}\Big) \\ =& \mathbf{E}_{x,\lambda,t}\Big[\varphi(x(T), \lambda(T))\, e^{-\beta \int_t^T \nabla_\lambda V\big(x(u),\lambda(u)\big) \circ\, d\lambda(u)}\Big]\,, \end{aligned} \tag{19}$$

for fixed $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$ and $0 \le t \le T$, where $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a bounded and continuous test function, $\mathbf{E}_{x,\lambda,t}$ denotes the conditional expectation with respect to the path ensemble of the dynamics (1), (3) starting from $x(t) = x$ and $\lambda(t) = \lambda$ at time t. The following lemma is a direct application of the Feynman-Kac formula [53], and we provide its proof for completeness.

Lemma 1. *Consider the dynamics $x(s), \lambda(s)$ given in (1), (3). The function g defined in (19) satisfies the equation*

$$\begin{aligned}&\partial_t g + \mathcal{L}_1 g + \mathcal{L}_2 g - 2\epsilon\beta\big(\alpha^T \nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda g\big) + \Big(\epsilon\beta^2|\alpha^T\nabla_\lambda V|^2 - \beta\mathcal{L}_2 V\Big) g = 0\,, \quad 0 \le t < T\,,\\ &g(\cdot,\cdot,T) = \varphi\,,\end{aligned} \tag{20}$$

where $\mathcal{L}_1$ is the operator defined in (6), which is the infinitesimal generator of the dynamics (1) for $x(s)$ when $\lambda \in \mathbb{R}^m$ is fixed, and $\mathcal{L}_2$ is the operator defined in (4) for the process $\lambda(s)$ when $x \in \mathbb{R}^n$ is fixed.

Proof. Using the tower property of the conditional expectation, we have

$$\begin{aligned} g(x,\lambda,t) =& \mathbf{E}_{x,\lambda,t}\Big[\varphi(x(T),\lambda(T))\, e^{-\beta\int_t^T \nabla_\lambda V\big(x(u),\lambda(u)\big)\circ d\lambda(u)}\Big]\\ =& \mathbf{E}_{x,\lambda,t}\Big[e^{-\beta\int_t^s \nabla_\lambda V\big(x(u),\lambda(u)\big)\circ d\lambda(u)} g(x(s),\lambda(s),s)\Big]\,,\end{aligned} \tag{21}$$

for all time $s \in [t,T]$. Let us define $Y(s) = e^{-\beta\int_t^s \nabla_\lambda V\big(x(u),\lambda(u)\big)\circ d\lambda(u)}$. Changing Stratonovich integration into Ito integration as in (18) and applying Ito's formula to the process $Y(s)$, we get

$$dY(s) = Y(s)\Big[-\beta\mathcal{L}_2 V\, ds + \epsilon\beta^2\big|\alpha^T\nabla_\lambda V\big|^2 ds - \sqrt{2\epsilon}\beta\big(\alpha^T\nabla_\lambda V\big)\cdot\, dw^{(2)}(s)\Big]\,.$$

In a similar way, applying Ito's formula to $g(x(s),\lambda(s),s)$, gives

$$dg = \big(\partial_t g + \mathcal{L}_1 g + \mathcal{L}_2 g\big)ds + \sqrt{2\beta^{-1}}\big(\sigma^T\nabla g\big)\cdot dw^{(1)}(s) + \sqrt{2\epsilon}\big(\alpha^T\nabla_\lambda g\big)\cdot\, dw^{(2)}(s)\,.$$

Note that, here and in the following, we drop the dependence of the functions on the states $x(s)$, $\lambda(s)$ and the time s in order to simplify notation. Applying Ito's formula to the product $Y(s)g(x(s),\lambda(s),s)$, we obtain

$$\begin{aligned} &e^{-\beta\int_t^s \nabla_\lambda V\big(x(u),\lambda(u)\big)\circ d\lambda(u)} g(x(s),\lambda(s),s)\\ =& g(x,\lambda,t) + \int_t^s Y(u)\Big(-\beta\mathcal{L}_2 V + \epsilon\beta^2|\alpha^T\nabla_\lambda V|^2\Big) g(x(u),\lambda(u),u)\, du\\ &+ \int_t^s Y(u)\big(\partial_t g + \mathcal{L}_1 g + \mathcal{L}_2 g\big)\, du - 2\epsilon\beta\int_t^s Y(u)\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda g\big)\, du + M(s)\,,\end{aligned} \tag{22}$$

where $M(s)$ is a (local) martingale. Taking expectations in (22) and using (21), we get

$$\begin{aligned}\mathbf{E}_{x,\lambda,t}\Big[&-\beta\int_t^s Y(u)(\mathcal{L}_2 V) g\, du + \epsilon\beta^2\int_t^s Y(u)|\alpha^T\nabla_\lambda V|^2 g\, du\\ &+\int_t^s Y(u)\big(\partial_t g + \mathcal{L}_1 g + \mathcal{L}_2 g\big)\, du - 2\epsilon\beta\int_t^s Y(u)\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda g\big)\, du\Big] = 0\,.\end{aligned}$$

Notice that $Y(t) = 1$, $x(t) = x$ and $\lambda(t) = \lambda$ at time t. Dividing the last equation by $(s-t)$ and letting $s \to t+$, we obtain (20) which concludes the proof. □

Now we can prove the Jarzynski equality as stated below.

Theorem 1 (Generalized Jarzynski equality)**.** *Let $x(s)$ and $\lambda(s)$ be given by (1) and (2), respectively. Then, for any bounded smooth test function $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we have*

$$\mathbf{E}_{\lambda(0),0}\Big[\varphi(x(t),\lambda(t))\, e^{-\beta W(t)}\Big] = \mathbf{E}_{\lambda(0),0}\Big[e^{-\beta\big(F(\lambda(t))-F(\lambda(0))\big)}\mathbf{E}_{\mu_{\lambda(t)}}\varphi(\cdot,\lambda(t))\Big]\,, \tag{23}$$

where $F(\cdot)$ is the free energy in (11) and $W(t) = W_{(0,t)}$ is the work defined in (17) on the time interval $[0,t]$. $\mathbf{E}_{\mu_{\lambda(t)}}$ denotes the expectation with respect to the probability measure $\mu_{\lambda(t)}$ on $\mathbb{R}^n$. And $\mathbf{E}_{\lambda(0),0}$ denotes the conditional expectation over the realizations of $x(s)$ and $\lambda(s)$, starting from fixed $\lambda(0) \in \mathbb{R}^m$ and the initial distribution $x(0) \sim \mu_{\lambda(0)}$. In particular, choosing $\varphi \equiv 1$, we have

$$\mathbf{E}_{\lambda(0),0}\Big[e^{-\beta W(t)}\Big] = \mathbf{E}_{\lambda(0),0}\Big[e^{-\beta\big(F(\lambda(t))-F(\lambda(0))\big)}\Big]\,. \tag{24}$$

Proof. It suffices to prove the equality (23) for $t = T$. From the definitions of the function g in (19) and the function $Z(\lambda)$ in (10), it is easy to see that (23) is equivalent to

$$\int_{\mathbb{R}^n} g(x,\lambda(0),0)e^{-\beta V(x,\lambda(0))}dx = \mathbf{E}_{\lambda(0),0}\bigg[\int_{\mathbb{R}^n} g(x,\lambda(T),T)\, e^{-\beta V(x,\lambda(T))}\, dx\bigg]\,. \tag{25}$$

Noticing that the process $\lambda(s)$ in (2) is independent of $x(s)$ and motivated by the form of (25), we consider the quantity $\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))} g(x,\lambda(s),s)dx$ as a function of time s. Applying Ito's formula, we compute

$$\begin{aligned} & d\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))} g(x,\lambda(s),s)dx\bigg] \\ =& \bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\Big(\partial_t g + \mathcal{L}_2 g + \big(\epsilon\beta^2|\alpha^T\nabla_\lambda V|^2 - \beta\mathcal{L}_2 V\big)g - 2\epsilon\beta\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda g\big)\Big)\, dx\bigg]ds \\ & + \sqrt{2\epsilon}\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\alpha^T\big(\nabla_\lambda g - \beta\nabla_\lambda V\, g\big)dx\bigg]\cdot dw^{(2)}(s)\,, \end{aligned} \tag{26}$$

where the functions under the integral above are evaluated at $(x,\lambda(s),s)$. Since the function g satisfies the equation (20) in Lemma 1, we find

$$\begin{aligned} & d\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))} g(x,\lambda(s),s)dx\bigg] \\ =& -\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\mathcal{L}_1 g\, dx\bigg]ds + \sqrt{2\epsilon}\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\alpha^T\big(\nabla_\lambda g - \beta\nabla_\lambda V\, g\big)dx\bigg]\cdot dw^{(2)}(s)\,. \end{aligned} \tag{27}$$

Recalling that μ_λ in (8) and $\mathcal{L}_1$ are the invariant measure and the infinitesimal generator of dynamics (5), we have $\mathcal{L}_1^*\big(e^{-\beta V(x,\lambda)}\big) = 0$, where $\mathcal{L}_1^*$ is the formal L^2 adjoint of $\mathcal{L}_1$. Integrating by parts, we conclude that the first term on the right hand side of equation (27) vanishes and therefore

$$d\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))} g(x,\lambda(s),s)dx\bigg] = \sqrt{2\epsilon}\bigg[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\alpha^T\big(\nabla_\lambda g - \beta\nabla_\lambda V\, g\big)dx\bigg]\cdot dw^{(2)}(s)\,.$$

Taking expectation and noticing that $g(\cdot,\cdot,T) \equiv \varphi$, we obtain (25) and the equality (23) readily follows. □

Remark 1. 1. *While Lemma 1 holds in both cases when the control protocol $\lambda(s)$ satisfies either dynamics (2) or dynamics (3), a close examination reveals that the proof of Theorem 1 above is valid only when the process $\lambda(s)$ is independent of the process $x(s)$, i.e., when $\lambda(s)$ satisfies dynamics (2).*

2. *When $\epsilon = 0$, the control protocol is deterministic and the work becomes*

$$W(t) = \int_0^t \nabla_\lambda V\big(x(s), \lambda(s)\big) \cdot \dot{\lambda}(s)\, ds = \int_0^t \nabla_\lambda V\big(x(s), \lambda(s)\big) \cdot f(\lambda(s), s)\, ds\,. \tag{28}$$

In this case, we recover the standard Jarzynski equality [37, 38, 40], since (24) becomes

$$\mathbf{E}_{\lambda(0),0}\Big[e^{-\beta W(t)}\Big] = e^{-\beta \Delta F(t)}\,, \tag{29}$$

where

$$\Delta F(t) = F\big(\lambda(t)\big) - F\big(\lambda(0)\big) \tag{30}$$

is the free energy difference and the conditional expectation is taken with respect to dynamics (1), starting from the equilibrium distribution $\mu_{\lambda(0)}$.

3. *Besides the Jarzynski's equality, the thermodynamic integration identity is another well known representation of the free energy that can be used to calculate free energy differences [24]. Based on the argument in this subsection, in Appendix A we will derive the thermodynamic integration identity from Jarzynski's equality, and therefore provide connections of these two methods.*

In [63], the authors proposed the escorted free energy calculation method based on an identity for dynamics involving an extra force term. In the following, we briefly discuss this identity and provide a proof of it using the same argument of Theorem 1. Let us consider the dynamics

$$d\bar{x}(s) = b(\bar{x}(s), \lambda(s))\, ds + u(\bar{x}(s), \lambda(s))\, ds + \sqrt{2\beta^{-1}}\sigma(\bar{x}(s), \lambda(s))\, dw^{(1)}(s)\,, \quad s \geq 0\,, \tag{31}$$

where $u : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a smooth vector field with compact support and $\lambda(s)$ satisfies (2). We define the modified work

$$\overline{W}_{(t_1,t_2)} = \int_{t_1}^{t_2} \nabla_\lambda V(\bar{x}(s), \lambda(s)) \circ d\lambda(s) + \int_{t_1}^{t_2} \Big(u \cdot \nabla V - \frac{1}{\beta}\nabla \cdot u\Big)(\bar{x}(s), \lambda(s))\, ds\,, \tag{32}$$

for $0 \leq t_1 \leq t_2 \leq T$.

Corollary 1. *Let $\bar{x}(s)$ and $\lambda(s)$ be given by (31) and (2), respectively. Then, for any bounded smooth test function $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we have*

$$\overline{\mathbf{E}}_{\lambda(0),0}\Big[\varphi(\bar{x}(t), \lambda(t))\, e^{-\beta \overline{W}(t)}\Big] = \overline{\mathbf{E}}_{\lambda(0),0}\Big[e^{-\beta\big(F(\lambda(t)) - F(\lambda(0))\big)}\mathbf{E}_{\mu_{\lambda(t)}}\varphi(\cdot, \lambda(t))\Big]\,, \tag{33}$$

$\forall\ 0 \leq t \leq T$, where $F(\cdot)$ is the free energy in (11) and $\overline{W}(t) = \overline{W}_{(0,t)}$ is the modified work in (32). $\mathbf{E}_{\mu_{\lambda(t)}}$ denotes the expectation with respect to the probability measure $\mu_{\lambda(t)}$ on $\mathbb{R}^n$, while $\overline{\mathbf{E}}_{\lambda(0),0}$ denotes the conditional expectation over the realizations of $\bar{x}(s)$ and $\lambda(s)$, starting from fixed $\lambda(0) \in \mathbb{R}^m$ and the initial distribution $\bar{x}(0) \sim \mu_{\lambda(0)}$. In particular, choosing $\varphi \equiv 1$, we have

$$\overline{\mathbf{E}}_{\lambda(0),0}\Big[e^{-\beta \overline{W}(t)}\Big] = \overline{\mathbf{E}}_{\lambda(0),0}\Big[e^{-\beta\big(F(\lambda(t)) - F(\lambda(0))\big)}\Big]\,. \tag{34}$$

Proof. We only sketch the proof since it is similar to the proof of Theorem 1. Similar to (19), we introduce the function

$$g(x,\lambda,t) = \overline{\mathbf{E}}_{x,\lambda,t}\Big(\varphi(\bar{x}(T),\lambda(T))\, e^{-\beta\overline{W}_{(t,T)}}\Big)\,, \tag{35}$$

where $\bar{x}(t) = x \in \mathbb{R}^n$, $\lambda(t) = \lambda \in \mathbb{R}^m$ and $t \in [0,T]$. Using the same argument of Lemma 1, we can verify that g satisfies the PDE

$$\begin{aligned}&\partial_t g + \mathcal{L}_1 g + \mathcal{L}_2 g + u\cdot\nabla g - 2\epsilon\beta\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda g\big)\\ &+ \Big(\epsilon\beta^2|\alpha^T\nabla_\lambda V|^2 - \beta\mathcal{L}_2 V - \beta u\cdot\nabla V + \nabla\cdot u\Big) g = 0\,, \quad 0 \le t < T\,,\end{aligned} \tag{36}$$

with the terminal condition $g(\cdot,\cdot,T) = \varphi$. Applying Ito's formula as we did in Theorem 1, we can get

$$\begin{aligned}&d\Big[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))} g(x,\lambda(s),s)\,dx\Big]\\ =& -\Big[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\Big(\mathcal{L}_1 g + u\cdot\nabla g - (\beta u\cdot\nabla V)g + (\nabla\cdot u)g\Big)\,dx\Big]\,ds\\ &+ \sqrt{2\epsilon}\Big[\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\alpha^T\big(\nabla_\lambda g - \beta\nabla_\lambda V\, g\big)dx\Big]\cdot dw^{(2)}(s)\,.\end{aligned} \tag{37}$$

Since u is smooth and has compact support, the first term on the right hand side above vanishes using integration by parts formula. (33) is obtained following the same argument in the proof of Theorem 1. □

2.3 Fluctuation theorem

In this subsection we study the fluctuation theorem in the alchemical transition case. Note that the main result below (Theorem 2) has been obtained in [10], where comprehensive analysis as well as several concrete examples have been presented. The main purpose of this subsection is to provide a both concise and mathematical derivation which directly leads to Theorem 2. A different proof which is similar (but shorter) to the argument in [10] can be found in Appendix C.

First of all, we introduce the "reversed" dynamics, which is closely related to the dynamics $x(s)$ in (1), or its vector form (15). Notice that different reversals of stochastic dynamics have been studied in the literature in both mathematics and physics communities. We refer to [32, 10] and the references therein. In our case, we consider the dynamics $x^R(s)$ on the time interval $s \in [0,T]$, which is governed by

$$dx^R(s) = \Big(-J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\big(x^R(s),\lambda^R(s)\big)ds + \sqrt{2\beta^{-1}}\sigma\big(x^R(s),\lambda^R(s)\big)\,dw^{(1)}(s)\,, \tag{38}$$

where $\lambda^R(s)$ is the control protocol satisfying the SDE

$$\begin{aligned}d\lambda^R(s) =& -f\big(x^R(s),\lambda^R(s),T-s\big)\,ds + 2\epsilon\big(\nabla_\lambda\cdot(\alpha\alpha^T)\big)\big(x^R(s),\lambda^R(s),T-s\big)\,ds\\ &+ \sqrt{2\epsilon}\,\alpha\big(x^R(s),\lambda^R(s),T-s\big)\,dw^{(2)}(s)\,.\end{aligned} \tag{39}$$

Comparing to dynamics (3), we note that there is an extra term $\nabla_\lambda\cdot(\alpha\alpha^T)$ in (39). The infinitesimal generator of the system (38) and (39) is given by

$$\begin{aligned}\mathcal{L}^R =&\Big(-J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla + \frac{1}{\beta}a : \nabla^2 + \Big(2\epsilon\nabla_\lambda\cdot(\alpha\alpha^T) - f\Big)\cdot\nabla_\lambda + \epsilon\,\alpha\alpha^T : \nabla_\lambda^2\\ =&\mathcal{L}_1^R + \mathcal{L}_2^R\,,\end{aligned} \tag{40}$$

where $\mathcal{L}_1^R$ is the infinitesimal generator of the dynamics (38) when $\lambda^R(s)$ is fixed, and similarly $\mathcal{L}_2^R$ is the infinitesimal generator of the dynamics (39) when $x^R(s)$ is fixed. We will also use the notation $\mathcal{L}^R_{(x,\lambda,T-t)}$ to emphasize that functions in the operator (40) are evaluated at $(x, \lambda, T-t)$.

The following fluctuation result concerns the relation between dynamics (15), (3) and the reversed ones (38), (39).

Theorem 2. *Let $0 \le t' < t \le T$, $x, x' \in \mathbb{R}^n$ and $\lambda, \lambda' \in \mathbb{R}^m$. For any continuous function $\eta \in C\big(\mathbb{R}^n \times \mathbb{R}^m \times [0,T]\big)$ with compact support, we have*

$$\begin{aligned}
&e^{-\beta V(x',\lambda')}\,\mathbf{E}^R_{x',\lambda',t'}\bigg[\exp\bigg(\int_{t'}^{t} \eta\big(x^R(s),\lambda^R(s),T-s\big)ds\bigg)\delta\big(x^R(t)-x\big)\,\delta\big(\lambda^R(t)-\lambda\big)\bigg]\\
=&e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,T-t}\bigg[e^{-\beta\mathcal{W}}\exp\bigg(\int_{T-t}^{T-t'} \eta\big(x(s),\lambda(s),s\big)ds\bigg)\delta\big(x(T-t')-x'\big)\delta\big(\lambda(T-t')-\lambda'\big)\bigg]\,,
\end{aligned}\tag{41}$$

where

$$\mathcal{W} = \int_{T-t}^{T-t'} \nabla_\lambda V\big(x(s),\lambda(s)\big)\circ d\lambda(s) - \frac{1}{\beta}\int_{T-t}^{T-t'}\Big[div_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big)\Big]\big(x(s),\lambda(s),s\big)ds\,,\tag{42}$$

$x^R(\cdot), \lambda^R(\cdot)$ satisfy the dynamics (38), (39), and $x(\cdot), \lambda(\cdot)$ satisfy the dynamics (15), (3), respectively. Here, $\delta(\cdot)$ denotes the Dirac delta function (see Remark 2 below) and div_λ denotes the divergence operator with respect to $\lambda \in \mathbb{R}^m$. $\mathbf{E}^R_{x',\lambda',t'}$ is the conditional expectation with respect to the path ensemble of the dynamics (38), (39) starting from $x^R(t') = x'$ and $\lambda^R(t') = \lambda'$ at time t', while $\mathbf{E}_{x,\lambda,T-t}$ is the conditional expectation with respect to the dynamics (15) and (3).

Proof. We consider the quantities on both sides of the equality (41). For the left hand side of (41), let us fix the values $(x', \lambda', t') \in \mathbb{R}^n \times \mathbb{R}^m \times [0,T]$ and define the function u by

$$u\big(x,\lambda,t\,;x',\lambda',t'\big) = \mathbf{E}^R_{x',\lambda',t'}\bigg[\exp\bigg(\int_{t'}^{t}\eta\big(x^R(s),\lambda^R(s),T-s\big)ds\bigg)\delta\big(x^R(t)-x\big)\delta\big(\lambda^R(t)-\lambda\big)\bigg]\,,\tag{43}$$

for $(x,\lambda,t) \in \mathbb{R}^n \times \mathbb{R}^m \times [0,T]$. It is known that u satisfies the PDE

$$\begin{aligned}
&\frac{\partial u}{\partial t} = \big(\mathcal{L}^R_{(x,\lambda,T-t)}\big)^* u + \eta(x,\lambda,T-t)\,u\,, \quad \forall\,(x,\lambda,t) \in \mathbb{R}^n \times \mathbb{R}^m \times (t',T]\,,\\
&u(x,\lambda,t\,;x',\lambda',t') = \delta(x-x')\,\delta(\lambda-\lambda')\,, \quad \text{if } \ t = t'\,,
\end{aligned}\tag{44}$$

where the operator $\mathcal{L}^R_{(x,\lambda,T-t)}$ is defined in (40) and $\big(\mathcal{L}^R_{(x,\lambda,T-t)}\big)^*$ denotes its formal L^2 adjoint. Direct calculation shows that, after some cancellation, we have

$$\begin{aligned}
\big(\mathcal{L}^R_{(x,\lambda,T-t)}\big)^*\phi =& \Big[\mathrm{div}(J + a\nabla V) + \mathrm{div}_\lambda\Big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\Big)\Big]\phi + \Big(J + a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla\phi\\
&+\frac{1}{\beta}a:\nabla^2\phi + f\cdot\nabla_\lambda\phi + \epsilon\,\alpha\alpha^T:\nabla^2_\lambda\phi\,,
\end{aligned}\tag{45}$$

for a smooth function ϕ.

For the right hand side of (41), we define the function g for fixed (x', λ', t') as

$$
\begin{aligned}
g(x,\lambda,t) = \mathbf{E}_{x,\lambda,T-t}\bigg[&e^{-\beta\mathcal{W}} \exp\bigg(\int_{T-t}^{T-t'} \eta\big(x(s),\lambda(s),s\big)ds\bigg) \\
&\times \delta\big(x(T-t')-x'\big)\delta\big(\lambda(T-t')-\lambda'\big)\bigg]\,,
\end{aligned}
$$

where $\mathcal{W}$ is defined in (42), and the dynamics $x(\cdot), \lambda(\cdot)$ satisfies SDEs (15), (3). Using the same argument as in Lemma 1, we can verify that the function g satisfies the PDE

$$
\begin{aligned}
&\frac{\partial g}{\partial t} = \overline{\mathcal{L}}_{(x,\lambda,T-t)}\, g\,, \qquad \forall\, (x,\lambda,t) \in \mathbb{R}^n \times \mathbb{R}^m \times (t', T]\,, \\
&g(x,\lambda,t) = \delta(x-x')\delta(\lambda-\lambda')\,, \qquad \text{if } \ t = t'\,,
\end{aligned}
\tag{46}
$$

where the operator $\overline{\mathcal{L}}_{(x,\lambda,T-t)}$ is defined as

$$
\begin{aligned}
\overline{\mathcal{L}}_{(x,\lambda,T-t)}\, \phi =& \Big[\epsilon\beta^2|\alpha^T\nabla_\lambda V|^2 - \beta\mathcal{L}_2 V + \mathrm{div}_\lambda\Big(f - \epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\Big) + \eta\Big]\phi \\
&+ \mathcal{L}_1\phi + \mathcal{L}_2\phi - 2\epsilon\beta\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda\phi\big)
\end{aligned}
\tag{47}
$$

for a smooth function ϕ, and the functions in (47) are evaluated at $(x,\lambda,T-t)$. Motivated by the right hand side of (41), now a key step is to consider the function $\omega(x,\lambda,t) = e^{-\beta V(x,\lambda)}g(x,\lambda,t)$. Recalling the relation (13), a direct calculation shows that

$$
\begin{aligned}
e^{-\beta V}\mathcal{L}_1 g =& e^{-\beta V}\Big(J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla\big(e^{\beta V}\omega\big) + \frac{e^{-\beta V}}{\beta} a : \nabla^2\big(e^{\beta V}\omega\big) \\
=& \Big(J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla\omega + \beta\Big[\big(J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\big)\cdot\nabla V\Big]\omega \\
&+ \frac{1}{\beta}a : \nabla^2\omega + 2(a\nabla V)\cdot\nabla\omega + \frac{e^{-\beta V}\omega}{\beta}a : \nabla^2\big(e^{\beta V}\big) \\
=& \Big[\mathrm{div}(J + a\nabla V)\Big]\omega + \Big(J + a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla\omega + \frac{1}{\beta}a : \nabla^2\omega\,, \\
e^{-\beta V}\mathcal{L}_2 g =& e^{-\beta V}\Big[f\cdot\nabla_\lambda(e^{\beta V}\omega) + \epsilon\,\alpha\alpha^T : \nabla^2_\lambda(e^{\beta V}\omega)\Big] \\
=& \mathcal{L}_2\omega + \beta(\mathcal{L}_2 V)\omega + 2\epsilon\beta\big(\alpha^T\nabla_\lambda V\big)\cdot\big(\alpha^T\nabla_\lambda\omega\big) + \epsilon\beta^2|\alpha^T\nabla_\lambda V|^2\,\omega\,, \\
e^{-\beta V}\nabla_\lambda g =& e^{-\beta V}\nabla_\lambda\big(e^{\beta V}\omega\big) = \beta\big(\nabla_\lambda V\big)\omega + \nabla_\lambda\omega\,.
\end{aligned}
\tag{48}
$$

Combining (40), (46), (47), (48), we can conclude that ω satisfies PDE

$$
\begin{aligned}
&\frac{\partial\omega}{\partial t} = e^{-\beta V}\overline{\mathcal{L}}_{(x,\lambda,T-t)}\, g = \big(\mathcal{L}^R_{(x,\lambda,T-t)}\big)^*\,\omega + \eta(x,\lambda,T-t)\,\omega\,, \quad \forall\,(x,\lambda,t)\in\mathbb{R}^n\times\mathbb{R}^m\times(t',T]\,, \\
&\omega(x,\lambda,t) = e^{-\beta V(x',\lambda')}\delta(x-x')\delta(\lambda-\lambda')\,, \quad \text{if } \ t = t'\,.
\end{aligned}
$$

Comparing the latter with (44), we obtain that $e^{-\beta V(x',\lambda')}u(x,\lambda,t\,;x',\lambda',t') = \omega(x,\lambda,t)$, which is equivalent to the equality (41). □

Remark 2. *We have adopted the Dirac delta function both in Theorem 2 and in its proof above, in order to simplify the derivations. Precisely, (41) should be understood in the sense of* *distributions, or equivalently,*

$$\begin{aligned}&\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} e^{-\beta V(x',\lambda')}\,\mathbf{E}^R_{x',\lambda',t'}\Big[\exp\Big(\int_{t'}^{t}\eta\big(x^R(s),\lambda^R(s),T-s\big)ds\Big)\varphi\big(x^R(t),\lambda^R(t),x',\lambda'\big)\Big]dx'd\lambda'\\ =&\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,T-t}\Big[e^{-\beta\mathcal{W}}\exp\Big(\int_{T-t}^{T-t'}\eta\big(x(s),\lambda(s),s\big)ds\Big)\varphi\big(x,\lambda,x(T-t'),\lambda(T-t')\big)\Big]dx\,d\lambda\,,\end{aligned} \tag{49}$$

for all test functions $\varphi(x,\lambda,x',\lambda')$ which are smooth enough with compact support. We emphasize that the above proof can be reformulated more rigorously, by introducing test functions and applying integration by parts.

From fluctuation theorems to Jarzynski's equality. It is well known that Jarzynski's equality can be obtained from the fluctuation theorem [10]. In the remaining part of this subsection, we consider the case when the control protocol $\lambda(s)$ satisfies the dynamics (2) and show that Theorem 1 is a consequence of Theorem 2. In this case, (39) governing the reversed protocol $\lambda^R(\cdot)$ simplifies to

$$\begin{aligned}d\lambda^R(s) =& -f\big(\lambda^R(s),T-s\big)\,ds + 2\epsilon\big(\nabla_\lambda\cdot(\alpha\alpha^T)\big)\big(\lambda^R(s),T-s\big)\,ds\\ &+\sqrt{2\epsilon}\,\alpha\big(\lambda^R(s),T-s\big)\,dw^{(2)}(s)\,,\end{aligned} \tag{50}$$

and therefore is independent of the process $x^R(\cdot)$ in (38). For simplicity, we only prove the equality (23) for $t=T$.

In order to derive the equality (23) in Theorem 1, we set $t'=0, t=T$ and $\eta=-\mathrm{div}_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big)$, which is a function independent of $x\in\mathbb{R}^n$. Multiplying $\varphi(x',\lambda')$ on both sides of the equality (41), integrating with respect to x, x', λ', and recalling the definition (17) of the work W, we obtain

$$\begin{aligned}&\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,0}\Big(\varphi(x(T),\lambda(T))\,e^{-\beta W}\Big)\,dx\\ =&\int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\varphi(x',\lambda')\,e^{-\beta V(x',\lambda')}\,\mathbf{E}^R_{x',\lambda',0}\Big[\exp\Big(\int_0^T\eta\big(\lambda^R(s),T-s\big)ds\Big)\delta(\lambda^R(T)-\lambda)\Big]dx'd\lambda'\,.\end{aligned} \tag{51}$$

Notice that the conditional expectation on the right hand side of (51) is actually independent of x' (This is only true when the control protocol doesn't depend on the dynamics. See Remark 1.). We have

$$\begin{aligned}&\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,0}\Big(\varphi(x(T),\lambda(T))\,e^{-\beta W}\Big)\,dx\\ =&\int_{\mathbb{R}^m}\big[\mathbf{E}_{\mu_{\lambda'}}\varphi(\cdot,\lambda')\big]\,Z(\lambda')\mathbf{E}^R_{\lambda',0}\Big[\exp\Big(\int_0^T\eta\big(\lambda^R(s),T-s\big)ds\Big)\delta(\lambda^R(T)-\lambda)\Big]\,d\lambda'\,,\end{aligned} \tag{52}$$

where $Z(\cdot)$ is the normalization constant in (10).

More generally, let us define the function

$$\psi(\lambda,t)=\int_{\mathbb{R}^m}\big[\mathbf{E}_{\mu_{\lambda'}}\varphi(\cdot,\lambda')\big]\,Z(\lambda')\mathbf{E}^R_{\lambda',0}\Big[\exp\Big(\int_0^{T-t}\eta(\lambda^R(s),T-s)ds\Big)\delta(\lambda^R(T-t)-\lambda)\Big]\,d\lambda'\,.$$

Similarly to the function u in (43) which satisfies the PDE (44), we know that ψ satisfies

$$\begin{aligned}&\frac{\partial\psi}{\partial t} + \big(\mathcal{L}_2^R\big)^*\psi - \Big[\mathrm{div}_\lambda\Big(f - \epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\Big)\Big]\psi = 0\,, \quad \forall\ (\lambda,t)\in\mathbb{R}^m\times[0,T)\,,\\ &\psi(\lambda,T) = Z(\lambda)\mathbf{E}_{\mu_\lambda}\varphi(\cdot,\lambda)\ ,\end{aligned} \tag{53}$$

where $\mathcal{L}_2^R = \big(2\epsilon\nabla_\lambda\cdot(\alpha\alpha^T) - f\big)\cdot\nabla_\lambda + \epsilon\,\alpha\alpha^T : \nabla_\lambda^2$, and the functions in (53) are evaluated at (λ,t). Calculating $(\mathcal{L}_2^R)^*$, one can conclude that (53) is equivalent to

$$\begin{aligned}&\frac{\partial\psi}{\partial t} + \mathcal{L}_2\psi = 0\,, \quad \forall\ (\lambda,t)\in\mathbb{R}^m\times[0,T)\,,\\ &\psi(\lambda,T) = Z(\lambda)\mathbf{E}_{\mu_\lambda}\varphi(\cdot,\lambda)\,,\end{aligned} \tag{54}$$

where $\mathcal{L}_2$ is the infinitesimal generator defined in (4) for the dynamics (2), and therefore the Feynman-Kac formula implies that

$$\psi(\lambda,t) = \mathbf{E}_{\lambda,t}\Big[Z\big(\lambda(T)\big)\mathbf{E}_{\mu_{\lambda(T)}}\varphi(\cdot,\lambda(T))\Big].$$

Combining this with the identity in (52), we conclude that

$$\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,0}\Big(\varphi(x(T),\lambda(T))\,e^{-\beta W}\Big)dx = \psi(\lambda,0) = \mathbf{E}_{\lambda,0}\big[Z\big(\lambda(T)\big)\mathbf{E}_{\mu_{\lambda(T)}}\varphi(\cdot,\lambda(T))\big]\,,$$

which is equivalent to the equality (23) in Theorem 1 for $t = T$. □

In the above analysis, we have assumed that the control protocol $\lambda(s)$ is perturbed by noise. Let us now consider the case when $\lambda(s)$ is deterministic, i.e., when $\epsilon = 0$ in dynamics (2). In this case, we have

$$\dot\lambda(s) = f(\lambda(s),s)\,, \quad 0\le s\le T\,, \tag{55}$$

and $\lambda^R(s) = \lambda(T-s)$. It is well known that Crooks's relations [16] can be derived from the fluctuation relation [10, 64]. In the following remark, for simplicity we will only state Crooks's relations for the escorted dynamics (31). Results corresponding to the original dynamics (1) can be recovered by choosing $u\equiv 0$.

Remark 3 (Crooks's relations for the escorted dynamics)**.** *Consider the reversed version of the escorted dynamics (31), which satisfies*

$$\begin{aligned}d\bar x^R(s) =&\Big(-J - a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\big(\bar x^R(s),\lambda^R(s)\big)\,ds - u(\bar x^R(s),\lambda^R(s))\,ds\\ &+\sqrt{2\beta^{-1}}\sigma(\bar x^R(s),\lambda^R(s))\,dw^{(1)}(s)\,, \quad s\ge 0\,.\end{aligned} \tag{56}$$

By slightly modifying the proof of Theorem 2, we can prove

$$\begin{aligned}&e^{-\beta V(x',\lambda(T))}\,\overline{\mathbf{E}}^R_{x',0}\Big[\exp\Big(\int_0^T \eta\big(\bar x^R(s),T-s\big)ds\Big)\delta\big(\bar x^R(T)-x\big)\Big]\\ =&e^{-\beta V(x,\lambda(0))}\,\overline{\mathbf{E}}_{x,0}\Big[e^{-\beta\overline{W}}\exp\Big(\int_0^T \eta\big(\bar x(s),s\big)\,ds\Big)\delta\big(\bar x(T)-x'\big)\Big]\,, \quad \forall\ x,x'\in\mathbb{R}^n\,,\end{aligned} \tag{57}$$

where $\overline{W} = \overline{W}_{(0,T)}$ is the modified work in (32) and $\eta\in C\big(\mathbb{R}^n\times[0,T]\big)$ is continuous with compact support. The notations $\overline{\mathbf{E}}_{x,0}$ and $\overline{\mathbf{E}}^R_{x',0}$ denote the ensemble averages with respect to *the escorted dynamics $\bar{x}(\cdot)$ in (31) and its reversed counterpart $\bar{x}^R(\cdot)$ in (56) starting from fixed state at time $s=0$, respectively.*

Since any (bounded) continuous function $\mathcal{G}$ on the path space can be approximated by linear combinations of functions which are of the form $\exp\big(\int_0^T \eta(\bar{x}(s),s)\,ds\big)$ (for instance, by discretizing $[0,T]$ into subintervals), integrating (57) gives

$$\frac{\overline{\mathbf{E}}_{\lambda(0),0}\big(e^{-\beta\overline{W}}\mathcal{G}\big)}{\overline{\mathbf{E}}^R_{\lambda(T),0}\,(\mathcal{G}^R)} = e^{-\beta\Delta F(T)}\,, \tag{58}$$

where $\mathcal{G}^R\big(x(\cdot)\big) = \mathcal{G}\big(x(T-\cdot)\big)$ for all path $x(\cdot)\in C\big([0,T],\mathbb{R}^n\big)$, and $\Delta F(T)$ is the free energy difference in (30). The notation $\overline{\mathbf{E}}_{\lambda(0),0}$ is the path ensemble average of the forward dynamics $\bar{x}(s)$ starting from $\bar{x}(0)\sim\mu_{\lambda(0)}$, and $\overline{\mathbf{E}}^R_{\lambda(T),0}$ is defined similarly for the reversed dynamics $\bar{x}^R(s)$. If we formally write $\overline{\mathcal{P}}[\bar{x}(\cdot)\,|\,\bar{x}(0)]$, $\overline{\mathcal{P}}^R[\bar{x}^R(\cdot)\,|\,\bar{x}^R(0)]$ as the probability densities on the path space for the dynamics $\bar{x}(s)$, $\bar{x}^R(s)$ starting from $\bar{x}(0)$ and $\bar{x}^R(0)$ respectively, we obtain from (58) that

$$\frac{\overline{\mathcal{P}}[x(\cdot)\,|\,x(0)]}{\overline{\mathcal{P}}^R[x(T-\cdot)\,|\,x(T)]} = e^{-\beta(\Delta\mathcal{U}(T)-\overline{W})}\,,\quad \forall\ x(\cdot)\in C\big([0,T],\mathbb{R}^n\big)\,, \tag{59}$$

where $\Delta\mathcal{U}(T)$ is the change of internal energy in (17).

Furthermore, notice that for the work function $\mathcal{G}\big(x(\cdot)\big)=\overline{W}$ in (32), we have

$$\begin{aligned}\mathcal{G}^R\big(x(\cdot)\big) =& \mathcal{G}\big(x(T-\cdot)\big)\\ =& \int_0^T \Big(\nabla_\lambda V\cdot f + u\cdot\nabla V - \frac{1}{\beta}\nabla\cdot u\Big)(x(s),\lambda(T-s),T-s)\,ds\\ =& -\int_0^T \Big(\nabla_\lambda V\cdot\dot{\lambda}^R + (-u)\cdot\nabla V - \frac{1}{\beta}\nabla\cdot(-u)\Big)(x(s),\lambda^R(s),s)\,ds\\ =& -\overline{W}^R,\end{aligned}$$

where $\overline{W}^R$ is the modified work of the reversed dynamics (56). Therefore, (58) implies

$$\frac{\overline{\mathbf{E}}_{\lambda(0),0}\big(e^{-\beta\overline{W}}\phi(\overline{W})\big)}{\overline{\mathbf{E}}^R_{\lambda(T),0}\,(\phi(-\overline{W}^R))} = e^{-\beta\Delta F(T)}\,,\quad \forall\ \phi\in C_b(\mathbb{R})\,. \tag{60}$$

Readers can recognize that the identities (59), (58) and (60) are the counterparts of the microscopic reversibility and Crooks's relations in [16, 64] for (escorted) continuous-time Markovian processes. It was already pointed out in [16] that these relations (in particular the microscopic reversibility) hold for general Markov chains out of equilibrium without reversibility assumption. The derivations above show that this is also true for the continuous-time process $\bar{x}(s)$ in (31) with the control protocol in (55).

2.4 Change of measure and information-theoretic formulation

In this subsection, we explore the idea of importance sampling [72, 31] to study the Jarzynski's equality. We focus on the case when the control protocol $\lambda(s)$ is deterministic and satisfies the ODE (55), i.e. $\epsilon=0$ in dynamics (2). For simplicity, we also assume that the coefficient matrix σ in dynamics (1) is an invertible $n\times n$ matrix. Denote $\mathbf{P}$, $\mathbf{E}$ as the probability measure and the mathematical expectation on path space $C\big([0,T],\mathbb{R}^n\big)$ with respect to paths of the process (15) starting from $x(0)\sim\mu_{\lambda(0)}$, where $\lambda(s)$ satisfies (55) with fixed $\lambda(0)\in\mathbb{R}^m$. Then the Jarzynski's equality (24) reads

$$\mathbf{E}\Big[e^{-\beta W}\Big] = e^{-\beta\Delta F}\,, \tag{61}$$

where $\Delta F = F\big(\lambda(T)\big) - F\big(\lambda(0)\big)$, with

$$W = \int_0^T \nabla_\lambda V\big(x(s),\lambda(s)\big)\cdot f\big(\lambda(s),s\big)ds\,. \tag{62}$$

See Remark 1 for related discussions.

Let $\overline{\mathbf{P}}$ be another probability measure on the space $C\big([0,T],\mathbb{R}^n\big)$ which is equivalent to $\mathbf{P}$ and let $\overline{\mathbf{E}}$ be the corresponding expectation. Applying a change of measure in (61), together with Jensen's inequality, we can deduce

$$\begin{aligned}\Delta F =& -\beta^{-1}\ln\overline{\mathbf{E}}\Big(e^{-\beta W}\frac{d\mathbf{P}}{d\overline{\mathbf{P}}}\Big)\\ \le&\,\overline{\mathbf{E}}\Big(W+\beta^{-1}\ln\frac{d\overline{\mathbf{P}}}{d\mathbf{P}}\Big)\\ =&\,\overline{\mathbf{E}}(W)+\beta^{-1}D_{KL}\big(\overline{\mathbf{P}}\,\|\,\mathbf{P}\big)\,,\end{aligned} \tag{63}$$

where $D_{KL}\big(\,\cdot\,\|\,\cdot\,\big)$ denotes the Kullback-Leibler divergence of two probability measures [47, 7]. Notice that the inequality (63) can be interpreted as a generalization of the second law of thermodynamics [8]. In particular, under certain conditions on the work W, the equality in (63) can be attained by the optimal probability measure $\mathbf{P}^*$, which is determined by

$$\frac{d\mathbf{P}^*}{d\mathbf{P}} = e^{-\beta(W-\Delta F)}\,, \qquad \mathbf{P}^* - a.s. \tag{64}$$

In other words, the optimal change of measure tilts the original path probabilities exponentially according to the differences between the work W and the free energy difference ΔF. In particular, the probability of paths with smaller work W (compared to ΔF) increases under the optimal measure.

Meanwhile, the importance sampling Monte Carlo estimator for the free energy difference ΔF based on the identity

$$\Delta F = -\beta^{-1}\ln\mathbf{E}^*\Big(e^{-\beta W}\frac{d\mathbf{P}}{d\mathbf{P}^*}\Big) \tag{65}$$

will achieve zero variance. More generally, inspired by the last line in (63), we define

$$\Phi(\overline{\mathbf{P}}) := \overline{\mathbf{E}}\big(W\big)+\beta^{-1}D_{KL}\big(\overline{\mathbf{P}}\,\|\,\mathbf{P}\big)\,, \tag{66}$$

for a general probability measure $\overline{\mathbf{P}}$ which is equivalent to $\mathbf{P}$. Then the above discussions imply the following variational principle

$$\begin{aligned}\Delta F =& \inf_{\overline{\mathbf{P}}\sim\mathbf{P}}\Big[\overline{\mathbf{E}}\big(W\big)+\beta^{-1}D_{KL}\big(\overline{\mathbf{P}}\,\|\,\mathbf{P}\big)\Big]\\ =& \inf_{\overline{\mathbf{P}}\sim\mathbf{P}}\Phi(\overline{\mathbf{P}}) = \Phi(\mathbf{P}^*)\,,\end{aligned} \tag{67}$$

where '$\sim$' denotes the equivalence relation between two probability measures. In other words, the optimal probability measure $\mathbf{P}^*$ in (64) can be characterized as the minimizer of the minimization problem (67) and the corresponding minimum equals to ΔF. Furthermore, using (64) and (66), we can verify the following simple relation

$$\begin{aligned}
\Phi(\overline{\mathbf{P}}) =&\overline{\mathbf{E}}\Big(W + \beta^{-1}\ln\frac{d\overline{\mathbf{P}}}{d\mathbf{P}}\Big)\\
=&\mathbf{E}^*\bigg[\Big(W + \beta^{-1}\ln\frac{d\overline{\mathbf{P}}}{d\mathbf{P}}\Big)\frac{d\overline{\mathbf{P}}}{d\mathbf{P}^*}\bigg]\\
=&\mathbf{E}^*\bigg[\Big(\Delta F + \beta^{-1}\ln\frac{d\mathbf{P}}{d\mathbf{P}^*} + \beta^{-1}\ln\frac{d\overline{\mathbf{P}}}{d\mathbf{P}}\Big)\frac{d\overline{\mathbf{P}}}{d\mathbf{P}^*}\bigg]\\
=&\Delta F + \beta^{-1}\mathbf{E}^*\bigg[\Big(\ln\frac{d\overline{\mathbf{P}}}{d\mathbf{P}^*}\Big)\frac{d\overline{\mathbf{P}}}{d\mathbf{P}^*}\bigg]\\
=&\Delta F + \beta^{-1}D_{KL}\big(\overline{\mathbf{P}}\,\|\,\mathbf{P}^*\big)\,,
\end{aligned} \tag{68}$$

for a general probability measure $\overline{\mathbf{P}}$ such that $\overline{\mathbf{P}} \sim \mathbf{P}$. It becomes apparent from the last expression in (68) that ΔF is the global minimum of the function Φ and is attained by the (unique) probability measure $\mathbf{P}^*$, since $D_{KL}\big(\overline{\mathbf{P}}\,\|\,\mathbf{P}^*\big) \geq 0$ and the equality is achieved if and only if $\overline{\mathbf{P}} = \mathbf{P}^*$. Furthermore, minimizing the function Φ is equivalent to minimizing the Kullback-Leibler divergence $D_{KL}\big(\,\cdot\,\|\,\mathbf{P}^*\big)$.

In the following, we show that the optimal change of measure $\mathbf{P}^*$ can be characterized more transparently. To this end, let $\mathbf{P}_{x,t}$, $\mathbf{E}_{x,t}$ denote the path measure and the conditional expectation of the process (15) starting from a fixed state $x \in \mathbb{R}^n$ at time t. Notice that, by the disintegration theorem [3, Theorem 5.3.1], we can write the path measure $\mathbf{P}$ as

$$\mathbf{P} = \int_{\mathbb{R}^n} \mathbf{P}_{x,0}\, d\mu_{\lambda(0)}(x).$$

Defining the function

$$g(x,t) = \mathbf{E}_{x,t}\big(e^{-\beta W_{(t,T)}}\big)\,, \tag{69}$$

analogously to (19), Jarzynski's equality (61) implies that

$$\Delta F = -\beta^{-1}\ln\big(\mathbf{E}_{\mu_{\lambda(0)}} g(\cdot,0)\big)\,. \tag{70}$$

Sampling an expectation value whose form is similar to (69) using importance sampling Monte Carlo method has been studied in previous work [20, 61, 66, 72, 30, 31]. In particular, we know from the Feynman-Kac formula that g solves the PDE

$$\partial_t g + \mathcal{L}_1 g - \beta(f\cdot\nabla_\lambda V)g = 0\,, \quad g(\cdot,T) = 1\,, \tag{71}$$

where $\mathcal{L}_1$ is the infinitesimal generator in (6) with $\lambda = \lambda(\cdot)$ being dependent on time t. Introducing $U = -\beta^{-1}\ln g$, it follows from (71) that U satisfies a Hamilton-Jacobi-Bellman equation

$$\begin{aligned}
&\partial_t U + \min_{c\in\mathbb{R}^n}\Big\{\mathcal{L}_1 U + \sigma c\cdot\nabla U + \frac{|c|^2}{4} + (f\cdot\nabla_\lambda V)\Big\} = 0\,,\\
&U(\cdot,T) = 0\,,
\end{aligned} \tag{72}$$

and one can show [23] that U is the value function of the optimal control problem

$$U(x,t) = \inf_{u_s} \mathbf{E}^u_{x,t}\bigg[\int_t^T \Big(\nabla_\lambda V\big(x^u(s),\lambda(s)\big)\cdot f(\lambda(s),s) + \frac{|u_s|^2}{4}\Big)ds\bigg]\,, \tag{73}$$

where $u_s \in \mathbb{R}^n$ is the control policy, $x^u(s)$ is the controlled process given by

$$dx^u(s) = b(x^u(s),\lambda(s))ds + \sigma(x^u(s),\lambda(s))u_s\,ds + \sqrt{2\beta^{-1}}\sigma(x^u(s),\lambda(s))\,dw^{(1)}(s)\,, \tag{74}$$

and $\mathbf{E}^u_{x,t}$ denotes the corresponding conditional expectation starting from $x^u(t) = x$ at time t.

In particular, it is well known that the feedback control policy

$$u^*_s(x) = -2\sigma^T(x,\lambda(s))\nabla U(x,s) = 2\beta^{-1}\frac{\sigma^T(x,\lambda(s))\nabla g(x,s)}{g(x,s)}\,,\quad (x,s)\in\mathbb{R}^n\times[0,T] \tag{75}$$

leads to the zero-variance importance sampling Monte Carlo estimator for the path ensemble average in (69) [29]. Based on these facts and the equality (70), it is not difficult to conclude that the optimal probability measure to sample the free energy ΔF in (65) is given by the disintegration expression

$$\mathbf{P}^* = \int_{\mathbb{R}^n} \mathbf{P}^*_{x,0}\, d\mu^*_0(x)\,, \tag{76}$$

where μ^*_0 is the probability measure on $\mathbb{R}^n$ such that

$$\frac{d\mu^*_0}{dx} \propto e^{-\beta V(x,\lambda(0))}g(x,0)\,, \tag{77}$$

and $\mathbf{P}^*_{x,0}$ is the probability measure corresponding to the controlled dynamics (74) starting from $x^u(0) = x$, with $u^*_s = u^*_s(x^u(s))$ which is defined in (75) for $s \in [0,T]$. In other words, the importance sampling estimator (65) for the free energy ΔF will achieve zero-variance, if we generate trajectories from dynamics (74) with the control u^*_s starting from the initial distribution $x^u(0) \sim \mu^*_0$.

Remark 4. *In the following, we make a comparison with other relevant directions in the literature.*

1. *(Optimal control protocol) In the importance sampling approach above, where the main purpose is to improve the numerical efficiency of free energy calculation, we assumed that the control protocol $\lambda(s)$ is fixed and the dynamics of the original nonequilibrium process is modified by adding an extra (additive) control force. In contrast to this, the problem of minimizing either the average work or the average heat by varying the control protocols has been considered in several recent works in the study of thermodynamics for small systems [62, 60, 2, 4]. Motivated by these studies, it may be also interesting to optimize the control protocols in order to minimize the variance of the Monte Carlo estimators. This problem is beyond the scope of the current paper but we would like to consider it in the future.*

2. *(Escorted free energy simulation) The idea of further adding an extra control force to the nonequilibrium processes in order to improve the efficiency of free energy calculation has also been explored in the escorted free energy simulation method [63, 64]. In this*

method [63], the authors derived the identity (34) for the modified dynamics (31), and suggested to apply it to compute the free energy difference ΔF by choosing the vector field u in (31) properly (such that the "lag" is reduced). There also exists an optimal vector field, at least formally, such that the Monte Carlo estimator in the escorted simulation method achieves zero variance. Despite of these similarities, we emphasize that the importance sampling method in this subsection and the escorted free energy simulation method rely on different identities (of the nonequilibrium processes with extra control). In other words, the change of measure identity in the first line of (63) and the identity (34) can not be derived from one to the other straightforwardly. Furthermore, unlike the escorted free energy simulation method where the initial distribution is fixed, in importance sampling one has the freedom to change the initial distribution as well. In particular, this is the case for the optimal change of measure, since μ_0^* in (77) is typically different from the equilibrium distribution $\mu_{\lambda(0)}$.

3. *(Bidirectional sampling, Bennett's acceptance ratio method) It is known in the literature [16, 67, 50, 64] that free energy estimators based on Crooks's relation (60), using trajectories of both the forward and backward processes, perform much better than estimators based on the Jarzynski's equality (61), which only use trajectories of the forward process. The optimal choice of the function ϕ in (60) is known [6], given the numbers of both forward and backward trajectories. It is interesting to consider how one can apply the importance sampling idea to further improve the efficiency of estimators which use trajectories of both forward and backward processes. We leave this question in future study.*

2.5 Cross-entropy method

From the previous subsection, we know that the probability measure $\mathbf{P}^*$ in (64), or equivalently in (76), is optimal in the sense that the importance sampling estimator (65) has zero-variance. However, in practice it is often difficult to compute $\mathbf{P}^*$ or u_s^*. In this subsection, we briefly outline a numerical approach to sample the free energy difference ΔF using the importance sampling Monte Carlo method [72, 56]. The main idea is to approximate the optimal measure $\mathbf{P}^*$ within a family of parameterized probability measures $\{\mathbf{P}_{\boldsymbol{\omega}}\,|\,\boldsymbol{\omega}\in\mathbb{R}^k\}$, with the hope that the closer $\mathbf{P}_{\boldsymbol{\omega}}$ is to $\mathbf{P}^*$, the more efficient the importance sampling estimator will be (in the sense that variance is small). Different from the importance sampling method studied in [68, 51] which requires Monte Carlo sampling in path space with an acceptance-rejection procedure, the method proposed below can be implemented at the SDE level.

We recall that the probability measure $\mathbf{P}$ corresponds to the trajectories of processes (1) and (55). Now let $\bar{\mu}_0$ be the probability measure on $\mathbb{R}^n$, possibly different from $\mu_{\lambda(0)}$. Given a parameter $\boldsymbol{\omega}=(\omega_1,\omega_2,\cdots,\omega_k)^T\in\mathbb{R}^k$, we define $\mathbf{P}_{\boldsymbol{\omega}}$ as the probability measure corresponding to the trajectories of the process

$$dx(s)=b\big(x(s),\lambda(s)\big)ds+\sigma\big(x(s),\lambda(s)\big)\Big(\sum_{l=1}^{k}\omega_l\phi^{(l)}\big(x(s),\lambda(s),s\big)\Big)\,ds+\sqrt{2\beta^{-1}}\sigma\big(x(s),\lambda(s)\big)\,dw(s)\,, \tag{78}$$

and the control protocol (55), starting from $x(0)\sim\bar{\mu}_0$, where $\phi^{(l)}:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^+\rightarrow\mathbb{R}^n$, $1\le l\le k$, are k ansatz functions. Clearly, we have $\mathbf{P}_{\boldsymbol{\omega}}=\mathbf{P}$ when $\boldsymbol{\omega}=\mathbf{0}\in\mathbb{R}^k$ and $\bar{\mu}_0=\mu_{\lambda(0)}$.

As a special choice of ansatz functions, we can take $\phi^{(l)} = -\sigma^T \nabla V^{(l)}$, where $V^{(l)} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $1 \le l \le k$, are k potential functions. In this case, recalling that dynamics (1) can be written equivalently as (15), we see that dynamics (78) becomes

$$dx(s) = \Big[J - a\nabla\Big(V + \sum_{l=1}^{k} \omega_l V^{(l)}\Big) + \frac{1}{\beta}\nabla \cdot a\Big](x(s), \lambda(s))\, ds + \sqrt{2\beta^{-1}}\sigma\big(x(s), \lambda(s)\big)\, dw(s)\,,$$

i.e., probability measure $\mathbf{P}_{\boldsymbol{\omega}}$ corresponds to the dynamics under the modified potential $V + \sum\limits_{l=1}^{k} \omega_l V^{(l)}$.

The optimal approximation of the probability measure $\mathbf{P}^*$ within the set $\big\{\mathbf{P}_{\boldsymbol{\omega}}\,|\, \boldsymbol{\omega} \in \mathbb{R}^k\big\}$ is defined as the minimizer of the minimization problem

$$\min_{\boldsymbol{\omega} \in \mathbb{R}^k} D_{KL}\big(\mathbf{P}^* \,\|\, \mathbf{P}_{\boldsymbol{\omega}}\big)\,. \tag{79}$$

Note that, comparing to the minimization of the function Φ in (66), which is equivalent to minimizing $D_{KL}(\cdot \,\|\, \mathbf{P}^*)$ by (68), approximations have been introduced in (79), i.e., we have first switched the order of the two arguments in $D_{KL}(\cdot \,\|\, \cdot)$ and then confined ourselves on a parameterized subset of probability measures with fixed starting distribution $\bar{\mu}_0$. Using (64), we can write the objective function in (79) more explicitly as

$$D_{KL}\big(\mathbf{P}^* \,\|\, \mathbf{P}_{\boldsymbol{\omega}}\big) = D_{KL}\big(\mathbf{P}^* \,\|\, \mathbf{P}\big) - e^{\beta\Delta F}\mathbf{E}\Big(e^{-\beta W} \ln \frac{d\mathbf{P}_{\boldsymbol{\omega}}}{d\mathbf{P}}\Big)\,, \tag{80}$$

where the parameter $\boldsymbol{\omega}$ only appears in the second term on the right hand side of the above equality. Applying Girsanov's theorem [53], we have

$$\frac{d\mathbf{P}_{\boldsymbol{\omega}}}{d\mathbf{P}} = \frac{d\bar{\mu}_0}{d\mu_{\lambda(0)}}\big(x(0)\big) \times \exp\left[\frac{\beta}{2}\int_0^T \Big(\sum_{l=1}^{k} \omega_l \phi^{(l)}\Big) \cdot \sigma^{-1}\big(dx(s) - b\, ds\big) - \frac{\beta}{4}\int_0^T \Big|\sum_{l=1}^{k} \omega_l \phi^{(l)}\Big|^2 ds\right], \tag{81}$$

where the dependence of the functions $b, \sigma, \phi^{(l)}$ on $x(s), \lambda(s), s$ is omitted for simplicity. Substituting (81) into equality (80), we can observe that the objective function in (79) is in fact quadratic with respect to the parameter $\boldsymbol{\omega} \in \mathbb{R}^k$. Taking derivatives, we conclude that the minimizer of (79) is determined by the linear equation $A\boldsymbol{\omega}^* = R$, where

$$A_{ll'} = \mathbf{E}\Big[e^{-\beta W}\int_0^T \phi^{(l)} \cdot \phi^{(l')}\, ds\Big]\,, \quad R_l = \mathbf{E}\Big[e^{-\beta W}\int_0^T \phi^{(l)} \cdot \sigma^{-1}\big(dx(s) - b\, ds\big)\Big]\,, \tag{82}$$

for $1 \le l, l' \le k$.

In practice, we can estimate entries of A and R in (82) by simulating a relatively small number of trajectories, and compute $\boldsymbol{\omega}^*$ by solving the linear equation $A\boldsymbol{\omega}^* = R$. After this, the free energy difference ΔF can be estimated using importance sampling by simulating a large number of trajectories corresponding to $\mathbf{P}_{\boldsymbol{\omega}^*}$. Also notice that, instead of computing A and R using the original dynamics and solving $\boldsymbol{\omega}^*$ directly, it is helpful to solve $\boldsymbol{\omega}^*$ in an iterative manner starting from a higher temperature (small β) or running a different dynamics (importance sampling). We refer readers to the previous studies [56, 72] for more algorithmic details.

Remark 5. *More generally, instead of keeping the starting distribution $\bar{\mu}_0$ fixed, we could also optimize $\bar{\mu}_0$ within a parameterized set of probability measures on $\mathbb{R}^n$ by solving an optimization* *problem which is similar to (79). In this case, while the optimal parameter $\boldsymbol{\omega}^*$ can still be obtained from the same linear equation $A\boldsymbol{\omega}^* = R$, a nonlinear equation needs to be solved in order to get the optimal $\bar{\mu}_0$. We expect to develop algorithms which adaptively optimize $\boldsymbol{\omega}^*$ and $\bar{\mu}_0$ in an alternative manner. This will be considered in future work.*

Choices of ansatz functions. Clearly, the efficiency of the importance sampling Monte Carlo method crucially depends on the choices of ansatz functions used in the cross-entropy method. From Jarzynski's equality (61) and the optimal change of measure (64), we can expect that an importance sampling estimator will have better performance if paths with smaller work W (comparing to ΔF) are sampled more frequently. Accordingly, the ansatz functions used in the cross-entropy method should be chosen such that the work W can be decreased by the control forces. A similar idea has been used in the previous work [31], where several ways of choosing ansatz functions have been proposed.

In the current situation where the work W is given in (62), we can see that W will be large if the potential increases along the movement of the parameter λ. Actually, this already explains the reason why a standard Monte Carlo simulation of fast-switching dynamics based on Jarzynski's equality is likely to have poor efficiency. To elucidate this point more clearly, we consider a special situation when the expression of the work W becomes simpler and allows us to have some insights on how to choose ansatz functions. Specifically, let $\lambda \in [0,1]$ and suppose that we are interested in the free energy differences corresponding to potentials $V(x,0)$ and $V(x,1)$, $x \in \mathbb{R}^n$. Then a simple way is to consider the linear interpolation [68]

$$V(x,\lambda) = (1-\lambda)V(x,0) + \lambda V(x,1)\,, \quad \lambda \in [0,1]\,, \tag{83}$$

and the control protocol $\lambda(s) = s$ on the time interval $s \in [0,1]$. In this case, the expression of work in (62) as a path functional becomes as simple as

$$W = \int_0^1 \Big(V(x(s),1) - V(x(s),0)\Big) ds\,. \tag{84}$$

It is not difficult to see that paths simulated by a standard Monte Carlo method will typically have large work due to the fact that, starting from the Boltzmann distribution of the potential $V(x,0)$ and on the finite time interval $[0,1]$, the nonequilibrium process $x(s)$ is likely to stay within the region where potential $V(x,1)$ is large, in particular when the low potential regions of $V(x,0)$ and $V(x,1)$ do not overlap (see [39] for more detailed discussions). Accordingly, the importance sampling can improve the efficiency of the standard Monte Carlo estimator if we place ansatz functions in a way such that, after optimization using the cross-entropy method, transitions of the controlled dynamics (78) from low energy regions of $V(x,0)$ to low energy region of $V(x,1)$ within time $[0,1]$ become easier. Similar idea (i.e., to reduce the "lag") has been used to guide the choice of the vector field in the escorted free energy simulation method [63, 64]. Readers are referred to Subsection 4.1 for numerical study of the ideas discussed above.

3 Jarzynski-like equality and fluctuation theorem : reaction coordinate case

Different from the situation in Section 2 where the free energy in (11) is defined as a function of the parameter λ through the invariant measure μ_λ on $\mathbb{R}^n$, in this section we assume a function

$\xi : \mathbb{R}^n \rightarrow \mathbb{R}^d$ is given and the free energy is defined as a function of $z \in \mathbb{R}^d$ through the invariant measure μ_z on the level set $\xi^{-1}(z)$. In the literature, such a function ξ is often termed as *reaction coordinate function* or *collective variable* [26, 28, 43, 12, 45, 48].

In this context, we point out that a Jarzynski-like equality has been obtained in the previous work [44], and a Jarzynski-Crooks fluctuation identity has been derived for the constrained Langevin dynamics in [46]. In this section, following the analysis in Section 2, we will prove a fluctuation theorem (Theorem 3) which is similar to Theorem 2, and then we obtain the Jarzynski-like equality (Theorem 4) by applying the fluctuation theorem. Importance sampling and variance reduction issues will be discussed in Subsection 3.4.

3.1 Mathematical setup

First of all, we recall some notations as well as some results from the work [70, 69] in order to introduce the problem under investigation.

Let $\xi : \mathbb{R}^n \rightarrow \mathbb{R}^d$ be a C^2 function with components $\xi = (\xi_1, \xi_2, \cdots, \xi_d)^T \in \mathbb{R}^d$, where $1 \leq d < n$. Given $z \in \operatorname{Im}\xi \subseteq \mathbb{R}^d$, which is a regular value of the map ξ, we define the level set

$$\Sigma_z = \xi^{-1}(z) = \Big\{ y \in \mathbb{R}^n \,\Big|\, \xi(y) = z \in \mathbb{R}^d \Big\} . \tag{85}$$

It is known from the regular value theorem [5] that Σ_z is a smooth $(n-d)$-dimensional submanifold of $\mathbb{R}^n$. Let ν_z denote the surface measure on Σ_z which is induced from the Euclidean metric on $\mathbb{R}^n$, and $\nabla\xi$ denote the $n \times d$ matrix whose entries are $(\nabla\xi)_{i\gamma} = \frac{\partial \xi_\gamma}{\partial y_i}$, $1 \leq i \leq n$, $1 \leq \gamma \leq d$.

Given a smooth function $V : \mathbb{R}^n \rightarrow \mathbb{R}$, we consider the probability measure on the submanifold Σ_z defined as

$$d\mu_z = \frac{1}{Q(z)} e^{-\beta V} \Big[\det\big(\nabla\xi^T \nabla\xi\big)\Big]^{-\frac{1}{2}} d\nu_z \, , \tag{86}$$

where $Q(z)$ is the normalization constant. The probability measure μ_z arises in many situations and plays an important role in the free energy calculation along a reaction coordinate [12, 13, 43, 70, 45, 69]. The free energy for fixed $z \in \operatorname{Im}\xi \subseteq \mathbb{R}^d$ is defined as

$$\begin{aligned} F(z) =& -\beta^{-1} \ln Q(z) \\ =& -\beta^{-1} \ln \int_{\Sigma_z} e^{-\beta V} \Big[\det\big(\nabla\xi^T \nabla\xi\big)\Big]^{-\frac{1}{2}} d\nu_z \\ =& -\beta^{-1} \ln \int_{\mathbb{R}^n} e^{-\beta V(y)} \delta\big(\xi(y) - z\big)\, dy \, , \end{aligned} \tag{87}$$

where the last equality follows from the co-area formula [22, 41]. Let $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n\times n}$ be an $n \times n$ matrix valued function such that the function $a(\cdot) := (\sigma\sigma^T)(\cdot)$ is uniformly elliptic on $\mathbb{R}^n$. Let $\Psi = \nabla\xi^T a \nabla\xi$ be the invertible $d \times d$ matrix whose entries are

$$\Psi_{\gamma\gamma'} = (\nabla\xi_\gamma)^T a \nabla\xi_{\gamma'} \, , \quad 1 \leq \gamma, \gamma' \leq d \, , \tag{88}$$

where $\nabla\xi_\gamma$ is the usual gradient of the function ξ_γ. Let $P = \mathrm{id} - a\nabla\xi\Psi^{-1}\nabla\xi^T$ be the projection matrix, with entries

$$P_{ij} =\delta_{ij} - (\Psi^{-1})_{\gamma\gamma'} a_{il} \partial_l \xi_\gamma \, \partial_j \xi_{\gamma'} \, , \quad 1 \leq i,j \leq n \, . \tag{89}$$

Notice that in the above δ_{ij} is the Kronecker delta function and Einstein's summation convention is used here and in the following. From (89), we can directly verify that

$$\begin{aligned}&P^2 = P\,, \quad P^T\nabla\xi_\gamma = 0\,, \quad 1 \le \gamma \le d\,,\\ &(aP^T)_{ij} = (Pa)_{ij} = a_{ij} - (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i(a\nabla\xi_{\gamma'})_j\,, \quad 1 \le i,j \le n\,,\end{aligned} \tag{90}$$

i.e., P is the orthogonal projection w.r.t. the scalar product $\langle u, v\rangle_{a^{-1}} = u^Ta^{-1}v$, for $u, v \in \mathbb{R}^n$.

It is shown in [69] that, starting from $y(0) \in \Sigma_z$, the process

$$dy_i(s) = -\,(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial(Pa)_{ij}}{\partial y_j}\,ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,, \quad 1 \le i \le n\,, \tag{91}$$

where $w(s)$ is an n-dimensional Brownian motion, will remain on the submanifold Σ_z and has a unique invariant measure μ_z which is defined in (86). In particular, denoting by $\mathcal{L}^\perp$ the infinitesimal generator of the process (91), i.e.,

$$\mathcal{L}^\perp = -(Pa)_{ij}\frac{\partial V}{\partial y_j}\frac{\partial}{\partial y_i} + \frac{1}{\beta}\frac{\partial(Pa)_{ij}}{\partial y_j}\frac{\partial}{\partial y_i} + \frac{1}{\beta}(Pa)_{ij}\frac{\partial^2}{\partial y_i\partial y_j}\,, \tag{92}$$

it is easy to verify that $\mathcal{L}^\perp\xi_\gamma \equiv 0$, for $1 \le \gamma \le d$.

3.2 Fluctuation theorem

In order to state the fluctuation theorem, we further introduce a "controlled" process as well as its time-reversed counterpart based on the process (91). Specifically, we let $f = (f_1, f_2, \cdots, f_d)^T : \mathbb{R}^n \times [0,T] \to \mathbb{R}^d$ be a bounded smooth function and consider the process

$$dy_i(s) = -\,(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial(Pa)_{ij}}{\partial y_j}\,ds + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\,f_{\gamma'}\,ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,, \tag{93}$$

for $1 \le i \le n$ on the time interval $[0,T]$. The infinitesimal generator of the process (93) is given by

$$\mathcal{L} = \mathcal{L}^\perp + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f_{\gamma'}\frac{\partial}{\partial y_i}\,, \tag{94}$$

where the operator $\mathcal{L}^\perp$ is defined in (92), and a simple application of Ito's formula implies that

$$d\xi(y(s)) = f(y(s), s)\,ds\,. \tag{95}$$

Similarly, the time-reversed process of the dynamics (93) on the time interval $[0,T]$ is defined as

$$dy_i^R(s) = -\,(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial(Pa)_{ij}}{\partial y_j}\,ds - (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\,f^-_{\gamma'}\,ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,, \tag{96}$$

where $1 \le i \le n$, $f^-_{\gamma'}(\cdot, s) = f_{\gamma'}(\cdot, T-s)$, and the infinitesimal generator is

$$\mathcal{L}^R = \mathcal{L}^\perp - (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\frac{\partial}{\partial y_i}\,. \tag{97}$$

Using a similar argument as in the proof of Theorem 2, we obtain the following fluctuation theorem which concerns the relation between the dynamics (93) and the time-reversed one (96).

Theorem 3. *Let $0 \le t' < t \le T$ and $y, y' \in \mathbb{R}^n$. For any continuous function $\eta \in C\big(\mathbb{R}^n \times [0,T]\big)$ with compact support, we have*

$$\begin{aligned} & e^{-\beta V(y')}\, \mathbf{E}^R_{y',t'}\Big[\exp\Big(\int_{t'}^{t} \eta(y^R(s), T-s) ds\Big) \delta\big(y^R(t) - y\big)\Big] \\ =& e^{-\beta V(y)}\, \mathbf{E}_{y,T-t}\Big[e^{-\beta \mathcal{W}} \exp\Big(\int_{T-t}^{T-t'} \eta(y(s), s) ds\Big) \delta\big(y(T-t') - y'\big)\Big]\,, \end{aligned} \tag{98}$$

where

$$\mathcal{W} = \int_{T-t}^{T-t'} \Big[(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f_{\gamma'} \frac{\partial V}{\partial y_i} - \frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f_{\gamma'}\Big)\Big] ds\,, \tag{99}$$

$y^R(\cdot)$, $y(\cdot)$ satisfy the dynamics (96) and (93), respectively. $\mathbf{E}^R_{y',t'}$ is the conditional expectation with respect to the path ensemble of the dynamics (96) starting from $y^R(t') = y'$ at time t'. And $\mathbf{E}_{y,T-t}$ is the conditional expectation with respect to the dynamics (93) starting from $y(T-t) = y$ at time $T-t$.

The proof of Theorem 3 can be found in Appendix D. Similar to Theorem 2, the identity (98) should be understood in the sense of distributions. We refer to Remark 2 for further discussions.

3.3 Jarzynski-like equality

In this subsection, we assume that there is a function $\widetilde{f} = (\widetilde{f}_1, \widetilde{f}_2, \cdots, \widetilde{f}_d)^T : \mathbb{R}^d \times [0,T] \rightarrow \mathbb{R}^d$, such that

$$f(y,s) = \widetilde{f}(\xi(y), s), \quad \forall (y,s) \in \mathbb{R}^n \times [0,T]\,. \tag{100}$$

Fix $t \in [0,T]$ and suppose that both the ODE

$$\dot{\zeta}(s\,;z) = \widetilde{f}(\zeta(s\,;z), s), \quad s \in [0,t]\,, \tag{101}$$

starting from $\zeta(0\,;z) = z$, and the ODE

$$\dot{\zeta}^R(s\,;z) = -\widetilde{f}(\zeta^R(s\,;z), T-s), \quad s \in [T-t, T]\,, \tag{102}$$

starting from $\zeta^R(T-t\,;z) = z$, have a unique solution for any $z \in \mathbb{R}^d$. Under this assumption, it is not difficult to conclude that

$$\zeta^R(s\,;\zeta(t\,;z)) = \zeta(T-s\,;z)\,, \quad \zeta(T-s\,;\zeta^R(T\,;z)) = \zeta^R(s\,;z)\,, \quad s \in [T-t, T]\,,$$

which in turn implies that the map $\zeta^R(T\,;\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is invertible and its inverse is given by $\zeta(t\,;\cdot)$.

Consider the process $y(s)$ in (93) on the time interval $[0,t]$, and process $y^R(s)$ in (96) on the time interval $[T-t, T]$, respectively. Assume that $\xi(y(0)) = z$ and $\xi(y^R(T-t)) = z'$, where $z, z' \in \mathbb{R}^d$. Similar to (95), we can obtain

$$d\xi(y(s)) = \widetilde{f}\big(\xi(y(s)), s\big)\, ds, \qquad d\xi(y^R(s)) = -\widetilde{f}\big(\xi(y^R(s)), T-s\big)\, ds\,,$$

which imply that

$$\xi(y(s)) = \zeta(s\,;z)\,, \quad \xi(y^R(T-s)) = \zeta^R(T-s\,;z')\,, \qquad \forall\, s \in [0,t]\,. \tag{103}$$

Applying Theorem 3, we can obtain the following Jarzynski-like equality for the free energy difference in the reaction coordinate case.

Theorem 4 (Jarzynski-like equality)**.** *Let $y(s)$ be the dynamics in (93) with the function f in (100) and $z(s)$ solve the ODE (101). For any smooth and bounded test function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ and $t \in [0,T]$, we have*

$$\mathbf{E}_{z(0),0}\Big[\varphi(y(t))\, e^{-\beta W(t)}\Big] = e^{-\beta\big(F(z(t))-F(z(0))\big)} \int_{\Sigma_{z(t)}} \varphi\, d\mu_{z(t)}\,, \tag{104}$$

where $F(\cdot)$ is the free energy in (87) and $W(t)$ is defined as

$$W(t) = \int_0^t \Big[(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i \frac{\partial V}{\partial y_i} - \frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\Big)\Big]\, \dot{z}_{\gamma'}(s)\, ds\,. \tag{105}$$

$\mathbf{E}_{z(0),0}$ denotes the conditional expectation with respect to the dynamics $y(s)$, starting from the initial distribution $y(0) \sim \mu_{z(0)}$ on $\Sigma_{z(0)}$. In particular, taking $\varphi \equiv 1$, we have

$$\mathbf{E}_{z(0),0}\Big[e^{-\beta W(t)}\Big] = e^{-\beta\big(F(z(t))-F(z(0))\big)}\,. \tag{106}$$

Proof. Let div_z denote the divergence operator with respect to $z \in \mathbb{R}^d$. Notice that from the definitions of Ψ in (88) and the function f in (100) we can compute

$$(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i \frac{\partial f_{\gamma'}}{\partial y_i} = (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i \frac{\partial \widetilde{f}_{\gamma'}}{\partial z_j}\frac{\partial \xi_j}{\partial y_i} = (\mathrm{div}_z \widetilde{f}\,)(\xi(y), s)\,.$$

Choosing $\eta(y,s) = -(\mathrm{div}_z \widetilde{f}\,)(\xi(y), s)$ in the equality (98) of Theorem 3, we obtain

$$\begin{aligned} & e^{-\beta V(y')}\, \mathbf{E}^R_{y',T-t}\Big[\exp\Big(-\int_{T-t}^T (\mathrm{div}_z \widetilde{f}\,)\big(\xi(y^R(s)), T-s\big)ds\Big)\delta\big(y^R(T)-y\big)\Big] \\ =& e^{-\beta V(y)}\, \mathbf{E}_{y,0}\Big[e^{-\beta W(t)}\delta\big(y(t)-y'\big)\Big]\,. \end{aligned} \tag{107}$$

Let $\tau > 0$ and multiply both sides of (107) by $\varphi(y')e^{-\beta\frac{|\xi(y)-z(0)|^2}{\tau}}$. Integrating with respect to y, y', yields

$$\begin{aligned} & \int_{\mathbb{R}^n} e^{-\beta\big(V(y)+\frac{|\zeta^R(T\,;\,\xi(y))-z(0)|^2}{\tau}\big)}\, \exp\Big(-\int_{T-t}^T (\mathrm{div}_z \widetilde{f}\,)\big(\zeta^R(s\,;\xi(y)), T-s\big)ds\Big)\, \varphi(y)\, dy \\ =& \int_{\mathbb{R}^n} e^{-\beta\big(V(y)+\frac{|\xi(y)-z(0)|^2}{\tau}\big)}\, \mathbf{E}_{y,0}\Big[e^{-\beta W(t)}\varphi(y(t))\Big]\, dy\,. \end{aligned} \tag{108}$$

Notice that, on the left hand side above, we have used the fact that $\xi(y^R(s))$ under the conditional expectation is deterministic and is given by (103).

We can rewrite the left hand side of (108) by applying the co-area formula

$$\begin{aligned} & \int_{\mathbb{R}^n} e^{-\beta\big(V(y)+\frac{|\zeta^R(T\,;\,\xi(y))-z(0)|^2}{\tau}\big)}\, \exp\Big(-\int_{T-t}^T (\mathrm{div}_z \widetilde{f}\,)\big(\zeta^R(s\,;\xi(y)), T-s\big)ds\Big)\, \varphi(y)\, dy \\ =& \int_{\mathbb{R}^d} e^{-\beta\frac{|z'-z(0)|^2}{\tau}} \Big[\int_{\{y\,|\,\zeta^R(T\,;\,\xi(y))=z'\}} e^{-\beta V(y)}\, \varphi(y)\, \exp\Big(-\int_{T-t}^T (\mathrm{div}_z \widetilde{f}\,)\big(\zeta^R(s\,;\xi(y)), T-s\big)ds\Big) \\ & \qquad \times \Big[\det\Big(\big(\nabla\zeta^R(T\,;\xi(y))\big)^T \nabla\zeta^R(T\,;\xi(y))\Big)\Big]^{-\frac{1}{2}} \nu^R_{z'}(dy)\Big]\, dz'\,, \end{aligned} \tag{109}$$

where $\nu^R_{z'}$ is the volume measure on the level set $\big\{y \in \mathbb{R}^n \,|\, \zeta^R(T\,;\xi(y)) = z'\big\}$, $\nabla\zeta^R(s\,;\xi(y))$ denotes the $n\times d$ matrix with components $\big(\nabla\zeta^R(s\,;\xi(y))\big)_{i\gamma} = \frac{\partial\zeta^R_\gamma(s\,;\xi(y))}{\partial y_i}$, for $s \in [T-t,T]$, $1 \le \gamma \le d$ and $1 \le i \le n$.

To simplify the above expressions, let $\nabla_z\zeta^R(s\,;z)$ denote the $d\times d$ matrix with components $(\nabla_z\zeta^R(s\,;z))_{ij} = \frac{\partial\zeta^R_i(s\,;z)}{\partial z_j}$ for $1 \le i,j \le d$, i.e., the differentiations with respect to the initial value at time $T-t$. Furthermore, since $\zeta^R(T\,;\cdot)$ is invertible, we can deduce that $\zeta^R(s\,;\cdot)$ is invertible for all $s\in[T-t,T]$, which then implies that the matrix $\nabla_z\zeta^R(s\,;z)$ has full rank for $s\in[T-t,T]$. Applying chain rule, we have $\nabla\zeta^R(s\,;\xi(y)) = \nabla\xi\nabla_z\zeta^R(s\,;\xi(y))$ and therefore

$$\Big[\det\Big(\big(\nabla\zeta^R(T\,;\xi(y))\big)^T\nabla\zeta^R(T\,;\xi(y))\Big)\Big]^{-\frac{1}{2}} = \Big[\det\Big(\nabla_z\zeta^R(T\,;\xi(y))\Big)\Big]^{-1}\Big[\det\big(\nabla\xi^T\nabla\xi\big)(y)\Big]^{-\frac{1}{2}}.$$

Combining the above identity, the equation (109), and applying Lemma 2 below, we know that equation (108) can be simplified as

$$\begin{aligned}&\frac{1}{Z_\tau}\int_{\mathbb{R}^n} e^{-\beta\big(V(y)+\frac{|\xi(y)-z(0)|^2}{\tau}\big)}\,\mathbf{E}_{y,0}\Big[e^{-\beta W(t)}\varphi(y(t))\Big]\,dy\\ =&\frac{\big(\frac{\pi\tau}{\beta}\big)^{\frac{d}{2}}}{Z_\tau}\Big(\frac{\beta}{\pi\tau}\Big)^{\frac{d}{2}}\int_{\mathbb{R}^d} e^{-\beta\frac{|z'-z(0)|^2}{\tau}}\Big[\int_{\{y\,|\,\zeta^R(T\,;\xi(y))=z'\}} e^{-\beta V(y)}\;\varphi(y)\Big[\det\big(\nabla\xi^T\nabla\xi\big)\Big]^{-\frac{1}{2}}\nu^R_{z'}(dy)\Big]\,dz'\,,\end{aligned} \tag{110}$$

where $Z_\tau = \int_{\mathbb{R}^n} e^{-\beta\big(V(y)+\frac{|\xi(y)-z(0)|^2}{\tau}\big)}dy$ is the normalization constant. Letting $\tau\to0$ and applying [69, Proposition 3], we obtain

$$\begin{aligned}&\int_{\Sigma_{z(0)}}\mathbf{E}_{y,0}\Big[e^{-\beta W(t)}\varphi(y(t))\Big]\,\mu_{z(0)}(dy)\\ =&\frac{1}{Q(z(0))}\int_{\big\{y\,\big|\,\zeta^R(T\,;\xi(y))=z(0)\big\}} e^{-\beta V(y)}\;\varphi(y)\Big[\det\big(\nabla\xi^T\nabla\xi\big)\Big]^{-\frac{1}{2}}\nu^R_{z(0)}(dy)\,,\end{aligned} \tag{111}$$

where $Q(\cdot)$ is the normalization constant in (86). Since the inverse of the map $\zeta^R(T\,;\cdot)$ is $\zeta(t\,;\cdot)$, we know

$$\big\{y\in\mathbb{R}^n\,\big|\,\zeta^R(T\,;\xi(y))=z(0)\big\} = \big\{y\in\mathbb{R}^n\,\big|\,\xi(y)=\zeta(t\,;z(0))=z(t)\big\} = \Sigma_{z(t)}\,,$$

and therefore (111) becomes

$$\int_{\Sigma_{z(0)}}\mathbf{E}_{y,0}\Big[e^{-\beta W(t)}\varphi(y(t))\Big]\,\mu_{z(0)}(dy) = \frac{Q(z(t))}{Q(z(0))}\int_{\Sigma_{z(t)}}\varphi(y)\mu_{z(t)}(dy)\,, \tag{112}$$

which is equivalent to the identity (104). □

We have used the following result in the above proof.

Lemma 2. *Let $\zeta^R(s\,;z)$ be the solution of the ODE (102) for $s\in[T-t,T]$, starting from $z\in\mathbb{R}^d$ at time $s=T-t$. $\nabla_z\zeta^R(s\,;z)$ denotes the $d\times d$ matrix where $(\nabla_z\zeta^R(s\,;z))_{ij} = \frac{\partial\zeta^R_i(s\,;z)}{\partial z_j}$ for $1\le i,j\le d$ and $T-t\le s\le T$. Suppose that $\nabla_z\zeta^R(s\,;z)$ is invertible for $T-t\le s\le T$, then we have*

$$\det\Big(\nabla_z\zeta^R(s\,;z)\Big) = e^{-\int_{T-t}^{s}(\mathrm{div}_z\widetilde{f})(\zeta^R(s'\,;z),T-s')\,ds'}\,,\quad s\in[T-t,T]\,. \tag{113}$$

Proof. Differentiating both sides of the ODE (102) with respect to z, we obtain the matrix equation

$$\frac{d\big(\nabla_z\zeta^R(s\,;z)\big)}{ds} = -\nabla_z\zeta^R(s\,;z)\,\nabla_z\widetilde{f}(\zeta^R(s\,;z),T-s)\,, \quad s\in[T-t,T]\,, \tag{114}$$

with the initial condition $\nabla_z\zeta^R(T-t\,;z)=\mathrm{id}$. Applying Jacobi's formula, we know that the determinant of $\nabla_z\zeta^R(s\,;z)$ satisfies

$$\begin{aligned}
&\frac{d\big[\det\big(\nabla_z\zeta^R(s\,;z)\big)\big]}{ds}\\
=&\det\big(\nabla_z\zeta^R(s\,;z)\big)\,\mathrm{tr}\Big(\Big(\nabla_z\zeta^R(s\,;z)\Big)^{-1}\frac{d\big(\nabla_z\zeta^R(s\,;z)\big)}{ds}\Big)\\
=&-\det\big(\nabla_z\zeta^R(s\,;z)\big)\,\mathrm{tr}\Big(\nabla_z\widetilde{f}(\zeta^R(s\,;z),T-s)\Big)\\
=&-\det\big(\nabla_z\zeta^R(s\,;z)\big)\,\big(\mathrm{div}_z\widetilde{f}\,\big)(\zeta^R(s\,;z),T-s)\,.
\end{aligned}$$

The expression (113) is obtained by integrating the above equation. □

Remark 6. 1. *In the special case when the reaction coordinate $\xi\in\mathbb{R}$ is scalar, matrix $a=\sigma=id$, we have $\Psi=|\nabla\xi|^2$ and it can be checked that the work (105) becomes*

$$\begin{aligned}
W(t) =& \int_0^t \Big[\frac{\nabla\xi}{|\nabla\xi|^2}\cdot\nabla V - \frac{1}{\beta}div\Big(\frac{\nabla\xi}{|\nabla\xi|^2}\Big)\Big]\dot{z}(s)\,ds\\
=& \int_0^t \frac{\nabla\xi}{|\nabla\xi|^2}\cdot\Big[\nabla\Big(V+\frac{1}{\beta}\ln|\nabla\xi|\Big)+\frac{1}{\beta}H\Big]\,\dot{z}(s)\,ds\,,
\end{aligned} \tag{115}$$

where $H=-div\Big(\frac{\nabla\xi}{|\nabla\xi|}\Big)\frac{\nabla\xi}{|\nabla\xi|}$ is the mean curvature vector (field) of the surface Σ_z [44]. Notice that the free energy (87) is different from the one considered in [44]. In fact, from the second expression in (115), we see that Theorem 4 is identical to the Feynman-Kac fluctuation equality Theorem of [44] for the potential $V+\frac{1}{2\beta}\ln(\det\Psi)$.

2. *As in the alchemical transition case, one can also study the escorted dynamics and Crooks's relations in the reaction coordinate case. For simplicity, we will omit the discussions on the escorted dynamics and only briefly summarize the Crooks's relations. In fact, by modifying the proof of Theorem 4, we can show that*

$$\frac{\mathbf{E}(e^{-\beta W}\mathcal{G})}{\mathbf{E}^R(\mathcal{G}^R)} = e^{-\beta\Delta F(T)}\,, \tag{116}$$

for any bounded smooth function $\mathcal{G}$ on the path space, where $W=W(T)$ is the work in (105), $\mathcal{G}^R(y(\cdot))=\mathcal{G}(y(T-\cdot))$ for any path $y(\cdot)$, $\mathbf{E}$ and $\mathbf{E}^R$ are the expectation with respect to the process $y(\cdot)$ in (93) starting from $y(0)\sim\mu_{z(0)}$ on $\Sigma_{z(0)}$, and the expectation with respect to the process $y^R(\cdot)$ in (96) starting from $y^R(0)\sim\mu_{z(T)}$ on $\Sigma_{z(T)}$, respectively. In particular, this implies

$$\frac{\mathbf{E}\big(e^{-\beta W}\phi(W)\big)}{\mathbf{E}^R(\phi(-W^R))} = e^{-\beta\Delta F(T)}\,, \quad \forall\,\phi\in C_b(\mathbb{R})\,, \tag{117}$$

where W^R is the work for the time-reversed process $y^R(\cdot)$ in (96). We refer to Remark 3 for comparisons.

3. Similarly as in the alchemical transition case, by considering the Jarzynski-like equality (106) for the dynamics

$$\begin{aligned} dy_i(s) =& -\frac{1}{\tau}(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta\tau}\frac{\partial (Pa)_{ij}}{\partial y_j}\,ds + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\, f_{\gamma'}\,ds \\ & + \sqrt{\frac{2\beta^{-1}}{\tau}}\,(P\sigma)_{ij}\,dw_j(s)\,, \end{aligned} \tag{118}$$

as $\tau \rightarrow 0$, we can recover the thermodynamic integration identity in the reaction coordinate case. See Appendix A and B for details.

3.4 Information-theoretic formulation and numerical considerations

In this subsection, we study the information-theoretic formulation of the Jarzynski-like equality (106) in the reaction coordinate setting. Numerical issues related to computing free energy differences will be discussed as well. Since the analysis is similar to Subsection 2.4 and Subsection 2.5, the discussion in this subsection will be brief and mainly focus on the changes.

First of all, let $\mathbf{P}$, $\mathbf{E}$ denote the probability measure and the expectation of the path ensemble corresponding to the dynamics (93) starting from $y(0) \sim \mu_{z(0)}$, with the function f given in (100). We can rewrite the equality (106) as

$$\Delta F = -\beta^{-1}\ln \mathbf{E}\left(e^{-\beta W}\right), \tag{119}$$

where $\Delta F = F(z(T)) - F(z(0))$ is the free energy difference and $W = W(T)$ is defined in (105). Let $\overline{\mathbf{P}}$ be another probability measure on the path space which is equivalent to $\mathbf{P}$ and $\overline{\mathbf{E}}$ denote the corresponding expectation. Applying a change of measure in (119), we have

$$\Delta F = -\beta^{-1}\ln \overline{\mathbf{E}}\left(e^{-\beta W}\frac{d\mathbf{P}}{d\overline{\mathbf{P}}}\right). \tag{120}$$

Following the same argument in Subsection 2.4, we can deduce exactly the same inequality (63), as well as the expression for the optimal measure $\mathbf{P}^*$, which is characterized by (64), such that the Monte Carlo estimator based on (65) will achieve zero variance. The derivations (66), (67), (68) in Subsection 2.4 carry over to the current setting as well.

On the other hand, since the trajectories of the dynamics (93) satisfy $\xi(y(t)) = z(t)$ for $t \in [0,T]$, it is important to notice that the probability measure $\mathbf{P}$ concentrates on the set of paths

$$\Big\{y(\cdot)\,\Big|\, y(\cdot) \in C([0,T],\mathbb{R}^n),\ y(t) \in \Sigma_{z(t)},\ 0 \le t \le T\Big\}. \tag{121}$$

Accordingly, the probability measure $\overline{\mathbf{P}}$ used to perform the change of measure in (120) should also concentrate on the set (121) in order to assure that it is equivalent to $\mathbf{P}$.

The optimal measure $\mathbf{P}^*$ can be characterized more transparently by considering the HJB equation. Specifically, define

$$g(y,t) = \mathbf{E}\Big(e^{-\beta W_{(t,T)}}\,\Big|\, y(t) = y\Big), \quad \forall\, y \in \Sigma_{z(t)}, \tag{122}$$

where $y(\cdot)$ satisfies (93) and $W_{(t,T)}$ is similarly defined as in (105) except that the integration is from t to T. It follows from the Feynman-Kac formula that g satisfies

$$\begin{aligned} &\partial_t g + \mathcal{L}g - \beta\Big[(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\frac{\partial V}{\partial y_i} - \frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\Big)\Big]\, f_{\gamma'} g = 0\,, \\ &g(\cdot,T) = 1\,. \end{aligned} \tag{123}$$

where $\mathcal{L}$ is the infinitesimal generator defined in (94) for the process $y(\cdot)$. And a simple calculation shows that $U = -\beta^{-1}\ln g$ satisfies the HJB equation

$$\begin{aligned}
&\partial_t U + \min_{c\in\mathbb{R}^n}\Big\{\mathcal{L}U + (P\sigma c)\cdot\nabla U + \frac{|c|^2}{4} \\
&\qquad + \Big[(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\frac{\partial V}{\partial y_i} - \frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\Big)\Big]\,f_{\gamma'}\Big\} = 0\,, \\
&U(\cdot,T) = 0\,,
\end{aligned} \tag{124}$$

from which we conclude that the optimally controlled dynamics satisfies

$$\begin{aligned}
dy_i(s) =& -(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j}\,ds + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\,f_{\gamma'}\,ds \\
&+ \big[P\sigma u^*_s(y(s))\big]_i\,ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,,\quad 1\le i\le n\,,
\end{aligned} \tag{125}$$

where the optimal feedback control $u^*_s(y) = -2(P\sigma)^T\nabla U$, starting from the distribution μ^*_0 which is determined by $\frac{d\mu^*_0}{d\mu_{z(0)}} \propto g(\cdot,0)$.

Cross-entropy method. In the following, we briefly discuss the cross-entropy method following Subsection 2.5. Consider a family of parameterized probability measures $\{\mathbf{P}_{\boldsymbol{\omega}}\,|\,\boldsymbol{\omega}\in\mathbb{R}^k\}$, where, for given $\boldsymbol{\omega} = (\omega_1,\omega_2,\cdots,\omega_k)^T \in \mathbb{R}^k$, $\mathbf{P}_{\boldsymbol{\omega}}$ is the probability measure of paths corresponding to the dynamics

$$\begin{aligned}
dy_i(s) =& -(Pa)_{ij}\frac{\partial V}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j}\,ds + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\,f_{\gamma'}\,ds \\
&+ (P\sigma)_{ij}\Big(\sum_{l=1}^k \omega_l\phi^{(l)}_j\Big)ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,,\quad 1\le i\le n\,,
\end{aligned} \tag{126}$$

where $\phi^{(l)} = (\phi^{(l)}_1,\phi^{(l)}_2,\cdots,\phi^{(l)}_n)^T : \mathbb{R}^n\times[0,T]\rightarrow\mathbb{R}^n$ are k ansatz functions, $1\le l\le k$. As a special choice, we consider $\phi^{(l)} = -\sigma^T\nabla V^{(l)}$ where $V^{(l)}:\mathbb{R}^n\rightarrow\mathbb{R}$, $1\le l\le k$, are smooth and linearly independent potential functions, by which (126) becomes

$$\begin{aligned}
dy_i(s) =& -(Pa)_{ij}\frac{\partial\big(V + \sum_{l=1}^k\omega_l V^{(l)}\big)}{\partial y_j}\,ds + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j}\,ds \\
&+ (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i\,f_{\gamma'}\,ds + \sqrt{2\beta^{-1}}\,(P\sigma)_{ij}\,dw_j(s)\,,\quad 1\le i\le n\,,
\end{aligned} \tag{127}$$

i.e., paths are sampled with the modified potential function $V + \sum\limits_{l=1}^k \omega_l V^{(l)}$.

Applying Ito's formula as in (95), we can verify that trajectories of the dynamics (126), starting from $y(0)\in\Sigma_{z(0)}$, satisfy $\xi(y(t)) = z(t)$ for $t\in[0,T]$ as well. Therefore, the probability measures $\mathbf{P}_{\boldsymbol{\omega}}$ indeed concentrate on the set (121). Applying Girsanov's theorem, we obtain

$$\frac{d\mathbf{P}_{\boldsymbol{\omega}}}{d\mathbf{P}} = \exp\left[\sqrt{\frac{\beta}{2}}\int_0^T\Big(\sum_{l=1}^k\omega_l\phi^{(l)}\Big)\cdot dw(s) - \frac{\beta}{4}\int_0^T\Big|\sum_{l=1}^k\omega_l\phi^{(l)}\Big|^2 ds\right], \tag{128}$$

where $w(s)$ is the Brownian motion in the original dynamics (93) (i.e., under the probability measure $\mathbf{P}$). Following the same argument as in Subsection 2.5, we know that the minimizer of the optimization problem (79) is given by the unique solution of the linear equation $A\boldsymbol{\omega}^* = R$, where

$$A_{ll'} = \mathbf{E}\Big(e^{-\beta W}\int_0^T\phi^{(l)}\cdot\phi^{(l')}\,ds\Big)\,,\quad R_l = \sqrt{2\beta^{-1}}\mathbf{E}\Big[e^{-\beta W}\int_0^T\phi^{(l)}\cdot dw(s)\Big]\,, \tag{129}$$

for $1 \leq l, l' \leq k$.

Variance reduction by increasing mixing. In practice, however, due to the complicate expressions of work W in (105) or (115), it becomes difficult to have an intuitive idea to guide the choices of ansatz functions, which play a crucial role in the cross-entropy method above. In the following, we briefly discuss another idea that can be explored in order to reduce the variance in the free energy calculation based on Jarzynski-like identity.

Different from the importance sampling method which improves the efficiency of Monte Carlo method by increasing the sampling frequency of paths with small work, the idea here, which is inspired by the analysis in Appendix A and Appendix B, is to compute free energy differences based on trajectories of the dynamics (118) with a small τ (similar idea has also been investigated in [18, 33]). The observation is that the standard Monte Carlo estimator based on Jarzynski-like identity typically sample trajectories with large work (therefore low efficiency) because the nonequilibrium dynamics do not have enough time to equilibrate under nonequilibrium force. Therefore, by decreasing τ in (118), the mixing of the "equilibrium part" of the nonequilibrium system becomes faster at each fixed nonequilibrium force. Numerically, the work W of the sampled trajectories is likely to be both smaller and more concentrated. From the analysis in Appendix A and Appendix B, we know that the free energy calculation method based on Jarzynski-like identity (106) reduces to the thermodynamic integration method when $\tau \rightarrow 0$. In practice, τ should be chosen not very small since otherwise the system will become more stiff and a smaller time step-size has to be used in numerical integration. Readers are referred to Subsection 4.2 for numerical study of free energy calculation using different τ.

4 Numerical examples

We consider two simple examples and study the efficiency of Monte Carlo methods for free energy computation.

4.1 Example 1: 1D example in alchemical transition case

In this example, we consider one-dimensional potentials

$$V(x, \lambda) = (1-\lambda)\frac{(x+1)^2}{2} + \lambda\Big(\frac{(x^2-1)^2}{4} - 0.4x\Big)\,, \tag{130}$$

where $x \in \mathbb{R}$ and $\lambda \in [0,1]$. As λ increases from 0 to 1, $V(\cdot, \lambda)$ varies from a quadratic potential centered at $x = -1$ to a tilted double well potential (Figure 1(a)). Recalling the free energy F defined in (11), (10), we will compute free energy differences $\Delta F(\lambda) = F(\lambda) - F(0)$, using Monte Carlo based on Jarzynski's identity (61). We fix $\beta = 5.0$ and the SDE

$$dx(s) = -\frac{\partial V}{\partial x}(x(s), \lambda(s))\, ds + \sqrt{2\beta^{-1}} dw(s)\,, \tag{131}$$

with control protocol $\lambda(s) = s$, $s \in [0,1]$, will be considered in the Monte Carlo simulations. Clearly, for the initial distribution $\mu_0 = \mu_{\lambda(0)}$, we have $\frac{d\mu_0}{dx} \propto \exp\big(-\beta\frac{(x+1)^2}{2}\big)$.

In fact, since the problem is one dimensional in space, we can directly compute the normalization constant $Z(\lambda)$ by numerically integrating (10) and therefore obtain the free energy differences $\Delta F(\lambda)$, which are shown in Figure 6(a). In particular, we obtain $\Delta F(1) = F(1) - F(0) =$

-3.44×10^{-1} and this will be our reference solution. Furthermore, we can also approximate the optimal change of measure $\mathbf{P}^*$ in (76) by computing the optimal control force u^* and the optimal initial distribution μ_0^* according to (75), (77), respectively. For this purpose, we need to compute the function $g(x,t) = \mathbf{E}_{x,t}\big(e^{-\beta W_{(t,T)}}\big)$ in (69) which satisfies (71). Notice that, in the current setting, we have $T=1$ and (71) becomes

$$\begin{aligned}
&\frac{\partial g}{\partial t} - \frac{\partial V}{\partial x}\frac{\partial g}{\partial x} + \frac{1}{\beta}\frac{\partial^2 g}{\partial x^2} - \beta\big(V(x,1) - V(x,0)\big)g = 0\,, \quad 0 \le t < 1\,,\\
&g(\cdot,1) = 1\,.
\end{aligned} \tag{132}$$

To compute g, we truncate the space of (x,t) to $[-5.0, 5.0] \times [0,1]$ and discretize the PDE (132) on a uniform grid of size 10000×10000, following a similar way that was described in [30, 71]. The solution g is obtained by solving the discretized system backwardly from $t=1$ to 0. The function $U = -\beta^{-1}\ln g$ is displayed in Figure 1(b) and the profile of $g(\cdot,0)$ at $t=0$ is shown in Figure 3(a). Based on these results, we can obtain the optimal control potentials (which is $V+2U$ according to (74) and (75)) and the optimal initial distribution μ_0^*. These results are shown in Figure 2(a), Figure 3(b) and Figure 4, respectively. In particular, combining the expression (77) with Figure 3(a) and Figure 4, it can be observed that, due to the strong inhomogeneity of $g(\cdot,0)$, the high probability density region of the optimal initial distribution μ_0^* is shifted along the positive x axis and has little overlap with that of the distribution μ_0.

Now we turn to discuss the performance of Monte Carlo methods. First of all, we apply the standard Monte Carlo method to estimate free energy differences. SDE (131) is discretized with time step-size $\Delta s = 5 \times 10^{-4}$ and we repeat the simulation 10 times. For each independent run, the estimator

$$\mathcal{I}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} e^{-\beta W_i(\lambda)} \tag{133}$$

is computed by generating $N = 5 \times 10^5$ trajectories of dynamics (131) starting from μ_0, where $W_i(\lambda)$ is the numerical approximation of (84) on $[0,\lambda]$ for the ith trajectory. The free energy differences are then estimated by

$$\Delta F(\lambda) \approx -\beta^{-1}\ln \mathcal{I}(\lambda)\,, \tag{134}$$

which is asymptotically unbiased when $N \to +\infty$. The results are summarized in Figure 6(a), Figure 6(b) as well as in the last row of Table 1. We can observe that the estimations of free energy differences have very large fluctuations within the 10 runs and the standard Monte Carlo estimator (133) has a very large (sample) standard deviation.

Noticing that the initial distribution μ_0 in fact is very different from the optimal initial distribution μ_0^*, we have also used the probability measure $\bar{\mu}_0$, which is given by $\frac{d\bar{\mu}_0}{dx} \propto \exp\big(-\beta\frac{(x-0.5)^2}{2}\big)$, as the initial distribution in importance sampling Monte Carlo methods. From the profiles of their probability density functions in Figure 4, we expect that the importance sampling Monte Carlo estimators using $\bar{\mu}_0$ will have better performance than estimators using μ_0. Besides the change of measure in the initial distribution, the controlled dynamics

$$dx(s) = -\frac{\partial V}{\partial x}(x(s),\lambda(s))\,ds + \sum_{l=1}^{k}\omega_l\phi^{(l)}(x(s),s)\,ds + \sqrt{2\beta^{-1}}\,dw(s) \tag{135}$$

is used to generate trajectories instead of dynamics (131), which leads to a further change of measure on path space. In (135), $\phi^{(l)}$ are ansatz functions which we choose to be either piecewise linear functions or Gaussian functions [31]. In the case of piecewise linear ansatz function, we divide the domain $[-1.3, 1.3]$ uniformly into 30 Voronoi cells $\mathcal{C}_l$ and the ansatz functions are defined as $\phi^{(l)}(x,t) = (1-t)\mathbf{1}_{\mathcal{C}_l}(x)$, $1 \le l \le 30$, where $\mathbf{1}_{\mathcal{C}_l}$ denotes the characteristic function of cell $\mathcal{C}_l$. In the case of Gaussian ansatz function, we choose two functions $\phi^{(l)}(x,t) = \frac{\partial V^{(l)}}{\partial x}(x,t)$, where $l = 1, 2$ and

$$V^{(1)}(x,t) = (1-t)\exp\Big(-\frac{x^2}{2}\Big), \quad V^{(2)}(x,t) = (1-t)\exp\Big(-\frac{(x-1.2)^2}{4.5}\Big). \tag{136}$$

In both cases, the ansatz functions are chosen based on the idea discussed in Subsection 2.5 and the dependence on time t is included since we know that the optimal control force, which is proportional to $\frac{\partial g}{\partial x}$, vanishes at time $t = 1$, due to the Dirichlet boundary condition in (132).

After these preparations, we apply the cross-entropy method discussed in Subsection 2.5 to optimize the coefficients ω_l in (135) by simulating 10^5 trajectories. The control forces at time $t = 0$, as well as the control potentials in Gaussian ansatz case are shown Figure 3(b) and Figure 2(b), respectively. Apparently, although the control forces are different from the optimal one, all of them can help drive the system along the positive x axis. Similarly as in the standard Monte Carlo case, we estimate the free energy differences using importance sampling Monte Carlo method for 10 times where $N = 5 \times 10^5$ trajectories of the controlled dynamics (135) are simulated for each run. Instead of (133), estimator

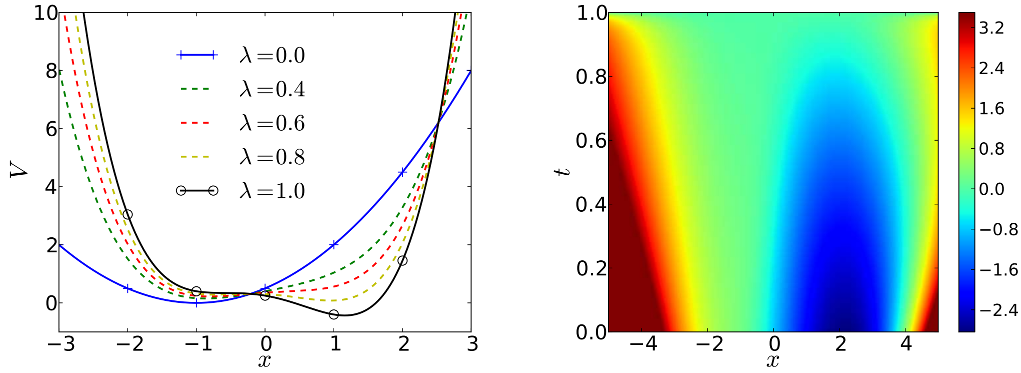
$$\mathcal{I}(\lambda) = \frac{1}{N}\sum_{i=1}^{N} e^{-\beta W_i(\lambda)}\, r_i \tag{137}$$

is computed, where r_i is the likelihood ratio given by Girsanov's theorem (see (81)). The results are shown in Figure 6(a), Figure 6(b), as well as in Table 1. Comparing to the standard deviation of the standard Monte Carlo estimator (133), we observe that the standard deviations of the importance sampling Monte Carlo estimators $\mathcal{I}(\lambda)$ in (137) are significantly reduced when we applied a change of measure both in the initial distribution and in the dynamics, i.e., when the controlled dynamics (135) with initial distribution $\bar{\mu}_0$ is used. And both types of ansatz functions exhibit comparable performances. To better understand the efficiency of Monte Carlo methods, the probability density functions and the mean values of work within the 10 runs of simulations are shown in Figure 5(a), Figure 5(b) and Table 1 for each Monte Carlo estimators. Clearly, by applying importance sampling both in the initial distribution and in the dynamics, trajectories with low work value are more efficiently sampled, leading to a much better efficiency of the Monte Carlo estimators.

4.2 Example 2: reaction coordinate case

In the second example, we study free energy calculation in the reaction coordinate case considered in Section 3. A similar example has been considered in [43], where the main focus was the approximation quality of effective dynamics. The system consists of three two-dimensional particles A, B, C whose positions are at x_A, x_B, x_C, with potential

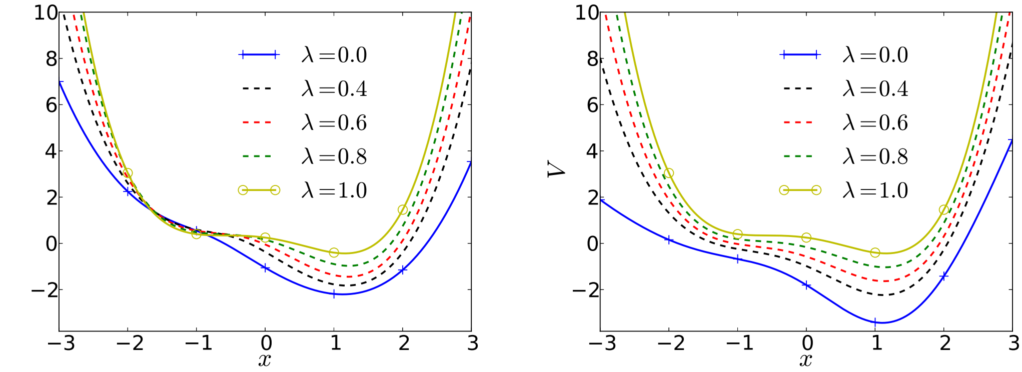
$$V(x_A, x_B, x_C) = \frac{1}{2\epsilon}\Big\{r_{BC} - \Big[1 + \kappa\Big(\sin(\theta_{ABC}) - \frac{1}{2}\Big)\Big]l_{eq}\Big\}^2 + \frac{1}{2\epsilon}\big(r_{AB} - l_{eq}\big)^2 + V_3(\theta_{ABC}), \tag{138}$$



(a) $V(x,\lambda)$

(b) $U = -\beta^{-1} \ln g$

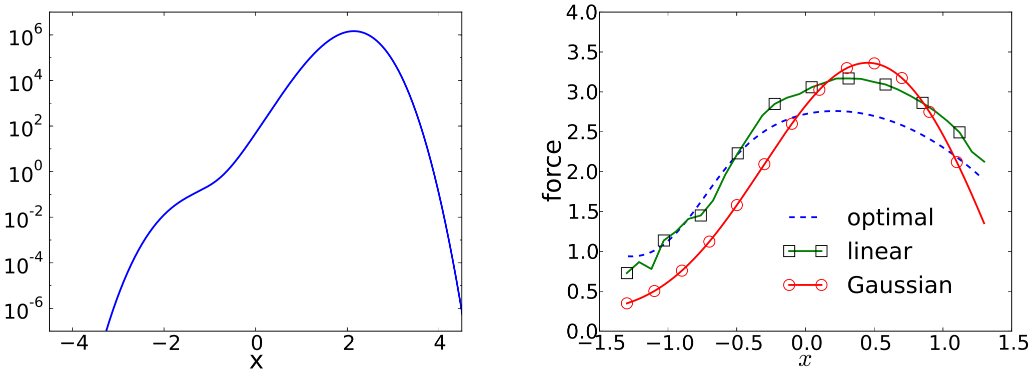
Figure 1. Example 1. (a) Potential $V(x,\lambda)$ in (130). (b) Function $U = -\beta^{-1} \ln g$, where $\beta = 5.0$ and g solves PDE (132).



(a) Optimally biased potentials

(b) Biased potentials using Gaussian ansatz

Figure 2. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0,1]$. (a) Optimally biased potential $(V + 2U)$. (b) Biased potentials computed from cross-entropy method with Gaussian ansatz functions (136).



(a) $g(x,0)$

(b) Control forces at $t = 0$

Figure 3. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0,1]$. (a) Profile of the function $g(x,0) = \mathbf{E}_{x,0}(e^{-\beta W})$ where $\beta = 5.0$ and g solves PDE (132). (b) Profiles of control forces at time $t = 0$. Curves with Labels "optimal", "linear" and "Gaussian" correspond to the optimal control u^*, the control forces obtained from the cross-entropy method using piecewise linear and Gaussian ansatz functions.

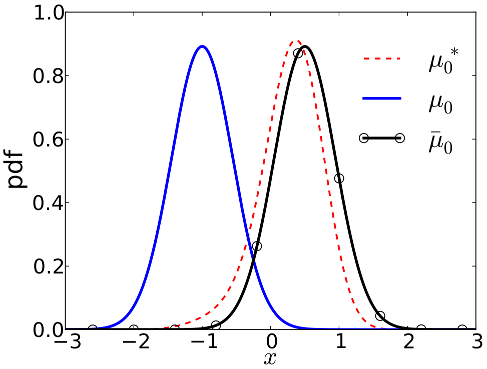
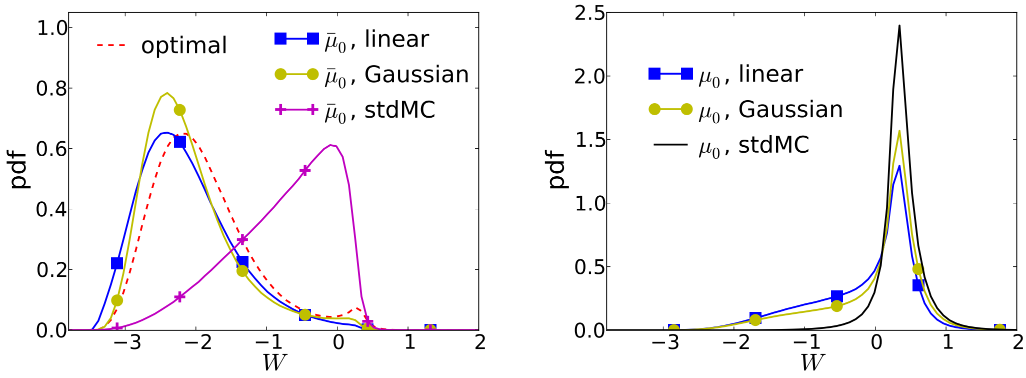


Figure 4. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0,1]$. Probability density functions of different initial distributions used in Monte Carlo methods for $\beta = 5.0$. The corresponding densities are $\frac{d\mu_0}{dx} \propto \exp\big(-\beta\frac{(x+1)^2}{2}\big)$, $\frac{d\bar{\mu}_0}{dx} \propto \exp\big(-\beta\frac{(x-0.5)^2}{2}\big)$, and $\frac{d\mu_0^*}{dx} \propto \exp\big(-\beta\frac{(x+1)^2}{2}\big)g(x,0)$, which is given by (77).



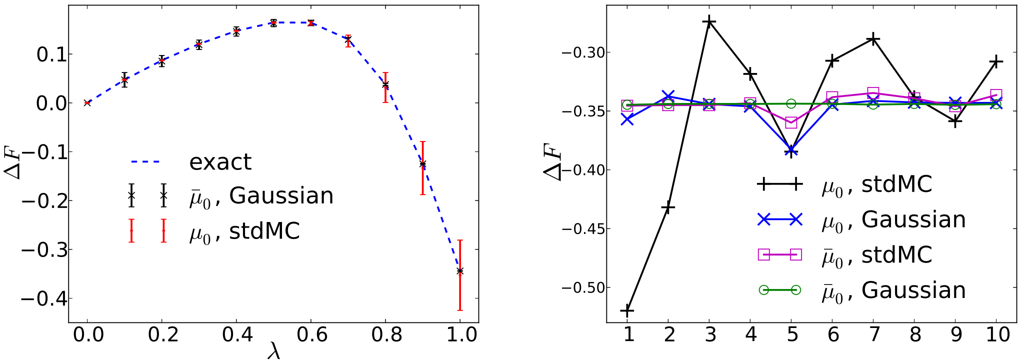
(a)

(b)

Figure 5. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0, 1]$. Probability density functions of work along trajectories estimated from 10 independent runs of Monte Carlo simulations where 5×10^5 trajectories are simulated for each run. (a) "optimal" corresponds to the importance sampling estimator with control u^* starting from the distribution μ_0^*. The other three curves correspond to Monte Carlo estimators with initial distribution $\bar{\mu}_0$, using either the controlled dynamics (135) with piecewise linear ansatz functions (Label "$\bar{\mu}_0$, linear"), Gaussian ansatz functions (Label "$\bar{\mu}_0$, Gaussian"), or the uncontrolled dynamics (131) (Label "$\bar{\mu}_0$, stdMC"). (b) Results correspond to Monte Carlo estimators with initial distribution μ_0, using either the controlled dynamics (135) with piecewise linear ansatz functions (Label "μ_0, linear"), Gaussian ansatz functions (Label "μ_0, Gaussian"), or the uncontrolled dynamics (131) (Label "μ_0, stdMC").

initial	control	mean $\mathcal{I}$	SD $\mathcal{I}$	mean ΔF	SD ΔF	mean W
μ_0^*	optimal	5.58	8.4×10^{-2}	-3.44×10^{-1}	2.4×10^{-4}	-1.85
$\bar{\mu}_0$	linear	5.59	6.0×10^{0}	-3.44×10^{-1}	3.4×10^{-4}	-2.08
	Gaussian	5.59	7.1×10^{0}	-3.44×10^{-1}	3.5×10^{-4}	-2.05
	stdMC	5.51	9.8×10^{1}	-3.41×10^{-1}	5.4×10^{-3}	-0.71
μ_0	linear	5.74	2.2×10^{2}	-3.49×10^{-1}	1.0×10^{-2}	-0.08
	Gaussian	5.71	2.6×10^{2}	-3.48×10^{-1}	1.3×10^{-2}	0.06
	stdMC	6.28	1.7×10^{3}	-3.53×10^{-1}	7.2×10^{-2}	0.40

Table 1. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0, 1]$. Estimations of free energy difference for $\lambda = 1$ using different (importance sampling) Monte Carlo methods. Direct calculation of (10) and (11) gives the reference value $\Delta F = -3.44 \times 10^{-1}$. Column "initial" specifies the initial distribution that are used to generate trajectories in Monte Carlo simulations. Column "control" specifies the different dynamics (different control forces) and the meaning of each name is the same as those appeared in Figure 5. Columns "mean $\mathcal{I}$", "SD $\mathcal{I}$" show the mean and the sample standard deviation of estimators (133) or (137). Columns "mean ΔF", "SD ΔF" show the mean and the sample standard deviation of 10 independent runs of the free energy difference estimations $\Delta F(1)$ using (134). The mean values of work W for different Monte Carlo methods are shown in Column "mean W".



(a) $\Delta F(\lambda)$

(b) $\Delta F(1)$

Figure 6. Example 1 with the control protocol $\lambda(s) = s$, for $s \in [0, 1]$. Labels of different curves have the same meaning as those appeared in Figure 5. (a) Profiles of free energy differences $\Delta F(\lambda)$ for $\lambda \in [0, 1]$. Standard deviations of the free energy difference estimations for 10 independent runs are shown in vertical error bar for different λ. "exact" corresponds to results obtained by directly integrating the normalization constant $Z(\lambda)$ from (10). (b) Mean values of free energy differences at $\lambda = 1$ for 10 independent runs using different (importance sampling) Monte Carlo methods. For each run, 5×10^5 trajectories of either SDE (131) or the controlled SDE (135) are generated with time step-size $\Delta t = 5 \times 10^{-4}$. Results corresponding to piecewise linear ansatz functions are not shown here since they are very similar to those corresponding Gaussian ansatz functions.

where r_{AB}, r_{BC} are the distances between particles A and B, B and C, respectively. θ_{ABC} is the angle spanned by the bonds AB and BC, and V_3 is the potential of angle given by

$$V_3(\theta) = \frac{k_\theta}{2}\Big((\theta-\theta_0)^2 - (\delta\theta)^2\Big)^2 - k_{\theta,1}(\theta-\theta_0)\,, \tag{139}$$

with $k_\theta > 0$. Furthermore, in order to remove rigid body motion invariance, we fix the position of particle B ($x_B = 0$) and particle A is only allowed to move along horizontal axis. For parameters, we take $\theta_0 = \frac{\pi}{3}$, $\delta\theta = \frac{\pi}{6}$, $\epsilon = 0.1$, $k_\theta = 20$, $k_{\theta,1} = 0.3$, and $l_{eq} = 5.0$.

The system essentially has three degree of freedom, i.e., the position of $x_C = (y_1, y_2)$ and the position of $x_A = (y_3, 0)$ on the x-axis. The free energy is defined according to (87), where we take

$$\xi(y_1, y_2, y_3) = \theta_{ABC} = \arctan\frac{y_2}{y_1} \tag{140}$$

as the reaction coordinate function and $\beta = 5.0$. In order to calculate free energy differences, we consider the dynamics $y(s) = (y_1(s), y_2(s), y_3(s))$ in (118) during the time interval $[0, 1]$ with $a = \sigma = \mathrm{id}$, and $f \equiv \frac{\pi}{3}$, starting from $\theta(y(0)) = \frac{\pi}{6}$ at time $s = 0$. In this case, the projection matrix in (89) can be directly computed as

$$P = \begin{pmatrix} \frac{y_1^2}{y_1^2+y_2^2} & \frac{y_1y_2}{y_1^2+y_2^2} & 0 \\ \frac{y_1y_2}{y_1^2+y_2^2} & \frac{y_2^2}{y_1^2+y_2^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{141}$$

and we have $\Psi = |\nabla\xi|^2 = \frac{1}{y_1^2+y_2^2}$ in (88). The angle θ_{ABC} of the system $y(s)$ evolves uniformly during time $s \in [0, 1]$ from $\frac{\pi}{6}$ to $\frac{\pi}{2}$ and the free energy at $\theta_{ABC} = \frac{\pi}{6}$ is taken as reference. The free energy differences are calculated based on the Jarzynski-like identity (106), where the work W is given in (115) and becomes as simple as

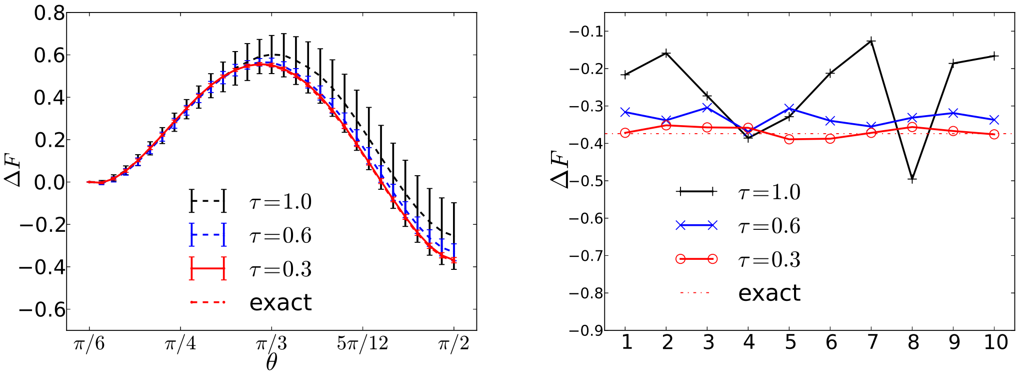
$$W(t) = \int_0^t \Big(-y_2\frac{\partial V}{\partial y_1} + y_1\frac{\partial V}{\partial y_2}\Big)(y(s))\,ds\,. \tag{142}$$

In the numerical experiment below, we take $\kappa = 0.3$, 0.6 in the potential V in (138) and the performance of the Monte Carlo estimator is tested using different values $\tau = 1.0$, 0.6, 0.3 in dynamics (118). In each case, we estimate the free energy differences based on 10 independent runs of Monte Carlo sampling of

$$\Delta F(\theta(t)) \approx -\beta^{-1}\ln\mathcal{I}(\theta(t)) = -\beta^{-1}\ln\Big(\frac{1}{N}\sum_{i=1}^{N} e^{-\beta W_i(t)}\Big)\,, \tag{143}$$

where $\theta(t) = \frac{\pi}{6} + \frac{\pi}{3}t$. In each run, $N = 5\times10^5$ trajectories of dynamics (118) are simulated using time step-size $\Delta t = 10^{-4}$, where W_i denotes the work (142) of the ith trajectory.

The numerical results are shown in Figure 7 , Figure 8 (results for $\kappa = 0.3$ are similar and therefore are not displayed) and Table 2. From both Figure 7 and Table 2, we can observe that the free energy calculation using $\tau = 1.0$ lead to large fluctuations and inaccurate estimations. On the other hand, by decreasing τ to 0.3, the variance of 10 independent runs of free energy calculation decreases significantly and the results become stable. Based on the 10 runs of Monte Carlo simulations of the nonequilibrium dynamics, we can also estimate the probability density functions of the work (142) and the results are shown in Figure 8. It can be seen that, as τ decreases, the probability density functions shift along the negative horizontal axis and become more concentrated. This indicates that the work of the sampled paths becomes smaller on average and the variance decreases. All these results confirm that variance of the Monte Carlo estimator can be reduced by decreasing the value of τ (see discussions at the end of Subsection 3.4).



(a) $\Delta F(\theta)$

(b) $\Delta F(\frac{\pi}{2})$

Figure 7. Example 2 for $\kappa = 0.6$. (a) Profiles of free energy differences $\Delta F(\theta)$ for $\theta = \theta_{ABC} \in [\frac{\pi}{6}, \frac{\pi}{2}]$ computed using different τ in (118). Standard deviations of the free energy difference estimations for 10 independent runs are shown in vertical error bar for different θ. "exact" corresponds to the reference results obtained by directly integrating the normalization constants $Q(\cdot)$ appeared in (86). Curves with Label "$\tau = 0.3$" and Label "exact" almost coincide. (b) Mean values of free energy differences at $\theta = \frac{\pi}{2}$ for 10 runs of Monte Carlo simulations using different values of τ in (118). The horizontal line with Label "exact" corresponds to the reference value $\Delta F(\frac{\pi}{2}) = -3.74 \times 10^{-1}$. For each run, 5×10^5 trajectories of SDE (118) are generated with time step size $\Delta t = 10^{-4}$.

5 Conclusions

In this work, we have studied nonequilibrium theorems for diffusion processes. Jarzynski's equalities and fluctuation theorems are proved for quite general types of diffusion processes in both the alchemical transition case and the reaction coordinate case. The information-theoretic formulation of the Jarzynski's equality, as well as variance reduction approaches are discussed in both cases. Our mathematical tools to derive these nonequilibrium relations are from the theory of stochastic differential equation, in particular the Feynman-Kac formula and the Girsanov's theorem. An advantage of the approach is that, it enables us to elucidate the connections between Jarzynski's equality and the thermodynamic integration identity, which were often treated as two distinct free energy calculation methods.

Two variance reduction approaches for Monte Carlo methods have been studied in order to compute free energy differences using Jarzynski's equality. As demonstrated by simple examples, these approaches can largely improve the efficiency of Monte Carlo estimators in both

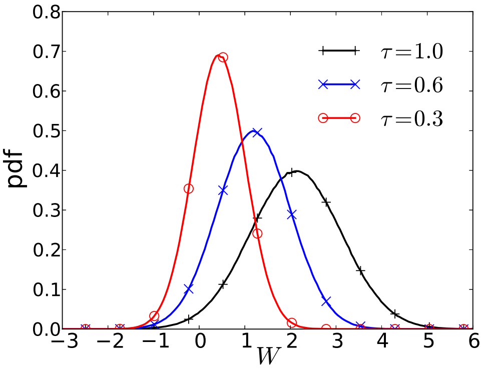


Figure 8. Example 2 for $\kappa = 0.6$. Probability density functions of the work W (142) along trajectories of (118) for different values $\tau = 1.0, 0.6, 0.3$. For each τ, the probability density function is estimated from 10 runs of Monte Carlo simulations where 5×10^5 trajectories are simulated in each run.

κ	τ	mean $\mathcal{I}$	SD $\mathcal{I}$	mean ΔF	SD ΔF	mean W
	1.0	5.46	9.4×10^1	-3.39×10^{-1}	1.4×10^{-2}	0.29
0.3	0.6	5.67	4.0×10^1	-3.47×10^{-1}	1.1×10^{-2}	0.05
	0.3	5.52	1.5×10^1	-3.42×10^{-1}	2.9×10^{-3}	-0.13
	1.0	4.27	2.0×10^3	-2.55×10^{-1}	1.6×10^{-1}	2.14
0.6	0.6	5.28	5.1×10^2	-3.32×10^{-1}	4.0×10^{-2}	1.22
	0.3	6.33	2.3×10^2	-3.69×10^{-1}	1.3×10^{-2}	0.46

Table 2. Example 2. Estimations of free energy difference for $\theta = \frac{\pi}{2}$ using Monte Carlo methods for different values κ and τ. Direct calculation of (10) and (11) gives the reference value $\Delta F(\frac{\pi}{2}) = -3.42 \times 10^{-1}$ and $\Delta F(\frac{\pi}{2}) = -3.74 \times 10^{-1}$ for $\kappa = 0.3$ and 0.6, respectively. Columns "mean $\mathcal{I}$", "SD $\mathcal{I}$" show the mean and the sample standard deviation of the estimator $\mathcal{I}$ in (143). Columns "mean ΔF", "SD ΔF" show the mean and the sample standard deviation of 10 runs of free energy difference estimations $\Delta F(\frac{\pi}{2})$ using (143). The mean values of the work W for Monte Carlo simulations using different κ and τ are shown in the Column "mean W".

the alchemical transition case and the reaction coordinate case. One of the key findings is that variance reduction by a change of measure requires to change both the initial distribution and the equation of the dynamics. We expect that our simple numerical studies can provide some insights into the source of sampling variances.

While the current work focuses on diffusion processes, the mathematical tools may be applicable to other types of stochastic processes, such as Markov chains, particle systems or networks, whose evolution depends on external parameters. In future work, we will also investigate free energy calculation for high-dimensional applications using the variance reduction approaches proposed in this work, together with the recent techniques of solving high-dimensional PDEs [17, 9, 21].

Acknowledgement

The authors acknowledge financial support by the Einstein Center of Mathematics (ECMath) through project CH21.

A Connections with thermodynamic integration and adiabatic switching : Alchemical transition case

In this appendix, we study two (essentially equivalent) asymptotic regimes of nonequilibrium processes using formal arguments. In particular, we will derive the thermodynamic integration identity from Jarzynski's identity, therefore bridging these two different free energy calculation methods. Let us point out that such a connection is indeed known in physics community [14], although we are not aware of its mathematical derivation in the literature. For simplicity, we only consider the alchemical transition case studied in Section 2 and assume the protocol $\lambda(\cdot)$ is deterministic with $\epsilon = 0$.

From Jarzynski's equality to thermodynamic integration Thermodynamic integration is a well known method and has been widely used to compute free energy differences [24]. From the definition of the normalization constant $Z(\cdot)$ in (10), we can derive the thermodynamic integration identity by the simple argument

$$\begin{aligned}\Delta F(T) =& F(\lambda(T)) - F(\lambda(0))\\ =& -\beta^{-1} \ln \frac{Z(\lambda(T))}{Z(\lambda(0))}\\ =& -\beta^{-1} \int_0^T \frac{d}{ds}\Big(\ln \frac{Z(\lambda(s))}{Z(\lambda(0))}\Big)\, ds\\ =& \int_0^T \Big(\frac{\int_{\mathbb{R}^n} e^{-\beta V(x,\lambda(s))}\nabla_\lambda V(x,\lambda(s))\,dx)}{Z(\lambda(s))}\Big) \cdot f(\lambda(s),s)\, ds\\ =& \int_0^T \big(\mathbf{E}_{\mu_{\lambda(s)}}(\nabla_\lambda V)\big) \cdot f(\lambda(s),s)\, ds\,. \end{aligned} \tag{144}$$

In the following, using a formal argument, we show that the identity (144) corresponds to the

Jarzynski's equality (29) in certain asymptotic limit. For this purpose, we consider the dynamics

$$dx(s) = \frac{1}{\tau} b(x(s), \lambda(s))\, ds + \sqrt{\frac{2\beta^{-1}}{\tau}}\sigma(x(s), \lambda(s))\, dw^{(1)}(s)\,, \tag{145}$$

on $s \in [0, T]$, where $0 < \tau \ll 1$ and $\lambda(s)$ satisfies the ODE

$$\dot{\lambda}(s) = f(\lambda(s), s)\,. \tag{146}$$

Clearly, dynamics (145) is related to (1) by rescaling time with the parameter $0 < \tau \ll 1$, and its infinitesimal generator is $\frac{1}{\tau}\mathcal{L}_1$, where $\mathcal{L}_1$ is defined in (6) with $\lambda(\cdot)$ being time dependent. The main observation is that, repeating the argument from Subsection 2.2, the Jarzynski's equality (29) holds for (145) and (146) for any $\tau > 0$. As a consequence,

$$e^{-\beta \Delta F(T)} = \mathbf{E}_{\mu(\lambda(0))}\Big(g(\cdot, \lambda(0), 0)\Big)\,, \tag{147}$$

where the function g now satisfies

$$\begin{aligned} &\partial_t g + \frac{1}{\tau}\mathcal{L}_1 g + f \cdot \nabla_\lambda g - \beta\big(f \cdot \nabla_\lambda V\big) g = 0\,, \quad 0 \le t < T\,, \\ &g(\cdot, \cdot, T) = 1\,. \end{aligned} \tag{148}$$

To show that (147) reduces to the thermodynamic integration identity (144) as $\tau \to 0$, it is enough to study the asymptotic limit of (148). To this end, we consider the formal asymptotic expansion

$$g = g_0 + \tau g_1 + \tau^2 g_2 + \cdots$$

as $\tau \to 0$, where $g_0, g_1, \cdots$ are functions independent of τ. Substituting this expansion into (148) and comparing terms of different powers of τ, we can conclude that $g_0 = g_0(\lambda, t)$ is independent of x and satisfies

$$\begin{aligned} &\partial_t g_0 + \mathcal{L}_1 g_1 + f \cdot \nabla_\lambda g_0 - \beta(f \cdot \nabla_\lambda V) g_0 = 0\,, \quad 0 \le t < T \\ &g_0(\cdot, T) = 1\,. \end{aligned} \tag{149}$$

Taking the expectation with respect to μ_λ on both sides of (149) and noticing that $\mathbf{E}_{\mu_\lambda}(\mathcal{L}_1 g_1) = 0$, we obtain

$$\begin{aligned} &\partial_t g_0 + f \cdot \nabla_\lambda g_0 - \beta\big(f \cdot \mathbf{E}_{\mu_\lambda}(\nabla_\lambda V)\big) g_0 = 0\,, \quad 0 \le t < T \\ &g_0(\cdot, T) = 1\,. \end{aligned} \tag{150}$$

It is easy to verify that the solution of (150) is given by

$$g_0(\lambda, t) = e^{-\beta \int_t^T \big(\mathbf{E}_{\mu_{\lambda(s)}}(\nabla_\lambda V)\big)\cdot f(\lambda(s), s)\, ds}\,, \tag{151}$$

where $\lambda(s)$ satisfies (146) with initial value $\lambda(t) = \lambda$. Taking the limit $\tau \to 0$ in (147) then yields

$$e^{-\beta \Delta F(T)} = \lim_{\tau \to 0} \mathbf{E}_{\mu(\lambda(0))}\Big(g(\cdot, \lambda(0), 0)\Big) = g_0(\lambda(0), 0) = e^{-\beta \int_0^T \big(\mathbf{E}_{\mu_{\lambda(s)}}(\nabla_\lambda V)\big)\cdot f(\lambda(s), s)\, ds}\,, \tag{152}$$

which is equivalent to the thermodynamic integration identity (144).

Adiabatic switching Now we turn to another (equivalent) asymptotic regime where the protocol $\lambda(\cdot)$ is switched infinitely slowly. Specifically, given $\lambda_0, \lambda_1 \in \mathbb{R}^m$, the protocol $\lambda(\cdot)$ satisfying $\lambda(0) = \lambda_0$ and $\lambda(T) = \lambda_1$ as $T \to +\infty$ is called adiabatic switching. For the nonequilibrium process $x(\cdot)$ in (1) under adiabatic switching, it is well known that we have

$$F(\lambda_1) - F(\lambda_0) = \lim_{T\to+\infty} \mathbf{E}_{\lambda_0,0}\big(W(T)\big) = \lim_{T\to+\infty} \mathbf{E}_{\lambda_0,0}\Big(\int_0^T \nabla_\lambda V(x(s),\lambda(s))\cdot f(\lambda(s),s)\,ds\Big)\,, \tag{153}$$

i.e., the free energy difference equals to the average work performed during the switching. In the following we provide a formal mathematical argument to derive the above identity. For this purpose, we define

$$u(x,\lambda,t) = \mathbf{E}\Big(\int_t^T \nabla_\lambda V(x(s),\lambda(s))\cdot f(\lambda(s),s)\,ds\,\Big|\, x(t)=x, \lambda(t)=\lambda\Big)\,, \tag{154}$$

which, by the Feynman-Kac formula, satisfies

$$\begin{aligned} &\partial_t u + \mathcal{L}_1 u + f\cdot\nabla_\lambda u + f\cdot\nabla_\lambda V = 0\,,\\ &u(\cdot,\cdot,T) = 0\,. \end{aligned} \tag{155}$$

Notice that, as $T \to +\infty$, the switching becomes infinitely slow and $\dot\lambda(t) = f$ goes to zero. Instead, we rescale the time by $\bar t = \frac{t}{T} \in [0,1]$ and define $\bar\lambda(\bar t) = \lambda(\frac{\bar t}{\tau})$, where $\tau = \frac{1}{T} \to 0$. $\bar\lambda(\cdot)$ satisfies $\bar\lambda(0) = \lambda_0, \bar\lambda(1) = \lambda_1$ and

$$\frac{d\bar\lambda}{d\bar t} = \bar f(\bar\lambda(\bar t), \bar t)\,, \tag{156}$$

where $\bar f(\cdot,\bar t) = \frac{1}{\tau} f(\cdot, \frac{\bar t}{\tau})$ is a function of $\mathcal{O}(1)$. Under this time scaling, PDE (155) becomes

$$\begin{aligned} &\partial_{\bar t} u + \frac{1}{\tau}\mathcal{L}_1 u + \bar f\cdot\nabla_\lambda u + \bar f\cdot\nabla_\lambda V = 0\,, \quad 0 \le \bar t < 1\,,\\ &u \equiv 0\,, \quad \bar t = 1\,. \end{aligned} \tag{157}$$

Consider the expansion $u = u_0 + \tau u_1 + \tau^2 u_2 + \cdots$, then the same argument as above yields that the function u_0 is independent of x and satisfies

$$\begin{aligned} &\partial_{\bar t} u_0 + \bar f\cdot\nabla_\lambda u_0 + \bar f\cdot\mathbf{E}_\lambda\big(\nabla_\lambda V\big) = 0\,, \quad 0 \le \bar t < 1\,,\\ &u_0 \equiv 0\,, \quad \bar t = 1\,. \end{aligned} \tag{158}$$

The solution of (158) can be directly computed:

$$u_0(\lambda, \bar t) = \int_{\bar t}^1 \mathbf{E}_{\bar\lambda(s)}\big(\nabla_\lambda V\big)\cdot \bar f(\bar\lambda(s), s)\,ds\,, \tag{159}$$

where $\bar\lambda(\cdot)$ satisfies (156) on $[\bar t, 1]$ with $\bar\lambda(\bar t) = \lambda$. In particular, taking $\bar t = 0$ and applying the thermodynamic integration identity (144), gives

$$u_0(\lambda_0, 0) = \int_0^1 \mathbf{E}_{\bar\lambda(s)}\big(\nabla_\lambda V\big)\cdot \bar f(\bar\lambda(s), s)\,ds = F(\lambda_1) - F(\lambda_0)\,. \tag{160}$$

Therefore,

$$\begin{aligned}&\lim_{T\rightarrow+\infty}\mathbf{E}_{\lambda_0,0}\Big(\int_0^T \nabla_\lambda V(x,\lambda(s))\cdot f(\lambda(s),s)\,ds\Big)\\=&\lim_{\tau\rightarrow 0}\mathbf{E}_{\mu_{\lambda_0}}\big(u(\cdot,\lambda_0,0)\big)\\=&u_0(\lambda_0,0)=F(\lambda_1)-F(\lambda_0)\,,\end{aligned}$$

which concludes the proof of (153).

B Thermodynamic integration identity in the reaction coordinate case

In the reaction coordinate case considered in Section 3, connections between the thermodynamic integration identity and the Jarzynski's equality as well as the adiabatic switching regime can be studied using the same asymptotic argument as in Appendix A. In this section, we omit the derivation and only provide the thermodynamic integration identity. We emphasize that both the identity and its proof can be found in the literature, e.g., [45, 43]. The result is included for readers' convenience.

Recall the definition of the probability measure μ_z in (86), where the normalization constant is given by

$$Q(z)=\int_{\mathbb{R}^n} e^{-\beta V(y)}\delta\big(\xi(y)-z\big)\,dy\,,\quad z\in\mathbb{R}^d\,,\tag{161}$$

and the free energy is defined in (87). Let $z(s)\in\mathbb{R}^d$ satisfy the ODE (101) on $[0,T]$. Similar to the derivations in (144), and using Lemma 3 below, we can compute

$$\begin{aligned}&F(z(T))-F(z(0))\\=&-\beta^{-1}\ln\frac{Q(z(T))}{Q(z(0))}\\=&-\beta^{-1}\int_0^T\frac{d}{ds}\Big(\ln\frac{Q(z(s))}{Q(z(0))}\Big)\,ds\\=&-\beta^{-1}\int_0^T\Big(\frac{1}{Q}\frac{\partial Q}{\partial z_\gamma}\Big)\big(z(s)\big)\,\dot{z}_\gamma(s)\,ds\\=&\int_0^T\mathbf{E}_{\mu_{z(s)}}\Big[(a\nabla\xi_{\gamma'})_i(\Psi^{-1})_{\gamma'\gamma}\frac{\partial V}{\partial y_i}-\frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((a\nabla\xi_{\gamma'})_i(\Psi^{-1})_{\gamma\gamma'}\Big)\Big]\,\dot{z}_\gamma(s)\,ds\,,\end{aligned}\tag{162}$$

where Einstein's summation convention has been used.

Lemma 3. *Let the function Q be defined in (161). For $1\le\gamma\le d$, we have*

$$\frac{\partial Q}{\partial z_\gamma}(z)=-\beta Q(z)\int_{\Sigma_z}\Big[(a\nabla\xi_{\gamma'})_i(\Psi^{-1})_{\gamma\gamma'}\frac{\partial V}{\partial y_i}-\frac{1}{\beta}\frac{\partial}{\partial y_i}\Big((a\nabla\xi_{\gamma'})_i(\Psi^{-1})_{\gamma\gamma'}\Big)\Big]\,\mu_z(dy)\,.$$

Proof. Let $\varphi:\mathbb{R}^d\rightarrow\mathbb{R}$ be a smooth test function with compact support. For $1\le\gamma\le d$, integrating by parts and using (161), we have

$$\int_{\mathbb{R}^d}\varphi(z)\frac{\partial Q}{\partial z_\gamma}(z)\,dz=-\int_{\mathbb{R}^d}\frac{\partial\varphi}{\partial z_\gamma}(z)Q(z)\,dz=-\int_{\mathbb{R}^n}\frac{\partial\varphi}{\partial z_\gamma}(\xi(y))e^{-\beta V(y)}\,dy\,.\tag{163}$$

On the other hand, from the relation

$$\frac{\partial\big(\varphi\circ\xi\big)}{\partial y_i}(y) = \frac{\partial\varphi}{\partial z_{\gamma'}}(\xi(y))\frac{\partial\xi_{\gamma'}}{\partial y_i}(y), \quad 1 \le i \le n\,,$$

and the definition of the $d\times d$ matrix Ψ in (88), we obtain

$$\frac{\partial\varphi}{\partial z_\gamma}(\xi(y)) = \Big[\frac{\partial\big(\varphi\circ\xi\big)}{\partial y_i}a_{ij}\frac{\partial\xi_{\gamma'}}{\partial y_j}(\Psi^{-1})_{\gamma\gamma'}\Big](y)\,. \tag{164}$$

Therefore, integrating by parts, (163) simplifies to

$$\begin{aligned}&\int_{\mathbb{R}^d}\varphi(z)\frac{\partial Q}{\partial z_\gamma}(z)\,dz\\ =&\int_{\mathbb{R}^n}\varphi(\xi(y))\frac{\partial}{\partial y_i}\Big(a_{ij}\frac{\partial\xi_{\gamma'}}{\partial y_j}(\Psi^{-1})_{\gamma\gamma'}e^{-\beta V(y)}\Big)\,dy\\ =&\int_{\mathbb{R}^d}\varphi(z)\Big[\int_{\mathbb{R}^n}\frac{\partial}{\partial y_i}\Big(a_{ij}\frac{\partial\xi_{\gamma'}}{\partial y_j}(\Psi^{-1})_{\gamma\gamma'}e^{-\beta V(y)}\Big)\,\delta(\xi(y)-z)dy\Big]dz\,,\end{aligned}$$

from which we can conclude after simplification. □

C An alternative proof of Theorem 2

In this appendix, we provide an alternative proof of Theorem 2. Different from the proof in Subsection 2.3 where only the Feynman-Kac formula has been used, the proof below relies on the combination of both the Feynman-Kac formula and Girsanov's Theorem. While the idea is inspired by the derivations in [10], the proof below is shorter.

Alternative proof of Theorem 2. First of all, we recall the definition of u in (43) as well as the equations (40), (44), (45) used in the proof of Theorem 2 in Subsection 2.3. In accordance with (45), we define

$$\overline{\mathcal{L}} = \Big(J + a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\cdot\nabla + \frac{1}{\beta}a:\nabla^2 + f\cdot\nabla_\lambda + \epsilon\,\alpha\alpha^T:\nabla^2_\lambda\,, \tag{165}$$

and consider the function $\omega(x,\lambda,t) = u\big(x,\lambda,T-t\,;x',\lambda',t'\big)$. From (44) and (45), we know that ω satisfies

$$\begin{aligned}&\frac{\partial\omega}{\partial t} + \overline{\mathcal{L}}_{(x,\lambda,t)}\omega + \Big[\mathrm{div}(J+a\nabla V) + \mathrm{div}_\lambda\Big(f - \epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\Big) + \eta\Big]\omega = 0\,, \quad \forall t\in[0,T-t')\,,\\ &\omega(x,\lambda,t) = \delta(x'-x)\delta(\lambda'-\lambda)\,, \quad t = T-t'\,,\end{aligned} \tag{166}$$

where $(x,\lambda)\in\mathbb{R}^n\times\mathbb{R}^m$ and $\overline{\mathcal{L}}_{(x,\lambda,t)}$ is the operator (165) evaluated at (x,λ,t). On the other hand, applying the Feynman-Kac formula to (166), we observe that

$$\begin{aligned}\omega(x,\lambda,t) =&\overline{\mathbf{E}}_{x,\lambda,t}\bigg[\exp\bigg(\int_t^{T-t'}\Big(\mathrm{div}\big(J+a\nabla V\big) + \mathrm{div}_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big) + \eta\Big)\big(\bar{x}(s),\bar{\lambda}(s),s\big)ds\bigg)\\ &\times\delta\big(\bar{x}(T-t')-x'\big)\delta\big(\bar{\lambda}(T-t')-\lambda'\big)\bigg]\,,\end{aligned} \tag{167}$$

where $\overline{\mathbf{E}}_{x,\lambda,t}$ denotes the conditional expectation under the path ensemble of the dynamics

$$d\bar{x}(s) =\Big(J + a\nabla V + \frac{1}{\beta}\nabla\cdot a\Big)\big(\bar{x}(s),\bar{\lambda}(s)\big)\,ds + \sqrt{2\beta^{-1}}\sigma\big(\bar{x}(s),\bar{\lambda}(s)\big)\,dw^{(1)}(s) \tag{168}$$

and the control protocol

$$d\bar{\lambda}(s) =f(\bar{x}(s),\bar{\lambda}(s),s)\,ds + \sqrt{2\epsilon}\,\alpha\big(\bar{x}(s),\bar{\lambda}(s),s\big)dw^{(2)}(s)\,, \tag{169}$$

starting from $\bar{x}(t) = x$ and $\bar{\lambda}(t) = \lambda$ at time t. Note that the infinitesimal generator of the dynamics (168) and (169) is given by the operator $\overline{\mathcal{L}}$ in (165).

Now we apply Girsanov's theorem to change the probability measure in (167) from the path ensemble of the dynamics (168), (169) to the path ensemble of the dynamics (15), (3). Specifically, starting from (x,λ) at time t, let $\mathbf{P}_{x,\lambda}$ and $\overline{\mathbf{P}}_{x,\lambda}$ denote the path measures on the time interval $[t, T-t']$ corresponding to (15), (3) and (168), (169), respectively. Applying Girsanov's theorem, we obtain after some straightforward calculations

$$\begin{aligned}\frac{d\mathbf{P}_{x,\lambda}}{d\overline{\mathbf{P}}_{x,\lambda}}\big(x(\cdot),\lambda(\cdot)\big) =&\exp\bigg[-\beta\int_t^{T-t'}\nabla V\big(x(s),\lambda(s)\big)\cdot dx(s)\\ &+\beta\int_t^{T-t'}\Big(\nabla V\cdot\big(J+\frac{1}{\beta}\nabla\cdot a\big)\Big)\big(x(s),\lambda(s)\big)\,ds\bigg]\,.\end{aligned} \tag{170}$$

Therefore, changing the probability measure in (167) from $\overline{\mathbf{P}}_{x,\lambda}$ to $\mathbf{P}_{x,\lambda}$, using (170), (13), we find

$$\begin{aligned}&u(x,\lambda,T-t) = \omega(x,\lambda,t)\\ =&\mathbf{E}_{x,\lambda,t}\bigg[\exp\bigg(\int_t^{T-t'}\Big(\mathrm{div}\big(J+a\nabla V\big)+\mathrm{div}_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big)+\eta\Big)\big(x(s),\lambda(s),s\big)ds\bigg)\\ &\times\delta\big(x(T-t')-x'\big)\delta\big(\lambda(T-t')-\lambda'\big)\frac{d\overline{\mathbf{P}}_{x,\lambda}}{d\mathbf{P}_{x,\lambda}}\big(x(\cdot),\lambda(\cdot)\big)\bigg]\\ =&\mathbf{E}_{x,\lambda,t}\bigg[\exp\bigg(\beta\int_t^{T-t'}\nabla V(x(s),\lambda(s))\cdot dx(s)+\int_t^{T-t'}\big(a:\nabla^2 V\big)\big(x(s),\lambda(s)\big)ds\\ &+\int_t^{T-t'}\Big(\mathrm{div}_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big)+\eta\Big)\big(x(s),\lambda(s),s\big)ds\bigg)\delta\big(x(T-t')-x'\big)\,\delta\big(\lambda(T-t')-\lambda'\big)\bigg]\\ =&\mathbf{E}_{x,\lambda,t}\bigg[\exp\bigg(\beta\int_t^{T-t'}\nabla V\big(x(s),\lambda(s)\big)\circ dx(s)+\int_t^{T-t'}\Big(\mathrm{div}_\lambda\big(f-\epsilon\nabla_\lambda\cdot(\alpha\alpha^T)\big)+\eta\Big)\big(x(s),\lambda(s),s\big)ds\bigg)\\ &\times\delta\big(x(T-t')-x'\big)\,\delta\big(\lambda(T-t')-\lambda'\big)\bigg]\,.\end{aligned}$$

Note that in the last equality above, we have converted Ito integration to Stratonovich integration according to (15). Substituting t by $T-t$, integrating by parts, and recalling the expression (43), we obtain

$$\begin{aligned}&e^{-\beta V(x',\lambda')}\,\mathbf{E}^R_{x',\lambda',t'}\bigg[\exp\bigg(\int_{t'}^{t}\eta(x^R(s),\lambda^R(s),T-s)ds\bigg)\delta\big(x^R(t)-x\big)\delta\big(\lambda^R(t)-\lambda\big)\bigg]\\ =&e^{-\beta V(x,\lambda)}\,\mathbf{E}_{x,\lambda,T-t}\bigg[e^{-\beta\mathcal{W}}\exp\bigg(\int_{T-t}^{T-t'}\eta(x(s),\lambda(s),s)ds\bigg)\delta\big(x(T-t')-x'\big)\delta\big(\lambda(T-t')-\lambda'\big)\bigg]\,,\end{aligned}$$

where $\mathcal{W}$ is defined in (42). □

D Proof of Theorem 3

In this appendix, we provide the proof of Theorem 3 in Subsection 3.2.

Proof of Theorem 3. We consider the quantities on both sides of the equality (98). For the left hand side of (98), let us fix $(y',t') \in \mathbb{R}^n \times [0,T]$ and define the function u by

$$u(y,t\,;y',t') = \mathbf{E}^R_{y',t'}\bigg[\exp\bigg(\int_{t'}^{t} \eta\big(y^R(s), T-s\big)ds\bigg)\delta\big(y^R(t)-y\big)\bigg]\,, \tag{171}$$

for $(y,t) \in \mathbb{R}^n \times [t',T]$. It is known that u satisfies the PDE

$$\begin{aligned}&\frac{\partial u}{\partial t} = \big(\mathcal{L}^R\big)^* u + \eta(y,T-t)\,u\,, \quad \forall\ (y,t) \in \mathbb{R}^n \times (t',T]\,,\\ &u(y,t\,;y',t') = \delta(y-y')\,, \quad \text{if} \ \ t=t'\,,\end{aligned} \tag{172}$$

where the operator $\mathcal{L}^R$ is defined in (97) and $\big(\mathcal{L}^R\big)^*$ denotes its formal L^2 adjoint. A direct calculation shows that

$$\begin{aligned}\big(\mathcal{L}^R\big)^*\phi =& \bigg[\frac{\partial}{\partial y_i}\Big((Pa)_{ij}\frac{\partial V}{\partial y_j}\Big) + \frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\Big)\bigg]\phi\\ &+\bigg[(Pa)_{ij}\frac{\partial V}{\partial y_j} + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j} + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\bigg]\frac{\partial \phi}{\partial y_i} + \frac{1}{\beta}(Pa)_{ij}\frac{\partial^2\phi}{\partial y_i\partial y_j}\,,\end{aligned} \tag{173}$$

for a smooth function ϕ.

For the right hand side of (98), fixing $(y',t') \in \mathbb{R}^n \times [0,T]$, we define the function g for $(y,t) \in \mathbb{R}^n \times [t',T]$ as

$$g(y,t) = \mathbf{E}_{y,T-t}\bigg[e^{-\beta\mathcal{W}}\exp\bigg(\int_{T-t}^{T-t'} \eta\big(y(s),s\big)ds\bigg)\delta\big(y(T-t')-y'\big)\bigg]\,,$$

where $\mathcal{W}$ is defined in (99), and the dynamics $y(\cdot)$ satisfies the SDE (93). Using the same argument as in Lemma 1, we can verify that g satisfies the PDE

$$\begin{aligned}&\frac{\partial g}{\partial t} = \overline{\mathcal{L}}\,g + \eta(\cdot,T-t)g\,, \qquad \forall\,(y,t) \in \mathbb{R}^n \times (t',T]\,,\\ &g(y,t) = \delta(y-y')\,, \qquad \text{if} \ \ t=t'\,,\end{aligned} \tag{174}$$

where the operator $\overline{\mathcal{L}}$ is defined as

$$\begin{aligned}\overline{\mathcal{L}}\,\phi =& \bigg[-\beta(\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\frac{\partial V}{\partial y_i} + \frac{\partial}{\partial y_i}\Big((\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\Big)\bigg]\phi\\ &+\mathcal{L}^\perp\phi + (\Psi^{-1})_{\gamma\gamma'}(a\nabla\xi_\gamma)_i f^-_{\gamma'}\frac{\partial\phi}{\partial y_i}\end{aligned} \tag{175}$$

for a smooth function ϕ. Now consider the function $\omega(y,t) = e^{-\beta V(y)}g(y,t)$. A direct calculation shows that

$$\begin{aligned}e^{-\beta V}\mathcal{L}^\perp g =& e^{-\beta V}\bigg[-(Pa)_{ij}\frac{\partial V}{\partial y_j}\frac{\partial\big(e^{\beta V}\omega\big)}{\partial y_i} + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j}\frac{\partial\big(e^{\beta V}\omega\big)}{\partial y_i} + \frac{1}{\beta}(Pa)_{ij}\frac{\partial^2\big(e^{\beta V}\omega\big)}{\partial y_i\partial y_j}\bigg]\\ =&\bigg[\frac{\partial}{\partial y_i}\Big((Pa)_{ij}\frac{\partial V}{\partial y_j}\Big)\bigg]\omega + \bigg[(Pa)_{ij}\frac{\partial V}{\partial y_j} + \frac{1}{\beta}\frac{\partial (Pa)_{ij}}{\partial y_j}\bigg]\frac{\partial\omega}{\partial y_i} + \frac{1}{\beta}(Pa)_{ij}\frac{\partial^2\omega}{\partial y_i\partial y_j}\,,\\ e^{-\beta V}\frac{\partial g}{\partial y_i} =& e^{-\beta V}\frac{\partial\big(e^{\beta V}\omega\big)}{\partial y_i} = \beta\frac{\partial V}{\partial y_i}\omega + \frac{\partial\omega}{\partial y_i}\,.\end{aligned} \tag{176}$$

Combining (97), (174), (175), (176), it follows that the function ω satisfies the PDE

$$\begin{aligned}
&\frac{\partial \omega}{\partial t} = e^{-\beta V}\Big[\overline{\mathcal{L}}\, g + \eta(\cdot, T-t) g\Big] = \big(\mathcal{L}^R\big)^* \,\omega + \eta(y, T-t)\,\omega\,, \quad \forall\, (y,t) \in \mathbb{R}^n \times (t', T]\,,\\
&\omega(y,t) = e^{-\beta V(y')}\delta(y-y')\,, \quad \text{if}\ \ t = t'\,.
\end{aligned}$$

Comparing this with the equation of function u in (172), we obtain

$$e^{-\beta V(y')} u(y,t\,; y', t') = \omega(y,t),$$

which is equivalent to (98). □

References

[1] C. Abrams and G. Bussi. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, 16(1):163–199, 2014.

[2] D. Abreu and U. Seifert. Extracting work from a single heat bath through feedback. *EPL (Europhysics Letters)*, 94(1):10001, 2011.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. Birkhäuser, 2005.

[4] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi. Optimal protocols and optimal transport in stochastic thermodynamics. *Phys. Rev. Lett.*, 106:250601, 2011.

[5] A. Banyaga and D. Hurtubise. *Lectures on Morse Homology*. Texts in the Mathematical Sciences. Springer Netherlands, 2004.

[6] C. H. Bennett. Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.*, 22(2):245 – 268, 1976.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] H. Callen. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, 1985.

[9] P. Chaudhari, A. Oberman, S. Osher, S. Soatto, and G. Carlier. Deep relaxation: partial differential equations for optimizing deep neural networks. 2017.

[10] R. Chetrite and K. Gawędzki. Fluctuation relations for diffusion processes. *Commun. Math. Phys.*, 282(2):469–518, 2008.

[11] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.*, 31(8):1569–1582, 2010.

[12] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem*, 6(9):1809–1814, 2005.

[13] G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden. Projection of diffusions on submanifolds: Application to mean force computation. *Comm. Pure Appl. Math.*, 61(3):371–408, 2008.

[14] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.*, 90(5):1481–1487, 1998.

[15] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721–2726, 1999.

[16] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61:2361–2366, 2000.

[17] J. Darbon and S. Osher. Algorithms for overcoming the curse of dimensionality for certain hamilton–jacobi equations arising in control theory and elsewhere. *Res. Math. Sci.*, 3(1):19, 2016.

[18] M. de Koning, W. Cai, A. Antonelli, and S. Yip. Efficient freeenergy calculations by the simulation of nonequilibrium processes. *Computing in Science & Engineering*, 2(3):88–96, 2000.

[19] C. Dellago and G. Hummer. Computing equilibrium free energies using non-equilibrium molecular dynamics. *Entropy*, 16(1):41, 2014.

[20] P. Dupuis, K. Spiliopoulos, and H. Wang. Importance sampling for multiscale diffusions. *Multiscale Model. Simul.*, 10(1):1–27, 2012.

[21] W. E, J. Han, and A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat.*, 5(4):349–380, 2017.

[22] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, 1991.

[23] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*. Springer, 2006.

[24] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science series. Elsevier Science, 2001.

[25] H. Ge and D.-Q. Jiang. Generalized Jarzynski's equality of inhomogeneous multidimensional diffusion processes. *J. Stat. Phys.*, 131(4):675–689, 2008.

[26] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.

[27] J. Gore, F. Ritort, and C. Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U.S.A.*, 100(22):12564–12569, 2003.

[28] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Ito differential. *Probab. Th. Rel. Fields*, 71(4):501–516, 1986.

[29] C. Hartmann, L. Richter, C. Schütte, and W. Zhang. Variational characterization of free energy: Theory and algorithms. *Entropy*, 19(11), 2017.

[30] C. Hartmann, C. Schütte, M. Weber, and W. Zhang. Importance sampling in path space for diffusion processes with slow-fast variables. *Probab. Th. Rel. Fields*, 170:177–228, 2017.

[31] C. Hartmann, C. Schütte, and W. Zhang. Model reduction algorithms for optimal control and importance sampling of diffusions. *Nonlinearity*, 29(8):2298–2326, 2016.

[32] U. G. Haussmann and E. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.

[33] D. A. Hendrix and C. Jarzynski. A "fast growth" method of computing free energy differences. *J. Chem. Phys.*, 114(14):5974–5981, 2001.

[34] J. M. Horowitz and S. Vaikuntanathan. Nonequilibrium detailed fluctuation theorem for repeated discrete feedback. *Phys. Rev. E*, 82:061120, 2010.

[35] G. Hummer and I. G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations. *J. Chem. Phys.*, 118(23):10762–10773, 2003.

[36] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 98(7):3658–3661, 2001.

[37] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018–5035, 1997.

[38] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[39] C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73:046105, 2006.

[40] C. Jarzynski. Nonequilibrium work relations: foundations and applications. *Eur. Phys. J. B*, 64(3):331–340, 2008.

[41] S. G. Krantz and H. R. Parks. *Geometric Integration Theory.* Birkhäuser Boston, 2008.

[42] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(1):255, 1966.

[43] F. Legoll and T. Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23(9):2131–2163, 2010.

[44] T. Lelièvre, M. Rousset, and G. Stoltz. Computation of free energy differences through nonequilibrium stochastic dynamics: The reaction coordinate case. *J. Comput. Phys.*, 222(2):624 – 643, 2007.

[45] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations : a mathematical perspective.* London Hackensack, N.J. Imperial College Press, 2010.

[46] T. Lelièvre, M. Rousset, and G. Stoltz. Langevin dynamics with constraints and computation of free eneregy differences. *Math. Comput.*, 81(280):2071–2125, 2012.

[47] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, New York, NY, USA, 2002.

[48] L. Maragliano and E. Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.*, 426(13):168–175, 2006.

[49] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani. Fluctuationdissipation: Response theory in statistical physics. *Phys. Rep.*, 461(4):111–195, 2008.

[50] D. D. L. Minh and J. D. Chodera. Optimal estimators and asymptotic variances for nonequilibrium path-ensemble averages. *J. Chem. Phys.*, 131(13), 2009.

[51] H. Oberhofer and C. Dellago. Optimum bias for fast-switching free energy calculations. *Comput. Phys. Commun.*, 179(13):41 – 45, 2008. Special issue based on the Conference on Computational Physics 2007CCP 2007.

[52] H. Oberhofer, C. Dellago, and P. L. Geissler. Biased sampling of nonequilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods? *J. Phys. Chem. B*, 109(14):6902–6915, 2005.

[53] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications.* Springer, 5th edition, 2000.

[54] M. Ponmurugan. Generalized detailed fluctuation theorem under nonequilibrium feedback control. *Phys. Rev. E*, 82:031129, 2010.

[55] M. Rousset and G. Stoltz. Equilibrium sampling from nonequilibrium dynamics. *J. Stat. Phys.*, 123(6):1251–1272, 2006.

[56] R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning (Information Science and Statistics).* Springer, 1 edition, 2004.

[57] T. Sagawa and M. Ueda. Generalized Jarzynski equality under nonequilibrium feedback control. *Phys. Rev. Lett.*, 104:090602, 2010.

[58] T. Sagawa and M. Ueda. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.*, 109:180602, 2012.

[59] T. Sagawa and M. Ueda. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E*, 85:021104, 2012.

[60] T. Schmiedl and U. Seifert. Optimal finite-time processes in stochastic thermodynamics. *Phys. Rev. Lett.*, 98:108301, 2007.

[61] K. Spiliopoulos. Large deviations and importance sampling for systems of slow-fast motion. *Appl. Math. Optim.*, 67:123–161, 2013.

[62] H. Then and A. Engel. Computing the optimal protocol for finite-time processes in stochastic thermodynamics. *Phys. Rev. E*, 77:041105, 2008.

[63] S. Vaikuntanathan and C. Jarzynski. Escorted free energy simulations: Improving convergence by reducing dissipation. *Phys. Rev. Lett.*, 100:190601, 2008.

[64] S. Vaikuntanathan and C. Jarzynski. Escorted free energy simulations. *J. Chem. Phys.*, 134(5):054107, 2011.

[65] E. Vanden-Eijnden. Some recent techniques for free energy calculations. *J. Comput. Chem.*, 30(11):1737–1747, 2009.

[66] E. Vanden-Eijnden and J. Weare. Rare event simulation of small noise diffusions. *Comm. Pure Appl. Math.*, 65(12):1770–1803, 2012.

[67] F. M. Ytreberg, R. H. Swendsen, and D. M. Zuckerman. Comparison of free energy methods for molecular systems. *J. Chem. Phys.*, 125(18), 2006.

[68] F. M. Ytreberg and D. M. Zuckerman. Single-ensemble nonequilibrium path-sampling estimates of free energy differences. *J. Chem. Phys.*, 120(23):10876–10879, 2004.

[69] W. Zhang. Ergodic SDEs on submanifolds and related numerical sampling schemes. *submitted*, 2018.

[70] W. Zhang, C. Hartmann, and C. Schütte. Effective dynamics along given reaction coordinates, and reaction rate theory. *Faraday Discuss.*, 195:365–394, 2016.

[71] W. Zhang and C. Schütte. Reliable approximation of long relaxation timescales in molecular dynamics. *Entropy*, 19(7), 2017.

[72] W. Zhang, H. Wang, C. Hartmann, M. Weber, and C. Schütte. Applications of the cross-entropy method to importance sampling and optimal control of diffusions. *SIAM J. Sci. Comput.*, 36(6):A2654–A2672, 2014.