COMMUN. MATH. SCI.
Vol. 15, No. 8, pp. 2279–2302

MULTISCALE GENTLEST ASCENT DYNAMICS FOR SADDLE POINT IN EFFECTIVE DYNAMICS OF SLOW-FAST SYSTEM*

SHUTING GU† AND XIANG ZHOU‡

Abstract. Here we present a multiscale method to calculate the saddle point associated with the effective dynamics arising from a stochastic system which couples slow deterministic drift and fast stochastic dynamics. This problem is motivated by the transition states on free energy surfaces in chemical physics. Our method is based on the gentlest ascent dynamics which couples the position variable and the direction variable and has the local convergence to saddle points. The dynamics of the direction vector is derived in terms of the covariance function with respective to the equilibrium distribution of the fast stochastic process. We apply multiscale numerical methods to efficiently solve the obtained multiscale gentlest ascent dynamics, and discuss the acceleration techniques based on an adaptive idea. The examples of stochastic ordinary and partial differential equations are presented.

Keywords. saddle point, gentlest ascent dynamics, multiscale method.

AMS subject classifications. 65K05, 82B05.

1. Introduction

The following slow-fast system $(X(t),Y(t))\in\mathcal{X}\times\mathcal{Y}$ is a typical dynamic system with two disparate time scales:

$$\begin{cases}\dot{X}^\varepsilon(t)=f(X^\varepsilon,Y^\varepsilon), & \text{(1.1a)}\\ \dot{Y}^\varepsilon(t)=\dfrac{1}{\varepsilon}b(X^\varepsilon,Y^\varepsilon)+\dfrac{1}{\sqrt{\varepsilon}}\sigma(X^\varepsilon,Y^\varepsilon)\eta(t), & \text{(1.1b)}\end{cases}$$

where ε is a small parameter. X^ε is the slow variable and Y^ε is the fast variable. For simplicity, we assume $\mathcal{X}=\mathbb{R}^n$ and $\mathcal{Y}=\mathbb{R}^m$. The functions $f(x,y)=(f_1,f_2,\ldots,f_n):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$ and $b(x,y)=(b_1,b_2,\ldots,b_m):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^m$ are two smooth vector fields. The $m\times m$ diffusion matrix $\sigma(x,y)$ is assumed non-degenerate for all x and y. $\eta(t)$ is the zero-mean Gaussian noise in $\mathbb{R}^m$ with a certain covariance function $\mathrm{E}(\eta(t)\eta(t'))$. For example, $\mathrm{E}(\eta(t)\eta(t'))=\delta(t-t')$ means that η is white noise $\dot{W}$. We mainly consider this white noise case in this paper for easy presentation, though the extension to the case with correlation structure is not difficult. Note that in our model, equation (1.1a) contains no diffusion term and the only random source comes from the Y-dependency of the function f.

Much interest in the above multiscale system is concerned with the effective dynamics of the slow variable, when the fast dynamics can be slaved and eliminated as the parameter ε tends to zero. Indeed, for any fixed $T<\infty$, the slow variable in (1.1), $X^\varepsilon(t)$, converges to a deterministic function $\bar{X}(t)$ satisfying the averaged equation

$$\dot{\bar{X}}(t)=F(\bar{X}(t)),$$

*Received: November 15, 2016; accepted (in revised form): August 23, 2017. Communicated by Qiang Du.

†Department of Mathematics, City University of Hong Kong, Tat Chee Ave, Kowloon, Hong Kong SAR. (shutinggu2-c@my.cityu.edu.hk).

‡Department of Mathematics, City University of Hong Kong,Tat Chee Ave, Kowloon, Hong Kong SAR. (xiang.zhou@city.edu.hk).

The research of XZ was supported by the grants from the Research Grants Council of the Hong Kong SAR, China (Project No. CityU 11304314, 109113 and 11304715).

during the time interval $t \in [0,T]$, for some function $F : \mathcal{X} \to \mathcal{X}$. Under certain stronger conditions, this convergence is also uniform in T. This is a typical result of the averaging principle, which has been developed in many works in the classic mathematical literature (see [19,20,27] and [5,12]). On the application side, one of the most important examples is the *extended Lagrangian method* for the coarse-grained molecular dynamics (refer to [9, 17] and [5]). In this application, the slow variable X usually corresponds to the coarse-grained variables that describe the macroscopic features of the underlying complex system, and the fast variable Y corresponds to the microscopic variables of the full atomistic coordinates. To map the full complex energy surface in $\mathcal{Y}$ into a low dimensional free energy surface in $\mathcal{X}$, it is essential to solve the averaged dynamics of the coarse-grained variable $\bar{X}$ with efficient numerical methods such as the heterogeneous multiscale method (HMM) in [6,7,9]. To faithfully recover the fluctuation for a finite ε, [18] used multiple replicas of the fast process in HMM simulations.

In this note we are interested in the computation of the saddle point in the effective dynamics F rather than the equilibrium fixed point. The exploration of the saddle point may help to investigate the phase space structure for F, which has no closed form at all. Specifically, we are interested in index-1 saddle points for the effective dynamics associated with the slow-fast system (1.1). By index-1, we mean that the unstable manifold of the flow F at the saddle point is one dimensional. The search of saddle point for the effective dynamics F is not an easy job, considering the lack of any analytical form of this dynamics. Thus, any Newton-type method can not work in practice. We shall use the gentlest ascent dynamics (GAD) developed in [10] to formulate a new dynamical system associated with F. This new system can be viewed as an *effective* dynamics for a new multi-scale slow-fast system. To efficiently solve this new multi-scale system, which is called MsGAD for convenience, we apply the heterogeneous multiscale method or the seamless coupling method.

The original idea of GAD is based on the min-mode or eigenvector-following methodology (see, e.g., [4,32]). The numerical developments based on this methodology, such as the dimer method [16], the activation-relaxation techniques [26], etc. have been used for quite a few applications on potential energy surface. The applications of GAD include [30], [11], and [24]. The authors in [2] and [1] have some further discussions on the properties of GAD. As an acceleration technique for GAD, a new iterative formulation and its algorithmic implementation for gradient systems has been recently developed in [13] and [14], respectively. The optimization-based idea is also implemented in [34]. One of the advantages of GAD is that its form is a continuous-time dynamical system, so it is very easy to fit in our current framework for slow-fast dynamics (1.1). The application of GAD to sample the free energy surface has been discussed in [29] for temperature-accelerated molecular dynamics. A review of various saddle-point methods in computational material sciences can found in [35].

The saddle point problem is closely related to the transition-state calculation for the randomly perturbed system with additive noise: $\dot{x} = F(x) + \sqrt{\varepsilon}\dot{w}$. The optimal transition path in this additive noise case is determined by the quadratic form of the rate function in the traditional Freidlin–Wentzell large deviation theory [12] and much research has shown that the bottleneck on the optimal path is usually in the form of saddle points, especially for the gradient system. So, many applications actually look for saddle points first rather than compute the path directly. However, the large deviation rate function for the slow-fast system (1.1) has a much more complicated exponential form [3,12,31] and from a mathematical viewpoint, little is known about the relationship between the saddle point and the transition state for the slow-fast dynamics. In this

paper, we do not intend to resolve the complicated issue of finding optimal transition paths for the slow-fast dynamics. We leave the efficient numerical scheme for the path calculation as the next project. However, from the viewpoint of practical applications of studying activated processes on free energy surfaces [17, 22, 25], our pursuit of the saddle point on the effective dynamics F, which is the gradient of the free energy, is of significance if the transition state theory is used for rate calculation. For example, the calculation of the nucleation rate from the bulk liquid to a crystalline solid (e.g., [28]) requires the location of the saddle point on the free energy surface and the flux across the barrier at the saddle point.

The rest of the paper is organized as follows. In Section 1.1 and Section 1.2, we review the averaging theorem for the slow-fast system (1.1) and the gentlest ascent dynamics. In Section 2 we derive the gentlest ascent dynamics for the averaged system, develop multiscale methods for computations and discuss the algorithmic details, and discuss an adaptive implementation and make a remark on the connection to the central limit result. Section 3 is devoted to the gradient system where an extended potential energy function exists, in view of the practical importance of this class of models. Two examples are presented in Section 4 to demonstrate our method and the concluding remarks are given in Section 5.

1.1. Averaging principle of slow-fast stochastic dynamics. The averaging principle to derive the effective dynamics of the slow-fast system (1.1) is based on the ergodicity assumption of the fast process Y^ε. Let $\widetilde{Y}^x$ be the solution of the equation

$$\dot{\widetilde{Y}}^x = b(x,\widetilde{Y}^x) + \sigma(x,\widetilde{Y}^x)\eta(t), \tag{1.2}$$

for any fixed parameter x. This is named the virtually fast process. Assume that the virtually fast process $\widetilde{Y}^x$ is ergodic at every x and its unique invariant measure $\mu_x(\mathrm{d}y)$ has a density function $\rho(x,y)$:

$$\mu_x(\mathrm{d}y) = \rho(x,y)\mathrm{d}y = \frac{1}{Z(x)} e^{-U(x,y)}\mathrm{d}y, \tag{1.3}$$

where the normalization factor $Z(x)$ is

$$Z(x) \doteq \int_{\mathcal{Y}} e^{-U(x,y)}\mathrm{d}y. \tag{1.4}$$

REMARK 1.1. If $b(x,y) = -\nabla_y U_1(x,y)$ and $\sigma(x,y) \equiv \sigma I$ for some potential energy function $U_1(x,y)$ and a constant σ, i.e., the fast dynamics is a gradient system, then $U(x,y) = \frac{2}{\sigma^2} U_1(x,y)$ and the equilibrium distribution $\rho(x,y) = e^{-\frac{2}{\sigma^2}U_1(x,y)} / \int e^{-\frac{2}{\sigma^2}U_1(x,y)}\mathrm{d}y$. In other cases, we simply set $U(x) = -\log\rho(x,y)$ and $Z(x) \equiv 1$.

By the ergodicity assumption, for any integrable function u, the expectation with respect to μ_x can be estimated by the time average,

$$\int u(y)\mu_x(\mathrm{d}y) = \lim_{T\to\infty} \frac{1}{T}\int_0^T u(\widetilde{Y}(t))\mathrm{d}t.$$

The averaging principle (cf. [5, 12] and references therein) states that as $\varepsilon \downarrow 0$, the slow component of the system (1.1), $X^\varepsilon(t)$, has a limit $\bar{X}(t)$ satisfying the following ordinary differential equation,

$$\dot{\bar{X}} = F(\bar{X}), \quad \text{where } F(x) \doteq \int f(x,y)\mu_x(dy). \tag{1.5}$$

In most cases, the averaging function F above has no closed formula, and the solution $\bar{X}$ has to be approximated by numerical methods.

1.2. Gentlest ascent dynamics (GAD). For a dynamical system $\dot{x}(t) = \varphi(x(t))$ where the flow φ is C^2-smooth, the gentlest ascent dynamics locally converges to saddle point of φ by coupling the position variable and the direction variable. This dynamics, as a solution to the saddle point problem, can be viewed as a counterpart to the steepest descent dynamics for searching for local minima. In this note, we are only interested in the index-1 saddle point, i.e., the unstable manifold of the saddle point is one dimensional.

The GAD for the flow $\dot{x}(t) = \varphi(x(t))$ is the following extended system for (x, v, w),

$$\begin{cases} \dot{x}(t) = \varphi(x) - 2\dfrac{\langle \varphi(x), w\rangle}{\langle w, v\rangle} v, & \text{(1.6a)} \\ \gamma \dot{v}(t) = D\varphi(x)v - \alpha v, & \text{(1.6b)} \\ \gamma \dot{w}(t) = D\varphi(x)^{\mathsf{T}} w - \beta w, & \text{(1.6c)} \end{cases}$$

where $\langle \cdot, \cdot \rangle$ is the inner product, $D\varphi(x)$ is the Jacobi matrix $(D\varphi)_{ij} \doteq \frac{\partial \varphi_i}{\partial x_j}$. Two scalars α and β are Lagrangian multipliers to impose certain normalization conditions for v and w. For instance, if one uses the normalization condition $\langle v, v\rangle = \langle w, v\rangle = 1$, then $\alpha = \langle v, D\varphi(x)v\rangle$ and $\beta = 2\langle w, D\varphi(x)v\rangle - \alpha$. By setting $\langle v(0), v(0)\rangle = \langle v(0), w(0)\rangle = 1$, the GAD flow (1.6) for such choices of α and β then will preserve these two normalization equations. This technique to determine α and β will be applied later in many same situations; we shall not repeat the calculation of these Lagrangian multipliers.

The modified force in (1.6a) has the effect of inverting the direction of the original force $\varphi(x)$ on the direction v to sustain "ascent" flow around the index-1 saddle point. The dynamics (1.6b) and (1.6c), if x is frozen and as time goes to infinity, tend to the left-eigenvector w and the right-eigenvector v of the Jacobi $D\varphi(x)$ corresponding to the largest eigenvalue, respectively. The coupling of x and (v, w) is actually relaxed by a finite positive number γ in (1.6), which introduces a separation of time scale artificially. As $\gamma \downarrow 0$, equation (1.6) becomes a two-scale system where x is slow variable and (v, w) are fast variables. When the fast variables v and w have a single limit state as time goes to infinity, denoted by $v(x)$ and $w(x)$, respectively, $v(x)$ and $w(x)$ are the right and left eigenvector of $D\varphi(x)$. In this case, at the limit $\gamma \downarrow 0$, the effective dynamics of (1.6) is

$$\dot{x}(t) = \varphi(x) - 2\frac{\langle \varphi(x), w(x)\rangle}{\langle v(x), w(x)\rangle} v(x). \tag{1.7}$$

The rigorous proof of the local convergence of the GAD (1.6) to a nearby index-1 saddle point is presented in [10] for any finite γ and in the appendix of [14] for the limit of vanishing γ.

REMARK 1.2. For the PDE case, the transpose of the Jacobi matrix $D\varphi$ becomes the adjoint of the variational derivative operator. For instance, if the flow reads $u_t = u_{xx} + u_x + f(u)$ with periodic boundary condition in 1-D, then the Jacobi matrix is $\partial_{xx} + \partial_x + f'(u)$ and its adjoint is $\partial_{xx} - \partial_x + f'(u)$.

Since the Jacobi matrix $D\varphi$ is generally asymmetric, the right direction v and the left direction w are both required in (1.6) to obliquely project the force φ onto $\operatorname{span}\{v\}$ and $\operatorname{span}\{v\}^{\perp}$, except for the gradient system $\varphi(x) = -\nabla V(x)$, where the Hessian is symmetric. The GAD for a gradient system involves only v:

$$\begin{cases} \dot{x}(t) = -\nabla V(x) + 2\dfrac{\langle \nabla V(x), v\rangle}{\langle v, v\rangle} v, & \text{(1.8a)} \\ \gamma \dot{v}(t) = -\nabla^2 V(x) v + \langle v, \nabla^2 V(x) v\rangle v, & \text{(1.8b)} \end{cases}$$

where $\nabla^2 V$ is the Hessian matrix of the potential energy function $V(x)$. Equation (1.8b) is in fact the steepest descent flow (rescaled by γ) to minimize the Rayleigh quotient, $\min_{\|v\|=1} v^{\mathsf{T}} H v$, for the Hessian matrix $H \doteq \nabla^2 V(x)$. So, the steady state $v(x)$ is actually the eigenvector of the Hessian H at the lowest eigenvalue, the so called "min-mode".

In evolving the vector $v(t)$ in (1.8b), the full Hessian matrix may not be needed in practice. The computation of the Hessian-vector multiplication $\nabla^2 V(x) v$ is usually done by the finite difference method as in the dimer method [16, 33, 36], because this multiplication is exactly the directional derivative along the v direction:

$$\nabla^2 V(x) v = \frac{\mathrm{d}}{\mathrm{d}h} \nabla V(x + hv)|_{h=0} \approx h^{-1}(\nabla V(x + hv) - \nabla V(x)).$$

2. Multiscale gentlest ascent dynamics for slow-fast system

2.1. Formulation of GAD for slow-fast system. We intend to extend the GAD to the slow-fast dynamics (1.1) to calculate index-1 saddle points of the effective flow $\dot{\bar{X}} = F(\bar{X})$ defined in equation (1.5). The direct application of the GAD (1.6) to equation (1.5) gives

$$\begin{cases} \dot{x}(t) = F(x) - 2\dfrac{\langle F(x), w\rangle}{\langle v, w\rangle} v, & \text{(2.1a)} \\ \gamma \dot{v}(t) = DF(x) v - \alpha v, & \text{(2.1b)} \\ \gamma \dot{w}(t) = DF(x)^{\mathsf{T}} w - \beta w, & \text{(2.1c)} \end{cases}$$

where $F(x) = \int f(x,y)\rho(x,y)\mathrm{d}y$ by definition, and $DF(x)$ is the Jacobi matrix of $F(x)$. Note the density $\rho(x,y) = Z(x)^{-1} e^{-U(x,y)}$. We introduce

$$g(x,y) \doteq -\nabla_x U(x,y), \tag{2.2}$$

and

$$G(x) \doteq \int g(x,y)\rho(x,y)\mathrm{d}y. \tag{2.3}$$

By the definition of $Z(x)$ in equation (1.4), we simply see

$$\nabla_x \log Z(x) = Z^{-1}(x)\nabla_x Z(x) = \int g(x,y)\rho(x,y)\mathrm{d}y = G(x). \tag{2.4}$$

We calculate the Jacobi matrix as follows,

$$\begin{aligned} (DF)_{ij}(x) &= \frac{\partial F_i}{\partial x_j}(x) = \frac{\partial}{\partial x_j}\left(\int f_i(x,y) Z^{-1}(x) e^{-U(x,y)}\mathrm{d}y\right) \\ &= \int \left(\partial_{x_j} f_i(x,y) + f_i(x,y) g_j(x,y) - f_i(x,y)\partial_{x_j} Z(x) Z^{-1}(x)\right)\rho(x,y)\mathrm{d}y \\ &= \int \left(\partial_{x_j} f_i(x,y) + f_i(x,y) g_j(x,y) - f_i(x,y) G_j(x)\right)\rho(x,y)\mathrm{d}y \\ &= \overline{\partial_{x_j} f_i}(x) + \overline{f_i g_j}(x) - F_i(x) G_j(x). \end{aligned}$$

To ease presentation, the overlined symbol $\overline{\theta}(x)$ for a bivariate function $\theta(x,y)$ is used to define the expectation with respective to $\mu_x(\mathrm{d}y)$, that is,

$$\overline{\theta}(x) \doteq \int \theta(x,y)\mu_x(\mathrm{d}y).$$

So, $\overline{f}(x) = F(x)$ and $\overline{g}(x) = G(x)$ by this definition.

The Jacobi matrix of the effective dynamics is thus given by

$$DF(x) = \overline{D_x f}(x) + \overline{C}(x), \tag{2.5}$$

where $D_x f(x,y)$ is the Jacobi matrix of $f(x,y)$ with respect to the variable x and

$$C(x,y) \doteq f(x,y) \otimes g(x,y) - F(x) \otimes G(x). \tag{2.6}$$

The tensor product $u \otimes v$ for any two vectors u and v corresponds to the matrix $[u_i v_j]$.

The term $\overline{C}(x)$ in equation (2.5) comes from the x-dependency of the equilibrium distribution $\mu_x(\mathrm{d}y)$. $\overline{C}(x)$ is actually the covariance of f and g with respect to the distribution $\mu_x(\mathrm{d}y)$ because it is easy to verify that $\overline{C}(x) = \overline{(f-F)\otimes(g-G)}(x)$. This means that one can use an alternative form of equation (2.6) as follows:

$$C(x,y) = (f(x,y) - F(x)) \otimes (g(x,y) - G(x)), \tag{2.7}$$

if only the average quantity $\overline{C}$ is concerned. One can also verify that the choice of $C(x,y) = f \otimes g - F \otimes g$ also generates the same expectation $\overline{C}$ as (2.6).

2.2. The multiscale GAD. We shall address how to construct multiscale schemes for system (2.1). We have obtained the expression of the Jacobi DF in equation (2.5), which is the ensemble average of the matrix $D_x f + C$ with respect to μ_x. This important feature allows us to view system (2.1) as an averaged equation of a multiscale system involving the original fast variable y^ε:

$$\begin{cases} \dot{x}^\varepsilon(t) = f(x^\varepsilon, y^\varepsilon) - 2\dfrac{\langle f(x^\varepsilon, y^\varepsilon), w^\varepsilon\rangle}{\langle w^\varepsilon, v^\varepsilon\rangle} v^\varepsilon, & \text{(2.8a)} \\ \dot{y}^\varepsilon(t) = \dfrac{1}{\varepsilon} b(x^\varepsilon, y^\varepsilon) + \dfrac{\sigma(x^\varepsilon, y^\varepsilon)}{\sqrt{\varepsilon}} \eta(t), & \text{(2.8b)} \\ \gamma \dot{v}^\varepsilon(t) = (D_x f(x^\varepsilon, y^\varepsilon) + C(x^\varepsilon, y^\varepsilon))\, v^\varepsilon - \alpha^\varepsilon v^\varepsilon, & \text{(2.8c)} \\ \gamma \dot{w}^\varepsilon(t) = (D_x f(x^\varepsilon, y^\varepsilon) + C(x^\varepsilon, y^\varepsilon))^{\mathsf{T}}\, w^\varepsilon - \beta^\varepsilon w^\varepsilon, & \text{(2.8d)} \end{cases}$$

where the Lagrangian multipliers α^ε and β^ε can be defined as before to enforce certain normalization conditons. Here for any constant γ, as $\varepsilon \to 0$, the slow variables are $(x^\varepsilon, v^\varepsilon, w^\varepsilon)$ and the fast variable is y^ε. We name the multiscale system (2.8) the MsGAD (multiscale gentlest ascent dynamics). Notice that the covariance term C is rank-1, so the matrix-vector multiplication is actually calculated simply as the inner product, for instance, $C(x,y)v = \langle g(x,y), v\rangle f(x,y) - \langle G(x), v\rangle F(x)$.

In the expressions of (2.6) or (2.7), *i.e.*, $C(x,y) = f(x,y)\otimes g(x,y) - F(x)\otimes G(x)$, or $C(x,y) = (f(x,y) - F(x)) \otimes (g(x,y) - G(x))$, the averaged terms $F(x) = \bar{f}(x)$ and $G(x) = \bar{g}(x)$ already involve the invariant measure μ_x. Hence, equations (2.1) and (2.8) are not in the "standard" forms like the equations (1.5) and (1.1) defined in Section 1.1. This issue will not bring any essential challenges if the HMM is used since F and G can be estimated first from the sampling average (see below). But in the seamless

coupling method to be introduced later in Section 2.2.2, it is necessary to write the system (2.8) in the "standard" slow-fast multiscale form as equation (1.1). It turns out that a single fast variable y is not sufficient: one needs to introduce another process $\hat{Y}_t^x$ as an independent copy of the virtual fast process $\widetilde{Y}_t^x$ to accommodate the expectation $F=\bar{f}$ (or $G=\bar{g}$). Based on the fact that $\overline{C}=\overline{f\otimes g-F\otimes g}$, we introduce

$$C_2(x,y,z)\doteq f(x,y)\otimes g(x,y)-f(x,z)\otimes g(x,y), \tag{2.9}$$

then it is obvious that $\overline{C}(x)=\mathrm{E}_y\mathrm{E}_z[C_2(x,y,z)]$ where E_y and E_z are, respectively, the expectations with respect to the independent random variables y and z, which both follow the same law μ_x. So, one uses the new process $\hat{Y}_t^x$ to calculate the expectation $F=\bar{f}$ with respect to z, and the old $\widetilde{Y}_t^x$ to calculate the expectation $\overline{C}=\overline{f\otimes g-F\otimes g}$ with respect to y as before. This approach is based on the equivalent form of $\overline{C}$:

$$\overline{C}(x)=\int\left[f(x,y)\otimes g(x,y)-\left(\int f(x,z)\hat{\mu}_x(\mathrm{d}z)\right)\otimes g(x,y)\right]\mu_x(\mathrm{d}y),$$

where $\hat{\mu}_x(=\mu_x)$ is the equilibrium distribution of the iid copy $\hat{Y}^x$.

However, for the alternative form in equation (2.7): $C(x,y)=(f(x,y)-F(x))\otimes(g(x,y)-G(x))$, $\overline{C}=\overline{(f-\bar{f})\otimes(g-\bar{g})}$ contains three expectations. One can run three fast processes y,z,w simultaneously as iid copies, and then letting $C_3(x,y,z,w)\doteq(f(x,y)-f(x,z))\otimes(g(x,y)-g(x,w))$, one has $\mathrm{E}_y\mathrm{E}_z\mathrm{E}_w C_3=\overline{C}$. It is interesting to find that it is possible to run only two processes y and z in this case by choosing $\hat{C}_2(x,y,z)\doteq(f(x,y)-f(x,z))\otimes(g(x,y)-g(x,z))$. Then the calculation shows $\mathrm{E}_y\mathrm{E}_z\left[\hat{C}_2(x,y,z)\right]=2(\overline{f\otimes g}-F\otimes G)=2\overline{C}$. Thus, one should use $\frac{1}{2}\hat{C}_2$ rather than $\hat{C}_2$ to estimate the covariance matrix $\overline{C}$. In summary, there may be different unbiased estimators of $\overline{C}$ if multiple streams of the fast processes are used. The variances of these estimator could be further analyzed analytically or compared numerically. We do not further pursue this issue and simply use the scheme based on the expression (2.9) for numerical examples in this paper.

2.2.1. HMM. Now we discuss the framework of the HMM (*heterogeneous multiscale method*, [6, 7]) for our averaged GAD system (2.1) and the MsGAD (2.8). There are two parameters ε and γ in the multiscale GAD. We let ε tend to zero in (2.8) to obtain equation (2.1) by the averaging principle. We can also further select a small γ in (2.1) to obtain an equation like (1.7).

The procedures of the HMM for the MsGAD are as follows. Select a macroscopic time step size Δt for evolving x and $\Delta\tau$ for evolving v and w (usually $\Delta\tau=\Delta t$); select a microscopic time step size δt for evolving y. The HMM scheme with forward Euler consists of the following steps.

(1) Use the macro-solver

$$x_{n+1}=x_n+(F_n-2v_n\langle F_n,v_n\rangle/\langle w_n,v_n\rangle)\Delta t,$$

where x_n is the approximation value of $\bar{X}(n\Delta t)$ and F_n, v_n, w_n are estimated below.

(2) Apply the micro-solver with time step size δt to M micro-steps for the fast dynamics

$$y_{n,m+1}=y_{n,m}+\frac{\delta t}{\varepsilon}b(x_n,y_{n,m})+\frac{\sigma(x_n,y_{n,m})}{\sqrt{\varepsilon}}\sqrt{\delta t}\eta_{n,m},$$

$m=0,1,2\ldots,M-1$. Here $\{\eta_{n,m}\}$ are iid $\mathcal{N}(0,1)$ random variables. The initial is chosen as the "warm start": $y_{n,0}\doteq y_{n-1,M}$. Note that only the ratio $\delta t/\varepsilon$ is needed here; equivalently one can view this ratio as the true step size for the virtual fast process $\tilde{Y}^x$ defined in equation (1.2).

(3) Estimate F_n, G_n and the Jacobi matrix $(DF)_n$:

$$F_n=\frac{1}{M}\sum_{m=1}^{M} f(x_n,y_{n,m}),\quad G_n=\frac{1}{M}\sum_{m=1}^{M} g(x_n,y_{n,m}),$$
$$(DF)_n=\frac{1}{M}\sum_{m=1}^{M}\Big(D_x f(x_n,y_{n,m})+f(x_n,y_{n,m})\otimes g(x_n,y_{n,m})\Big)-F_n\otimes G_n.$$

(4) Solve the right and left direction v_n and w_n for K steps by using the time step size $\Delta\tau$ and the initial $v_{n,0}=v_{n-1}$, $w_{n,0}=w_{n-1}$:

$$\hat{v}_{n,k+1}=v_{n,k}+\Delta\tau(DF)_n v_{n,k},\quad \hat{w}_{n,k+1}=w_{n,k}+\Delta\tau(DF)_n^{\mathsf{T}} w_{n,k},$$
$$v_{n,k+1}=\frac{\hat{v}_{n,k+1}}{|\hat{v}_{n,k+1}|},\quad w_{n,k+1}=\frac{\hat{w}_{n,k+1}}{\langle v_{n,k+1},\hat{w}_{n,k+1}\rangle},$$

$k=0,1,2\ldots,K-1$. Then $v_n=v_{n,K}$ and $w_n=w_{n,K}$.

The microscopic time step size has to be much smaller than ε; $\delta t/\varepsilon$ is the effective step size in solving the virtually fast process $\widetilde{Y}$. The time length, $M\times\delta t$, should be sufficient long for the fast process Y^ε to relax toward the equilibrium distribution. The number of steps M also should be large enough to reduce the statistical error in estimating the averaged quantities F, G and DF. The number of steps K is to take care of a possible small constant γ. $K=1$ actually works in principle for many cases. A larger K can give a better accuracy in computing the eigenvector v, though it also has more computational burden. There is no requirement for $\Delta\tau$ as long as the ODE solver is stable; the time step size $\Delta\tau$ can be simply set as Δt.

REMARK 2.1. On the choice of the macro-solver for x and v,w, an explicit scheme with larger stability region is preferred, such as the stablized Runge-Kutta methods. On the micro-solver for the virtually fast process, a numerical scheme for SDE with higher order weak convergence rate for the *long* time integration is preferred to capture the equilibrium distribution better. For instance, when σ is a constant, one can use the stochastic Heun method ([21]) which requires two force evaluations at each time step or a non-Markovian scheme proposed recently in [23]:

$$y_{n,m+1}=y_{n,m}+\frac{\delta t}{\varepsilon}b(x_n,y_{n,m})+\frac{\sigma\sqrt{\delta t}}{2\sqrt{\varepsilon}}(\eta_{n,m}+\eta_{n,m+1}),\quad m=0,1,\ldots,$$

where $\{\eta_{n,m}\}$ are iid $\mathcal{N}(0,1)$ random variables.

2.2.2. Seamless coupling method (SCM). The seamless coupling strategy proposed in [8] does not need the back and forth communication of the macro- and micro-states of the system. It was thought as a boosting algorithm: by increasing the small parameter ε to a larger number ε' in the slow-fast system, the seamless scheme simultaneously solves this boosted system with the time step size $\widetilde{\Delta}t$ which is smaller than the macro-time step size Δt in the HMM. It is potentially more efficient than the

HMM if the micro-model is difficult. Following this idea, we increase ε in system (2.8) to a larger number, say $\varepsilon' = \varepsilon\lambda$ for a constant $\lambda > 1$. As stated in previous section, to use equation (2.7), we need to introduce an independent copy z^ε for the fast process to handle the double expectation in the covariance matrix $\overline{C}$. In summary, the seamless coupling method solves the following system:

$$\begin{cases} \dot{x}^\varepsilon(t) = f(x^\varepsilon, y^\varepsilon) - 2\dfrac{\langle f(x^\varepsilon, y^\varepsilon), w^\varepsilon\rangle}{\langle w^\varepsilon, v^\varepsilon\rangle} v^\varepsilon, & \text{(2.10a)} \\ \dot{y}^\varepsilon(t) = \dfrac{1}{\varepsilon\lambda} b(x^\varepsilon, y^\varepsilon) + \dfrac{1}{\sqrt{\varepsilon\lambda}} \eta_1(t), & \text{(2.10b)} \\ \dot{z}^\varepsilon(t) = \dfrac{1}{\varepsilon\lambda} b(x^\varepsilon, z^\varepsilon) + \dfrac{1}{\sqrt{\varepsilon\lambda}} \eta_2(t), & \text{(2.10c)} \\ \gamma\dot{v}^\varepsilon(t) = (D_x f(x^\varepsilon, y^\varepsilon)) v^\varepsilon + f(x^\varepsilon, y^\varepsilon)\langle g(x^\varepsilon, y^\varepsilon), v^\varepsilon\rangle \\ \qquad\qquad - f(x^\varepsilon, z^\varepsilon)\langle g(x^\varepsilon, y^\varepsilon), v^\varepsilon\rangle - \alpha^\varepsilon v^\varepsilon, & \text{(2.10d)} \\ \gamma\dot{w}^\varepsilon(t) = (D_x f(x^\varepsilon, y^\varepsilon))^{\mathsf{T}} w^\varepsilon + g(x^\varepsilon, y^\varepsilon)\langle f(x^\varepsilon, y^\varepsilon), w^\varepsilon\rangle \\ \qquad\qquad - g(x^\varepsilon, z^\varepsilon)\langle f(x^\varepsilon, y^\varepsilon), w^\varepsilon\rangle - \beta^\varepsilon v^\varepsilon, & \text{(2.10e)} \end{cases}$$

where η_1 and η_2 are two i.i.d. copies of η. The joint fast processes y^ε and z^ε correspond to the equilibrium distribution $\mu_x(\mathrm{d}y) \times \mu_x(\mathrm{d}z)$. It is clear that as $\varepsilon \to 0$, the effective dynamics of system (2.10) is just equation (2.1). System (2.10) is solved by any standard ODE/SDE solver, such as the Euler scheme, with a common time step size $\widetilde{\Delta}t$ for all five components. [5] suggests a time step size $\widetilde{\Delta}t = \Delta t/M$, where M is the micro-step number previously used in the HMM. After a sufficient long time for relaxation, the solution $x^\varepsilon(t)$ will go close to the index-1 saddle point of the flow F, then one may average $x^\varepsilon(t)$ within a time interval to improve the accuracy.

To end this subsection, we discuss the difference between the HMM and the seamless coupling method (SCM). One can refer to the book [5] and references therein for general discussions. For the application of these two methods to the search of stationary points here, we emphasize some key and special differences below. Firstly, the HMM and SCM solve two related but different systems: equation (2.1) and equation (2.10), respectively. The HMM solves the effective dynamics (2.1) by simulating its multiscale version (2.8) at the vanishing ε limit, whereas the SCM solves the pre-asymptotic system (2.10) with an effective parameter $\varepsilon' = \varepsilon\lambda > 0$. So, as ε' tends to 0, the solution of the SC tends to the solution of the HMM. In other words, the result of the SCM is an approximate solution to the true saddle point of the effective force F and the error is proportional to $\sqrt{\varepsilon'}$ (see Section 2.6). In the long time run, this also means that, with a fixed $\varepsilon' > 0$, the trajectory of the SCM will eventually hop between multiple saddle points (if they exist) as the time goes to infinity. Secondly, in the HMM, the two expectations in the covariance matrix $\overline{C}$ are computed via the sample average from a single stream of random samples from μ_x, whereas in the SCM, this double expectation is realized by two streams of random samples for the virtual fast process. Therefore, the understanding of the seamless coupling here is more than the classic boosting idea since the classic seamless coupling for slow-fast two-scale system does not have to run two independent streams of the virtual fast process. Thirdly, the HMM needs calculation of the sample average, which has high computational costs but controls variance very well, whereas the SCM does not take sample averages at each iteration, which is fast but is more noisy.

2.3. Adaptive strategy. The quantity of our interest is the saddle point, not the whole trajectory. This focus on the final destination, regardless of the accuracy of

the trajectories at the early stage, facilitates the application of adaptivity ideas to fine tune some parameters in our algorithms. In order to reduce the variance of the solution and improve the practical efficiency, a very simple idea is to apply the SCM first and then switch to the HMM after an appropriate time period. We generally discuss a few more useful methodologies below.

(1) Average the output: $\tilde{x}(t) \doteq \frac{1}{t-t_0}\int_{t_0}^{t} x(s)\mathrm{d}s$ after a period of "burn-in" time t_0. This is the widely used technique in Monte Carlo simulation and the stochastic gradient dynamics to reduce the variance of the original noisy output. This principle works for both HMM (particularly for a small M) and the SCM without any extra computational burden, but for the SCM, the effect of improvement is more significant.

(2) Choose the sample size M adaptively in the HMM: Use a small M at the first stage for a quick search of the area of the solution, then use a large M to suppress the variance when the fluctuation starts to dominate the errors. One might simply link the increasing rate of M to the time, say $M \propto t^r$ for some $r > 0$.

(3) Reduce the small parameter ε' adaptively in the SCM: The errors of the seamless coupling method depend on the small parameter of $\varepsilon' := \varepsilon\lambda$ and the step size $\widetilde{\Delta}t$. A smaller ε' means a smaller deviation of x^ε from the true solution (associated with F). So, one can decrease the value of ε' as time runs. Since the effective step size for the virtual fast processes is $\widetilde{\Delta}t/\varepsilon'$, then one also has to decrease the step size $\widetilde{\Delta}t$ so that $\widetilde{\Delta}t/\varepsilon'$ either is fixed or decreases too. For example, the typical decay of $\widetilde{\Delta}t$ at time t in stochastic optimization corresponds to the scaling $\widetilde{\Delta}t \propto t^{-1}$; equivalently, the step size at k-th iteration is $\widetilde{\Delta}t_k = \widetilde{\Delta}t_1/\sqrt{k}$. Then the decay of ε' may be set as $\varepsilon'(t) \propto t^{-1+c}$ for a small constant $0 \le c < 1$, so that $\widetilde{\Delta}t/\varepsilon' \propto t^{-c}$.

We shall demonstrate the effectiveness of the above three ideas in the numerical section.

2.4. Numerical calculation of the right direction v. In the algorithms presented above, we formally use the full matrix of the Jacobi DF in the dynamics for the right and left directions, $v(t)$ and $w(t)$, respectively. If the dynamics is in the form of PDEs, then as pointed out in Remark 1.2, the Jacobi DF and its transpose are just the variational derivative and its adjoint, both of which are not difficult to derive in most cases. However, in some real applications such as atomistic models with long range interaction, it is not efficient or even feasible to store this Jacobi matrix element by element. In what follows, we discuss this challenge and the numerical remedies according to the specific features of slow-fast system. We start from the right direction v. It is noted that in many cases, the dimension of $\mathcal{X}$ is much lower than the dimension of $\mathcal{Y}$. This means that $D_x f(x,y)$ is a small scale matrix and its all entry-wise values can be calculated and stored with reasonable cost.

If the fast dynamics is of gradient type, i.e., $b(x,y) = -\nabla_y U_1(x,y)$ and $\sigma(x,y) \equiv \sigma I$, then $U = \frac{2}{\sigma^2}U_1$ and $g(x,y) = -\frac{2}{\sigma^2}\nabla_x U_1(x,y)$. The covariance matrix is $C(x,y) = -\frac{2}{\sigma^2} f(x,y) \otimes \nabla_x U_1(x,y)$. Assume the analytic form of $\nabla_x U_1(x,y)$ is available in practice, then the rank-1 matrix C is easy to calculate: $C(x,y)v = -\frac{2}{\sigma^2} f(x,y)\langle \nabla_x U_1(x,y), v\rangle$.

However, for non-gradient dynamics, the equilibrium distribution $\rho(x,y)$ has no closed formula available. For the dynamics of v in (2.8c), the finite difference scheme

$$\begin{aligned}(DF(x))v &= \lim_{h\to 0}(F(x+hv) - F(x))/h \\ &= \lim_{h\to 0} h^{-1}\int f(x+hv,y)\rho(x+hv,y) - f(x,y)\rho(x,y)\mathrm{d}y \end{aligned} \tag{2.11}$$

can be applied to compute the matrix-vector multiplication $(DF(x))v$. If one knows the closed form of the Jacobi matrix $D_x f$, $(D_x f(x,y))v$ thus can be directly computed by the matrix-vector multiplication. The remaining term $(DF - D_x f)v = \overline{C}(x)v$ can be computed by the finite difference scheme. By Remark 1.1 and equations (2.2) and (2.3), we have $g = \nabla_x \log\rho$ and $C(x,y) = f(x,y) \otimes (\nabla_x \log\rho(x,y)) = \rho^{-1} f(x,y) \otimes (\nabla_x \rho(x,y))$. Then,

$$\begin{aligned} C(x,y)v &= \rho^{-1} f(x,y) \langle v, \nabla_x \rho(x,y) \rangle \\ &= \rho^{-1} f(x,y) \lim_{h\to 0} (\rho(x+hv,y) - \rho(x,y))/h. \end{aligned}$$

It follows that

$$\begin{aligned} \overline{C}(x)v &= \int C(x,y)v\rho(x,y)\mathrm{d}y \\ &= \lim_{h\to 0} h^{-1} \int f(x,y)(\rho(x+hv,y) - \rho(x,y))\mathrm{d}y. \end{aligned} \tag{2.12}$$

So, equation (2.11) can be rewritten as

$$(DF(x))v = \int D_x f(x,y)v\rho(x,y)\mathrm{d}y + \lim_{h\to 0} \frac{1}{h} \int f(x,y)(\rho(x+hv,y) - \rho(x,y))\mathrm{d}y. \tag{2.13}$$

In practice, a finite step size h is used in equations (2.11) or (2.13). Then, in the MsGAD, besides the process $\widetilde{Y}_t^x$ slaved at x, a second independent virtually fast process $\widetilde{Y}_t^{x+hv}$ is also required to obtain the distribution $\rho(x+hv,\cdot)$. The second integration in equation (2.13) is computed from the time averages from $\rho(x+hv,\cdot)$ and $\rho(x,\cdot)$.

2.5. Numerical calculation of the left direction w. In many cases, we have to evolve the dynamics for the w direction in equation (2.8d), except that the original slow-fast system (1.1) is of the gradient type jointly (see Section 3 below). It is worthwhile to note that even if both the slow and fast components are of the (parametrized) gradient types separately, but the coupled system (1.1) is not jointly driven by a single potential energy function, then the resulted averaging equation (1.5) might still not be a gradient system and the Jacobi matrix $DF(x)$ is not symmetric. Refer to Section 4.1 for such an example.

If the fast dynamics is of the gradient type, as we mentioned above for the $C(x,y)v$ calculation, then it is also quite simple to calculate $C(x,y)^{\mathsf{T}}w$ by using $\nabla_x U_1(x,y)$.

Yet, when the fast dynamics is of the non-gradient type, the matrix-vector multiplication $(DF(x))^{\mathsf{T}}w$, in particular $C(x,y)^{\mathsf{T}}w$, will impose a very severe computational challenge in this rather general situation, because it is not possible to apply the trick of directional derivative. We do not have any perfect solution for this non-gradient case and the only method to evaluate the derivatives is numerical evaluation. For $DF = \overline{D_x f + C}$, we assume that the full matrix $D_x f(x,y)$ has an expression to compute and its transpose $(D_x f(x,y))^{\mathsf{T}}$ is obtained by a numerical transpose operation (or simply the adjoint operator in the infinite dimensional setting of the PDE case). Then the term

$$\overline{C^{\mathsf{T}}w} = \int \theta(x,y)\nabla_x \rho(x,y)\mathrm{d}y, \quad \text{where} \quad \theta(x,y) \doteq \langle f(x,y), w \rangle,$$

involves the sensitivity analysis for the equilibrium distribution $\nabla_x \rho(x,y)$. Although this sensitivity could be analytically derived in some case, it can be approximated by the numerical derivative of ρ

$$\partial_{x_j} \rho(x,y) = \lim_{h\to 0} (\rho(x+he_j,y) - \rho(x,y))/h, \qquad j = 1,\ldots,m, \tag{2.14}$$

where e_j is the unit vector along x_j axis. Then the j-th component is

$$(\overline{C^{\mathsf{T}}w})_j = \lim_{h\to 0}\int \theta(x,y)(\rho(x+he_j,y)-\rho(x,y))/h\,\mathrm{d}y.$$

By choosing a finite number $h \ll 1$, this brute-force calculation will have to add m independent fast processes $\widetilde{Y}^{x+he_j}$, $j=1,\ldots,m$, to account for $\rho(x+he_j,\cdot)$. The scheme (2.14) is computationally feasible only when the dimension of $\mathcal{X}$ is low. In the next subsection, we show that the x-derivative of $\nabla_x\rho$ can be transformed to the y-derivative of some function by the perturbation analysis for the equilibrium density $\rho(x,y)$. However, the numerical challenge in this new form might still exist.

2.6. Connection with the central limit theorem. In the last part of this section, we shall give a remark on the connection with the normal deviation from the averaged system and how this normal derivation theory can be useful in theory to estimate the fluctuation in the SCM. In the MsGAD, the Jacobi $DF(x)$ is important: it determines the linearization of the averaged dynamics F. This linearization also plays an important role in approximating $X^\varepsilon - \bar{X}$, the difference of the solutions to the multiscale system (1.1) and the averaged system (1.5). Define the normalized difference

$$\xi_t^\varepsilon \doteq \frac{1}{\sqrt{\varepsilon}}(X_t^\varepsilon - \bar{X}_t). \tag{2.15}$$

By Theorem 3.1, §7.3 in [12], under the assumption of strong mixing, as $\varepsilon\to 0$, the normalized difference converges weakly to the solution of the following SDE

$$\dot{\xi}(t) = (DF(\bar{X}_t))\xi(t) + \beta(\bar{X}_t)\eta(t), \tag{2.16}$$

where $DF(\bar{X}_t)$ is the Jacobi matrix DF at the averaged solution $\bar{X}$ and the diffusion term β is the $m\times m$ matrix such that

$$\beta(x)\beta(x)^{\mathsf{T}} = A(x) \doteq \lim_{T\to\infty}\frac{1}{T}\int_0^T\int_0^T (f(x,\widetilde{Y}_s^x)-F(x))\otimes(f(x,\widetilde{Y}_t^x)-F(x))\mathrm{d}t\mathrm{d}s,$$

which can be estimated by running a long trajectory of the fast process $\widetilde{Y}^x$.

If the formal asymptotic expansion is applied to derive the equation (2.16), the drift term in (2.16) has a different form, denoted as B,

$$B(x) = \overline{D_x f}(x) + \int\left(\int_0^\infty \nabla_y \mathrm{E}^y f(x,\widetilde{Y}_\tau^x)\mathrm{d}\tau\right) D_x b(x,y)\ \mu_x(\mathrm{d}y), \tag{2.17}$$

where E^y is the expectation for the distribution of $\widetilde{Y}^x$ with initial $\widetilde{Y}_0^x = y$. Refer to the appendix in [3] for this formula. Comparing (2.17) with (2.5), we find that $\overline{C}(x)$ should be equal to the double integral term on the right-hand side of (2.17). The proof of this fact is attached in the appendix of this paper. Numerically, the formula (2.17) is not friendly due to the differential operator ∇_y for the expectation term E^y.

In summary, the normalized difference ξ in (2.16) in the central limit theorem shares the same drift flow as our dynamics for the right direction v in (2.1). Hence, the numerical methods we developed here for the MsGAD may be useful to the calculation of ξ. The more interesting observation is that the above result can aid in understanding the fluctuations in the seamless coupling method, at least in theory.

REMARK 2.2. The above convergence theorem for $\xi_t^\varepsilon \to \xi(t)$ as t tends to infinity can also be applied to the effective GAD system (2.1) and its multiscale system (2.10). That is to define $\bar{X}$ and X^ε in (2.15) as the trajectories of the GAD system (2.1) (computed from the HMM) and that of (2.10) (computed from the SCM), respectively. Then the corresponding normalized deviation satisfies the equation in the form of equation (2.16) (after redefining the F and β terms accordingly for the GAD rather than for the original dynamics), which characterizes the fluctuations of the SCM. Furthermore, one can linearize equation (2.16) around the saddle point x^* to obtain a rough estimate of the fluctuation around the true solution x^*.

3. The MsGAD for gradient system

As mentioned earlier, one important example of the slow-fast system (1.1) in practice is the extended Lagrangian method for the coarse-grained molecular dynamics. In this example, an energy potential $U(x,y)$ exists in the extended space $\mathcal{X}\times\mathcal{Y}$ to drive the slow-fast system. For this gradient system, we shall see that the averaged system $\dot{\bar{X}} = F(\bar{X})$ is also a gradient system, i.e., a (free energy) function $W(x)$ exists in $\mathcal{X}$ such that $F(x) = \nabla W(x)$. The general discussions for the GAD above can be greatly simplified much for this gradient case. Now we assume the multiscale system (1.1) has the special form

$$\begin{cases} \dot{X}^\varepsilon = -\nabla_x U(X^\varepsilon, Y^\varepsilon), & (3.1a)\\ \dot{Y}^\varepsilon = -\dfrac{1}{\varepsilon}\nabla_y U(X^\varepsilon, Y^\varepsilon) + \dfrac{1}{\sqrt{\varepsilon}}\sigma\eta(t), & (3.1b) \end{cases}$$

where σ is assumed a scalar constant and η is a standard white noise. $U(x,y)$ is a potential energy function for $(X^\varepsilon, Y^\varepsilon)$. The equilibrium measure of the virtually fast process is

$$\mu_x(\mathrm{d}y) = \rho(x,y)\mathrm{d}y = \frac{1}{Z(x)} e^{-\frac{2}{\sigma^2}U(x,y)}\mathrm{d}y, \quad Z(x) \doteq \int e^{-\frac{2}{\sigma^2}U(x,y)}\mathrm{d}y. \tag{3.2}$$

As $\varepsilon \downarrow 0$, the averaged equation for $\bar{X}$ is

$$\dot{\bar{X}} = F(\bar{X}), \tag{3.3}$$

where

$$F(x) = -\int \nabla_x U(x,y)\rho(x,y)\mathrm{d}y.$$

By definition of $Z(x)$, we can get

$$\nabla_x \log Z(x) = \frac{2}{\sigma^2}\int -\nabla_x U(x,y)\rho(x,y)\mathrm{d}y = \frac{2}{\sigma^2}F(x).$$

Thus, the averaged equation (3.3) can be rewritten as a gradient system

$$\dot{\bar{X}} = -\nabla_x W(\bar{X}), \tag{3.4}$$

where the effective potential is

$$W(x) = -\frac{\sigma^2}{2}\log Z(x) = -\frac{\sigma^2}{2}\log\left(\int e^{-2U(x,y)/\sigma^2}\mathrm{d}y\right). \tag{3.5}$$

By the calculation of (2.5), the Hessian matrix of $W(x)$ is

$$\nabla_x^2 W(x) = -DF(x) = \overline{\nabla_x^2 U}(x) - \frac{2}{\sigma^2}\overline{\nabla_x U \otimes \nabla_x U}(x) + \frac{2}{\sigma^2}\overline{\nabla_x U}(x) \otimes \overline{\nabla_x U}(x). \tag{3.6}$$

The right hand side contains the Fisher information matrix of the invariant measure $\mu_x (= \rho(x,y)\mathrm{d}y)$:

$$-\mathrm{E}_{\mu_x}\left[\nabla_x^2 \log \rho\right] = \frac{4}{\sigma^2}\left(\overline{\nabla_x U \otimes \nabla_x U}(x) - \overline{\nabla_x U}(x) \otimes \overline{\nabla_x U}(x)\right).$$

3.1. Example. The following extended potential is widely used in free energy sampling and the coarse grained molecular dynamics simulation [9]:

$$U(x,y) = V(y) + \frac{1}{2}\kappa |x - q(y)|^2, \tag{3.7}$$

where $q = (q_1(y), \ldots, q_n(y))$ is a given function mapping fast variables in $\mathcal{Y}$ to the space $\mathcal{X}$, which defines coarse-grained variables. $\kappa > 0$ is a parameter coupling the potential of the microscopic system and the coarse-grained variables. Ideally, κ should be infinitely large, but it is a large constant in practice. The slow-fast dynamic system associated with equation (3.7) is

$$\begin{cases} \dot{X}^\varepsilon = -\kappa(X^\varepsilon - q(Y^\varepsilon)), & \text{(3.8a)} \\ \dot{Y}^\varepsilon = -\dfrac{1}{\varepsilon}\Big(\nabla V(Y^\varepsilon) - \kappa (Dq(Y^\varepsilon))^{\mathsf{T}}(X^\varepsilon - q(Y^\varepsilon))\Big) + \dfrac{1}{\sqrt{\varepsilon}}\sigma\eta(t). & \text{(3.8b)} \end{cases}$$

where $Dq(y)$ is the Jacobi matrix $(\partial_{y_j} q_i)$. In order to sample the space $\mathcal{X}$ with the correct marginal equilibrium distribution in the extended Lagrangian method, the slow dynamics (3.8a) actually should also be independently driven by an Brownian motion. Since here we only study the saddle point rather than the distribution of X^ε, we are only concerned with the deterministic steepest descent drift flow in (3.8a).

For this example, we have that $\nabla_x U(x,y) = \kappa(x - q(y))$ and $\nabla_x^2 U(x,y) \equiv \kappa I$, where I is the identity matrix. It follows that $\overline{\nabla_x U} = \kappa(\bar{x} - Q(x))$ and $\overline{\nabla_x^2 U} \equiv \kappa I$, where

$$\bar{x} = \int x\rho(x,y)\mathrm{d}y, \qquad Q(x) \doteq \bar{q}(x) = \int q(y)\rho(x,y)\mathrm{d}y,$$

and

$$\rho(x,y) = Z(x)^{-1} e^{-\frac{2}{\sigma^2}V(y)} e^{-\frac{\kappa}{\sigma^2}(x - q(y))^2}.$$

Then, the effective potential W is

$$W(x) = -\frac{\sigma^2}{2}\log Z(x) = -\frac{\sigma^2}{2}\log\left(\int e^{-\frac{2}{\sigma^2}V(y)} e^{-\frac{\kappa}{\sigma^2}(x - q(y))^2}\mathrm{d}y\right).$$

and the effective force for the slow variable is

$$F(x) = -\nabla_x W(x) = -\kappa(\bar{x} - Q(x)).$$

The Hessian matrix of W for this example, by the result in equation (3.6), is

$$\nabla_x^2 W(x) = \kappa I - \frac{2\kappa^2}{\sigma^2}\left(\overline{(x-q)\otimes(x-q)}(x) - (\bar{x} - Q(x)) \otimes (\bar{x} - Q(x))\right), \tag{3.9}$$

and hence

$$(\nabla_x^2 W)v = \kappa v - \frac{2\kappa^2}{\sigma^2}\left(\overline{\langle x-q, v\rangle (x-q)} - \langle \bar{x}-Q, v\rangle (\bar{x}-Q)\right).$$

The dynamics for direction v in the GAD, $\gamma\dot{v} = -(\nabla^2 W)v + \alpha v$, is

$$\gamma\dot{v} = \frac{2\kappa^2}{\sigma^2}\left(\overline{\langle x-q, v\rangle (x-q)} - \langle \bar{x}-Q(x), v\rangle (\bar{x}-Q(x))\right) + \alpha' v$$

by absorbing the term κ into the new Lagrangian multiplier α'.

In summary, for this example, the GAD for the averaged system is

$$\begin{cases} \dot{x}(t) = -\kappa(\bar{x}-Q(x)) + 2\dfrac{\langle \kappa(\bar{x}-Q(x)), v\rangle}{\langle v, v\rangle} v, & \text{(3.10a)} \\ \gamma\dot{v}(t) = \dfrac{2\kappa^2}{\sigma^2}\left(\overline{\langle x-q, v\rangle (x-q)} - \langle \bar{x}-Q(x), v\rangle (\bar{x}-Q(x))\right) + \alpha' v. & \text{(3.10b)} \end{cases}$$

If $\kappa \to \infty$, then $\rho(x,y) \to \widetilde{Z}(x)^{-1}\delta(x-q(y))e^{-\frac{2}{\sigma^2}V(y)}$, where $\widetilde{Z}(x) \doteq \int \delta(x-q(y))e^{-\frac{2}{\sigma^2}V(y)}\mathrm{d}y$. And it follows that $W(x) \to \mathcal{F}(x) \doteq -\frac{\sigma^2}{2}\log \widetilde{Z}(x)$ (up to a constant), where $\mathcal{F}$ is the free energy surface of the coarse-grained variable x. For further details on the GAD for this type of question and the applications in coarse-grained molecular dynamics simulation, refer to the publication [29].

3.2. The dimer method for gradient system. The GAD for the gradient system (3.4) requires the calculation of the product of the Hessian $\nabla_x^2 W$ and the vector v, which can be numerically approximated by the finite difference scheme when the Hessian itself is difficult to obtain. This idea has been widely used in the dimer method [16,33]. The one-side finite difference scheme is

$$(\nabla_x^2 W)\, v \approx \frac{\nabla_x W(x+hv) - \nabla_x W(x)}{h} = -\frac{F(x+hv) - F(x)}{h}.$$

To evaluate $F(x+hv)$, which is $\int f(x+hv, y)\rho(x+hv, y)\mathrm{d}y$, one simple approach is to run a second independent fast process $\widetilde{Y}_t^{x+hv}$ to obtain the density $\rho(x+hv, \cdot)$ parametrized at $x+hv$. When the central finite difference scheme is used, the third process for $x-hv$ is required. This approach leads to an undesired extra burden of simulating multiple fast processes. In contrast, by using the formula derived in (3.6), only one trajectory for the virtually fast process is required.

4. Numerical examples

To illustrate how the MsGAD works, we present two numerical examples below. The first is a system consisting of a two dimensional ordinary differential equation and a two dimensional stochastic ordinary differential equation. It is not a gradient system. We apply the HMM (Section 2.2.1) in the MsGAD. The second is a system of an Allen–Cahn partial differential equation and a stochastic Allen–Cahn partial differential equation. This second system has an extended potential functional. We apply and compare both the HMM and the seamless coupling scheme (Section 2.2.2) to this second example.

4.1. A two-dimensional example. We consider the following system on $\mathcal{X} \times \mathcal{Y} = \mathbb{R}^2 \times \mathbb{R}^2$

$$\begin{cases} \dot{X}_i = -\displaystyle\sum_j D_{ij} X_j + Y_i^2, & \text{(4.1a)} \\ \dot{Y}_i = -\dfrac{1}{\varepsilon}\dfrac{Y_i}{\Gamma_i(X)} + \dfrac{1}{\sqrt{\varepsilon}}\sigma\eta(t). & \text{(4.1b)} \end{cases}$$

The vector field $\Gamma(x)=(\Gamma_1(x),\Gamma_2(x)):\mathcal{X}\to\mathcal{Y}$ is given. $D=(D_{ij})$ is a 2×2 symmetric matrix. σ is a constant.

The processes $\{Y_i(t)\}$ are independent Ornstein–Uhlenbeck processes parametrized by $X=x$. The equilibrium distribution of $Y=(Y_1,Y_2)$ is the product measure of $\mathcal{N}(0,\sigma^2\Gamma_i(x)/2)$. The calculation shows that the limit equation has a closed form

$$\dot{\bar{X}}_i=-\sum_j D_{ij}\bar{X}_j+\frac{\sigma^2}{2}\Gamma_i(\bar{X}). \tag{4.2}$$

Note that if D is positive-definite, then equation (4.1a) can be rewritten as $\dot{X}=-\nabla_X(\frac{1}{2}X^{\mathsf{T}}DX-\sum_i Y_i^2X_i)$ and equation (4.1b) also becomes $\dot{Y}=-\nabla_Y(\sum_i \frac{Y_i^2}{2\Gamma_i(X)})$. However, even in this positive-definite case, it is easy to see that there is no single potential for system (4.1) in the extended space $\mathcal{X}\times\mathcal{Y}$, for whatever choice of Γ. This means that we do not know *a priori* if the averaging system is gradient or not. The analytical form of the averaged system (4.2) (with a positive-definite D) shows that this system is gradient if and only if the vector field Γ is gradient, i.e., there exists a scalar function R such that

$$\Gamma(x)=(\Gamma_1(x),\dots,\Gamma_n(x))=\nabla_x R(x). \tag{4.3}$$

This suggests that the existence of an extended potential function is a sufficient, but not a necessary condition for the averaged dynamics to be gradient.

Next, we show the numerical results of the MsGAD for this example. First, we have to run both directions $v(t)$ and $w(t)$ in our MsGAD scheme, because, as previously mentioned, we do not know DF is symmetric *a priori*. To validate our result, we not only show the convergence to the saddle point, but also compare the trajectory $x(t)$ in the MsGAD with that of the classic GAD applied to the known limit equation (4.2). The parameters we used in the numerical tests are the following. $\sigma^2=10$, $D=\begin{bmatrix}0.8 & -0.2\\ -0.2 & 0.5\end{bmatrix}$, and for $x=(x_1,x_2)$,

$$R(x)=\sum_i \arctan(x_i-5),\quad \Gamma_i(x)=\left(1+(x_i-5)^2\right)^{-1},\ i=1,2.$$

The averaged equation (4.2) becomes

$$\dot{\bar{X}}=-\nabla W(\bar{X}),\quad \text{where } W(x)=x^{\mathsf{T}}Dx-\frac{\sigma^2}{2}R(x). \tag{4.4}$$

W has three local minima $m_1=(0.4643,0.6985)$, $m_2=(2.2038,5.9804)$, $m_3=(5.7109,6.2369)$, and two saddle points $s_1=(1.2841,3.4483)$ and $s_2=(3.5689,6.0735)$. See Figure 4.1 below.

In the HMM scheme of the MsGAD, we use the forward Euler solver for the whole system. We take the macro-time step size $\Delta t=0.01$ for both x^ε and $v^\varepsilon,w^\varepsilon$ and set $\tau=1.0$ and $K=1$ for this example. The micro-time step size is $\delta t=\varepsilon\times 0.01$ and the total sampling time $T=10$ is used to estimate the effective force and the Jacobi matrix. The initial values for x are set on three local minima, respectively. The initial values for the directions (v,w) and the fast processes are arbitrarily chosen. Figure 4.1 shows the four GAD trajectories of the x component (dashed line) starting from three local minima. Depending on the initial values of x, these four trajectories converge to the different neighboring saddle points. Two of them which start from m_2, converge to the

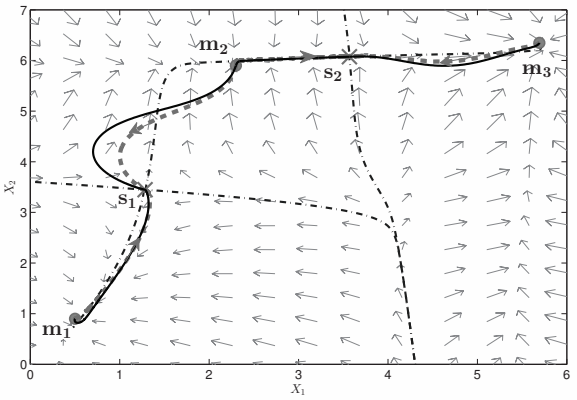


FIG. 4.1. *GAD trajectories from three different local minima (m_1,m_2 and m_3) to two different saddle points (s_1 and s_2). The flow indicated by the arrows corresponds to the averaged gradient system (4.4). The dash-dotted curves are the stable/unstable manifolds of the two saddle points; they determine the basin boundaries of the three local wells. The thick dashed curves with arrows marked are the trajectories of MsGAD by the HMM. As comparison, the thin solid curves are the trajectories of the GAD directly applied to the closed form (4.4) of the averaged system.*

saddle point s_1 and s_2 respectively, due to the different initial values for the directions v and w.

In the end, we conduct a numerical investigation of the variance in estimating the force $F=\bar{f}$ and the Jacobi matrix $DF=\overline{D_x f}+\overline{C}$. Specifically, we look at $\mathrm{Var}_y(f(x,y))$ and $\mathrm{Var}_y(D_x f(x,y)+C(x,y))$, where the variance acts component-wise. The calculation shows that $\mathrm{Var}_y(f(x,y))=\mathrm{Var}_y(y_i^2)=\frac{1}{2}\sigma^4\Gamma_i(x)^2$ and $\mathrm{Var}_y(D_x f(x,y)+C(x,y))=\mathrm{Var}_y(C(x,y))$. For each component, $\mathrm{Var}_y(C_{ij}(x,y))=\mathrm{Var}_y(f_i(x,y)g_j(x,y))$, one can find that $f_i(x,y)g_j(x,y)=-(D_{i1}x_1+D_{i2}x_2+y_i^2)(2y_j^2/\sigma^2+\Gamma_j(x))(x_j-5)$ after plugging in $\Gamma_i'(x_i)=-2(x_i-5)\Gamma_i^2(x_i)$. Thus, we have the analytical result that $\mathrm{Var}_y(C_{ij}(x,y))=w_{ij}(x)(x_j-5)^2$. The expressions of $w_{ij}(x)$ are very long and there is no need to write them down. The key observation we want to draw attention to is that $\mathrm{Var}_y(C_{ij}(x,y))$ is mainly dominated by the term $(x_j-5)^2$, while the variance $\mathrm{Var}_y(f_i(x,y))$ is proportional to $\Gamma_i(x_i)=(1+(x_i-5)^2)^{-1}$. In Figure 4.2, we plot the variance of all components for the force and the covariance matrix and it clearly shows that the variance of C_{i2} near $x_2\approx 5$ are distinctively different from that of F or the other two components. Therefore, there seems to be no general conclusion that the variance of the Jacobi matrix would be larger than the force due to the non-trivial dependency of μ_x on x.

4.2. A coupled Allen–Cahn system. Our second example is a system of stochastic partial differential equations of $(u^\varepsilon(x,t),\phi^\varepsilon(x,t))$ in the Hilbert space $L^2([0,1])$, satisfying Allen–Cahn-type equations with Neumann boundary condition:

$$\begin{cases} \partial_t u^\varepsilon = \kappa^2\Delta u^\varepsilon + u^\varepsilon - (u^\varepsilon)^3 + \mu\phi^\varepsilon, & \text{(4.5a)} \\ \partial_t \phi^\varepsilon = \dfrac{1}{\varepsilon}[\Delta\phi^\varepsilon - \phi^\varepsilon + \mu u^\varepsilon] + \dfrac{\sigma}{\sqrt{\varepsilon}}\dot{W}(t), & \text{(4.5b)} \end{cases}$$

where $\Delta=\partial_x^2$ and $W(t)$ is an $L^2([0,1])$-valued Wiener process with a positive-definite (spatial) covariance operator Q. $\dot{W}$ is white noise in time, κ is the diffusion coefficient

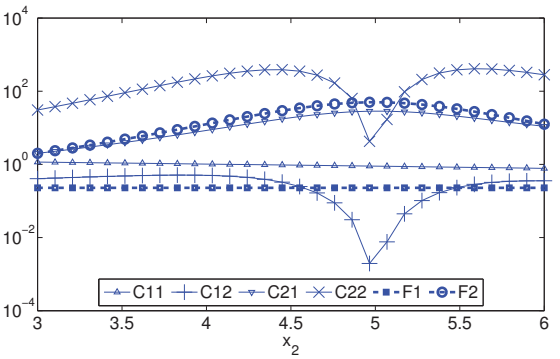


FIG. 4.2. *The variances of the estimator for the effective force F and the Jacobi matrix DF. In our example, the variance of DF equals the variance of C. The plot is the variance for each component along a vertical line segment in the $\mathcal{X}$ plane with $x_2 \in [3,6]$ and $x_1 = 1.2841$. The saddle point s_1 is on this line segment.*

in the slow dynamics, μ is the coupling constant between the slow and the fast dynamics, and σ is the noise intensity. For any fixed $u^\varepsilon = u$, the equilibrium distribution of the SPDE (4.5b) is the Gaussian measure $\mathcal{N}(\mu(I-\Delta)^{-1}u, \sigma^2(I-\Delta)^{-1}Q/2)$ on the Cameron–Martin space. We simply choose Q as identity, i.e., formally, $\dot{W}$ is the spatio-temporal white noise. Then, it is easy to see that the averaged equation for the limit solution $\bar{u}$ is

$$\begin{cases} \partial_t \bar{u} = \kappa^2 \Delta \bar{u} + \bar{u} - \bar{u}^3 + \mu^2 (I-\Delta)^{-1} \bar{u}, & \text{(4.6a)} \\ \dfrac{\partial \bar{u}}{\partial \vec{n}}\Big|_{x=0} = 0, \quad \dfrac{\partial \bar{u}}{\partial \vec{n}}\Big|_{x=1} = 0. & \text{(4.6b)} \end{cases}$$

We can find an energy functional $U(u,\phi)$ jointly for the pair (u,ϕ):

$$U(u,\phi) = \int_\Omega \frac{\kappa^2}{2} u_x^2 + \frac{1}{4}(u^2-1)^2 - \mu u\phi + \frac{1}{2}\phi_x^2 + \frac{1}{2}\phi^2 \, \mathrm{d}x. \tag{4.7}$$

Since this is a gradient system, the corresponding MsGAD only involves one direction (see Section 3):

$$\begin{cases} \partial_t u^\varepsilon = -\delta_u U(u^\varepsilon, \phi^\varepsilon) + 2\dfrac{\langle \delta_u U(u^\varepsilon, \phi^\varepsilon), v^\varepsilon \rangle}{\langle v^\varepsilon, v^\varepsilon \rangle} v^\varepsilon, & \text{(4.8a)} \\ \partial_t \phi^\varepsilon = -\dfrac{1}{\varepsilon}\delta_\phi U(u^\varepsilon, \phi^\varepsilon) + \dfrac{\sigma}{\sqrt{\varepsilon}}\dot{W}(t), & \text{(4.8b)} \\ \partial_t v^\varepsilon = -\delta_u^2 U(u^\varepsilon, \phi^\varepsilon) v^\varepsilon + C v^\varepsilon - \alpha^\varepsilon v^\varepsilon, & \text{(4.8c)} \end{cases}$$

where $\delta_u U$ and $\delta_\phi U$ are the Fréchet derivative of $U(u,\phi)$, $\delta_u^2 U$ is the Hessian, and $C = -\frac{2}{\sigma^2}\delta_u U \otimes \delta_u U + \frac{2}{\sigma^2}\overline{\delta_u U} \otimes \overline{\delta_u U}$.

In this example, the parameters are set as $\kappa = 0.01$, $\sigma = 0.3$ and $\mu = 1$. The two local minima of the effective dynamics are $u \equiv \pm 1$ for $\mu = 0$, and $u \equiv \pm 1.4142$ for $\mu = 1$, respectively. The saddle points are shown in Figure 4.3.

Equation (4.8) is solved by the HMM and the SCM. The time-discretization scheme for equation (4.8a) is a convex-splitting scheme for saddle point search [15] to allow a stable time step size Δt. The spatial discretization is the central finite difference method

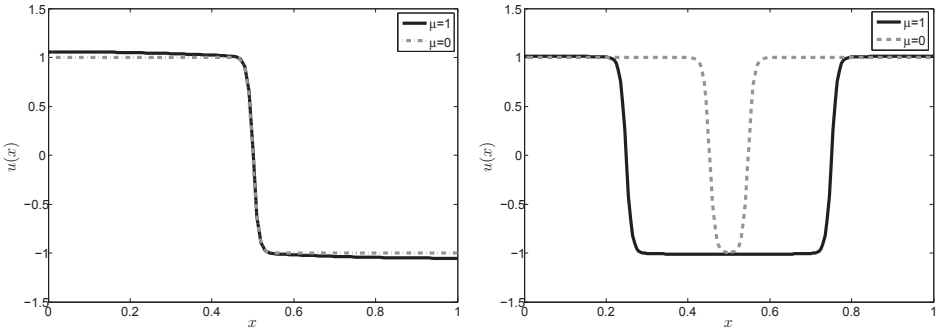


FIG. 4.3. *Two types of index-1 saddle points for $\mu=0$ (dashed lines) and $\mu=1$ (solid lines). The saddle point in (a) has a lower energy than the saddle point in (b).*

with the uniform mesh grid size $\Delta x=1/200$. The term $\delta_u^2 U(u,\phi)v$ in system (4.8) is calculated by finite difference approximation

$$\delta_u^2 U(u,\phi)v=\frac{1}{h}[\delta_u U(u+hv,\phi)-\delta_u U(u,\phi)]$$

with $h=0.001$.

For the seamless coupling method to solve the MsGAD (4.8), according to Section 2.2.2, the boosted system is

$$\begin{cases}\partial_t u^\varepsilon=-\delta_u U(u^\varepsilon,\phi^\varepsilon)+2\dfrac{\langle\delta_u U(u^\varepsilon,\phi^\varepsilon),v^\varepsilon\rangle}{\langle v^\varepsilon,v^\varepsilon\rangle}v^\varepsilon, & \text{(4.9a)}\\ \partial_t\phi^\varepsilon=-\dfrac{1}{\varepsilon\lambda}\delta_\phi U(u^\varepsilon,\phi^\varepsilon)+\dfrac{\sigma}{\sqrt{\varepsilon\lambda}}\dot{W}_1(t), & \text{(4.9b)}\\ \partial_t\psi^\varepsilon=-\dfrac{1}{\varepsilon\lambda}\delta_\psi U(u^\varepsilon,\psi^\varepsilon)+\dfrac{\sigma}{\sqrt{\varepsilon\lambda}}\dot{W}_2(t), & \text{(4.9c)}\\ \partial_t v^\varepsilon=-\delta_u^2 U(u^\varepsilon,\phi^\varepsilon)v^\varepsilon+f(u^\varepsilon,\phi^\varepsilon)\langle g(u^\varepsilon,\phi^\varepsilon),v^\varepsilon\rangle & \\ \qquad\qquad -f(u^\varepsilon,\psi^\varepsilon)\langle g(u^\varepsilon,\phi^\varepsilon),v^\varepsilon\rangle-\alpha^\varepsilon v^\varepsilon, & \text{(4.9d)}\end{cases}$$

where W_1 and W_2 are two i.i.d. copies of W, and $\langle\cdot,\cdot\rangle$ is the inner product in $L^2([0,1])$.

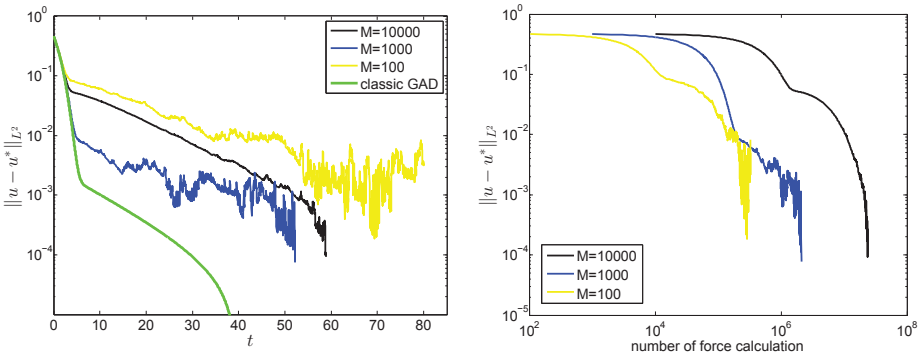
Define the errors as follows,

$$\mathrm{err}_H(t)=\|u_H(x,t)-u^*(x)\|_{L^2},\qquad \mathrm{err}_S(t)=\|u_S(x,t)-u^*(x)\|_{L^2},$$

where $u_H(x,t)$ represents the result of system (4.8) solved by the HMM and $u_S(x,t)$ is the result of system (4.9) solved by the SCM. The true solution $u^*(x)$ is obtained from the classic GAD applied to the closed form of the averaged system (4.6) with a very fine time step size and a sufficiently small tolerance.

To test our HMM and SCM, we calculate the saddle point with the lowest energy, i.e., the profile in the left panel of Figure 4.3. The initial guess is the function $\cos(\pi x)$.

In the HMM, the macro-time step size for u^ε and v^ε is $\Delta t=0.025$ and the micro-time step size for ϕ^ε is $\delta t=0.01\times\varepsilon$. To demonstrate the effect of the sample size M in the HMM, we use different M and run the simulation up to the error 10^{-4} and plot in Figure 4.4 the errors with respect to the time as well as with respect to the



(a) (b)

FIG. 4.4. *(a) shows the evolution of the error with the physical time for the classic GAD applied to (4.6) and for the HMM with various sample sizes $M = 1000$ and 10000. (b) is the decay of the errors with the number of force calculations.*

number of force calculations (including the cost in calculating the direction variables). It is observed that a smaller sample size leads to a relatively larger fluctuation and the fluctuation appears earlier in time. The reduced cost from a smaller M is quite obvious in the subfigure (b). One can expect a further cost reduction if M decreases further; but to attain the error as small as 10^{-4}, one should either switch to a larger M when approaching the saddle point or average the output in time in order to reduce the variance.

Next, we present the details of the SCM and its adaptive version. In the standard SCM, the constant time step size $\widetilde{\Delta}t = 1.0 \times 10^{-2}$ is applied for all components and $\varepsilon = 1.0 \times 10^{-4}, \lambda = 10$ (or effectively, $\varepsilon' = \varepsilon\lambda = 1.0 \times 10^{-3}$). In the adaptive SCM, the same values of $\widetilde{\Delta}t$ and ε' are used up to the time t_a, where the error err_S attains a given threshold 10^{-2}. From the time t_a, the adaptive time step size $\widetilde{\Delta}t$ and the adaptive ε' are then applied. Specifically, we use $\widetilde{\Delta}t_k = k^{-1/2}\widetilde{\Delta}t$ and $\varepsilon'_k = k^{-1/2}\varepsilon'$ where the subindex k counts the number of time steps starting from t_a. As stated in Section 2.3, the final output of the SCM and adaptive SCM are the time-averaged solution $\tilde{u}(x,t) \doteq \frac{1}{t-t_0}\int_{t_0}^{t} u(x,s)\mathrm{d}s$, where $t_0 = 10$ in this example. This continuous-time integral can be easily implemented on the discrete time points in an iterative way.

Figure 4.5 shows the effectiveness of the strategy of decreasing ε' at time marches. For a fixed $\varepsilon' = 10^{-3}$, the error of the SCM cannot reach the level of 10^{-3}. This discrepancy quickly shrinks when we start to decrease the value of ε' gradually. The error eventually can decrease to the level 10^{-5}. However, this comes with a price: as ε' shrinks, the problem becomes stiffer and the step size $\widetilde{\Delta}t$ has to decrease at the same speed. The log-log plot in Figure 4.6 shows the fast growth of the computational cost when ε' decreases in the adaptive SCM. Thus, the SCM is not desired for a tiny ε', though it can quickly settle down to a region of importance with cheap cost. Figure 4.6 also illustrates the performance of the HMM and the SCM. Note that since the time step sizes used here in HMM and adaptive SCM are different, this figure is only a qualitative picture. But what we can learn from the above numerical explorations for this example is that the best strategy in practice perhaps is to start with a boosted value ε' in the adaptive SCM and then switch to the HMM with an adaptive choice of M. The optimal choices of the underlying parameters are of both practical importance

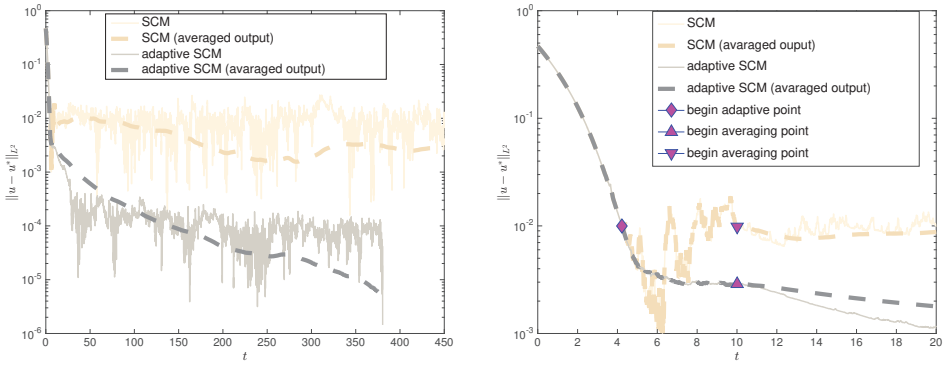


FIG. 4.5. *The evolution of the error with time in the seamless coupling method. The right panel is the same plot in the early stage before time* $t = 20$. *The solutions are averaged with respect to time after the burn-in time* $t_0 = 10$, *which are shown by the two triangular symbols in the plot. The resulted new outputs are shown in dashed thick lines. The value of* ε' *and the step size in the adaptive SCM start to decay when the error decreases to* 0.01, *which is shown by the diamond symbol in the right plot. The "SCM" means the standard SCM with a fixed* ε' *and a constant time step size.*

and theoretic interests and are left for future study.

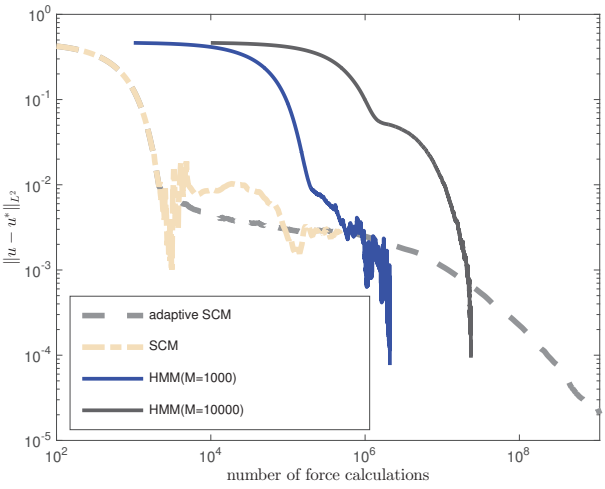


FIG. 4.6. *The illustration of the computational costs for the SCM, the adaptive version of SCM (both with averaged output) and the HMM. The plot is in log-log scale.*

5. Conclusion

We have proposed the multiscale method based on the GAD in slow-fast systems. In order to calculate the saddle point of effective dynamics, we derived a new slow-fast system, the MsGAD, whose effective dynamics is consistent with the GAD for the effective dynamics. By applying the multiscale numerical method such as the HMM or the seamless coupling method, we efficiently compute saddle points on averaged dynamics.

Some adaptive techniques are presented and illustrated by numerical experiment.

Appendix A. To show that B is the Jacobi matrix DF, we just need to prove that

$$\int (f(x,y)-F(x))\otimes\nabla_x \log\rho(x,y)\rho(x,y)(\mathrm{d}y)$$
$$=\int\left(\int_0^\infty \nabla_y \mathrm{E}^y f(x,\widetilde{Y}_\tau^x)\mathrm{d}\tau\right) D_x b(x,y)\rho(x,y)\mathrm{d}y.$$

To show this matrix equality, we just need to show

$$\int (f(x,y)-F(x))\,\langle\nabla_x\rho(x,y),e\rangle\,\mathrm{d}y$$
$$=\int\left(\int_0^\infty \nabla_y \mathrm{E}^y f(x,\widetilde{Y}_\tau^x)\mathrm{d}\tau\right)(D_x b(x,y))e\,\rho(x,y)\mathrm{d}y$$

for arbitrary vector e. Write the infinitesimal perturbation $\widetilde{\rho}:=\rho(x+he,y)-\rho(x,y)\approx h\nabla_x\rho(x,y)\cdot e$ and $\widetilde{b}(x,y)=b(x+he,y)-b(x,y)\approx h(D_x b(x,y))e$ for $h\ll 1$, and assume e is independent of x, then the lemma below tells us that

$$\int (f(x,y)-F(x))\,\langle\nabla_x\rho(x,y),e\rangle\,\mathrm{d}y=-\int\langle\nabla_y u,(D_x b(x,y))e\rangle\,\rho(x,y)\mathrm{d}y,$$

where u satisfies $\mathcal{L}_y u(x,y)=f(x,y)-F(x)$. By the Feymann–Kac formula, we have the representation $u(x,y)=\int_0^\infty \mathrm{E}^y[-f(x,\widetilde{Y}_t^x)+F(x)]\mathrm{d}t$, so $\nabla_y u=-\int_0^\infty \nabla_y \mathrm{E}^y[f(x,\widetilde{Y}_t^x)]\mathrm{d}t$. This completes our proof.

LEMMA A.1. *If $\rho(y)$ is the unique equilibrium probability density function of the SDE*

$$\mathrm{d}Y=b(Y)\mathrm{d}t+\sigma(Y)\mathrm{d}W,$$

$\mathcal{L}$ is the infinitesimal generator, and the infinitesimal perturbation is applied for the drift term: $b\to b+\widetilde{b}$ and the diffusion term $a\doteq\sigma\sigma^T\to a+\widetilde{a}$, where $\widetilde{b}$ and $\widetilde{a}$ are small terms, and let $\Theta\doteq\int\theta(y)\rho(y)\mathrm{d}y$, then the perturbation of Θ is

$$\delta\Theta\doteq\int\theta(y)\widetilde{\rho}(y)\mathrm{d}y=-\int\widetilde{b}\cdot(\rho\nabla u)(y)\mathrm{d}y-\frac{1}{2}\int\widetilde{a}:\rho\nabla\nabla u(y)\mathrm{d}y,$$

where u is the solution of the adjoint equation $\mathcal{L}u=\theta(y)-\Theta$ and decays to 0 at infinity.

Proof. The infinitesimal generator is $\mathcal{L}=b(y)\cdot\nabla+\frac{1}{2}a(y):\nabla\nabla$. The density ρ satisfies the equation $\mathcal{L}^*\rho=0$, where $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$. By linearizing the perturbed equation, we have $\mathcal{L}^*\widetilde{\rho}=\nabla\cdot(\rho\widetilde{b})-\nabla\nabla:(\rho\widetilde{a})$. Multiply this equation by u, then from the integration by parts, we have $\delta\Theta=\int\widetilde{\rho}(y)\theta(y)\mathrm{d}y=\int\widetilde{\rho}(y)\mathcal{L}u\mathrm{d}y=-\int\widetilde{b}\cdot(\rho\nabla u)(y)\mathrm{d}y-\frac{1}{2}\int\widetilde{a}:\rho\nabla\nabla u(y)\mathrm{d}y$. ∎

REFERENCES

[1] J.M. Bofill and W. Quapp, *The variational nature of the gentlest ascent dynamics and the relation of a variational minimum of a curve and the minimum energy path*, Theoretical Chemistry Accounts, 135:1–14, 2015.

[2] J.M. Bofill, W. Quapp, and E. Bernuz, *Some remarks on the model of the extended gentlest ascent dynamics*, Journal of Mathematical Chemistry, 53:41–57, 2014.

[3] F. Bouchet, T. Grafke, T. Tangarife, and E. Vanden-Eijnden, *Large deviations in fast–slow systems*, J. Stat. Phys., 162:793–812, 2016.

[4] G.M. Crippen and H.A. Scheraga, *Minimization of polypeptide energy: XI. the method of gentlest ascent*, Arch. Biochem. Biophys., 144:462–466, 1971.
[5] W. E, *Principle of Multiscale Modeling*, Cambridge University Press, 2011.
[6] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1:87–132, 2003.
[7] W. E, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, *The heterogeneous multiscale method: A review*, Commun. Comput. Phys., 2007.
[8] W. E, W. Ren, and E. Vanden-Eijnden, *A general strategy for designing seamless multiscale methods*, J. Comput. Phys., 228:5437–5453, 2009.
[9] W. E and E. Vanden-Eijnden, *Metastability, conformation dynamics, and transition pathways in complex systems*, in Multiscale Modelling and Simulation, S. Attinger and P. Koumoutsakos, eds., Lecture Notes in Computational Science and Engineering, Springer Berlin Heidelberg, 39:35–68, 2004.
[10] W. E and X. Zhou, *The gentlest ascent dynamics*, Nonlinearity, 24:1831, 2011.
[11] W. E, X. Zhou, and X. Cheng, *Subcritical bifurcation in spatially extended systems*, Nonlinearity, 25(3):761–779, 2012.
[12] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, New York, Third Edition , 2012.
[13] W. Gao, J. Leng, and X. Zhou, *An iterative minimization formulation for saddle point search*, SIAM J. Numer. Anal., 53:1786–1805, 2015.
[14] W. Gao, J. Leng, and X. Zhou, *Iterative minimization algorithm for efficient calculations of transition states*, J. Comput. Phys., 309:69–87, 2016.
[15] S. Gu and X. Zhou, *Convex splitting method for the calculation of transition states of energy functional*, arXiv:1610.07153, 2016.
[16] G. Henkelman and H. Jónsson, *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives*, J. Chem. Phys., 111:7010–7022, 1999.
[17] M. Iannuzzi, A. Laio, and M. Parrinello, *Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics*, Phys. Rev. Lett., 90:238302, 2003.
[18] D. Kelly and E. Vanden-Eijnden, *Fluctuations in the heterogeneous multiscale methods for fast-slow systems*, arXiv:1601.02147, 2016.
[19] R. Z. Khasaminskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory of Probability & Its Applications, 11:390–406, 1966.
[20] R. Z. Khasminskii and G. Yin, *On averaging principles: An asymptotic expansion approach*, SIAM J. Math. Anal., 35:1534–1560, 2004.
[21] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, 2013.
[22] A. Laio and M. Parrinello, *Escaping free-energy minima*, Proceedings of the National Academy of Sciences, 99:12562–12566, 2002.
[23] B. Leimkuhler, C. Matthews, and G. Stoltz, *The computation of averages from equilibrium and nonequilibrium langevin molecular dynamics*, IMA Journal of Numerical Analysis, 36:13–79, 2016.
[24] C. Li, J. Lu, and W. Yang, *Gentlest ascent dynamics for calculating first excited state and exploring energy landscape of Kohn-Sham density functionals*, The Journal of Chemical Physics, 143:224110, 2105.
[25] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *String method in collective variables: Minimum free energy paths and isocommittor surfaces*, J. Chem. Phys., 125:024106, 2006.
[26] N. Mousseau and G. Barkema, *Traveling through potential energy surfaces of disordered materials: the activation-relaxation technique*, Phys. Rev. E, 57:2419–2424, 1998.
[27] G. Papanicolaou, *Introduction to the asymptotic analysis of stochastic equations*, in Lectures in Appl. Math., Amer. Math. Soc., Providence, R.I., 16:109–147, 1977.
[28] P. Rein ten Wolde, M.J. Ruiz-Montero, and D. Frenkel, *Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling*, J. Chem. Phys., 104:9932–9947, 1996.
[29] A. Samanta, M. Chen, T.-Q. Yu, M. Tuckerman, and W. E, *Sampling saddle points on a free energy surface*, J. Chem. Phys., 140:164109, 2014.
[30] A. Samanta and W. E, *Atomistic simulations of rare events using gentlest ascent dynamics*, J. Chem. Phys., 136:124104, 2012.
[31] A. Veretennikov, *On large deviations for SDEs with small diffusion and averaging*, Stoch. Proc. Appl., 89:69–79, 2000.
[32] D.J. Wales, *Energy Landscapes with Application to Clusters, Biomolecules and Glasses*, Cambridge University Press, 2003.

[33] J. Zhang and Q. Du, *Shrinking dimer dynamics and its applications to saddle point search*, SIAM J. Numer. Anal., 50:1899–1921, 2012.

[34] L. Zhang, Q. Du, and Z. Zheng, *Optimization-based shrinking dimer method for finding transition states*, SIAM J. Sci. Comput., 38:A528–A544, 2016.

[35] L. Zhang, W. Ren, A. Samanta, and Q. Du, *Recent developments in computational modeling of nucleation in phase transformations*, Computational Materials, 2:16003, 2016.

[36] L. Zhang, J. Zhang, and Q. Du, *Finding critical nuclei in phase transformations by shrinking dimer dynamics and its variants*, Commun. Comput. Phys., 16:781–798, 2014.