COMMUN. MATH. SCI.
Vol. 15, No. 6, pp. 1507–1525

SENSITIVITY ANALYSIS AND OPTIMIZATION OF REACTION RATE*

SHUTING GU†, LING LIN‡, AND XIANG ZHOU§

Abstract. The chemical reaction rate from reactant to product depends on the geometry of potential energy surface (PES) as well as the temperature. We consider a design problem of how to choose the best PES from a given family of smooth potential functions in order to maximize (or minimize) the reaction rate for a given chemical reaction. By utilizing the transition-path theory, we relate reaction rate to committor functions which solves boundary-value elliptic problems, and perform the sensitivity analysis of the underlying elliptic equations via adjoint approach. We derive the derivative of the reaction rate with respect to the potential function. The shape derivative with respect to the domains defining reactant and product is also investigated. The numerical optimization method based on the gradient is applied for two simple numerical examples to demonstrate the feasibility of our approach.

Keywords. rare event; reaction rate; transition path theory; sensitivity analysis.

AMS subject classifications. 49Q12; 82C80.

1. Introduction

The understanding of rare events is one of the most fundamental problems in chemistry. The typical examples of rare events are chemical reactions, conformation changes of molecules and so on. A chemical reaction is conveniently discussed in terms of a reaction energy profile [1–3]. For systems with smooth energy landscapes, the well-known transition state theory [4–6] or Kramers rate theory [2] gives a sufficiently accurate description of the transition process and the corresponding chemical reaction rate. The main object of interest in these theories is transition state, which is saddle point on potential energy landscape. The transition rate is related to the energy barrier in the form of Arrhenius law, which works asymptotically well on smooth potential and at low temperature.

Recently, the transition-path theory (TPT) has been developed in [7] to better address the challenges in systems with rugged potential energy landscapes, or when entropic effects matter. See also [8–10] and the review in [11]. As a theoretical tool for analyzing the transition-path ensemble (in practice, these paths are usually collected from the transition path sampling (TPS) technique [12]) or other methods), the TPT not only characterizes the probability distribution of the particles in the transition-path ensemble and the associated probability current, but also provides an exact formula for transition rate, regardless of the complexity of reaction energy profile. The TPT-rate is given by the committor functions, which are also referred to as capacitance functions [13, 14], with a clear probability meaning of heading toward the product state before returning to the reactant state. The committor function satisfies a boundary-value problem for an elliptic partial differential equation.

*Received: February 23, 2016; accepted (in revised form): August 16, 2016. Communicated by David Anderson.
This research of XZ was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11304314, 109113,11304715). L. Lin acknowledges the partial financial support of the DRS Fellowship Program of Freie Universität Berlin.
†Department of Mathematics, City University of Hong Kong, Tat Chee Ave, Kowloon, Hong Kong.
‡Department of Mathematics, City University of Hong Kong, Tat Chee Ave, Kowloon, Hong Kong.
§Department of Mathematics, City University of Hong Kong, Tat Chee Ave, Kowloon, Hong Kong (xizhou@cityu.edu.hk).

Since the various tools for calculating the reaction rate have been developed, it would be convenient in practice to explore different structures or morphologies of molecules so that the control of the reaction rate can be achieved to design better chemical products. One main approach is to test many possibilities of the underlying potential energy surface and select the desired one by optimization. The basic task to perform such optimization procedures is to determine the sensitivity of the rate with respect to the perturbation of the current potential energy, i.e., to calculate the "derivative" of the objective function — the reaction rate — with respect to the free parameters which can be tuned in a family of potential surfaces. This paper serves a mathematical derivation of the sensitivity analysis when the potential function and/or the domain of states are perturbed. Then based on the derivative information, we apply it to the optimization problems of maximizing or minimizing the rate. Mathematically, we work on the sensitivity analysis of the elliptic PDE with respect to the coefficients and the domain by the adjoint method.

The adjoint approach for sensitivity analysis was carried out for a Fokker–Planck equation in a climate model in [15]. Our methods share some elementary common things with this work since the adjoint idea is the standard approach for sensitivity analysis, but our problems for the reaction rate have different setup of the boundary-value problems and have more complicated objective functionals. Moreover, more advanced mathematical generalizations are involved here; for example, the use of variational structure of our problem (i.e., detailed balance) significantly simplifies some calculations. Furthermore, our calculations include the shape derivative by using the technique in [16], which helps evaluate the proper size of the domains defining the product or reactant states. This is the first time, to the best knowledge of the authors, to consider the effect of the defining domains for reactant and product on the evaluation of reaction rate. In real application problems, there might be different type of kinetic measurements which define the product and reactant states differently, then the shape derivative of the reaction rate provides a useful criterion to determine which definitions of the product and reactant states are most appropriate. The key element in our final result is a mean first passage time of hitting either the reactant state or the product state (see Equation (3.9)).

In the end, we want to remark that we are allowed to change the potential function itself rather than its gradient force. In some areas such as the large deviation or importance sampling, the controlled process for the underlying stochastic differential equation (SDE) corresponds to an extra drift term (cf. [17–19]); when this drift term is driven as the gradient of another function (such as value function or quasi-potential, which is quite common in many cases), then there is an analogy between our problem in this paper and those works. In those works, the drift is being changed via an absolutely continuous change of measure with the goal of simulating the system under a different measure. The goal of those papers is to minimize the variance of the importance sampling Monte Carlo method by picking up the optimal drift, which is usually approximated in the large deviation asymptotic region via Hamilton–Jacobi–Bell equation. Our analysis in this paper is entirely based on the elliptic PDE in the non-asymptotic region and would help solve these problems in a different perspective, beyond the contribution to the optimization of the reaction rate.

The paper is organized as follows. In Section 2, we introduce the TPT and the definitions and calculations of two rates defined in the TPT. Section 3 is our main result about the sensitivity analysis. Section 4 presents two numerical examples. Section 5 contains our concluding discussion.

2. Reaction rate and transition path theory

We consider the following diffusion process X_t defined by an overdamped Langevin equation[1] in $\mathbb{R}^n$ associated with a potential function U,

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W, \tag{2.1}$$

where W is the standard Brownian motion in $\mathbb{R}^n$ and $\varepsilon = k_B T$ is the thermal noise amplitude. The infinitesimal generator of this diffusion process is

$$\mathcal{L}f = -\nabla U \cdot \nabla f + \varepsilon \Delta f,$$

and the formal adjoint of $\mathcal{L}$ is

$$\mathcal{L}^* f = \nabla \cdot (\nabla U f) + \varepsilon \Delta f.$$

The equilibrium probability density function (pdf) of equation (2.1) is given by

$$\rho(x) = Z^{-1} e^{-U(x)/\varepsilon}, \quad \text{where} \quad Z = \int e^{-U(x)/\varepsilon}\,\mathrm{d}x.$$

As a convention, the integral domain is $\mathbb{R}^n$ if not specified. This pdf ρ is the unique solution to the Fokker–Planck equation

$$\mathcal{L}^* \rho = \nabla \cdot (\rho \nabla U) + \varepsilon \Delta \rho = 0 \tag{2.2}$$

satisfying the normalization condition

$$\int \rho(x)\,\mathrm{d}x = 1. \tag{2.3}$$

It is worthwhile noting that $\mathcal{L}$ has the following divergence form with the aid of ρ,

$$\mathcal{L}f = -\nabla U \cdot \nabla f + \varepsilon \Delta f = \frac{\varepsilon}{\rho} \nabla \cdot (\rho \nabla f). \tag{2.4}$$

Let A and B be two disjoint closed domains (with piecewise continuous boundaries) in $\mathbb{R}^n$ and they represent the reactant and the product, respectively, in a chemical reaction of concern. Denote the set $D = \mathbb{R}^n \setminus (A \cup B)$ or $(A \cup B)^c$.

The *transition rate* between A and B in the TPT is defined as the number of transition events from the reactant set A to the product set B in unit time and is calculated as follows via the ergodicity property (see [11]),

$$\nu = \varepsilon \int_D \rho(x) \left|\nabla q(x)\right|^2 \mathrm{d}x, \tag{2.5}$$

where the comittor function q is the solution to the following boundary-value problem on D

$$\begin{cases} \mathcal{L}q = 0, & x \in D, \\ q|_A = 0, \quad q|_B = 1. \end{cases} \tag{2.6}$$

[1]The underdamped Langevin dynamics can also be studied by the TPT. If one assumes the committor function $q(x,p)$ can be represented by a function of the positions $q(x)$ only, then the problem exactly reduces to the overdamped case. See Section 3.5 in [11] for details.

The above rate ν should not be confused with the two reaction rates from A to B and B to A defined, respectively, as $\kappa_{A,B}=\nu/m_A$ and $\kappa_{B,A}=\nu/m_B$ where $m_A=\int\rho(x)(1-q(x))\,\mathrm{d}x=1-\int\rho(x)q(x)\,\mathrm{d}x$ and $m_B=\int\rho(x)q(x)\,\mathrm{d}x$. The meanings of m_A and m_B are the time fractions during which a generic reactive trajectory was assigned to A or B([11]). The rates $\kappa_{A,B}$ and $\kappa_{B,A}$ are the phenomenological chemical reaction rates for the reaction direction from A to B, and from B to A, respectively. We call $\kappa_{A,B}$ the A-B reaction rate and $\kappa_{B,A}$ the B-A reaction rate.

For the given sets A and B, we only need consider A-B reaction rate $\kappa_{A,B}$ mathematically since $\kappa_{B,A}$ can be done in exactly the same way. So, to be concrete, we only study $\kappa_{A,B}$ and denote it simply as κ throughout the paper. Thus, the rate of our main interest is κ, defined as

$$\kappa=\frac{\nu}{m},\tag{2.7}$$

where

$$m=\int\rho(x)(1-q(x))\,\mathrm{d}x=1-\int\rho(x)q(x)\,\mathrm{d}x.\tag{2.8}$$

We shall use the notation $\nu[U,A,B]$ later (similarly $\kappa[U,A,B]$, $m[U,A,B]$, etc.) to emphasise the dependence on the potential U or the sets A and B.

3. Sensitivity analysis

The question of interest is how to tune the potential function U in certain given range to optimize the A-B reaction rate κ. To achieve this goal, we first need to calculate the Fréchet derivative of the reaction rate κ with respect to the potential function U for the fixed domains A and B. In the following, we use $\kappa[U]$ to emphesize the dependence on the function U, i.e., we regard κ as a functional of U. Denote the L^2 inner product in $\mathbb{R}^n$ by $\langle\cdot,\cdot\rangle$. Assume that the admissible set of the potential U is $\mathbb{U}$. For example, $\mathbb{U}$ is the set of all C^1 functions on $\mathbb{R}^n$ with a lower bound and grows to infinity sufficiently fast when $|x|\to\infty$. For such potentials, the density $\rho=Z^{-1}e^{-U/\varepsilon}$ then belongs to $L^1(\mathbb{R}^n)$ to make sense. For specific applications, the admissible set $\mathbb{U}$ may contain some extra constraints from physical reality. We assume all potential functions U in consideration belong to $\mathbb{U}$. We first have the following Fréchet derivative for the equilibrium pdf ρ.

LEMMA 3.1. *Given a function f on $\mathbb{R}^n$ such that*

$$F(U):=\langle f,\rho[U]\rangle<\infty,\quad \textit{where } \rho[U]=e^{-U/\varepsilon}/\Big(\int e^{-U/\varepsilon}\,\mathrm{d}x\Big),$$

then

$$\frac{\delta F}{\delta U}=\varepsilon^{-1}(F-f)\rho.\tag{3.1}$$

Proof. Denote the infinitesimal change of U by $\widetilde{U}$ and the infinitesimal perturbation of the equilibrium measure ρ by $\widetilde{\rho}$. Then $\delta F=\langle f,\widetilde{\rho}\rangle$. We first linearize the elliptic Equation (2.2) by expanding

$$\mathcal{L}^*[U+\widetilde{U}](\rho+\widetilde{\rho})=\nabla\cdot\Big((\nabla U+\nabla\widetilde{U})(\rho+\widetilde{\rho})\Big)+\varepsilon\Delta(\rho+\widetilde{\rho})=0$$

and neglecting the high order term $\nabla\cdot(\tilde{\rho}\nabla\tilde{U})$, and then after reduction we get

$$\nabla\cdot\Big(\rho\nabla\widetilde{U}\Big)+\mathcal{L}^*\widetilde{\rho}=0,$$

where $\mathcal{L}^*$ is associated with U. Now multiply the above by a smooth function g which decays sufficiently fast at infinity (g and ∇g vanishes at infinity), and integrate by parts twice, then we have that

$$\left\langle \nabla\cdot(\rho\nabla g), \widetilde{U}\right\rangle + \langle \mathcal{L}g, \widetilde{\rho}\rangle = 0.$$

By the equivalent form of $\mathcal{L}$ in equation (2.4), the above equation leads to

$$\left\langle \mathcal{L}g, \rho\widetilde{U}\right\rangle/\varepsilon + \langle \mathcal{L}g, \widetilde{\rho}\rangle = 0. \tag{3.2}$$

By choosing the function g such that

$$\langle \widetilde{\rho}, f - \mathcal{L}g\rangle = 0 \tag{3.3}$$

holds for all $\widetilde{\rho}$ satisfying the constraints $\int \widetilde{\rho}\mathrm{d}x = 0$, we then have an equation for g:

$$f - \mathcal{L}g = c, \tag{3.4}$$

where c is a constant. Then, by equations (3.2),(3.3) and (3.4), it holds that

$$\begin{aligned}\delta F &= \langle f, \widetilde{\rho}\rangle - 0 = \langle \widetilde{\rho}, f - \mathcal{L}g\rangle - \left\langle \mathcal{L}g, \rho\widetilde{U}\right\rangle/\varepsilon \\ &= -\left\langle \mathcal{L}g, \rho\widetilde{U}\right\rangle/\varepsilon = \left\langle \rho(c-f), \widetilde{U}\right\rangle/\varepsilon.\end{aligned} \tag{3.5}$$

The constant c is actually equal to F since multiplying equation (3.4) by ρ and integration by parts show that

$$F = \langle f, \rho\rangle = c + \langle \mathcal{L}^*\rho, g\rangle = c.$$

So, we showed that

$$\delta F = \left\langle \rho(F-f)/\varepsilon, \widetilde{U}\right\rangle. \tag{3.6}$$

□

Our main result is the following theorem.

THEOREM 3.1. *The Fréchet derivative of ν is*

$$\frac{\delta\nu}{\delta U} = \rho(\nu/\varepsilon - |\nabla q|^2). \tag{3.7}$$

The Fréchet derivativeof m is

$$\frac{\delta m}{\delta U} = \frac{\rho}{\varepsilon}(q + m - 1) - \rho(\nabla q\cdot\nabla w), \tag{3.8}$$

where $w : D \to \mathbb{R}$ is the solution of the equation

$$\begin{cases} \mathcal{L}w = -1, & x \in D, \\ w|_{\partial D} = 0. \end{cases} \tag{3.9}$$

It follows that the Fréchet derivative of κ is

$$\frac{\delta\kappa}{\delta U} = \frac{1}{m^2}\left(m\frac{\delta\nu}{\delta U} - \nu\frac{\delta m}{\delta U}\right) = \frac{\rho}{m}\left(\kappa(1-q)/\varepsilon + \kappa\nabla q\cdot\nabla w - |\nabla q|^2\right). \tag{3.10}$$

REMARK 3.1. The probabilistic interpretation of $w(x)$ is the mean first passage of entering the domain $A\cup B$ of the diffusion process (2.1) for $X_0 = x \in D$. If A and B are two wells of the potential energy surface U, this function $w(x)$ is much less than the transition time between two wells A and B. This means the direct stochastic simulation of the mean first passage time w is much faster than the direct simulation of A-B transition.

REMARK 3.2. From the proof of this theorem, one can easily verify that $\int_{\mathbb{R}^n} \frac{\delta\kappa}{\delta U}\,\mathrm{d}x = 0$ with the Fréchet derivative given in equation (3.10). So any perturbation in the form of $U(x) \to U(x) + c$ for a constant c leads to no change of κ.

REMARK 3.3. Equation (3.10) shows that, since $\nabla q \equiv 0$ on $A\cup B$, $\frac{\delta\kappa}{\delta U}$ vanishes inside B and equals to $\frac{\kappa\rho}{\varepsilon m}$ inside A. Note that $\frac{\kappa\rho}{\varepsilon m}$ is always positive, then it implies that just increasing U inside of the set A always increases κ.

REMARK 3.4. There is no guarantee that ∇q or ∇w is continuous on the boundary $\partial D = \partial A \cup \partial B$.

Proof. **(Proof of Theorem 3.1.)** Denote the infinitesimal perturbation of the equilibrium measure ρ by $\widetilde{\rho}$ and the infinitesimal perturbation of the committor function q by $\widetilde{q}$, respectively. Then from equation (2.5),

$$\delta\nu = \delta\nu_1 + \delta\nu_2 := \int \varepsilon\widetilde{\rho}|\nabla q|^2\,\mathrm{d}x + 2\int \varepsilon\rho\nabla q\cdot\nabla\widetilde{q}\,\mathrm{d}x.$$

The first term $\delta\nu_1$ is given by Lemma 3.1 by letting $f = \varepsilon|\nabla q|^2$. To prove the equality (3.7), it suffices to show that $\delta\nu_2 := 2\varepsilon\int \rho\nabla q\cdot\nabla\widetilde{q}\,\mathrm{d}x$ actually vanishes. This is from the fact that the committor function q minimizes the functional $\int \rho|\nabla q|^2\,\mathrm{d}x$ (see [11]). For any test function ϕ vanishing on ∂D, $\langle\rho\nabla q, \nabla\phi\rangle = 0$ holds for the minimizer q. The choice of $\phi = \widetilde{q}$ gives the conclusion $\frac{\delta\nu_2}{\delta U} = 0$.

Next we prove the equality (3.8). By the definition of m in equation (2.8), we write

$$\delta m = \delta m_1 + \delta m_2 := -\int q\widetilde{\rho}\,\mathrm{d}x - \int_D \rho\widetilde{q}\,\mathrm{d}x.$$

Applying Lemma (3.1) with $f = -q$ gives that

$$\delta m_1 = \varepsilon^{-1}\left\langle \rho(q+m-1), \widetilde{U}\right\rangle.$$

For δm_2, linearizing $\mathcal{L}[U+\widetilde{U}](q+\widetilde{q}) = 0$ gives

$$\mathcal{L}\widetilde{q} = \nabla q\cdot\nabla\widetilde{U}.$$

By equation (2.4) and $\mathcal{L}q = 0$, then we can calculate that

$$\rho\mathcal{L}\widetilde{q} = \rho\nabla q\cdot\nabla\widetilde{U} = \nabla\cdot(\widetilde{U}\rho\nabla q) - \widetilde{U}\nabla\cdot(\rho\nabla q) = \nabla\cdot(\widetilde{U}\rho\nabla q). \tag{3.11}$$

From this result and noting that $\mathcal{L}w = \frac{\varepsilon}{\rho}\nabla\cdot(\rho\nabla w) = -1$, i.e., $-\rho = \varepsilon\nabla\cdot(\rho\nabla w)$, as well as the boundary conditions of w, we obtain

$$\begin{aligned}\delta m_2 &= -\int_D \tilde{q}\rho\,\mathrm{d}x = \int_D \varepsilon\tilde{q}\nabla\cdot(\rho\nabla w)\,\mathrm{d}x\\ &= \varepsilon\int_D w\nabla\cdot(\rho\nabla\tilde{q})\,\mathrm{d}x \qquad \text{(integration by parts)}\end{aligned}$$

$$
\begin{aligned}
&= \int_D w\rho\mathcal{L}\widetilde{q}\,\mathrm{d}x \\
&= \int_D w\nabla\cdot(\widetilde{U}\rho\nabla q)\,\mathrm{d}x \qquad \because (3.11) \\
&= -\int_D \widetilde{U}\rho\nabla q\cdot\nabla w\,\mathrm{d}x. \qquad (3.12)
\end{aligned}
$$

Therefore, the Fréchet derivative of m in (3.8) holds. The result (3.10) is a trivial calculation based on the Fréchet derivatives (3.7) and (3.8). ∎

A corollary of Theorem 3.1 is on the derivative of the reaction rate κ with respect to the noise amplitude ε.

COROLLARY 3.1.

$$
\frac{\partial\kappa}{\partial\varepsilon} = \varepsilon^{-1}\kappa - \left\langle \frac{\rho}{m}\left(\varepsilon^{-2}\kappa(1-q)+\varepsilon^{-1}\kappa\nabla q\cdot\nabla w-\varepsilon^{-1}|\nabla q|^2\right), U\right\rangle. \qquad (3.13)
$$

Here all functions are associated with the potential U and the noise amplitude ε.

Proof. Now we write $\rho=\rho[U,\varepsilon]$, $q=q[U,\varepsilon]$ and $w=w[U,\varepsilon]$ to denote the dependence on both the potential U and the noise size ε. The key observations are the following simple scaling relations: $\rho[U,\varepsilon]=\rho[U/\varepsilon,1]$, $q[U,\varepsilon]=q[U/\varepsilon,1]$. Then it follows that the same scaling holds for ν/ε, m and consequently, $\kappa[U,\varepsilon]=\varepsilon\kappa[U/\varepsilon,1]$. Denote the Fréchet derivative in (3.10) by $g[U,\varepsilon]$ to avoid confusion. Then, the derivative in (3.13) is

$$
\frac{\partial\kappa[U,\varepsilon]}{\partial\varepsilon} = \kappa[U/\varepsilon,1]+\varepsilon\frac{\partial\kappa[U/\varepsilon,1]}{\partial\varepsilon} = \frac{\kappa[U,\varepsilon]}{\varepsilon} - \varepsilon^{-1}\langle g[U/\varepsilon,1],U\rangle.
$$

Since $w[U/\varepsilon,1]=\varepsilon w[U,\varepsilon]$, then the result (3.13) follows immediately. ∎

The next theorem is the sensitivity analysis including the shape derivative of the reaction rate with respect to the perturbation of the domain. This result is useful to determine if the choice of A and B is appropriate: a too large shape derivative implies a bad choice of these sets to represent the reactant and product states. The proof of this Theorem 3.2 is more technical than the proof of Theorem 3.1, so we place it in the Appendix.

THEOREM 3.2. *Let U be infinitesimally perturbed as $U'=U+\widetilde{U}$, where $|\widetilde{U}|\ll 1$. For $\Omega=A$, B, or D, denote the perturbed domains by Ω' and we define $\delta\Omega$ to be the symmetric difference $\Omega'\triangle\Omega=(\Omega'\setminus\Omega)\cup(\Omega\setminus\Omega')$. Let $n(x)$ be the outer normal unit vector (for the domain D) at $x\in\partial D$. We represent the infinitesimal perturbation of the boundary ∂D by an infinitesimal function $\alpha(x)$ defined on the boundary ∂D as follows: any $x\in\partial D$ is transformed to the new point x' on $\partial D'$, defined as $x'=x+\alpha(x)n(x)$. See Figure 3.1.*

Then the infinitesimal perturbation of the rate κ, $\delta\kappa[U,A,B]:=\kappa[U',A',B']-\kappa[U,A,B]$ has the following expression

$$
\delta\kappa = \left\langle \frac{\rho}{m}\left(\kappa(1-q)/\varepsilon+\kappa\nabla q\cdot\nabla w-|\nabla q|^2\right),\widetilde{U}\right\rangle + \left\langle \frac{\varepsilon\rho}{m}\left(\kappa\nabla q\cdot\nabla w-|\nabla q|^2\right),\alpha\right\rangle_{\partial D} \qquad (3.14)
$$

where $\langle\cdot,\cdot\rangle_{\partial D}$ means the boundary integration on ∂D.

REMARK 3.5. By the strong maximum principle and the Hopf Lemma, it is true that

$$
\frac{\partial w}{\partial n} < 0 \quad \text{on } \partial D.
$$

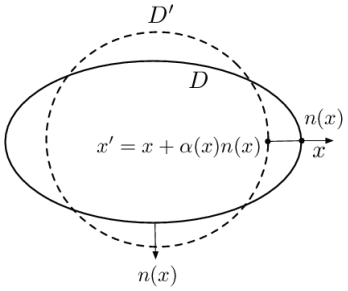

FIG. 3.1. *Illustration of the perturbation of the domain D. The solid and dashed curves are the boundaries of the original domain D and the perturbed domain D' respectively. $n(x)$ is the outer normal unit vector on the boundary of D at $x \in \partial D$. The infinitesimal perturbation is specified by the mapping $x \in \partial D \to x' \in \partial D'$ in which $x' = x + \alpha(x)n(x)$. For the point x plotted in the above figure for illustration, which belongs to $(\partial D) \setminus D'$, the sign of $\alpha(x)$ is negative.*

Note that $n \parallel \nabla q$ and their directions are opposite on ∂A and the same on ∂B. Thus

$$\nabla q \cdot \nabla w = \pm |\nabla q| \frac{\partial w}{\partial n} \quad \text{on } \partial D \tag{3.15}$$

with the minus sign on ∂A and the plus sign on ∂B. So the variation corresponding to the change of B

$$\varepsilon \frac{\rho}{m} \left(\kappa |\nabla q| \frac{\partial w}{\partial n} - |\nabla q|^2 \right)$$

is always negative. This means that increasing the set B (leading to shrinking D and $\alpha < 0$ on ∂B) will always increase κ.

4. Numerical methods and examples

Assume that the family of the potential functions can be parametrized as $U(x;a)$ where $a = (a_1, a_2, \cdots, a_p)$ is the parameter in $\mathbb{R}^p$. Then the infinitesimal perturbation of U is

$$\delta U = \nabla_a U(x,a) \cdot \delta a$$

and it follows that

$$\delta \kappa = \int_{D \cup A} \frac{\delta \kappa}{\delta U}(x) \nabla_a U(x,a) \cdot \delta a \, \mathrm{d}x = \int_{D \cup A} \frac{\delta \kappa}{\delta U}(x) \nabla_a U(x,a) \, \mathrm{d}x \cdot \delta a.$$

The function $\frac{\delta \kappa}{\delta U}(x)$ is given in equation (3.10). The integral domain here is $D \cup A = B^c$ due to Remark 3.3. Thus, the gradient of κ with respect to the parameter a is

$$\nabla_a \kappa(a) = \int_{D \cup A} \frac{\delta \kappa}{\delta U}(x) \nabla_a U(x,a) \, \mathrm{d}x, \tag{4.1}$$

i.e.,

$$\frac{\partial \kappa(a)}{\partial a_j} = \int_{D \cup A} \frac{\delta \kappa}{\delta U}(x) \frac{\partial U(x;a)}{\partial a_j} \, \mathrm{d}x.$$

In many cases, the numerical value of κ is too small so it is beneficial to work on $\log(\kappa)$. Then

$$\frac{\partial \log \kappa(a)}{\partial a_j} = \int_{D\cup A} \frac{1}{\kappa}\frac{\delta \kappa}{\delta U}(x)\frac{\partial U(x;a)}{\partial a_j}\,\mathrm{d}x,$$

where

$$\frac{1}{\kappa}\frac{\delta \kappa}{\delta U} = \frac{\rho}{m}\left(\frac{1-q}{\varepsilon} + \nabla q\cdot\nabla w\right) - \frac{\rho}{\nu}|\nabla q|^2.$$

And also,

$$\frac{\partial \log \kappa(a,\varepsilon)}{\partial \varepsilon} = \varepsilon^{-1} - \left\langle \frac{\rho}{m}\left(\varepsilon^{-2}(1-q) + \varepsilon^{-1}\nabla q\cdot\nabla w\right) - \frac{\rho}{\nu}\varepsilon^{-1}|\nabla q|^2, U\right\rangle.$$

Here $\nu = \kappa m$ is used.

The design problem is an optimization problem of $\min \kappa(a)$ or $\max \kappa(a)$ subject to certain constraints. For example, the constraint $\|U(x;a) - U(x;a_0)\| \le c$ for a threshold c corresponds to the penalty of changing the potential energy surface away from a particular surface at a_0. Based on the result (4.1), any gradient-based numerical optimization method can be used to solve this design problem. In the next, we present two examples to show how this method works.

4.1. 1D double-well potential perturbed by periodic field. We consider the following superposition of the well-known double-well potential function and a periodic potential from a given family parametrized by the parameter a:

$$U(x;a) = U_0(x) + U_1(x;a) := \frac{1}{4}(x^2-1)^2 + \frac{2}{5}(1+\sin(a_1 x - a_2)).$$

The parametrized periodic potential is applied on the system as an external field. The parameter is $a = (a_1, a_2)$ where $a_1 > 0$ is the frequency and a_2 is the phase. Since the two local minima of the double-well potential U_0 is at ± 1, we then fix the set $A = (-\infty, -0.85]$ and $B = [0.85, \infty)$ to represent the reactant and product states, respectively. So the domain $D = (-0.85, 0.85)$. Set $\varepsilon = 0.03$. We want to maximize the A-B reaction rate κ.

The elliptic PDEs for q and w were solved on D by the finite difference scheme. The differentiation (4.1) was calculated by the standard numerical quadrature from the numerical values of ∇q and ∇w. We use the MATLAB routine `fmincon` as our numerical solver for the optimization and find the optimal $a^* = (1.643, 2.521)$. In the left panel of Figure 4.1, the search path from an initial guess is shown with the contour plot of the objective function $\kappa(a)$ as a function of $a = (a_1, a_2)$. The right panel of Figure 4.1 shows how the original symmetric potential U_0 is changed by adding the optimal periodic $U_1(x;a^*)$. The half period at the optimal solution is $\pi/a_1^* \approx 2$, roughly equal to the distance between the two wells located at ± 1.

Figure 4.2 shows the partial derivatives of $\kappa(a)$ with respect to a_1 for $a_2 = a_2^* = 2.521$ and to a_2 for $a_1 = a_1^* = 1.643$. The optimal $a^* = (1.643, 2.521)$ is confirmed furthermore clearly by this figure that the partial derivatives $\partial\kappa/\partial a_i$ are positive for $a_i < a_i^*$ and negative for $a_i > a_i^*$, $i = 1, 2$. Figure 4.3 shows the partial derivative of $\log \kappa(a, \varepsilon)$ with respect to ε at the optimal $a^* = (1.643, 2.521)$. It is observed that $\partial \log \kappa/\partial \varepsilon$ is always positive, therefore the rate κ is an increasing function of the noise amplitude ε, as we expected. Figure 4.4 shows the result if the problem is to maximize the transition rate ν. It is observed that the optimal profile of the potential function has a flat region

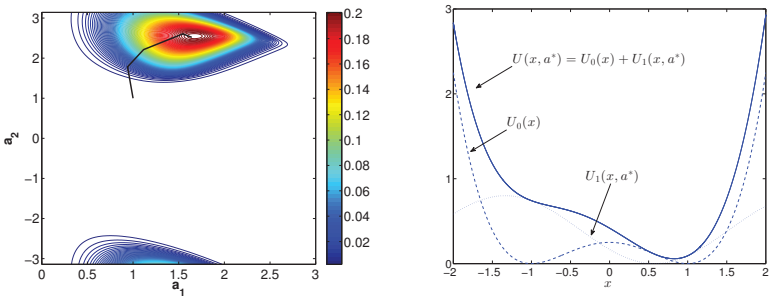

FIG. 4.1. *left: Contour plot of* $\kappa(a)$ *for* $\varepsilon = 0.03$. *The black lines are the search path of the optimization algorithm from the initial* $(1,1)$ *toward the optimal solution* $a^* = (1.643, 2.521)$. *right: The optimally tilted potential function* $U(x;a^*)$. $U_0(x)$ *and* $U_1(x;a^*)$ *are also plotted as the dashed and dotted curves, respectively.*

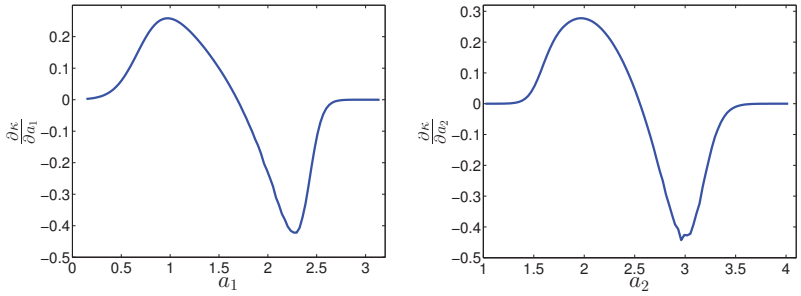

FIG. 4.2. *left: The partial derivative of* $\kappa(a)$ *with respect to* a_1 *for* $a_2 = a_2^* = 2.521$ *and* $\varepsilon = 0.03$. *right: The partial derivative of* $\kappa(a)$ *with respect to* a_2 *for* $a_1 = a_1^* = 1.643$ *and* $\varepsilon = 0.03$.

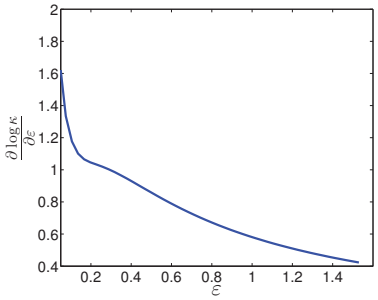

FIG. 4.3. *The partial derivative of* $\log\kappa(a,\varepsilon)$ *with respect to* ε *for the optimal* $a^* = (1.643, 2.521)$.

between two original local minima. This is different from the result for optimal κ, where the optimal profile is tilted from double wells to nearly one well. Given that the difference between κ and ν is due to m, the time fraction that the system is in the well A, these numerical results are what we expected since the titled potential will also decrease m to increase κ.

The shape derivative calculated from the formula in Theorem 3.2 is validated by comparing with the numerical results from the finite difference scheme. In Figure 4.5,

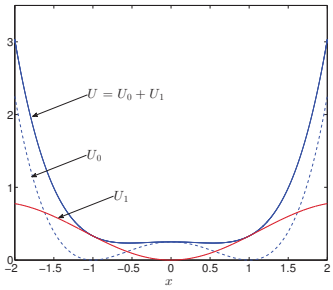

FIG. 4.4. *The optimal profile which maximizes the transition rate* ν. *The optimal solution is* $a = (1.398, 1.571)$.

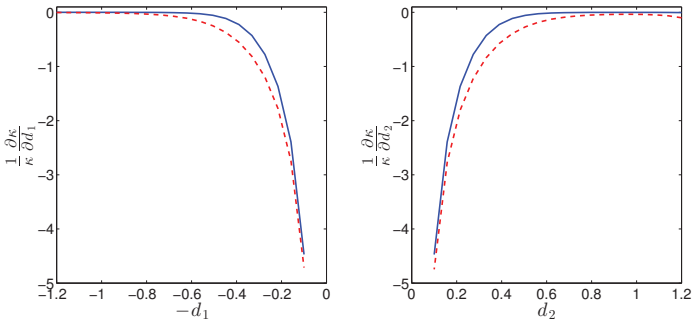

FIG. 4.5. *The relative perturbation* $\delta\kappa/\kappa$ *for* $D = [-d_1, d_2]$ *at* $d_2 = d_1 = d$ *when* d *is varied from* 0.1 *to* 1.2. *The horizontal axes are for* $-d_1$ *and* d_2, *respectively. The left panel shows* $\frac{1}{\kappa}\frac{\partial \kappa}{\partial d_1}$ *and the right panel shows* $\frac{1}{\kappa}\frac{\partial \kappa}{\partial d_2}$. *The solid curves are for* $\varepsilon = 0.03$ *and the dashed are for* $\varepsilon = 0.06$.

we show the shape derivative for the double well U_0 by setting the domain $D = (-d_1, d_2)$, i.e., $A = (-\infty, -d_1]$ and $B = [d_2, +\infty)$. Then the relative changes, $\frac{1}{\kappa}\frac{\partial \kappa}{\partial d_i}$ for $i = 1, 2$ are calculated and plotted. Figure 4.5 shows that the smaller the distance of the domain boundary to the transition state ($x = 0$), the more sensitive the reaction rate with respect to the domain. In addition, the smaller noise amplitude ε shows a less sensitivity. This observation is consistent with the fact in the vanishing noise limit, the reactant set A can be safely defined as the whole well except a boundary layer near separatrix with width $O(\sqrt{\varepsilon})$. If we use the criteria that the size of the relative change $\frac{1}{\kappa}\frac{\partial \kappa}{\partial d_1}$ should be less than 1 to determine the domain A, we could choose $A = (-\infty, -0.25)$ for $\varepsilon = 0.03$ and $A = (-\infty, -0.30)$ for $\varepsilon = 0.06$ to ensure that the change of A would not significantly affect the reaction rate.

4.2. Müller potential. We consider the following two-dimensional energy potential surface

$$U(x,y) = U_0(x,y) + h\sum_{i=1}^{M} \exp\left(-\frac{1}{2l^2}\left((x-x_i)^2 + (y-y_i)^2\right)\right), \tag{4.2}$$

where U_0 is the Müller potential [20] multiplied by 5×10^{-3}. The potential is perturbed by M Gaussian modes with centers (x_i, y_i). We fix $h = 0.25$ and $l = 0.25$ first. Then

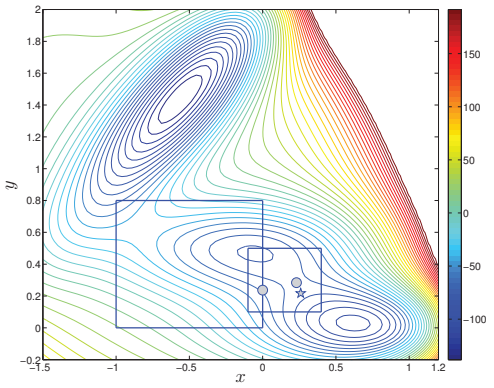

FIG. 4.6. *The Müller potential* (4.2) *is perturbed by adding a mixture of Gaussian modes which are restricted in the two rectangles shown in this figure. The contour plot is for* $U_0(x,y)$. *The star-shaped marker* (★) *is the optimal location of the center for* $M=1$. *The two ball-shaped markers* (●) *are the optimal locations of two centers for* $M=2$.

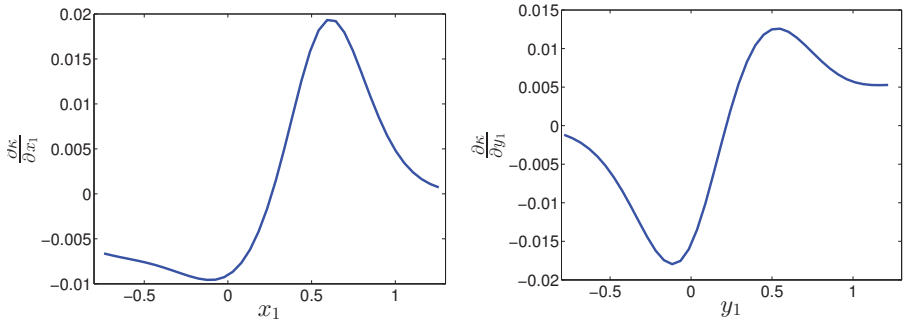

FIG. 4.7. *The partial derivatives of* κ *with respect to* x_1 *at* $y_1=y_1^*=0.217$ *(left) and to* y_1 *at* $x_1=x_1^*=0.260$ *(right) for the case* $M=1$. *The other parameters* $\varepsilon=0.25$ *and* $h=l=0.25$.

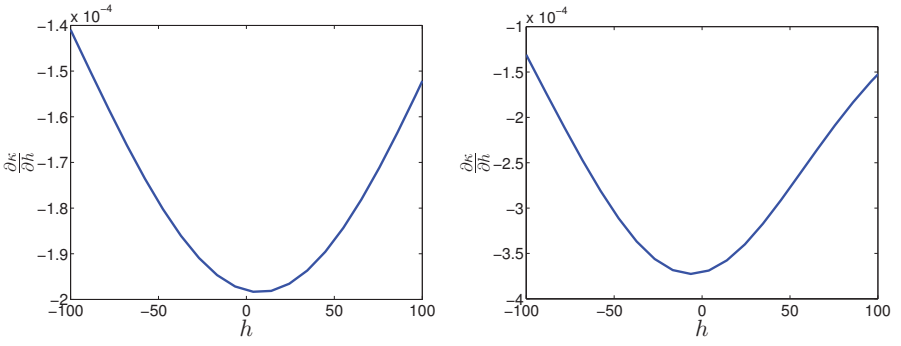

FIG. 4.8. *The partial derivatives of* κ *with respect to the parameter* h *for the case* $M=1$ *when* (x_1,y_1) *is at the optimal center* $(0.260,0.217)$ *(left) and the case* $M=2$ *at the optimal centers* $(0.231,0.284)$ *and* $(0.000,0.237)$ *(right). The other parameters are fixed at* $\varepsilon=0.25$ *and* $l=0.25$.

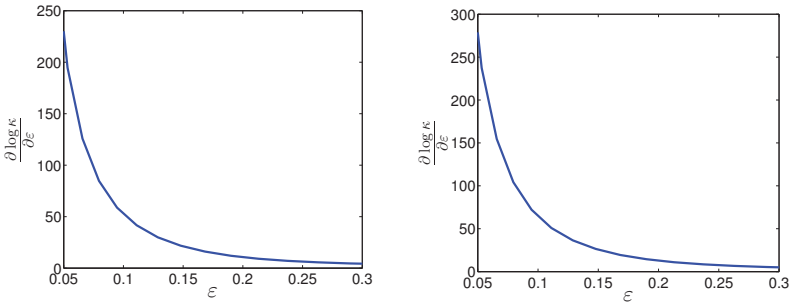

FIG. 4.9. *The partial derivatives of* $\log\kappa$ *with respect to the noise amplitude* ε *for the case* $M=1$ *(left) with the center* $(0.260,0.217)$ *and the case* $M=2$ *(right) with the optimal centers* $(0.231,0.284)$ *and* $(0.000,0.237)$. *The other parameters are fixed at* $h=l=0.25$.

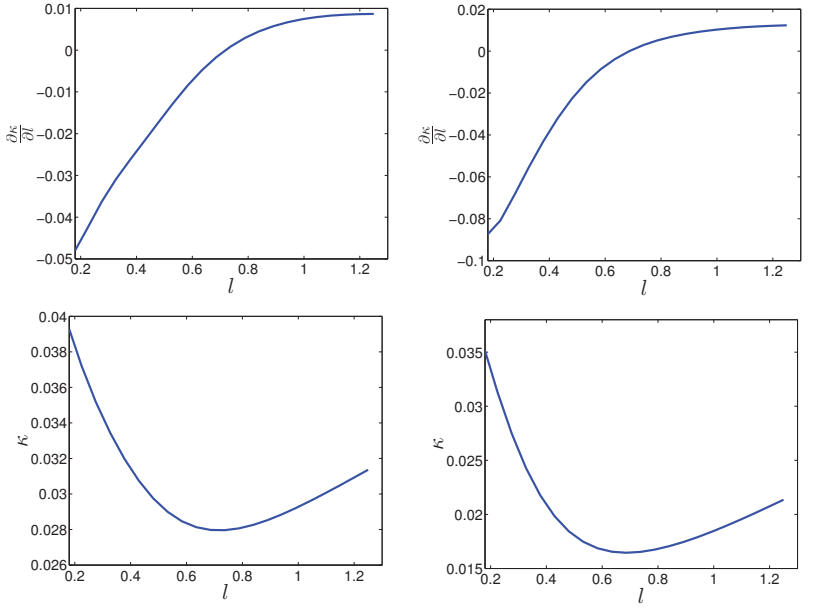

FIG. 4.10. *The values (top row) and the partial derivatives (bottom row) of* κ *with respect to the parameter* l *for the case* $M=1$ *(left column) with the center* $(0.260,0.217)$ *and the case* $M=2$ *(right column) with the optimal centers* $(0.231,0.284)$ *and* $(0.000,0.237)$. *The other parameters are fixed at* $\varepsilon=0.25$ *and* $h=0.25$.

the only free parameters to tune are (x_i,y_i). Our goal is to minimize the A-B reaction rate κ. Here the set A is a ball with radius 0.2 and its center is a local minimum at $(-0.558,1.442)$; the set B is another ball with the same radius and its center is another local minimum at $(0.624,0.028)$.

The numerical results shown below are for $M=1$ and $M=2$. For $M=1$, the two-dimensional parameter $a=(x_1,y_1)$. We restrict a inside a box $B_1=[-0.1,0.4]\times[0.1,0.5]$ where one saddle point lies. For $M=2$, the four-dimensional parameter $a=(x_1,y_1,x_2,y_2)$. The first center (x_1,y_1) is inside the box B_1 and the second center (x_2,y_2) is inside the box $B_2=[-1,0]\times[0,0.8]$. Refer to Figure 4.6 for the two rectangles B_1 and B_2. The noise size ε is 0.25 in this example. The elliptic PDEs for the functions

q and w are solved on a truncated domain which is sufficiently large (at $\varepsilon = 0.25$ used here, the domain is a square with side size 7×7). When there is no perturbation and only the Müller potential U_0 is applied, the reaction rate $\kappa_0 = 0.046$. We find that for the case $M = 1$, the optimal center is unique and is equal to $(0.260, 0.217)$ with the minimum value $\kappa^* = 0.036$; for the case $M = 2$, after trying different initial data for (x_2, y_2) the optimal pair centers are $(0.231, 0.284)$ and $(0.000, 0.237)$ with $\kappa^* = 0.030$. For this example with the given setting of two constraint boxes, our findings of the minimizers are unique for both $M = 1$ and $M = 2$. It was observed that the second optimal center (x_2, y_2) in B_2 is still close to the saddle point on the right, even for initial guesses chosen near the saddle point on the left or the intermediate local minimum in the middle part. This might be due to the higher energy barrier associated with the saddle point on the right.

The optimal center $(x_1^*, y_1^*) = (0.260, 0.217)$ for the case $M = 1$ is confirmed again by Figure 4.7, which shows the partial derivatives of κ for x_1 and y_1, respectively, at $\varepsilon = 0.25$. To better understand the effects of other parameters in this toy model, the partial derivatives of κ or $\log \kappa$ with respect to the three parameters h, ε and l are plotted respectively in Figures 4.8, 4.9 and 4.10 , for the both cases of $M = 1, 2$ with (x_i, y_i) fixed at the optimal center(s). The effects of h and ε on the rate κ are transparent from the figures: the increasing strength h of the Gaussian-peak potential makes the transition more difficult and a higher temperature ε certainly increases the reaction rate. So, the derivatives for h and ε in Figures 4.8 and 4.9 are negative and positive, respectively. Figure 4.10 of the derivative for l, the width of the peak, is quite interesting: there is a minimal value of κ at $l = l^* \approx 0.73$. See the bottom row of this figure, and this critical l^* is also confirmed by the plot of the derivative in the top row. When $l < l^*$, a wider peak gives smaller reaction rate, which shows more wide coverage of the Gaussian mode to fill the saddle region. However, for the sufficiently large $l > l^*$, the Gaussian modes are too flat and global, which might also affect the wells with lower energy to partially generate an opposite effect from the $l < l^*$ case.

5. Conclusion

By using the transition path theory to describe the reaction rate, and using an adjoint method for boundary-value elliptic problems, we have derived the Fréchet derivative of the rates with respect to the potential surface by introducing the adjoint equation related to the first passage time. The shape derivative for the domains defining the product and reactant is also derived and it requires the same auxiliary equation. These formulae of sensitivity analysis are useful to tune the free parameters of the potential surface to optimize the desired rate. One and two dimensional examples have been presented to show the success of this approach. To apply this method to high dimensional problems, the numerical difficulty arises since it is impossible to solve the corresponding PDEs in high dimension. The possible method for improvement could be to consider the random-walk interpretation for these elliptic PDEs, or to use dimension reduction technique by restricting the solution supported on a tube for certain problems. For a general SDE which may lack of potential energy (without detailed balance) and involve the state-dependent diffusion term, the main strategies in the current work can be generalized in principle to such non-gradient systems.

Appendix A. Proof of Theorem 3.2.

Proof. It should be noted that ∇q may have a jump discontinuity on the boundary ∂D. $\nabla q(x)$ is trivially equal to zero from the outside of D from its boundary condition of (2.6). We will refer to ∇q on the boundary ∂D as the limits from inside of D.

For the perturbed quantities, we write $\rho' = \rho[U']$, $q' = q[U', A', B']$, $\kappa' = \kappa[U', A', B]$, $m' = m[U', A', B']$ and $\nu' = \nu[U', A', B']$ to denote the dependency on the coefficient function U' and the domains A', B'. $D' = (A' \cup B')^c$. Then

$$\begin{aligned}\kappa' - \kappa &= \kappa[U', A', B'] - \kappa[U, A, B] = \frac{\nu'}{m'} - \frac{\nu}{m} \\ &= \frac{\nu' - \nu}{m} + \frac{\nu'(m - m')}{m'm} \\ &\simeq \frac{\nu' - \nu}{m} + \frac{\nu}{m^2}(m - m'). \end{aligned} \tag{A.1}$$

The symbol "$\simeq$" means that the difference of two sides is at the order of $o(\widetilde{U})$ or $o(\alpha)$ where $\widetilde{U}$ is the infinitesimal perturbation of the function U and α is the infinitesimal perturbation of the domain D. The discontinuities of $\nabla q'$ on $\partial D'$ and ∇q and ∂D imply that the derivative difference, $\nabla q' - \nabla q$, may not be the first-order infinitesimal perturbation on δD.

To ease the presentation, we summarize the following obvious fact in the lemma below, which will be used repeatedly later.

LEMMA A.1. *Suppose that $\omega = 0$ on D^c and $\omega' = 0$ on D'^c. Also assume the difference $\omega' - \omega$ is of the order $o(1)$ on $D \cap D'$. Then*

$$\int \omega' \widetilde{U} \, \mathrm{d}x = \int_{D'} \omega' \widetilde{U} \, \mathrm{d}x \simeq \int_{D' \cap D} \omega' \widetilde{U} \, \mathrm{d}x \simeq \int_{D \cap D'} \omega \widetilde{U} \, \mathrm{d}x \simeq \int_{D} \omega \widetilde{U} \, \mathrm{d}x = \int \omega \widetilde{U} \, \mathrm{d}x, \tag{A.2}$$

where $\widetilde{U}$ is defined as above.

Now we first compute

$$\begin{aligned}\nu' - \nu &= \nu[U', A', B'] - \nu[U, A, B] \\ &= (\nu[U', A', B'] - \nu[U, A', B']) + (\nu[U, A', B'] - \nu[U, A, B]). \end{aligned} \tag{A.3}$$

For the first part on the above RHS, by equations (3.7) and (A.2), we have that

$$\begin{aligned}&\nu[U', A', B'] - \nu[U, A', B'] \\ &\simeq \int \frac{\delta \nu}{\delta U}[U, A', B'] \widetilde{U} \, \mathrm{d}x = \int \rho'(\nu'/\varepsilon - |\nabla q'|^2) \widetilde{U} \, \mathrm{d}x \\ &\simeq \int \rho(\nu/\varepsilon - |\nabla q|^2) \widetilde{U} \, \mathrm{d}x. \end{aligned} \tag{A.4}$$

For the term $\nu[U, A', B'] - \nu[U, A, B]$ involving shape derivative, we carry out the following steps. Set $q^* = q[U, A', B']$ and $\hat{q} = q^* - q$. Define $q^\dagger$ on $\delta D = (D \setminus D') \cup (D' \setminus D)$ as follows: $q^\dagger$ is equal to q on $D \setminus D'$ and q^* on $D' \setminus D$. Then,

$$\begin{aligned}\nu[U, A', B'] - \nu[U, A, B] &= \int_{D'} \varepsilon \rho |\nabla q^*|^2 \, \mathrm{d}x - \int_{D} \varepsilon \rho |\nabla q|^2 \, \mathrm{d}x \\ &= \int_{\delta D} \varepsilon \rho |\nabla q^\dagger|^2 \, \mathrm{d}x + \int_{D \cap D'} \varepsilon \rho \left(|\nabla q^*|^2 - |\nabla q|^2 \right) \mathrm{d}x, \end{aligned} \tag{A.5}$$

where the integrand on δD is taken with a plus sign over $D' \setminus D$ (where $\alpha(x) > 0$) and a minus sign over $D \setminus D'$ (where $\alpha(x) < 0$). Refer to Figure 3.1 for illustration. Since α is small, from the mean value theorem, we obtain

$$\int_{\delta D} \rho |\nabla q^\dagger|^2 \, \mathrm{d}x = \int_{D' \setminus D} \rho |\nabla q^*|^2 \, \mathrm{d}x - \int_{D \setminus D'} \rho |\nabla q|^2 \, \mathrm{d}x$$

$$\simeq \int_{(\partial D)\cap D'} \alpha\rho|\nabla q^*|^2 \,\mathrm{d}\sigma_{\partial D} + \int_{(\partial D)\setminus D'} \alpha\rho|\nabla q|^2 \,\mathrm{d}\sigma_{\partial D}$$
$$\simeq \int_{\partial D} \alpha\rho|\nabla q|^2 \,\mathrm{d}\sigma_{\partial D}. \tag{A.6}$$

For the second term of RHS in (A.5), since ∇q^* and ∇q, are continuous in the closure of $D\cap D'$, we then have

$$\int_{D\cap D'} \rho\left(|\nabla q^*|^2 - |\nabla q|^2\right) \mathrm{d}x \simeq 2\int_{D\cap D'} \rho\nabla q\cdot\nabla\hat{q}\,\mathrm{d}x$$
$$= 2\int_{D\cap D'} \nabla\cdot(\hat{q}\rho\nabla q)\,\mathrm{d}x = 2\int_{\partial(D\cap D')} \hat{q}\rho\nabla q\cdot n\,\mathrm{d}\sigma_{\partial D}. \tag{A.7}$$

where the fact $\nabla\cdot(\rho\nabla q)=\varepsilon^{-1}\rho\mathcal{L}q=0$ on D is used for the first equality on the second line. $n(x)$ is the outer normal unit vector at $x\in\partial D$. The boundary $\partial(D\cap D')$ can be decomposed into two pieces $(\partial D)\cap D'$ and $(\partial D')\cap D$. Note that for $x\in\partial D$, $x'=x+\alpha(x)n(x)\in\partial D'$, and $q^*(x')$ is equal to $q(x)$, which is either constant zero or constant one. Thus for $x\in(\partial D)\cap D'$, it holds that

$$\hat{q}(x) = q^*(x) - q(x) = q^*(x) - q^*(x')$$
$$\simeq -\alpha(x)\nabla q^*(x)\cdot n(x)$$
$$\simeq -\alpha(x)\nabla q(x)\cdot n(x). \tag{A.8}$$

In addition, for $x\in(\partial D)\setminus D'$, then q^* has the same boundary value at both x and x'. Then

$$\hat{q}(x) = q^*(x) - q(x) = q^*(x) - q^*(x') = 0. \tag{A.9}$$

It then follows that by noting $n\parallel\nabla q$ on ∂D,

$$2\int_{(\partial D)\cap D'} \hat{q}\rho\nabla q\cdot n\,\mathrm{d}\sigma_{\partial D} \simeq -2\int_{(\partial D)\cap D'} \alpha\rho|\nabla q|^2\,\mathrm{d}\sigma_{\partial D}. \tag{A.10}$$

Similarly, for $x'\in(\partial D')\cap D$, then $x\in(\partial D)\setminus D'$ and

$$\hat{q}(x') = q^*(x') - q(x') = q(x) - q(x') \simeq -\alpha(x)\nabla q(x)\cdot n(x).$$

It follows too that

$$2\int_{(\partial D')\cap D} \hat{q}\rho\nabla q\cdot n\,\mathrm{d}\sigma_{\partial D'} \simeq -2\int_{(\partial D)\setminus D'} \alpha\rho|\nabla q|^2\,\mathrm{d}\sigma_{\partial D}. \tag{A.11}$$

Consequently, from equations (A.5)–(A.11) we have obtained that

$$\nu[U,A',B'] - \nu[U,A,B] \simeq -\int_{\partial D} \varepsilon\alpha\rho|\nabla q|^2\,\mathrm{d}\sigma_{\partial D},$$

and together with equations (A.3) and (A.4), we have

$$\nu' - \nu \simeq \int_D \rho\left(\nu/\varepsilon - |\nabla q|^2\right)\widetilde{U}\,\mathrm{d}x - \int_{\partial D} \varepsilon\alpha\rho|\nabla q|^2\,\mathrm{d}\sigma_{\partial D}. \tag{A.12}$$

In the remaining part of proof, we calculate $m-m'$.

$$\begin{aligned} m-m' &= \int \rho' q' \,\mathrm{d}x - \int \rho q \,\mathrm{d}x \\ &= \left(\int \rho' q' \,\mathrm{d}x - \int \rho q^* \,\mathrm{d}x\right) + \left(\int \rho q^* \,\mathrm{d}x - \int \rho q \,\mathrm{d}x\right). \end{aligned} \tag{A.13}$$

The first term of RHS above is determined by equations (3.8) and (A.2):

$$\begin{aligned} \int \rho' q' \,\mathrm{d}x - \int \rho q^* \,\mathrm{d}x &\simeq \int \left(-\frac{\rho}{\varepsilon}(q^* + m[U,A',B'] - 1) + \rho \nabla q^* \cdot \nabla w[U,A',B']\right) \widetilde{U} \,\mathrm{d}x \\ &\simeq \int \left(-\frac{\rho}{\varepsilon}(q + m - 1) + \rho \nabla q \cdot \nabla w\right) \widetilde{U} \,\mathrm{d}x, \end{aligned} \tag{A.14}$$

where $w = w[U,A,B]$ is defined by equations (3.9). For the second term of RHS in (A.13), using the boundary conditions of q^* and q, we decompose the integration as follows

$$\begin{aligned} &\int \rho q^* \,\mathrm{d}x - \int \rho q \,\mathrm{d}x = \int_{D'} \rho q^* \,\mathrm{d}x + \int_{B'} \rho \,\mathrm{d}x - \int_D \rho q \,\mathrm{d}x - \int_B \rho \,\mathrm{d}x \\ &= \left(\int_{D'} \rho q^* \,\mathrm{d}x - \int_D \rho q^* \,\mathrm{d}x\right) + \left(\int_D \rho q^* \,\mathrm{d}x - \int_D \rho q \,\mathrm{d}x\right) + \int_{B'} \rho \,\mathrm{d}x - \int_B \rho \,\mathrm{d}x \\ &= \int_{\delta D} \rho q^* \,\mathrm{d}x + \int_D \rho \hat{q} \,\mathrm{d}x + \int_{\delta B} \rho \,\mathrm{d}x. \end{aligned}$$

Applying the mean value theorem again, we have

$$\begin{gathered} \int_{\delta D} \rho q^* \,\mathrm{d}x \simeq \int_{\delta D} \rho q \,\mathrm{d}x \simeq \int_{\partial D} \alpha \rho q \,\mathrm{d}\sigma_{\partial D} = \int_{\partial B} \alpha \rho \,\mathrm{d}\sigma_{\partial B}, \\ \int_{\delta B} \rho \,\mathrm{d}x \simeq \int_{\partial B} -\alpha \rho \,\mathrm{d}\sigma_{\partial B}. \end{gathered}$$

Thus it follows that

$$\int \rho q^* \,\mathrm{d}x - \int \rho q \,\mathrm{d}x \simeq \int_D \rho \hat{q} \,\mathrm{d}x. \tag{A.15}$$

Recall that by equations (2.4) and (3.9), $\rho = \rho \times (-\mathcal{L}w) = -\varepsilon \nabla \cdot (\rho \nabla w)$ in D and $w = 0$ on ∂D. Therefore

$$\int_D \rho \hat{q} \,\mathrm{d}x = -\int_D \varepsilon \hat{q} \nabla \cdot (\rho \nabla w) \,\mathrm{d}x = -\int_{\partial D} \varepsilon \rho \hat{q} \nabla w \cdot n \,\mathrm{d}\sigma_{\partial D} + \int_D \varepsilon \rho \nabla w \cdot \nabla \hat{q} \,\mathrm{d}x. \tag{A.16}$$

From equations (A.8) and (A.9) and the simple fact that $n \parallel \nabla q$ on ∂D,

$$-\int_{\partial D} \rho \hat{q} \nabla w \cdot n \,\mathrm{d}\sigma_{\partial D} \simeq \int_{(\partial D) \cap D'} \alpha \rho (\nabla q \cdot n)(\nabla w \cdot n) \,\mathrm{d}\sigma_{\partial D} = \int_{(\partial D) \cap D'} \alpha \rho \nabla q \cdot \nabla w \,\mathrm{d}\sigma_{\partial D}. \tag{A.17}$$

For the second term of RHS in (A.16), since $\nabla \hat{q}$ is not continuous on D, we cannot apply integration by parts directly. So we first split it into two parts

$$\int_{D \cap D'} \rho \nabla w \cdot \nabla \hat{q} \,\mathrm{d}x + \int_{D \setminus D'} \rho \nabla w \cdot \nabla \hat{q} \,\mathrm{d}x.$$

Observing that $\nabla\hat{q} = -\nabla q$ on $D \setminus D'$ and applying the mean value theorem again, we obtain

$$\int_{D\setminus D'} \rho\nabla w\cdot\nabla\hat{q}\,\mathrm{d}x = -\int_{D\setminus D'} \rho\nabla w\cdot\nabla q\,\mathrm{d}x \simeq \int_{(\partial D)\setminus D'} \alpha\rho\nabla q\cdot\nabla w\,\mathrm{d}\sigma_{\partial D}. \tag{A.18}$$

Applying integration by parts and using $\mathcal{L}\hat{q} = 0$ on $D \cap D'$ and $w = 0$ on ∂D, we are led to

$$\begin{aligned}
\int_{D\cap D'} \rho\nabla w\cdot\nabla\hat{q}\,\mathrm{d}x &= \int_{\partial(D\cap D')} w\rho\nabla\hat{q}\cdot n\,\mathrm{d}\sigma_{\partial(D\cap D')} - \int_{D\cap D'} w\nabla\cdot(\rho\nabla\hat{q})\,\mathrm{d}x \\
&= \left(\int_{(\partial D)\cap D'} + \int_{(\partial D')\cap D}\right) w\rho\nabla\hat{q}\cdot n\,\mathrm{d}\sigma_{\partial(D\cap D')} - \int_{D\cap D'} w\rho\varepsilon^{-1}\mathcal{L}\hat{q}\,\mathrm{d}x \\
&= \int_{(\partial D')\cap D} w\rho\nabla\hat{q}\cdot n\,\mathrm{d}\sigma_{\partial D'} \\
&\simeq 0.
\end{aligned} \tag{A.19}$$

The last equality above uses the fact that w and $\nabla\hat{q}$ are both small on $\partial D' \cap D$. Consequently, we deduce from equations (A.15)–(A.19) that

$$\int \rho q^*\,\mathrm{d}x - \int \rho q\,\mathrm{d}x \simeq \int_{\partial D} \varepsilon\alpha\rho\nabla q\cdot\nabla w\,\mathrm{d}\sigma_{\partial D}.$$

This together with equations (A.13) and (A.14) yields

$$m - m' \simeq \int \left(-\frac{\rho}{\varepsilon}(q+m-1) + \rho\nabla q\cdot\nabla w\right)\widetilde{U}\,\mathrm{d}x + \int_{\partial D} \varepsilon\alpha\rho\nabla q\cdot\nabla w\,\mathrm{d}\sigma_{\partial D}, \tag{A.20}$$

and eventually from equations (A.1), (A.12), and (A.20) we conclude

$$\begin{aligned}
\kappa' - \kappa &= \kappa[U', A', B'] - \kappa[U, A, B] \\
&\simeq \int \left(-\frac{\rho}{m}|\nabla q|^2 + \rho\kappa/\varepsilon\right)\widetilde{U}\,\mathrm{d}x - \int_{\partial D} \varepsilon\alpha\frac{\rho}{m}|\nabla q|^2\,\mathrm{d}\sigma_{\partial D} \\
&\quad + \frac{\kappa}{m}\left(-\int \rho(q+m-1)/\varepsilon\tilde{U}\,\mathrm{d}x + \int_D \rho\nabla q\cdot\nabla w\widetilde{U}\,\mathrm{d}x + \int_{\partial D} \varepsilon\alpha\rho\nabla q\cdot\nabla w\,\mathrm{d}\sigma_{\partial D}\right) \\
&= \int \frac{\rho}{m}\left[-|\nabla q|^2 - \kappa(q-1)/\varepsilon + \kappa\nabla q\cdot\nabla w\right]\widetilde{U}\,\mathrm{d}x \\
&\qquad + \int_{\partial D} \varepsilon\alpha\frac{\rho}{m}\left(-|\nabla q|^2 + \kappa\nabla q\cdot\nabla w\right)\mathrm{d}\sigma_{\partial D}.
\end{aligned}$$

∎

REFERENCES

[1] H.A. Karmers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica, 7:284–304, 1940.

[2] P. Hänggi, P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after Kramers*, Rev. Mod. Phys., 62(2):251–341, 1990.

[3] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library, North Holland, Third Edition, 2007.

[4] H. Eyring, *The activated complex and the absolute rate of chemical reactions*, Chem. Rev., 17:65–77, 1935.

[5] J. Horiuti, *On the statistical mechanical treatment of the absolute rate of chemical reaction*, Bull. Chem. Soc. Jpn., 13:210–216, 1938.
[6] E. Wigner, *The transition state method*, Trans. Faraday Soc., 34:29–41, 1938.
[7] W. E and E. Vanden-Eijnden, *Towards a theory of transition paths*, J. Stat. Phys., 123:503–523, 2006.
[8] E. Vanden-Eijnden, *Transition path theory*, in Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology, M. Ferrario, G. Ciccotti, and K. Binder (eds.), Springer, 1:453–493, 2006.
[9] P. Metzner, C. Schüttle, and E. Vanden-Eijnden, *Illustration of transition path theory on a collection of simple examples*, J. Chem. Phys., 125(8):084110, 2006.
[10] P. Metzner, C. Schütte, and E. Vanden-Eijnden, *Transition path theory for Markov jump processes*, Multiscale Model. Simul., 7:1192–1219, January 2009.
[11] W. E and E. Vanden-Eijnden, *Transition path theory and path-finding algorithms for the study of rare events*, Annual Review of Physical Chemistry, 61:391–420, 2010.
[12] C. Dellago, P.G. Bolhuis, and P.L. Geissler, *Transition path sampling*, Adv. Chem. Phys., 123:1–78, 2002.
[13] A.-S. Sznitman, *Brownian Motion, Obstacles and Random Media*, Springer-Verlag, 1998.
[14] J. Doob, *Classical Potential Theory and Its Probabilistic Counterpart*, Springer-Verlag, 2001.
[15] J. Thuburn, *Climate sensitivities via a fokkerCplanck adjoint approach*, Q. J. R. Meteorol. Soc., 131:73–92, 2005.
[16] O. Pironneau, *Optimal Shape Design of Elliptic Systmes*, Springer Series in Computational Physics, Springer-Verlag, 1984.
[17] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stoch. Stoch. Reports, 76:481–508, 2004.
[18] P. Dupuis, K. Spiliopoulos, and X. Zhou, *Escape from an attractor: importance sampling and rest points: I*, Ann. Appl. Prob., 25(5):2909–2958, 2015.
[19] 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.
[20] K. Müller, *Reaction paths on multidimensional energy hypersurfaces*, Angewandte Chemie International Edition in English, 19(1):1–13, 1980.