

Stability, Accuracy and Cost of NEB and String Methods

Congming Jin*

*Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Beijing 100080, P.O. Box 2719, China.*

Received 20 December 2006; Accepted (in revised version) 25 April 2007

Available online 2 August 2007

Abstract. In this paper, the zero-temperature string method and the nudged elastic band method for computing the transition paths and transition rates between metastable states are investigated. The stability, accuracy as well as computational cost of the two methods are discussed. The results are verified by numerical experiments.

AMS subject classifications: 92-08, 92E20, 65M12

Key words: Transition path, transition state, string method, nudged elastic band method, stability, accuracy and cost.

1 Introduction

The string method [4,5] and the nudged elastic band (NEB) method [9] have been widely used in the study of transition paths and transition rates between metastable states. Both methods have been successfully applied to continuous models, empirical potential models and first-principles calculations, see [4, 10, 18–20] for the string method, and [2, 8, 12, 13, 21–23] for the NEB method.

In this paper, we focus on the zero-temperature string (ZTS) method and the NEB method. The ZTS method and the NEB method have some similarities. First, both methods evolve a chain of images of the system between the initial state and the final state. Second, the potential forces are decomposed into components normal and tangential to the path in both methods. Third, both methods minimize the energy in the plane normal to the path at each image. On the other hand, the two methods are different in several aspects. In the NEB method, extra spring interaction between the adjacent images is added to ensure continuity of the path. The systems move in a force field which is a combination of the normal component of the potential force and the tangential component of the

*Corresponding author. *Email address:* jincm@1sec.cc.ac.cn (C. Jin)

spring force. In the ZTS method, the images are points on a string, i.e., a smooth curve with intrinsic parametrization such as arc length or energy-weighted arc length, which connects two metastable states. The string evolves to the minimal energy path (MEP) under the normal component of the potential force subject to some constraint.

We give a detailed theoretical analysis of the ZTS and the NEB methods, including the stability, accuracy and computational cost. An adaptive time step is obtained from the stability conditions for each method. Then the computational cost is estimated. A good choice for the elastic constant is obtained for the NEB method. These choices of the parameters make the methods more efficient or more accurate. As for the accuracy of the transition path, both methods have first order accuracy under L_2 -norm and L_∞ -norm. Two techniques to improve the accuracy at the transition state are discussed.

The rest of this paper is organized as follows: In Section 2, the ZTS method and the NEB method are briefly reviewed. In Section 3, the stability conditions of the ZTS method and the NEB method are provided. We analyze the accuracy in Section 4. Estimates of the computational cost of the two methods are presented in Section 5. We conclude the paper in Section 6.

2 ZTS method and NEB method

In this section, we briefly review the ZTS method and the NEB method (see [4,9] for more details).

2.1 ZTS method

Consider the example of a system modelled by the following stochastic equation

$$\dot{X}^\varepsilon = -\nabla V(X^\varepsilon) + \sqrt{2\varepsilon}\dot{W}, \quad (2.1)$$

where $V(X)$ is the potential energy of the system, \dot{W} is a white noise, and ε is a parameter representing the strength of the white noise. Suppose the potential energy has two minima A and B . Let φ be a smooth curve, i.e., a string, connecting the two minima of the potential energy, A and B . By definition, φ is a MEP if

$$0 = (\nabla V(\varphi))^\perp, \quad (2.2)$$

where

$$(\nabla V(\varphi))^\perp = \nabla V(\varphi) - (\nabla V(\varphi) \cdot \hat{\tau})\hat{\tau},$$

with $\hat{\tau} = \varphi_\alpha / |\varphi_\alpha|$ being the unit tangent vector along φ , and α the intrinsic parameter of the string. Equivalently, the MEP φ is a curve that minimizes V in the hyperplane normal to itself. One way of finding solutions of Eq. (2.2) is to follow the dynamics determined by

$$\varphi_t = -(\nabla V(\varphi))^\perp + \gamma\hat{\tau}, \quad (2.3)$$

where the scalar field $\gamma \equiv \gamma(\alpha, t)$ is a Lagrange multiplier determined by the parametrization of the string. The simplest choice is to parameterize φ by normalized arc length so that $\alpha = 0$ at A and $\alpha = 1$ at B . In this case, Eq. (2.3) is subject to the constraint

$$(|\varphi_\alpha|)_\alpha = 0, \quad (2.4)$$

which determines γ . Other parametrizations can also be straightforwardly implemented by modifying the constraint (2.4). For instance, a parametrization by energy-weighted arc length which increases the resolution at the transition states is achieved using the constraint

$$[f(V(\varphi))|\varphi_\alpha]_\alpha = 0,$$

where $f(z)$ is some suitable monitor function satisfying $f'(z) > 0$.

2.2 NEB method

In the NEB method, a set of images $\{\varphi_i\}_{i=0}^N$ of the system are connected by springs to simulate a path. Here φ_0 and φ_N are two known stable states A and B , respectively. The force acting on image i ($i = 1, 2, \dots, N-1$) is

$$F_i = -(\nabla V(\varphi_i))^\perp + (\tilde{F}_i)^\parallel, \quad (2.5)$$

where $\tilde{F}_i = k(\varphi_{i+1} - 2\varphi_i + \varphi_{i-1})$ is the spring force, k is the elastic constant of the springs, and $(\tilde{F}_i)^\parallel = (\tilde{F}_i, \hat{\tau}_i) \hat{\tau}_i$ is the component of \tilde{F}_i parallel to the path. The images move under the forces defined by Eq. (2.5).

In a continuum formulation, the NEB method can be written as

$$\varphi_t = -(\nabla V(\varphi))^\perp + \tilde{k}(\varphi_{\alpha\alpha})^\parallel, \quad (2.6)$$

where the path is parameterized by the arc length parameter α and

$$\tilde{k} = \lim_{\Delta\alpha \rightarrow 0} k(\Delta\alpha)^2. \quad (2.7)$$

For simplicity of notation, we will use k for this \tilde{k} in the following discussion.

3 Stability

Usually, Eqs. (2.3) and (2.6) are solved by finite difference schemes. In this section, we consider the stability of these two equations. Based on the stability conditions, we present the stable finite difference schemes and adaptive time steps.

Since Eqs. (2.3) and (2.6) are nonlinear partial differential equations, we linearize them and present the stability conditions for the linearized equations. Consider a potential $V(x)$ where x is in n -dimensional space. Let φ be a path parameterized by α . The MEP is denoted by φ_0 .

3.1 The stability of the ZTS method

The ZTS method can be written as

$$\begin{cases} \frac{\partial \varphi}{\partial t} = -(\nabla V)^\perp = -\nabla V + (\nabla V, \hat{\tau})\hat{\tau}, \\ \frac{\partial}{\partial \alpha} \left| \frac{\partial \varphi}{\partial \alpha} \right|^2 = 0. \end{cases} \quad (3.1)$$

The second equation of (3.1) is used to enforce the constraint. The unit tangent of the path φ is

$$\hat{\tau} = \frac{\partial \varphi}{\partial \alpha} / \left| \frac{\partial \varphi}{\partial \alpha} \right|.$$

Denote the unit tangent of the MEP by

$$\hat{e}_1 \equiv \hat{\tau}_0 = \frac{\partial \varphi_0}{\partial \alpha} / \left| \frac{\partial \varphi_0}{\partial \alpha} \right|.$$

Let $\varphi = \varphi_0 + \delta\varphi$, where $\delta\varphi$ is a small variation of φ_0 . We look for the equation of $\delta\varphi$. From Taylor expansion, we obtain

$$\left| \frac{\partial \varphi}{\partial \alpha} \right| = \left| \frac{\partial \varphi_0}{\partial \alpha} + \frac{\partial \delta\varphi}{\partial \alpha} \right| = \left| \frac{\partial \varphi_0}{\partial \alpha} \right| + \frac{\partial \delta\varphi_1}{\partial \alpha} + \mathcal{O}(|\delta\varphi|^2), \quad (3.2)$$

where $\delta\varphi_1$ is the length of the projection of $\delta\varphi$ on \hat{e}_1 , i.e., $\delta\varphi_1 = (\delta\varphi, \hat{e}_1)$. Substituting (3.2) in the second equation of (3.1), we obtain the linearized equation of the constrained equation

$$\frac{\partial^2 \delta\varphi_1}{\partial \alpha^2} = 0. \quad (3.3)$$

Now consider the first equation of (3.1). Note that the tangent can be written as

$$\hat{\tau} = \frac{\partial}{\partial \alpha}(\varphi_0 + \delta\varphi) \left| \frac{\partial}{\partial \alpha}(\varphi_0 + \delta\varphi) \right|^{-1}. \quad (3.4)$$

Hence by some simple calculations, we obtain

$$\begin{aligned} \hat{\tau} &= \left(\frac{\partial \varphi_0}{\partial \alpha} + \frac{\partial}{\partial \alpha} \delta\varphi \right) \left| \frac{\partial \varphi_0}{\partial \alpha} + \frac{\partial}{\partial \alpha} \delta\varphi \right|^{-1} = \left(\hat{e}_1 + \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta\varphi}{\partial \alpha} \right) \left| \hat{e}_1 + \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta\varphi}{\partial \alpha} \right|^{-1} \\ &= \hat{e}_1 - \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta\varphi_1}{\partial \alpha} \hat{e}_1 + \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta\varphi}{\partial \alpha}, \end{aligned} \quad (3.5)$$

and

$$\delta \hat{\tau} = \hat{\tau} - \hat{\tau}_0 = - \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta \varphi_1}{\partial \alpha} \hat{e}_1 + \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} \frac{\partial \delta \varphi}{\partial \alpha}. \tag{3.6}$$

For simplicity, we will use the tensor product. Let $a = (a_1, a_2, \dots, a_n) \in R^n$ and $b = (b_1, b_2, \dots, b_n) \in R^n$ be two n -dimensional vectors, and the *tensor product* of a and b be a matrix $a \otimes b$ whose elements are

$$(a \otimes b)_{ij} = a_i * b_j, \quad \forall i, j = 1, 2, \dots, n. \tag{3.7}$$

Suppose $c = (c_1, c_2, \dots, c_n)$ is another n -dimensional vector. It is easy to verify that

$$(a, b)c = (c \otimes a)b. \tag{3.8}$$

Since $\delta \varphi_1 = (\hat{e}_1, \delta \varphi)$, the variation of the unit tangent can be written as

$$\delta \hat{\tau} = \hat{\tau} - \hat{\tau}_0 = \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} (I - \hat{e}_1 \otimes \hat{e}_1) \frac{\partial \delta \varphi}{\partial \alpha}. \tag{3.9}$$

Using the Taylor expansion, we obtain

$$\nabla V(\varphi_0 + \delta \varphi) = \nabla V(\varphi_0) + \nabla^2 V(\varphi_0) \delta \varphi, \tag{3.10}$$

$$(\nabla V(\varphi_0 + \delta \varphi), \hat{\tau}) \hat{\tau} = (\nabla V(\varphi_0) + \nabla^2 V(\varphi_0) \delta \varphi, \hat{e}_1 + \delta \hat{\tau}) (\hat{e}_1 + \delta \hat{\tau}). \tag{3.11}$$

We have

$$\nabla V(\varphi_0) = (\nabla V(\varphi_0), \hat{e}_1) \hat{e}_1, \tag{3.12}$$

where the fact that

$$(\nabla V(\varphi_0))^\perp = 0 \tag{3.13}$$

is used. Substituting (3.10) and (3.11) in the first equation of (3.1) and using (3.9) and (3.12), we obtain the linearization of the first equation of (3.1) as follows

$$\frac{\partial \delta \varphi}{\partial t} = -\nabla^2 V(\varphi_0) \delta \varphi + (\nabla^2 V(\varphi_0) \delta \varphi, \hat{e}_1) \hat{e}_1 + (\nabla V(\varphi_0), \hat{e}_1) \delta \hat{\tau} + (\nabla V(\varphi_0), \delta \hat{\tau}) \hat{e}_1,$$

or

$$\frac{\partial \delta \varphi}{\partial t} = (\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) \delta \varphi + (\nabla V(\varphi_0), \hat{e}_1) \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} (I - \hat{e}_1 \otimes \hat{e}_1) \frac{\partial \delta \varphi}{\partial \alpha},$$

where the fact that

$$(\hat{e}_1 \otimes \hat{e}_1)^2 = \hat{e}_1 \otimes \hat{e}_1 \tag{3.14}$$

is used. Thus, the linearized equations of (3.1) can be written as

$$\begin{cases} \frac{\partial \delta \varphi}{\partial t} = (\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) \delta \varphi + (\nabla V(\varphi_0), \hat{e}_1) \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} (I - \hat{e}_1 \otimes \hat{e}_1) \frac{\partial \delta \varphi}{\partial \alpha}, \\ \frac{\partial^2 \delta \varphi_1}{\partial \alpha^2} = 0. \end{cases} \quad (3.15)$$

For the zeroth order term including $\delta \varphi$ (where no derivative in α is involved), the general stability condition is stated as follows:

$$\max\{\|(\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0)\|\} \Delta t \leq \text{const}. \quad (3.16)$$

In principle, the constant can be any value if we want to compute on a finite time interval. However, we are interested in path convergent to the steady state, so we shall require that Δt be in the stability region of the time-stepping scheme, i.e.,

$$\max\{\|(\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) - I\|\} \Delta t < 1. \quad (3.17)$$

Considering the first order term involving $\frac{\partial \delta \varphi}{\partial \alpha}$, we should use upwind scheme. It can be verified that $B = I - \hat{e}_1 \otimes \hat{e}_1$ satisfies

$$B = B^2.$$

Thus if λ is an eigenvalue of B then λ^2 is also an eigenvalue of B . So the eigenvalue of B is either 1 or 0. When we use an upwind scheme, the stability condition for the first order term is

$$\max\left\{\left|\frac{(\nabla V(\varphi_0), \hat{e}_1)}{\left|\frac{\partial \varphi_0}{\partial \alpha}\right| \Delta \alpha}\right|\right\} \Delta t \leq 1. \quad (3.18)$$

Consequently, the stability condition for the ZTS method is

$$\max\left\{\|(\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) - I\|, \left|\frac{(\nabla V(\varphi_0), \hat{e}_1)}{\left|\frac{\partial \varphi_0}{\partial \alpha}\right| \Delta \alpha}\right|\right\} \Delta t \leq 1. \quad (3.19)$$

3.2 Stability of the NEB method

The equation for the NEB method is written as

$$\varphi_t = -(\nabla V(\varphi))^\perp + k \left(\frac{\partial^2 \varphi}{\partial \alpha^2} \right)^\parallel, \quad (3.20)$$

where α is the arc length parameter.

Let $\varphi = \varphi_0 + \delta\varphi$, where $\delta\varphi$ is a small variation of φ_0 . Similar to the analysis of the ZTS method, the spring force can be written as

$$\begin{aligned} \left(\frac{\partial^2 \varphi}{\partial \alpha^2}\right)^\parallel &= \left(\frac{\partial^2 \varphi}{\partial \alpha^2}, \hat{\tau}\right) \hat{\tau} \\ &= \left(\frac{\partial^2(\varphi_0 + \delta\varphi)}{\partial \alpha^2}, \hat{e}_1 + \delta\hat{\tau}\right) (\hat{e}_1 + \delta\hat{\tau}) \\ &= \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1\right) \hat{e}_1 + \left(\frac{\partial^2 \delta\varphi}{\partial \alpha^2}, \hat{e}_1\right) \hat{e}_1 + \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1\right) \delta\hat{\tau} + \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \delta\hat{\tau}\right) \hat{e}_1. \end{aligned} \quad (3.21)$$

Here $\hat{\tau}$, $\delta\hat{\tau}$, and \hat{e}_1 are defined in Section 3.1. Substituting (3.10), (3.11) and (3.21) in (3.20), we are led to the linearized equation

$$\begin{aligned} \frac{\partial \delta\varphi}{\partial t} &= k(\hat{e}_1 \otimes \hat{e}_1) \frac{\partial^2 \delta\varphi}{\partial \alpha^2} + (\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) \delta\varphi \\ &\quad + \left[(\nabla V(\varphi_0), \hat{e}_1) I + k \left(\hat{e}_1, \frac{\partial^2 \varphi_0}{\partial \alpha^2} \right) I + k \left(\hat{e}_1 \otimes \frac{\partial^2 \varphi_0}{\partial \alpha^2} \right) \right] (I - \hat{e}_1 \otimes \hat{e}_1) \frac{\partial \delta\varphi}{\partial \alpha} \\ &= I_1 + I_2 + I_3, \end{aligned} \quad (3.22)$$

where the leading order terms cancel out since φ_0 is a MEP.

Now we are going to analyze these three terms, respectively. It is seen the first term

$$I_1 = k(\hat{e}_1 \otimes \hat{e}_1) \frac{\partial^2 \delta\varphi}{\partial \alpha^2} \quad (3.23)$$

looks like a heat equation term. As a result, the general stability condition should be

$$\max \left\{ \frac{k}{(\Delta\alpha)^2} \right\} \Delta t < \frac{1}{2}. \quad (3.24)$$

As for the second term

$$I_2 = (\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) \delta\varphi, \quad (3.25)$$

which is of zeroth order, the general stability condition is

$$\max \left\{ |(\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) - I| \right\} \Delta t \leq 1. \quad (3.26)$$

Finally it is easy to get that

$$\begin{aligned} I_3 &= \left\{ (\nabla V(\varphi_0), \hat{e}_1) I + k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1 \right) I \right. \\ &\quad \left. + \hat{e}_1 \otimes \left[-(\nabla V(\varphi_0), \hat{e}_1) \hat{e}_1 - k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2} \right)^\parallel + k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2} \right)^\perp \right] \right\} \frac{\partial \delta\varphi}{\partial \alpha}, \end{aligned} \quad (3.27)$$

where

$$\left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}\right)^{\parallel} = \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1\right) \hat{e}_1$$

is the tangential component and

$$\left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}\right)^{\perp} = \frac{\partial^2 \varphi_0}{\partial \alpha^2} - \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}\right)^{\parallel}$$

is the normal component of $\frac{\partial^2 \varphi_0}{\partial \alpha^2}$. Hence the third term is of first order, and we should use the upwind scheme to solve it. Denote

$$\begin{aligned} A = & (\nabla V(\varphi_0), \hat{e}_1) I + k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1\right) I \\ & + \hat{e}_1 \otimes \left[-(\nabla V(\varphi_0), \hat{e}_1) \hat{e}_1 - k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}\right)^{\parallel} + k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}\right)^{\perp} \right]. \end{aligned}$$

Then A has n eigenvalues,

$$\lambda_1 = (\nabla V(\varphi_0), \hat{e}_1) + k \left(\frac{\partial^2 \varphi_0}{\partial \alpha^2}, \hat{e}_1\right) \text{ and } \lambda_2 = \dots = \lambda_n = 0. \quad (3.28)$$

Consequently, the stability condition takes the form

$$\max \left\{ \left| \frac{\lambda_1}{\Delta \alpha} \right| \right\} \Delta t \leq 1. \quad (3.29)$$

From (3.24), (3.26) and (3.29), we obtain the stability condition of the NEB method as follows,

$$\max \left\{ \frac{2k}{(\Delta \alpha)^2}, |(\hat{e}_1 \otimes \hat{e}_1 - I) \nabla^2 V(\varphi_0) - I|, \left| \frac{\lambda_1}{\Delta \alpha} \right| \right\} \Delta t \leq 1. \quad (3.30)$$

3.3 Stable finite difference schemes

As we have seen, upwind schemes are needed to solve both the ZTS equation and the NEB equation, which has also been illustrated in [6] from physical and numerical points of view. Suppose there are $N+1$ images $\{x_j\}_{j=0}^N$ on the string of the ZTS method or the band of the NEB method. The energies of the images are denoted by $V(x_j)$ ($j=0, 1, \dots, N$). For simplicity, we denote the potential force by $f_j = -\nabla V(x_j)$.

Since the eigenvalues of the matrix $B = I - \hat{e}_1 \otimes \hat{e}_1$ are all non-negative, we know the upwind scheme of the tangent for Eq. (3.15) in the ZTS method should be

$$\hat{\tau}_j = \begin{cases} \frac{x_{j+1} - x_j}{\Delta \alpha_j} & \text{if } (f_j, \hat{\tau}_j) \leq 0 \\ \frac{x_j - x_{j-1}}{\Delta \alpha_j} & \text{if } (f_j, \hat{\tau}_j) > 0 \end{cases} \quad j = 1, 2, \dots, N-1, \quad (3.31)$$

where $\Delta\alpha_j$ is the distance between x_{j+1} and x_j if $(f_j, \hat{\tau}_j) \leq 0$ (or between x_{j-1} and x_j if $(f_j, \hat{\tau}_j) > 0$), which is an approximation of the arc length. For a smooth potential surface, the upwind scheme is

$$\hat{\tau}_j = \begin{cases} \frac{x_{j+1} - x_j}{\Delta\alpha_j} & \text{if } V_{j-1} < V_j < V_{j+1} \\ \frac{x_j - x_{j-1}}{\Delta\alpha_j} & \text{if } V_{j-1} > V_j > V_{j+1} \end{cases} \quad j=1,2,\dots,N-1. \quad (3.32)$$

When x_j is a minimum or maximum along the path, i.e., $V_j < V_{j-1}$ and $V_j < V_{j+1}$; or $V_j > V_{j-1}$ and $V_j > V_{j+1}$, the tangent is computed by

$$\hat{\tau}_j = (x_j - x_{j-1})\Delta V_{j+1} + (x_{j+1} - x_j)\Delta V_j, \quad (3.33)$$

where $\Delta V_j = |V_j - V_{j-1}|$. Finally, we normalize the tangents so that their lengths satisfy $|\hat{\tau}_j| = 1$ ($j=1,2,\dots,N-1$).

For the NEB method, from (3.22), we know that the upwind scheme of the tangent relates not only to the potential, but also to the forces of the springs. The upwind scheme of the NEB method is

$$\hat{\tau}_j = \begin{cases} \frac{x_{j+1} - x_j}{\Delta\alpha_j} & \text{if } (f_j + \tilde{f}_j, \hat{\tau}_j) \leq 0 \\ \frac{x_j - x_{j-1}}{\Delta\alpha_j} & \text{if } (f_j + \tilde{f}_j, \hat{\tau}_j) > 0 \end{cases} \quad j=1,2,\dots,N-1, \quad (3.34)$$

where

$$\tilde{f}_j = k \frac{x_{j+1} - 2x_j + x_{j-1}}{(\Delta\alpha_j)^2}$$

is an approximation of the spring force at image j . When the points on the path distribute uniformly, the forces of the springs are zero. It is known that the forces of the springs make the points on the path distribute uniformly. The spring force becomes very small when the path converges to the MEP. So in practice, we also use Eqs. (3.32) and (3.33) to approximate the tangents in the NEB method.

Now we give the algorithms for the ZTS method and the NEB method. For the ZTS method, because of the intrinsic description of the string, it is very simple to implement an efficient algorithm which solves (2.3) using a time-splitting scheme. The string is discretized into a chain of images which move under the potential force $-(\nabla V(\varphi))^\perp$. After a number of steps depending on the accuracy for the constraint, a reparametrization step is applied to conserve the constraint. This is illustrated in Algorithm 3.1.

For the NEB method, the implementation of the NEB method is presented in Algorithm 3.2.

Algorithm 3.1:

-
1. Give an initial discrete string, for example, interpolate the two metastable states linearly. Denote the images on the string by $\{x_j^0\}_{j=0}^N$.
 2. Move the images using the following finite difference scheme

$$\frac{x_j^{n+1} - x_j^n}{\Delta t} = f_j^n - (f_j^n, \hat{\tau}_j^n) \hat{\tau}_j^n, \quad j=1,2,\dots,N-1, \quad (3.35)$$

where $\hat{\tau}_j$ is given by the upwind scheme of the ZTS method. If $\max_{0 < j < n} |f_j^n - (f_j^n, \hat{\tau}_j^n) \hat{\tau}_j^n| < \varepsilon$, $\varepsilon \ll 1$, then stop.

3. If $\frac{\min \Delta \alpha_j}{\max \Delta \alpha_j} < c$, $0 < c < 1$, reparameterize the string using interpolation to enforce the constraint.
 4. Goto step 2.
-

Algorithm 3.2:

-
1. Give an initial discrete path, for example, interpolate the two metastable states linearly. We denote the images on the path by $\{x_j^0\}_{j=0}^N$.
 2. Move the images using the following difference scheme

$$\frac{x_j^{n+1} - x_j^n}{\Delta t} = f_j^n - (f_j^n, \hat{\tau}_j^n) \hat{\tau}_j^n + (\tilde{f}_j^n, \hat{\tau}_j^n) \hat{\tau}_j^n, \quad j=1,2,\dots,N-1, \quad (3.36)$$

where $\tilde{f}_j^n = k(x_{j+1}^n - 2x_j^n + x_{j-1}^n) / (\Delta \alpha_j)^2$ is the force of the spring and $\hat{\tau}_j$ is computed by the upwind scheme of the NEB method. If $\max_{0 < j < N} |f_j^n - (f_j^n, \hat{\tau}_j^n) \hat{\tau}_j^n| < \varepsilon$, $\varepsilon \ll 1$, then stop.

3. Goto step 2.
-

3.4 The choice for the time step

For a not very complex problem we can obtain $\max_{0 < j < N} |(\nabla V(x_j))^\perp| < \varepsilon$ after hundreds or thousands of steps. We choose the time step according to (3.18) for the ZTS method, and (3.24) and (3.29) for the NEB method, respectively, which is

$$dt = C \min_{0 < j < N} \left\{ \frac{\Delta \alpha_j}{|(f_j, \hat{\tau}_j)|} \right\} \quad (3.37)$$

for the ZTS method, and

$$dt = C \min_{0 < j < N} \left\{ \frac{\Delta \alpha_j}{|-(f_j, \hat{\tau}_j) + k(\frac{\partial^2 \varphi}{\partial a^2}, \hat{\tau}_j)|}, \frac{(\Delta \alpha_j)^2}{2k} \right\} \quad (3.38)$$

for the NEB method, where $C \in (0,1)$ is a constant.

When $\Delta\alpha_j$ decreases, the time step of the NEB method decreases rapidly since the time step is proportional to $(\Delta\alpha_j)^2$, and the number of steps increases rapidly. To achieve a reasonable time step for the NEB method, we need to choose a small elastic constant k . However, when k is very small, the band has a low order accuracy. Thus it is difficult to choose an optimal elastic constant. We have to choose different elastic constants for different problems. Even for the same problem, we have to choose different elastic constants for bands with different number of images to achieve better accuracy.

Although the stability condition of the linearized equation is not the same as that of the corresponding nonlinear equation, it is usually used as a reference for the corresponding nonlinear equation. Here we illustrate this for the ZTS method and the NEB method with the double well potential

$$V(x_1, x_2) = (x_1^2 - 1)^2 + x_2^2. \tag{3.39}$$

The MEP can be computed exactly

$$\begin{cases} (\varphi_0(\alpha))_1 = -1 + 2\alpha \\ (\varphi_0(\alpha))_2 = 0 \end{cases} \quad (0 \leq \alpha \leq 1), \tag{3.40}$$

where $(\varphi_0(\alpha))_i (i=1,2)$ is the i th element of $\varphi_0(\alpha)$. Denote

$$\begin{aligned} f &= -\nabla V(\varphi_0) = \begin{pmatrix} -4x_1(x_1^2 - 1) \\ 0 \end{pmatrix}, \\ H &= \nabla^2 V(\varphi_0) = \begin{pmatrix} 4(3x_1^2 - 1) & 0 \\ 0 & 2 \end{pmatrix}, \\ \hat{e}_1 &= \frac{\partial \varphi_0}{\partial \alpha} \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \end{aligned}$$

We can rewrite the linearized equation for the NEB method as

$$\begin{aligned} \frac{\partial \delta \varphi}{\partial t} &= k(\hat{e}_1 \otimes \hat{e}_1) \frac{\partial^2 \delta \varphi}{\partial \alpha^2} + (\hat{e}_1 \otimes \hat{e}_1 - I) H \delta \varphi - (f, \hat{e}_1) \left| \frac{\partial \varphi_0}{\partial \alpha} \right|^{-1} (I - \hat{e}_1 \otimes \hat{e}_1) \frac{\partial \delta \varphi}{\partial \alpha} \\ &= C_0 \delta \varphi + C_1 \frac{\partial \delta \varphi}{\partial \alpha} + k C_2 \frac{\partial^2 \delta \varphi}{\partial \alpha^2}, \end{aligned}$$

where

$$C_0 = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}, \quad C_1 = -\frac{(f, \hat{e}_1)}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad C_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus we obtain a simple version

$$\begin{cases} \left(\frac{\partial \delta \varphi}{\partial t} \right)_1 = k \left(\frac{\partial^2 \delta \varphi}{\partial \alpha^2} \right)_1 \\ \left(\frac{\partial \delta \varphi}{\partial t} \right)_2 = -2(\delta \varphi)_2 - \frac{f_1}{2} \left(\frac{\partial \delta \varphi}{\partial \alpha} \right)_2 \end{cases} \tag{3.41}$$

For simplicity, we use φ for $\delta\varphi$ in the finite difference equation. The finite difference equation for (3.41) is as follows

$$\left\{ \begin{array}{l} \varphi_j^{m+1}(1) = \varphi_j^m(1) + C_2(1,1)k \frac{\varphi_{j+1}^m(1) - 2\varphi_j^m(1) + \varphi_{j-1}^m(1)}{(\Delta\alpha_j)^2} \Delta t, \\ \varphi_j^{m+1}(2) = \varphi_j^m(2) + C_0(2,2)\varphi_j^m(2)\Delta t + \begin{cases} C_1(2,2) \frac{\varphi_{j+1}^m(2) - \varphi_j^m(2)}{\Delta\alpha_j} \Delta t, & \text{if } C_1(2,2) \geq 0, \\ C_1(2,2) \frac{\varphi_j^m(2) - \varphi_{j-1}^m(2)}{\Delta\alpha_j} \Delta t, & \text{if } C_1(2,2) < 0, \end{cases} \\ j = 1, 2, \dots, N-1, \end{array} \right. \quad (3.42)$$

where $\varphi_j^m(i)$ ($i=1,2$) is the i th element of φ_j^m . Here we have used the upwind scheme for the second equation. These are linear equations and the corresponding stability condition can be found by using the Fourier method. For the first equation of (3.42), the optimal time step is

$$\Delta t_1 = \frac{(\Delta\alpha_j)^2}{2kC_2(1,1)}. \quad (3.43)$$

Using the Fourier method to the second equation of (3.42), we achieve the optimal time step for the second equation

$$\Delta t_2 = \min \left\{ 1, \frac{-2C_0(2,2) + 2\frac{|C_1(2,2)|}{\Delta\alpha_j}}{2\frac{C_1(2,2)^2}{(\Delta\alpha_j)^2} + C_0(2,2)^2 - 2\frac{|C_1(2,2)|C_0(2,2)}{\Delta\alpha_j}} \right\}. \quad (3.44)$$

So the optimal time step for the NEB method is

$$\Delta t = \min\{\Delta t_1, \Delta t_2\}. \quad (3.45)$$

The linearized equation for the ZTS method is

$$\left\{ \begin{array}{l} \left(\frac{\partial\delta\varphi}{\partial t} \right)_1 = 0, \\ \left(\frac{\partial\delta\varphi}{\partial t} \right)_2 = -2(\delta\varphi)_2 - \frac{f_1}{2} \left(\frac{\partial\delta\varphi}{\partial\alpha} \right)_2', \\ \left(\frac{\partial^2\delta\varphi}{\partial\alpha^2} \right)_1 = 0. \end{array} \right. \quad (3.46)$$

The time step of Eq. (3.46) is just determined by the second equation and it is just the same as the second equation of (3.41). So the optimal time step for the linearized equation of the ZTS method for the double well potential is

$$\Delta t = \Delta t_2, \quad (3.47)$$

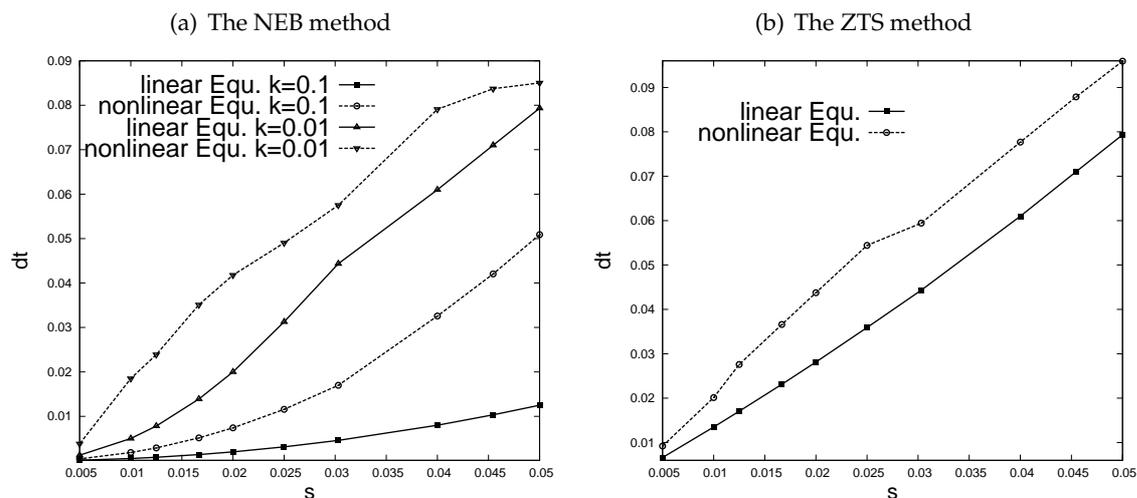


Figure 1: The relation between the optimal time step and the normalized arc length in the NEB method and the ZTS method for the double well potential. (a): The NEB method, and (b): the ZTS method.

where Δt_2 is given by Eq. (3.44).

In our numerical experiments, we choose the time step as $dt = C * \Delta t$, ($0 < C < 1$), where Δt is the optimal time step.

So far we have obtained the optimal time step for the linearized equation. The optimal time step of the corresponding nonlinear equation usually can be approximated by that of the linearized equation, which is illustrated in Fig. 1.

This example has some special property. The diffusion term and the lower order term are separable for the NEB method. Thus we can obtain an optimal elastic constant. If we choose the time step from the second equation of (3.42), we obtain an optimal elastic constant

$$k = 0.5 \frac{(\Delta \alpha)^2}{C_2(1,1) \Delta t}. \quad (3.48)$$

Under the stability condition, for the same Δt , the smaller the k is, the bigger the time step is and the faster the band converges to the solution. This is illustrated by numerical experiments in Fig. 1(a).

4 Accuracy

Using Algorithms 3.1 and 3.2 we can obtain an approximate MEP. In this section, we shall analyze the accuracy of the numerical MEP and the transition state (TS) which is the most important point on the MEP.

4.1 Pathway

For the accuracy of the numerical MEP, the most natural way is to measure the distance between the numerical MEP and the exact MEP. The steady state equation of the ZTS method and the NEB method is

$$0 = -(\nabla V)^\perp = -\nabla V + (\nabla V, \hat{\tau}) \hat{\tau}. \tag{4.1}$$

Let $f(\varphi) = -\nabla V(\varphi)$. Then (4.1) becomes

$$0 = f(\varphi) - (f(\varphi), \hat{\tau}) \hat{\tau}. \tag{4.2}$$

The finite difference equation of (4.1) is formulated as

$$\begin{cases} f(x_j) - \left(f(x_j), \frac{x_{j+1} - x_j}{\Delta\alpha} \right) \frac{x_{j+1} - x_j}{\Delta\alpha} = 0 & \text{if } f(x_j) \leq 0 \\ f(x_j) - \left(f(x_j), \frac{x_j - x_{j-1}}{\Delta\alpha} \right) \frac{x_j - x_{j-1}}{\Delta\alpha} = 0 & \text{if } f(x_j) > 0 \end{cases} \quad j=1,2,\dots,N-1. \tag{4.3}$$

Here we use the upwind scheme and assume the distances $\Delta\alpha$ between the adjacent points are equal. Let $\{\varphi_j\}_{j=0}^N$ be $N+1$ points on the exact MEP, that is,

$$0 = f(\varphi_j) - (f(\varphi_j), \hat{\tau}_j) \hat{\tau}_j.$$

Suppose $\Delta\alpha \equiv |\varphi_j - \varphi_{j-1}|$ ($j=1,2,\dots,N$) are equal. Then we have

$$\begin{aligned} & f(\varphi_j) - \left(f(\varphi_j), \frac{\varphi_{j+1} - \varphi_j}{\Delta\alpha} \right) \frac{\varphi_{j+1} - \varphi_j}{\Delta\alpha} \\ &= f(\varphi_j) - \left(f(\varphi_j), \frac{\varphi_j + \Delta\alpha \hat{\tau}_j + \mathcal{O}((\Delta\alpha)^2) - \varphi_j}{\Delta\alpha} \right) \frac{\varphi_j + \Delta\alpha \hat{\tau}_j + \mathcal{O}((\Delta\alpha)^2) - \varphi_j}{\Delta\alpha} \\ &= f(\varphi_j) - \left(f(\varphi_j), \frac{\Delta\alpha \hat{\tau}_j + \mathcal{O}((\Delta\alpha)^2)}{\Delta\alpha} \right) \frac{\Delta\alpha \hat{\tau}_j + \mathcal{O}((\Delta\alpha)^2)}{\Delta\alpha} \\ &= f(\varphi_j) - (f(\varphi_j), \hat{\tau}_j) \hat{\tau}_j - (f(\varphi_j), \hat{\tau}_j) \mathcal{O}(\Delta\alpha) - (f(\varphi_j), \mathcal{O}(\Delta\alpha)) \hat{\tau}_j. \end{aligned} \tag{4.4}$$

Thus

$$f(\varphi_j) - \left(f(\varphi_j), \frac{\varphi_{j+1} - \varphi_j}{\Delta\alpha} \right) \frac{\varphi_{j+1} - \varphi_j}{\Delta\alpha} = \mathcal{O}(\Delta\alpha). \tag{4.5}$$

The same result can be obtained for the second equation of (4.3). Then we know that the difference equation is consistent and the accuracy is of first order.

Usually the exact MEP is unknown and it is very difficult to obtain an analytical one. We give a numerical MEP with enough points as the exact MEP. Here we compute a MEP of 1200 points as the exact one, denoted by $\Phi = \{\varphi_j\}_{j=0}^{1200}$, and compare the MEP of less

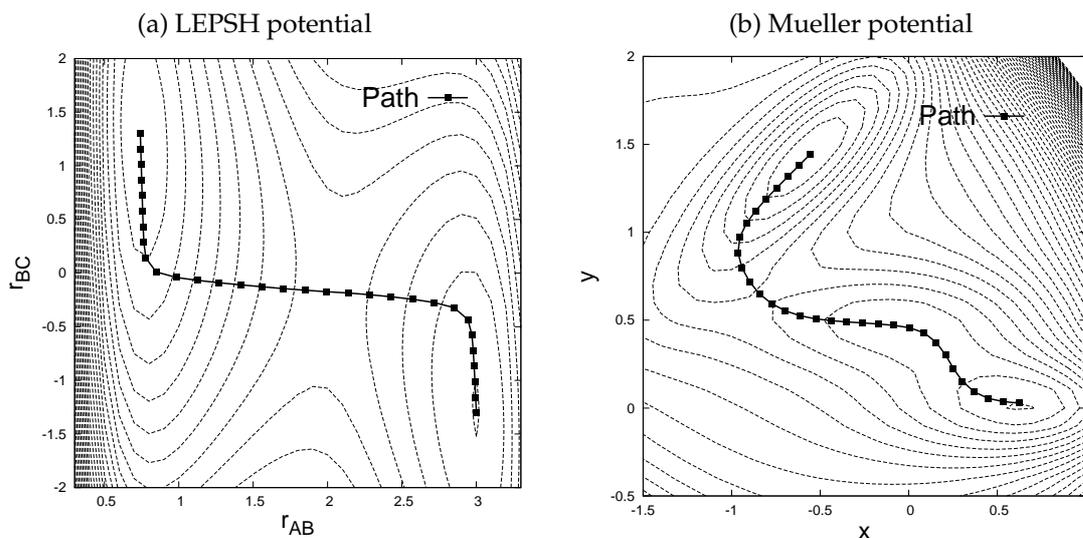


Figure 2: Two typical two-dimensional potential contour.

points $X = \{x_j\}_{j=0}^N$ ($N \ll 1200$) to the MEP Φ . We define the error under the L_∞ -norm as the distance between two MEPs, that is

$$\|\Phi - X\|_{L_\infty} = \max_{0 < j < N} \min_{0 < i < 1200} \{|\varphi_i - x_j|\}. \quad (4.6)$$

The error under the L_2 -norm is defined by

$$\|\Phi - X\|_{L_2} = \left(\sum_{j=1}^{N-1} \min_{0 < i < 1200} \{|\varphi_i - x_j|^2 \Delta \alpha_j\} \right)^{\frac{1}{2}}. \quad (4.7)$$

In our experiments, we choose N such that the normalized arc length interval is $1/N = j/300$, $j=1, 2, \dots, 10$. From (4.5) we may expect that the ZTS method and the NEB method have a first order accuracy for the errors under the L_∞ -norm and the L_2 -norm. Here we use two typical two-dimensional examples to test our results. The two typical examples are:

- (LEPS potential coupled with harmonic oscillator). We consider the system involving four atoms A , B , C and D confined in a line. Atom B can form a chemical bond with either A or C , and can interact with the fourth atom D in a harmonic way. The form of the potential can be found in [9]. A contour plot of the potential surface is given in Fig. 2.
- (Mueller potential). Mueller potential is invented as a nontrivial test example for reaction path algorithms [15, 16]. A contour plot of the potential surface is given in Fig. 2.

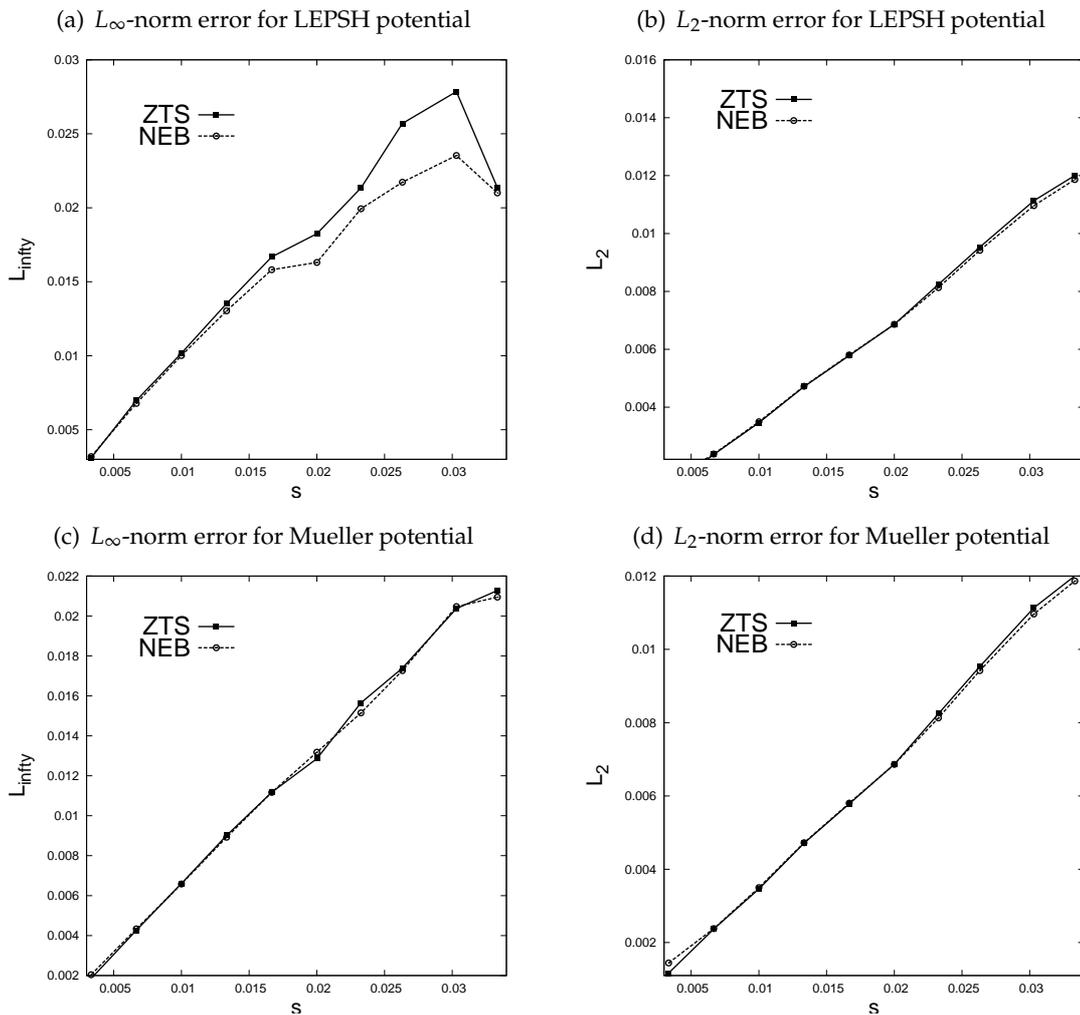


Figure 3: The horizontal coordinate is the normalized arc length interval $1/N$ and the vertical coordinate is the error. The solid lines are the results of the ZTS method and the dashed lines are those of the NEB method.

The numerical results are demonstrated in Fig. 3. From our numerical experiments, we can see that the errors under the L_∞ -norm and the L_2 -norm indeed have first order accuracy.

Remark 4.1. The L_∞ -norm error for LEPSH potential has some problems in the 30-point and 33-point results. This is because the path of the 30-point string cuts the corner where the error of the point is large and is just the L_∞ -norm error of the path. The L_∞ -norm error, which is an individual point behavior, sometimes can not describe the error of the path accurately. We can see the L_2 -norm error, which is an average behaviour of all the points, makes a correction to this problem.

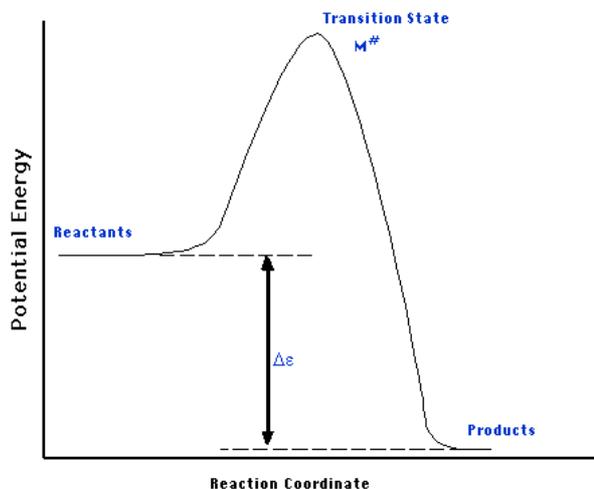


Figure 4: The MEP in reaction coordinate.

4.2 Transition state (TS)

The transition state is a special saddle point, which is very important in the dynamics of complex systems. There are many papers trying to find the saddle points in the energy potential surface (see [17] and references there). Once we find the saddle points, it is easy to find the MEP. The TS also has a close relation to the transition rate.

To understand the importance of the transition state more clearly, we first give a brief introduction to the transition state. Murrell and Laidler defined a *transition state* as a stationary point with a single negative Hessian eigenvalue [14]. Stationary points with one or more negative Hessian eigenvalues are often called *saddle points*, because they have a local maximum in one or more degrees of freedom. Fig. 4 gives an illustration for the transition state, which has appeared in many papers or books. The transition state is then identified by the highest energy point on the MEP [1]. We refer the reader to [24] and the references cited therein for more details.

The energy and the negative eigenvalue of the Hessian matrix at the TS are related to the transition rate. Suppose A and B are two metastable states, and there is only one transition state C on the MEP which connects A and B . The transition rate for Eq. (2.1) from A to B can be expressed as [19]

$$k_{AB} = \frac{(\lambda_m |\lambda_s|)^{1/2}}{2\pi} \sqrt{\frac{\det H^\perp(A)}{\det H^\perp(C)}} e^{-\frac{1}{\hbar} \Delta V}, \quad (4.8)$$

where λ_m and λ_s are the eigenvalues of the Hessian matrix H corresponding to the eigenvectors $\varphi_\alpha(0)$ and $\varphi_\alpha(\alpha^*)$, $\varphi(0)$ and $\varphi(\alpha^*)$ are the parameter representation of the stable state A and the transition state C , and ΔV is the energy barrier defined by

$$\Delta V = V(C) - V(A). \quad (4.9)$$

Thus we need to analyze the following errors:

- The error of the energy barrier.

Since $V(A)$ is known, the error of ΔV is introduced by the approximation of the energy at the transition state $V(C)$. We define the error of the energy barrier by

$$D_e(X) = |V(C) - V(x_i)|, \quad (4.10)$$

where x_i is the approximation of the TS.

- The error of the eigenvalue and $\det H^\perp(C)$.

We have obtained an approximate MEP using the ZTS method or the NEB method. With this approximate MEP, we can use interpolation to give an estimation of the TS. Suppose energies $V_j = V(x_j)$ ($j=0,1,\dots,N$) have the maximum value at x_{imax} on the numerical MEP, where $imax \in [1,N]$ is an integer. Let the coordinates and the energy be functions of arc length α , that is $x(\alpha)$ and $V(\alpha)$. On the four points x_i ($i=imax-2,imax-1,imax,imax+1$), we define the arc length parameters

$$\begin{aligned} \alpha_1 &= 0, \\ \alpha_2 &= |x(imax-1) - x(imax-2)|, \\ \alpha_3 &= |x(imax) - x(imax-1)| + \alpha_2, \\ \alpha_4 &= |x(imax+1) - x(imax)| + \alpha_3. \end{aligned}$$

We choose $\alpha_1, \alpha_2, \alpha_3$ and α_4 as the interpolation points, and use Lagrange cubic polynomials $Px(\alpha)$ and $PV(\alpha)$ to approximate $x(\alpha)$ and $V(\alpha)$, respectively. We first find the point α^* where the Lagrange cubic polynomial $PV(\alpha)$ attains its maximum, then take the point $Px(\alpha^*)$ as the approximate maximum point of $V(x(\alpha))$, which is the TS. From numerical experiments, this interpolation is better than what has been proposed in [6], which uses the forces at the points.

Now we estimate the accuracy of the TS using this interpolation method.

Lemma 4.1. ([11]) *Let $f: [a,b] \rightarrow \mathbb{R}$ be $(n+1)$ -times continuously differentiable. Then the remainder $R_n f := f - p_n$ for polynomial interpolation with $n+1$ distinct points $x_0, \dots, x_n \in [a,b]$ can be represented in the form*

$$(R_n f)(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^n (x - x_j), \quad x \in [a,b], \quad (4.11)$$

for some $\xi \in [a,b]$ depending on x .

Following the above lemma, if the interpolation step is $h_j = j * h$ for some fixed h , the remainder is $R_3^j f(x) \approx C j^4 h^4$, where C is a constant depending on x and j . So

$$\log(R_3^j f(x)) \approx \log(C) + 4\log(j) + 4\log(h),$$

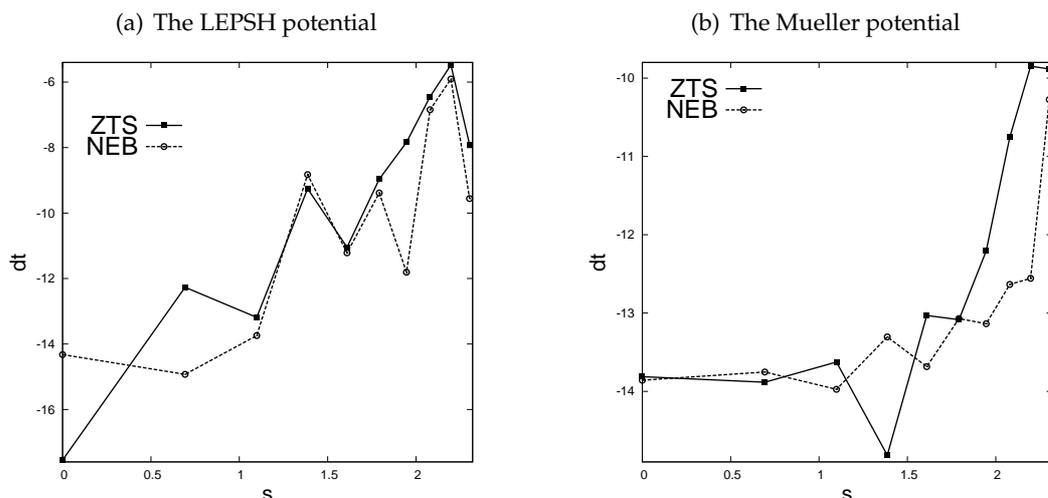


Figure 5: The horizontal coordinate is $\log(j)$, where $j \cdot h$ is the normalized arc length interval. The vertical coordinate is the logarithm of the remainder $\log(R_3^j f(x))$.

which means the logarithm of the remainder and the logarithm of j have a linear relationship. The relations of the logarithm of the errors at the transition state and the logarithm of $\{j, j=1,2,\dots,10\}$ are shown in Fig. 5 for the Mueller potential and the LEPESH potential. We can see the results of the Mueller potential are consistent well with the interpolation lemma. That is the logarithm of the errors at the transition state and the logarithm of $\{j, j=1,2,\dots,10\}$ have almost a linear relationship. However the results of the LEPESH potential are not consistent with the theory. This is possibly because the fourth derivative of the potential becomes bigger near the transition state.

4.3 Some improvement on the transition state

In this subsection, we introduce two methods to give a better TS.

- Climbing image method.

This method was developed for the NEB method in [7]. Here we use it also for the ZTS method. The transition state is a maximum along the MEP and a minimum in the hyperplane normal to the path. Using this property, we change the movement of the points which have a local maximum energy along the path. We move these points along the force in the hyperplane normal to the path and against the force along the path. When the path is reparameterized, the points that have local maximum along the path do not change. This method is just like the eigenvector-following method [3]. Near the transition state, the direction of the path is an approximation of the eigenvector of the negative Hessian eigenvalue. This method costs only little more time than the standard ZTS method. Using this method, the accuracy of the energy and the negative Hessian eigenvalue is improved greatly.

Table 1: Errors for energy barrier for the LEPSH potential.

N	Interpolation	CI-ZTS	WA-ZTS
30	5.10×10^{-5}	5.77×10^{-15}	1.19×10^{-9}
60	2.19×10^{-6}	5.22×10^{-15}	3.48×10^{-11}
100	1.21×10^{-6}	5.33×10^{-15}	3.22×10^{-11}
150	9.35×10^{-7}	5.11×10^{-15}	1.19×10^{-9}
300	1.00×10^{-6}	5.22×10^{-15}	1.20×10^{-9}

- Weighted arc length method.

Since the transition state is very important, we give some changes in the ZTS method near the transition state. The basic idea is to use a non-equivalent arc length string. We use small arc length interval near the TS. Here we propose a simple and efficient choice for the arc length parameter. Let $\{x_i\}_{i=0}^N$ be the string and x_{imax} be the saddle point. When reparameterizing the path, we define the parameter (the normalized arc length) at the points as follows:

$$\begin{cases} \alpha_{imax} = \frac{imax}{N}, \\ \alpha_{imax-1} = \alpha_{imax} - C/N, \\ \alpha_{imax+1} = \alpha_{imax} + C/N, \\ \alpha_i = i * \alpha_{imax-1} / (imax-1), \quad i = 1, 2, \dots, imax-2, \\ \alpha_i = \alpha_{i-1} + (1.0 - \alpha_{imax+1}) / (n - imax - 1), \quad i = imax+2, \dots, N, \end{cases} \quad (4.12)$$

where C is a constant, for example it can be $1/20$ or $1/30$. Since the force near the TS is very small, from Eq. (3.18), we know the time step is not affected in this case. We need only little more time using this method than using the standard ZTS method. We only change the motion of the points near the saddle point. It is very simple to implement.

The weighted arc length method of the NEB method is implemented by using different elastic constant for different intervals. More precisely, the forces of the springs are given by

$$\tilde{f}_i = k_{i+1}(x_{i+1} - x_i) + k_i(x_i - x_{i-1}), \quad i = 1, 2, \dots, N-1,$$

where the elastic constant k_i is big near the saddle point and is small near the end point. A choice for the elastic constant is given in [7].

We use these techniques to the ZTS method and the NEB method for the LEPSH potential and the Mueller potential. We show the results in Tables 1, and 2 for the ZTS method. We can see that the accuracy has been greatly improved. The results are similar for the NEB method.

The weighted arc length ZTS (WA-ZTS) method does not have accuracy as high as that of the climbing image ZTS (CI-ZTS) method for these two examples. The weighted

Table 2: Errors for energy barrier for the Mueller potential.

N	Interpolation	CI-ZTS	WA-ZTS
30	3.69×10^{-4}	7.10×10^{-15}	6.35×10^{-10}
60	1.60×10^{-5}	7.10×10^{-15}	2.01×10^{-9}
100	1.88×10^{-6}	7.10×10^{-15}	3.29×10^{-10}
150	4.70×10^{-6}	7.10×10^{-15}	6.25×10^{-13}
300	2.41×10^{-8}	7.10×10^{-15}	1.21×10^{-13}

arc length method aims to accurately calculate the pathway around the saddle point, in particular the saddle point and the unstable direction. The climbing image method is a local search technique like the steepest descent method or Newton's method for local minima, therefore it is only for the saddle point. When the potential surface is rough, which is the case in most of the problems that we are interested in, the climbing image method does not work any more, while the weighted arc length method can still work.

Remark 4.2. In principle, the errors in Tables 1 and 2 for the CI-ZTS should be the machine error (10^{-15} or 10^{-16} if double precision is used), which is independent of the problem or the parameter N .

5 Computational cost

In this section, we shall estimate the computational cost of the ZTS method and the NEB method. Suppose there are $N+1$ points $\{x_j\}_{j=0}^N$ on the path. When the iteration number that makes the forces perpendicular to the path satisfy

$$\max_{0 < j < N} \{ |(\nabla V(x_j))^\perp| \} < \varepsilon$$

(we choose $\varepsilon = 10^{-7}$ in practice) is m , the computational cost can be measured by $m \times (N-1)$. In the ZTS method, the time step is chosen according to Eq. (3.37) for which we choose $C = 0.6$, i.e.,

$$\Delta t = 0.6 \min_{0 < i < N} \left\{ \frac{\Delta \alpha_i}{(f_i, \hat{\tau}_i)} \right\}. \quad (5.1)$$

The iteration number m is expected to be proportional to N , which is inversely proportional to the arc length interval $\Delta \alpha_i$. For the NEB method, the time step depends on the potential energy and the elastic constant. On the one hand, when the elastic constant is too small, the path has a lower accuracy. On the other hand, when the elastic constant is too big, we need to use very small time step to guarantee the stability of the difference scheme. We have carried out many numerical experiments for the Mueller potential and the LEPSH potential. We found that the elastic constant can be chosen as $k = c/N$ for

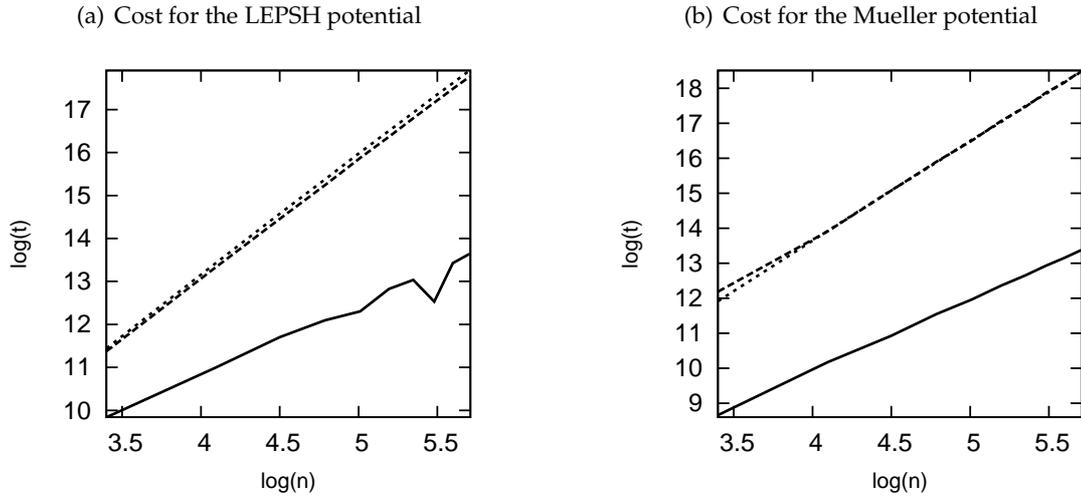


Figure 6: The solid curve gives the computational cost of the ZTS method, the long dashed line shows the computational cost for the NEB method and the short dashed line represents the computational cost of the NEB method with the force given by (5.3). The horizontal coordinate is $\log(N)$, where $1/N$ is the normalized arc length interval $\Delta\alpha$ and the vertical coordinate is the computational cost $\log(t)$. Here $t = m \times (N - 1)$.

accuracy. From (3.24) and (3.29), the elastic constant $k = c/N$ can also guarantee a reasonable time step. We used $c = 0.5$ in the LEP SH potential and $c = 30$ in the Mueller potential. Then we chose the time step in the NEB method according to (3.38), which is

$$\Delta t = 0.6 \min_{0 < i < N} \left\{ \frac{\Delta\alpha_i}{|-(f_i, \hat{\tau}_i) + (\tilde{f}_i, \hat{\tau}_i)|}, \frac{(\Delta\alpha_i)^2}{2k} \right\}. \tag{5.2}$$

With these choices, in principle, the iteration number m for the NEB method is also proportional to N . We give the relations between $\log(m \times (N - 1))$ and $\log N$ in Fig. 6 for the ZTS method and the NEB method. The tangent of the line of the NEB method is about 2.8 for the Mueller potential and 2.75 for the LEP SH potential. The tangent of the line of the ZTS method is about 1.65 for the Mueller potential and 2.03 for the LEP SH potential. For the two typical examples, the computational cost of the ZTS method is less than that of the NEB method. Especially for the Mueller potential, the computational cost of the ZTS method is ten times less than that of the NEB method.

In [6], the spring force in the NEB method is given by

$$\tilde{f}_i = k(|x_{i+1} - x_i| - |x_i - x_{i-1}|)\hat{\tau}_i, \quad i = 1, 2, \dots, n - 1, \tag{5.3}$$

where $|x_{i+1} - x_i|$ is the distance between x_{i+1} and x_i . In numerical experiments, the computational cost of this NEB method is almost the same as that of the standard NEB method.

6 Conclusions

In this paper, we have presented a detailed theoretical analysis of the ZTS method and the NEB method. First, we have given the stability conditions. From the stability condition, we have given an adaptive time step for each method, which makes the computation more efficient. Second, we have estimated the accuracy of the ZTS method and the NEB method. The errors of the MEP under the L_∞ -norm and the L_2 -norm have first order accuracy in both methods. Finally, we have estimated the computational cost and found that the ZTS method costs much less than the NEB method. In addition, the ZTS method is very simple to implement. While, the NEB method has an additional elastic constant parameter which is difficult to choose in the implementation.

Another focus of this paper is the discussion of the properties of the transition state. We have approximated the transition state using interpolation. Two techniques have also been proposed for the ZTS method to improve the accuracy at the transition state. When the potential surface is smooth, the climbing image method is recommended, because it produces the best transition state. However, when the potential surface is very rough, the climbing image method does not work any more, while the weighted arc length method can still work.

Acknowledgments

This work was partially supported by the National Science Foundation of China under the grant 10425105 and the National Basic Research Program under the grant 2005CB321704. The author thanks Professor Weinan E for his supervision, Professor Weiqing Ren for providing code of the ZTS method, and Professor Yang Xiang for improving the English writing.

References

- [1] D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation, *J. Chem. Phys.*, 68 (1978), 2959-2970.
- [2] I. M. Ciobica and T. A. van Santen, A DFT study of CH_x chemisorption and transition states for C-H activation on the Ru(1120) surface, *J. Phys. Chem. B*, 106 (2002), 6200-6205.
- [3] G. M. Crippen and H. A. Scheraga, Minimization of polypeptide energy. XI. The method of gentlest ascent, *Arch. Biochem. Biophys.*, 144 (1971), 462-466.
- [4] W. E, W. Ren and E. Vanden-Eijnden, String method for the study of rare events, *Phys. Rev. B*, 66 (2002), 52301-52304.
- [5] W. E, W. Ren and E. Vanden-Eijnden, Energy landscape and thermally activated switching of submicron-sized ferromagnetic elements, *J. Appl. Phys.*, 93 (2003), 2275-2282.
- [6] G. Henkelman and H. Jónsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, *J. Chem. Phys.*, 113 (2000), 9978-9985.

- [7] G. Henkelman, B. P. Uberuaga and H. Jónsson, A climbing image nudged elastic band method for finding saddle points and minimum energy paths, *J. Chem. Phys.*, 113 (2000), 9901-9904.
- [8] A. Heyden, B. Peters, A. T. Bell and F. J. Keil, Comprehensive DFT study of nitrous oxide decomposition over Fe-ZSM-5, *J. Phys. Chem. B*, 109 (2005), 1857-1873.
- [9] H. Jónsson, G. Mills and K. W. Jacobsen, Nudged elastic band method for finding minimum energy paths of transitions, in: B. J. Berne, G. Ciccoti and D. F. Coker (Eds.), *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, 1998.
- [10] Y. Kanai, A. Tilocca, A. Selloni and R. Car, First-principles string molecular dynamics: An efficient approach for finding chemical reaction pathways, *J. Chem. Phys.*, 121 (2004), 3359-3367.
- [11] R. Kress, *Numerical Analysis*, Springer-Verlag, 1998.
- [12] P. Maragakis, S. A. Andreev, Y. Brumer, et al., Adaptive nudged elastic band approach for transition state calculation, *J. Chem. Phys.*, 117 (2002), 4651-4658.
- [13] D. H. Mathews and D. A. Case, Nudged elastic band calculation of minimal energy paths for the conformational change of a GG non-canonical pair, *J. Mol. Biol.*, 357 (2006), 1683-1693.
- [14] J. N. Murrell and K. J. Laidler, Symmetries of activated complexes, *Trans. Faraday Soc.*, 64 (1968), 371-377.
- [15] K. Mueller, Reaktionswege auf mehrdimensionalen Energiehyperflächen, *Angew. Chem.*, 19 (1980), 1-14.
- [16] K. Mueller and L. D. Brown, Location of saddle points and minimum energy paths by a constrained simplex optimization procedure, *Theor. Chim. Acta.*, 53 (1979), 75-93.
- [17] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson and H. Jónsson, Comparison of methods for finding saddle points without knowledge of the final states, *J. Chem. Phys.*, 121 (2004), 9776-9792.
- [18] W. Quapp, A growing string method for the reaction pathway defined by a Newton trajectory, *J. Chem. Phys.*, 122 (2005), 174106-174116.
- [19] W. Ren, *Numerical Methods for the Study of Energy Landscapes and Rare Events*, PhD thesis, New York University, 2002.
- [20] W. Ren, E. Vanden-Eijnden, P. Maragakis and W. E, Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide, *J. Chem. Phys.*, 123 (2005), 134109-134120.
- [21] M. R. Sorensen, K. W. Jacobsen and H. Jónsson, Thermal diffusion processes in metal tip-surface interactions: Contact formation and adatom mobility, *Phys. Rev. Lett.*, 77 (1996), 5067-5070.
- [22] B. P. Uberuaga, M. Leskovar, A. P. Smith, H. Jónsson and M. Olmstead, Diffusion of Ge below the Si(100) surface: Theory and experiment, *Phys. Rev. Lett.*, 84 (2000), 2441-2444.
- [23] M. Villarba and H. Jónsson, Diffusion mechanisms relevant to metal crystal growth: Pt/Pt(111), *Surf. Sci.*, 317 (1994), 15-36.
- [24] D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge University Press, 2004.