The geometric minimum action method for computing minimum energy paths

Eric Vanden-Eijnden[1,a)] and Matthias Heymann[2,b)]
[1]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*
[2]*Mathematics Department, Duke University, Durham, North Carolina 27708, USA*

(Received 10 September 2007; accepted 14 December 2007; published online 14 February 2008)

An algorithm is proposed to calculate the minimum energy path (MEP). The algorithm is based on a variational formulation in which the MEP is characterized as the curve minimizing a certain functional. The algorithm performs this minimization using a preconditioned steepest-descent scheme with a reparametrization step to enforce a constraint on the curve parametrization. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2833040]

In the description of rare reactive events, the minimum energy paths (MEPs) play a special role. The MEPs are paths in configuration space which connect two minima of the energy via saddle points and are everywhere tangent to the potential force. The MEPs are also the paths of maximum likelihood by which transitions between these minima occur, at least if the dynamics of the system is overdamped (this property is explained later). Several computational methods have been developed for finding the MEPs, among which the nudged elastic band (NEB) method[1,2] and the (zero-temperature) string method[3,4] probably are the most successful.

Here we propose a new technique for computing MEPs based on the geometric minimum action method (gMAM) recently introduced in Refs. 5 and 6. gMAM was primarily designed for systems without an underlying potential (like, e.g., chemical kinetic systems), but it can be specialized to compute MEPs. The gMAM is slightly more complicated to implement than the string method, but it has the advantage of having a variational formulation in which the MEP is the minimizer of an action functional. This action had already been derived and used in Ref. 7, as discussed in more detail later.

To introduce the technique, let us consider a system modeled by the overdamped Langevin equation

$$\dot{x} = -\nabla V(x) + \sqrt{2\beta^{-1}}\,\eta(t), \tag{1}$$

where $x(t) \in \mathbb{R}^n$ denotes the instantaneous position of the system, $V(x)$ is the potential energy, β is the inverse temperature, and $\eta(t)$ is a Gaussian white noise with mean 0 and covariance $\langle \eta_i(t)\,\eta_j(t')\rangle = \delta_{ij}\delta(t-t')$. For convenience, we have written Eq. (1) in units of time in which the friction coefficient is unity. The probability density of a trajectory solution of Eq. (1) on the time interval $[0,T]$ is formally given by

$$\exp(-\beta S_T(x)), \tag{2}$$

where we defined the action

$$S_T(x) = \int_0^T |\dot{x}(t) + \nabla V(x(t))|^2 dt. \tag{3}$$

If we pick two points in configuration space, x_1 and x_2, the path of maximum likelihood by which the solution of Eq. (1) connects these two points is the one which maximizes the probability density Eq. (2), i.e., the path which

$$\text{minimizes } S_T(x) \text{ subject to } x(0) = x_1, \quad x(T) = x_2, \tag{4}$$

where the minimization is taken over both the path x itself and its length in time, T. The minimization over T ensures that the likelihood is maximized also over all possible times the path can take to travel from x_1 to x_2, consistent with viewing Eq. (2) as a probability density in x and T. The minimization over T can be performed analytically, which will give an action for curves whose minimizers are the MEPs. To show this, notice first that

$$S_T(x) \geq 2\int_0^T |\dot{x}||\nabla V(x)|dt + 2(V(x_2) - V(x_1)), \tag{5}$$

where we expanded the integrand in Eq. (3) and used $|\dot{x}|^2 + |\nabla V(x)|^2 \geq 2|\dot{x}||\nabla V(x)|$ and $\int_0^T \dot{x}\cdot\nabla V(x)dt = V(x_2) - V(x_1)$ [assuming that $x(0)=x_1$ and $x(T)=x_2$ as in Eq. (4)]. We now make two key observations. First, the minimum of $S_T(x)$ subject to $x(0)=x_1$ and $x(T)=x_2$ is equal to the minimum of its lower bound in Eq. (5) subject to these constraints. Indeed, since we minimize over all possible paths leading from x_1 to x_2, we can use specific paths parametrized so that $|\dot{x}| = |\nabla V(x)|$. In this case $|\dot{x}|^2 + |\nabla V(x)|^2 = 2|\dot{x}||\nabla V(x)|$, implying equality in the aforementioned estimate. The second key observation is that $\int_0^T |\dot{x}||\nabla V(x)|dt$ is the line integral of $|\nabla V|$ along the curve $\gamma = \{x(t) : t \in [0,T]\}$, i.e., it is equal to

$$\hat{S}(\gamma) = \int_\gamma |\nabla V| ds, \tag{6}$$

and as such it is independent of the way the curve γ is parametrized and, in particular, of T. Consequently, denoting by $S^\star$ the minimum of $S_T(x)$ over all x and T subject to $x(0)=x_1$ and $x(T)=x_2$, we have

a)Electronic mail: eve2@cims.nyu.edu.
b)Electronic mail: heymann@math.duke.edu.

 © 2008 American Institute of Physics

Downloaded 11 Jan 2009 to 67.159.68.103. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

$$S^\star = 2\min_\gamma \hat{S}(\gamma) + 2(V(x_2) - V(x_1)), \tag{7}$$

where the minimization in Eq. (7) is performed over all curves γ connecting x_1 to x_2. The new action $\hat{S}(\gamma)$ will be our starting point for the numerical procedure presented later to identify the MEP because the (local) minimizer(s) of $\hat{S}(\gamma)$ are nothing but the MEP(s) connecting x_1 to x_2 [assuming from now on that x_1 and x_2 are the locations of two minima of $V(x)$]. This is a simple consequence of the facts that (i) $\int_\gamma |\nabla V| ds \geq \Delta V_{1,2} + \Delta V_{2,1}$, where $\Delta V_{1,2}$ is the minimum energy barrier to cross when going from x_1 to x_2 and $\Delta V_{2,1}$ the one when going from x_2 to x_1, and (ii) this lower bound is only attained if γ is the MEP since in this case γ is either parallel to ∇V (on the way uphill) or antiparallel to ∇V (on the way downhill), meaning that $\int_{\gamma_u} |\nabla V| ds = \int_{\gamma_u} (dV/ds) ds = \Delta V_{1,2}$ on the piece γ_u of γ going uphill and $\int_{\gamma_d} |\nabla V| ds = -\int_{\gamma_d} (dV/ds) ds = \Delta V_{2,1}$ on the piece γ_d of γ going downhill.[8]

The action Eq. (6) had already been derived in Ref. 7 by an argument similar to the one earlier: To the best of our knowledge, these authors were the first to realize that the MEP can be geometrically characterized as the curve minimizing an action for curves. An algorithm for minimizing this action was also proposed in Ref. 7. Next, we propose a more efficient algorithm based on gMAM.

First, we need to reintroduce a parametrization of the curve γ. In the spirit of the string method, we will use $\gamma = \{\varphi(\alpha) : \alpha \in [0,1]\}$, where $\varphi(\alpha)$ satisfies $|\varphi'(\alpha)| \equiv$ length of $\gamma = cst$ [here the prime in $\varphi'(\alpha)$ denotes the derivative with respect to α, so $\varphi'(\alpha)$ is a vector tangent to the curve γ at the point $\varphi(\alpha)$]. This choice amounts to parametrizing γ by normalized arclength. Expressed in terms of φ, the action Eq. (6) becomes

$$\hat{S}(\gamma) = \hat{S}(\varphi) = \int_0^1 |\nabla V(\varphi(\alpha))| |\varphi'(\alpha)| d\alpha. \tag{8}$$

The minimizer of this action must satisfy the Euler–Lagrange equation $\delta\hat{S}/\delta\varphi = 0$, where $\delta\hat{S}/\delta\varphi$ denotes the functional derivative of $\hat{S}(\varphi)$. A direct calculation using $|\varphi'(\alpha)| \equiv cst$ [and, hence, $|\varphi'(\alpha)|' \equiv 0$] shows that it can be expressed as

$$\delta\hat{S}/\delta\varphi = -g\varphi'' + g^{-1} \nabla \nabla V(\varphi) \nabla V(\varphi) - g'\varphi', \tag{9}$$

where $\nabla\nabla V$ is the Hessian tensor of V and we defined

$$g(\alpha) = |\nabla V(\varphi(\alpha))| / |\varphi'(\alpha)|. \tag{10}$$

To find a solution of $\delta\hat{S}/\delta\varphi = 0$, we will use a preconditioned steepest-descent algorithm in the direction of $-g\delta\hat{S}/\delta\varphi$ with a reparametrization step to enforce the constraint that $|\varphi'(\alpha)| \equiv cst$. Specifically, given an initial guess for the path, $\varphi^0(\alpha)$ with $\varphi^0(0) = x_1$ and $\varphi^0(1) = x_2$, we discretize it into $N+1$ points as $\varphi_j^0 = \varphi^0(j/N)$ with $j = 0, \ldots, N$, and for $k \geq 0$ we update these points by iterating upon the following steps until convergence:

(1) For $j = 1, \ldots, N-1$, we obtain φ_j^* by solving the tridiagonal linear algebraic system

$$\frac{\varphi_j^* - \varphi_j^k}{\tau} = (g_j^k)^2 \frac{\varphi_{j+1}^* + \varphi_{j-1}^* - 2\varphi_j^*}{(1/N)^2} - \nabla\nabla V(\varphi_j^k) \nabla V(\varphi_j^k) + g_j^k \frac{(g_{j+1}^k - g_j^k)(\varphi_{j+1}^k - \varphi_j^k)}{2/N^2} + g_j^k \frac{(g_j^k - g_{j-1}^k)(\varphi_j^k - \varphi_{j-1}^k)}{2/N^2}, \tag{11}$$

with the boundary conditions $\varphi_0^* = x_1$ and $\varphi_N^* = x_2$. Here $\tau > 0$ is a parameter playing the role of the updating time step and $g_j^k = |\nabla V(\varphi_j^k)| / L^k$, where $L^k = \sum_{j=1}^N |\varphi_j^k - \varphi_{j-1}^k|$ is the length of the curve (and, hence, the discrete approximation of $|\varphi'|$ since $|\varphi'| \equiv$ length of γ).

(2) We interpolate a piecewise linear curve through $\{\varphi_j^*\}_{j=0,\ldots,N}$ and redistribute points at equal distance along this curve to obtain $\{\varphi_j^{k+1}\}_{j=0,\ldots,N}$.

This algorithm is gMAM, specialized to the action Eq. (6). Step 1 is the preconditioned steepest-descent step, corresponding to motion in the direction of $-g\delta\hat{S}/\delta\varphi$ with a time step τ since the right-hand side of Eq. (11) is $-g_j^k$ times the gradient of the following discretized approximation of Eq. (8): $E(\varphi^k) = \frac{1}{2}\sum_{j=1}^N |\nabla V(\varphi_j^k) + \nabla V(\varphi_{j-1}^k)| L^k / N$. We multiplied $\delta\hat{S}/\delta\varphi$ by $g \geq 0$ in order to avoid problems at the critical points where $|\nabla V| = 0$ since the factor g^{-1} in $\delta\hat{S}/\delta\varphi$ diverges at these points. The preconditioning amounts to using the semi-implicit updating step in Eq. (11) which guarantees that the code is stable with a time step τ which is independent of N (i.e., τ does not have to be decreased when N is increased). The system Eq. (11) can be solved in $O(N)$ operations by the Thomas algorithm (see, e.g., Sec. 2.9 in Ref. 10). Step 1 also requires the computation of the product $\nabla\nabla V \nabla V$. This can be done without computing the Hessian $\nabla\nabla V$ by using the discrete approximation

$$\varepsilon^{-1}(\nabla V(\varphi_j^k + \varepsilon \nabla V(\varphi_j^k)) - \nabla V(\varphi_j^k)) = \nabla\nabla V(\varphi_j^k) \nabla V(\varphi_j^k) + O(\varepsilon), \tag{12}$$

where $\varepsilon > 0$ is a small parameter.

Step 2 is the same as the reparametrization step used in the string method and is added to enforce the discrete equivalent of $|\varphi'| \equiv cst$. For completeness, we recall the details of this reparametrization step at the end of this paper.

Going back to the algorithm proposed in Ref. 7 we can now explain in which way the earlier algorithm is superior to it. Instead of imposing the constraint that $|\varphi'| \equiv cst$ by a reparametrization step as we do, in Ref. 7 this constraint was imposed approximately by adding the penalty term $\lambda \int_0^1 |\varphi'|^2 d\alpha$ to Eq. (8), where $\lambda > 0$ is a parameter. The problem with this strategy is that if λ is too small then the constraint that $|\varphi'| \equiv cst$ is not imposed accurately enough, which may lead to numerical instabilities and/or inaccuracies. But if λ is taken too big, the additional term makes the problem stiff which slows down convergence. Here we avoid these difficulties by enforcing the constraint $|\varphi'| \equiv cst$ by a reparametrization step.

Because of the way we approximate the derivatives in Eq. (11) and because we use linear interpolation, our algo-

Downloaded 11 Jan 2009 to 67.159.68.103. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

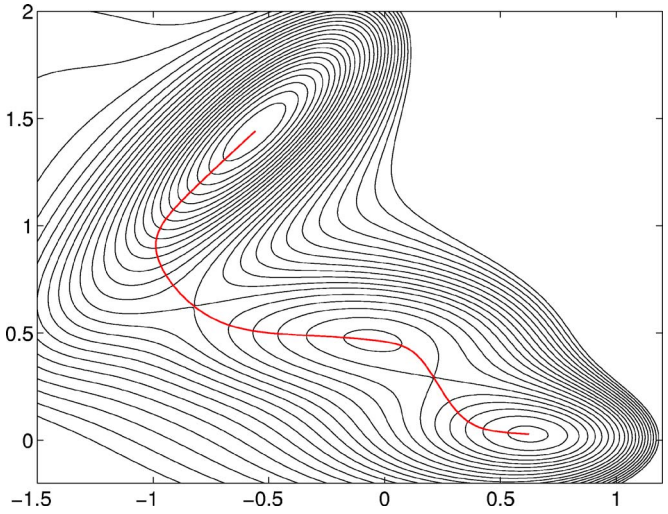

FIG. 1. (Color online) The MEP in the Müller potential calculated with $N=10^2$ discretization points.

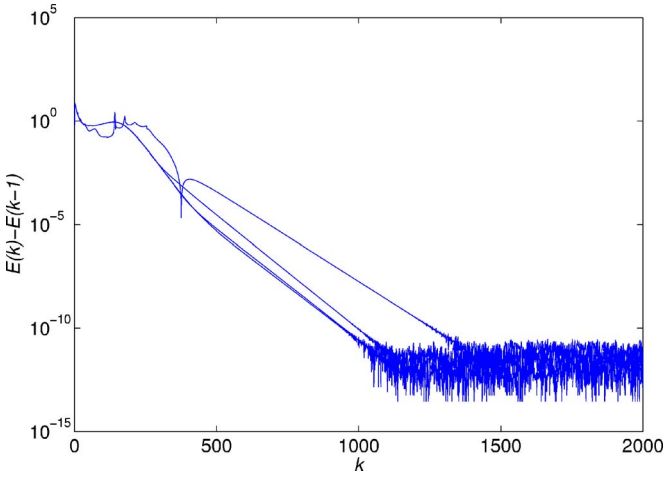

FIG. 2. (Color online) Convergence rate at $N=20,10^2,10^3,10^4$ (the convergence rate increases slightly with N, i.e., the curve decreasing the fastest is the one for $N=10^4$). The quantity plotted here is $E(\varphi^k)-E(\varphi^{k-1})$ where k is the iteration number and $E(\varphi^k)=\frac{1}{2}\Sigma_{j=1}^{N}|\nabla V(\varphi_j^k)+\nabla V(\varphi_{j-1}^k)|L^k/N$ is the discrete approximation of the functional $\hat{S}(\varphi)$. It should be stressed that the large number of steps shown here are those needed to reach convergence at machine precision, but at the end of the calculation the changes in φ_j^k are minute, and for all practical purposes the calculation could be stopped after 200 iterations or so.

rithm is second-order accurate in $1/N$, meaning that the piecewise linear curve interpolated across the points φ_j^k is, at convergence, $O(1/N^2)$ away from a true MEP. Since the time step τ can be chosen independently of N, the number of steps to achieve convergence to that accuracy is $O(\log N)$ (assuming exponential convergence in time), and since every iteration requires $O(N)$ operations, the total cost of the scheme is $O(N \log N)$. If the required accuracy is $\delta>0$, then to achieve this accuracy we must take $N=O(\delta^{-1/2})$ and the cost is $O(\delta^{-1/2} \log \delta^{-1})$. The same efficiency can be achieved with the string method, but it requires one to use higher-order procedures for the time stepping and the reparametrization step (see Ref. 4 for details). As for NEB, its cost is higher, $O(\delta^{-1} \log \delta^{-1})$, because stability requirements impose further constraints on the time step.[4]

To test our algorithm, we used the two-dimensional Müller potential.[9] As the initial condition for φ we took a straight line between the two minima of V located at $x_1 \approx (-0.56, 1.44)$ and $x_2 \approx (0.62, 0.03)$. We tested the scheme at various values of N between 20 and 10^4. We always used the same time step $\tau=10^{-7}$ (larger time steps led to instabilities at all N) and we set $\varepsilon=10^{-7}$. Figure 1 shows the MEP obtained using $N=10^2$. A simple measure for the error of the MEP is

$$\text{error}(N) = \max_{j=0,\ldots,N} \text{dist}(\varphi_j^k, \text{true MEP}), \tag{13}$$

where the true MEP was inferred from the computation with $N=10^4$. We obtained $\text{error}(100)=4.89\times10^{-4}$. Figure 2 shows the convergence rate of the algorithm at various values of N. As expected, the number of steps needed until convergence is roughly independent of N.

In conclusion, we have proposed a new scheme to identify the MEP based on gMAM. Because this scheme is slightly more complicated to implement than the string method and is similar in efficiency, it is not clear whether it has any definite edge on the string method. Yet, the new algorithm has the advantage of being based on a variational formulation. This means that its efficiency could potentially be improved by using more sophisticated minimization techniques than the preconditioned steepest-descent scheme we used, like, e.g., preconditioned conjugate-gradient or multigrid methods.[10] These options are currently under investigation.

Details of the reparametrization step 2. Denote by $L(j)$ the length up to φ_j^* of the piecewise linear curve interpolated across these points, i.e., $L(0)=0$ and for $j=1,\ldots,N$,

$$L(j) = \sum_{j'=1}^{j} |\varphi_{j'}^* - \varphi_{j'-1}^*|.$$

Then $L(N)$ is the total length of the piecewise linear curve and thus $\ell(j)=L(N)j/N$, $j=1,\ldots,N$, is the distance along this curve from the starting point x_1 up to the point at which we need to put the new jth point φ_j^{k+1} in order to make these new points equidistant along the curve. This is done by setting $\varphi_0^{k+1}=\varphi_0^*$ and $\varphi_N^{k+1}=\varphi_N^*$ (by construction the end points are unaffected by the reparametrization step) and for $j=1,\ldots,N-1$ taking

$$\varphi_j^{k+1} = \varphi_{i(j)}^* + (\ell(j) - L(i(j)))\frac{\varphi_{i(j)+1}^* - \varphi_{i(j)}^*}{|\varphi_{i(j)+1}^* - \varphi_{i(j)}^*|},$$

where $i(j) \in \{0,\ldots,N-1\}$ is the index fulfilling $L(i(j)) < \ell(j) \le L(i(j)+1)$. We then have $|\varphi_{j+1}^{k+1}-\varphi_j^{k+1}|=|\varphi_j^{k+1}-\varphi_{j-1}^{k+1}|$ up to $O(1/N^2)$ for all $j=1,\ldots,N-1$. Note that the reparametrization step can be done in $O(N)$ operations.

We thank Ron Elber for pointing out Ref. 7 to us. We are also grateful to Giovanni Ciccotti for useful comments. This work was partially supported by NSF Grant Nos. DMS02-09959 and DMS02-39625 and by ONR Grant No. N00014-04-1-0565.

Downloaded 11 Jan 2009 to 67.159.68.103. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp

[1] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998).
[2] G. Henkelman and H. Jónsson, J. Chem. Phys. **113**, 9978 (2000).
[3] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[4] W. E, W. Ren and E. Vanden-Eijnden, J. Chem. Phys. **126**, 164103 (2007).
[5] M. Heymann, Ph.D. thesis, New York University, 2007.
[6] M. Heymann and E. Vanden-Eijnden, Commun. Pure Appl. Math. (in press).
[7] R. Olender and R. Elber, J. Mol. Struct.: THEOCHEM **398–399**, 63 (1997); Proceeding of the WATOC Symposium.
[8] This argument assumes that there is only one saddle point and no other critical point along the MEP joining x_1 to x_2, but it can be easily generalized to more complicated situations and the conclusion remains valid: The minimizers of Eq. (6) are MEPs.
[9] K. Müller, Angew. Chem. **92**, 1 (1980).
[10] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations—An Introduction*, 2nd ed. (Cambridge University Press, Cambridge, 2005).

Downloaded 11 Jan 2009 to 67.159.68.103. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/jcp/copyright.jsp