Published in final edited form as:

Theor Chem Acc. 2017 January ; 136: . doi:10.1007/s00214-016-2041-3.

A minimization principle for transition paths of maximum flux for collective variables

Robert D. Skeel[1,*], **Ruijun Zhao**[2], and **Carol Beth Post**[3]

[1]Department of Computer Science, Purdue University, West Lafayette, IN, 47906 USA

[2]Department of Mathematics and Statistics, Minnesota State University, Mankato, MN, 56001 USA

[3]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN, 47906 USA

Abstract

Considered is the construction of transition paths of conformational changes for proteins and other macromolecules, using methods that do not require the generation of dynamics trajectories. Special attention is given to the use of a reduced set of collective variables for describing such paths. A favored way to define transition paths is to seek channels through the transition state having cross sections with a high reactive flux (density of last hitting points of reactive trajectories). Given here is a formula for reactive flux that is independent of the parameterization of "collective variable space." This formula is needed for the principal curve of the reactive flux (as in the revised finite temperature string method) and for the maximum flux transition (MaxFlux) path. Additionally, a resistance functional is derived for narrow tubes, which when minimized yields a MaxFlux path. A strategy for minimization is outlined in the spirit of the string method. Finally, alternative approaches based on determining trajectories of high probability are considered, and it is observed that they yield paths that depend on the parameterization of collective variable space, except in the case of zero temperature, where such a path coincides with a MaxFlux path.

Keywords

Reaction path; Minimum energy path

1. Introduction

The calculation of transition paths of conformational change for proteins and other biomolecules is a subject of great interest, given the difficulty of experimental characterization of paths and their functional importance. Protein function often requires transitioning from one stable state to another. Experiment characterizes the structure of these stable states, but the structural changes required for going between the states are generally unknown. This article concerns the question of how to define and characterize the objects

Phone +1-765-494-9025, rskeel@purdue.edu.

that are to be computed. Informally, the goal is to compute representative paths from one stable state A to another stable state B. A representative path is the "center" of an isolated cluster of trajectories. The aim of this article is to identify sound formulations of such paths together with some consideration on how to compute them.

Transition paths might not cluster adequately—in full configuration space. Moreover, reduction to a small number of variables is needed for comprehending the result. Therefore, assume there is a smaller set of *collective variables*, functions of the configuration x,

$$\zeta_1=\xi_1(x),\zeta_2=\xi_2(x),\ldots,\xi_\nu=\xi_\nu(x),$$

abbreviated as $\zeta = \xi(x)$, such that in ζ-space, reactive trajectories cluster into one or several distinct isolated channels connecting two separated subsets A_ξ and B_ξ of collective variable space corresponding to A, B. As an example, one might use ϕ and ψ torsion angles.

There are several general approaches to this problem. One approach is to generate an ensemble of reactive trajectories, using (say) transition path sampling [2], and then cluster them. A different approach is to directly construct a representative path, by optimizing a functional that measures its quality or by some similar process. (For this approach, there remains the question of finding a multiplicity or network of transition paths.) Hybrid approaches, e.g., milestoning [25] and Markov state models [22], generate shorter trajectories to build discrete models, which can then be processed to find transition paths.

Considered in this article is the direct construction of transition paths. There are several ways of defining such paths:

1. Centers of channels of high reactive flux, i.e., channels with high concentrations *of distinct* reactive trajectories. It is natural [16] to construct such channels with cross sections that are isocommittors (see Sect. 3). A maximum flux transition (MaxFlux) path uses a point of maximum flux on each [10, 27]. The principal curve, suggested for the finite temperature string (FTS) method [10], uses the mean on each cross section.
2. Centers of channels of high reactive probability density. The original FTS method [16] employs the principal curve.
3. Trajectories of greatest probability, based on minimizing the Onsager– Machlup action [19], also called the Friedlin–Wentzell action [6].
4. Reaction paths based on mean first-passage times [21].

The second of these approaches has been supplanted by the first. The third one is shown in Sect. 6 to be flawed, except in the special case of zero temperature, where it coincides with the first.

The focus of this article is the MaxFlux path, which can be described as the center of a narrow tube of uniform cross section that maximizes the flux of reactive trajectories. A main result of this article is the derivation of a (correct) formulation of this concept for collective

variables. It is beyond the scope of this article to specify how to choose collective variables or, for the most part, how to compute MaxFlux paths.

The principal contributions of this article are as follows: First, several existing formulations for defining transition paths are shown to be unphysical in the sense that they depend on the coordinates chosen to represent the degrees of freedom in the system. As an example, it should not matter whether distances or their squares are chosen as coordinates. Second, a formula is given for the flux intensity in collective variables thus providing a generalization of the *revised* string method [10] to collective variables. Third, the MaxFlux path *in collective variables* is formulated as a minimization principle, making it easier to design quality numerical methods for computing the path. With a functional to minimize, it is possible to check whether an update is in a direction of descent. Also, functionals are relatively straightforward to discretize accurately, compared to differential equations. The latter often requires “fixes”, such as upwind differencing.

1.1. Prior work

The seminal article by Berkowitz, Morgan, McCammon, and Northrup [1] introduces the MaxFlux path as a minimum resistance path for Brownian dynamics trajectories, suggesting an analogy with electrical circuits. The path is defined as the minimum of a generalized resistance functional, and Euler–Lagrange equations are derived using the calculus of variations. However, the suggested generalization to collective variables is not invariant under a change of variables in collective variable space, see Sect. 4.2. Later a “MaxFlux” algorithm based on minimizing a discretization of this functional is proposed [13]. A yet later article [14] contrasts the MaxFlux path with the minimum free energy path (MFEP), which is its zero-temperature limit: Fig. 2 of that article illustrates for alanine dipeptide how the MFEP can develop a cusp at an intermediate local minimum, and Fig. 6a shows the density of dynamical trajectories obtained from transition path sampling and how they agree much better with the MaxFlux path.

The committor concept is an attractive tool for defining transition paths. The seminal article on the finite temperature string method [16] constructs the path from the point on each committor that is the arithmetic mean of hitting points from reactive trajectories. The method is later revised [10] to use last hitting points (reactive flux) instead, building on an expression derived for last hitting point density [9, 23]. It is noted in Ref. [10], Eq. (48), that, for cartesian coordinates, if the point of maximum reactive flux is used for the center instead of the arithmetic mean, this results in a MaxFlux path.

Based on the maximum reactive flux criterion, Ref. [27] derives a differential equation (but no minimization principle) for the MaxFlux path in collective variables, which differs from that of Ref. [1]. It is noted that the proposed path is independent of how collective variable space is parameterized, unlike that of Ref. [1]. Specifically, if one uses instead variables η, where $\zeta = \chi(\eta)$, the two paths that are determined also satisfy this relation. A numerical algorithm is constructed, tested, and later applied [12] to a double basin Gō model of CDK2 kinase, involving 813 DOFs reduced to 7 collective variables, obtaining the same final results for 3 different initial paths.

Ref. [4], Sect. 2.1, gives a formula for the “MaxFlux functional”, which is an improvement over the integral of Ref. [1], However, the revised formula is still not invariant under a change of variables in collective variable space, see Sect. 4.2. That same article also gives, in Fig. 6b, a vivid visualization of the flux for phi/psi angles for alanine dipeptide. Finally, an intriguing method for finding globally optimal MaxFlux paths is presented, further described in Sect. 4.4.

There is a large literature on computational methods, which is not surveyed here except for mention of continuous path dynamics [5] as a strategy for determining transition paths.

Olender and Elber [19] propose defining as a transition path a Brownian dynamics trajectory of maximum probability, which under some assumptions is a path of steepest descent, often called a minimum energy path. Credit for this idea is attributed [3] to Wolynes [26]. This has been generalized to a minimum free energy path [16]. A more recent article [11] proposes an alternative functional (involving the Laplacian of the free energy) to represent relative probabilities of different trajectories.

1.2. Outline

Section 2 presents a Brownian dynamics model for the dynamics of the collective variables. It shows how the inverse of the diffusion tensor defines a natural metric for collective variable space.

Section 3 derives the reactive flux (last hitting point density) for collective variables. This is useful not only as one way of deriving the MaxFlux path but also for obtaining the collective variable formulation of the principal curve of reactive flux, which is the arithmetic mean of the reactive flux on each isocommittor (used by the revised finite temperature string method).

Section 4 obtains a formula for the flow rate of distinct reactive trajectories through a narrow tube of uniform cross section. The flow rate is the *reciprocal* of the “resistance”

$$R[Z]=\frac{C_\xi}{d_0 a}\int_0^1 e^{\beta F(Z)}(\det G(Z))^{1/2}|Z_s|_Z \mathrm{d}s, \quad (1)$$

where $\zeta = Z(s)$, $0 \le s \le 1$, is the path, $Z_s(s) = (\mathrm{d}/\mathrm{d}s)Z(s)$, β is inverse temperature, $F(\zeta)$ is the free energy function (potential of mean force), $G(\zeta)$ is the metric tensor, $|Z_s(s)|_{Z(s)}\mathrm{d}s = (Z_s(s)^\top G(Z(s))Z_s(s))^{1/2}ds$ is arc length, and the prefactor $C_\xi/(d_0 a)$ is a constant independent of Z. The flow rate given by this formula is *independent of how collective variable space is parameterized*. It is shown that the flow rate is maximized by a MaxFlux path. The variational characterization of the MaxFlux path enables a global optimum to be determined via the Hamilton–Jacobi equation.

Section 5 suggests an approach to minimizing the integral by embedding in dynamics. The computational advantage is explained, and the mathematical soundness is confirmed.

Section 6 demonstrates that the maximum probability criterion based on the Onsager–Machlup action is flawed except in the zero-temperature limit. Only in this case is the path independent of the parameterization of collective variable space.

1.3. Discussion and conclusions

A main contention of this article is that the reactive flux concept, because it is a geometric concept, is currently the best approach to constructing representative transition paths without actually generating dynamics trajectories. In particular, the idea of instead selecting a path of maximum probability based on the Onsager–Machlup action is problematic. For finite temperature, such a path is not a geometric construct (since it depends on the parameterization of collective variable space). Also, the most probable path depends on the duration T of the trajectories, and it is not meaningful to compare probabilities for different durations. For zero temperature and infinite duration, the most probable path coincides with the MaxFlux path. Less importantly, the functional that is to be minimized to determine a most probable path contains the gradient of the free energy function, which is computationally inconvenient for minimization.

The principal curve of reactive flux is an attractive object for representing a cluster of reactive trajectories. Nonetheless, a MaxFlux path is simpler and less costly to compute. In any case, the latter can serve as a starting point for more ambitious calculations. The flow rate expression (1) enables comparison of flow rates among several isolated channels of comparable widths. Moreover, this expression can be interpreted as a distance measure of the path. Therefore, the MaxFlux path can be viewed as a geodesic, enabling the use of the Hamilton–Jacobi equation to find a globally optimal path—for a modest number of collective variables. Additionally, a minimization principle can contribute to the construction of a robust computational method by providing a search direction that points downhill.

Another contention of this article is that the temperature used to construct a path in collective variable space should be the same as that used to define the free energy function. And it happens that finite temperature paths are easier to compute, due to the smoothing property of the second derivative term in the Euler–Lagrange equation.

Existing formulations have undoubtedly produced useful results and have been instrumental for more recent developments. Nonetheless, going forward, it is attractive to use more rigorous formulations that have less cause for concern.

Two computational issues are addressed here. The first is difficulties that might arise from arbitrariness in the curve parameterization. For this, the use of constraints is more straightforward than the selection of an appropriate penalty term in the objective function. The second is difficulties encountered in the use of standard accelerated continuous optimization techniques to minimize an integral of an exponential function. This can be circumvented by building an artificial dynamical system that pushes the path in a direction of descent for the functional, though generally not the direction of steepest descent. Additionally, this requires only values of derivatives of the free energy function, which can be obtained in the typical way by approximating a Dirac delta function by a highly peaked Gaussian and then sampling by (restrained) molecular dynamics.

2. Brownian dynamics model

Assume the molecular system obeys Newtonian dynamics with potential energy function $U(x)$ and diagonal matrix M of atomic masses. Then positions x and momenta p satisfy $x = X(t)$, $p = P(t)$ where $(\mathrm{d}/\mathrm{d}t)X(t) = M^{-1}P(t)$ and $(\mathrm{d}/\mathrm{d}t)P(t) = -\nabla U(X(t))$. Initial values are drawn from a Boltzmann–Gibbs distribution $\rho(x,p)$: positions x from probability density $\propto e^{-\beta u(x)}$ and momenta p from a Gaussian distribution.

The probability density function for $\xi(x)$ is

$$\rho_\xi(\zeta)=\langle\delta(\xi(x)-\zeta)\rangle$$

where $\langle O(x)\rangle = \int\int O(x)\rho(x,p)\mathrm{d}x\mathrm{d}p$ and $\delta(\zeta) = \delta(\zeta_1)\delta(\zeta_2)\cdots\delta(\zeta_\nu)$. An "effective energy" function $F(\zeta)$, sometimes called *a free energy function*, is defined by

$$C_\xi e^{-\beta F(\zeta)}=\rho_\xi(\zeta) \quad 2$$

where C_ξ is an unspecified constant.

Assume the collective variables relax more slowly than other DOFs and momenta, so that paths are well approximated by Brownian dynamics trajectories $\zeta = \hat{\zeta}(\tau)$ defined by (Ref. [16], Eq. (17))

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\hat{\zeta}=-D(\hat{\zeta})\beta\nabla F(\hat{\zeta})+\mathrm{div}D(\hat{\zeta})+\sqrt{2}D(\hat{\zeta})^{1/2}\frac{\mathrm{d}}{\mathrm{d}\tau}W(\tau)$$

where $W(\tau)$ is a standard Wiener process and $D(\zeta)$ is a diffusion tensor given by

$$D(\zeta)=d_0 m_{\mathrm{tot}}\left\langle\frac{\partial\xi}{\partial x}(x)M^{-1}\frac{\partial\xi}{\partial x}(x)^\top\right\rangle_{\xi(x)=\zeta}. \quad 3$$

Here, m_{tot} is the sum of the masses, d_0 is an unknown scaling factor, and $\langle O(x)\rangle_{\xi=\zeta}$ is the expectation for the conditional probability density $\rho(x,p|\xi(x) = \zeta)$:

$$\langle O(x)\rangle_{\xi=\zeta}=\frac{\langle\delta(\xi(x)-\zeta)O(x)\rangle}{\delta(\xi(x)-\zeta)}.$$

An alternative to this analytical formulation of Brownian dynamics is a first-principles empirical construction employing "swarms of trajectories" [15, 17, 20].

It is useful to have a metric

$$|\delta\zeta|_\zeta=(\delta\zeta^\top G(\zeta)\delta\zeta)^{1/2}, \quad 4$$

with metric tensor $G(\zeta)$, to measure distance from ζ to $\zeta + \delta\zeta$, in collective variable space. (The subscript ζ is to be understood as a value that provides the argument of G.) The use of a tensor ensures that a change of variables *within* collective variable space, $\zeta = x(\eta)$, leaves distances (and angles) unchanged.

The choice

$$G(\zeta)=d_0 D(\zeta)^{-1} \quad 5$$

proves to be convenient. This has a useful geometric interpretation. Each point ζ of collective variable space represents a manifold $\mathcal{M} : \xi(x) = \zeta$, of dimension $3N - \nu$ in cartesian configuration space. Let x be a point on $\mathcal{M}$ and let $x + \delta x$ be the nearest point to x on the manifold $\mathcal{M}' : \xi(x') = \zeta + \delta\zeta$. The natural metric for cartesian space is the mass-weighted RMS norm $\|\delta x\| = (\delta x^\top M \delta x / m_{\text{tot}})^{1/2}$. It can be shown that $\|\delta x\| = (\delta\zeta^\top \Gamma(x)\delta\zeta)^{1/2}$ where

$$\Gamma(x)=\left(m_{\text{tot}}\frac{\partial\xi}{\partial x}(x)M^{-1}\frac{\partial\xi}{\partial x}(x)^\top\right)^{-1}.$$

That is, $(\delta\zeta^\top \Gamma(x)\delta\zeta)^{1/2}$ is the distance between the two infinitesimally close manifolds at the point x. The tensor $G(\zeta)=\langle\Gamma(x)^{-1}\rangle^{-1}_{\xi(x)=\zeta}$ is a Boltzmann-weighted harmonic average of $\Gamma(x)$ over all x on $\mathcal{M}$. Thus does the distance $|\delta\zeta|_\zeta$ measure cartesian RMSD between two infinitesimally close manifolds in configuration space.

3. Centers of high reactive flux

Given metastable states A_ξ and B_ξ, the problem is to find a path in collective variable space through the center of a channel with a high flow rate of distinct *reactive* trajectories.

To measure the progress of a transition, there is a natural reaction coordinate, known as the *committor:* The committor $q(\zeta)$ is the probability that a trajectory starting from ζ reaches B_ξ before A_ξ. It is known (Ref. [16], Eq. (15)) that $f = q$ minimizes $I(f)$ subject *to* $f(\zeta) = 0$ on the boundary of A_ξ and $f(\zeta) = 1$ on the boundary of B_ξ where

$$I(f)=\int_\Omega e^{-\beta F(\zeta)}\nabla f(\zeta)^\top G(\zeta)^{-1}\nabla f(\zeta)d\zeta. \quad 6$$

3.1. Reactive flux

The flow rate (for last hitting points) across an arbitrary surface Σ is given by (Ref. [18], Eq. (A12))

$$\text{flow rate} = \int_{\sum} J(\zeta)^{\top} \hat{n}(\zeta) \mathrm{d}S_{\zeta} \quad 7$$

where $\hat{n}(\zeta)$ is the unit outward normal at ζ and

$$J(\zeta) = \rho_{\xi}(\zeta) D(\zeta) \nabla q(\zeta) = d_0 \rho_{\xi}(\zeta) G(\zeta)^{-1} \nabla q(\zeta)$$

is the probability current density (Ref. [18], Eqs. (A13), (A14), (14), (A1)). Flux is flow rate per unit surface area, and a formula for surface area for collective variables is shown in Appendix “Area metric” to be

$$a = \int_{\sum} (\det G(\zeta))^{1/2} |G(\zeta)^{-1} \hat{n}(\zeta)|_{\zeta} \mathrm{d}S_{\zeta}. \quad 8$$

Applying Eqs. (7) and (8) to an infinitesimal surface $\delta\Sigma$ gives the reactive flux

$$j(\zeta) = d_0 \rho_{\xi}(\zeta) (\det G(\zeta))^{1/2} |G(\zeta)^{-1} \nabla q(\zeta)|_{\zeta}. \quad 9$$

Assume the bulk of reactive trajectories cluster into isolated channels. For any such channel, define the path to intersect each isocommittor at a center of the reactive flux $j(\xi)$. The center might be defined to be (i) a local maximum of the reactive flux or (ii) its centroid. The first characterization, adopted in Ref. [27], is from a lecture by Vanden-Eijnden building on Ref. [18], Sec. III.C, and published in Ref. [10], Eq. (48), for cartesian coordinates. For the centroid to be meaningful, its calculation should be restricted to some subset of the isocommittor. It could be a tube T of constant radius around an initial approximation or around the current approximation—provided the radius of T is not too small. Let $\zeta = Z(s)$ be a path along which the committor increases, and let $\Sigma(s)$ be the isocommittor passing through $\zeta = Z(s)$, Each point $Z(s)$ on the principal curve minimizes $\int_{T(s)} \mathrm{dist}(\zeta, Z(s))^2 j(\zeta) \mathrm{d}S_{\zeta}$ where $T(s)$ is the intersection of T with $\Sigma(s)$. The centroid of the reactive flux, in the case of cartesian coordinates, is used by the revised FTS method (Ref. [10], Sec. 3.6).

The original finite temperature string method (Ref. [7], Sec. II.B) uses instead the reactive probability (for all hitting points) $q(\zeta)(1 - q(\zeta))\rho_{\xi}(\zeta)\mathrm{d}\zeta$. With a volume element $(\det G(\zeta)^{1/2}\mathrm{d}\zeta$, this gives a reactive probability density proportional to $\mathrm{e}^{-\beta F(\zeta)}(\det G(\zeta)^{-1/2}$ on an isocommittor $q(\zeta) = \text{constant}$.

3.2. Channels of restricted width

Assume that the channel is narrow enough to justify the following simplification:

Assumption 1 Each isocommittor is planar within the channel.

In particular, it is assumed that these planes do not intersect within the channel. Assume a trial path $\zeta = Z(s)$ is given for which $q(Z(s))$ is strictly increasing. Then the isocommittor through $Z(s)$ is a plane

$$\Pi(s):\hat{n}(s)^{\top}(\zeta - Z(s))=0, \quad 10$$

where $\hat{n}(s)$ is a unit outward normal. This assumption has the dramatic effect of converting the high-dimensional problem of minimizing $I(f)$ over all $f(\zeta)$, $\zeta \in \Omega$, to the one-dimensional problem of solving for $\hat{n}(s)$ and $Z(s)$, $1 \le s \le 1$.

It is common to make one further simplification:

Assumption 2 The flux $J(Z)$ is parallel to the direction Z_s of the path.

The result of these two assumptions is shown in Appendix “Path of maximum reactive flux” to be the set of ordinary differential equations

$$(I - P(Z, Z_s))\left(-\frac{\nabla_\zeta \gamma(Z)}{\gamma(Z)} - \frac{\nabla_\zeta c(Z, Z_s)}{c(Z, Z_s)} + - \frac{(\nabla_\omega c(Z, Z_s))_s}{c(Z, Z_s)}\right) = 0 \quad 11$$

for $\zeta = Z(s)$ where

$$\begin{aligned} \gamma(\zeta) &= e^{\beta F(\zeta)} (\det G(\zeta))^{1/2}, \\ c(\zeta, \omega) &= |\omega|_\zeta = (\omega^{\top} G(\zeta) \omega)^{1/2}, \\ P(\zeta, \omega) &= c(\zeta, \omega)^{-2} G(\zeta) \omega \omega^{\top}. \end{aligned} \quad 12$$

(Eq. (11) above is the same as Eq. (11) of Ref. [27].) Note that $P(Z,Z_s)$ is a projector, making the equations underdetermined. This is due to the arbitrariness of the parameterization.

The $\beta \to \infty$ limit of Eq. (11) for Z, but with β held fixed in $F(Z;\beta)$ and $D(Z;\beta)$, is $(I - P(Z,Z_s))\nabla F = 0$, which is equivalent to the condition $Z_s \parallel G(Z)^{-1}\nabla F(Z)$ for the minimum free energy path given in Sect. 6.

4. Variational principle for MaxFlux paths

Given here is a derivation from a variational principle of the MaxFlux path given by Eq. (11). Equivalence is confirmed at the end of this section.

The problem considered here is to find a *narrow* tube in collective variable space from A_ξ to B_ξ with a circular cross section of specified hyperarea a for which the flow rate of distinct *reactive* trajectories is maximized under the assumption that all dynamics is confined to the tube. This is different from generating a (huge) ensemble of trajectories and finding a tube that will contain the largest number of them. Rather, the goal is to find a set of points visited by the largest number of reactive trajectories.

Define the tube in terms of an enclosed trial path $\zeta = Z(s)$, $0 \le s \le 1$, connecting A_ξ and B_ξ. It is desirable to define a cross section of the tube as an intersection of the tube with an isocommittor. With this definition the ends of the tube will be surfaces of A_ξ and B_ξ, since they are isocommittors. Use of the metric $|\mathrm{d}\zeta|_\zeta$ to measure cross-sectional area ensures that the flow rate is invariant under a change of coordinates.

4.1. Flow rate through a narrow tube

Consider first the general case with no restriction on dynamics. Assume a path $\zeta = Z(s)$ for which $q(Z(s))$ is strictly increasing, and define $\hat{q}(s)$ and $\sigma(\zeta)$ by

$$\hat{q}(s)=q(Z(s)) \text{ and } \hat{q}(\sigma(\zeta))=q(\zeta). \quad 13$$

Appendix “Narrow tube flow rate formula” derives the formula

$$\text{flow rate}=C_\xi d_0\left(\int_0^1 \frac{\mathrm{d}s}{Q(s)}\right)^{-1} \quad 14$$

where

$$Q(s)=\int_{\sum(s)} e^{-\beta F(\zeta)}\nabla\sigma(\zeta)^\top G(\zeta)^{-1}\hat{n}(\zeta)\mathrm{d}S_\zeta. \quad 15$$

Let $\mathcal{T}(s)$ be the intersection of the tube with $\Sigma(s)$. For a sufficiently narrow tube, $\nabla\sigma(\zeta)$, $G(\zeta)$, and $F(\zeta)$ are nearly constant on $\zeta \in \mathcal{T}(s)$. Hence, the formula (15) for $Q(s)$ simplifies to

$$Q(s)=e^{-\beta F(Z)}\nabla\sigma(Z)^\top G(Z)^{-1}\hat{n}(Z)\int_{T(s)}\mathrm{d}S_\zeta.$$

Also, formula (8) gives the area of a cross section of the tube as

$$a=|G(Z)^{-1}\hat{n}(Z)|_Z(\det G(Z))^{1/2}\int_{T(s)}\mathrm{d}S_\zeta.$$

Therefore, using $\hat{n}(Z) \parallel \nabla\sigma(Z)$, one gets

$$Q(s)=ae^{-\beta F(Z)}|G(Z)^{-1}\nabla\sigma(Z)|_Z(\det G(Z))^{-1/2}.$$

Because the flux $J(\zeta)$ is assumed to be constant across $T(s)$, it must be parallel to $Z_s(s)$, so $G(Z)^{-1}\nabla\sigma(Z) \parallel Z_s$. Together with $(\nabla\sigma(Z))^\top Z_s = 1$, this implies that $\nabla\sigma(Z)=G(Z)Z_s/(Z_s^\top G(Z)Z_s)$, so

$$Q(s)=a|Z_s(s)|^{-1}_{Z(s)}e^{-1\beta F(Z(s))}\det G(Z(s))^{-1/2}. \quad 16$$

Using Eqs. (14) and (16), one obtains $1 / R[Z]$, with $R[Z]$ given by Eq. (1), for the narrow tube flow rate in the case of a single path. If there are several isolated tubes, the flow rate is the sum of the individual contributions, and $1 / R[Z]$ gives the relative rate for each tube.

4.2. MaxFlux path

The MaxFlux path $\zeta = Z(s)$, $0 \le s \le 1$, minimizes the resistance $R[Z]$ of the enclosing tube, given by Eq. (1), with $Z(0)$ on the boundary of A_ξ and $Z(1)$ on that of B_ξ.

For simplicity, one might shrink A_ξ and B_ξ to points. Normally, A_ξ and B_ξ each contain local minima of $F(\zeta)$, and it would be natural to use the local minima for such points, which is conveniently obtained with boundary conditions $\nabla F(Z(1)) = \nabla F(Z(0)) = 0$. Nonetheless, for $\beta < +\infty$, arbitrary Dirichlet boundary conditions can be imposed.

In addition to $\zeta = Z(s)$, one should define a *reference path* $x = X(s)$ in cartesian coordinates. If one simply chooses $X(s) = \text{argmin}_{\xi(x)=Z(s)} U(x)$, the path could have jumps in it. Instead, one might ask that $X(s)$ be a MaxFlux path subject to the constraint $\xi(X(s)) = Z(s)$.

4.3. The geometric property

The form of the integral for $R[Z]$—and hence the path—does not depends on how collective variable space is parameterized.

This is to be contrasted with the minimum resistance path of Ref. [1], which minimizes

$$\int_0^1 \exp(\beta F(Z))|Z_s|^{-1}|Z_s|_Z^2 \mathrm{d}s,$$

where $|\mathrm{d}\zeta| = (\mathrm{d}\zeta^\top \mathrm{d}\zeta)^{1/2}$, and does depend on how collective variable space is parameterized. The two integrals are identical for $G(\zeta) = I$, however.

Ref. [4], Sec. 2.1, offers instead the formula

$$\text{resistanve}=\int_0^1 \exp(\beta F(Z))(Z_s^\top G(Z) Z_s)^{1/2}\,\mathrm{d}s,$$

with a proof given for the special case $\xi(x) = G^{1/2}x$ where G is constant. As a check, with the change of variables $\zeta = x(\eta)$,

$$\text{resistanve}=\int_0^1 \exp(\beta \overline{F}(\overline{Z}))(\overline{Z}_s^\top \overline{G}(\overline{Z})\overline{Z}_s)^{1/2}\,\mathrm{d}s$$

where $Z(s) = \chi(\tilde{Z}(s))$, $\bar{F}(\eta) = F(\chi(\eta))$, and $\bar{G}(\eta) = ((\partial\chi/\partial\eta)(\eta))^\top\ G(x(\eta))(\partial x/\partial\eta)(\eta)$. This is formally correct, except that $F(\chi(\eta)) - (1/\beta)\log|\det((\partial\chi/\partial\eta)(\eta))|$ is the free energy function, not simply $F(\chi(\eta))$ (cf, Ref. [15], Eq. (9)).

4.4. Using the Hamilton–Jacobi equation to find global minimum

The resistance is the length of a path with metric tensor $G^\times = e^{2\beta F}(\det G)G$, so the MaxFlux path is a *geodesic* from A_ξ to B_ξ. Let $g(\zeta)$ be the distance from a point ζ to A_ξ. Then $g(\zeta)$ can be shown [4] to satisfy the first order partial differential equation

$$((\nabla g(\zeta))^\top G^\times(\zeta)\nabla g(\zeta))^{1/2}=1.$$

A level set of $g(\zeta)$ is a wavefront of points equally distant from A_ξ. As g increases, the front moves away from A_ξ. Where it first reaches B_ξ is the end of the global MaxFlux path. To construct the path, follow the gradient back to A_ξ.

Figure 4a of Ref. [4] illustrates the level curves advancing to state B_ξ for phi-psi collective variables for alanine dipeptide. That article acknowledges that the method is practical only in low dimensions.

4.5. Parameterizing the path

It is convenient to remove the arbitrariness in parameterizing the optimal path $\zeta = Z(s)$. A convenient normalization is to choose s to be relative arclength so that $|Z_s(s)|_{Z(s)}$ is constant (length of entire path) and $(\mathrm{d}/\mathrm{d}s)|Z_s(s)|_{Z(s)} = 0$.

For *reliably* computing a minimizing path $\zeta = Z(s)$, it is necessary to enforce the parameterization. There are two principal ways (Ref. [8], Sec. III) to do this: (i) use of a constraint, as in the string method, and (ii) addition of a term to the objective function that penalizes deviation from the constraint, as in the nudged elastic band method. It is difficult to properly weight a penalty function, so the use of a constraint is considered here.

There is a drawback to the imposition of a constraint and that is that the path is then required to have second derivatives that are square integrable. This is a serious matter in the limiting case $\beta \to \infty$ of zero temperature, for which case the following conditions must be imposed:

(i) ζ_A and ζ_B must be free energy local minima, and (ii) the constraint must be omitted at intermediate local minima.

4.6. The Euler–Lagrange Equation

With the concise notation of Eq. (12), the variational principle for the MaxFlux path is to choose $Z(s)$ to minimize

$$\int_0^1 \gamma(Z)c(Z,Z_s)\mathrm{d}s, \quad 17$$

where $\gamma(\zeta)$ and $c(\zeta, \omega)$ are defined by Eq. (12), subject to given boundary values for $Z(0)$ and $Z(1)$ and normalization constraint $c(Z,Z_s)_s = 0$. The Lagrangian is

$$\int_0^1 \gamma(Z)c(Z,Z_s)\mathrm{d}s + \int_0^1 \lambda \frac{\mathrm{d}}{\mathrm{d}s} c(Z,Z_s)\mathrm{d}s. \quad 18$$

Notation for the remainder of this section omits function arguments Z, Z_s, and Z_{ss}.

As an intermediate step, the following expression is derived for the first variation for functional (18):

$$\int_0^1 \delta Z^\top (h - \lambda_s b + \lambda_{ss}\gamma\nabla_\omega c)ds + \Big[(\delta Z)_s^\top \lambda \nabla_\omega c\Big]_0^1. \quad 19$$

where

$$h = (I-P)(c\nabla_\zeta \gamma + \gamma b) \text{ and } b = \nabla_\zeta c - \frac{\mathrm{d}}{\mathrm{d}s}\nabla_\omega c.$$

To see this, first integrate by parts and apply boundary conditions to get

$$\int_0^1 \delta Z^\top \left(\nabla_\zeta((\gamma-\lambda_s)c) - \frac{\mathrm{d}}{\mathrm{d}s}((\gamma-\lambda_s)\nabla_\omega c)\right)\mathrm{d}s + \Big[(\delta Z)_s^\top \lambda \nabla_\omega c\Big]_0^1.$$

Then note that

$$Z_s^\top b = Z_s^\top \nabla_\zeta c - (Z_s^\top \nabla_\omega c)_s + Z_{ss}^\top \nabla_\omega c = 0, \quad 20$$

whence $b = (I - P)b$.

To get the Euler–Lagrange equations, consider all possible perturbations δZ. The perturbation δZ can be defined to force $\lambda\nabla_\omega c$ to vanish at $s = 0, 1$, implying that $\lambda(0) = \lambda(1) = 0$. Also, one has

$$h - \lambda_s b + \lambda_{ss}\gamma\nabla_\omega c = 0,$$

so, using Eq. (20),

$$0 = Z_s^\top (h - \lambda_s b + \lambda_{ss}\gamma\nabla_\omega c) = \lambda_{ss}\gamma c.$$

Therefore, $\lambda_{ss} = 0$ and $\lambda = 0$. In conclusion, the Euler-Lagrange equations of the variational principle are

$$h = 0, \quad c_s = 0, \quad \lambda = 0,$$

with $Z(0)$, $Z(1)$ given.

Note that without the constraint, the equation is simply $h = 0$, with $Z(0)$ and $Z(1)$ given, which is the same as Eq. (11), showing the equivalence of the two characterizations of a MaxFlux path. The alternative characterization introduced here is that the path is the center of a narrow tube of uniform cross-sectional area that maximizes the reactive flow rate *if* the free energy function is set to $+\infty$ outside the tube. Equivalence is dependent on the intersections of the tube with isocommittors having uniform area using the appropriate area metric.

The terms $c\nabla_\zeta\gamma$ and γb of $h(Z, Z_s, Z_{ss})$ expand to

$$c\nabla_\zeta\gamma = ce^{\beta F}(\det G)^{1/2}\left(\beta\nabla F + \frac{1}{2}\nabla_\zeta \mathrm{tr}(G(Z)^{-1}G(\zeta))|_{\zeta=Z}\right) \quad 21$$

and

$$\gamma b = e^{\beta F}(\det G)^{1/2}\left(\nabla_\zeta (Z_s^\top G(\zeta) Z_s)|_{\zeta=Z} + \frac{\mathrm{d}}{\mathrm{d}s}(GZ_s)\right). \quad 22$$

5. Continuous path dynamics

Though it can be helpful to have the Euler–Lagrange equations, the equations themselves do not distinguish between minima and maxima, so it is also good to stay close to the minimization formulation. Also, spatial discretization of the unknown path can be effected

in a more principled way if applied to the original functional. Unfortunately, accelerated optimization methods achieve their speed by assuming locally quadratic behavior, which is inappropriate for the resistance functional (1)/(18). This argues for minimization based on gradient descent dynamics. Indeed, dynamics in (continuum) path space is proposed for the string method [5], yielding a partial differential equation to be integrated until (stable) stationarity is attained. It is worth examining the specifics of this approach even if one chooses instead to apply dynamics to a discrete path.

Consider a path evolving in time:

$$\zeta = Z(s,t),\ t \geq 0.$$

The goal is to define a partial differential equation

$$\frac{\partial}{\partial t} Z = f(Z, Z_s, Z_{ss}),\ 0 < s < 1,$$

with $Z(0, t)$, $Z(1, t)$ given, so that

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_0^1 \gamma(Z) c(Z, Z_s)\mathrm{d}s < 0 \text{ and } (|Z_s|_Z)_s = 0,$$

and integrate until the path attains a stable stationary state. The constraint requires that $(f^\top \nabla_\zeta c + f_s^\top \nabla_\omega c)_s = 0$, where again arguments Z, Z_s, and Z_{ss} are omitted. This can be rewritten as

$$(f^\top b)_s + (f^\top \nabla_\omega c)_{ss} = 0. \quad 23$$

From Eq. (19) with $\lambda = 0$, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 \gamma c \mathrm{d}s = \int_0^1 f^\top h \mathrm{d}s. \quad 24$$

One might choose for f the direction of steepest descent,

$$f = -G^{-1}(h - \lambda_s b + \lambda_{ss}\nabla_\omega c),\ \lambda(0) = \lambda(1) = 0, \quad 25$$

which is derived in Appendix “Direction of steepest descent”. Choosing the direction of steepest descent f has two drawbacks:

i. the range of f is very great due to the presence of the factor $e^{\beta F}$ in its first term, and,

ii. less importantly, the constraint terms in f are complicated.

Therefore, instead, it is proposed to use

$$f = -S^{-1}G^{-1}h + \lambda Z_s$$

where $S(Z, Z_s)$ is a positive scale factor. (The constraint remains Eq. (23) as before.) From Eqs. (21) and (22), it seems that a good choice is $S = c\beta e^{\beta F}(\det G)^{1/2}$, for which $f = -G^{-1}((I - P)(\nabla F + \cdots) + \lambda Z_s)$. In any case, Eq. (24) becomes

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_0^1 \gamma c \mathrm{d}s = \int_0^1 S^{-1}h^\top G^{-1}h \mathrm{d}s < 0,$$

since $Z_s^\top h = 0$. The constraint, Eq. (23), gives

$$(c\lambda_s - b^\top G^{-1}S^{-1}h)_s = 0$$

for the Lagrange multiplier λ. (As a check, note that in the infinite-time limit as $h \to 0$, the constraint becomes $\lambda_{ss} = 0$.)

Note. The requirement that $f = 0$ at $s = 0,1$ is achieved by Z_{ss} attaining suitable point values at $s = 0,1|$. For the limiting case $\beta \to \infty$, Z_{ss} is not present in f, so this condition must be met by having $\nabla F = 0$ at $s = 0, 1$.

6. Paths of greatest probability

For a given duration T of a Brownian dynamics trajectory, there is a measure of the relative probability [16] of different trajectories $\zeta = \hat{\zeta}(\tau)$, $0 \le \tau \le T$, namely the action functional

$$S|\hat{\zeta}| = \int_0^T \left| \frac{d}{d\tau}\hat{\zeta} + \beta D(\hat{\zeta})\nabla F(\hat{\zeta}) - \mathrm{div} D(\hat{\zeta}) \right|_{\hat{\zeta}}^2 \mathrm{d}\tau, \quad (26)$$

where definitions (4)/(5) are used in defining the metric $|\mathrm{d}\zeta|_\zeta$. Olender and Elber [19] propose defining a reaction path, in the case of cartesian coordinates, as a path that minimizes this. An obvious drawback is its dependence on duration T; it is not sensible to compare $S[\hat{\zeta}]$ for different durations T.

In Appendix “Most probable path”, it is shown that by optimizing the kinetics for a given path $\zeta = Z(s)$, $0 \le s \le 1$, the problem reduces to minimizing

$$\overline{S}[Z]=\int_0^1 \left(\frac{2|f|_Z^2+C}{(|f|_Z^2+C)^{1/2}}|Z_s|_Z+2d_0 f^\top D^{-1} Z_s \right) \mathrm{d}s \quad 27$$

where

$$f(\zeta)=\beta D(\zeta)\nabla F(\zeta) - \mathrm{div} D(\zeta)$$

and C satisfies

$$\int_0^1 (|f|_Z^2+C)^{-1/2}|Z_s|_Z \mathrm{d}s=T. \quad 28$$

The functional $\bar{S}[Z]$ is not a geometric quantity. Appendix "Most probable path" shows that the path depends not only on the choice of collective variable space but also on how this space is parameterized.

The functional $\bar{S}[Z]$ (scaled by $1/\beta$) is, however, a geometric quantity in the zero-temperature limit $\beta \to \infty$. In addition, if the end points are specified, the functional simplifies to

$$\int_0^1 \frac{2|D\nabla F|_Z^2+C}{(|D\nabla F_Z^2+C)}|Z_s|_Z \mathrm{d}s+2d_0(F(Z(1)) - F(Z(0))).$$

For the limiting case $\beta \to \infty$, it is reasonable to select $T = +\infty$, which, from Eq. (28), can be attained for a path of finite length only if $C = 0$. Hence, the problem simplifies to that of minimizing

$$\int_0^1 |D(Z)\nabla F(Z)|_Z |Z_s|_Z \mathrm{d}s.$$

Invariably such a path passes through a critical point of $F(\zeta)$ giving a duration $T = +\infty$. This result is obtained for cartesian coordinates in Refs. [19, 24] and is employed to show that the minimizer of this functional is the minimum energy path, for which $Z_s \| D(Z)\nabla F(Z)$. For collective variables, the minimizer is a minimum free energy path (Ref. [16], Appendix "Area metric"). The generalization to free energy, however, is subject to an inconsistency, in that, in the limit $\beta \to \infty$, β is held fixed in $F(\zeta; \beta)$ and $D(\zeta; \beta)$.

Acknowledgments

This material is based upon work supported by Grant R01GM083605 from the National Institute of General Medical Sciences.

7. Appendix

7.1. Area metric

A formula is derived for measuring surface area with a metric tensor $G(\zeta)$. An infinitesimal surface $\delta\Sigma$ enclosing a point ζ_0 can be expressed as the intersection of an infinitesimal region $\delta\Omega$ surrounding ζ_0 and the plane $\hat{n}\cdot(\zeta-\zeta_0)=0$. To determine the area of $\delta\Sigma$, introduce variables $\zeta' = G^{1/2}(\zeta-\zeta_0)$, for which distances and angles are locally Euclidean. In the new coordinates, the surface becomes the intersection of some transformed region $\delta\Omega'$ and the plane $\hat{n}'\cdot\zeta'=0$ where $\hat{n}' = G^{-1/2}\hat{n}/|G^{-1/2}n(x0302)|$ and $|\delta\zeta| = (\delta\zeta^{\top}\delta\zeta)^{1/2}$. Its surface area is

$$\begin{aligned}\int_{\delta\sum'}\mathrm{d}S_{\zeta'} &= \int_{\delta\Omega'}\delta(\hat{n}'\cdot\zeta')d\zeta' \\ &= \int_{\delta\Omega}\delta(\hat{n}'\cdot G^{1/2}(\zeta-\zeta_0))\det G^{1/2}\mathrm{d}\zeta \\ &= \int_{\delta\Omega}\delta((\hat{n}/|G^{-1/2}\hat{n}|)\cdot(\zeta-\zeta_0))(\det G)^{1/2}\mathrm{d}\zeta \\ &= \int_{\delta\Omega}|G^{-1/2}\hat{n}|\delta(\hat{n}\cdot(\zeta-\zeta_0))(\det G)^{1/2}\mathrm{d}\zeta \\ &= \int_{\sum\Omega}|G^{-1/2}\hat{n}|(\det G)^{1/2}\mathrm{d}S_{\zeta}\end{aligned}$$

where the first and last equality follow by applying the coarea formula This result is restated in Eq. (8) using the collective variable metric.

7.2. Path of maximum reactive flux

Define $\hat{q}(s) = q(Z(s))$, and define $\sigma(\zeta)$ implicitly by $q(\zeta) = \hat{q}(\sigma(\zeta))$. The assumption Eq. (10) that isocommittors are planar implies that $\sigma(\zeta)$ satisfies

$$\hat{n}(\sigma(\zeta))^{\top}(\zeta - Z(\sigma(\zeta)))=0,$$

so on each isocommittor $\Pi(s)$,

$$\nabla\sigma(\zeta)|_{\sigma(\zeta)=s}=\frac{\hat{n}(s)}{Z_s(s)^{\top}\hat{n}(s)-\hat{n}_s(s)^{\top}(\zeta-Z(s))}. \quad 29$$

It is convenient to use Eq. (12) to write Eq. (9) as

$$j(\zeta)=d_0C_{\xi}\gamma(\zeta)^{-1}c(\zeta,G(\zeta)^{-1}\nabla q(\zeta)). \quad 30$$

Using Eq. (29) and $q(\zeta) = \hat{q}(\sigma(\zeta))$, Eq. (30) becomes

$$j(\zeta)|_{\sigma(\zeta)=s}=d_0 C_\xi \hat{q}_s(s)\frac{\gamma(\zeta)^{-1}c(\zeta,G(\zeta)^{-1}\hat{n}(s))}{\hat{n}(s)^\top Z_s(s)-\hat{n}_s(s)^\top(\zeta-Z(s))}.$$

Maximizing $j(\zeta)$ over ζ subject to $\hat{n}(s)^\top(\zeta - Z(s)) = 0$ gives the condition

$$-\frac{\nabla_\zeta\gamma(Z)}{\gamma(Z)}-\frac{\nabla_\zeta c(Z,G(Z)^{-1}\hat{n})}{c(Z,G(Z)^{-1}\hat{n})}+\frac{\hat{n}_s}{\hat{n}^\top Z_s}\parallel\hat{n} \quad 31$$

for the value $Z(s)$ that maximizes $j(\zeta)$. (See Ref. [27], Sec. 3.4.)

The assumption that the flux $J(Z)$ is parallel to the direction Z_s of the path implies $| G(Z)^{-1}\hat{n} \parallel Z_s$. Noting that $\nabla_\omega c(\zeta, \omega) = G(\zeta)\omega/c(\zeta, \omega)$, condition (31) becomes

$$-\frac{\nabla_\zeta\gamma(Z)}{\gamma(Z)}-\frac{\nabla_\zeta c(Z,Z_s)}{c(Z,Z_s)}+\frac{(\nabla_\omega c(Z,Z_s))_s}{c(Z,Z_s)}\parallel\nabla_\omega c(Z,Z_s).$$

(See Ref. [27], Eq. (10).) This can be written as an equation by expressing the left-hand side as an undetermined scalar times the right-hand side and then solving for the scalar after premultiplying this equation by $Z_s^\top$. The result is Eq. (11).

7.3. Narrow tube flow rate formula

Using Eq. (6) and Eq. (18) of Ref. [18], the flow rate can be expressed as a volume integral:

$$\text{flow rate}=\int_\Omega \rho_\xi(\zeta)\nabla q(\zeta)^\top D(\zeta)\nabla q(\zeta)\mathrm{d}\zeta=C_\xi d_0 I(q). \quad 32$$

Foliating Ω in Eq. (6) gives

$$I(q)=\int_0^1\int_\Omega e^{-\beta F(\zeta)}\nabla q(\zeta)^\top G(\zeta)^{-1}\nabla q(\zeta)\delta(\sigma(\zeta)-s)\mathrm{d}\zeta\mathrm{d}s$$

where the decomposition $\hat{q}(\sigma(\zeta) = q(\zeta)$ is defined by Eq. (13). Using the coarea formula gives

$$I(q)=\int_0^1\int_{\sum(s)}\frac{e^{-\beta F(\zeta)}\nabla q(\zeta)^\top G(\zeta)^{-1}\nabla q(\zeta)}{|\nabla\sigma(\zeta)|}\mathrm{d}S_\zeta\mathrm{d}s.$$

Using $\nabla q(\zeta) = \hat{q}_s(s)\nabla\sigma(\zeta)$ gives

$$I(q)=\int_0^1 \hat{q}_s(s)^2 \int_{\sum(s)} e^{-\beta F(\zeta)} \nabla\sigma(\zeta)^\top G(\zeta)^{-1} \hat{n}(\zeta) \mathrm{d}S_\zeta \mathrm{d}s$$
$$=\int_0^1 \hat{q}_s(s)^2 Q(s) \mathrm{d}s$$

where $\hat{n}(\zeta) = \nabla\sigma(\zeta)/|\nabla\sigma(\zeta)|$ is the outer normal at ζ on $\Sigma(s)$ and $Q(s)$ is given by Eq. (15). Minimizing over $\hat{q}$ for fixed σ gives the Euler–Lagrange equation $-(2\hat{q}_s Q(s))_s = 0$, whose solution is given by

$$q(Z(s))=\hat{q}(s)=I(q)\int_0^s \frac{\mathrm{d}s}{Q(s)}.$$

Therefore,

$$I(q)=\left(\int_0^1 \frac{\mathrm{d}s}{Q(s)}\right)^{-1}. \quad 33$$

From Eqs. (33) and (32), one obtains Eq. (14).

7.4. Direction of steepest descent

Let $(\int_0^1 \delta Z(s)^\top G(Z(s))\delta Z(s)\mathrm{d}s)^{1/2}$ measure the size of a deviation $\delta Z(s)$ from a path $Z(s)$. For the direction of steepest descent f for the functional (17) subject to the constraint (23) the Lagrangian is

$$\int_0^1 f^\top h \mathrm{d}s+\mu\left(\int_0^1 f^\top G f \mathrm{d}s - \mathrm{const}\right)+\int_0^1 \lambda((f^\top b)_s+(f^\top \nabla_\omega c)_{ss})\mathrm{d}s.$$

Its first variation is

$$\int_0^1 (\delta f)^\top (h - \lambda_s b+\lambda_{ss}\nabla_\omega c+2\mu G f)\mathrm{d}s+[(df)_s^\top \lambda \nabla_w c]_0^1.$$

The Euler-Lagrange equations give $\lambda(0) = \lambda(1) = 0$ and $f = -(2\mu G)^{-1}(h - \lambda_s b + \lambda_{ss}\nabla_\omega c)$, in addition to the two original constraints. The value of const can be chosen to make $\mu=\frac{1}{2}$ giving Eq. (25). With this choice, Eq. (23) becomes

$$(-\lambda_{sss}+b^\top G^{-1} b\lambda_s - b^\top G^{-1} h)_s=0.$$

where $b^\top G^{-1} \nabla_\omega c = 0$ follows from Eq. (20). Also, using Eq. (20), one has for $s = 0,1$ that

$$0=Z_s^\top Gf=-\lambda_{ss}Z_s^\top\nabla_\omega c=\lambda_{ss}c^2,$$

thus obtaining two more boundary conditions for λ, namely $\lambda_{ss}(0) = \lambda_{ss}(1) = 0$.

7.5. Most probable path

As a first step in maximizing the action, given by Eq. (26), separate the path of the trajectory from its kinetics by a reparameterization $t = t(s)$, $0 \le s \le 1$, and define the path by

$$Z(s)=\hat{\zeta}(\tau(s)).$$

(For example, choose $\tau(s)$ so that s is relative arclength, meaning that $|Z_s(s)| = L$ where L is the length of the path.) With this change of variables, the action becomes

$$S[\hat{\zeta}]=\int_0^1|\tau_s^{-1}Z_s+f(Z)|_Z^2\tau_s\mathrm{d}s.$$

The schedule $t = \tau(s)$ that minimizes the action subject to $\tau(0) = 0$ and $\tau(1) = T$ is determined using the calculation of variations. The result is that $\tau_s=|Z_s|_Z(|f|_Z^2+C)^{-1/2}$ where C satisfies Eq. (28). With this choice, the action becomes $\bar{S}[Z]$, given in Eq. (27).

Consider now the effect of a change of variables on the functional. In particular, if one use instead variables η, where

$$\zeta=\chi(\eta),$$

one has $Z = \chi(\bar{Z})$ and $\bar{F}(\eta)$, $\bar{D}(\eta)$ given by

$$\begin{aligned}\overline{F}&=F(\chi)-\tfrac{1}{\beta}\log\left|\det\tfrac{\partial\chi}{\partial\eta}\right|,\\ \overline{D}&=\left(\tfrac{\partial\chi}{\partial\eta}\right)^{-1}D(\chi)\left(\tfrac{\partial\chi}{\partial\eta}\right)^{-\top},\end{aligned}$$

where the last two equations are derived from definitions given by Eqs. (2) and (3). Therefore, $\bar{Z}_s = ((\partial\chi/\partial\eta)(\bar{Z}))^{-1}Z_s$ and $\bar{s}[\bar{Z}] = \bar{S}[Z]$ *provided* $\bar{f} = (\partial\chi/\partial\eta)^{-1}f(\chi)$ where

$$\overline{f}=\beta\overline{D}\nabla_\eta\overline{F}-\mathrm{div}_\eta\overline{D}.$$

Differentiating the expression for $\bar{F}$ gives

$$\left(\frac{\partial\chi}{\partial\eta}\right)^{\top}(\nabla_{\zeta}F)(\chi)=\nabla_{\eta}\left(\overline{F}+\frac{1}{\beta}\log\left|\det\frac{\partial\chi}{\partial\eta}\right|\right),$$

whence

$$f(\chi)=\beta D(\chi)(\nabla_{\zeta}F)(\chi)-(\mathrm{div}D)(\chi)=\frac{\partial\chi}{\partial\eta}\overline{f}+g,$$

and

$$g=\frac{\partial\chi}{\partial\eta}\overline{D}\nabla_{\eta}\left(\log\left|\det\frac{\partial\chi}{\partial\eta}\right|\right)-(\mathrm{div}D)(\chi)+\frac{\partial\chi}{\partial\eta}\mathrm{div}\overline{D}.$$

One has

$$\left(\frac{\partial\chi}{\partial\eta}\overline{D}\nabla\left(\log\left|\det\frac{\partial\chi}{\partial\eta}\right|\right)\right)^{i}=\chi^{i}_{j}\overline{D}^{jk}\left(\log\left|\det\frac{\partial\chi}{\partial\eta}\right|\right)_{k}=\chi^{i}_{j}\overline{D}^{jk}\mathrm{tr}\left(\left(\frac{\partial\chi}{\partial\eta}\right)^{-1}\left(\frac{\partial\chi}{\partial\eta}\right)_{k}\right),$$

where tensor notation is used, with subscripts for differentiation. Differentiating | $D^{ij}(\chi)$ gives

$$D^{ij}_{k}(\chi)\chi^{k}_{l}=(\chi^{i}_{m}\overline{D}^{mn}\chi^{j}_{n})_{l},$$

whence

$$D^{ij}_{k}(\chi)=(\chi^{i}_{m}\overline{D}^{mn}\chi^{j}_{n})_{l}\overline{\chi}^{l}_{k},$$

where $\overline{\chi}^{i}_{j}$ is element ij of $(\partial\chi/\partial\eta)^{-1}$, so

$$(\mathrm{div}D)^{i}(\chi)=(\chi^{i}_{m}\overline{D}^{mn}\chi^{j}_{n})_{l}\overline{\chi}^{l}_{j}=\chi^{i}_{m}\overline{D}^{ml}+\chi^{i}_{m}\overline{D}^{ml}_{l}+\chi^{i}_{m}\overline{D}^{mn}\chi^{i}_{m}\overline{\chi}^{k}_{j}.$$

Therefore,

$$g^{i}=\mathrm{tr}\left(\overline{D}\frac{\partial^{2}\chi^{i}}{\partial\eta\partial\eta}\right).$$

The extra term g changes the form of the functional to be minimized, so if one minimizes the integral using instead variables η, the resulting path $\eta = \bar{Z}(s)$ does not satisfy $Z(s) = \chi(\bar{Z}(s))$.

References

1. Berkowitz M, Morgan JD, McCammon JA, Northrup SH. Diffusion-controlled reactions: a variational formula for the optimum reaction coordinate. J Chem Phys. 1983; 79:5563–5565.
2. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu Rev Phys Chem. 2002; 53:291–318. [PubMed: 11972010]
3. Burger SK, Yang W. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. J Chem Phys. 2006; 124:054109–1. 054109–13. [PubMed: 16468853]
4. Cameron MK. Estimation of reactive fluxes in gradient stochastic systems using an analogy with electric circuits. J Comput Phys. 2013; 247:137–152.
5. Weinan E, Ren W, Vanden-Eijnden E. String method for the study of rare events. Phys Rev B. 2002; 66:6688–6693.
6. Weinan E, Ren W, Vanden-Eijnden E. Minimum action method for the study of rare events. Commun Pure Appl Math. 2004; 57(5):637–656.
7. Weinan E, Ren W, Vanden-Eijnden E. Finite temperature string method for the study of rare events. J Phys Chem B. 2005; 109(5):6688–6693. [PubMed: 16851751]
8. Weinan E, Ren W, Vanden-Eijnden E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. J Chem Phys. 2007; 126:164103–1. 164103–8. [PubMed: 17477585]
9. Weinan E, Vanden-Eijnden E. Towards a theory of transition paths. J Stat Phys. 2006; 123(3):503–523.
10. Weinan E, Vanden-Eijnden E. Transition-path theory and pathfinding algorithms for the study of rare events. Annu Rev Phys Chem. 2010; 61(1):391–420. [PubMed: 18999998]
11. Faccioli P, Sega M, Pederiva F, Orland H. Dominant pathways in protein folding. Phys Rev Lett. 2006; 97:108101. [PubMed: 17025856]
12. Huang H, Zhao R, Dickson BM, Skeel RD, Post CB. αC helix as a switch in the conformational transition of Src/CDK-like kinase domains. J Phys Chem B. 2012; 116(15):4465–4475. DOI: 10.1021/jp301628r [PubMed: 22448785]
13. Huo S, Straub JE. The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body system at finite temperature. J Chem Phys. 1997; 107(13):5000–5006.
14. Jiménez A, Crehuet R. Can the MaxFlux algorithm describe bifurcating paths? Theor Chem Acc. 2007; 118:769–775.
15. Johnson ME, Hummer G. Characterization of a dynamic string method for the construction of transition pathways in molecular reactions. J Phys Chem B. 2012; 116:8573–8583. [PubMed: 22616575]
16. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: minimum free energy paths and isocommittor surfaces. J Chem Phys. 2006; 125(8):024106–1. 024106–15.
17. Maragliano L, Roux B, Vanden-Eijnden E. Comparison between mean forces and swarms-of-trajectories string methods. J Chem Theory Comput. 2014; 10(2):524–533. [PubMed: 26580029]
18. Metzner P, Schütte C, Vanden-Eijnden E. Illustration of transition path theory on a collection of simple examples. J Chem Phys. 2006; 125(8):084110–1. 084110–17. [PubMed: 16965004]
19. Olender R, Elber R. Yet another look at the steepest descent path. J Mol Struct THEOCHEM. 1997:398–399. 63–71.
20. Pan AC, Sezer D, Roux B. Finding transition pathways using the string method with swarms of trajectories. J Phys Chem B. 2008; 112(11):3432–3440. [PubMed: 18290641]
21. Park S, Sener MK, Lu D, Schulten K. Reaction paths based on mean first-passage times. J Chem Phys. 2003; 119(3):1313–1319.

22. Singhal N, Pande VS. Error analysis and efficient sampling in Markovian state models for molecular dynamics. J Chem Phys. 2005; 123(20):4909.

23. Vanden-Eijnden, E. Transition path theory. In: Binder, K.Ciccotti, G., Ferrario, M., editors. Computer simulations in condensed matter: from materials to chemical biology. Vol. 2. Lecture notes in physics Springer; Berlin: 2006. p. 453-493.

24. Vanden-Eijnden E, Heymann M. The geometric minimum action method for computing minimum energy paths. J Chem Phys. 2008; 128(6):61103.

25. West AM, Elber R, Shalloway D. Extending molecular dynamics timescales with milestoning: example of complex kinetics in a solvated peptide. J Chem Phys. 2007; 126:145104. [PubMed: 17444753]

26. Wolynes, P. Chemical reaction dynamics in complex molecular systems. In: Stein, DL., editor. Lectures in the sciences of complexity: the proceedings of the 1988 complex systems summer school held June–July, 1988 in Santa Fe, New Mexico. Santa Fe institute studies in the sciences of complexity lectures; Addison-Wesley, Reading; 1989.

27. Zhao R, Shen J, Skeel RD. Maximum flux transition paths of conformational change. J Chem Theory Comput. 2010; 6(8):2411–2423. [PubMed: 20890401]