

# The String Method as a Dynamical System

Maria Cameron<sup>1</sup>, Robert V. Kohn<sup>1</sup>, and Eric Vanden-Eijnden<sup>1</sup>

January 4, 2010

## Abstract

This paper is devoted to the theoretical analysis of the zero-temperature string method, a scheme for identifying minimum energy paths (MEP's) on a given energy landscape. By definition, MEP's are curves connecting critical points on the energy landscape which are everywhere tangent to the gradient of the potential except possibly at critical points. In practice, MEP's are mountain pass curves that play a special role e.g. in the context of rare reactive events that occur when one considers a steepest descent dynamics on the potential perturbed by a small random noise. The string method aims to identify MEP's by moving each point of the curve by steepest descent on the energy landscape. Here we address the question whether such a curve evolution necessarily converges to a MEP. Surprisingly, the answer is no, for an interesting reason: MEP's may not be isolated, in the sense that there may be families of them that can be continuously deformed into one another. This degeneracy is related to the presence of critical points of Morse index 2 or higher along the MEP. In this paper, we elucidate this issue and characterize completely the limit set of a curve evolving by the string method. We establish rigorously that the limit set of such a curve is again a curve when the MEP's are isolated. We also show under the same hypothesis that the string evolution converges to an MEP. However, we identify and classify situations where the limit set is not a curve and may contain higher-dimensional parts. We present a collection of examples where the limit set of a path contains a 2D region, a 2D surface, or a region of an arbitrary dimension up to the dimension of the space. In some of our examples the evolving path wanders around without converging to its limit set. In other examples it fills a region, converging to its limit set which is not an MEP.

## 1 Introduction

Many problems from structural biology, computational chemistry, and condensed matter physics involve dissipative dynamical systems perturbed by thermal fluctuations. Such systems often spend most of their time in metastable states, making rare thermally-activated transitions between these states. Typically, the transitions are not accessible to direct numerical simulation due to the long waiting times. Therefore the identification of transition pathways and rates is a cross-cutting challenge of great scientific importance.

The simplest version of this problem involves identification of a mountain pass separating two local minima in a potential energy landscape  $V(x)$ . The associated minimum energy path (MEP) – which does steepest ascent to the saddle point then steepest descent thereafter –

---

<sup>1</sup>Courant Institute of Mathematical Science, Department of Mathematics, New York University, 251 Mercer Street, New York, NY 10012, [cameron@cims.nyu.edu](mailto:cameron@cims.nyu.edu), [kohn@cims.nyu.edu](mailto:kohn@cims.nyu.edu), [eve2@cims.nyu.edu](mailto:eve2@cims.nyu.edu)

represents the pathway of the transition, and the height of the saddle determines its rate via the Arrhenius law (see Section 1.3). In the desired applications there can be thousands of degrees of freedom, and the potential landscape is only available numerically. So the mountain pass must also be found numerically.

Numerous algorithms have been proposed in the physics and chemistry literature for finding mountain passes. The string method [E et al, 2002, E et al, 2007] and the closely related nudged elastic band method [Jonsson et al, 1998, Sheppard et al, 2008] are perhaps the most widely used, but other algorithms include the method of [Ulitsky and Elber, 1990], the conjugate peak refinement method [Fischer and Karplus, 1992], the dimer method [Henkelman and Jonsson, 1999], discrete path sampling [Wales, 2002], the activation-relaxation technique [Barkema and Mousseau, 2001], and the step-and-slide method [Miron and Fichthorn, 2001]. Though these and other algorithms have been applied very effectively in a wide range of examples, their mathematical and numerical analysis has hardly been studied at all.

This paper provides a mathematical analysis of the string method. We focus on the continuous-time, continuous-space setting, where the method amounts to an evolution law for curves in state space. Thus our goal is not the numerical analysis of the string method, but rather a study of the associated dynamical system.

The stationary states of the string method are curves that do only steepest descent or steepest ascent, i.e. that trace a sequence of heteroclinic orbits for the gradient dynamical system  $\dot{x} = -\nabla V(x)$  forward or backward in time. These curves are traditionally called minimum energy paths (MEP's) and they can be characterized geometrically as the solutions of

$$[\nabla V(\gamma)]^\perp = 0 \tag{1}$$

where  $\gamma$  denotes the curve and  $[\nabla V(\gamma)]^\perp$  is the component of  $\nabla V$  perpendicular to  $\gamma$ ; Equation (1) must be satisfied pointwise along the curve, and is supplemented by the condition that the curve connects two minima of  $V(x)$ . If  $V(x)$  possesses at least two minima and satisfies certain conditions (see Sec. 1.2), the existence of solutions to Eq. (1) is easy to establish. These solutions, however, typically fail to be unique and they may do so in an interesting way. Indeed, they may be non-isolated in the sense that families of solutions to Eq. (1) may exist in which every member of the family can be continuously deformed into another. This property is intrinsically linked to the presence of critical points of index  $\geq 2$  in the energy landscape.

As we show below, the lack of uniqueness of the solutions of Eq. (1) has important consequences for the string evolution. In particular, when non-isolated solutions exist, the string evolution may not converge to an MEP – instead, the large-time dynamics of the string can be much more complicated. Below we will give some examples of such complicated large-time behavior, including one where the evolution is time-periodic and another where it fills a region of the space. However, we will show that the limit set of the string evolution is always a union of heteroclinic orbits. Thus, though the large-time behavior of the string can be complicated, it always provides information about the structure of the potential landscape.

We have discussed the link to thermally-activated transitions, which was the original motivation for the string method. But our work can also be viewed from a different, more dynamical-system-oriented viewpoint. For any autonomous system  $\dot{x} = F(x)$ , it is standard to consider the image of a point (a trajectory) and its limit points (the  $\omega$ -limit set). In studying the string method, we are considering instead the image of a *curve* and its limit points.

We turn now to a systematic discussion of the string method (Section 1.1), before giving a more complete summary of our results (Section 1.2). The introduction closes with a brief explanation of the connection between MEP's and thermally-activated switching (Section 1.3).

## 1.1 The string method

In our continuous-time, continuous-space setting, the string method is an evolution law for curves in  $\mathbb{R}^n$  under the influence of a potential  $V : \mathbb{R}^n \rightarrow \mathbb{R}$ . In this work, a curve is a continuous image of a unit interval in  $\mathbb{R}^n$ . Moreover, the ends of the curve are fixed at two distinct critical points of the potential. Suppose the curve is differentiable and its configuration at time  $t$  is parametrized

nondegenerately by  $\alpha \in (A(t), B(t))$ , i.e. the curve at time  $t$  is  $\{\phi(\alpha, t) : \alpha \in (A(t), B(t))\}$  and its unit tangent is  $\hat{\tau}(\alpha, t) = \phi_\alpha / |\phi_\alpha|$ . Then string method evolves it so that *the normal velocity is the projection of  $\nabla V$  normal to the curve*:

$$\phi_t(\alpha, t) = -\nabla V^\perp + \lambda \hat{\tau} \quad (2)$$

where

$$\nabla V^\perp = \nabla V(\phi(\alpha, t)) - \langle \nabla V(\phi(\alpha, t)), \hat{\tau}(\alpha, t) \rangle \hat{\tau}(\alpha, t). \quad (3)$$

The evolution as a geometric curve is well-defined, though  $\lambda = \lambda(\alpha, t)$  is not unique. Different choices of  $\lambda$  correspond to different parametrizations. This is easy to see using chain rule: if  $\phi(\alpha, t)$  and  $\tilde{\phi}(\tilde{\alpha}, t)$  represent the same evolving curve then  $\phi(\alpha, t) = \tilde{\phi}(\tilde{\alpha}(\alpha, t), t)$ , and differentiation gives

$$\phi_t = \tilde{\phi}_t + \tilde{\phi}_{\tilde{\alpha}} \tilde{\alpha}_t = \tilde{\phi}_t + \mu \hat{\tau}.$$

with  $\mu = |\tilde{\phi}_{\tilde{\alpha}}| \tilde{\alpha}_t$ .

It is often convenient to choose a particular parametrization. For a robust numerical solution, a good choice is the unit-speed parametrization; then  $|\phi_\alpha| = 1$ ,  $\lambda(\alpha, t)$  is fully determined, and  $\alpha$  ranges over  $(0, l(t))$  where  $l(t)$  is the length of the curve at time  $t$ . Moreover, it is advantageous to avoid evaluation of  $\nabla V^\perp$  [E et al, 2007]. If we set

$$\beta = \lambda + \langle \nabla V(\phi(\alpha, t)), \hat{\tau}(\alpha, t) \rangle,$$

then the evolution law is given by

$$\phi_t(\alpha, t) = -\nabla V + \beta \hat{\tau} \quad (4)$$

where  $\beta(\alpha, t)$  is again uniquely determined by the requirement that  $|\phi_\alpha| = 1$ . To be stationary under Eqs. (2) or (4), a curve must satisfy  $\nabla V^\perp = 0$  pointwise. When this happens, Eq. (2) says that  $\phi_t$  is everywhere tangent to the curve; therefore its image (as a geometric curve) doesn't change, though the parametrization may change in time. Thus: a piecewise smooth curve passing through a sequence of critical points  $x_1, \dots, x_N$  is a stationary state of the string method if the curve is everywhere tangent to  $\nabla V$ . Put differently: the curve is stationary if it consists of a sequence of critical points connected by ‘‘heteroclinic orbits’’ (solutions of  $\dot{x} = -\nabla V(x)$  traced forward or backward in time). As noted above, such curves are called *minimum energy paths*, though the string evolution does not do steepest descent for any energy functional defined on curves. Other evolutions that do have a steepest descent character can be written down and exploited numerically, see [Olender and Elber, 1997, Vanden-Eijnden and Heymann, 2008], but we will not consider them here.

The preceding discussion assumed differentiability, to be sure that the unit tangent  $\hat{\tau}$  was well-defined. It is, however, important to relax this assumption, permitting a sharp change of direction when the path passes through a critical point of  $V$  and allowing the initial path to have sharp turns. Under this extended definition, we require that

- the curve passes through a finite number of points where  $\nabla V = 0$ ;
- except at these points the curve is differentiable and Eq. (2) holds;
- $\phi_t = 0$  whenever  $\phi(\alpha, t)$  is a critical point of  $V$ .

This is consistent with Eq. (2), if we take the view that  $\lambda$  must vanish at a corner (since  $\hat{\tau}$  is undefined) and  $\nabla V^\perp = 0$  at a critical point (since it is a projection of  $\nabla V = 0$ ).

Similarly, instead of using Eqs. (2) or (4), for our analysis it will be more convenient to work with

$$\phi_t = -\nabla V(\phi), \quad \phi(\alpha, 0) = \phi_0(\alpha). \quad (5)$$

In this representation, the parametrization is fixed by the initial condition and each point of the curve moves independently under the steepest descent

$$\dot{x} = -\nabla V(x). \quad (6)$$

In summary, the evolution of a curve as a geometrical object is determined by its normal velocity. The normal velocity is the same for Eqs. (2), (4), and (5) and equals  $\nabla V^\perp(\phi)$ . Therefore, the curve evolution according to Eqs. (2), (4), and (5) is equivalent.

In considering the large-time behavior of the evolution by Eq. (5), it is natural to consider its *limit set*. The definition should be independent of parametrization, so we cannot simply fix  $\alpha$  and study the large-time behavior of  $t \mapsto \phi(\alpha, t)$ . But it is nevertheless convenient to formulate the definition in terms of parametrized curves:

**Definition 1.** We say  $x_0$  is a limit point of a parametrized, evolving curve  $\phi(\alpha, t)$  if there exist sequences  $\{\alpha_n\}_{n=1}^\infty$  and  $\{t_n\}_{n=1}^\infty$  such that  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$  and

$$\lim_{n \rightarrow \infty} \phi(\alpha_n, t_n) = x_0. \quad (7)$$

The collection of all limit points of the evolving curve is its limit set.

In other words,  $x_0$  is a limit point of a curve if every neighborhood is visited by the evolving curve infinitely often.

We prove in Section Appendix A that the limit set is closed and connected, and is a union of critical points and heteroclinic trajectories. However it does not need to be a curve; instead it can be multidimensional.

If the evolving curve stabilizes as  $t \rightarrow \infty$  then it must approach its limit set, which must in turn be an MEP. This happens, for example, for the simple double-well potential shown in Fig. 1(a). But the evolving curve need not, in general, stabilize. We shall give examples of more

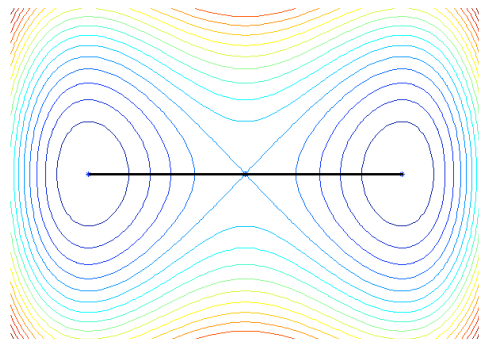


Figure 1: For this symmetric double-well potential, there is a unique MEP connecting the two minima via the saddle. For any initial curve starting at one minimum and ending at the other, the evolving curve converges to this MEP as  $t \rightarrow \infty$ . This is perhaps visually clear; it will be a consequence of Theorem 2. Let us emphasize the significance of letting  $\alpha$  change in Definition 1. Indeed, for this potential, every point evolved by  $\dot{x} = -\nabla V$  approaches one of the minima, or the saddle.

complicated behavior in Section 3.

When we work with parametrized curves “stabilization” is an awkward notion, since the geometric curve can stabilize though the parametrization might not. To avoid this issue, it is convenient to focus instead on whether the evolving curve *converges to its limit set*. The following definition is clearly parametrization-independent:

**Definition 2.** Let  $\phi(\alpha, t)$  be an evolving parametrized curve, and let  $L$  be its limit set. We say the curve converges to its limit if the Hausdorff distance  $d_H(\phi, L)$  between the curve and its limit set tends to zero as  $t \rightarrow \infty$ . In other words, for any  $\epsilon > 0$  there is  $t_\epsilon$  such that for any  $t > t_\epsilon$

- (i) for any  $x \in L$   $B_\epsilon(x) \cap \phi \neq \emptyset$ ; and
- (ii) for any  $y \in \phi$   $B_\epsilon(y) \cap L \neq \emptyset$ .

## 1.2 Summary of main results

We assume throughout this paper that the potential  $V$

- (1) is twice continuously differentiable,
- (2) has isolated critical points, each nondegenerate in the sense that the Hessian is invertible;
- (3) grows at infinity, in the sense that the sublevel sets  $S(C) = \{x : V(x) \leq C\}$  are compact for every  $C \in \mathbb{R}$ .

Condition (1) assures that the steepest-descent ODE  $\dot{x} = -\nabla V(x)$  has a unique solution, and condition (3) assures that its trajectories remain bounded as  $t \rightarrow \infty$ . Condition (2) implies that each trajectory approaches a critical point of  $V$  as  $t \rightarrow \infty$ . Condition (3) often can be replaced with a weaker condition (3'). Let us denote the connected component of the set  $S(C)$  containing a given point  $x$  by  $S(x, C)$ . Then the condition (3) can be replaced by:

- (3') the sets  $S(x, V(x))$  are compact for any point  $x$  of the initial path.

In Section 2 we announce those of our results that are affirmative and easy to comprehend. In Section 2.1 we state the basic properties of the limit set. The limit set of any bounded subset of  $\mathbb{R}^n$  is closed, connected, and consists of critical points and heteroclinic trajectories. Since the limit set is roughly speaking the  $\omega$ -limit set of the initial path, it is not surprising that our arguments resemble those used in a standard ODE setting to study the  $\omega$ -limit set of a trajectory. In Section 2.2 we ask whether the evolution converges to an MEP. The answer is affirmative if there are only finitely many MEP's connecting the critical points in the limit set. This is the case, for example, if all the critical points in the limit set are local minima or saddle points of index 1.

In Section 3 we construct examples where the evolution does not stabilize and the limit set is not an MEP. We start with an example demonstrating the significance of condition (3) for the potential. Then we present examples where a path revolves endlessly around some axis and where it fills a 2D region. Then we present a construction in which the limit set of a path is  $k$ -dimensional, where  $3 \leq k \leq n$ . Finally we turn to examples demonstrating the significance of the saddles of Morse index  $\geq 2$ . In these examples, the initial path is nowhere close to the index-2 saddles, and it does not intersect their stable manifolds. We show that the limit set can nevertheless be very sensitive to the initial path or can include multidimensional parts.

The more complex behavior seen in Section 3 is evidently linked to the presence of saddle points of index  $\geq 2$ , and to the existence of a continuum of MEP's. In such settings the question whether or not a given curve evolves to an MEP depends on the details of its relation to the steepest-descent flow. We elucidate this dependence in Section 4.

We were initially surprised to see that the string method does not always converge to an MEP. But perhaps we should not have been. When the MEP's are isolated there is little freedom and the evolution converges. When the MEP's are not isolated there is much more freedom and the evolution can be more complicated.

Is the possible failure of convergence a problem for the original application, which involved finding the "mountain pass" between two local minima? Not particularly. A mountain pass is, by definition, a saddle point of index 1. Therefore if the two minima are separated by a single mountain pass, the string method will probably find it – unless, due to an unfortunate choice of initial condition, the string wanders near some other critical point with index  $\geq 2$ .

## 1.3 Thermally-activated transitions and minimum energy paths

This section discusses briefly why (certain) MEP's are relevant to thermally-activated switching. We include it as motivation; the information here will not be used in the rest of the paper.

Consider the stochastic differential equation (SDE)

$$dx = -\nabla V(x)dt + \sqrt{2\epsilon} dw, \quad (8)$$

where  $\epsilon$  is a parameter. When  $\epsilon$  is small, the solution of Eq. (8) spends most of its time near the local minima of the potential  $V$ , with occasional excursions away from these local minima which eventually lead to transitions from one local minimum to another. This is the simplest model for how thermal fluctuation produces “rare events,” making a physical system evolve through a sequence of metastable states [Freidlin and Wentzell, 1998].

The rates and pathways of rare events can be understood (asymptotically as  $\epsilon \rightarrow 0$ ) in terms of the rate functional of large deviation theory:

$$S_T(z) = \int_0^T |\dot{z}(t) + \nabla V(z(t))|^2 dt. \quad (9)$$

The essential reason is this. Formally, the probability that the SDE follows the path  $z$  is proportional to

$$\exp(-S_T(z)/\epsilon). \quad (10)$$

Suppose  $x_1$  and  $x_2$  are metastable states, i.e. local minima of  $V$ . Then among all paths  $z$  that start at  $x_1$  at time 0 and end at  $x_2$  at time  $T$ , the one with maximum likelihood is the rate functional minimizer, i.e. the solution of

$$\min_{\gamma(0)=x_1, \gamma(T)=x_2} S_T(\gamma). \quad (11)$$

This is mainly useful when  $\epsilon$  is small, since the probability density proportional to Eq. (10) is then sharply concentrated near the action-minimizing path(s). The preceding discussion can be made rigorous using large deviation theory [Freidlin and Wentzell, 1998].

Suppose  $x_1$  and  $x_2$  are separated by a single “mountain pass,” in the sense that the path minimizing  $\max_{0 < t < T} V(z(t))$  goes directly uphill from  $x_1$  to a saddle, then directly downhill to  $x_2$ . Then this MEP achieves the minimum action in the limit  $T \rightarrow \infty$ . To see why, consider any path  $z(t)$ , and let  $\tau$  be the time when  $V(z(t))$  achieves its maximum. Then

$$\begin{aligned} S_T(z) &\geq \int_0^\tau |\dot{z} + \nabla V(z)|^2 dt \\ &= \int_0^\tau (|\dot{z} - \nabla V(z)|^2 + 4\dot{z} \cdot \nabla V(z)) dt \\ &\geq 4(V(z(\tau)) - V(x_1)) \\ &\geq 4(V(x_{\text{mp}}) - V(x_1)) \end{aligned}$$

where  $x_{\text{mp}}$  is the mountain pass (the saddle achieving the min/max). To achieve this bound,  $z$  would have to pass through  $x_{\text{mp}}$ , doing steepest ascent ( $\dot{z} = \nabla V(z)$ ) until its arrival then steepest descent ( $\dot{z} = -\nabla V(z)$ ) thereafter. This is not possible for finite  $T$ —both steepest ascent and steepest descent take an infinite amount of time to leave one critical point and arrive at another. However an approximation argument shows that the bound is achieved in the limit  $T \rightarrow \infty$ .

In general, a pair of local minima may not be separated by a single mountain pass. They may instead be separated by a sequence of mountain passes. Then, when the system leaves  $x_1$  it will typically cross a single mountain pass and go to a metastable state other than  $x_2$ . One can of course look for a sequence of such transitions leading ultimately to  $x_2$ . Since such a sequence determines an MEP, the string method can be useful as a search tool.

## 2 Limit sets and sufficient conditions for convergence

In this section we state the basic properties of the limit sets such as closedness, connectedness, and being a union of critical points and heteroclinic trajectories. We also formulate a sufficient condition for convergence to an MEP.

## 2.1 Basic properties of limit sets

For our analysis in Section 4 we will need to consider the time evolution not only of curves but also of other compact and connected subsets of  $\mathbb{R}^n$  according to  $\phi_t = -\nabla V(\phi)$ . Therefore, we will establish the basic properties of the limit sets for an arbitrary compact (initially) and connected subset of  $\mathbb{R}^n$ . First we will extend the definition of the limit set. Suppose  $\phi$  is a subset of  $\mathbb{R}^n$  evolving in time  $t$ , and  $\phi = \phi_0$  at  $t = 0$ . We shall use the following notation. If  $x \in \phi_0$  is a fixed point then  $\phi(x, t)$  is the orbit of the point  $x$ .

**Definition 3.**  $x_0$  is a limit point of a set  $\phi$  if there exist sequences  $\{x_n\}_{n=1}^\infty$  and  $\{t_n\}_{n=1}^\infty$  such that  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$  and

$$\lim_{n \rightarrow \infty} \phi(x_n, t_n) = x_0. \quad (12)$$

The collection of limit points of the set  $\phi$  is its limit set.

**Theorem 1.** Suppose the potential  $V(x)$  satisfies conditions (1), (2) and (3') from Section 1.2. Let the set  $\phi(x_0, t)$  be evolving according to Eq. (5) and  $\phi(x_0, 0) = \phi_0 \subset \mathbb{R}^n$  be the initial set. Then the limit set of  $\phi$  is closed, compact and consists only of critical points and heteroclinic trajectories.

This theorem follows from Lemmas A-4 and A-6 in Appendix A.

## 2.2 Sufficient conditions for convergence to an MEP

Theorem 1 shows that the limit set of a path is closed, connected, and consists of the critical points and heteroclinic trajectories. However, the limit set of a path is not necessarily a curve. It can be multidimensional, as we will show in Section 3. Then the path might either fail to converge to its limit set as in Sections 3.2 and 3.5, or it might converge and hence fill the entire higher dimensional set as in Section 3.3.

We start by showing that if a potential has only a finite number of critical points and all of them are of Morse index  $\leq 1$ , then the limit set  $L$  of any path is a curve.

**Theorem 2.** Suppose a potential  $V(x)$  satisfies conditions (1), (2) and (3') from Section 1.2. Suppose the path  $\phi(\alpha, t)$  evolves according to Eq. (5) and satisfies  $\phi(\alpha, 0) = \phi_0(\alpha)$ . Suppose  $V(x)$  has only a finite number of critical points, and all of its critical points are of Morse index  $\leq 1$ . Then the limit set  $L$  of the path  $\phi$  is a curve.

*Proof.* The limit set  $L$  of the path  $\phi$  may consist only of heteroclinic trajectories and critical points according to Theorem 1. Each heteroclinic trajectory emanates from some critical point of Morse index 1. Since there is a finite number of Morse index 1 critical points, and only two trajectories emanate from each of them, there can be at most a finite number of heteroclinic trajectories. Hence  $L$  contains at most a finite number of heteroclinic trajectories. Since  $L$  is connected, it must be a curve.  $\square$

Next we state a theorem asserting that if the limit set  $L$  of a path  $\phi$  is a curve, then the path  $\phi$  converges to  $L$ .

**Theorem 3.** Suppose a potential  $V(x)$  and an initial path  $\phi_0(\alpha)$  are such that conditions (1), (2), and (3') from Section 1.2 hold. Suppose that the limit set  $L$  of the path  $\phi$  is a curve. Then the path  $\phi$  converges to its limit set  $L$ .

**Corollary 4.** Let all of the conditions of Theorem 2 hold. Then the limit set of the path is a curve and the path converges to it.

The proof of Theorem 3 and further results on the limit sets of paths will be presented in Section 4. These results include a criterion for the limit set to be a curve and involve a theory of tubes of trajectories. Prior to introducing this material we would like to provide some examples illustrating the possible complex behavior of a path evolving according to Eq. (4).

### 3 Examples

This section provides examples in which the string evolution by Eq. (5) does not converge to an MEP. We start, in Section 3.1, with an easy example showing the importance of the hypothesis that the sublevel sets of  $V$  are compact. Then we turn, in Sections 3.2-3.4, to more interesting examples in which the evolution is periodic and the limit set is a 2D surface (Section 3.2) or the evolution is recurrent and the limit set is a solid region (Sections 3.3 and 3.4 ). Finally, we give examples in Section 3.5 showing how the influence of an index  $\geq 2$  critical point can be very important even for initial paths that don't go anywhere near it.

The examples in Sections 3.2–3.5 all involve systems with a continuum of MEP's. This is no accident: as we have shown in Section 2.2, if critical points in the limit set are joined by just a finite number of MEP's then the string method must converge to an MEP.

#### 3.1 What if the sublevel sets are noncompact?

In this section we present an example justifying the significance of condition (3') from Section 1.2 that the sets  $S(\phi_0(\alpha), V(\phi_0(\alpha)))$  s are compact for every point  $\phi_0(\alpha)$  of the initial path. Suppose the potential is given by

$$V(x, y) = -\frac{(x+1)^2 + (y^2 - 1)^2}{x^3}, \quad (x, y) \in (-\infty, 0) \times \mathbb{R}. \quad (13)$$

This potential is bounded from below, symmetric w.r.t the  $x$ -axis and has local minima at  $(-1, 1)$  and  $(-1, -1)$  and saddles at  $(-3, 1)$  and  $(-3, -1)$ . Let the initial path be a straight line segment connecting the two minima. The set  $S((-1, 0), V(-1, 0))$  is noncompact (see Fig. 2). Hence condition (3') does not hold. The vector field associated with this potential is given by

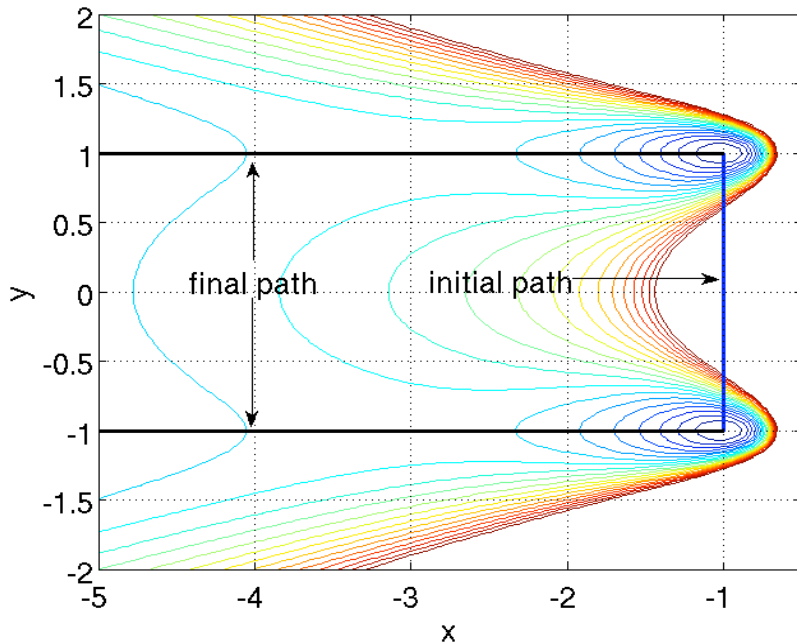


Figure 2: The time evolution of a path in the potential given by Eq. (13). The path goes to infinity.



$$\begin{aligned}
-V_x &= \frac{2(x+1)}{x^3} - 3\frac{(x+1)^2 + (y^2-1)^2}{x^4}, \\
-V_y &= \frac{4y(y^2-1)}{x^3}.
\end{aligned}$$

We consider the evolution of the path according to Eq. (5). In particular, we consider the evolution of the point of the path lying on the negative  $x$ -semiaxis. It is easy to see that  $V_y = 0$  on the negative  $x$ -semiaxis. Hence the point  $(-1, 0)$  of the initial path will stay on the negative  $x$ -semiaxis. Therefore, its time evolution is given by

$$x_t = -\frac{x^2 + 4x + 6}{x^4}, \quad x(0) = -1.$$

Then

$$\begin{aligned}
-\int \frac{x^4}{x^2 + 4x + 6} &= t + C, \\
-\frac{x^3}{3} + 2x^2 - 10x + 8 \log((x+2)^2 + 2) + \frac{28}{\sqrt{2}} \arctan \frac{x+2}{\sqrt{2}} &= t + C.
\end{aligned} \tag{14}$$

Eq. (14) shows that if  $t \rightarrow \infty$  then  $x \rightarrow -\infty$ . The limit set of the path consists of the saddles at  $(-3, 1)$  and  $(-3, -1)$ , the minima at  $(-1, 1)$  and  $(-1, -1)$ , and four trajectories emanating from the saddles, two of which go to the minima, and the other two - to infinity. Geometrically, the limit set consists of two half-lines (Fig. 2). The path converges to every point of its limit set which is 1D, however, the limit set is not connected and not of finite length.

## 3.2 Examples of revolving paths

One can easily visualise the meaning of the sentence “a path converges to its limit set” in the sense of Definitions 1 and 2. This means that the motion of the curve  $C$  corresponding to the path slows down and the curve, as a geometrical object, approaches its limit set  $L$  (though every point of the curve might keep moving within the curve). Does the evolution of the curve always slow down? The answer is no. First, in Section 3.2.1 we will give a simple 2D example of an infinite logarithmic spiral which endlessly revolves in a quadratic potential. This example serves only as a motivation, as the initial path does not satisfy the requirement of being of finite length and having ends at two distinct minima. This potential has only one minimum, and the spiral has one end at the minimum, while the other one lies at infinity. Nonetheless the rotation of the spiral gave us a hint for a more complicated construction of an endlessly revolving path of finite length in 3D. Roughly speaking, our revolving path in 3D is a stereographic projection of the 2D spiral onto a deformed sphere. We present it in Section 3.2.2.

### 3.2.1 Motivation

Suppose the potential is given by

$$V(x, y) = \frac{x^2}{2} + \frac{y^2}{2},$$

so  $-\nabla V(x, y) = (-x, -y)$ . Consider the logarithmic spiral, given in polar coordinates by

$$r(\theta) = e^{-\theta \cot \alpha}.$$

with  $0 < \alpha < \pi/2$  (so  $\cot \alpha > 0$ , and  $r \rightarrow 0$  as  $\theta \rightarrow \infty$ ) (Fig. 3). One of the properties of logarithmic spiral is that it has the same angle  $\alpha$  with all lines passing through the origin.

We claim that the spiral evolves according to Eq. (4) by clockwise rotation at constant angular velocity  $\omega = \tan \alpha$ .

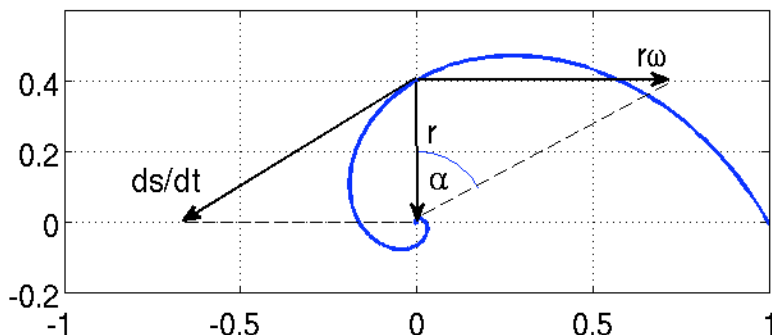


Figure 3: The rotating spiral. The motion of each point of the path, representing the spiral can be decomposed into the sum of rotation around the origin with the constant angular velocity  $\omega = \tan \alpha$  and contraction along itself toward the origin with speed  $\frac{ds}{dt} = \frac{r}{\cos \alpha}$ .

According to Eq. (5) each point of the spiral moves toward the origin with speed  $r = \sqrt{x^2 + y^2}$ . In order to understand how the spiral evolves as a geometrical object, we decompose the velocity vector of each point of the spiral into a sum of the following two vectors (Fig. 3). One of them, with length  $\frac{ds}{dt} = \frac{r}{\cos \alpha}$ , is tangent to the spiral, and the other one, with length  $r\omega = r \tan \alpha$ , is normal to the vector field. Since the spiral has the same angle  $\alpha$  with all of the lines through the origin, the angular velocity  $\omega = \tan \alpha$  is the same at every point of the spiral. Therefore, the motion of the spiral can be viewed as a sum of rotation around the origin with the constant angular velocity  $\omega$  and contraction along itself toward the origin with speed  $\frac{r}{\cos \alpha}$ . Therefore the motion of the spiral according to Eq. (4) reduces to rotation around the origin with a constant angular velocity  $\omega = \tan \alpha$ . The limit set of any path representing the spiral is the whole  $\mathbb{R}^2$ .

### 3.2.2 A revolving path in 3D

We would like to construct a path whose time evolution according to Eq. (4) is rotation around the  $z$ -axis at a constant given angular velocity  $\omega$ .

Take a cylindrically symmetric and twice continuously differentiable potential in  $\mathbb{R}^3$

$$V(x, y, z) = -\cos \sqrt{x^2 + y^2} \sin 2z. \quad (15)$$

It is convenient to use cylindrical coordinates  $(x, y, z) = (r \cos \theta, r \sin \theta, z)$ . Since  $V$  is cylindrically symmetric with a minimum at  $(0, 0, \pi/4)$  and a maximum at  $(0, 0, -\pi/4)$  the heteroclinic orbits of  $(x_t, y_t, z_t) = -\nabla V$  are obtained by rotation of the family of the heteroclinic orbits of the corresponding potential

$$V(r, z) = -\cos r \sin 2z$$

around the  $z$ -axis. Let us pick one such orbit shown in Fig. 4(a). Let  $(r(s), z(s))$  be a unit speed parametrization of this curve with  $s = 0$  at its midpoint. Set  $g(s) = |\nabla V(r(s), z(s))|$ . Now consider a path in  $\mathbb{R}^3$  lying on the surface of revolution of the curve  $(r(s), z(s))$  and given by

$$\phi(s) = (x(s), y(s), z(s)) = (r(s) \cos \theta(s), r(s) \sin \theta(s), z(s)), \quad (16)$$

where  $\theta(s)$  is to be determined as follows. Since the path rotates with the given constant angular velocity  $\omega$ ,

$$g(s) \tan \alpha(s) = r(s)\omega,$$

where  $\alpha(s)$  is the angle between the tangent vector to the curve at the point  $(r(s), \theta(s), z(s))$  and the gradient of the potential at this point (Figure 4(b)). Then  $\theta(s)$  is found from

$$\theta(s) = \int_0^s \frac{\tan \alpha(s') ds'}{r(s')} = \int_0^s \frac{\omega}{g(s')} ds'. \quad (17)$$

Note that since  $g$  vanishes linearly at each endpoint,  $\theta(s) \rightarrow \pm\infty$  as  $s$  approaches the ends of its interval of definition. This means that the path performs an infinite number of revolutions as it approaches the critical points (Fig. 4(b)).

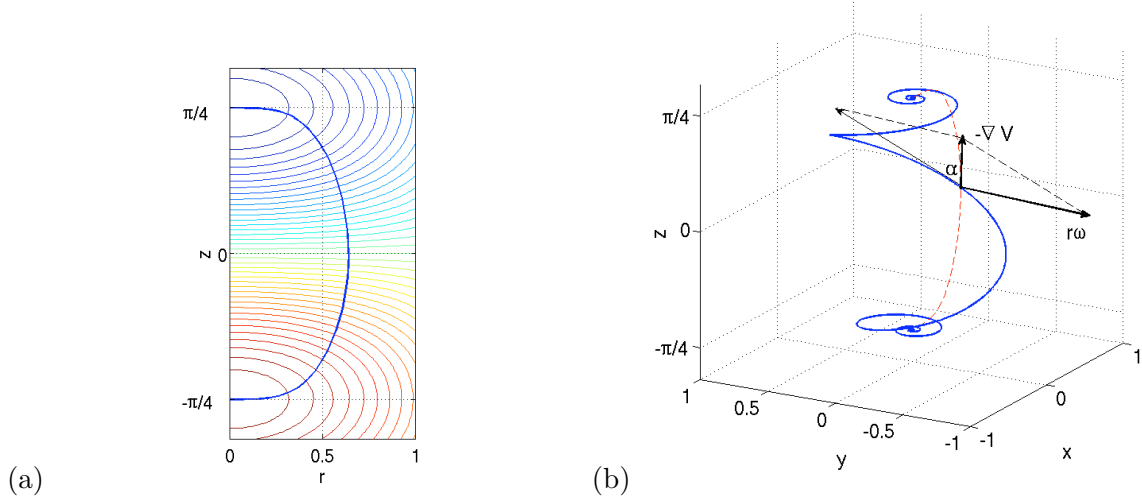


Figure 4: Construction of the initial path such that its evolution according to Eq. (4) is rotation around the  $z$ -axis with constant angular velocity  $\omega$ . (a) The heteroclinic trajectory  $(r(s), z(s))$ . (b) The path  $(r(s), \theta(s), z(s))$  constructed for  $\omega = 5$  and the decomposition of the velocity vector  $-\nabla V$  into two components. One of them, of length  $r\omega$ , is normal to  $-\nabla V$  and tangent to the surface of revolution of the trajectory  $(r(s), z(s))$ . The second one is tangent to the path.

An example of such a path for  $\omega = 5$  is shown in Fig. 4(b). By construction, the motion of this path according to the gradient flow can be decomposed into rotation around the  $z$ -axis with the constant angular velocity  $\omega$  and contraction along itself toward the minimum with the speed  $\frac{g(s)}{\cos \alpha(s)}$ . Then the motion of the path according to Eq. (4) is merely rotation around the  $z$ -axis.

The limit set of this path is the surface of revolution of the heteroclinic trajectory in Fig. 4a. Clearly this path does not converge to its limit set.

Let us show that this path has finite length. Indeed,

$$\begin{aligned} \phi_s &= (r_s \cos \theta, r_s \sin \theta, z_s) + (-r \sin \theta, r \cos \theta, 0) \theta_s \\ &= \frac{-\nabla V}{g(s)} + \frac{\omega}{g(s)} (-y, x, 0). \end{aligned}$$

Then

$$|\phi_s| = \left| \frac{-\nabla V}{g(s)} + \frac{\omega}{g(s)} (-y, x, 0) \right| \leq 1 + \omega \frac{r(s)}{g(s)},$$

Now observe that both  $g(s)$  and  $r(s)$  tend to zero linearly as  $s$  approaches either endpoint of its interval of definition. Hence the length which is the integral of  $|\phi_s|$  over the interval of definition of  $s$  is finite. In fact, as  $\phi$  approaches  $(0, 0, \pm\pi/4)$  its shape is asymptotically a logarithmic spiral.

### 3.3 A path which fills a region in 2D

We have seen in Section 3.2 that one can construct a path whose limit set is 2D and the path does not converge to its limit set.

In this section we will present an example where the limit set of a path of finite length is a 2D region and the path converges to its limit set, i.e., it fills the region.

Consider a potential in 2D given by

$$V(x, y) = x^4 - 2x^2 + y^4 - 2y^2. \quad (18)$$

This potential has four minima at  $(-1, -1)$ ,  $(1, -1)$ ,  $(-1, 1)$  and  $(1, 1)$ , a maximum at  $(0, 0)$ , and four saddle points at  $(-1, 0)$ ,  $(0, -1)$ ,  $(1, 0)$  and  $(0, 1)$ . We will construct the initial path as shown in Figure 5(a). Suppose the initial path  $\phi_0$  starts at the minimum  $(-1, -1)$ , passes

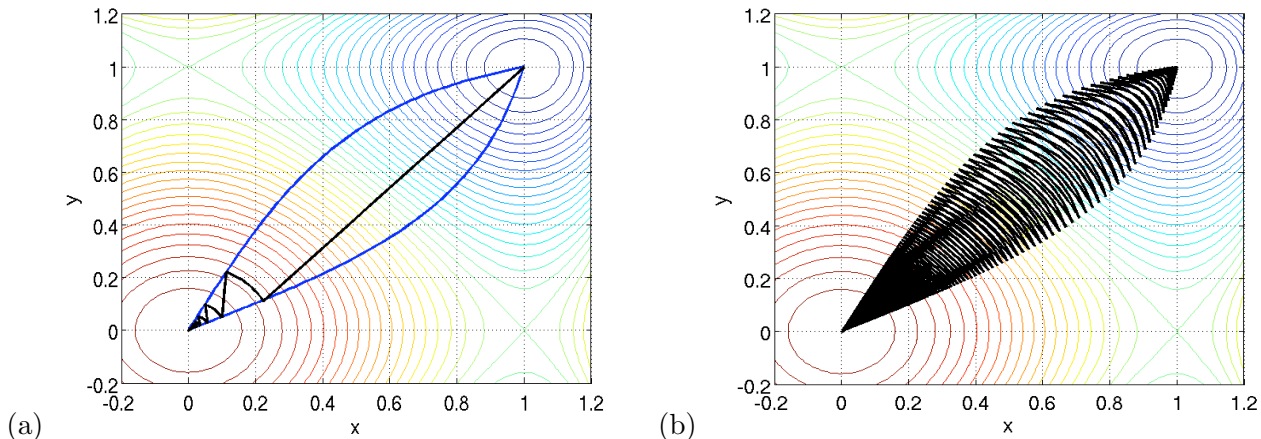


Figure 5: An example where the limit set of a path is a 2D region and the path converges to its limit set. (a) The initial path. (b) The path at some time  $t > 0$ .

through the maximum at  $(0, 0)$ , and ends at the minimum  $(1, 1)$ . As  $\phi_0$  passes through the maximum, the evolving path  $\phi$  passes through it at all times. We can choose the initial path between  $(-1, -1)$  and  $(0, 0)$  to be a segment of a straight line. It is easy to check that it is an integral curve, hence, this part of the path will not evolve in time. Hence we can focus only on the path between  $(0, 0)$  and  $(1, 1)$ .

We take two integral curves connecting the maximum at  $(0, 0)$  and the minimum at  $(1, 1)$  and symmetric w. r. t. the line  $x = y$ . Then we construct the initial path from the arcs of the circles

$$x^2 + y^2 = \frac{1}{n^4}, \quad \text{radii } r_n = \frac{1}{n^2}, \quad n = 2, 3, 4, \dots \quad (19)$$

lying between these two integral curves, and segments of straight lines connecting them as shown in Fig. 5(a). This initial path has finite length. Indeed, it follows from the triangle inequality that the total length of its linear segments does not exceed the total length of its circular segments plus  $\sqrt{2}$  (the upper bound for the largest linear segment (see Fig. 5(a))). The total length of its circular segments is less than the convergent sum

$$\sum_{n=1}^{\infty} \frac{\pi}{2n^2} < \infty.$$

Therefore, the length of the initial path is finite.

As  $t \rightarrow \infty$  the path fills the region between these two integral curves (Fig. 3.3(b)). A rigorous proof that this path converges to its limit set is provided in Appendix B.

### 3.4 A path whose limit set is multi-dimensional

In this section we show that the limit set of a path can be multi-dimensional. First we explain the idea of the construction, then we provide an example.

Suppose a potential  $V(x)$  defined in  $\mathbb{R}^n$  has a critical point  $S$  of Morse index  $3 \leq k \leq n$  and a minimum  $M$  such that one can choose a  $k$ -dimensional body  $B$  consisting of heteroclinic trajectories starting at  $S$  and arriving at  $M$ . For simplicity we assume that the body lies completely in a  $k$ -dimensional subspace of  $\mathbb{R}^n$ . Let  $(x_1, \dots, x_{k-1}, z)$  be Cartesian coordinates in the subspace. Let  $S = (0, \dots, 0, -z_0)$  and  $M = (0, \dots, 0, z_0)$  (Fig. 6). Suppose the plane  $\{z = 0\}$  intersects the body  $B$  so that each heteroclinic trajectory in  $B$  crosses the plane exactly once. Let us choose a continuous curve

$$\gamma = \{\gamma(\theta) \mid 0 \leq \theta < \infty\} = \{x_1^0(\theta), x_2^0(\theta), \dots, x_{k-1}^0(\theta), 0 \mid 0 \leq \theta < \infty\} \quad (20)$$

such that

- (i) the curve  $\gamma$  lies in the section  $\sigma = B \cap \{z = 0\}$ ,
- (ii) any infinite interval of the curve  $\gamma$  of the form  $\{\gamma(\theta) \mid \theta_0 \leq \theta < \infty\}$  is dense in  $\sigma$ , and
- (iii) the velocity vector of the curve  $\frac{d}{d\theta}\gamma$  is bounded.

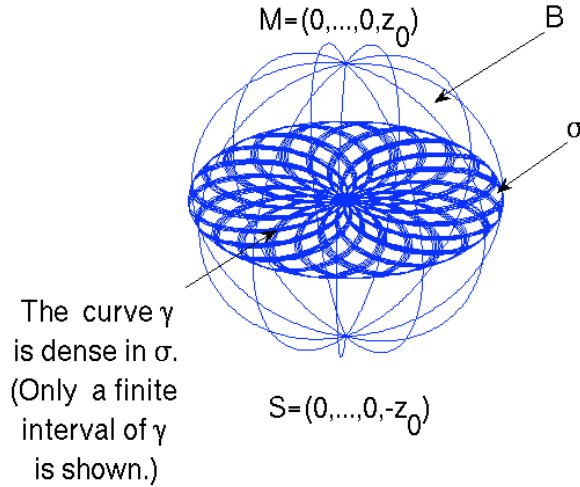


Figure 6: Illustration for Section 3.4. The body  $B$  consists of heteroclinic trajectories starting at the critical point  $S = (0, \dots, 0, -z_0)$  and arriving at the minimum  $M = (0, \dots, 0, z_0)$ . The curve  $\gamma$  lies in the section  $\sigma$  of the body  $B$  by the plane  $\{z = 0\}$  and is dense in  $\sigma$ .

Then we choose an initial path  $\phi_0$  of finite length such that the path  $\phi(t)$  evolving according to

$$\phi_t = -\nabla V(\phi), \quad \phi(0) = \phi_0, \quad (21)$$

intersects the plane  $\alpha = \{z = 0\}$  following the curve  $\gamma$ . In other words, at each moment of time  $t$  the curve  $\phi(t)$  intersects the section  $\sigma$  at some point  $\gamma(\theta(t))$  of the curve. Our choice of the initial path is such that the function  $\theta(t)$  is strictly increasing (for simplicity),  $\theta(0) = 0$  and  $\theta(t) \rightarrow \infty$  as  $t \rightarrow \infty$ . In terms of the flow  $g^t$  corresponding to Eq. (21) this implies that

$$g^{[0, \infty)}\phi_0 \cap \sigma = \gamma. \quad (22)$$

We will prove the existence of such a path later. First we prove

**Claim 1.** *In the construction above, the limit set of the path  $\phi$  is the whole body  $B$ .*

*Proof.* First we foliate the body  $B$  with the images of the section  $\sigma$  under the flow  $g^t$ ,  $t \in \mathbb{R}$ , i.e,

$$B = \bigcup_{t \in \mathbb{R}} g^t \sigma. \quad (23)$$

Therefore, for every point  $x^* \in B$  there is  $t^* \in \mathbb{R}$  such that  $x^* \in g^{t^*} \sigma$ . From Eq. (22) we have

$$g^{[t^*, \infty)} \phi_0 \cap g^{t^*} \sigma = g^{t^*} \gamma. \quad (24)$$

Since the evolution of the path  $\phi$  starts at time zero, we need to take care of the case  $t^* < 0$ . In this case, the evolving path follows not the whole curve  $g^{t^*} \gamma$  but its infinite interval

$$\{g^{t^*} \gamma(\theta) \mid \theta(-t^*) \leq \theta < \infty\}.$$

It follows from Condition (ii) on the curve  $\gamma$  that this interval is still dense in  $g^{t^*} \sigma$ . Therefore, for any  $t^* \in \mathbb{R}$  we have that

$$g^{[0, \infty)} \phi_0 \cap g^{t^*} \sigma$$

is dense in  $g^{t^*} \sigma$ . Then it follows from Eq. (23) that  $g^{[0, \infty)} \phi_0$  is dense in  $B$ .  $\square$

Now it remains to show that such a curve  $\phi_0$  exists. Suppose that the heteroclinic trajectories whose union is the body  $B$ , can be parametrized by the  $z$ -coordinate and the points  $(x_1^0, x_2^0, \dots, x_{k-1}^0)$  where they meet the plane  $\{z = 0\}$ :

$$x_j = x_j(z, x_1^0, \dots, x_{k-1}^0).$$

First we choose a function  $z(\theta)$  such that

$$z(0) = 0 \quad \text{and} \quad z(\theta) \rightarrow -z_0 \quad \text{as} \quad \theta \rightarrow \infty. \quad (25)$$

Then we take this  $z(\theta)$  and the dense curve  $\gamma(\theta) = (x_1^0(\theta), \dots, x_{k-1}^0(\theta))$  (Eq. (20)) and construct the piece of the initial path from the critical point  $S$  to the plane  $\{z = 0\}$  as follows

$$\phi_0^1(\theta) = (z(\theta), x_1(z(\theta), x_1^0(\theta), \dots, x_{k-1}^0(\theta)), \dots, x_{k-1}(z(\theta), x_1^0(\theta), \dots, x_{k-1}^0(\theta))). \quad (26)$$

For this choice of  $\phi_0^1$  we have

$$g^{[0, \infty)} \phi_0^1 \cap \sigma = \gamma.$$

The second piece of the initial path  $\phi_0^2$  can be an arbitrary path of finite length from the plane  $\{z = 0\}$  to the minimum  $M$ . Then the initial path  $\phi_0$  is the union of these two paths, i.e.,

$$\phi_0 = \phi_0^1 \cup \phi_0^2.$$

Now we show that the initial path can be of finite length. We will present an example of a potential and a function  $z(\theta)$  where this is so.

Consider a potential  $V(x)$ ,  $x = (x_1, \dots, x_n)$ , is given by

$$V(x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_n) = -\cos \sqrt{x_1^2 + \dots + x_{k-1}^2} \sin z + x_{k+1}^2 + \dots + x_n^2 \quad (27)$$

in the domain

$$\Omega = \{x \in \mathbb{R}^n \mid x_1^2 + \dots + x_{k-1}^2 < 1\}.$$

The domain  $\Omega$  is chosen so that  $V(x)$  is a Morse potential in it.

This potential has an index  $k$  critical point  $S$  at  $(0, \dots, 0, -\frac{\pi}{2}, 0, \dots, 0)$  and a minimum  $M$  at  $(0, \dots, 0, \frac{\pi}{2}, 0, \dots, 0)$ . All of the heteroclinic trajectories starting at  $S$  and arriving at  $M$  lie in the  $k$ -dimensional subspace  $(x_1, \dots, x_{k-1}, z, 0, \dots, 0)$ . Therefore, our construction will take place only in this subspace, and we will ignore the other coordinates. We define the body  $B$  as

a union of all heteroclinic trajectories from  $S$  to  $M$  passing through the unit disc centered at the origin and normal to the  $z$ -axis

$$D = \{(x_1, \dots, x_{k-1}, 0) \mid x_1^2 + \dots + x_{k-1}^2 \leq 1\}.$$

These trajectories are readily found in the generalized cylindric coordinates

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \vdots \\ x_{k-2} \\ x_{k-1} \\ z \end{pmatrix} = \begin{pmatrix} r \cos \theta_1 \\ r \sin \theta_1 \cos \theta_2 \\ \vdots \\ r \sin \theta_1 \dots \sin \theta_{m-1} \cos \theta_m \\ \vdots \\ r \sin \theta_1 \dots \sin \theta_{k-3} \cos \theta_{k-2} \\ r \sin \theta_1 \dots \sin \theta_{k-3} \sin \theta_{k-2} \\ z \end{pmatrix} \quad (28)$$

where  $r = \sqrt{x_1^2 + \dots + x_{k-1}^2}$ . Indeed, the gradient flow for the potential (27) is given by

$$\begin{cases} \dot{r} = -\sin r \sin z, \\ \dot{z} = \cos r \cos z, \\ \dot{\theta}_j = 0, \quad j = 1, \dots, k-2. \end{cases}$$

Hence  $\theta_j(t) = \text{const}$ ,  $j = 1, \dots, k-2$ , and  $r$  can be found as a function of  $z$  from

$$\frac{dr}{dz} = -\frac{\sin r \sin z}{\cos r \cos z}.$$

Integrating this equation we get

$$r(z, r_0) = \arcsin(\sin r_0 \cos z), \quad (29)$$

where  $r_0$  is the value of  $r$  at  $z = 0$  or the  $r$ -coordinate of the point at which the trajectory crosses the unit disc  $D$ .

Next, we choose a curve

$$\gamma = \{r_0(\theta), \theta_1(\theta), \dots, \theta_{k-2}(\theta) \mid 0 \leq \theta < \infty\} \quad (30)$$

dense in the disc  $D$ . Then we define the initial path parametrized by  $\theta$ . We pick a function  $z(\theta)$  such that  $z(0) = 0$  and  $z(\theta) \rightarrow -\frac{\pi}{2}$  as  $\theta \rightarrow \infty$ , e.g.,

$$z(\theta) = -\frac{\pi}{2} + \frac{\pi}{2}e^{-\theta}. \quad (31)$$

Next, we define  $r(\theta)$  using Eqs. (29) and (31) by

$$r(\theta) = r(z(\theta), r_0(\theta)) = \arcsin(\sin r_0(\theta) \cos z(\theta)). \quad (32)$$

Finally, we take the functions  $\theta_j(\theta)$ ,  $j = 1, \dots, k-2$  from Eq. (30). Thus, the piece of the initial path from the critical point  $S$  to the plane  $\{z = 0\}$  is given by

$$\phi_0^1(\theta) = (z(\theta), r(z(\theta), r_0(\theta)), \theta_1(\theta), \dots, \theta_{k-2}(\theta)), \quad (33)$$

where  $\theta_j$ ,  $j = 1, \dots, k-2$ ,  $z$ , and  $r$  are defined in Eqs. (30), (31), and (32) respectively. The remaining piece of the initial path  $\phi_0^2$  from the plane  $\{z = 0\}$  to the minimum  $M$  can be an arbitrary path of finite length lying in the body  $B$ . For example, it can be chosen to follow the half-trajectory from the point  $(r_0(0), \theta_1(0), \dots, \theta_{k-2}(0))$  of the disc  $D$  to the minimum  $M$ . It

remains to show that the initial path just constructed is of a finite length. Since  $\phi_0 = \phi_0^1 \cup \phi_0^2$  and the length of  $\phi_0^2$  is finite it suffices to check that the length of  $\phi_0^1$  given by

$$l(\phi_0^1) = \int_0^\infty \sqrt{\left(\frac{dx_1}{d\theta}\right)^2 + \dots + \left(\frac{dx_{k-1}}{d\theta}\right)^2 + \left(\frac{dz}{d\theta}\right)^2} d\theta \quad (34)$$

is finite. It follows from Eq. (28) that

$$\frac{dx_j}{d\theta} = \frac{\partial x_j}{\partial r} r_\theta + r \sum \left( \prod \left\{ \begin{array}{c} \pm \sin \theta_m \\ \pm \cos \theta_m \end{array} \right\} \frac{d\theta_j}{d\theta} \right). \quad (35)$$

Squaring Eq. (35) and adding, we get

$$\left(\frac{dx_1}{d\theta}\right)^2 + \dots + \left(\frac{dx_{k-1}}{d\theta}\right)^2 \leq r_\theta^2 + r^2 \sum \left(\frac{d\theta_j}{d\theta}\right)^2. \quad (36)$$

The crossterms from the right-hand side of Eq. (35) involving  $rr_\theta$  add to zero. Furthermore, it follows from Eq. (29) that

$$r_\theta = \frac{\partial r}{\partial z} \frac{dz}{d\theta} + \frac{\partial r}{\partial r_0} \frac{dr_0}{d\theta} \quad (37)$$

where

$$\frac{\partial r}{\partial z} = \frac{-\sin r_0 \sin z}{\sqrt{1 - \sin^2 r_0 \cos^2 z}}, \quad (38)$$

$$\frac{\partial r}{\partial r_0} = \frac{\cos r_0 \cos z}{\sqrt{1 - \sin^2 r_0 \cos^2 z}}. \quad (39)$$

Since  $r_0 < 1 < \pi/2$ , the denominator in Eqs. (38) and (39) is bounded. Moreover,

$$|\cos z(\theta)| = \left| \cos \left( -\frac{\pi}{2} + \frac{\pi}{2} e^{-\theta} \right) \right| = \left| \sin \left( \frac{\pi}{2} e^{-\theta} \right) \right| \leq \frac{\pi}{2} e^{-\theta}. \quad (40)$$

Therefore, we can bound  $|r_\theta|$  using Eqs. (31) and (37)-(40)

$$|r_\theta| < C e^{-\theta} \left( \left| \frac{\partial r}{\partial z} \right| + \left| \frac{dr_0}{d\theta} \right| \right). \quad (41)$$

Then from Eqs. (34), (36), and (41) we have

$$l(\phi_0^1) = \int_0^\infty d\theta C e^{-\theta} \left[ \left( \left| \frac{\partial r}{\partial z} \right| + \left| \frac{dr_0}{d\theta} \right| \right)^2 + \sum_{j=1}^{k-2} \left| \frac{d\theta_j}{d\theta} \right|^2 + 1 \right]^{1/2}. \quad (42)$$

In summary, to make the initial curve of finite length it suffices that the integral in Eq. (42) be finite. Condition (iii) on the curve  $\gamma$  and Eq. (38) guarantee that the expression in square brackets in Eq. (42) is bounded. This implies that the integral in Eq.(42) is finite. Thus we have shown that for given  $k$  and  $n$ ,  $3 \leq k \leq n$ , we can find a potential in  $\mathbb{R}^n$  and construct an initial path of finite length such that the limit set of of the path is  $k$ -dimensional.

If  $k = 3$  the cylindrical coordinates are  $(z, r, \theta)$ . We can choose the curve  $\gamma$  given by

$$r_0(\theta) = \frac{1}{2} + \frac{1}{2} \cos \beta \theta \quad (43)$$

where  $\beta$  is irrational. This curve is dense in the unit disc, and  $\frac{dr_0}{d\theta}$  is bounded.

**Remark** In the construction just presented we have chosen the function  $z(t) = -\frac{\pi}{2} + \frac{\pi}{2} e^{-\theta}$ . This choice makes the path  $\phi = g^t \phi_0$  endlessly wander in its limit set  $B$  without convergence. However, another choice of  $z(t)$  might cause the path to fill the body  $B$ , i.e., converge to its limit set.



### 3.5 Potentials with critical points of Morse index $\geq 2$

In this section, we present examples illuminating the impact of a critical point of Morse index 2 on the limit set of a path. First, we show how a path, initially not passing through the stable manifold of a saddle of Morse index  $\geq 2$  can end up passing through such a saddle. Second, we demonstrate that a Morse index  $\geq 2$  critical point might make the limit set very sensitive to the initial path even if the limit set is a curve. Finally, we give an example of a path whose limit set contains a 2D surface while the initial path neither passes through a Morse index two saddle nor intersects its stable manifold.

#### 3.5.1 The ‘‘Umbrella’’ potential

Consider the potential in 3D given by

$$V(x, y, z) = 10u^2(u^2 - 1) + 0.07x^2 + y^2 + \frac{15}{19\sqrt{19}}z^2, \quad \text{where } u = \frac{x^2 + z^2}{4} + y. \quad (44)$$

This potential has the following seven critical points:

1. a minimum at  $M_1 = \left(0, -\frac{1.5}{\sqrt{5}}, 0\right)$ ;
2. minima  $M_2 = (-\sqrt{2.24}, 0.14, 0)$  and  $M_3 = (\sqrt{2.24}, 0.14, 0)$ ;
3. a Morse index 1 saddle  $S_1 = (0, 0, 0)$  with unstable direction  $y$ ;
4. a Morse index 2 saddle  $A = \left(0, \frac{1.5}{\sqrt{5}}, 0\right)$  with unstable directions  $x$  and  $z$ ;
5. Morse index 1 saddles  $S_2 = \left(0, \frac{30}{19\sqrt{19}}, \sqrt{\frac{108}{19\sqrt{19}}}\right)$  and  $S_3 = \left(0, \frac{30}{19\sqrt{19}}, -\sqrt{\frac{108}{19\sqrt{19}}}\right)$  with unstable direction  $x$ .

All numbers in the list of critical points above are exact. The coefficients in the potential (44) have been chosen so that (i) the critical points are nondegenerate and can be found analytically, and (ii) there are only seven critical points rather than more.

The collection of the critical points and heteroclinic trajectories of potential (44) is shown in Fig. (7). The image in this figure is reminiscent of an umbrella. The Morse index 2 saddle  $A = (0, 1.5/\sqrt{5}, 0)$  is at the top of its canopy. One of the minima,  $M_1 = (0, -1.5/\sqrt{5}, 0)$  is at the end of its handle, the other two,  $M_2 = (-\sqrt{2.24}, 0.14, 0)$  and  $M_3 = (\sqrt{2.24}, 0.14, 0)$ , are at the opposite points on the edge of its canopy. The Morse index 1 saddle  $S_1 = (0, 0, 0)$  is at the middle of its handle, and the other two,  $S_2 = (0, 30/\sqrt{19\sqrt{19}}, \sqrt{108}/\sqrt{19\sqrt{19}})$  and  $S_3 = (0, 30/\sqrt{19\sqrt{19}}, -\sqrt{108}/\sqrt{19\sqrt{19}})$ , are at the opposite points on the edge of its canopy.

#### 3.5.2 Sliding to a Morse index 2 saddle

First, we demonstrate that a path not passing originally through a stable manifold of a saddle of Morse index 2 or more can evolve to a curve passing through it. Furthermore, we show that the final path depends on the eigenvalues of the critical points.

The stable manifold of the saddle  $S_1$  is a bullet-shaped surface. To one side lies the basin of attraction of the minimum  $M_1$ . The rest of space is divided by the vertical plane  $x = 0$  into the basins of attraction of the minima  $M_2$  and  $M_3$ .

We take three very close initial paths as shown in Fig. 8(a-c). These paths connect the minima  $M_1$  and  $M_2$  and consist of two line segments with the intermediate vertices at  $(-\epsilon, 0, 1.5)$ ,  $(0, 0, 1.5)$ , and  $(\epsilon, 0, 1.5)$  where  $\epsilon$  is a small positive number. Thus, the path in Fig. 8(a) lies in the union of stable manifolds of minima  $M_1$  and  $M_2$  and the saddle  $S_1$ ; the path in Fig. 8(b) lies in the union of stable manifolds of minima  $M_1$  and  $M_2$  and the saddles  $S_1$  and  $S_2$ ; and the path in Fig. 8(c) lies in the union of stable manifolds of minima  $M_1$ ,  $M_2$  and  $M_3$  and the saddles  $S_1$  and  $S_2$ . We emphasize that none of these paths intersects the stable manifold of the Morse index 2 saddle which is the set

$$\{(x, y, z) \mid x = 0, z = 0, y > 0\}. \quad (45)$$

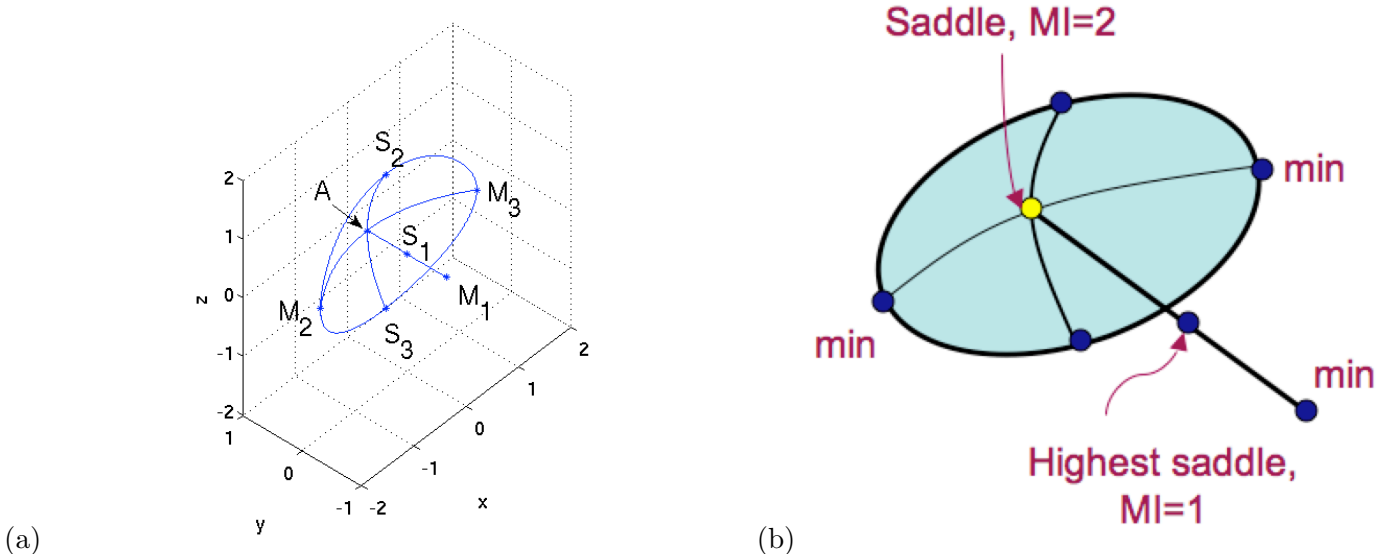


Figure 7: (a) The critical points and selected heteroclinic trajectories of the potential given by Eq. (44). (b) A schematic picture of the heteroclinic structure. The shaded area consists of heteroclinic trajectories.

However, there are only two heteroclinic trajectories coming out of the saddle  $S_1$ . One of them goes to the minimum at  $M_1$ , while the other one goes to the Morse index 2 saddle at  $A$ . Hence the final path must pass through the index 2 saddle  $A$  for any initial path connecting the minima  $S_1$  and  $S_2$ . The final paths are shown in Fig. 8(a-c). As we see, all of them indeed pass through the Morse index 2 saddle  $A$ .

It is unsurprising that the final paths pass through the particular sets of critical points. However the choice of the heteroclinic trajectory from the saddle  $A$  to saddle  $S_2$  for the final path in fig. 8(a) and to saddle  $S_3$  for the final path in fig. 8(b) is not so obvious. There are infinitely many heteroclinic trajectories from the saddle  $A$  to the saddles  $S_2$  and  $S_3$  each of which are equally suitable from the point of view of the Large Deviation Theory [Freidlin and Wentzell, 1998]. The explanation for these choices of heteroclinic trajectories is the following.

The eigenvalues of  $-\nabla\nabla V$  at the saddle  $S_1$  are  $(\lambda_x = -0.14, \lambda_y = 18, \lambda_z \approx -0.36)$ . Therefore, almost all of the trajectories entering the saddle  $S_1$  do it along the  $x$ -direction. The eigenvalues at the Morse index 2 saddle  $A$  are  $(\lambda_x \approx 0.53, \lambda_y = -36, \lambda_z \approx 0.31)$ . Therefore, almost all of the trajectories exiting this saddle do so along the  $z$ -direction.

The choice of the heteroclinic trajectory from the saddle  $A$  to the minimum  $M_2$  (Fig. 8 (a)) or the minimum  $M_3$  (Fig. 8 (c)) is determined by the points of the initial path in the neighborhood of the stable manifold of the saddle  $S_1$  which is a bullet-shaped surface. Indeed, all other points soon reach neighborhoods of the minima  $M_1$  and  $M_2$ . The points of the initial path close the stable manifold of the saddle  $S_1$  first evolve toward the saddle and approach the  $y$ -axis along the  $x$ -direction, unless the initial path crosses the stable manifold within the  $yz$ -plane as it does in Fig. 8(b). This determines the choice of the heteroclinic trajectory coming out of the saddle  $A$ .

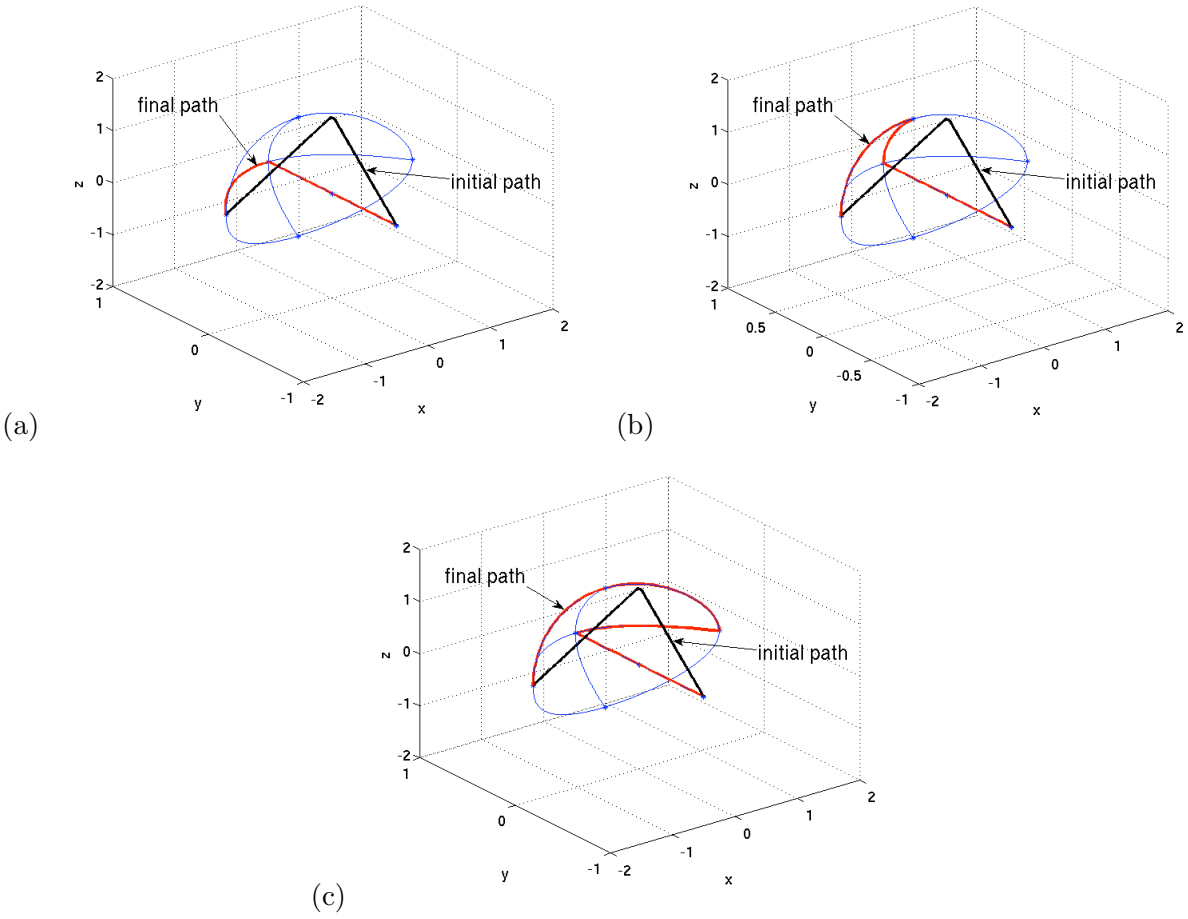


Figure 8: Examples of the initial and final paths in the potential given by Eq. (44). Three very close initial paths lead to very different final paths. The final paths are determined not only by which basins of attraction the initial paths go through, but also by the geometry of the network of the heteroclinic trajectories and the eigenvalues of the critical points.

### 3.5.3 The umbrella-shaped limit set

In the next example, we present a path whose limit set contains a 2D surface, while the initial path neither passes through the Morse index 2 saddle, nor intersects its stable manifold. The initial path, shown in Fig. 9(a), consists of the segment of a straight line from the minimum  $M_1$  to the saddle  $S_1$ , the spiral in the  $xz$ -plane

$$\phi_0 = \{(x, y, z) | x = e^{-0.1\theta} \cos \theta, y = 0, z = e^{-0.1\theta} \sin \theta, \theta \geq 0\},$$

and the segment of a straight line connecting free end of the spiral with the minimum at  $M_3$  (Fig. 9(a)). The initial path does not intersect the stable manifold of the Morse index 2 saddle given by Eq. (45). The limit set of this path is the the whole "umbrella" formed by the entire collection of the critical points and heteroclinic trajectories of the potential given by Eq. (44). The path at some moment of time  $t > 0$  is shown in Fig. 9(b).

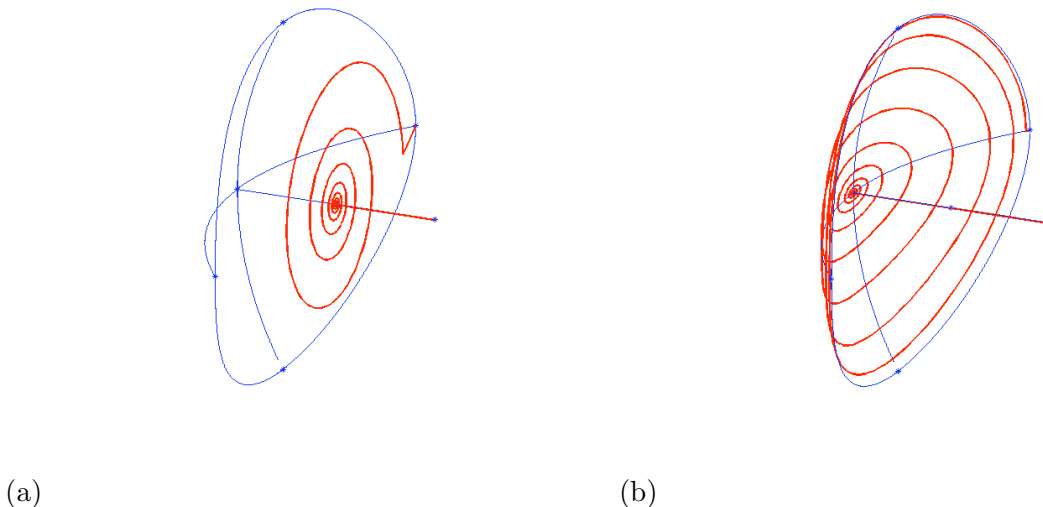


Figure 9: (a) The initial path. (b) The path at some moment of time  $t > 0$ .

As the path evolves in time, the coils of the spiral quickly get into the neighborhood of the "canopy" and then slowly expand toward its edge. This path does not converge to its limit set. However, if we would choose a similar initial path but replace the logarithmic spiral with a spiral with denser coils, the path might fill the "canopy", i.e., converge to its limit set.

## 4 Theorems on convergence

In this section we prove Theorem 3 from Section 2.2. We also state and prove a criterion for convergence to a single MEP, and a sufficient condition for such convergence which explains the success of the string method as a numerical algorithm. The proofs involve some results about "tubes of trajectories" which will be developed in Section 4.1.

We start this section with a technical lemma which simplifies the proofs of convergence to the limit set. This lemma replaces the requirement that the Hausdorff distance tend to zero with a condition that is simpler to check.

**Lemma 1.** *Let  $\phi$  be a set evolving continuously in time and remaining in some bounded region  $D$  at all times. Let  $L$  be its limit set. Suppose for any  $\epsilon > 0$  and every point  $x \in L$  there is a*

moment of time  $t_0 = t_0(\epsilon, x)$  such that for any  $t > t_0$   $B_\epsilon(x) \cap \phi \neq \emptyset$ . Then  $\phi$  converges to its limit set  $L$ .

*Proof.* First we show that for a given  $\epsilon$  one can find a  $t_1(\epsilon)$  that works for every  $x \in L$ , i.e., for any  $t > t_1(\epsilon)$  and every  $x \in L$   $B_\epsilon \cap \phi \neq \emptyset$ . Since  $L$  is compact, there is a finite subset  $\{x_i\}_{i=1}^N \subset L$  such that

$$\bigcup_{i=1}^N B_{\epsilon/2}(x_i)$$

covers  $L$ . If  $t \geq \max_i t_0(\epsilon/2, x_i)$  then each  $B_{\epsilon/2}(x_i)$  meets  $\phi$ . Since  $B_{\epsilon/2}(x_i)$  has diameter  $\epsilon$ , we conclude that when  $t > \max_{1 \leq i \leq N} t_0(\epsilon/2, x_i)$ , each point of  $L$  is at most distance  $\epsilon$  from some point of  $\phi$ . Then  $t_1(\epsilon) = \max_{1 \leq i \leq N} t_0(\epsilon/2, x_i)$  works for every  $x \in L$ .

Next we show that for any  $\epsilon > 0$  there is a  $t_2(\epsilon)$  such that for  $t > t_2(\epsilon)$ , every point of  $\phi$  is within distance  $\epsilon$  of some point of  $L$ . The argument is by contradiction. If the assertion is false then for some  $\epsilon > 0$ , the curve  $\phi$  wanders beyond an  $\epsilon$ -neighborhood of  $L$  at a sequence of times  $t_j \rightarrow \infty$ . Since  $\phi$  stays in a compact set  $D$ , it must have limit points outside the  $\epsilon$ -neighborhood of  $L$ . But this is a contradiction, since  $L$  is the limit set of  $\phi$ .  $\square$

## 4.1 Tubes of trajectories

**Definition 4.** Let  $x_0$  be a regular (non-critical) point. Let  $D(x_0)$  be an  $(n-1)$ -dimensional disc centered at  $x_0$  and normal to the trajectory passing through  $x_0$ . Moreover, assume that  $D(x_0)$  does not contain critical points and is nowhere tangent to the vector field. The set of all of the negative half-trajectories arriving at the disc  $D(x_0)$  is a tube  $T(D(x_0))$ , i.e.,

$$T(D(x_0)) = \{x(t), t \leq 0 \mid x(0) \in D(x_0)\}.$$

**Remark** The disc  $D(x_0)$  is an analog of the *segment without contact* in [Hurewicz, 1958].

**Definition 5.** A negative half-trajectory arriving at  $\partial D$  is a boundary trajectory of the tube. The collection of the boundary trajectories is the boundary of the tube.

**Definition 6.** Let  $x_0$  be a point on a trajectory  $C$ . Let  $\{D_r(x_0)\}$ ,  $0 < r < r_0$  be a nested family of  $(n-1)$ -dimensional discs normal  $C$ . We will call the corresponding family of tubes a nested family of tubes.

**Theorem 5.** Suppose a potential  $V(x)$  and an initial path  $\phi_0(\alpha)$  are such that conditions (1), (2), and (3') from Section 1.2 hold. Suppose the path  $\phi(\alpha, t)$  does not converge to its limit set  $L$ . Then there is a nested family of tubes of negative half-trajectories such that the initial path  $\phi_0$  leaves and enters each tube of the family infinitely many times (Fig.10 (a)).

*Proof.* Let  $L$  be the limit set of  $\phi$ . As we have shown in Section 1.1 the limit sets of the paths evolving according to Eqs. (4) and (5) coincide. Hence we can focus on the evolution according to Eq. (5). It follows from the compactness of sets  $S(\phi_0(\alpha), V(\phi_0(\alpha)))$  that  $L$  is nonempty. By Theorem 1,  $L$  consists of critical points and heteroclinic trajectories.

Since the path does not converge to its limit set  $L$ , there is a point  $x_0 \in L$  for which the condition in Lemma 1 is violated. Therefore, there exists an  $\epsilon > 0$  such that for any moment of time  $t_0 > 0$  there is a moment of time  $t_1 > t_0$  such that  $\phi(\alpha, t_1) \notin D_\epsilon(x_0)$  for any  $\alpha$ . Here  $D_\epsilon(x_0)$  is an  $\epsilon$ -disc centered in  $x_0$  and normal to the trajectory passing through  $x_0$ . Such a disc contains exactly one point of each trajectory passing through it if  $\epsilon$  is sufficiently small (we assume that this is the case). We also assume that this disc is sufficiently small to be a "disc without contact" [Hurewicz, 1958] (i.e., it contains no critical points and nowhere touches the vector field.) Note further, that we can construct an open neighborhood of  $x_0$  by mapping the disc by the flow associated with Eq. (6) forward and backward in time by some small time interval  $dt$ . Hence we have justified that we can think of such discs rather than balls in Definition 1 of the limit set and Lemma 1.

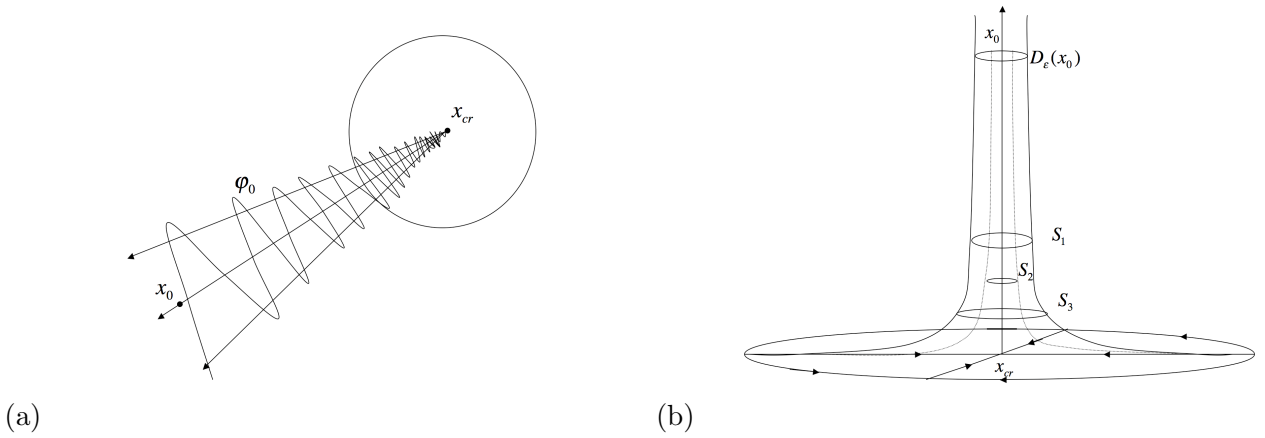


Figure 10: Illustration for Theorem 5. (a) The initial path  $\phi_0$  enters and leaves a tube of trajectories and any tube inside this tube infinitely many times in any neighborhood of a critical point of Morse index 2 or more. (b) Construction of the sets  $S_i$ ,  $i \in \mathbb{N}$ .

It follows from Theorem 1 that the point  $x_0$  belongs to some heteroclinic trajectory  $C$  emanating from a critical point  $x_{cr}$ . Take this  $\epsilon$ -disc and construct a tube around the trajectory  $C$ . Since  $x_0$  is a limit point of  $\phi$ , there are sequences  $\{\alpha_n\}$  and  $\{t'_n\} \rightarrow \infty$  such that  $\phi(\alpha_n, t'_n) \in D_{\epsilon 2^{-n}}(x_0)$ .

Therefore, we can build the following sequence of sets  $\{S_i\}$ . Pick  $t_1$  to be a moment of time at which no point of  $\phi$  lies in  $D_\epsilon(x_0)$ . If no point of  $\phi$  lies at  $D_\epsilon(x_0)$  at time  $t_1$  then no point of  $\phi$  at time 0, i.e., no point of  $\phi_0$  lies in the set  $g^{t_1}D_\epsilon(x_0)$  where  $g$  is the flow associated with solving Eq. (6) backward in time (see Fig. 10(b)). Pick the set  $S_1 = g^{t_1}B_\epsilon(x_0)$ . Then pick  $t_2 > t_1$  from the sequence  $\{t'_n\}$ . There is a point of  $\phi$  lying in  $D_{\epsilon 2^{-n(t_2)}}(x_0)$ . Hence, there is a point of  $\phi_0$  in  $g^{t_2}D_{\epsilon 2^{-n(t_2)}}(x_0)$ . Pick the set  $S_2 = D_{\epsilon 2^{-n(t_2)}}(x_0)$ . Next pick  $t_3 > t_2$  such that no point of  $\phi$  lies in  $D_\epsilon(x_0)$ . Then no point of  $\phi_0$  lies in  $g^{t_3}D_\epsilon(x_0)$ . Pick  $S_3 = g^{t_3}D_\epsilon(x_0)$ . Then pick  $t_4 > t_3$  from the sequence  $\{t'_n\}$ . And so on.

Thus we have built a sequence of disjoint sets  $\{S_i\}_{i=1}^\infty$  such that the path  $\phi_0$  must intersect the sets with even numbers and avoid the sets with the odd ones. By construction, the sets  $S_i$  are sections of the tube of the trajectories passing through  $D_\epsilon(x_0)$ . The initial path  $\phi_0$  must enter and exit the tube between every two sections with odd numbers.

The same construction can be done for any inner tube of the tube in hand. Hence there is a nested family of tubes such that the initial path enters and exits each of the tubes of the family infinitely many times.  $\square$

#### 4.1.1 Proof of Theorem 3

*Proof.* We will prove this theorem by contradiction. Assume that the path  $\phi$  does not converge to its limit set  $L$ . Then there is a point  $x_0 \in L$  for which the condition of Lemma 1 is violated.

Suppose first that  $-\nabla V(x_0) \neq 0$ . Then there exists an  $\epsilon > 0$  such that (i) for any moment of time  $t_0 > 0$  there is a moment of time  $t_1 > t_0$  such that  $\phi(\alpha, t_1) \notin B_\epsilon(x_0)$  for any  $\alpha$ . Since the limit set  $L$  is a curve consisting of heteroclinic trajectories and critical points, there is an  $\epsilon'$ -ball around  $x_0$  such that (ii) if a point  $x_1 \in (L \cap B_{\epsilon'}(x_0))$  then the point  $x_1$  belongs to the trajectory through the point  $x_0$ . Moreover, it follows from Lemma A-2 that (iii) if  $\epsilon'$  is sufficiently small, the trajectory through  $x_0$  enters and exits it not more than once. Hence we can find such  $\epsilon > 0$  that all of the conditions (i), (ii) and (iii) hold.

Let  $D_\epsilon(x_0)$  be the disc which is the section of the ball  $B_\epsilon(x_0)$  passing through  $x_0$  and normal

to the trajectory through  $x_0$ . According to Theorem 5 there exists a nested family of tubes of negative half-trajectories built from the disc  $D_\epsilon(x_0)$  such that the initial path  $\phi_0$  leaves and enters each tube of the family infinitely many times. Fix the disc  $D_{\epsilon/2}(x_0)$ . The initial path intersects the boundary of the tube built from it infinitely many times. Moreover, it follows from the construction in the proof of Theorem 5 that the travel times from these points of intersection to the boundary of the disc  $D_{\epsilon/2}(x_0)$  tend to infinity. Thus, there are sequences  $\{t_n\} \rightarrow \infty$  and  $\{\alpha_n\}$  such that  $\phi(\alpha_n, t_n) \in \partial D_{\epsilon/2}(x_0)$ . Since  $\partial D_{\epsilon/2}(x_0)$  is compact, there is a convergent subsequence  $\{\phi(\alpha_{n_k}, t_{n_k})\} \rightarrow x_1 \in \partial D_{\epsilon/2}(x_0)$ . Thus,  $x_1$  is another limit point of the path  $\phi$  lying in the disc  $D_\epsilon(x_0)$ . But by our construction, the only limit point in  $D_\epsilon(x_0)$  is  $x_0$ . Thus, the assumption that the path  $\phi$  does not approach  $x_0 \in L$  as  $t \rightarrow \infty$  is false. Hence it must approach  $x_0 \in L$  as  $t \rightarrow \infty$  if  $-\nabla V(x_0) \neq 0$ .

Now consider the case where  $x_0 \in L$  is a critical point. Since the limit set is connected there is a heteroclinic trajectory  $C \subset L$  approaching the critical point  $x_0$ . Then for every  $\epsilon_0 > 0$  there is a point  $x \in C$  in the interior of the  $\epsilon_0$ -ball centered at  $x_0$ . Surround  $x$  by a ball  $B_\epsilon(x)$  lying completely in  $B_{\epsilon_0}(x_0)$ . Since  $x$  is a regular point of the potential, the path approaches  $x$  as  $t \rightarrow \infty$  as we have proven above. I.e., there is a moment of time  $t_0$  such that for every  $t > t_0$   $\{\phi(\alpha, t) | \alpha \in [0, 1]\} \cap B_\epsilon(x) \neq \emptyset$ . Hence,  $\{\phi(\alpha, t) | \alpha \in [0, 1]\} \cap B_{\epsilon_0}(x_0) \neq \emptyset$ . Therefore, the path approaches  $x_0 \in L$  as  $t \rightarrow \infty$  in the case where  $-\nabla V(x_0) = 0$  as well.

Hence, the path converges to  $L$  if  $L$  is a curve.  $\square$

#### 4.1.2 Negative limit sets of the boundaries of the tubes

In Theorem 5 we have proven that if a path does not converge to its limit set then the initial path (of finite length) must enter and exit each tube of some nested family of tubes infinitely many times. Where can it do so? To answer this question we introduce the negative limit sets of a tube  $T(D(x_0))$  and of  $\partial T(D(x_0))$  its boundary.

The negative limit sets are defined by Definition 1 applied to the time evolution according to Eq. 6 with reversed time. Lemmas A-4 and A-5 are valid for the negative limit sets of tubes. Thus, *the negative limit set of a tube is closed, connected and consists of critical points and entire trajectories*. As for the limit set of the boundary of a tube, Lemmas A-4 and A-5 are valid, (i.e., it is closed and contains entire trajectories), if the boundary of the disc  $\partial D(x_0)$  is connected, i.e. if the dimension of the space is greater than 2. Otherwise, if  $n = 2$ , the negative limit set of the boundary of the tube is closed and contains whole trajectories.

An initial path of finite length can enter and exit each tube of some nested family of tubes infinitely many times only if the distance between the boundaries of any two tubes of the family is zero. Furthermore, this bound should be achieved at some finite point of the boundaries. Otherwise, the initial path cannot have finite length. The boundaries of any two tubes of a nested family can become arbitrarily close to each other only in the vicinity of their limit sets. Therefore we are interested in the negative limit sets of the boundaries of the tubes.

Let us describe some possible negative limit sets of the boundaries of the tubes. For brevity we will refer to them as limit sets.

We start with the following lemma.

**Lemma 2.** *Suppose a potential  $V(x)$  satisfies Conditions (1), (2), and (3) from Section 1.2. Suppose that all of the boundary trajectories of a tube come from infinity. Then the negative limit set of the boundary of the tube is empty. (Informally speaking, it completely lies at infinity.)*

*Proof.* We will prove this lemma by contradiction. Assume that the negative limit set of the boundary of the tube  $\partial T(D(x_0))$  contains a finite point  $x^*$ . Then there are sequences  $\{t_n\} \rightarrow \infty$  and  $\{x_n\} \subset \partial D(x_0)$  such that  $g^{t_n} x_n \rightarrow x^*$  where  $g^t$  is the flow associated with Eq. (6) with time reversed. Since  $\partial D(x_0)$  is compact, the sequence  $\{x_n\}$  contains a subsequence convergent to a point  $\hat{x} \in \partial D(x_0)$ . We claim that the trajectory through  $\hat{x}$  emanates from some critical point. By the theorem of the continuity of the solutions w.r.t. the initial data for any  $t_n > 0$  and  $\epsilon > 0$  there is a number  $M \in \mathbb{N}$  such that for any  $m > M$   $|g^{t_n} x_m - g^{t_n} \hat{x}| < \epsilon$ . Note that  $M$

might be larger than  $n$ . Then for any  $n$  we have

$$\begin{aligned} |g^{t^m} x_m - g^{t^n} \hat{x}| &= |g^{t^m} x_m - x^* + x^* - g^{t^n} x_m + g^{t^n} x_m - g^{t^n} \hat{x}| \\ &\leq |g^{t^m} x_m - x^*| + |x^*| + |g^{t^n} x_m| + |g^{t^n} x_m - g^{t^n} \hat{x}|. \end{aligned}$$

The term  $|g^{t^m} x_m - x^*|$  is bounded. The term  $|g^{t^n} x_m - g^{t^n} \hat{x}| < \epsilon$ . The term  $|g^{t^n} x_m|$  is bounded because of the following. By Lemma A-1 the potential decreases along the trajectories. Hence it increases along the trajectories if the time is reversed. Hence  $V(g^{t^m} x_m) > V(g^{t^n} x_m)$  as  $m \geq n$ . Since the sublevel sets of the potential are compact,  $g^{t^n} x_m$  lies in the compact set  $S(V(g^{t^m} x_m))$ . Therefore,  $|g^{t^n} x_m|$  is bounded. Thus, since  $|g^{t^m} x_m - g^{t^n} \hat{x}|$  is bounded and  $|g^{t^m} x_m|$  is bounded,  $|g^{t^n} \hat{x}|$  is bounded. Hence it must approach some critical point according to Lemma A-2. But according to the condition of the lemma, all boundary trajectories come from infinity. Thus, our assumption is false. Hence the limit set is empty.  $\square$

Examples of nonempty limit sets of tubes and their boundaries are shown in Fig. 11.

In Fig. 11(a) the limit sets of the tube and its boundary coincide and consist of a single point which must be a maximum.

In Fig. 11(b) the limit set of the tube is the stable manifold of the critical point  $x_1$  from which the trajectory through  $x_0$  emanates. The limit set of the boundary of the tube is the boundary of the stable manifold. All of the critical points belonging to the limit set of the boundary of the tube are the limit points for at least one boundary trajectory. This situation can occur only if the critical point  $x_1$  is of Morse index one.

A similar situation is shown in Fig. 11(c). However, there is a critical point in the limit set of the boundary of the tube which is not the limit point for any of the trajectories of the tube.

In Fig. 11(d) the case where the limit sets of the tube and its boundary go to infinity is shown. Again, the limit set of the tube is the stable manifold, while the limit set of the boundary of the tube is its boundary.

In Fig. 11(e) the limit sets of the tube and its boundary coincide and are the stable manifold of the critical point  $x_1$ . This situation can occur if  $x_1$  has Morse index not less than 2 but less than  $n$ .

In Fig. 11(f) the limit sets of the tube and its boundary coincide and consist of not only the stable manifold of  $x_1$  but also of some parts of stable manifolds of other critical points. These parts can be of higher dimension than the stable manifold of  $x_1$  as it is shown in this figure.

**Example** Let us revisit the example in Section 3.5.3. We consider a nested family of tubes built from a small disc around a point on the heteroclinic trajectory connecting the Morse index 2 saddle  $A$  and the minimum  $M_3$  lying in the plane  $\{z = 0\}$ . The limit sets of these tubes and their boundaries coincide and consist of the ray  $\{(x, y, z) | x = 0, y > 0, z = 0\}$  and the part of the stable manifold of the index 1 saddle at  $(0, 0, 0)$  for which  $x \geq 0$ . These coinciding limit sets are shown in Fig. 12. The initial path exits and enters this tube and every tube inside it in the neighborhood of the origin, belonging to the limit set of the boundaries of these tubes.

## 4.2 When is the limit set of a path a curve?

Now we are well-equipped to state and prove a criterion for when the limit set of a path is a curve.

**Theorem 6.** *Suppose a potential  $V(x)$  and an initial path  $\phi_0(\alpha)$  are such that conditions (1), (2), and (3') from Section 1.2 hold. Then the limit set of the path  $\phi(\alpha, t)$  is not a curve if and only if there is a nested family of tubes such that the initial path  $\phi_0$  leaves and enters each tube of the family infinitely many times.*

*Proof.*  $\Leftarrow$  Suppose there is a nested family of tubes of negative half-trajectories such that the initial path  $\phi_0$  enters and exits each tube of this family infinitely many times. Since the initial path has a finite length, this can happen only if the negative limit sets of the boundaries of each tube of this family coincide. In this case, the initial path does it in



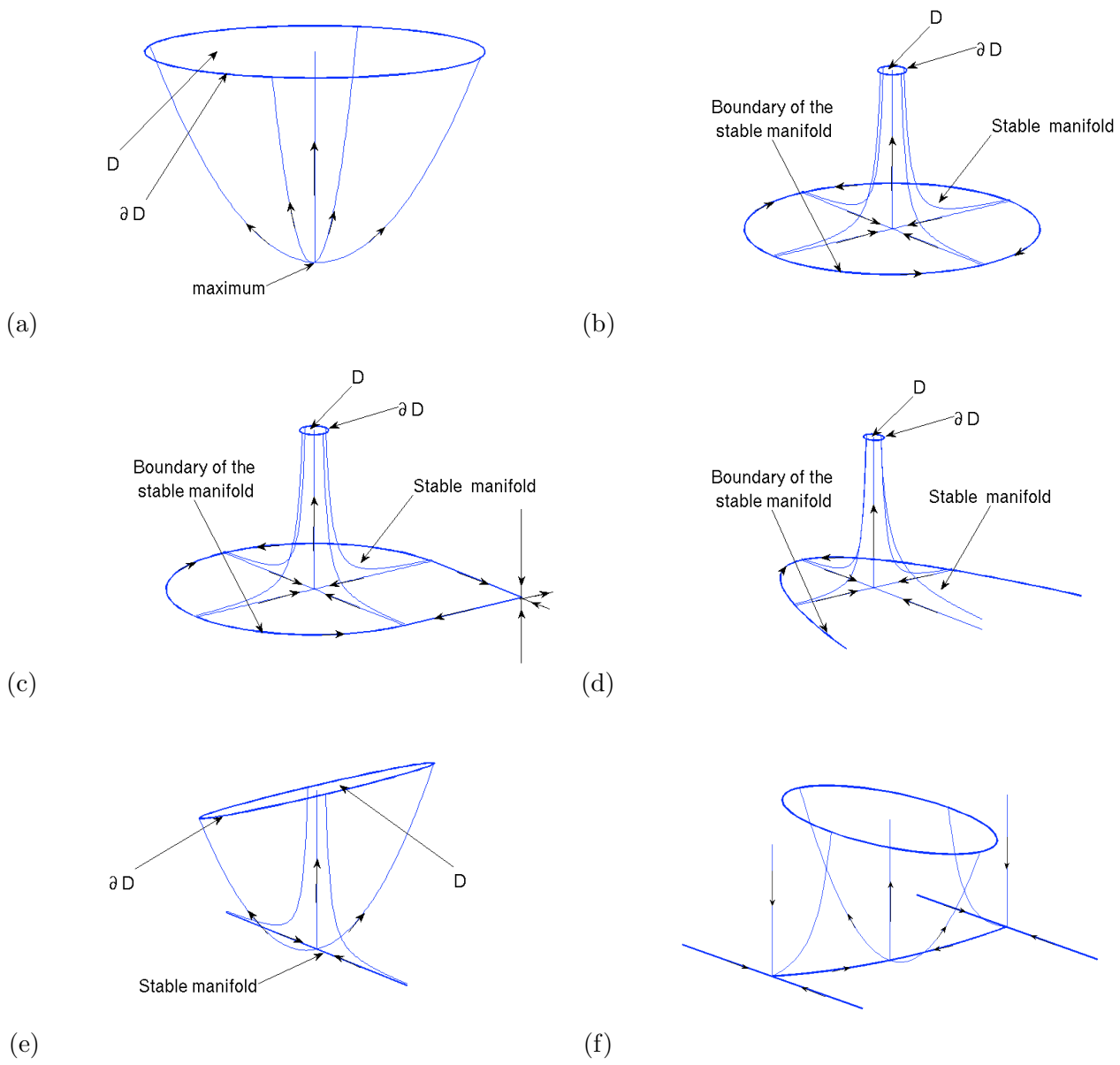


Figure 11: Examples of limit sets of a tube and its boundary.

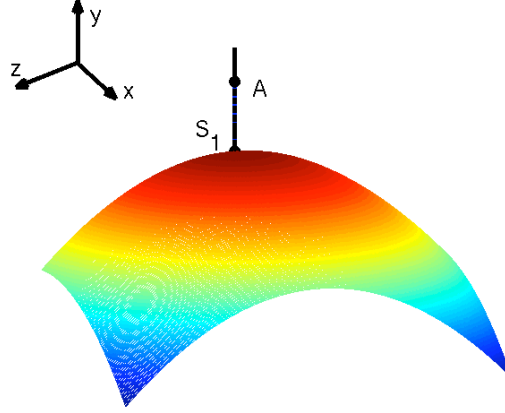


Figure 12: The limit sets of a tube and its boundary built from a small disc around a point on the heteroclinic trajectory connecting the Morse index 2 saddle  $A$  and the minimum  $M_3$  lying in the plane  $\{z = 0\}$  (see Fig. 7).

the neighborhood of the common negative limit set of the boundaries of the tubes of this nested family.

Suppose this nested family corresponds to some non-critical point  $x_0$  and a family of cocentered discs

$$\{D_r(x_0) \mid 0 < r \leq \epsilon\}$$

normal to the trajectory through  $x_0$ . Then for every disc  $D_r(x_0)$  there is a sequence of points of the initial path  $\{\phi_0(\alpha_i)\}$  such that  $\phi(\alpha_i, t_i) \in \partial D_r(x_0)$ . Since  $\partial D_r(x_0)$  is compact, there is a convergent subsequence  $\phi(\alpha_{i_k}, t_{i_k}) \rightarrow x_k \in \partial D_r(x_0)$ . Hence  $x_k$  belongs to the limit set  $L$  of the path  $\phi$ . Since the limit set contains entire trajectories, in every ball around  $x_0$  there are infinitely many trajectories belonging to the limit set  $L$ . Hence  $L$  is not a curve.

$\Rightarrow$  Suppose the limit set  $L$  of  $\phi$  is not a curve. Then there is a non-critical point  $x_0 \in L$  such that every neighborhood of it intersects a trajectory belonging to  $L$ , different from the one through  $x_0$ . Let  $D_\epsilon(x_0)$  be a disc centered at  $x_0$  and normal to the trajectory  $C_0$  through  $x_0$  such that it intersects every trajectory not more than once. Consider a neighborhood of  $x_0$  consisting of the segments of trajectories through this disc

$$U(x_0) = \{x(t) \mid -\Delta t < t < \Delta t, x(0) \in D_\epsilon(x_0)\}.$$

This neighborhood must contain a piece of trajectory  $C_1 \subset L$  different from  $C_0 \subset L$ . Hence there is a point  $x_1 \in D_\epsilon(x_0)$  different from  $x_0$  such that  $x_1 \in L$ . Let  $D_{|x_0-x_1|/3}(x_0) \subset D_\epsilon(x_0)$  and  $D_{|x_0-x_1|/3}(x_1) \subset D_\epsilon(x_0)$  be disjoint discs centered at  $x_0$  and  $x_1$  respectively. Since both  $x_0$  and  $x_1$  are limit points of  $\phi$  there are sequences  $\{\phi(\alpha_i, a_i)\} \subset D_{|x_0-x_1|/3}(x_0)$  and  $\{\phi(\beta_i, b_i)\} \subset D_{|x_0-x_1|/3}(x_1)$  where  $a_i$  and  $b_i$  are moments of time  $a_i \rightarrow \infty$  and  $b_i \rightarrow \infty$ . Thus we can make up a sequence  $\{\phi(\gamma_i, t_i)\}$  out of these two sequences such that for odd  $i$ 's, its elements lie in  $D_{|x_0-x_1|/3}(x_0)$  and for even  $i$ 's, in  $D_{|x_0-x_1|/3}(x_1)$ . Thus, by an argument similar to the one in the proof of Lemma 5, the initial path  $\phi_0$  must enter and exit the tube of negative half-trajectories built from the disc  $D_{|x_0-x_1|/3}(x_0)$  infinitely many times. The same argument can be applied to any disc  $D_\delta(x_0) \subset D_\epsilon(x_0)$  where  $0 < \delta < |x_0 - x_1|/3$ . Hence there is a nested family of negative half-trajectories which the initial path exits and enters infinitely many times if the limit set is not a curve.

□

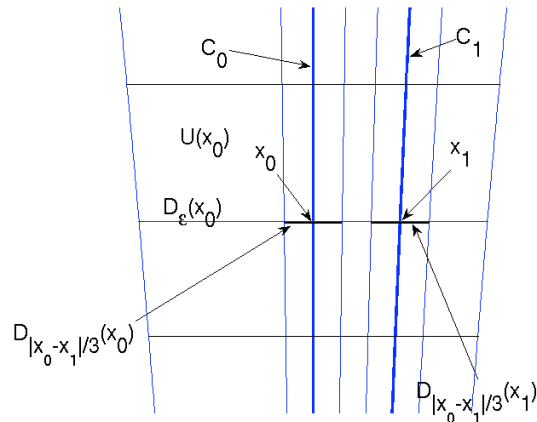


Figure 13: Illustration for Theorem 6. Construction of a neighborhood  $U(x_0)$ . Disc  $D_\epsilon(x_0)$ , trajectories  $C_0$  and  $C_1$ , and tubes of negative half-trajectories built from the discs  $D_{|x_0-x_1|/3}(x_0)$  and  $D_{|x_0-x_1|/3}(x_1)$

This criterion shows why the string method always leads to some path in natural science applications. The reason is stated in the following

**Corollary 7.** *Suppose a potential  $V(x)$  and an initial path  $\phi_0(\alpha)$  are such that conditions (1), (2), and (3') from Section 1.2 hold. Suppose that the potential is piecewise analytic and the initial path is piecewise analytic. Then the limit set of the path is a curve and the path converges to it.*

*Proof.* Since the potential is piecewise analytic, all of the boundaries of the tubes of negative half-trajectories are piecewise analytic hypersurfaces. A piecewise analytic path cannot have more than a finite number of intersections with a piecewise analytic surface. Hence, according to Theorem 6, its limit set is a curve. Then, by Theorem 3, it must converge to it. □

### 4.3 Comments on Examples from Section 3

In this section we show how Theorem 6 applies to the examples from Section 3.

#### 4.3.1 The revolving path in 3D

Consider the revolving path in 3D described in Section 3.2.2. The potential in this example is cylindrically symmetric, with a Morse index 3 maximum at  $(0, 0, -\frac{\pi}{4})$  and a minimum at  $(0, 0, \frac{\pi}{4})$ . There is a continuum of heteroclinic trajectories starting at the maximum and ending at the minimum. The path is constructed on the surface of revolution of one of these trajectories ( $x = r(s) > 0, y = 0, z = z(s)$ ). The limit set of this path is the entire surface of revolution, and the path does not converge to its limit set.

Let us consider a small tube of trajectories built from the disc around the midpoint of the trajectory ( $x = r(s), y = 0, z = z(s)$ ). This tube is horn-shaped with the apex at the maximum. It is easy to see from the construction of the initial path that it exits and enters this tube infinitely many times. The same is true for any smaller tube built around the same trajectory. Hence the initial path enters and exits a nested family of tubes infinitely many times. Therefore,

the fact that its limit set is the entire surface of revolution rather than some curve agrees with Theorem 6.

### 4.3.2 The path which fills a region

Now we revisit the example of the path which fills a 2D region described in Section 3.3. The potential has a Morse index 2 maximum at the origin. There is a continuous family of heteroclinic trajectories starting at the maximum and ending at the minimum at  $(1, 1)$ . The path is built between two selected trajectories as shown in Fig. 14. The limit set of the path is the entire region between these two trajectories. The path converges to its limit set.

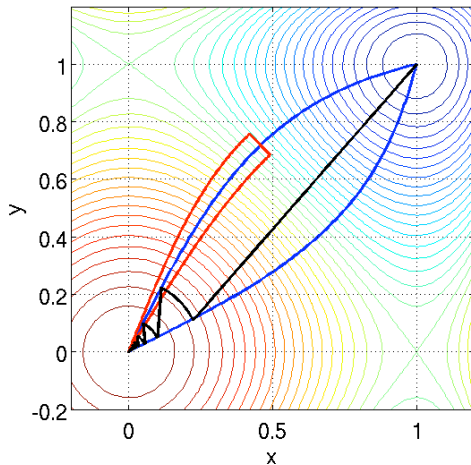


Figure 14: Illustration for Section 14. The initial path enters and exits the tube of trajectories infinitely many times.

Consider a small tube around one of trajectories between which the initial path is built. This tube is shown in Fig. 14. The trajectories of this tube come out of the maximum. By construction, the initial path enters and exits it infinitely many times. It is apparent that the same is true for any tube lying this tube. Hence the initial path enters and exits a nested family of tubes infinitely many times. Therefore, the fact that its limit set is the entire regions bounded by the two trajectories between which we built the initial path rather than some curve agrees with Theorem 6.

### 4.3.3 The path whose limit set is multidimensional

Now consider the example in Section 3.4 of a path whose limit set is  $k$ -dimensional, where  $3 \leq k \leq n$ . The potential has a Morse index  $k$  critical point and a continuous family of heteroclinic trajectories starting at it and ending at a minimum. A subset of these trajectories forms the body  $B$ . The path is constructed in this body so that its limit set is the entire body.

Let us take a small tube around any trajectory of the body  $B$ . We consider the intersection of this tube with the  $\mathbb{R}^k$  dimensional subspace where the body  $B$  is located. All of the negative half-trajectories lying in this tube start at the Morse index  $k$  critical point. It follows from the construction of the initial path that it enters and exits this tube and any tube inside it infinitely many times. Therefore, the fact that its limit set is the entire body  $B$  rather than some curve agrees with Theorem 6.

## 5 Concluding remarks

We have analysed the string method as a dynamical system. The string method provides an evolution law for paths. Thus we were interested in the time evolution and limit sets of entire paths rather than in those for single points, as is usual in dynamical systems. We have extended the definitions of the limit set and of convergence for the case of evolving paths. We have shown that the limit set of a path is closed, connected and may consist only of critical points and heteroclinic trajectories. However it is not necessarily a curve. We have provided a collection of examples demonstrating that it might be multidimensional, and the evolving path might never approach any single MEP.

It follows from our theorems (Corollary 4) that the limit set can be multidimensional only if there is continuum of MEP's connecting two given minima. When the MEP's are isolated, which is true in the absence of Morse index  $\geq 2$  critical points, any initial path must converge to one of them by the string evolution.

The present work neither improves upon nor detracts from the string method as a numerical procedure for finding MEP's. However it does provide insight into the underlying path evolution law. We have explained why the numerical algorithm always succeeds and when and why the exact evolution does not converge to a single MEP. One should not think of the cases where a path does not converge to a single MEP as a failure of an "analytical" string method. These cases involve a sort of physical ambiguity, and the limit sets of such paths may include more physically relevant transition paths.

Finally, it is worth pointing out that the difficulties associated with non-isolated MEP's may also appear in other techniques to identify MEP's, e.g. the ones mentioned in the introduction or the ones in [Olender and Elber, 1997, Vanden-Eijnden and Heymann, 2008] based on minimizing an energy functional for curves. For the latter, the existence of non-isolated MEP's simply means that this energy functional is flat in certain parts of curve space.

## 6 Acknowledgements

We would like to thank Mitch Luskin for interesting discussions when this project was at an early stage, and Lai-Sang Young for valuable discussions as it progressed. We also would like to thank the anonymous reviewers whose comments were helpful and stimulating. This work was partially supported by NSF grants DMS-0807347 and DMS-0708140, and ONR grant N00014-04-1-0565.

## References

- [Barkema and Mousseau, 2001] Barkema G.T., Mousseau N., The activation-relaxation technique: an efficient algorithm for sampling energy landscapes, *Computational Materials Science*, **20**, Number 3, pp. 285-292 (2001)
- [Conley, 1978] Conley, Charles, Isolated invariant sets and the Morse index, *C.B.M.S. Lecture Notes*, **38** (1978)
- [E et al, 2002] E, Weinan, Ren, Weiqing, and Vanden-Eijnden, Eric, String method for study of rare events, *Physical Review B* **66**,052301 (2002)
- [E et al, 2007] E, Weinan, Ren, Weiqing and Vanden-Eijnden, Eric, Simplified and improved string method for computing the minimum energy paths in barrier-crossing events, *The Journal of Chemical Physics* **126**, 164103 (2007)
- [Gardiner, 2004] Gardiner, C. W., *Handbook of Stochastic Methods*, Third Edition, Springer-Verlag Berlin Heidelberg New York, 2004
- [Fischer and Karplus, 1992] Fischer, S. and Karplus, M. Conjugate Peak Refinement : an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters* **194**, 252–261 (1992).

- [Freidlin and Wentzell, 1998] Freidlin, M. I. and Wentzell, A. D., Random Perturbations of Dynamical Systems, 2nd edition, Springer, New York, 1998
- [Henkelman and Jonsson, 1999] Henkelman, G. and Jonsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. The Journal of Chemical Physics **111**, 7010 (1999)
- [Hurewicz, 1958] Hurewicz, Witold, Lectures on Ordinary Differential Equations, The Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, INC., New York, 1958
- [Jonsson et al, 1998] Jonsson, H., Mills, G., and Jacobsen, K.W., Nudged elastic band method for finding minimum energy paths of transitions, Classical and Quantum Dynamics in Condensed Phase Simulations, edited by B. J. Berne, G. Ciccoli, and D. F. Coker, World Scientific, Singapore, p. 385, 1998
- [Miron and Fichthorn, 2001] Miron, R. A. and Fichthorn, K. A. The Step and Slide method for finding saddle points on multidimensional potential surfaces. The Journal of Chemical Physics **115**, 8742 (2001)
- [Olender and Elber, 1997] Olender, R. and Elber, R. Yet another look at the steepest descent path. J. Mol. Struct.: Theochem (Proceedings of the WATOC symposium), **398-399**, 63-71 (1997).
- [Ren, 2002] Ren, W. Numerical Methods for the Study of Energy Landscapes and Rare Events. Ph.D. Thesis, New York University, 2002
- [Robinson, 1999] Robinson, Clark, Dynamical Systems. Stability, Symbolic Dynamics, and Chaos, (Second Edition), CRC Press, 1999
- [Sheppard et al, 2008] Sheppard, D., Terrel, R. and Henkelman, G. Optimization methods for finding minimum energy paths. The Journal of Chemical Physics **128**, 134106 (2008)
- [Ulitsky and Elber, 1990] Ulitsky, A. and Elber, R. A new technique to calculate steepest descent paths in flexible polyatomic systems. J. Chem. Phys. **96**, 1510 (1990).
- [Vanden-Eijnden and Heymann, 2008] Vanden-Eijnden, Eric and Heymann, Matthias, The geometric minimum action method for computing minimum energy paths, The Journal of Chemical Physics **128**, 061103 (2008)
- [Wales, 2002] Wales, D.J., Discrete path sampling, Molecular Physics, **100**, Number 20, pp. 3285-3305 (2002)

## APPENDIX A

### A-1 Limit set of a single trajectory

The results stated in this Section are well-known and relatively easy, see e.g. [Robinson, 1999] (p. 357). Nevertheless, we state them to make this paper self-contained and accessible to the broadest possible audience.

**Lemma A-1.** *The function  $w(t) \equiv V(x(t))$  decreases along each nonstationary trajectory  $\{x(t) | t \geq 0\}$  of Eq. (6).*

*Proof.*  $w_t = \nabla V \cdot (-\nabla V) = -|\nabla V|^2 < 0$ . □

**Lemma A-2.** *Let  $V(x)$  be a potential satisfying conditions (1) and (2) from Section 1.2. Let  $\{x(t) | t \geq 0, x(0) = x_0\}$  be a positive half-trajectory of Eq. (6). Suppose lies in some compact subset of  $\mathbb{R}^n$ . Then  $x(t)$  must approach some critical point of Eq. (6) as  $t \rightarrow \infty$ .*

**Lemma A-3.** *Let  $V(x)$  be a potential satisfying conditions (1) and (2) from Section 1.2. Any trajectory  $\{x(t) | t \in \mathbb{R}\}$  of Eq. (6) completely contained in a compact set is heteroclinic, i.e., it approaches some critical points  $x_1$  as  $t \rightarrow -\infty$  and  $x_2$  as  $t \rightarrow \infty$ , and  $x_1 \neq x_2$ .*

## A-2 Lemmas on the limit set of a subset of $\mathbb{R}^n$ .

The following lemma contains known results, easily deducible e.g. from [Conley, 1978] (p. 29). Despite this we provide a complete proof since (i) it is short, and (ii) we want our paper to be self-contained and accessible to the broadest possible audience.

**Lemma A-4.** (1) *Let  $\phi_0$  be closed and connected subset of  $\mathbb{R}^n$ . Let  $g$  be a continuous flow. Then the limit set of a set  $\phi_0$  is closed and connected.*  
(2) *If  $g$  is the gradient flow forward in time:  $\dot{x} = -\nabla V(x)$  and the sets  $S(x, V(x))$  are compact for any  $x \in \phi_0$ . Then the limit set  $L$  of  $\phi$  is compact.*

*Proof.* (1) Let  $L$  be the limit set of  $\phi_0$ . It follows from the definition of the limit set that  $L$  can be viewed as

$$L = \bigcap_{\tau \geq 0} \overline{g^{[\tau, +\infty)} \phi_0}. \quad (\text{A-1})$$

Hence  $L$  is closed as it is an intersection of closed sets. Furthermore, the closure of a connected set is again connected, hence  $\overline{g^{[\tau, +\infty)} \phi_0}$  is connected. Then  $L$  is connected as an intersection of connected sets is connected.

(2) According to Lemma A-1 the potential decreases along trajectories (Lemma A-1). Hence all of the trajectories remain in a compact set. Therefore all of the sets  $\overline{g^{[\tau, +\infty)} \phi_0}$  are compact. Hence their intersection is also compact.  $\square$

Now we focus on the character of the limit set. The following result is an easy extension of a lemma from [Hurewicz, 1958].

**Lemma A-5.** *Suppose a point  $x_0$  belongs to the limit set of  $\phi$ . The whole trajectory  $\{x(t) | t \in \mathbb{R}\}$  containing the point  $x_0$  must belong to the limit set of  $\phi$ .*

*Proof.* Let  $x_0$  belong to a trajectory  $C = \{x(t) | -\infty < t < \infty, x(t_0) = x_0\}$ . Since  $x_0$  belongs to the limit set of  $\phi$  there are sequences  $\{x_n\}$  and  $\{t_n\} \rightarrow \infty$  such that  $\phi(x_n, t_n) \rightarrow x_0$  as  $n \rightarrow \infty$ . Each point of  $\phi$  moves along some trajectory. Fix some  $\delta > 0$  and consider the sequence  $\{\phi(x_n, t_n + \delta)\}$ . According to the theorem on the continuous dependence of the initial data, this sequence must converge to the point  $x(t_0 + \delta)$  of the trajectory  $C$ . Therefore, the point  $x(t_0 + \delta)$  of  $C$  belongs to the limit set of  $\phi$ . Since we can choose  $\delta$  arbitrarily, any point of the trajectory must be in the limit set of  $\phi$ .  $\square$

**Lemma A-6.** *Suppose the sets  $S(x, V(x))$ ,  $x \in \phi_0$  are compact. Then the limit set of  $\phi$  may consist only of critical points and heteroclinic trajectories.*

*Proof.* By Lemma A-1, the function  $w(t) = V(\phi(x, t))$  is a decreasing function for every  $x \in \phi_0$ . Therefore, every phase trajectory  $\phi(x, t)$  must stay in the set  $S(x, V(x))$ . All these sets are compact by the assumption. And all of them lie in the compact set  $S(x_0, V(x_0))$ , where  $x_0 \in \phi_0$  is such that  $V(x_0) = \max_{x \in \phi_0} V(x)$ . Therefore, all of the limit points must lie in this compact set. Then according to Lemmas A-3 and A-5 the limit set consists of the heteroclinic trajectories and the critical points.  $\square$

## APPENDIX B

Here we provide a detailed proof that the path in Section 3.3 converges to its limit set which is the region between two integral curves.

The phase trajectories for the potential given by Eq. (18) are the solutions of the system

$$\begin{aligned} x_t &= -4x^3 + 4x \\ y_t &= -4y^3 + 4y. \end{aligned}$$

Trajectories emanating from  $(0, 0)$  and going to  $(1, 1)$  are given by

$$\begin{aligned} x(t) &= \frac{1}{\sqrt{Ae^{-8t} + 1}}, \\ y(t) &= \frac{1}{\sqrt{Be^{-8t} + 1}} \end{aligned}$$

where  $A$  and  $B$  are positive constants. The corresponding integral curves are

$$y(x) = \frac{x}{\sqrt{x^2(1 - C^2) + C^2}}, \quad (\text{B-1})$$

where  $C$  is a constant. The derivative of  $y$  w.r.t.  $x$  at  $x = 0$  is  $1/|C|$ . Let the two integral curves used in the construction of the initial path corresponding to  $C = 2$  and  $C = 1/2$ .

Consider the evolution of this path according to Eq. 5. The initial path intersects each integral curve between the two chosen integral curves at the points which can be divided into two series. The first series corresponds to the intersections with the circles given by Eq. (19) and the second series corresponds to the intersections with the segments of straight lines. We will consider only the intersections with the circles, as it is easier and sufficient for our goals.

To prove that the path  $\phi$  converges to its limit set which is the region  $\Omega$  between the two integral curves it suffices to show by Lemma 1 that for any  $(x_0, y_0) \in \Omega$  and any  $\epsilon > 0$  there is  $t^* = t^*(x_0, \epsilon)$  such that for any  $t > t^*$   $B_\epsilon(x_0, y_0) \cap \phi \neq \emptyset$ .

Fix some  $\epsilon > 0$  and pick some point  $(x_0, y_0) \in \Omega$ . The integral curve passing through the point  $(x_0, y_0)$  is given by Eq. (B-1) where

$$C^2 = \frac{x_0^2(1 - y_0^2)}{y_0^2(1 - x_0^2)}. \quad (\text{B-2})$$

It intersects the the circles given by Eq. (19) at the points  $(x_n(0), \sqrt{n^{-4} - x_n(0)})$  where  $x_n(0)$  is found from

$$x_n^2(0) = \frac{-1 - C^2 + 1 - C^2 n^{-4} + \sqrt{(1 + C^2)^2 - 2(1 - C^2)^2 n^{-4} + (1 - C^2)^2 n^{-8}}}{2(1 - C^2)}$$

and  $C^2$  is given by Eq. (B-2). By the Taylor expansion we find that

$$x_n^2(0) = \frac{C^2}{n^4(1 + C^2)} + O(n^{-8}), \quad y_n^2(0) = \frac{1}{n^4(1 + C^2)} + O(n^{-8}) \quad \text{and} \quad (\text{B-3})$$

$$x_n(0) = \frac{1}{n^2\sqrt{\frac{1}{C^2} + 1}} + O(n^{-4}), \quad y_n(0) = \frac{1}{n^2\sqrt{1 + C^2}} + O(n^{-4}) \quad (\text{B-4})$$

The points of the initial path given by Eq. (B-4) evolve according to

$$x_n(t) = \left( \left( \frac{1}{x_n^2(0)} - 1 \right) e^{-8t} + 1 \right)^{-\frac{1}{2}}, \quad y_n(t) = \left( \left( \frac{1}{y_n^2(0)} - 1 \right) e^{-8t} + 1 \right)^{-\frac{1}{2}}. \quad (\text{B-5})$$

Each of them (maybe except the first few) reaches  $(x_0, y_0)$  at the time  $t_n$  such that

$$e^{-8t_n} = \frac{x_0^{-2} - 1}{x_n^{-2}(0) - 1}.$$

Let us show that if  $(x_n(t_n), y_n(t_n)) = (x_0, y_0)$  then  $(x_{n+1}(t_n), y_{n+1}(t_n))$  lies in the  $\epsilon$ -ball around



$(x_0, y_0)$  for sufficiently large  $n$ . Indeed,

$$\begin{aligned}
x_{n+1}(t_n) &= \left( \left( \frac{1}{x_{n+1}^2(0)} - 1 \right) e^{-8t_n} + 1 \right)^{-\frac{1}{2}} \\
&= \left( \left( \frac{1}{x_{n+1}^2(0)} - 1 \right) \frac{x_0^{-2} - 1}{x_n^{-2}(0) - 1} + 1 \right)^{-\frac{1}{2}} \\
&= \left( \frac{x_{n+1}^{-2}(0) - 1}{x_n^{-2}(0) - 1} \left( \frac{1}{x_0^2} - 1 \right) + 1 \right)^{-\frac{1}{2}}. \tag{B-6}
\end{aligned}$$

The sequence in Eq. (B-6) tends to  $x_0$  as  $n \rightarrow \infty$  if and only if

$$\frac{x_{n+1}^{-2}(0) - 1}{x_n^{-2}(0) - 1} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Using Eq. (B-3) we find that

$$\frac{x_{n+1}^{-2}(0) - 1}{x_n^{-2}(0) - 1} = \frac{(n+1)^4 + O(1)}{n^4 + O(1)} = 1 + O(n^{-1}) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Hence,  $x_{n+1}(t_n) \rightarrow x_0$  as  $n \rightarrow \infty$ . Similarly we can prove that  $y_{n+1}(t_n) \rightarrow y_0$  as  $n \rightarrow \infty$ . Therefore, starting from some  $n$ , if the point  $(x_n(t_n), y_n(t_n))$  of the path reaches the point  $(x_0, y_0)$ , then point  $(x_{n+1}(t_n), y_{n+1}(t_n))$  of the path lying on the same integral curve is in the  $\epsilon$ -ball around  $(x_0, y_0)$ . Since the above argument is valid for any point  $(x_0, y_0) \in \Omega$  and any  $\epsilon > 0$ , we have proved that the path converges to its limit set while the limit set is a 2D region.