ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

Computing Mountain Passes

Jorge J. Moré and Todd S. Munson

Mathematics and Computer Science Division

Preprint ANL/MCS-P957-0502

May 2002

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

Contents

1 Introduction **1**

2 Mathematics of Mountain Passes **3**

3 Variational Problems **7**

4 The Elastic String Algorithm **10**

5 Convergence Analysis: Unbounded Level Sets **13**

6 Convergence Analysis: Bounded Level Sets **20**

7 Computational Experiments **22**

8 Transitions States and Reaction Pathways **25**

9 Mountain Passes for Variational Problems **27**

Computing Mountain Passes*

Jorge J. Moré,† and Todd S. Munson†

Abstract

We propose the elastic string algorithm for computing mountain passes in finite-dimensional problems. We analyze the convergence properties and numerical performance of this algorithm for benchmark problems in chemistry and discretizations of infinite-dimensional variational problems. We show that any limit point of the elastic string algorithm is a path that crosses a critical point at which the Hessian matrix is not positive definite.

1 Introduction

A basic version of the mountain-pass theorem shows that if $x_a \in \mathbb{R}^n$ is a strict minimizer of a continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$, and f has no critical points at infinity, then there is a critical point x^* in a path between x_a and any $x_b \in \mathbb{R}^n$ with $f(x_b) \le f(x_a)$ such that the value of f at x^* is

$$\gamma = \inf_{p \in \Gamma} \left\{ \max \left\{ f[p(t)] : t \in [0,1] \right\} \right\}, \tag{1.1}$$

where Γ is the set of all paths that connect x_a with x_b. This result is intuitively clear in $\mathbb{R}^2$ because in this case we cross a mountain pass by following a path for which the maximal elevation is minimal.

The mountain-pass theorem forms the basis for calculating transition states in chemistry. The functions arising in the calculation of mountain passes for these problems are typically potential energy surfaces for a system with x_a and x_b associated with stable states. Analysis of the system requires knowing how the geometry and energy changes as the system transitions between stable states. The mountain pass is of interest because it provides the lowest energy required to transition between stable states.

There is a vast literature on algorithms for computing transition states. Early work includes Bell, Crighton, and Fletcher [4], while recent work is reviewed by Henkelman, Jóhannesson, and Jónsson [12]. We state here only that, as far as we know, the mountain-pass theorem is not mentioned in any work on transition states from the computational chemistry literature.

The mountain-pass theorem is also a fundamental tool in nonlinear analysis, where it is used to prove existence results for variational problems in infinite-dimensional dynamical

*This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38, and by the National Science Foundation (Information Technology Research) grant CCR-0082807.

†Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439 ({more,tmunson}@mcs.anl.gov).

systems. The extension of the mountain-pass theorem to infinite-dimensional spaces, established in 1973 by Ambrosetti and Rabinowitz [2], is widely considered to be a milestone in nonlinear analysis. Algorithms for the computation of mountain passes for infinite-dimensional spaces were not considered until Choi and McKenna [8] proposed an algorithm for functionals of the form

$$\int_{\mathcal{D}} \left(\tfrac{1}{2} \|\nabla u(s)\|^2 - h[s, u(s)] \right) \, ds, \tag{1.2}$$

where $\mathcal{D}$ is an open, bounded set in $\mathbb{R}^m$ for some $m \geq 1$. These functionals are of interest because the critical points are weak solutions to the semilinear partial differential equation $-\Delta u(s) = g(s, u)$, where $g(s,t) = \partial_2 h(s,t)$. Choi and McKenna [8] described their algorithm in an infinite-dimensional setting so that, for example, the computation of the steepest descent direction for the functional requires the solution of a linear partial differential equation. The work of Choi and McKenna has been extended and refined, but in all cases the algorithms have been formulated in an infinite-dimensional setting. Moreover, the algorithms are restricted to functionals of the form (1.2). Interesting numerical results with this infinite-dimensional approach have been obtained by Chen, Zhou, and Ni [7]. We refer to this paper and to the related papers [9, 14, 15] for additional information.

We propose the elastic string algorithm for the computation of mountain passes in finite-dimensional problems. This algorithm is derived from the mountain-pass characterization (1.1) by approximating Γ with the set of piecewise linear paths. We analyze the convergence and numerical performance of this algorithm for benchmark problems in chemistry and variational problems of the form (1.2).

Section 2 presents background for the mountain-pass theorem and related results. The main ingredients in the mountain-pass theorem are the characterization (1.1), the Palais-Smale condition, and the Ekeland variational principle.

Section 3 is dedicated to the variational problem (1.2). We analyze a class of functions that includes finite-dimensional approximations to the variational problem (1.2), and we show that these mappings satisfy the conditions of the mountain-pass theorem.

Section 4 presents the elastic string algorithm. The nudged elastic band algorithm mentioned by Henkelman, Jóhannesson, and Jónsson [12] is related, but this algorithm is based on an intuitive notion of how systems transition between stable stages. In particular, there is no clear relationship between the nudged elastic band algorithm and the characterization (1.1). The elastic string algorithm also differs from the algorithms (for example, [9, 7, 14, 15]) based on the infinite-dimensional approach of Choi and McKenna [8] for functionals of the form (1.2). These algorithms use only steepest descent searches and thus are unlikely to be efficient on general problems. On the other hand, the finite-dimensional approach based on the elastic string algorithm is applicable to any variational problem and can use a wide variety of optimization algorithms.

The analysis of the elastic string algorithm appears in Sections 5 and 6. The results in Section 5 are for functions with unbounded level sets, while Section 6 analyzes the case where the functions have compact level sets. The results are closely related, but the assumptions for functions with compact level sets are stronger. The main results in these two sections show that the iterates in the elastic string algorithm are bounded, and any limit point of the iterates is a critical point at which the Hessian matrix is *not* positive definite.

Sections 7 describes our computational experiments. We discuss the computational environment and note that the aim of these computational results is to show that the elastic string algorithm provides a reasonable approximation to a mountain pass by using a modest number of breakpoints in the elastic string. This is important because the number of variables in the optimization problem in the elastic string algorithm is mn, where m is the number of breakpoints in the piecewise linear path (elastic string).

Sections 8 and 9 present the numerical results for, respectively, transition states and variational problems of the form (1.2). We regard these results as preliminary, and for this reason the problems are of modest size. In particular, for the variational problems (1.2) we use meshes for $\mathcal{D}$ with roughly 400 grid points. On the other hand, we consider two reasonably interesting geometries for $\mathcal{D}$: a circle, and a circle with a smaller square removed from the center of the circle. These geometries give insight into the dependence of a mountain-pass solution on the geometry of $\mathcal{D}$.

2 Mathematics of Mountain Passes

Given a continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and two points x_a and x_b, we seek a critical point on a path between x_a and x_b. A path that connects x_a with x_b is a continuous function $p : [0,1] \mapsto \mathbb{R}^n$ such that $p(0) = x_a$ and $p(1) = x_b$. If we use

$$\Gamma = \{p \in C[0,1] : p(0) = x_a,\ p(1) = x_b\} \tag{2.1}$$

to denote all paths between x_a and x_b, then a set S separates x_a and x_b if every path $p \in \Gamma$ intersects S; that is, there is a $t \in (0,1)$ such that $p(t) \in S$. As an example of a separating set, note that the boundary $\partial B(x_a, r)$ of the ball

$$B(x_a, r) = \{x \in \mathbb{R}^n : \|x - x_a\| \le r\}$$

separates x_a and x_b if $r < \|x_b - x_a\|$. Separating sets may be unbounded. For example, the hyperplane $\{x : \langle x, v\rangle = \alpha\}$ separates x_a and x_b if $\langle x_a, v\rangle < \alpha$ and $\langle x_b, v\rangle > \alpha$.

We require that the value of f on the separating set S be sufficiently high. Specifically, we require that

$$\inf\{f(x) : x \in S\} > \max\{f(x_a), f(x_b)\}. \tag{2.2}$$

The following result provides additional insight into requirement (2.2).

Lemma 2.1 *Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuous on $\mathbb{R}^n$, and define*

$$\gamma = \inf_{p\in\Gamma} \{\max\{f[p(t)] : t \in [0,1]\}\}. \tag{2.3}$$

There is a set S that separates x_a and x_b, and (2.2) *holds if and only if* $\gamma > \max\{f(x_a), f(x_b)\}$. *Moreover, if* $\gamma > \max\{f(x_a), f(x_b)\}$, *then* $\{x \in \mathbb{R}^n : f(x) = \alpha\}$ *separates x_a and x_b for any α with* $\max\{f(x_a), f(x_b)\} < \alpha \le \gamma$.

Proof. Assume that $\gamma > \max\{f(x_a), f(x_b)\}$. We show that

$$f^{-1}(\alpha) = \{x \in \mathbb{R}^n : f(x) = \alpha\}$$

separates x_a and x_b for any α with $\max\{f(x_a), f(x_b)\} < \alpha \le \gamma$. If $p \in \Gamma$, then $f[p(t)] < \alpha$ for $t \in \{0, 1\}$ by definition. We cannot have $f[p(t)] < \alpha$ for all $t \in (0, 1)$ because then

$$\gamma \le \max\{f[p(t)] : t \in [0, 1]\} < \alpha,$$

and this contradicts that $\alpha \le \gamma$. Hence, $f[p(t)] \ge \alpha$ for some $t \in (0, 1)$, and by continuity $f[p(t)] = \alpha$ for some $t \in (0, 1)$. This shows that $f^{-1}(\alpha)$ separates x_a and x_b as desired.

Conversely, assume that S separates x_a and x_b and that (2.2) holds. If $p \in \Gamma$, then $p(t) \in S$ for some $t \in (0, 1)$, and hence

$$\max\{f[p(t)] : t \in [0, 1]\} \ge f[p(t)] \ge \inf\{f(x) : x \in S\}.$$

Since this holds for all $p \in \Gamma$, we obtain that

$$\gamma \ge \inf\{f(x) : x \in S\} > \max\{f(x_a), f(x_b)\}.$$

This is the desired result. ■

The function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ shown in Figure 2.1 illustrates Lemma 2.1 and motivates the mountain-pass theorem. If x_a and x_b are chosen as the two minimizers, then x_a and x_b are separated by a set S that satisfies (2.2). Indeed, the boundary of the ball $\partial B(x_c, r)$, where x_c is the minimizer with highest function value and r is sufficiently small, separates x_a and x_b and satisfies (2.2). It is also clear from Figure 2.2 that f has a critical point on a path that connects x_a with x_b, and that γ is a critical value.

We are concerned with mappings that may have unbounded level sets. Hence, we replace the usual compactness assumption on the level sets with the following assumption.

Definition 2.2 *Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable on the set C. The mapping f satisfies the Palais-Smale condition on C if the existence of a sequence $\{x_k\}$ in C such that*

$$\lim_{k\to\infty} f(x_k) = \alpha, \qquad \lim_{k\to\infty} \nabla f(x_k) = 0,$$

for some α, implies that $\{x_k\}$ has a convergent subsequence.

From an optimization viewpoint, the Palais-Smale condition rules out critical points at infinity, since an optimization algorithm is likely to generate a Palais-Smale sequence if f is bounded below but does not have compact level sets.

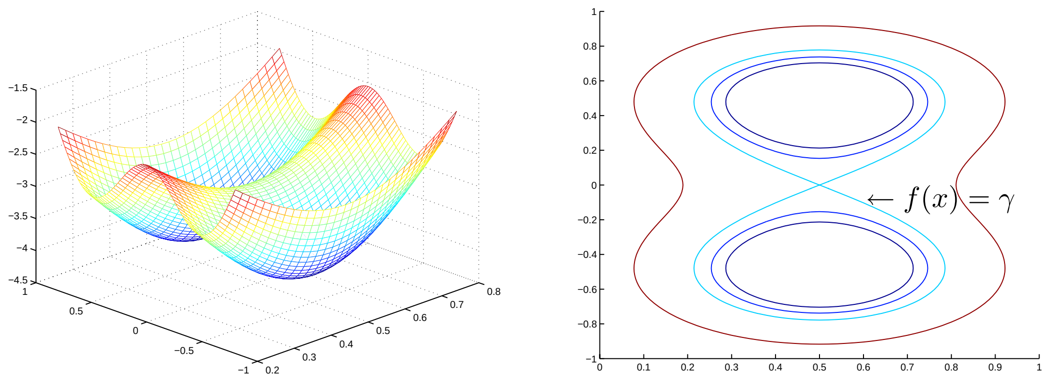

Figure 2.1: A mountain pass with x_a and x_b separated by $\{x \in \mathbb{R}^n : f(x) = \gamma\}$

If f has compact level sets, then f satisfies the Palais-Smale condition, but the converse does not hold. For example, the function $t \mapsto t^2 - t^4$ satisfies the Palais-Smale condition but does not have bounded level sets. Also note that if $|f| + \|\nabla f\|$ is coercive, then f satisfies the Palais-Smale condition. The following result provides additional insight into the role of the Palais-Smale condition.

Theorem 2.3 *Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable on $\mathbb{R}^n$. If f is bounded below and satisfies the Palais-Smale condition on $\mathbb{R}^n$, then f is coercive on $\mathbb{R}^n$ and achieves its minimum.*

Proof. The proof uses the differentiable version of the Ekeland variational principle: If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable and bounded below, and if x_ϵ satisfies

$$f(x_\epsilon) \le \inf\{f(x) : x \in \mathbb{R}^n\} + \epsilon,$$

for some $\epsilon \ge 0$, then for any $\delta > 0$ there is an $y \in \mathbb{R}^n$ such that

$$f(y) \le f(x_\epsilon), \qquad \|\nabla f(y)\| \le \frac{\epsilon}{\delta}, \qquad \|y - x_\epsilon\| \le \delta.$$

We use this result by defining appropriate sequences $\{x_k\}$, $\{\epsilon_k\}$, and $\{\delta_k\}$. Assume, on the contrary, that f is not coercive. Then there is a sequence $\{x_k\}$ such that $\|x_k\| \to \infty$ and $\{f(x_k)\}$ is bounded above. If $\epsilon_k = f(x_k) - \mu$, where $\mu = \inf\{f(x) : x \in \mathbb{R}^n\}$, and $\delta_k = \frac{1}{2}\|x_k\|$ in Ekeland's variational principle, then there is a sequence $\{y_k\}$ with

$$f(y_k) \le f(x_k), \qquad \|\nabla f(y_k)\| \le 2\frac{f(x_k) - \mu}{\|x_k\|}, \qquad \|y_k - x_k\| \le \tfrac{1}{2}\|x_k\|.$$

Thus $\|\nabla f(y_k)\|$ converges to zero, $\{f(y_k)\}$ is bounded, and since

$$\|y_k\| \ge \|x_k\| - \|y_k - x_k\| \ge \tfrac{1}{2}\|x_k\|,$$

the sequence $\{y_k\}$ does not have a convergent subsequence. This is not possible, however, when f satisfies the Palais-Smale condition. ■

Our proof that f is coercive on $\mathbb{R}^n$ in Theorem 2.3 is a variation on the original proof of Caklovic, Li, and Willem [6] but is included here because it uses the differentiable version of Ekeland's variational principle in an interesting fashion. The general version of the Ekeland variational principle applies to lower semicontinuous functions in metric spaces and is a key ingredient in the proof of the mountain-pass theorem. Aubin and Ekeland [3, Section 5.3] discuss several applications of this variational principle.

Theorem 2.4 (Mountain-pass) *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. If f satisfies the Palais-Smale condition on $\mathbb{R}^n$ and γ is defined by* (2.3)*, then γ is a critical value of f.*

The mountain-pass theorem is an essential tool in critical point theory where it is frequently used to prove the existence of nontrivial solutions of nonlinear problems. The proof of Theorem 2.4 can be found, for example, in Mawhin and Willem [16, Chapter 4] and Ekeland [10, Chapter IV]. This version is an improvement over the original version of Ambrosetti and Rabinowitz [2], where the separating set was the boundary of a ball centered on x_a.

The Palais-Smale condition is needed for the mountain-pass theorem. For example, consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by

$$f(\xi_1, \xi_2) = \xi_1^2 + (1 - \xi_1)^3 \xi_2^2,$$

and set $x_a = (0, 0)$ and $x_b = (2, 2)$. Then x_a is a strict minimizer of f and $f(x_a) = f(x_b)$. Thus, $\partial B(x_a, r)$ separates x_a and x_b for r small enough, and (2.2) holds. However, x_a is the only critical point of f, so $\gamma > 0$ is not a critical value of f. In this case the mountain-pass theorem fails to apply because the Palais-Smale condition does not hold.

The mountain-pass theorem guarantees the existence of a critical point but does not provide information on the eigenvalue structure of points in the critical point set, that is,

$$K_\gamma = \{x \in \mathbb{R}^n : f(x) = \gamma, \ \nabla f(x) = 0\},$$

where γ is defined by (2.3). The characterization (2.3) does show that any $x^* \in K_\gamma$ cannot be a strict minimizer, but the following result provides additional information.

Theorem 2.5 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on $\mathbb{R}^n$, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds and that f satisfies the Palais-Smale condition on $\mathbb{R}^n$. If $x^* \in K_\gamma$ and $\nabla^2 f(x^*)$ is nonsingular, then $\nabla^2 f(x^*)$ has precisely one negative eigenvalue.*

This result is due to Ambrosetti [1], but can also be derived from deeper results in Morse theory. See for example, Chapter 8 in Mawhin and Willem [16]. We also note that the number of negative eigenvalues of the Hessian $\nabla^2 f(x^*)$ is the Morse index of the critical point. Thus, Theorem 2.5 can be restated as saying that if x^* is a nondegenerate critical point in K_γ, then the Morse index of x^* is one. In our numerical results we use Theorem 2.5 to verify that the computed critical points are indeed mountain passes.

3 Variational Problems

We are interested in the development of algorithms for finding mountain passes of finite-dimensional problems. In particular, we focus on problems that arise as finite-dimensional approximations to the variational problem

$$\int_{\mathcal{D}} \left(\tfrac{1}{2}\|\nabla u(s)\|^2 - h[s, u(s)] \right) \, ds, \tag{3.1}$$

where $\mathcal{D}$ is an open, bounded set in $\mathbb{R}^m$ for some $m \geq 1$. We seek a function $u : \mathcal{D} \mapsto \mathbb{R}$ that minimizes (3.1) over all suitably smooth functions that satisfy the boundary data $u = u_D$. In most of the applications $u_D \geq 0$ on $\partial\mathcal{D}$, but in the remainder of the paper we assume, without loss of generality, that $u_D \equiv 0$.

Critical points of this variational problem are of interest because, under reasonable conditions on $\mathcal{D}$ and h, the critical points of (3.1) over $H_0^1(\mathcal{D})$ are precisely the weak solutions of the semilinear partial differential equation

$$-\Delta u(s) = g(s, u), \quad s \in \mathcal{D}, \qquad h(s, t) = \int_0^t g(s, \tau) \, d\tau. \tag{3.2}$$

Thus, if (3.1) has a critical point, then the existence of a solution to the differential equation (3.2) is guaranteed. Cases of interest include

$$g(s, t) = \begin{cases} (t)_\circ^p & \text{Lane-Emden } (p > 1) \\ (1/\varepsilon^2)(t^3 - t) & \text{Singularly perturbed Dirichlet} \\ 4\pi(t^2 + 2t)_\circ^{3/2} & \text{Chandrasekaran} \\ \|s\|^l (t)_\circ^p & \text{Henon } (l > 0 \text{ and } p > 1), \end{cases} \tag{3.3}$$

where $(\alpha)_\circ^p = |\alpha|^{p-1}\alpha$. These problems were considered by Chen, Zhou, and Ni [7], and we use them for the numerical results in Section 9.

The use of the mountain-pass theorem to obtain existence results for the semilinear differential equation (3.2) and related problems has been an active research area. Willem [20] discusses these results. For the semilinear problem (3.2) existence of a solution is guaranteed if g and h satisfy regularity assumptions that include growth conditions at $t = 0$ and $t = \infty$. In particular, the existence results require that

$$\limsup_{t \to 0} \frac{g(s, t)}{t} \leq 0, \quad \text{and} \quad 0 < q\, h(s, t) \leq g(s, t)t, \quad |t| \geq t_0, \; s \in \mathcal{D}, \tag{3.4}$$

for constants $q > 2$ and $t_0 > 0$. These conditions hold, in particular, for the equations listed in (3.3).

The variational problem (3.1) usually has multiple solutions even if (3.4) holds. For example, Struwe [18, Theorem 6.2] points out that (3.1) has at least two critical points $u_2 \geq 0 \geq u_1$. Struwe [18, Theorem 6.6] also points out that if h is an even function, then problem (3.1) has an infinite number of critical points. In our numerical results we explore the properties of the mountain-pass critical points of (3.1).

As noted in the introduction, the work of Choi and McKenna [8] led to a series of papers (for example, [9, 7, 14, 15]) on algorithms for the computation of mountain passes for the variational problem (3.1). Since these algorithms work directly with the functional (3.1), the computation of the steepest descent direction requires the solution of a linear partial differential equation. On the other hand, our algorithm is for general finite-dimensional problems; hence, we are not restricted to functionals of the form (3.1).

The following result will be used to show that finite-dimensional approximations to the variational problem (3.1) have mountain passes.

Theorem 3.1 *Assume that $A \in \mathbb{R}^{n\times n}$ is positive definite, and define $f : \mathbb{R}^n \mapsto \mathbb{R}$ by*

$$f(x) = \tfrac{1}{2}\langle x, Ax\rangle - H(x),$$

where $H : \mathbb{R}^n \mapsto \mathbb{R}$ is continuously differentiable with

$$\limsup_{x\to 0} \frac{H(x)-H(0)}{\|x\|^2} \le 0, \qquad \lim_{\|x\|\to\infty} \frac{H(x)-H(0)}{\|x\|^2} = +\infty. \tag{3.5}$$

Then f has a critical point x^* with $f(x^*) > f(0)$. In addition,

$$\lim_{\|x\|\to\infty} \frac{f(x)-f(0)}{\|x\|^2} = -\infty, \tag{3.6}$$

Proof. We prove that the assumptions of the mountain-pass theorem are satisfied with $x_a = 0$, any x_b with $\|x_b\|$ sufficiently large, and $S = \partial B(0, r)$, where $r > 0$ is sufficiently small.

We first show that $f(x) > f(0)$ for $\|x\| = r$ and $r > 0$ sufficiently small. Let $\sigma_1 > 0$ be the smallest eigenvalue of A, and note that

$$\frac{f(x)-f(0)}{\|x\|^2} \ge \sigma_1 - \frac{H(x)-H(0)}{\|x\|^2}.$$

Thus,

$$\liminf_{x\to 0} \frac{f(x)-f(0)}{\|x\|^2} \ge \sigma_1,$$

so that $f(x) > f(0)$ for $x \in \partial B(0, r)$ and r sufficiently small. Similarly,

$$\frac{f(x)-f(0)}{\|x\|^2} \le \sigma_n - \frac{H(x)-H(0)}{\|x\|^2},$$

where $\sigma_n = \|A\|_2$, and thus (3.6) holds. In particular, $f(x_b) < f(0)$ for $\|x_b\|$ sufficiently large. Hence, the assumptions on x_a, x_b, and S of the mountain-pass theorem are satisfied, and thus $\gamma > f(0)$.

We note that the Palais-Smale condition holds, since (3.6) shows that $\{f(x_k)\}$ is unbounded below if $\{\|x_k\|\}$ is unbounded. Hence, the mountain-pass theorem guarantees that there is a critical point x^* with $f(x^*) = \gamma > \max\{f(x_a), f(x_b)\} = f(0)$. ■

The simplest function that satisfies the assumptions of Theorem 3.1 is obtained if A is the identity matrix and H is a multiple of $\|x\|^p$ for $p > 2$. If

$$f(x) = \frac{1}{2}\|x\|^2 - \frac{1}{p}\|x\|^p,$$

then all the assumptions of Theorem 3.1 are satisfied. In this case all x with $\|x\| = 1$ are critical points, and there is another critical point at $x = 0$.

In general, the growth condition (3.5) near $x = 0$ implies that $\nabla H(0) = 0$, but the converse does not hold. Thus, under the assumptions of Theorem 3.1, f has a critical point at the origin and another critical point $x^* \neq 0$.

The growth conditions of Theorem 3.1 are satisfied when H is a finite-dimensional approximation to the variational problem (3.1). In this case, a finite-difference or finite-element approximation to the integral in (3.1) leads to a mapping H of the form

$$H(x) - H(0) = \sum_{i=1}^{n} \alpha_i h(s_i, x_i), \qquad \alpha_i > 0,$$

where α_i are the weights, $s_i \in \mathcal{D}$, x_i are approximations to $u(s_i)$, and $H(0)$ contains boundary data. We assume that the mapping h in (3.1) is continuous in $\mathcal{D} \times \mathbb{R}$ and satisfies

$$\limsup_{t \to 0} \frac{h(s,t)}{t^2} \leq 0, \quad \text{and} \quad \lim_{t \to \infty} \frac{h(s,t)}{t^2} = \infty, \;\; s \in \mathcal{D}. \tag{3.7}$$

If these assumptions hold, then the mapping H satisfies the assumptions of Theorem 3.1. We prove this by first noting that the mapping H has the desired behavior near $x = 0$ because for any index i we have

$$\limsup_{x \to 0} \frac{h(s,x_i)}{\|x\|^2} = \limsup_{x \to 0} \left(\frac{h(s,x_i)}{x_i^2} \right) \left(\frac{|x_i|}{\|x\|} \right)^2 \leq 0,$$

and $\alpha_i > 0$. Proving that H has the desired behavior at infinity requires a more elaborate argument. First we note that there is an $r > 0$ such that $h(s,t) \geq 0$ for $|t| \geq r$ and that by continuity of h there is a $\mu > 0$ such that $h(s,t) \geq -\mu$ for $|t| \leq r$. Hence,

$$\liminf_{\|x\| \to \infty} \frac{h(s_i,x_i)}{\|x\|^2} \geq 0$$

for all indices i. Moreover, there is at least one index i with

$$\limsup_{x \to 0} \frac{h(s,x_i)}{\|x\|^2} = \limsup_{x \to 0} \left(\frac{h(s,x_i)}{x_i^2} \right) \left(\frac{|x_i|}{\|x\|} \right)^2 = +\infty.$$

Since $\alpha_i > 0$ for all indices i, this proves that H has the desired behavior at infinity.

The growth conditions (3.7) on h are weaker than the growth conditions (3.4) required for the infinite-dimensional problem. A short computation shows that condition (3.4) on g

at $t = 0$ implies condition (3.7) on h at $t = 0$. For the conditions at $t = \infty$, integrating the second inequality in (3.4) shows that

$$h(s,t) \geq h(s,t_0)\left(\frac{|t|}{t_0}\right)^q, \qquad |t| \geq t_0.$$

Since $q > 2$, condition (3.7) at infinity clearly holds.

Verifying that the growth conditions (3.7) hold in particular cases is not difficult. Clearly, condition (3.7) at $t = 0$ is satisfied if $g(s,0) = 0$ and $\partial_2 g(s,0) \leq 0$. Moreover, $g(s,0) = 0$ and $\partial_2 g(s,0) \leq 0$ for all the mappings in (3.3).

The super-quadratic growth condition (3.7) at $t = \infty$ can be verified by direct computation. For example, for the Lane-Emden equation $h(s,t) = |t|^{p+1}/(p+1)$, and since $p > 1$, the growth is super-quadratic. In a similar fashion, for the Chandrasekaran function one can show that $h(s,t) \geq \alpha_1 |t|^3 + \alpha_2$ for some constants α_1 and α_2 with $\alpha_1 > 0$.

4 The Elastic String Algorithm

The elastic string algorithm is derived from the mountain-pass characterization (2.3) by restricting the paths connecting x_a with x_b to be piecewise linear with m breakpoints and choosing the breakpoints optimally. Thus, consider the optimization problem

$$\min_{p \in \Gamma_\pi} \left\{ \max \left\{ f[p(t_k)] : 1 \leq k \leq m \right\} \right\}, \tag{4.1}$$

where $\pi = \{t_0, \ldots, t_{m+1}\}$ is a partition of $[0,1]$ with $t_0 = 0$ and $t_{m+1} = 1$;

$$PL[0,1] = \Big\{ p \in C[0,1] : p(0) = x_a,\ p(1) = x_b,\ p \text{ linear on } (t_k, t_{k+1}),\ 0 \leq k \leq m \Big\}$$

is the set of piecewise linear paths that connect x_a with x_b; and

$$\Gamma_\pi = \left\{ p \in PL[0,1] : \int_0^1 \|p'(t)\|\, dt \leq L \right\}$$

is the set of piecewise linear paths that connect x_a with x_b with length bounded by L. If $p \in \Gamma_\pi$ is parameterized by the breakpoints $x_k = p(t_k)$ for $1 \leq k \leq m$, then

$$p(t) = \frac{t - t_k}{t_{k+1} - t_k} x_{k+1} + \frac{t_{k+1} - t}{t_{k+1} - t_k} x_k, \qquad t \in [t_k, t_{k+1}].$$

In particular, p' satisfies

$$\|p'(t)\| = \frac{\|x_{k+1} - x_k\|}{t_{k+1} - t_k}, \qquad t \in (t_k, t_{k+1}),$$

and thus p has length bounded by L if and only if

$$\|x_{k+1} - x_k\| \leq h_k, \qquad 0 \leq k \leq m,$$

for some bounds $h_k > 0$, where

$$\sum_{k=0}^{m} h_k = L. \tag{4.2}$$

This shows that the optimization problem (4.1) can be solved by generating the breakpoints $x_1, \ldots, x_m$ in the piecewise linear path via the optimization problem

$$\min \left\{ \nu(x) : \|x_{k+1} - x_k\| \le h_k,\ 0 \le k \le m \right\}, \tag{4.3}$$

where

$$\nu(x) = \max \left\{ f(x_1), \ldots, f(x_m) \right\}. \tag{4.4}$$

This is an optimization problem over vectors $x \in \mathbb{R}^{mn}$, where x has the (vector) components $x_1, \ldots, x_m$, and each component $x_i \in \mathbb{R}^n$.

We analyze the elastic string algorithm as m increases, but we usually suppress the dependence on m unless this is required for clarity. For emphasis, the term "level" is used to refer to m. We assume that the bound L satisfies $\|x_b - x_a\| \le L$ so that the compact, convex set

$$\Omega = \{ x \in \mathbb{R}^n : \|x - x_a\| \le L,\ \|x - x_b\| \le L \} = B(x_a, L) \cap B(x_b, L)$$

is nonempty. Further, the length L_m of the path is allowed to vary with m because we do not know a priori the length of a path passing through a mountain pass. We assume that

$$\lim_{m \to \infty} \frac{L_m}{m} = 0.$$

Although the bound L_m is allowed to increase with m, we show that under suitable conditions, L_m is increased only a finite number of times.

We assume that the bounds $\{h_k\}$ are chosen to be quasi-uniform in the sense that

$$\max\{h_k : 0 \le k \le m\} \le \kappa \min\{h_k : 0 \le k \le m\}, \tag{4.5}$$

for some constant $\kappa \ge 1$. The quasi-uniform condition (4.5) shows that

$$\frac{1}{\kappa} \left(\frac{L}{m+1} \right) \le \min\{h_k : 0 \le k \le m\} \le \max\{h_k : 0 \le k \le m\} \le \kappa \left(\frac{L}{m+1} \right).$$

Under this assumption, the mesh spacings $\{h_k\}$ converge to zero as m increases, as can be seen from the assumptions made on L_m and the inequality above that relates the maximal and minimal mesh spacing $\{h_k\}$.

If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuous on some open set that contains Ω, then the optimization problem (4.3) has a continuous objective function ν and a compact feasible set. Indeed, all the feasible points lie in the compact set

$$\{ x \in \mathbb{R}^{mn} : x_k \in \Omega \}$$

in $\mathbb{R}^{mn}$. Hence, (4.3) has a minimizer. In the elastic string algorithm the breakpoints $x_1, \ldots, x_m$ can be any critical point of the optimization problem (4.3).

The analysis of the elastic string algorithm is based on an equivalent formulation of the optimization problem (4.3). We consider the optimization problem

$$\min \{ \nu : f(x_k) \le \nu, \ 1 \le k \le m, \ \ c_k(x) \le 0, \ 0 \le k \le m \}, \tag{4.6}$$

where $c_k : \mathbb{R}^{mn} \mapsto \mathbb{R}$ are the constraints

$$c_k(x) = \tfrac{1}{2} \left(\|x_{k+1} - x_k\|^2 - h_k^2 \right).$$

The following result shows that the constraints of this optimization problem satisfy a constraint qualification.

Lemma 4.1 *If $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuously differentiable on some open set that contains Ω, then the active constraint gradients of problem* (4.6) *are positively linearly independent.*

Proof. If we define the constraints $d_k : \mathbb{R}^{mn} \mapsto \mathbb{R}$ by $d_k(x) = f(x_k) - \nu$ for $1 \le k \le m$, then we need to show that if

$$\sum_{k \in \mathcal{A}} \alpha_k \begin{pmatrix} -1 \\ \nabla d_k(x) \end{pmatrix} + \sum_{k \in \mathcal{B}} \beta_k \begin{pmatrix} 0 \\ \nabla c_k(x) \end{pmatrix} = 0, \qquad \alpha_k \ge 0, \quad \beta_k \ge 0,$$

where $\mathcal{A}$ is the set of active d_k constraints and $\mathcal{B}$ is the set of active c_k constraints, then $\alpha_k \equiv 0$ and $\beta_k \equiv 0$, Clearly, this condition implies that

$$\sum_{k \in \mathcal{A}} \alpha_k = 0,$$

and since $\alpha_k \ge 0$, we must have $\alpha_k \equiv 0$ for any choice of $\mathcal{A}$. We now prove that the active constraint gradients $\nabla c_k(x)$ are linearly independent, that is, if

$$\sum_{k \in \mathcal{B}} \beta_k \nabla c_k(x) = 0,$$

then $\beta_k \equiv 0$. We proceed by direct computation. We define $s_k = x_{k+1} - x_k$ and note that the representation $\nabla c_k(x) = (c_{i,k})$ in terms of components $c_{i,k} \in \mathbb{R}^n$ is $c_{i,k} = 0$ for $i < k$ or $i > k+1$, $c_{k,k} = -s_k$ and $c_{k+1,k} = s_k$. Thus, if l is the first index such that $\beta_l \ne 0$, then the above condition implies that $\beta_l s_l = 0$. Since $\|s_l\| = h_l > 0$ for $l \in \mathcal{B}$, we have $\beta_l = 0$. This contradiction shows that the active constraint gradients $\nabla c_k(x)$ are linearly independent. ■

Lemma 4.1 shows that the Kuhn-Tucker conditions for the optimization problem (4.6) hold with multipliers $\lambda_k \ge 0$ and $\mu_k \ge 0$. Thus, if we define the constraints $d_k : \mathbb{R}^{mn} \mapsto \mathbb{R}$

by $d_k(x) = f(x_k) - \nu$ for $1 \le k \le m$, then the first-order Kuhn-Tucker conditions are the $mn+1$ equations,

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \sum_{k=1}^{m} \lambda_k \begin{pmatrix} -1 \\ \nabla d_k(x) \end{pmatrix} + \sum_{k=0}^{m} \mu_k \begin{pmatrix} 0 \\ \nabla c_k(x) \end{pmatrix} = 0;$$

the restrictions $\lambda_k \ge 0$ and $\mu_k \ge 0$ on the sign of the multipliers; and the complementarity conditions

$$\lambda_k(f(x_k) - \nu) = 0, \quad 1 \le k \le m, \qquad \mu_k(\|s_k\| - h_k) = 0, \quad 0 \le k \le m, \tag{4.7}$$

where $s_k = x_{k+1} - x_k$. The first equation in the Kuhn-Tucker conditions implies that the multipliers λ_k satisfy the condition

$$\sum_{k=1}^{m} \lambda_k = 1, \tag{4.8}$$

while the remaining mn equations can be written in the form

$$\lambda_k \nabla f(x_k) = \mu_k s_k - \mu_{k-1} s_{k-1}, \qquad 1 \le k \le m. \tag{4.9}$$

These equations follow from the Kuhn-Tucker conditions by noting that the representation of $\nabla d_k(x) = (d_{i,k})$ in terms of components $d_{i,k} \in \mathbb{R}^n$ is $d_{k,k} = \nabla f(x_k)$ and $d_{i,k} = 0$ for $i \ne k$, while as noted in the proof of Lemma 4.1, the representation of $\nabla c_k(x) = (c_{i,k})$ in terms of components $c_{i,k} \in \mathbb{R}^n$ is $c_{i,k} = 0$ for $i < k$ or $i > k+1$, $c_{k,k} = -s_k$ and $c_{k+1,k} = s_k$.

In the next two sections we use the Kuhn-Tucker conditions (4.7–4.9) to analyze the convergence of the elastic string algorithm.

5 Convergence Analysis: Unbounded Level Sets

In our analysis of the elastic string algorithm we assume that we are given a continuously differentiable mapping $f : \mathbb{R}^n \mapsto \mathbb{R}$, two points x_a and x_b, and a closed set S that separates x_a and x_b with (2.2). Our aim is to provide a constructive proof of the mountain-pass theorem.

We label the indices where the multipliers $\lambda_k > 0$ by $k_1, \ldots, k_l$, where $k_i \le k_{i+1}$. The complementarity condition (4.7) implies that $f(x_k)$ is maximal for $k_1, \ldots, k_l$, that is,

$$f(x_k) = \nu = \max\{f(x_j) : 1 \le j \le m\}, \qquad k = k_1, \ldots, k_l. \tag{5.1}$$

As a consequence of this definition, $\lambda_k = 0$ for $k_i < k < k_{i+1}$. For the results below, we set $k_0 = 0$ and $k_{l+1} = m+1$.

Lemma 5.1 *Define $k_1, \ldots, k_l$ by requiring that $\lambda_k > 0$ if and only if $k = k_1, \ldots, k_l$. Then*

$$\frac{1}{\kappa}\mu_j \le \mu_{k_i} \le \kappa \mu_j, \qquad k_i \le j < k_{i+1}, \qquad 0 \le i \le l.$$

Proof. Since $\lambda_j = 0$ for $k_i < j < k_{i+1}$, the Kuhn-Tucker condition (4.9) shows that

$$\mu_j s_j = \mu_{j-1} s_{j-1}, \qquad k_i < j < k_{i+1}.$$

Hence, the complementarity condition (4.7) implies that

$$\mu_j h_j = \mu_{j-1} h_{j-1}, \qquad k_i < j < k_{i+1}.$$

In particular, $\mu_{k_i} h_{k_i} = \mu_j h_j$ for $k_i \le j < k_{i+1}$. The result now follows from the quasi-uniform condition (4.5) on $\{h_k\}$. ■

Lemma 5.1 shows, in particular, that if $\mu_k > 0$ for $k = k_0, \dots, k_l$, then all the multipliers μ_k are positive. We now analyze this situation and show that in this case the path p has at most l kinks, that is, breakpoints t_k where the derivative p' is not continuous. Since

$$\lim_{t \to t_k^+} p'(t) = \frac{x_{k+1} - x_k}{h_k} = \frac{s_k}{h_k},$$

the path p has a kink at t_k if and only if $h_{k-1} s_k \ne h_k s_{k-1}$.

Lemma 5.2 *If $\mu_{k_i} > 0$ for all $0 \le i \le l$, then the piecewise linear path p that connects x_a with x_b via $x_1, \dots, x_m$ has at most l kinks occurring at t_{k_i} for $1 \le i \le l$. Moreover,*

$$\sum_{i=0}^{l} \|x_{k_{i+1}} - x_{k_i}\| = L,$$

and thus the path p has length L.

Proof. Lemma 5.1 implies that $\mu_j > 0$ for $k_i < j < k_{i+1}$. Since $\lambda_j = 0$ for $k_i < j < k_{i+1}$, the Kuhn-Tucker condition (4.9) shows that

$$\mu_j s_j = \mu_{j-1} s_{j-1}, \qquad k_i < j < k_{i+1}.$$

In particular, $\mu_{k_i} s_{k_i} = \mu_j s_j$ for $k_i \le j < k_{i+1}$. Since all the multipliers μ_j are positive, the points x_j with $k_i \le j \le k_{i+1}$ are collinear, and thus the piecewise linear path p that connects x_a with x_b via $x_1, \dots, x_m$ has at most l kinks. In addition, since

$$\|x_{k_{i+1}} - x_{k_i}\| = \sum_{j=k_i}^{k_{i+1}-1} h_j,$$

the path p has length L. ■

The case where the assumptions of Lemma 5.2 hold should be considered rare, since in general we expect some of the multipliers μ_k to be zero. If the elastic string algorithm computes a path of length L, this usually indicates that the bound L is too small and that it should be increased when m is increased. Therefore, we set

$$L = \min\left\{ \eta L, m^{1/2} \|x_b - x_a\| \right\}, \qquad \eta > 1,$$

whenever $\mu_k > 0$ for all k. This updating rule guarantees that if $L_m \to \infty$, then $\{L_m/m\}$ converges to zero. Thus, as noted earlier, the mesh spacings $\{h_k\}$ converge to zero.

In the remainder of this section, we analyze the elastic string as the level changes over a sequence $\mathcal{M}$. We assume that the number l of indices where $\lambda_k > 0$ is bounded, independent of the level m. This is a technical assumption that is likely to be satisfied by most problems but can rule out problems where the mountain pass is degenerate. Since $f(x_k)$ approximates $f[p(t_k)]$, where p is a piecewise linear (continuous) function, we expect to have $l \le 2$ in most cases, and this is certainly what our numerical results show.

We show next that L is updated only a finite number of times. This requires a growth assumption on f. We assume that there is an $r > 0$ such that

$$f(x) \le \max\{f(x_a), f(x_b)\}, \qquad \|x - x_a\| \ge r. \tag{5.2}$$

This assumption clearly holds for functions that satisfy the assumptions of Theorem 3.1, since in this case $f(x) \to -\infty$ as $\|x\| \to \infty$.

Although assumption (5.2) is natural for functions considered in Section 2, this assumption rules out coercive functions. As we shall see, this case requires different assumptions, and thus we delay consideration of this case.

An important consequence of (5.2) is that the separating set S is compact. Indeed, (2.2) shows that (5.2) cannot hold for $x \in S$. Hence, $\|x - x_a\| < r$ for $x \in S$, and thus S is compact. We use the compactness of S in the next result.

Lemma 5.3 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on $\mathbb{R}^n$, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level. If* (5.2) *holds, then there is no sequence of levels $\mathcal{M}$ such that $\mu_{k_i} > 0$ for all i with $0 \le i \le l$.*

Proof. Assume, on the contrary, that there is a sequence of levels $\mathcal{M}$ such that $\mu_{k_i} > 0$ for all i with $0 \le i \le l$. Hence, the updating rules for L show that $L \to \infty$ as the level m increases. Since Lemma 5.2 guarantees that

$$\sum_{i=0}^{l} \|x_{k_{i+1}} - x_{k_i}\| = L,$$

and since $L \to \infty$, we obtain that for m large enough,

$$\max\left\{\|x_{k_i} - x_a\| : 1 \le i \le l\right\} \ge \frac{L}{(2l+1)} \ge r,$$

and thus assumption (5.2) on the behavior of f implies that

$$f(x_{k_i}) \le \max\{f(x_a), f(x_b)\},$$

for some index i. Since $f(x_{k_i}) \ge f(x_k)$, assumption (2.2) on S shows that we actually have

$$f(x_k) \le \max\{f(x_a), f(x_b)\} < \inf\{f(x) : x \in S\}, \qquad 1 \le k \le m.$$

This inequality is not possible because, as we now show, there is an index k such that some subsequence of $\{x_k\}$ converges to an element $x^* \in S$. We establish this claim by first noting that since all paths from x_a to x_b intersect S, there is a $\xi \in (0,1)$ such that $p(\xi) \in S$. Moreover, if t_k is the closest mesh point to ξ, then

$$\|x_k - p(\xi)\| = \|p(t_k) - p(\xi)\| \le h_k.$$

Since S is compact, $p(\xi) \in S$, and $\{h_k\}$ converges to zero, some subsequence of $\{x_k\}$ converges to an element $x^* \in S$. This establishes the claim and completes the proof. ■

Lemma 5.3 shows that for m large enough, L is eventually constant. Thus, Ω_m does not change for m large enough, and this guarantees that all the iterates remain in a compact set. In particular, all the iterates are bounded. Lemma 5.3 also guarantees that there is a sequence of levels $\mathcal{M}$ and an index i with $0 \le i \le l$ such that $\mu_{k_i} = 0$ for all levels in $\mathcal{M}$. This key result is the basis for the convergence results.

The behavior of the multipliers μ_{k_i} is another essential ingredient in the convergence proof. The next result is used to prove that all sequences $\{\mu_{k_i}\}$ are bounded.

Lemma 5.4 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on some open set that contains Ω, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level. If* (5.2) *holds, then $\{\mu_{k_i}\}$ is bounded on some sequence $\mathcal{M}$ of levels if and only if $\{\mu_{k_i-1}\}$ is bounded on $\mathcal{M}$.*

Proof. We first note that Lemma 5.3 shows that Ω is compact, since L is updated only a finite number of times. This fact is used in the remainder of the proof.

We rule out the case where $k_1 = 1$ or $k_l = m$ with an argument drawn from the proof of Lemma 5.3. Let p be the path that connects x_a with x_b with breakpoints $x_1, \dots, x_m$. Since all paths from x_a to x_b intersect S, there is an $\xi \in (0,1)$ such that $p(\xi) \in S$. Hence, using (2.2), we have

$$f[p(\xi)] \ge \inf\{f(x) : x \in S\} \ge \gamma > \max\{f(x_a), f(x_b)\}.$$

If t_k is the closest mesh point to ξ, then $\|x_k - p(\xi)\| \le h_k$. Hence, $\{x_k\}$ and $\{p(\xi)\}$ have a common subsequence that converges to an element $x^* \in S$ as $m \to \infty$ because $\{x_k\}$ lies in a compact set, $p(\xi) \in S$, and $\{h_k\}$ converges to zero. Thus, the above inequality shows that for any α such that $\max\{f(x_a), f(x_b)\} < \alpha < \gamma$,

$$f(x_k) \ge \alpha > \max\{f(x_a), f(x_b)\},$$

if the level m is sufficiently large. Since $f(x_{k_i}) \ge f(x_k)$ for $1 \le i \le l$, this proves, in particular, that x_{k_i} cannot converge to either x_a or x_b. Thus, $k_1 > 1$ and $k_l < m$ for all levels large enough.

We now show that the multipliers $\{\mu_{k_i}\}$ is bounded if $\{\mu_{k_i-1}\}$ is bounded. We bound μ_{k_i} by noting that

$$f(x_{k_i+1}) = f(x_{k_i}) + \langle \nabla f(x_{k_i}), s_{k_i} \rangle + \tfrac{1}{2}\langle s_{k_i}, A_{k_i} s_{k_i} \rangle,$$

where $A_{k_i} = \nabla^2 f(x_{k_i} + \xi_{k_i} s_{k_i})$ for some $\xi_{k_i} \in (0,1)$. The sequence $\{\|A_{k_i}\|\}$ is uniformly bounded because $\nabla^2 f$ is bounded on the compact, convex set Ω. Thus, since $k_i \le k_l < m$, we must have $f(x_{k_i+1}) \le f(x_{k_i})$ and

$$\langle \nabla f(x_{k_i}), s_{k_i} \rangle \le -\tfrac{1}{2} \langle s_{k_i}, A_{k_i} s_{k_i} \rangle \le \tfrac{1}{2} \sigma h_{k_i}^2,$$

where σ is an upper bound on $\|A_{k_i}\|$. This bound and (4.9) imply that

$$\mu_{k_i} \|s_{k_i}\|^2 = \mu_{k_i-1} \langle s_{k_i-1}, s_{k_i} \rangle + \lambda_{k_i} \langle \nabla f(x_{k_i}), s_{k_i} \rangle \le \mu_{k_i-1} h_{k_i-1} h_{k_i} + \tfrac{1}{2} \sigma h_{k_i}^2.$$

Hence, the complementarity condition (4.7) and the quasi-uniform condition (4.5) on $\{h_{k_i}\}$ imply $\mu_{k_i} \le \kappa \mu_{k_i-1} + \frac{1}{2}\sigma$. This shows that that $\{\mu_{k_i}\}$ is bounded if $\{\mu_{k_i-1}\}$ is bounded.

The proof of the converse is similar, with only minor changes in the argument. We assume that $\{\mu_{k_i}\}$ is bounded and show that $\{\mu_{k_i-1}\}$ is bounded. In this case we bound μ_{k_i-1} by noting that

$$f(x_{k_i-1}) = f(x_{k_i}) - \langle \nabla f(x_{k_i}), s_{k_i-1} \rangle + \tfrac{1}{2} \langle s_{k_i-1}, A_{k_i} s_{k_i-1} \rangle,$$

where $A_{k_i} = \nabla^2 f(x_{k_i} - \xi_{k_i} s_{k_i-1})$ for some $\xi_{k_i} \in (0,1)$. Since $k_i \ge k_1 > 1$, we can assert that $f(x_{k_i-1}) \le f(x_{k_i})$, and hence

$$\langle \nabla f(x_{k_i}), s_{k_i-1} \rangle \ge \tfrac{1}{2} \langle s_{k_i-1}, A_{k_i} s_{k_i-1} \rangle \ge -\tfrac{1}{2} \sigma h_{k_i-1}^2,$$

where σ is an upper bound on $\|A_{k_i}\|$. Thus, this bound and (4.9) imply that

$$\mu_{k_i-1} \|s_{k_i-1}\|^2 = \mu_{k_i} \langle s_{k_i-1}, s_{k_i} \rangle - \lambda_{k_i} \langle \nabla f(x_{k_i}), s_{k_i-1} \rangle \le \mu_{k_i} h_{k_i-1} h_{k_i} + \tfrac{1}{2} \sigma h_{k_i-1}^2.$$

Hence, the complementarity condition (4.7) and the quasi-uniform condition (4.5) on $\{h_{k_i}\}$ imply that $\mu_{k_i-1} \le \kappa \mu_{k_i} + \frac{1}{2}\sigma$, as desired. ■

We now use Lemma 5.4 to prove that all sequences $\{\mu_{k_i}\}$ are bounded, and thus $\|\nabla f(x_{k_i})\|$ converges to zero for some index i as $m \to \infty$.

Theorem 5.5 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on some open set that contains Ω, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level. If the growth condition* (5.2) *holds, then*

$$\lim_{m \to \infty} \min \{ \|\nabla f(x_{k_i})\| : 1 \le i \le l \} = 0. \tag{5.3}$$

Proof. We need to show that for any sequence there is a finer subsequence $\mathcal{M}$ such that $\|\nabla f(x_{k_i})\|$ converges to zero for some index i.

Lemma 5.3 shows that for any sequence there is a finer subsequence $\mathcal{M}$ and an index i with $0 \le i \le l$ such that $\mu_{k_i} = 0$ for all levels in $\mathcal{M}$. Actually, to reach the desired conclusion we need to know only that there is a sequence of levels $\mathcal{M}$ and an index i with $0 \le i \le l$

such that $\{\mu_{k_i}\}$ is bounded. We claim that all $\{\mu_{k_i}\}$ are bounded. Lemma 5.1 implies that $\{\mu_k\}$ is bounded for all $k_i \le k < k_{i+1}$. We can now use Lemma 5.4 to guarantee that $\{\mu_k\}$ is bounded for all $k_i - 1 \le k \le k_{i+1}$. Moreover, now that we know that $\{\mu_k\}$ is bounded for $k = k_i - 1$, we can conclude from Lemma 5.1 that $\{\mu_k\}$ is bounded for $k = k_{i-1}$. Hence, $\{\mu_k\}$ is bounded for $k = k_{i-1}$ and $k = k_{i+1}$. Since there are only l subsequences $k_i, \ldots, k_l$, this argument can be repeated to show that $\{\mu_{k_i}\}$ are bounded for all $0 \le i \le l$.

We now assert that if there is a sequence of levels $\mathcal{M}$ such that all sequences $\{\mu_{k_i}\}$ with $0 \le i \le l$ are bounded, then $\|\nabla f(x_{k_i})\|$ converges to zero for some index i. Condition (4.8) on the multipliers implies that there is an index i with $1 \le i \le l$ such that $\lambda_{k_i} \ge 1/(1+l) > 0$. Since l is independent of the level m, $\{\lambda_{k_i}\}$ is bounded away from zero. Moreover, (4.9) implies that

$$\lambda_{k_i} \|\nabla f(x_{k_i})\| \le \mu_{k_i} h_{k_i} + \mu_{k_i - 1} h_{k_i - 1}.$$

Since $\{h_{k_i}\}$ converges to zero and all μ_{k_i} are bounded for levels in $\mathcal{M}$, we have shown that $\|\nabla f(x_{k_i})\|$ converges to zero as desired. ■

We now investigate the behavior of the sequence of piecewise linear paths $\{p_m\}$ generated by the elastic string algorithm, where p_m is defined by setting $p_m(t_k) = x_k$ for $0 \le k \le m+1$, and $\{t_k\}$ is a quasi-uniform partition of $[0, 1]$.

Theorem 5.6 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on some open set that contains Ω, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level. If the growth condition* (5.2) *holds, then any limit point of the paths $\{p_m\}$ generated by the elastic string algorithm is a path p^* that crosses a critical point x^* of f. Moreover, $\{p_m\}$ has at least one limit point.*

Proof. Since all the iterates $\{x_k\}$ are bounded, the sequence of paths $\{p_m\}$ is uniformly bounded. Moreover, $\{p_m\}$ is equicontinuous, since $|t_\beta - t_\alpha| \le \max\{t_{k+1} - t_k\}$ implies that

$$\|p_m(t_\beta) - p_m(t_\alpha)\| \le \max\{h_k\},$$

and $\{h_k\}$ converges to zero. In particular, any limit point of $\{p_m\}$ is continuous. Moreover, the Arzela-Ascoli theorem shows that $\{p_m\}$ has a limit point $p^* \in C[0, 1]$. Hence, p^* is a path that connects x_a with x_b.

We now show that p^* crosses some critical point of f. Theorem 5.5 shows that $\|\nabla f(x_{k_i})\|$ converges to zero for some index i, and since all the iterates are bounded, $\{x_{k_i}\}$ has a limit point x^* with $\nabla f(x^*) = 0$. Thus, if we define t_{k_i} by $p_m(t_{k_i}) = x_{k_i}$, then $p^*(t) = x^*$ for some $t \in [0, 1]$ as desired. ■

We have shown that the elastic string algorithm is guaranteed to find a critical point, but not a critical point of least function value. That is, there is no guarantee that if ν^* is a limit point of $\{\nu_m\}$, then $\nu^* = \gamma$, where γ is defined by (2.3). We also know little about the eigenvalue structure of this critical point, but we now show that this critical point is not likely to be a minimizer.

Theorem 5.7 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on some open set that contains Ω, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level. If the growth condition* (5.2) *holds, then*

$$\limsup_{m\to\infty} \min \left\{ \Lambda_1 \left[\nabla^2 f(x_{k_i}) \right] : 1 \le i \le l \right\} \le 0, \tag{5.4}$$

where $\Lambda_1[A]$ is the smallest eigenvalue of the symmetric matrix A.

Proof. We need to show that for every sequence there is a finer subsequence $\mathcal{M}$ such that

$$\Lambda_1 \left[\nabla^2 f(x_{k_i}) \right] \le 0.$$

for some index i with $1 \le i \le l$. Lemma 5.3 shows that for every sequence of levels there is a finer subsequence $\mathcal{M}$ and an index i with $0 \le i \le l$ such that $\mu_{k_i} = 0$. Hence, either there is an index j with $1 \le j \le l$ such that $\mu_{k_{j-1}} = 0$ and $\mu_{k_j} > 0$, or $\mu_{k_j} = 0$ for all j with $1 \le j \le l$. We consider these two cases.

If there is an index j with $1 \le j \le l$ such that $\mu_{k_{j-1}} = 0$ and $\mu_{k_j} > 0$, then we argue as in the proof of Lemma 5.4 by first noting that the mean value theorem implies that

$$f(x_{k_j+1}) = f(x_{k_j}) + \langle \nabla f(x_{k_j}), s_{k_j} \rangle + \tfrac{1}{2} \langle s_{k_j}, A_{k_j} s_{k_j} \rangle,$$

where $A_{k_j} = \nabla^2 f(x_{k_j} + \xi_{k_j} s_{k_j})$ for some $\xi_{k_j} \in (0,1)$. Since $k_j < m$ and $f(x_{k_j+1}) \le f(x_{k_j})$,

$$\tfrac{1}{2} \langle s_{k_j}, A_{k_j} s_{k_j} \rangle \le -\langle \nabla f(x_{k_j}), s_{k_j} \rangle.$$

We have chosen k_j so that $\mu_{k_j} > 0$ and $\mu_{k_j-1} = 0$, and thus (4.9) implies that

$$\lambda_{k_j} \langle \nabla f(x_{k_j}), s_{k_j} \rangle = \mu_{k_j} \|s_{k_j}\|^2 = \mu_{k_j} h_{k_j}^2 > 0.$$

The last two inequalities show that $\langle s_{k_j}, A_{k_j} s_{k_j} \rangle < 0$, and thus, for sufficiently large m, we have that $\Lambda_1 \left[\nabla^2 f(x_{k_j}) \right] \le 0$, as desired.

The argument in the case where $\mu_{k_j} = 0$ for all j with $1 \le j \le l$ is similar. In this case, (4.9) implies that $\nabla f(x_{k_j}) = 0$, and thus the above argument yields that

$$f(x_{k_j+1}) = f(x_{k_j}) + \tfrac{1}{2} \langle s_{k_j}, A_{k_j} s_{k_j} \rangle.$$

Since l is bounded, independent of the level, we can choose k_j so that $f(x_{k_j+1}) < f(x_{k_j})$, and then $\langle s_{k_j}, A_{k_j} s_{k_j} \rangle < 0$ as desired. ∎

Theorem 5.7 shows that there is an index i such that $\Lambda_1 \left[\nabla^2 f(x^*) \right] \le 0$ for any limit point x^* of $\{x_{k_i}\}$, and thus if $\nabla^2 f(x^*)$ is nonsingular, then the Hessian matrix has at least one negative eigenvalue. This implies that the elastic string algorithm is likely to find a mountain pass. In future work we will examine the conditions under which the elastic string algorithm is guaranteed to find a critical point that satisfies the conditions of Theorem 2.5, that is, a critical point with precisely one negative eigenvalue.

6 Convergence Analysis: Bounded Level Sets

Our convergence results, Theorem 5.5 and 5.7, were obtained under the growth assumption (5.2), and, as we have observed, this assumption rules out coercive functions. We can generalize the convergence results by noting that the key result in the convergence theory of the elastic string algorithm is Lemma 5.3. Once we show that there is no sequence of levels $\mathcal{M}$ such that $\mu_{k_i} > 0$ for all i with $0 \leq i \leq l$, then all the convergence results follow. Lemma 5.3 implies, in particular, that L_m is updated a finite number of times, and thus Ω_m is fixed for m large enough. Hence, all the breakpoints $\{x_k\}$ are uniformly bounded, independent of the level m.

We can obtain a convergence theory for coercive functions by showing that the values of ν_m in the elastic string algorithm are bounded. In particular, assume that given $p_0 \in \Gamma$, the critical point of problem (4.6) chosen by the elastic string algorithm satisfies

$$\nu_m \leq \max \{ f[p_0(t)] : t \in [0,1] \}, \tag{6.1}$$

where ν_m is the value of the optimization problem (4.6) at a critical point. This is a natural requirement because ν_m satisfies (6.1) when p_0 is the piecewise linear path generated by the elastic string algorithm. We are requiring, however, that (6.1) hold for a fixed $p_0 \in \Gamma$.

We can satisfy (6.1) by imposing an additional requirement on the optimization algorithm used to solve (4.6) and by choosing the initial point in the optimization algorithm so that

$$\nu(x_0) \leq \max \{ f[p_0(t)] : t \in [0,1] \}, \tag{6.2}$$

where $\nu : \mathbb{R}^{mn} \mapsto \mathbb{R}$ is defined by (4.4). We now show how to satisfy (6.1) and (6.2) in terms of the optimization problem (4.3), that is,

$$\min \{ \nu(x) : c(x) \leq 0 \},$$

where $c : \mathbb{R}^{mn} \mapsto \mathbb{R}^{m+1}$ are the constraints in (4.3); the discussion carries over to (4.6).

We require that if a feasible starting point x_0 is chosen for (4.3), then the optimization algorithm determines a critical point x^* such that $\nu(x^*) \leq \nu(x_0)$. Clearly, if this requirement holds and if the starting point x_0 satisfies (6.2), then (6.1) holds. We can meet this requirement by using a feasible optimization algorithm such as the feasible sequential quadratic programming algorithm of Lawrence and Tits [13]. Another option is to use a general optimization algorithm on the problem

$$\min \{ \nu(x) : \nu(x) \leq \nu(x_0),\ c(x) \leq 0 \},$$

and assume that the critical point x^* satisfies $\nu(x^*) < \nu(x_0)$. This option is attractive in our case because the additional constraint $\nu(x) \leq \nu(x_0)$ can be satisfied by adding a bound to ν in the formulation (4.6).

Choosing a feasible starting point so that (6.2) holds requires some care because x_0 depends on the level. If, however, we can choose a partition $\{t_k\}$ of $[0,1]$ such that the starting point $x_0 = \{p_0(t_k)\}$ is feasible, then clearly (6.2) holds.

Lemma 6.1 *Assume that the bounds $h_k > 0$ satisfy* (4.2)*. If the path p_0 connecting x_a with x_b satisfies*

$$\int_0^1 \|p_0'(t)\|\, dt \le L, \tag{6.3}$$

then there is a partition $\{t_k\}$ of $[0,1]$ such that

$$\|p_0(t_{k+1}) - p_0(t_k)\| \le h_k, \qquad 0 \le k \le m. \tag{6.4}$$

Proof. We first define the partition $\{t_k\}$. Given t_k, define $\phi : \mathbb{R} \mapsto \mathbb{R}$ by

$$\phi_k(t) = ||p_0(t) - p(t_k)\|.$$

If $\phi_k(1) \le h_k$ then we set $t_j = 1$ for $k \le j \le m$. If, on the other hand, $\phi_k(1) > h_k$, then there is a $t_{k+1} \in (t_k, 1)$ such that $\phi_k(1) = h_k$. This defines the partition $\{t_k\}$. If at any stage of the construction $\phi_k(1) \le h_k$, then (6.4) holds, since $t_m = 1$. Otherwise,

$$\|p_0(t_{k+1}) - p_0(t_k)\| = h_k, \qquad 0 \le k < m,$$

and we need to show that $\|p_0(1) - p_0(t_m)\| \le h_m$. First we note that since

$$h_k = \|p_0(t_{k+1}) - p_0(t_k)\| \le \int_{t_k}^{t_{k+1}} \|p_0'(t)\|\, dt, \qquad 0 \le k < m,$$

the bound (6.3) on the length of p_0 implies that

$$\sum_{k=0}^{m-1} h_k \le \int_0^{t_m} \|p_0'(t)\|\, dt \le L - \int_{t_m}^1 \|p_0'(t)\|\, dt.$$

Hence, assumption (4.2) on the bounds h_k proves that

$$\|p_0(1) - p_0(t_m)\| \le \int_{t_m}^1 \|p_0'(t)\|\, dt \le L - \sum_{k=0}^{m-1} h_k = h_m,$$

as desired. ■

Lemma 6.1 shows that if p_0 is a path of finite length and if L is an upper bound on the length of p_0, then we can choose a partition $\{t_k\}$ of $[0,1]$ such that the starting point $x_0 = \{p_0(t_k)\}$ is feasible. Hence, (6.2) holds, and thus (6.1) is satisfied by any optimization algorithm that decreases the value of ν.

We now show that the convergence results hold for coercive functions by assuming that there is an $r > 0$ such that

$$f(x) > \max\{f[p_0(t)] : t \in [0,1]\}, \qquad \|x - x_a\| \ge r, \tag{6.5}$$

where $p_0 \in \Gamma$ is a fixed path. We could choose $p_0(t) = x_a + t(x_b - x_a)$, but this condition allows a more general choice of $p_0 \in \Gamma$. We also note that, unlike (5.2), assumption (6.5) does not imply that S is compact.

The following result is the analogue of Lemma 5.3 for condition (6.5).

Lemma 6.2 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on $\mathbb{R}^n$, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level, and that the elastic string algorithm generates iterates that satisfy* (6.1)*. If the growth condition* (6.5) *holds, then there is no sequence of levels $\mathcal{M}$ such that $\mu_{k_i} > 0$ for all i with $0 \le i \le l$.*

Proof. The proof is similar to Lemma 5.3. Assume, on the contrary, that there is a sequence of levels $\mathcal{M}$ such that $\mu_{k_i} > 0$ for all i with $0 \le i \le l$. Hence, the updating rules for L guarantee that $L \to \infty$ as the level m increases. Since Lemma 5.2 guarantees that

$$\sum_{i=0}^{l} \|x_{k_{i+1}} - x_{k_i}\| = L,$$

and since $L \to \infty$, we obtain that for m large enough,

$$\max \{\|x_{k_i} - x_a\| : 1 \le i \le l\} \ge \frac{L}{(2l+1)} \ge r,$$

and thus assumption (6.5) on the behavior of f implies that

$$\nu_m = f(x_{k_i}) > \max \{f[p_0(t)] : t \in [0,1]\},$$

for some index i. However, this contradicts requirement (6.1) on the elastic string algorithm. ■

As we have remarked, convergence of the elastic string algorithm for coercive functions is a direct consequence of Lemma 5.3. We note this result for future reference.

Theorem 6.3 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable on $\mathbb{R}^n$, and assume that the points x_a and x_b are separated by a closed set S such that* (2.2) *holds. Assume, in addition, that l is bounded, independent of the level and that the elastic string algorithm generates iterates that satisfy* (6.1)*. If the growth condition* (6.5) *holds, then the convergence results* (5.3) *and* (5.4) *hold.*

7 Computational Experiments

The elastic string algorithm can be used to find mountain passes by choosing the number of breakpoints m and solving the constrained optimization problem (4.3) using a general constrained optimization algorithm. Since this optimization problem has $nm + 1$ variables and $2m + 1$ constraints, large values of m can impose severe demands on the optimization algorithm, in particular, when n is large. This concern motivates our study of the behavior of the elastic string algorithm for modest values of m.

We expect that the elastic string algorithm will produce a rough approximation to a mountain pass for small values of m. In the computational experiments, we evaluate this hypothesis by determining when the approximation x_s^* produced by the elastic string algorithm is a suitable approximation to Newton's method for solving $\nabla f(x) = 0$.

Newton's method for $\nabla f(x) = 0$ is locally and quadratically convergent if x_s^* is sufficiently close to a nondegenerate mountain pass x^* and $\nabla^2 f$ is locally Lipschitzian. We implemented Newton's method with an Armijo line search, that is,

$$x_{k+1} = x_k - \alpha_k \nabla^2 f(x_k)^{-1} \nabla f(x_k),$$

where α_k is the line-search parameter. Each Newton direction was calculated by using a factorization of the Hessian matrix. Other methods for solving indefinite systems of equations could also be applied. Since we expect the Hessian matrix to have one negative eigenvalue at the mountain pass, using a Cholesky factorization or conjugate gradients is inappropriate. We terminate Newton's method when

$$\|\nabla f(x_k)\| \le \tau, \qquad \tau = 10^{-6}.$$

The mountain pass determined in this manner is denoted by x^*. In most of the test cases, the approximate mountain pass x_s^* is close enough to x^* so that the gradient norm $\|\nabla f(x_k)\|$ of the Newton iterates converges quadratically to zero.

Our benchmark problems were implemented in the AMPL [11] modeling language. The optimization problem (4.3) was solved by using the versions of KNITRO [5] and LOQO [19] available on the NEOS Server [17]. In general, both codes were able to find a solution to the optimization problem (4.3) within the alloted number of iterations; we note in the text any failures that occurred. We do not provide information on the relative performance on these problems because this is not relevant to the results.

We checked that the mountain-pass x^* had precisely one negative eigenvalue (Morse index 1), as predicted by Theorem 2.5, by computing the two smallest eigenvalues of the Hessian matrix. This analysis was performed in MATLAB using the Hessian evaluations supplied by AMPL.

We tested the elastic string algorithm on benchmark chemistry problems that require the computation of transition structures and reaction pathways. These problems are described in Section 8. We also tested the elastic string algorithm on variational problems associated with semilinear partial differential equations. These results are presented in Section 9.

We end this section by discussing the determination of mountain passes for the six-hump camel back function defined by

$$f(\xi_1, \xi_2) = \left(4 - 2.1\xi_1^2 + \frac{1}{3}\xi_1^4\right)\xi_1^2 + \xi_1\xi_2 + 4(\xi_2^2 - 1)\xi_2^2.$$

This function is frequently used to test global optimization algorithms, but it is also useful because it is simple and yet the results obtained are fairly typical of the results obtained for the application problems presented in the next two sections.

As shown in Figure 7.1, this function has six minimizers and six mountain passes. We have tested the elastic string algorithm by choosing $x_a = (-1.5, -0.6)$ and selecting various

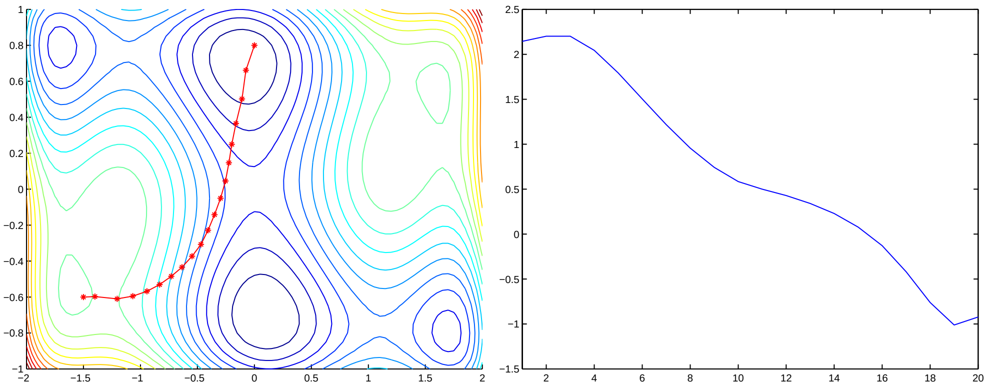

Figure 7.1: Contours for the six-hump camel back function and the path profile.

values for x_b, but we present results only for $x_b = (0.0, 0.8)$. We used a uniform partition with

$$h_k \equiv h = \frac{2}{m+1} \|x_b - x_a\|$$

and $m = 20$ breakpoints. The starting point for the constrained optimization algorithm (breakpoints for the piecewise linear path) were initialized to x_a.

Figure 7.1 presents a contour plot of the six-hump camel back function overlaid with the path calculated by the elastic string algorithm, along with a plot of the path profile, that is, a plot of the function values $f(x_k)$ for $1 \le k \le m$.

The elastic string algorithm found an approximate mountain pass x_s^* with a gradient norm of $\|\nabla f(x_s^*)\| = 0.49$. Newton's method from x_s^* required four iterations to obtain an accurate mountain pass, and for all iterations the step length $\alpha_k = 1$.

The results obtained for the six-hump camel back function are fairly typical of the behavior of the elastic string algorithm. For most problems the accuracy of x_s^* is low and directly related to the value of the bound h. However, the approximate mountain pass x_s^* provides a good starting point for Newton's method, and convergence is obtained in a few iterations, usually three or four iterations.

The path profile of function values in Figure 7.1 is also typical in the sense that two points with the same maximum function value are found straddling a mountain pass. This can be explained by noting that if the path p is known, then the elastic string algorithm can be viewed as choosing a partition $\{\tau_k\}$ for $[0, 1]$ that solves the optimization problem

$$\min \Big\{ \nu(\tau_1, \dots, \tau_m) : |\tau_{k+1} - \tau_k| \le h_k, \ 0 \le k \le m \Big\},$$

where h_k satisfies the quasi-uniform restriction (4.5),

$$\nu(\tau_1, \dots, \tau_m) = \max \{ \phi(\tau_k) : 1 \le k \le m \},$$

and ϕ is the mapping $t \mapsto f[p(t)]$. Assume that ϕ is continuously differentiable, and let ν^* be the value of this problem at an optimal partition. We cannot have $\phi(t) = \nu^*$ at a unique mesh point $\tau \in \{\tau_k\}$ because a small perturbation of the mesh point yields a smaller value for ν. This claim is clear if ϕ has a unique maximizer in $(0, 1)$, but holds in general. Hence, as shown in Figure 7.1, we have $\phi(t) = \nu^*$ for at least two mesh points. In most cases $\phi(t) = \nu^*$ at precisely two points unless ϕ is constant on some line segment or ϕ is periodic.

We note that the piecewise linear path calculated by the elastic string algorithm is not unique. Only the breakpoints with maximum function value are relevant when determining an approximate mountain pass, and little can be said about other local maximizers that are not global maximizers.

We also note that the eigenvalues of the Hessian at the approximate mountain pass x_s^* are usually relatively close to those of the mountain pass x^*. For the six-hump camel back function the two smallest eigenvalues at x_s^* were $(-3.0, 9.2)$. These values changed to $(-6.1, 9.6)$ for the accurate mountain pass.

8 Transitions States and Reaction Pathways

Computational chemists are interested in calculating mountain passes when the function is the potential energy surface of a chemical reaction. Such problems are of interest from an optimization viewpoint because these potential energy surfaces are usually highly nonlinear and have many minimizers.

We consider two problems proposed by Henkelman, Jóhannesson, and Jónsson [12]. In the first problem the potential energy function models a reaction involving three atoms with motion restricted to a line. The two local minimizers of this function correspond to the reactants and products in the reaction, and the mountain pass between the reactants and products is the transition state.

The potential energy function for this problem is defined in terms of functions Q that model Coulomb interactions, and functions J for quantum mechanical interactions:

$$\begin{aligned}
Q_{ab}(\xi_1, \xi_2) &= 2.260(1.5e^{-3.884(\xi_1-0.742)} - e^{-1.942(\xi_1-0.742)}) \\
Q_{bc}(\xi_1, \xi_2) &= 1.318(1.5e^{-3.884(\xi_2-0.742)} - e^{-1.942(\xi_2-0.742)}) \\
Q_{ac}(\xi_1, \xi_2) &= 1.605(1.5e^{-11.652} - e^{-5.826}) \\
J_{ab}(\xi_1, \xi_2) &= 1.130(e^{-3.884(\xi_1-0.742)} - 6e^{-1.942(\xi_1-0.742)}) \\
J_{bc}(\xi_1, \xi_2) &= 0.659(e^{-3.884(\xi_2-0.742)} - 6e^{-1.942(\xi_2-0.742)}) \\
J_{ac}(\xi_1, \xi_2) &= 0.820(e^{-11.652} - 6e^{-5.826}).
\end{aligned}$$

The LEPS potential energy function is then defined by

$$\begin{aligned}
V(\xi_1, \xi_2) = \; & Q_{ab}(\xi_1, \xi_2) + Q_{bc}(\xi_1, \xi_2) + Q_{ac}(\xi_1, \xi_2) - \\
& \Big(J_{ab}(\xi_1, \xi_2)^2 + J_{bc}(\xi_1, \xi_2)^2 + J_{ac}(\xi_1, \xi_2)^2 - \\
& J_{ab}(\xi_1, \xi_2) J_{bc}(\xi_1, \xi_2) - J_{ab}(\xi_1, \xi_2) J_{ac}(\xi_1, \xi_2) - J_{bc}(\xi_1, \xi_2) J_{ac}(\xi_1, \xi_2) \Big)^{1/2}.
\end{aligned}$$

The final potential energy function used in [12] adds an additional term so that the final potential energy function is

$$V_1(\xi_1, \xi_2) = V(\xi_1, 3.742 - \xi_1) + 0.405(\xi_1 + 0.867\xi_2 - 1.871)^2.$$

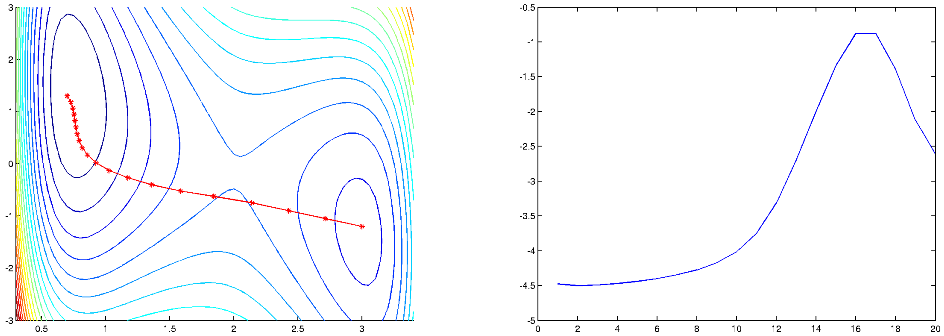

Figure 8.1: Contours and path profile for the LEPS potential.

See [12] for a more detailed description of this potential energy function.

The parameters for the elastic string algorithm were set as in the six-hump camel back problem of Section 7. We used

$$h_k \equiv h = \frac{2}{m+1}\|x_b - x_a\|,$$

with $m = 20$ breakpoints, and the breakpoints for the piecewise linear path were initialized to x_a.

Figure 8.1 presents a contour plot of the potential energy function with the path calculated by the elastic string algorithm along with a plot of the function values $f(x_k)$ for $1 \le k \le m$. These plots were obtained with x_a and x_b near the minimizers of f. We used $x_a = (0.7, 1.3)$ and $x_b = (3.0, -1.2)$.

The elastic string algorithm with either LOQO or KNITRO solved this problem, but LOQO found $f(x_s^*) = -0.87$, while KNITRO obtained $f(x_s^*) = -0.97$. The approximate mountain pass x_s^* had a relatively large gradient of $\|\nabla f(x_s^*)\| = 1.0$, but in spite of this Newton's method converged in four iterations. The two smallest eigenvalues of the Hessian matrix at x_s^* are $(-5.7, 0.68)$; these change to $(-8.0, 0.66)$ at the accurate mountain pass.

In the second benchmark problem that appears in [12], a Gaussian function is added to the LEPS potential near the site of the original mountain pass. The aim is to create two saddle points. The Gaussian potential is then

$$V_2(\xi_1, \xi_2) = V_1(\xi_1, \xi_2) + 1.5 \exp\left(-0.5\left(\left(\frac{\xi_1 - 2.02083}{0.10}\right)^2 + \left(\frac{\xi_2 + 0.27288}{0.35}\right)^2\right)\right).$$

The contour plot of the Gaussian potential energy function in Figure 8.2 shows that this potential energy function does have two saddle points.

The results obtained by the elastic string algorithm for this potential are similar to those for the LEPS potential. For this problem $\|\nabla f(x_s^*)\| = 2.1$, and the two smallest eigenvalues of the Hessian matrix are $(-1.6, 15.6)$ at x_s^* and $(-15.8, 4.3)$ at x^*. In this case, however,

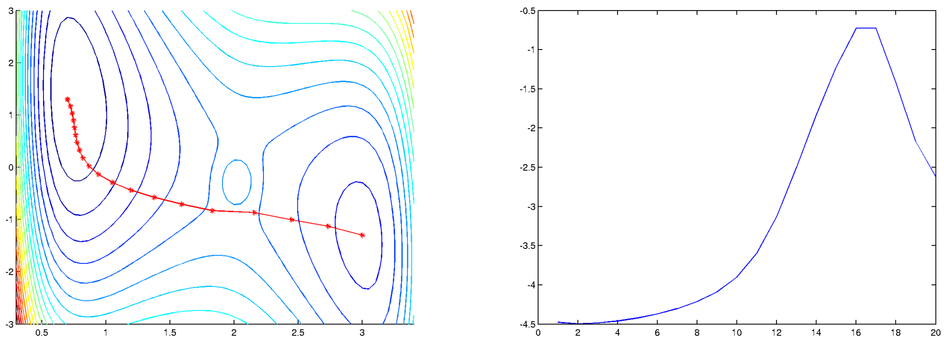

Figure 8.2: Contours and path profile for the Gaussian potential.

Newton's method required five iterations to converge from x_s^*, and the second iterate used a line-search parameter of 0.5.

9 Mountain Passes for Variational Problems

The next set of computational experiments explores the behavior of the elastic string algorithm on finite-dimensional approximations to the variational problem (3.1) discussed in Section 2. Our aim is to show that the elastic string algorithm, with small values of the level m, computes an approximation x_s^* to a mountain pass that is also a suitable starting point for Newton's method. We want Newton's method started at x_s^* to converge quadratically to a mountain pass x^*.

The choice of problems follows those of Chen, Zhou, and Ni [7], but as noted in the introduction, our approach to computing mountain passes is different. We consider various geometries for $\mathcal{D}$ to explore the behavior of the mountain passes for nonstandard geometries. In particular, we use the Lane-Emden problem over the unit square, the singularly perturbed Dirichlet problem over the unit circle, and the Henon problem on a domain consisting of the unit circle with a smaller square cut out of the center. The domains and meshes used for the singularly perturbed Dirichlet and Henon problems are shown in Figure 9.1. We used homogeneous Dirichlet boundary conditions for all of these test problems.

We discretize the variational problem by using difference approximations based on a triangularization of $\mathcal{D}$. The finite-dimensional approximation is of the form

$$f(x) = \sum_{k=1}^{n_e} f_k(x),$$

where n_e is the number of elements, $f_k : \mathbb{R}^n \mapsto \mathbb{R}$ is an approximation to the integral

$$\int_{\mathcal{T}_k} \left(\tfrac{1}{2} \|\nabla u(s)\|^2 - h[s, u(s)] \right) \, ds,$$

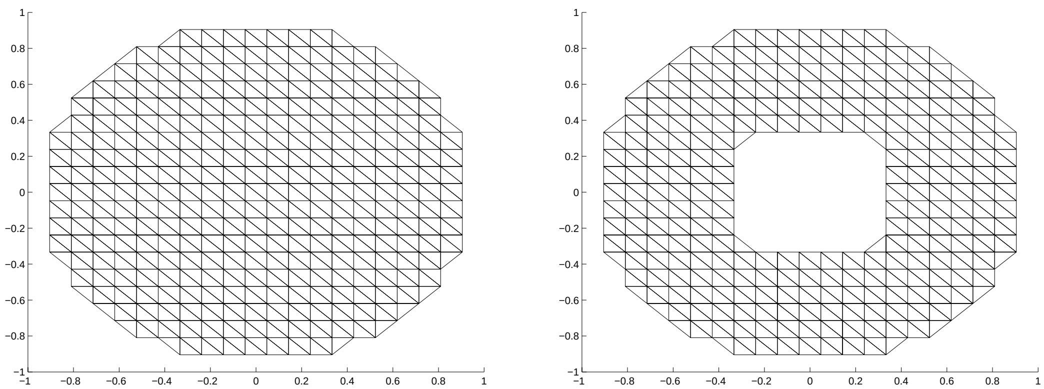

Figure 9.1: Meshes for the singularly perturbed Dirichlet (left) and Henon (right) problems.

and $\mathcal{T}_k$ is the kth triangular element. The function f_k depends only on the interior grid points associated with $\mathcal{T}_k$, while f depends on the n interior grid points.

As noted in Section 2, for these problems $x_a = 0$ is a local minimizer of the finite-dimensional approximation, and thus $S = \partial B(x_a, r)$ is a compact separating set for all $r > 0$ sufficiently small. Since $f(x_a) = 0$, any x_b with $f(x_b) \le 0$ satisfies the assumptions of the mountain-pass theorem. Theorem 3.1 shows that any x_b with $\|x_b\|$ large enough satisfies $f(x_b) \le 0$, but if $\|x_b\|$ is large, then $L \ge \|x_b - x_a\|$ must be large, and thus the bounds $\{h_k\}$ are small only with large values of m.

In the numerical results x_b is produced by first choosing an x_c with $f(x_c) < f(x_a)$, and then setting $\tau > 0$ to the smallest element in the set $\{\mu^k : k \ge 0\}$, where $\mu \in (0, 1)$, so that

$$f(x_a + \tau(x_c - x_a)) < f(x_a).$$

Defining $x_b = x_a + \tau(x_c - x_a)$ yields an x_b that can be significantly closer to x_a than x_c. As a result, the bounds $\{h_k\}$ are relatively small for modest values of m.

We did not experiment with an adaptive choice of the bounds $\{h_k\}$ because our intention is to present results for a basic implementation. We used a uniform partition with

$$h_k \equiv h = \frac{2}{m+1}\|x_b - x_a\|, \qquad 0 \le k \le m.$$

The constrained optimization algorithms used to solve problem (4.3) require a starting value (x_0, ν_0), and we used $x_0 = 0$ and $\nu_0 = 0$. This starting point violates the constraints associated with h_0 and h_m but is otherwise feasible. On the other hand, $x_0 = 0$ is not a good approximation to the mountain pass. The performance of the constrained optimization algorithms used to solve (4.3) might be improved by using a better starting point, but we wanted to show that the formulation (4.3) is suitable even if a poor starting point is chosen.

The literature on the critical points of the variational problem (3.1) is extensive. We noted in Section 2 that if (3.4) holds, then (3.1) has nontrivial critical points $u^+ \ge 0 \ge u^-$. In addition, Struwe [18, Theorem 6.6] points out that if $t \mapsto h(s, t)$ is even, then (3.1) has

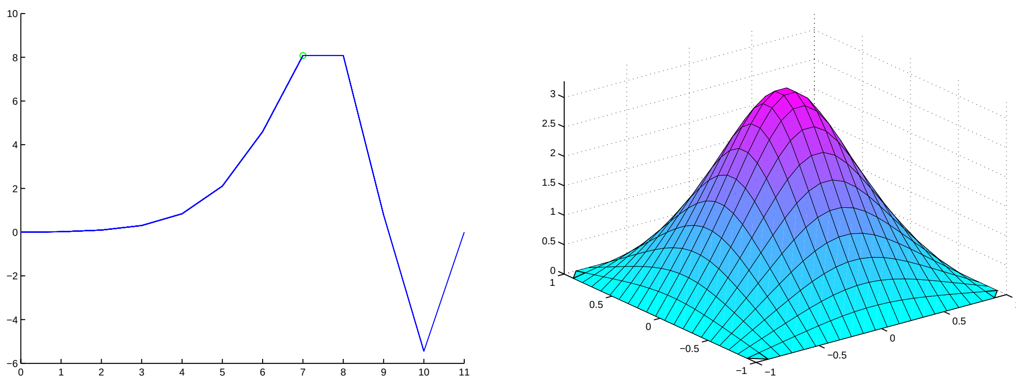

Figure 9.2: Path profile and mountain pass for the Lane-Emden problem.

an infinite number of critical points. This result is relevant because $t \mapsto h(s,t)$ is even for the problems in our computational experiments. Of course, the elastic string algorithm determines a mountain-pass solution, and this restricts the class of admissible solutions. For example, a mountain-pass solution does not change sign in $\mathcal{D}$ for the problems in this section. Chen, Zhou, and Ni [7] provide additional information on the theoretical properties of the problems in this section.

The first set of computational results is for the Lane-Emden equation $-\Delta u = u^3$ on the unit square $\mathcal{D} = (-1,1) \times (-1,1)$. As noted in Section 2, the variational functional associated with this problem is

$$\int_{\mathcal{D}} \left(\frac{1}{2} \|\nabla u(s)\|^2 - \frac{1}{4} u(s)^4 \right) ds.$$

The elastic string algorithm with $m = 10$ finds an approximate mountain pass x_s^* with $\|\nabla f(x_s^*)\| = 0.66$, but with $(-0.19, -0.0084)$ as the two smallest eigenvalues of the Hessian at x_s^*. Although the Hessian matrix had more than one negative eigenvalue, Newton's method converged in four iterations to the mountain pass x^* in Figure 9.2. The two smallest eigenvalues $(-0.10, 0.027)$ of the Hessian at x^* now have the proper signs.

The strategy of using Newton's method from x_s^* is questionable when the inertia of the Hessian matrix is wrong. In this case, however, the second smallest eigenvalue is relatively small, so it is not unreasonable to find an incorrect inertia. In the same vein, we note that the elastic string algorithm with $m = 20$ produces an x_s^* with the correct inertia.

Figure 9.2 presents a plot of the path profile drawn from the function values $f(x_k)$ for $0 \le k \le m+1$, where the circled iterates have the correct inertia. This plot shows that the maximum is achieved at two points, just as in the results of Sections 7 and 8. The ragged nature of the plot for the Lane-Emden problem is due to the use of $m = 10$; a smoother profile is obtained with higher values of m.

We next consider the singularly perturbed Dirichlet problem $-\varepsilon^2 \Delta u = u^3 - u$ with $\mathcal{D}$

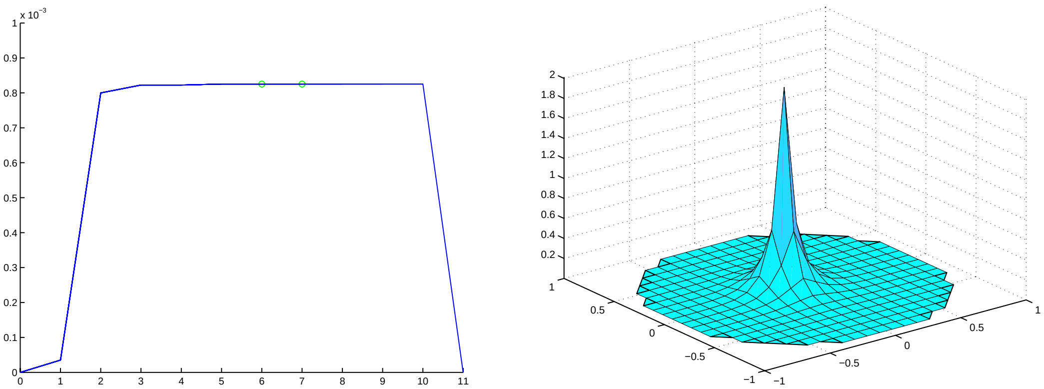

Figure 9.3: Path profile and mountain pass for the singularly perturbed Dirichlet problem.

the unit circle in $\mathbb{R}^2$. For this problem the variational formulation is

$$\int_{\mathcal{D}} \left(\frac{\varepsilon^2}{2} \|\nabla u(s)\|^2 + \frac{1}{2} u(s)^2 - \frac{1}{4} u(s)^4 \right) ds,$$

so we can expect numerical difficulties as $\varepsilon \to 0$. Chen, Zhou, and Ni [7] examine the the behavior of the mountain pass as $\varepsilon \to 0$, but we explore only the case $\varepsilon^2 = 10^{-2}$.

The elastic string algorithm with $m = 10$ obtained the mountain pass in Figure 9.3. In this case $\|\nabla f(x_s^*)\| = 0.14$, and the two smallest eigenvalues of the Hessian at x_s^* are $(-0.21, 0.0088)$. Thus, the Hessian matrix has one negative eigenvalue as desired. Moreover, Newton's method converged in four iterations to the solution shown in Figure 9.3, and the two smallest eigenvalues of the Hessian at x^* are $(-0.065, 0.0096)$.

The path profile shown in Figure 9.3 achieves the maximum at several points. This result is unexpected, since in all previous cases the maximum was achieved only at two values. This result, however, is due to the poor resolution with small values of m. For this problem the path profile is resolved with $m = 40$, and as seen in Figure 9.4, the path profile achieves the maximum at only two points. We note that the plot on the left of Figure 9.4 shows the path profile for the whole path, while the plot on the right of the path profile for breakpoints $5, \ldots, 20$ that lie in the interval $[-0.5, 0.5]$.

We also note that the path profile in Figure 9.3 is not well determined, since the mountain-pass level is small. In this case $f(x_s^*) = 8.2 \times 10^{-4}$; hence, the barrier to be crossed is unusually low and can create numerical difficulties. We can raise the value of the barrier in a natural way by considering the functional $u \mapsto \varepsilon^{-2} f(u)$, where f is the functional in the original formulation, but we did not explore this option.

Results obtained with larger values of m show that the value of $f(x_s^*)$ increases as m increases. In particular, $f(x_s^*) = 3.1 \times 10^{-2}$ for the singularly perturbed Dirichlet problem with $m = 40$. This increase is to be expected because the subspace Γ_π in (4.1) increases as m increases.

The final problem is the Henon equation $-\Delta u = \|s\| u^3$ on a domain $\mathcal{D}$ that is the unit circle with a smaller square cut out of the middle. Figure 9.1 (right) shows the mesh used

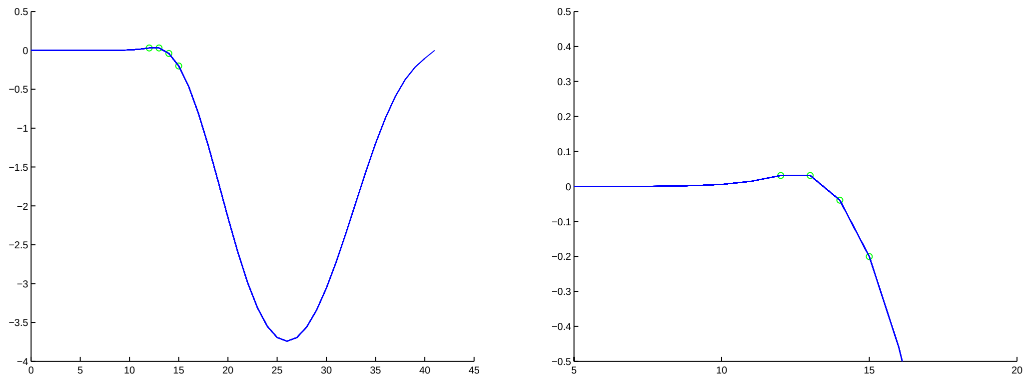

Figure 9.4: Path profile ($m = 40$) for the singularly perturbed Dirichlet problem.

to discretize the variational functional

$$\int_{\mathcal{D}} \left(\frac{1}{2} \|\nabla u(s)\|^2 - \frac{\|s\|}{4} u(s)^4 \right) \, ds$$

associated with the Henon equation. The plot of the mountain pass in Figure 9.5 shows that the solution lacks the symmetry of the preceding two problems. In general, the symmetry properties of the domain $\mathcal{D}$ are reflected in the symmetry properties of the solution. For example, Chen, Zhou, and Ni [7] noted that on an annular domain there is a solution that is not rotationally symmetric; and since the domain is rotationally symmetric, the mountain-pass solutions form a connected nontrivial set. In particular, the mountain-pass solutions are not isolated. For our domain, which is not rotationally symmetric, there seem to be four distinct mountain-pass solutions.

The elastic string algorithm with $m = 10$ produced an approximate mountain pass with $\|\nabla f(x_s^*)\| = 6.5$, with the two smallest eigenvalues of the Hessian matrix being $(-0.20, 0.20)$. Newton's method converged from x_s^* but required ten iterations, indicating that the accuracy of x_s^* was not adequate. The numerical results with $m = 20$ show an improvement in

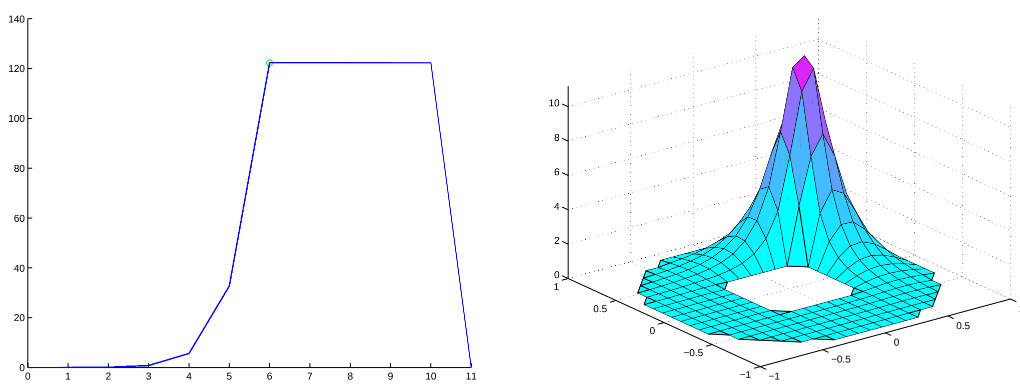

Figure 9.5: Path profile and mountain pass for the Henon equation.

the convergence of Newton's method: in this case convergence from x_s^* requires only five iterations, with all step lengths $\alpha_k = 1$.

The results obtained with $m = 10$ appear in Figure 9.5. The path profile shows that the maximum is achieved for more than two breakpoints because of the low accuracy of the approximate mountain pass x_s^*. We encountered a similar situation with the singularly perturbed Dirichlet problem. Results obtained with $m = 50$ show that the maximum is achieved at only two breakpoints.

Acknowledgments

Our work on mountain passes has benefited from discussions with various researchers, in particular, Mihai Anitescu and Andreas Griewank. In addition, Bob Vanderbei deserves special credit for his comments and insights on the elastic string algorithm during the Erice (July 2001) Conference on Optimization.

References

[1] A. Ambrosetti, *Elliptic equations with jumping nonlinearities*, J. Math. Phys. Sci., 18 (1984), pp. 1–12.

[2] A. Ambrosetti and P. H. Rabinowitz, *Dual variational methods in critical point theory and applications*, J. Funct. Anal., 14 (1973), pp. 349–381.

[3] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons, 1984.

[4] S. Bell, J. S. Crighton, and R. Fletcher, *A new efficient method for locating saddle points*, Chem. Phys. Lett., 82 (1982), pp. 122–126.

[5] R. Byrd, M. Hribar, and J. Nocedal, *An interior point method for large scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877–900.

[6] L. Caklovic, S. Li, and M. Willem, *A note on Palais-Smale condition and coercivity*, Differential Integral Equations, 3 (1990), pp. 799–800.

[7] G. Chen, J. Zhou, and W. Ni, *Algorithms and visualization for solutions of nonlinear elliptic equations*, Internat. J. Bifur. Chaos, 10 (2000), pp. 1565–1612.

[8] Y. S. Choi and P. J. McKenna, *A mountain pass method for the numerical solution of semilinar elliptic problems*, Nonlinear Anal., 20 (1993), pp. 417–437.

[9] Z. Ding, D. Costa, and G. Chen, *A high-linking algorithm for sign-changing solutions of semilinear elliptic equations*, Nonlinear Anal., 38 (1999).

[10] I. Ekeland, *Convexity Methods in Hamiltonian Mechanics*, Springer-Verlag, 1990.

[11] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, The Scientific Press, 1993.

[12] G. Henkelman, G. Jóhannesson, and H. Jónsson, *Methods for finding saddle points and minimum energy paths*, in Reviews in Computational Chemistry, D. Schwartz, ed., vol. 5, Kluwer, 2000.

[13] C. T. Lawrence and A. L. Tits, *A computationally efficient feasible sequential quadratic programming algorithm*, SIAM J. Optim., 11 (2001), pp. 1092–1118.

[14] Y. Li and J. Zhou, *A minimax method for finding multiple critical points and its applications to semilinear PDEs*, SIAM J. Sci. Comput., 23 (2001), pp. 840–865.

[15] ——, *Convergence results of a local minimax method for finding multiple critical points*, SIAM J. Sci. Comput., (2002). to appear.

[16] J. Mawhin and M. Willem, *Critical Point Theory and Hamiltonian Systems*, Springer-Verlag, 1989.

[17] *NEOS Server for Optimization Problems*. See `www-neos.mcs.anl.gov`.

[18] M. Struwe, *Variational Methods*, Springer-Verlag, 2000.

[19] R. J. Vanderbei, *LOQO user's manual – Version 3.10*, Opt. Meth. Software, 12 (1999), pp. 485–514.

[20] M. Willem, *Minimax Theorems*, Birkäuser, 1996.