Steklov Regularization and Trajectory Methods for Univariate Global Optimization

Orhan Arıkan* Regina S. Burachik† C. Yalçın Kaya‡

September 13, 2018

arXiv:1809.04530v1 [math.OC] 9 Sep 2018

Abstract

We introduce a new regularization technique, using what we refer to as the Steklov regularization function, and apply this technique to devise an algorithm that computes a global minimizer of univariate coercive functions. First, we show that the Steklov regularization convexifies a given univariate coercive function. Then, by using the regularization parameter as the independent variable, a trajectory is constructed on the surface generated by the Steklov function. For monic quartic polynomials, we prove that this trajectory does generate a global minimizer. In the process, we derive some properties of quartic polynomials. Comparisons are made with a previous approach which uses a quadratic regularization function. We carry out numerical experiments to illustrate the working of the new method on polynomials of various degree as well as a non-polynomial function.

Key words: Global optimization, mean filter, Steklov smoothing, Steklov regularization, scale-shift invariance, trajectory methods.

1 Introduction

Mean filter is a digital filtering technique in signal processing, which is used to remove noise. The technique can also be viewed as a smoothing procedure. In digital imaging, for example, this filtering technique is performed by replacing each pixel value by the mean value of its neighbours and itself in a "window" – see [16] and the references therein. The expected outcome is the removal of noise in the image and the smoothening of the image. The mean filter idea was originally proposed and has so far been used for the processing of discrete data.

In the present paper, we propose and analyse a similar idea in the setting of continuous optimization, involving a coercive univariate function instead of discrete data. When the averaging process described above is employed for a univariate function $f(x)$ over an interval (corresponding to a window) of variable size centred at x, i.e., $[x-t, x+t]$, with $t>0$, one obtains the well-known Steklov smoothing function [6–9], expressed in terms of the function f and the size of the interval, denoted here by $\mu(x,t)$. The Steklov smoothing function is typically used in getting an approximate solution to the problem of minimizing a nonsmooth f: The smooth function $\mu(x,t)$ is minimized over an interval, or window, with small t, so

*Electrical and Electronics Engineering Department, Bilkent University, Bilkent, 06800 Ankara, Turkey. E-mail: oarikan@ee.bilkent.edu.tr .

†School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, S.A. 5095, Australia. E-mail: regina.burachik@unisa.edu.au .

‡School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, S.A. 5095, Australia. E-mail: yalcin.kaya@unisa.edu.au .

that the solution of the smoothed problem is a close enough approximation to the solution of the original problem.

Although the properties of μ have very well been explored in the literature for small t, it has not yet been studied for large t. This is the point where our study steps in. In the present paper, first we show that for large enough t and certain coercive f, $\mu(\cdot, t)$ is strictly convex – see Theorem 1. In this sense, μ *regularizes* the function f for large t by convexifying (as well as smoothening) it, and that is the reason why we call it the *Steklov regularization function*. We note that, if $\mu(\cdot, t_0)$ is strictly convex for some $t_0 > 0$, then $\mu(\cdot, t_0)$ has a unique minimizer. The main aim of the current paper is to propose a method, namely Algorithm 1, based on constructing and following a trajectory between the unique minimizer of $\mu(\cdot, t_0)$ and a global minimizer of $f(x)$. The trajectory here is the solution of an ordinary differential equation (ODE) obtained by using $\mu(x, t)$.

Univariate global optimization has long been an active area of research – see [10, 11, 13] and the references therein. Most multidimensional iterative methods involve line searches at a given search direction, and these uni-dimensional searches are equivalent to the global minimization of a univariate function. Therefore, finding efficiently a global minimizer in such a line search has importance on its own, and can be crucial for the success of such iterative high-dimensional techniques. Thus the relevance of developing new, efficient univariate techniques. Moreover, the result of such a line search can be useful as a starting guess for the global minimizer of the original higher dimensional problem. Although the present paper focuses on the univariate case, an extension of our approach to the multi-variable case is under investigation.

Trajectory based methods are not new to optimization. The trajectories (to follow) in these methods are typically solutions of ODEs incorporating the gradient of $f(x)$. Convergence analyses for these types of methods have so far been given only for local minima – see, for example, [3, 5] and the references therein. Trajectory based methods have been proposed also for global optimization, albeit without a convergence proof, to the best knowledge of the authors – see, for example, [14].

We note one particular trajectory based technique for global optimization, the *backward differential flow* method, which was proposed by Zhu et al. in [17], where the trajectories are solutions of an ODE that emanates from the (classical) quadratic regularization function rather than the Steklov regularization function. We have recently illustrated that the backward differential flow method, given as Algorithm 3 in the current paper, may not yield a global minimizer, even in the case when the function is a quartic polynomial – see [1].

We provide a convergence proof of our approach for the case of quartic polynomials (see Algorithm 2 and Theorem 3). Our numerical experiments indicate that our method can be viewed as a better alternative to that given by Zhu et al. [17]. Indeed, our method converges to a global minimum in most of the (randomly generated) cases of even-higher-degree monic polynomials. On the other hand, the method by Zhu et al. fails to converge in the great majority of the cases – see Table 1.

In addition to the convexification and convergence results in Theorems 1 and 3, respectively, we provide auxiliary results, which are interesting in their own right. For example, we prove in Lemma 2 that, if Algorithm 1 can generate the global minimizer of a given function $f(x)$, then it can also generate the global minimizer of $f(\alpha\, x - a)$, where $a \in \mathbb{R}$ and $\alpha > 0$ are fixed. We refer to this property as the *scale-shift invariance* property. We note that, while Algorithm 1 (and thus Algorithm 2) is scale-shift invariant, Algorithm 3 is not. It is well-known that scale changes and translations can be used to simplify the expression of a function. For example, the third degree term of a quartic polynomial can be made to vanish after a simple horizontal

shift, which transforms the polynomial into the so-called *depressed* form.

We uncover certain properties of quartic polynomials, which are independent of the method we propose. Lemma 3 states that, if a quartic polynomial has two local minimizers, then its curvature at the global minimizer is greater. Moreover, its global minimizer is farther from the origin. Lemma 4, on the other hand, tells us at least how far from the origin the global minimizer will be located and what its sign is going to be. Lemma 7 presents a simple condition under which a monic depressed quartic polynomial is quasi-convex. Lemma 5 asserts the value t_0 that convexifies $\mu(\cdot, t_0)$ and states the minimizer x_0 of $\mu(\cdot, t_0)$, which are conveniently used in Algorithm 2. Proposition 4 provides a condition on t for quasi-convexity of $\mu(\cdot, t_0)$. Lemmas 8–11 provide some properties of the trajectories run in Algorithm 2 which in turn facilitate the proof of Theorem 3.

The paper is organized as follows. In Section 2, we introduce the Steklov regularization and prove certain properties, including convexification. In Section 3, we describe Algorithm 1 and prove the scale-shift invariance property. In Section 4, we derive some properties of quartic polynomials, provide Algorithm 2 and prove its convergence. In Section 5, we describe Algorithm 3, which uses the quadratic regularization. In Section 6, we carry out extensive numerical experiments using Algorithms 1 and 3 for polynomials of various degrees, including a non-polynomial example, and make comparisons.

2 Steklov Regularization

In an analogous way to the original (discrete) mean filter technique [16], first choose a "window" with centre x. In the case when x is a scalar, this window is just a finite interval. Then compute the *mean value* of a continuous function $f : \mathbb{R} \to \mathbb{R}$ over the window and assign this value as the value of an associated function at x. Furthermore, pass/shift the window across the whole domain of f, assigning values to the mean function at every x in the domain of f.

The window, or the interval, can typically be chosen to be centred at x, as $[x - t, x + t]$, where t is a fixed positive real number defining the window size. Therefore, we can regard the associated function as a function of not only x but also t.

The function we have just motivated with mean filter turns out to be already in use in the nonsmooth optimization literature, in obtaining smooth approximations of nondifferentiable objective functions, via a convolution integral, for t small enough. A well-known class of *mollifiers* in the convolution integral is referred to as the Steklov mollifiers [8]. A use of these mollifiers in the convolution integral in turn gives rise to the so-called *Steklov smoothing function*, definition and properties of which can be found in [6–9].

We note that the function we have motivated by means of mean filter is nothing but the Steklov smoothing function. Since our concern will be to *convexify* a given function for large enough t (rather than making it smooth for small t), we refer to the resulting function as the *Steklov regularization function*, or simply the *Steklov function*.

Definition 1 The *Steklov function* associated with a continuous function f is denoted by $\mu : \mathbb{R} \times (0, \infty) \to \mathbb{R}$ and defined as

$$\mu(x, t) := \frac{1}{2t} \int_{x-t}^{x+t} f(\tau)\, d\tau\,. \tag{1}$$

We also refer to $\mu(\cdot, \cdot)$ as the *Steklov regularization of* f. □

Remark 1 Since the function f is continuous, $\mu : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ is well defined and differentiable on $\mathbb{R} \times (0, \infty)$. □

We collect in the next lemma some useful properties of μ.

Lemma 1 *Given a continuous function $f : \mathbb{R} \to \mathbb{R}$, let $\mu : \mathbb{R} \times (0, \infty) \to \mathbb{R}$ be as in (1). The following equalities hold for μ.*

(i)

$$\mu_x(x,t) = \frac{1}{2t}\left(f(x+t) - f(x-t)\right), \tag{2}$$

where μ_x stands for $\partial\mu/\partial x$.

(ii)

$$\mu_{xx}(x,t) = \frac{1}{2t}\left(f'(x+t) - f'(x-t)\right), \tag{3}$$

where μ_{xx} stands for $\partial^2\mu/\partial x^2$.

(iii)

$$\mu_{tx}(x,t) = \frac{1}{t}\left[\frac{1}{2}\left(f'(x+t) + f'(x-t)\right) - \mu_x(x,t)\right], \tag{4}$$

where μ_{tx} stands for $\partial^2\mu/\partial t\,\partial x$.

Proof. Part (i) follows directly from the Fundamental Theorem of Calculus, and the remaining parts are obtained by differentiating μ_x with respect to x and t, respectively. □

The following theorem states general assumptions under which the Steklov function μ convexifies a coercive function f, and hence we regard the effect of μ as a regularization.

Theorem 1 (Convexification) *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that there exist two real numbers a and b, with $a < b$, for which the following conditions hold.*

(a) $f'(x) < 0$ for all $x \le a$ and $f'(x) > 0$ for all $x \ge b$.

(b) f' is strictly increasing and unbounded below on $(-\infty, a]$.

(c) f' is strictly increasing and unbounded above on $[b, \infty)$.

Then there exists $t_0 > 0$ such that $\mu(\cdot, t)$ is strictly convex for all $t \ge t_0$.

Proof. From part (a) and the fact that f' is continuous on $[a, b]$, there exist real numbers $\alpha < 0$ and $\beta > 0$ such that

$$\alpha := \min_{x\in[a,b]} f'(x) \le f'(a) < 0 < f'(b) \le \max_{x\in[a,b]} f'(x) =: \beta\,.$$

By parts (b) and (c), there exist $\widetilde{a} \le a$ and $\widetilde{b} \ge b$ such that $f'(x) < \alpha$ for all $x \le \widetilde{a}$, and $f'(x) > \beta$ for all $x \ge \widetilde{b}$. Let $t_0 \ge \widetilde{b} - \widetilde{a} > 0$. We will show that for every $t \ge t_0$ and every $x \in \mathbb{R}$, we have

$$f'(x+t) - f'(x-t) > 0. \tag{5}$$

By (3), this amounts to showing convexity of $\mu(\cdot, t)$ all $t \ge t_0$. Only the following cases are possible for the pair $x - t_0$, $x + t_0$.

(i) $x - t_0, x + t_0 \in (-\infty, \widetilde{a}]$.

(ii) $x - t_0, x + t_0 \in [\widetilde{b}, \infty)$.

(iii) $x - t_0 \in (-\infty, \widetilde{a}]$ and $x + t_0 \in (\widetilde{a}, \widetilde{b})$.

(iv) $x - t_0 \in (\widetilde{a}, \widetilde{b})$ and $x + t_0 \in [\widetilde{b}, \infty)$.

(v) $x - t_0 \in (-\infty, \widetilde{a}]$ and $x + t_0 \in [\widetilde{b}, \infty)$.

Note that the case $x - t_0$, $x + t_0 \in (\widetilde{a}, \widetilde{b})$ is not possible by the choice of t_0. Indeed, if $x - t_0$, $x + t_0 \in (\widetilde{a}, \widetilde{b})$ we can write

$$t_0 < x - \widetilde{a} \quad \text{and} \quad t_0 < \widetilde{b} - x\,,$$

so $t_0 < (\widetilde{b} - \widetilde{a})/2 < \widetilde{b} - \widetilde{a}$, contradicting the choice of t_0. We prove (5) by considering all the possible cases (i)–(v).

(i) By part (b) and the fact that $x + t_0 > x - t_0$, we have that $f'(x + t_0) - f'(x - t_0) > 0$. To complete the proof of (5), fix now $t > t_0$. We have the following sub-cases:

$$(\mathrm{i}_1)\ x + t \in (\widetilde{a}, a)\,, \qquad (\mathrm{i}_2)\ x + t \in [a, b]\,, \qquad (\mathrm{i}_3)\ x + t \in (b, +\infty)\,.$$

In case (i_1) we use part (b) and the fact that $x - t_0$, $x - t, x + t_0$, $x + t \in (-\infty, a]$ to write

$$\begin{aligned} f'(x + t_0) &< f'(x + t)\,, \\ f'(x - t_0) &> f'(x - t)\,, \end{aligned}$$

so $0 < f'(x + t_0) - f'(x - t_0) < f'(x + t) - f'(x - t)$, as desired. In case (i_2), we use the definition of α to write $f'(x + t) \geq \alpha$. Since $x + t_0 \in (-\infty, \widetilde{a}]$ we also have that $f'(x + t_0) < \alpha$. Using part (b) and the fact that $x - t < x - t_0 \leq \widetilde{a}$, we have

$$0 < f'(x + t_0) - f'(x - t_0) < \alpha - f'(x - t) \leq f'(x + t) - f'(x - t)\,,$$

as desired. In sub-case (i_3), we note that $f'(x + t) > \alpha$. Indeed, since $x + t > b$ we use part (c) to write $\alpha \leq f'(b) < f'(x + t)$. Altogether,

$$0 < f'(x + t_0) - f'(x - t_0) < \alpha - f'(x - t_0) < f'(x + t) - f'(x - t)\,,$$

where we also used (b) in the third inequality. This completes the proof for case (i). Due to symmetry, the proof for case (ii) is done in exactly the same way as for case (i), *mutatis mutandis*. We therefore omit the proof for case (ii).

(iii) As in (i), we consider three subcases:

$$(\mathrm{iii}_1)\ x + t_0 \in (\widetilde{a}, a)\,, \qquad (\mathrm{iii}_2)\ x + t_0 \in [a, b]\,, \qquad (\mathrm{iii}_3)\ x + t_0 \in (b, +\infty)\,.$$

Case (iii_1) implies that $x - t_0, x + t_0 \in (-\infty, a]$ and by part (b)

$$f'(x + t_0) - f'(x - t_0) > 0\,. \tag{6}$$

Case (iii_2) gives $x + t_0 \in [a, b]$ and $x - t_0 \in (-\infty, \widetilde{a}]$. So, again we have (6). Indeed,

$$f'(x + t_0) - f'(x - t_0) > \alpha - \alpha = 0\,.$$

In case (iii_3) we have $x + t_0 \in (b, \infty)$ and $x - t_0 \in (-\infty, \widetilde{a}]$. So by parts (b) and (c) we have that $\alpha \leq f'(b) < f'(x + t_0)$ and $f'(x - t_0) < \alpha$. As in case (iii_2) we obtain (6). To complete the proof for case (iii), fix $t \geq t_0$. As in case (i) we need to consider three sub-cases:

$$(\text{iii}_4)\ x+t \in (\widetilde{a}, a)\,, \qquad (\text{iii}_5)\ x+t \in [a, b]\,, \qquad (\text{iii}_6)\ x+t \in (b, +\infty)\,.$$

All three sub-cases are resolved exactly as in cases (iii_1), (iii_2) and (iii_3), respectively, with t_0 replaced by t. This completes the proof for case (iii).

(iv) Again, we consider three sub-cases:

$$(\text{iv}_1)\ x-t_0 \in (\widetilde{a}, a)\,, \qquad (\text{iv}_2)\ x-t_0 \in [a, b]\,, \qquad (\text{iv}_3)\ x-t_0 \in (b, \widetilde{b})\,.$$

In case (iv_1) we have $\widetilde{a} < x - t_0 < a$ so by part (b) $f'(x-t_0) < f'(a) \leq \beta$. Also in case (iv_2) we have $f'(x-t_0) \leq \beta$. In both cases, we can write

$$f'(x+t_0) - f'(x-t_0) > \beta - \beta = 0\,,$$

where we also used the fact that $f'(x+t_0) > \beta$. In case (iv_3), $x-t_0$, $x+t_0 \in (b, \infty)$ and we use directly part (c) to conclude that $f'(x+t_0) - f'(x-t_0) > 0$.
To complete the proof for case (iv), fix $t \geq t_0$. We always have that $x+t \in [\widetilde{b}, \infty)$ so $f'(x+t) > \beta$. We consider again the following sub-cases:

$$(\text{iv}_4)\ x-t \in (\widetilde{a}, a)\,, \qquad (\text{iv}_5)\ x-t \in [a, b]\,, \qquad (\text{iv}_6)\ x-t_0 \in (b, \widetilde{b})\,.$$

As in case (iii), all three sub-cases are resolved exactly as cases (iv_1), (iv_2) and (iv_3), respectively, with t_0 replaced by t. This completes the proof for case (iv).

(v) Use parts (b) and (c) to write

$$0 < \beta - \alpha < f'(x+t_0) - f'(x-t_0) < f'(x+t) - f'(x-t),$$

where we used the definition of $\widetilde{a}$ and $\widetilde{b}$ in the second inequality and parts (b) and (c) in the third. This completes the proof for case (v).

The proof of the theorem is complete. □

Remark 2 It is easy to check that, in Theorem 1, we can take $a := -R$ and $b := R$ for $R := \max\{|a|, |b|\}$. Note that $R > 0$ because $a < b$.

Remark 3 Monic polynomials of even degree are an important special case of functions which can be convexified by μ.

Graphical depictions of $\mu(x,t)$ for typical monic quartic polynomials and how convexification happens in each example case can be observed in Figure 1 on page **21** .

We focus our attention on functions which are *coercive*, in the sense of [12, Definition 3.25]. In our framework, this concept is stated as follows.

Definition 2 Let $f : \mathbb{R} \to \mathbb{R}$ be a function which is bounded below on bounded sets. We say that f is *coercive* if

$$\liminf_{|x|\to\infty} \frac{f(x)}{|x|} = \infty\,. \tag{7}$$

□

Coercive functions might be non-differentiable, and hence in general they may not verify the assumptions of Theorem 1. The following result shows that a function verifying the assumptions of Theorem 1 is coercive in the sense of Definition 2.

Proposition 1 (Coercivity) *Let f be as in Theorem 1. Then f is coercive in the sense of Definition 2.*

Proof. The statement on the boundedness of f is a direct consequence of the continuity of f. By Remark 2, we can assume that Theorem 1 holds with $a = -R$ and $b = R$ for some $R > 0$. To prove (7) we will show that for all $M > 0$ we have

$$\liminf_{|x|\to\infty} \frac{f(x)}{|x|} \geq M. \tag{8}$$

If (8) is not true, there exists $M_0 > 0$ and a sequence $(x_n) \subset \mathbb{R}$ such that $|x_n| > n$ for all $n \in \mathbb{N}$ and

$$\frac{f(x_n)}{|x_n|} < M_0. \tag{9}$$

Without loss of generality we can assume that the sequence $(x_n) \subset [R, \infty)$ and strictly monotone increasing, or $(x_n) \subset (-\infty, -R]$ and strictly monotone decreasing (otherwise we take a subsequence of the original sequence). Moreover, we can further assume that $(x_n) \subset [R, \infty)$ and strictly monotone increasing, because the proof for the latter case is identical to the one for the case in which $(x_n) \subset (-\infty, -R]$ and strictly monotone decreasing (*mutatis mutandis*). So it is enough to assume that $(x_n) \subset [R, \infty)$ and strictly monotone increasing. Since $x_n \uparrow +\infty$ and f' is strictly increasing and unbounded above in $[R, \infty)$ there exists n_0 such that $f'(x_n) > 2M_0$ for all $n \geq n_0$. Using the mean value theorem we can write for all $n > n_0$:

$$\begin{aligned} f(x_n) - f(x_{n_0}) &= \sum_{j=n_0}^{n-1} (f(x_{j+1}) - f(x_j)) = \sum_{j=n_0}^{n-1} f'(\theta_j)(x_{j+1} - x_j) \\ &> f'(x_{n_0}) \sum_{j=n_0}^{n-1} (x_{j+1} - x_j) > 2M_0(x_n - x_{n_0}), \end{aligned}$$

where we used that $R \leq x_{n_0} \leq x_j < \theta_j < x_{j+1}$ and the fact that f' is increasing in $[R, \infty)$ in the first inequality, and the definition of n_0 in the last one. Dividing the expression by $|x_n| = x_n$ and using (9) we obtain

$$M_0 - \frac{f(x_{n_0})}{x_n} > \frac{f(x_n)}{x_n} - \frac{f(x_{n_0})}{x_n} > 2M_0 \left(1 - \frac{x_{n_0}}{x_n}\right).$$

Taking limits for $n \to \infty$ and using the fact that $x_n \to +\infty$ we obtain

$$M_0 \geq 2M_0,$$

a contradiction. This completes the proof. □

The next proposition shows that $\mu(\cdot, t)$ is a good approximation of f at x for small values of t.

Proposition 2 (Limiting Functions) *Fix $x \in \mathbb{R}$ and $t_0 > 0$. Assume that f is twice continuously differentiable at x.*

$$\lim_{t\to 0} \mu(x,t) = f(x)\,, \quad \lim_{t\to 0} \mu_x(x,t) = f'(x)\,, \quad \lim_{t\to 0} \mu_{xx}(x,t) = f''(x)\,, \quad \lim_{t\to 0} \mu_{tx}(x,t) = 0\,. \tag{10}$$

Proof. The first limit is a consequence of l'Hôpital's rule:

$$\lim_{t\to 0} \mu(x,t) = \lim_{t\to 0} \frac{\displaystyle\int_{x-t}^{x+t} f(\tau)\, d\tau}{2\,t} = \lim_{t\to 0} \frac{f(x+t) + f(x-t)}{2} = f(x)\,.$$

The second and third limits are also a result of the application of l'Hôpital's rule on the limit, as $t \to 0$, of (2) and (3), respectively. Proving the last equality is more involved: Since f is $\mathcal{C}^2$ we can write

$$f(x+t) = f(x) + tf'(x) + o_1(t^2)\,,$$
$$f(x-t) = f(x) - tf'(x) + o_2(t^2)\,,$$

where $\lim_{t\to 0}(o_i(t^2)/t) = 0$, for $i = 1, 2$. Using the two equalities above we derive

$$\begin{aligned}\mu_x(x,t) &= \frac{1}{2t}\left(f(x+t) - f(x-t)\right) = \frac{1}{2t}\left(2tf'(x) + o_3(t^2)\right)\\ &= f'(x) + o(t)\,,\end{aligned} \tag{11}$$

where $\lim_{t\to 0}(o_i(t)/t) = 0$. Using (11) we have

$$\begin{aligned}\lim_{t\to 0}\mu_{tx}(x,t) &= \lim_{t\to 0}\frac{1}{2t}\left[f'(x+t) + f'(x-t)\right] - \frac{1}{t}\mu_x(x,t)\\ &= \lim_{t\to 0}\frac{1}{2t}\left[f'(x+t) + f'(x-t)\right] - \frac{1}{2t}f'(x) - \frac{1}{2t}f'(x) - \frac{o(t)}{t}\\ &= \lim_{t\to 0}\frac{1}{t}\left[\frac{f'(x+t) - f'(x)}{t} - \frac{[f'(x-t) - f'(x)]}{(-t)}\right] - \frac{o(t)}{t}\\ &= f''(x) - f''(x) = 0.\end{aligned}$$

□

3 A Trajectory Method Using Steklov regularization

The trajectory approach we formulate is based on constructing a continuously differentiable path through points where

$$\mu_x(x,t) = 0\,, \quad \forall t \in (0, t_0]\,. \tag{12}$$

We interpret the variable x as a function dependent on t, i.e., $x : [0, t_0] \to \mathbb{R}$, mapping $t \mapsto x(t)$. By taking the total derivative of both sides of (12) with respect to the independent variable t, we obtain

$$\mu_{xx}(x(t),t)\dot{x}(t) + \mu_{tx}(x(t),t) = 0\,, \text{ for a.e. } t \in (0, t_0]\,, \tag{13}$$

where $\dot{x}$ stands for dx/dt. In particular, we note that, for $(x_0, t_0) = (x(t_0), t_0)$, we have by (12) that $\mu_x(x_0, t_0) = 0$. After re-arranging (13), one obtains the initial value problem

$$\dot{x}(t) = -\frac{\mu_{tx}(x(t),t)}{\mu_{xx}(x(t),t)}\,, \quad \text{for a.e. } t \in (0, t_0]\,, \quad \text{with } x(t_0) = x_0\,, \tag{14}$$

provided that $\mu_{xx}(x(t), t) \neq 0$ a.e. in $(0, t_0]$.

Remark 4 Suppose that $x(\cdot)$ is a solution of the ODE in (14). Then Proposition 2 implies that, if $\lim_{t\to 0^+} f''(x(t)) \neq 0$, then $\lim_{t\to 0^+}\dot{x}(t) = 0$.

3.1 An algorithm for global optimization

We motivate our first method as follows. Assume that f is as in Theorem 1. Let $x_0 \in \mathbb{R}$ and $t_0 > 0$ be such that

$$f'(x+t_0) - f'(x-t_0) > 0\,, \quad \forall x \in \mathbb{R}\,, \quad \text{and} \quad f(x_0+t_0) - f(x_0-t_0) = 0\,.$$

From (2) and (3), the last two expressions imply that $\mu_{xx}(x,t_0) > 0$ and $\mu_x(x_0,t_0) = 0$, respectively. Using (12) and (3)–(4) in the IVP (14), we obtain

$$\dot{x}(t) = -\frac{f'(x(t)+t) + f'(x(t)-t)}{f'(x(t)+t) - f'(x(t)-t)}\,, \quad \text{for a.e. } t \in (0,t_0]\,, \quad \text{with } x(t_0) = x_0\,. \tag{15}$$

Algorithm 1 below serves to find a global minimizer of f.

Algorithm 1

Step 1 *Choose the parameter $t_0 > 0$ large enough so that $\mu(\cdot,t_0)$ is convex. Find the (global) minimizer x_0 of $\mu(\cdot,t_0)$, i.e., solve $f(x_0+t_0) - f(x_0-t_0) = 0$ for x_0.*

Step 2 *Solve the initial value problem in (15).*

Step 3 *Report $\lim_{t\to 0^+} x(t) =: x^*$ as a global minimizer of f.*

Algorithm 1 is said to be *well-defined* for the function f if there exist x_0 and $t_0 > 0$ such that Steps 1–3 of the algorithm can be carried out. This entails, in particular, that the solution of the IVP in Step 2 is obtained uniquely. Theorem 1 establishes assumptions on f under which Step 1 can be carried out.

In the following lemma, we show that Algorithm 1 is *scale-shift invariant*; i.e., if Algorithm 1 is well-defined for the function f, then it is also well-defined for any scale change and horizontal translation, of f.

Lemma 2 (Scale-Shift Invariance) *Fix $\alpha > 0$ and $a \in \mathbb{R}$. Assume that Algorithm 1 is well-defined for f, and let x_0 and t_0 be as in Step 1 for f. Let x^* be the global minimizer of f generated by Step 3 of Algorithm 1 for f. Set $g(x) := f(\alpha\, x - a)$ and denote the Steklov function associated with g by*

$$\widetilde{\mu}(x,t) := \frac{1}{2t}\int_{x-t}^{x+t} g(\tau)\, d\tau\,. \tag{16}$$

Then Algorithm 1, with μ replaced by $\widetilde{\mu}$ is well-defined for g and generates $z^* := \dfrac{x^*+a}{\alpha}$, which is a global minimizer of g. In this case, Step 1 can be carried out with $s_0 := t_0/\alpha$, and $z_0 := \dfrac{x_0+a}{\alpha}$.

Proof. Using the definition of g and (16) we can write

$$\begin{aligned} \widetilde{\mu}(x,t) \;&=\; \frac{1}{2t}\int_{x-t}^{x+t} f(\alpha\,\tau - a)\, d\tau \;=\; \frac{1}{2t}\int_{\alpha(x-t)-a}^{\alpha(x+t)-a} f(\eta)\, d\eta \\ &=\; \frac{\alpha}{2(\alpha\, t)}\int_{\alpha\, x-a-(\alpha t)}^{\alpha\, x-a+(\alpha t)} f(\eta)\, d\eta \;=\; \alpha\,\mu(\alpha\, x - a, \alpha\, t)\,, \end{aligned} \tag{17}$$

through a change of the dummy integration variable, $\eta = \alpha\,\tau - a$, and the definition in (1). Then, by taking partial derivatives of $\widetilde{\mu}$, where we employ the chain rule on the right-most term of the second line in (17), we get

$$\begin{aligned} &\widetilde{\mu}_x(x,t) = \alpha^2 \mu_x(\alpha\,x - a, \alpha\,t), \quad \widetilde{\mu}_{xx}(x,t) = \alpha^3 \mu_{xx}(\alpha\,x - a, \alpha\,t) \\ &\widetilde{\mu}_{tx}(x,t) = \alpha^3 \mu_{tx}(\alpha\,x - a, \alpha\,t)\,. \end{aligned} \tag{18}$$

Since Algorithm 1 is well-defined for f, Steps 1 and 2 of the algorithm can be executed, generating a global minimizer x_0 of $\mu(\cdot, t_0)$ in Step 1, a unique solution $x(\cdot)$ to the IVP in (14) in Step 2, where $\mu(\cdot, t_0)$ is convex and

$$\mu_x(x_0, t_0) = 0\,. \tag{19}$$

In Step 3, a global minimizer of f is obtained as $\lim_{t\to 0^+} x(t) = x^*$. Take $z_0 := (x_0 + a)/\alpha$ and $s_0 := t_0/\alpha$. We show now that Step 1 is well defined for g, for s_0 and z_0 in place of t_0 and x_0, respectively. Indeed, by Step 1 for f we know that $\mu(\cdot, t_0)$ is convex. Hence, the composition of $\mu(\cdot, t_0)$ with the linear function $L(x) = \alpha\,x - a$ is also convex. Namely, the function $\mu(\alpha\,(\cdot) - a, t_0)$ is convex, and hence any positive multiple of it is convex. Therefore, by (17) we deduce that $\widetilde{\mu}(\cdot, t_0)$ is convex. The first equality in (18), combined with (19) and the definitions of z_0 and t_0 give

$$0 = \mu_x(x_0, t_0) = \mu_x(\alpha z_0 - a, \alpha\,s_0) = \frac{1}{\alpha^2}\,\widetilde{\mu}_x(z_0, s_0)\,,$$

so that z_0 is a global minimizer of $\widetilde{\mu}(\cdot, s_0)$. This shows that Step 1 is well defined for g. We proceed now to show that Step 2 is well defined for g. Take $x(\cdot)$ to be the unique solution of the IVP in (14) obtained in Step 2 for f, and define $z(t) := (x(\alpha\,t) + a)/\alpha$, for all $t \in (0, s_0]$. We claim that $z(\cdot)$ solves the following IVP:

$$\dot{z}(t) = -\frac{\widetilde{\mu}_{tx}(z(t), t)}{\widetilde{\mu}_{xx}(z(t), t)}\,, \quad \text{for a.e. } t \in (0, s_0]\,, \quad z(s_0) = z_0\,, \tag{20}$$

which is the IVP in (14) with μ replaced by $\widetilde{\mu}$ and x_0 replaced by z_0. Indeed, take $t \in (0, s_0] = (0, t_0/\alpha]$. Then $\alpha\,t \in (0, t_0]$ and by (14) we can write

$$\dot{z}(t) = \dot{x}(\alpha\,t) = -\frac{\mu_{tx}(x(\alpha\,t), \alpha\,t)}{\mu_{xx}(x(\alpha\,t), \alpha\,t)} = -\frac{\mu_{tx}(\alpha\,z(t) - a, \alpha\,t)}{\mu_{xx}(\alpha\,z(t) - a, \alpha\,t)} = -\frac{\widetilde{\mu}_{tx}(z(t), t)}{\widetilde{\mu}_{xx}(z(t), t)},$$

where we have used the definition of z in the first equality, the definition of x as solution of (14) in the second equality. We have used the second and third equalities of (18) in the third equality above. The fact that $z(s_0) = z_0$ follows directly from the definition of z and the fact that $x(t_0) = x_0$. Therefore, z solves (20) and hence Step 2 is well defined for g. To check that the same holds for Step 3, take $x^* = \lim_{t\to 0^+} x(t)$ to be the global minimizer of f generated in Step 3 for f. The solution of the IVP in (20) will now result in

$$\lim_{t\to 0^+} z(t) = \lim_{t\to 0^+} \frac{x(\alpha\,t) + a}{\alpha} = \frac{x^* + a}{\alpha}\,.$$

Since x^* is a global minimizer of f, we have

$$g\left(\frac{x^* + a}{\alpha}\right) = f(x^*) \le f(\alpha\,x - a) = g(x), \quad ,\forall\, x \in \mathbb{R}\,,$$

so $(x^* + a)/\alpha$ is a global minimizer of g. Hence, Step 3 is well defined for g and the proof is complete. □

4 Quartic Polynomials

In this section, we consider the special case of monic *depressed* quartic polynomials, namely,

$$f(x) = x^4 + a_2\,x^2 + a_1\,x + a_0\,, \tag{21}$$

where a_0, a_1 and a_2 are real constants such that $a_2 < 0$ and $a_1 \neq 0$. Note that the depressed form is general enough. Indeed, given an arbitrary quartic polynomial, $g(y) = y^4 + b_3\,y^3 + b_2\,y^2 + b_1\,y + b_0$, the substitution $y = x - (b_3/4)$ reduces g to a depressed form. As for the assumption $a_2 < 0$, note that, if $a_2 \geq 0$ then $f''(x) = 12\,x^2 + 2\,a_2 \geq 0$ which yields that $f(\cdot)$ is convex. In this case, there is no need to apply Algorithm 1 to find a global minimum of $f(\cdot)$. Assumption $a_1 \neq 0$ is posed since if $a_1 = 0$ then $f(\cdot)$ has two global minimizers simply given by the set $\{-\sqrt{-a_2/2}, \sqrt{-a_2/2}\}$. Hence, the non-trivial case is when $a_2 < 0$ and $a_1 \neq 0$.

4.1 Properties of monic quartic polynomials

A quartic polynomial can have at most two local minima. The following lemma helps distinguish which of these two is the global minimum.

Lemma 3 (Curvature) *Let f be a monic quartic polynomial. Assume that $f'(x_1) = f'(x_2) = 0$ with $x_1 \neq x_2$. The following properties hold.*

(i) $f(x_1) < f(x_2)$ if, and only if, $f''(x_1) > f''(x_2)$, in particular, $|x_1| > |x_2|$.

(ii) $f(x_1) = f(x_2)$ if, and only if, $f''(x_1) = f''(x_2)$.

Proof. The proof of parts (i) and (ii) is done in two steps.
Step 1: In this step, we show that it is enough to prove the lemma for a depressed quartic polynomial. We prove the claim for part (i). The claim for part (ii) is proved in an identical way. Assume that part (i) of the lemma is true for depressed monic quartic polynomials and that we have a quartic polynomial $h(x) = x^4 + c_3x^3 + c_2x^2 + c_1x + c_0$ with $c_3 \neq 0$. Assume that $h'(x_1) = h'(x_2) = 0$ As noted above, the "shifted" polynomial $f(x) := h(x - (c_3/4))$ is (monic and) depressed. Using the chain rule we have

$$h'(x_1) = f'(x_1 + (c_3/4)) = h'(x_2) = f'(x_2 + (c_3/4)) = 0\,.$$

Since part (i) of the lemma is true for f we have

$$h(x_1) = f(x_1 + (c_3/4)) < f(x_2 + (c_3/4)) = h(x_2)$$

if and only if

$$h''(x_1) = f''(x_1 + (c_3/4)) > f''(x_2 + (c_3/4)) = h''(x_2)\,,$$

and hence part (i) of the lemma holds for h. As mentioned before, the proof of the fact that part (ii) of the lemma holds for h follows identical steps. Therefore, it is enough to prove the lemma for depressed quartic polynomials.
Step 2: In this step, we show that, if f is a depressed quartic polynomial such that $f'(x) = f'(y) = 0$, with $x \neq y$, then we have

$$12\,\frac{[f(x) - f(y)]}{(x-y)^2} = (f''(y) - f''(x)). \tag{22}$$

Note that parts (i) and (ii) of the lemma follow directly from (22). This is straightforward for part (ii). As for part (i), if (22) holds, the assumption on x_1 and x_2 implies that for $x := x_1$ and $y := x_2$ we have

$$\operatorname{sgn}\left[f(x_1) - f(x_2)\right] = \operatorname{sgn}\left[f''(x_2) - f''(x_1)\right] = -\operatorname{sgn}\left[f''(x_1) - f''(x_2)\right],$$

which is the statement of part (i) of the lemma, also observing that $f''(x_1) > f''(x_2)$ if and only if $12x_1^2 + 2a_2 > 12x_2^2 + 2a_2$, i.e., $|x_1| > |x_2|$. Hence, we proceed to prove (22) when $f'(x) = f'(y) = 0$ and f is a depressed quartic polynomial. The assumption on x and y and the Taylor development of f gives

$$\begin{aligned} f(x) - f(y) &= [f''(y)/2](x-y)^2 + [f'''(y)/6](x-y)^3 + (x-y)^4, \\ f(y) - f(x) &= [f''(x)/2](y-x)^2 + [f'''(x)/6](y-x)^3 + (y-x)^4, \end{aligned}$$

By subtracting side-by-side the second equality from the first one, and re-arranging the resulting expression we obtain

$$2\frac{[f(x) - f(y)]}{(x-y)^2} = \frac{[f''(y) - f''(x)]}{2} + \frac{[f'''(y) + f'''(x)]}{6}(x-y). \tag{23}$$

By direct calculation, the rightmost term in (23) can be written as follows

$$\begin{aligned} \frac{[f'''(y) + f'''(x)]}{6}(x-y) &= \frac{[24y + 24x]}{6}(x-y) = 4(x^2 - y^2) \\ &= \left[(12x^2 + 2a_2) - (12y^2 + 2a_2)\right]/3 \\ &= [f''(x) - f''(y)]/3 \\ &= -[f''(y) - f''(x)]/3\,. \end{aligned}$$

Using this in (23) yields

$$2\,\frac{[f(x) - f(y)]}{(x-y)^2} = \frac{[f''(y) - f''(x)]}{2} - \frac{[f''(y) - f''(x)]}{3} = \frac{[f''(y) - f''(x)]}{6}, \tag{24}$$

which is (22). The proof is complete. □

Remark 5 The previous lemma is not valid for higher degree polynomials. The function $f(x) = x^6 - \frac{8}{5}x^5 + \frac{2}{3}x^3$, with the local extrema $x_1 = 0$ and $x_2 = 1$, furnishes a counterexample.

Lemma 4 (Sign of a Minimizer) *Consider a monic depressed quartic polynomial f, with $a_1 \neq 0$ and $a_2 < 0$. If x_1 and x_2 are the local minimizers of f, then* $\operatorname{sgn}(x_1) = -\operatorname{sgn}(x_2)$. *Suppose that x^* is the global minimizer of f. Then $|x^*| > \sqrt{-a_2/6}$ and, in particular,* $\operatorname{sgn}(x^*) = -\operatorname{sgn}(a_1)$.

Proof. Suppose that x_1 and x_2 are the local minimizers of f. Note that $f''(x) = 12\,x^2 + 2\,a_2$ is an even function, i.e., $f''(-x) = f''(x)$. Since $a_2 < 0$, we have $f''(x) = 0$ when $x = \sqrt{-a_2/6} =: \widetilde{x}$ and $x = -\widetilde{x}$. Then, since $f''(x) > 0$ for $x < -\widetilde{x}$ and $x > \widetilde{x}$, one of the local minima is placed to the left of $-\widetilde{x}$ and the other to the right of $\widetilde{x}$, i.e., $x_1 < -\widetilde{x} < 0$ and $x_2 > \widetilde{x} > 0$. So $\operatorname{sgn}(x_1) = -\operatorname{sgn}(x_2)$ and $|x^*| > \sqrt{-a_2/6}$. Now, we can write

$$\begin{aligned} f'(x_1) &= f'(-\widetilde{x}) + \int_{-\widetilde{x}}^{x_1} f''(y)\,dy \\ &= f'(-\widetilde{x}) - \int_{\widetilde{x}}^{-x_1} f''(y)\,dy = 0 \end{aligned} \tag{25}$$

where, in (25), a change of variables and the fact that f'' is even have been used. We can also write

$$f'(x_2) = f'(\widetilde{x}) + \int_{\widetilde{x}}^{x_2} f''(y)\,dy = 0\,. \tag{26}$$

Adding Equations (25)-(26) side by side and using $f'(\widetilde{x}) + f'(-\widetilde{x}) = 2\,a_1$, one gets

$$2\,a_1 + \int_{\widetilde{x}}^{x_2} f''(y)\,dy - \int_{\widetilde{x}}^{-x_1} f''(y)\,dy = 0\,. \tag{27}$$

To complete the proof, we consider two cases: $a_1 < 0$ and $a_1 > 0$. If $a_1 < 0$, from (27) we have,

$$-2\,a_1 = \int_{\widetilde{x}}^{x_2} f''(y)\,dy - \int_{\widetilde{x}}^{-x_1} f''(y)\,dy > 0\,,$$

which implies that $x_2 > -x_1 > 0$, since $f''(y) > 0$ over both integration intervals. Since $f'''(x) = 24x > 0$ for $x > 0$, f'' is increasing in $(0, \infty)$ so $f''(x_2) > f''(-x_1) = f''(x_1)$. Then, by Lemma 3, $f(x_2) < f(x_1)$ and hence $x^* = x_2 > 0$ is the global minimizer. Therefore, $\operatorname{sgn}(x^*) = 1 = -\operatorname{sgn}(a_1)$.

Suppose now that $a_1 > 0$. Through similar steps, we get $-x_1 > x_2 > 0$, or $x_1 < -x_2 < 0$. Since $f'''(x) = 24x < 0$ for $x < 0$, f'' is decreasing in $(-\infty, 0)$ so $f''(x_1) > f''(-x_2) = f''(x_2)$. Then, by Lemma 3, $f(x_1) < f(x_2)$ and hence $x^* = x_1 < 0$ is the global minimizer. Therefore, $\operatorname{sgn}(x^*) = -1 = -\operatorname{sgn}(a_1)$, completing the proof. □

4.2 An algorithm for global minimization of quartic polynomials

In this section we consider the specific case of applying Algorithm 1 to quartic polynomials.

Proposition 3 *If $f(x)$ is a univariate quartic monic polynomial, then $\mu(x,t)$, as in (1), can be written as*

$$\mu(x,t) = f(x) + \frac{t^2}{6}\,f''(x) + \frac{t^4}{5}\,. \tag{28}$$

Proof. Let $f(x) := x^4 + a_3\,x^3 + a_2\,x^2 + a_1\,x + a_0$, where a_0, a_1, a_2 and a_3 are real numbers. Substitution of f into (1), followed by straightforward integration, expanding and rearranging, yield (28). □

Using (28), $\mu(x,t)$ and its derivatives can now be re-written for monic depressed polynomials as follows.

$$\mu(x,t) = f(x) + \frac{t^2}{6}\,f''(x) + \frac{t^4}{5} = x^4 + (a_2 + 2\,t^2)\,x^2 + a_1\,x + a_0 + \frac{t^2}{3} + \frac{t^4}{5}\,, \tag{29}$$

$$\mu_x(x,t) = f'(x) + \frac{t^2}{6}\,f'''(x) = 4\,x^3 + 2\,(a_2 + 2\,t^2)\,x + a_1\,, \tag{30}$$

$$\mu_{xx}(x,t) = f''(x) + 4\,t^2 = 12\,x^2 + 2\,(a_2 + 2\,t^2)\,, \tag{31}$$

$$\mu_{tx}(x,t) = \frac{t}{3}\,f'''(x) = 8\,t\,x\,. \tag{32}$$

In Step 1 of Algorithm 1, we need to find (i) some $t_0 > 0$ such that $\mu(\cdot, t_0)$ is convex, and (ii) the point x_0 which is the global minimizer of $\mu(\cdot, t_0)$. The next lemma finds these for the particular case of quartic polynomials.

Lemma 5 (Convexification of Quartic Polynomials) *Given any monic depressed quartic polynomial $f(x)$ with $a_1 \neq 0$ and $a_2 < 0$, $\mu(\cdot, t_0)$ is convex if*

$$t_0 := \sqrt{-a_2/2}\,.$$

The unique (global) minimizer of $\mu(\cdot, t_0)$ is

$$x_0 := -\sqrt[3]{a_1/4}\,.$$

Proof.

For convexity of $\mu(\cdot, t_0)$, we need to have

$$\mu_{xx}(x, t_0) = 12\,x^2 + 2\,(a_2 + 2\,t_0^2) \geq 0\,, \quad \forall x \in \mathbb{R},$$

which immediately follows if $t_0 = \sqrt{-a_2/2}$. The (global) minimizer x_0 of $\mu(\cdot, t_0)$ would then be found by solving

$$\mu_x(x_0, t_0) = 4\,x_0^3 + 2\,(a_2 + 2\,t_0^2)\,x_0 + a_1 = 0\,, \tag{33}$$

or, with $t_0 = \sqrt{-a_2/2}$,

$$4\,x_0^3 + a_1 = 0\,,$$

for x_0. Since $a_1 \neq 0$, we have $x_0 \neq 0$. This implies that $\mu_{xx}(x_0, t_0) > 0$, and hence $\mu(\cdot, t_0)$ is strictly convex around x_0, which implies that x_0 is the unique minimizer. This completes the proof. □

For the special case of monic depressed quartic polynomials, Algorithm 1 reduces to the following, using Lemma 5, (31) and (32).

Algorithm 2

Step 1 *Given the quartic polynomial in (21), let $t_0 = \sqrt{-a_2/2}$ and $x_0 = -\sqrt[3]{a_1/4}\,$.*

Step 2 *Solve the initial value problem*

$$\dot{x}(t) = -\frac{4\,t\,x(t)}{6\,x^2(t) + 2\,t^2 + a_2}\,, \quad \text{for a.e. } t \in [0, t_0]\,, \quad \text{with } x(t_0) = x_0\,. \tag{34}$$

Step 3 *Report $x(0)$ as the global minimizer of $f(x)$.*

By (12) and (14), IVP (34) can be derived under the assumption that $\mu_x(x(t), t) = 0$ and $\mu_{xx}(x(t), t) \neq 0$ for a.e. $t \in [0, t_0]$. Hence, it is worth investigating if, at all, the denominator of the right-hand side of the ODE in (34) vanishes, i.e., $\mu_{xx}(x(t), t) = 0$ for some $t \in [0, t_0]$. Since a solution $x(t)$ of the ODE in (34) satisfies $\mu_x(x(t), t) = 0$, the pathological situation happens at points (x, t) which satisfy the equations $\mu_x(x, t) = 0$ and $\mu_{xx}(x, t) = 0$ simultaneously. We investigate this situation in the following lemma.

Lemma 6 (Flatness) *Let f be a monic depressed quartic polynomial. Consider solutions $(x,t) \in \mathbb{R} \times [0,\infty)$ of the system*

$$\mu_x(x,t) = 0 \qquad \text{and} \qquad \mu_{xx}(x,t) = 0\,. \tag{35}$$

(a) If $a_2 > -3\,a_1^{2/3}/2$ then the system in (35) has no solution.

(b) If $a_2 \leq -3\,a_1^{2/3}/2$ then the system in (35) has a unique solution $(\widehat{x},\widehat{t}) \in \mathbb{R} \times [0,\infty)$ such that

$$\widehat{x} \;:=\; \frac{1}{2}\,\sqrt[3]{a_1}\,, \tag{36}$$

$$\widehat{t} \;:=\; \frac{1}{2}\,\sqrt{-\left(3\,a_1^{2/3} + 2\,a_2\right)}\,. \tag{37}$$

Proof. Start with

$$\mu_x(\widehat{x},\widehat{t}) \;=\; 4\,\widehat{x}^3 + 2\,(a_2 + 2\,\widehat{t}^2)\,\widehat{x} + a_1 = 0\,, \tag{38}$$

$$\mu_{xx}(\widehat{x},\widehat{t}) \;=\; 12\,\widehat{x}^2 + 2\,(a_2 + 2\,\widehat{t}^2) = 0\,. \tag{39}$$

Using (39) in (38) gives $\widehat{x} = \frac{1}{2}\,\sqrt[3]{a_1}$. Using this value of $\widehat{x}$ in (39) gives

$$0 = 6\,\left(\frac{1}{2}\,\sqrt[3]{a_1}\right)^2 + (a_2 + 2\,\widehat{t}^2) = \frac{3}{2}a_1^{2/3} + a_2 + 2\,\widehat{t}^2. \tag{40}$$

To prove part (a), note that $a_2 > -3\,a_1^{2/3}/2$ if and only if

$$0 = \frac{3}{2}a_1^{2/3} + a_2 + 2\,\widehat{t}^2 > 2\,\widehat{t}^2,$$

which entails a contradiction and therefore implies that the system has no solution. This proves (a). On the other hand, $a_2 \leq -3\,a_1^{2/3}/2$ if and only if there is a unique nonnegative solution of (40), given by (37). This proves part (b). □

We will consider in our analysis a notion which is weaker than convexity, called *quasi-convexity*.

Definition 3 The function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *quasi-convex* when all its level sets are convex, i.e., when for every $\alpha \in \mathbb{R}$ we have that the set $\{x \in \mathbb{R}^n \;:\; f(x) \leq \alpha\}$ is convex.

Definition 4 Let I be a (possibly infinite) interval in $\mathbb{R}$. Recall that $f : \mathbb{R} \to \mathbb{R}$ is *non-increasing in I* if for all $x, y \in I$ such that $x < y$ we have $f(x) \geq f(y)$. Similarly, f is *non-decreasing in I* if for all $x, y \in I$ such that $x < y$ we have $f(x) \leq f(y)$.

The following result is a trivial re-statement of Theorem 4.9.11 in [15].

Theorem 2 (Quasi-convexity) *A function $f : \mathbb{R} \to \mathbb{R}$ is quasi-convex if and only if there exists $m \in [-\infty, +\infty]$ such that f is non-increasing in $(-\infty, m]$ and non-decreasing in $[m, +\infty)$.*

Remark 6 When m is infinite in Theorem 2, one of the intervals is empty and this covers the case in which the function is everywhere non-increasing or everywhere non-decreasing. Functions as in the statement of Theorem 2 are sometimes called *unimodal*.

Lemma 7 (Quasi-convexity of a Quartic Polynomial) *Let h be a monic depressed quartic polynomial given by $h(x) := x^4 + b_2x^2 + b_1x + b_0$. Consider $\Delta := -16\,[8\,b_2^3 + 27\,b_1^2]$, i.e., Δ is the* discriminant *of $h'(x) = 4\,x^3 + 2\,b_2\,x + b_1$. Then h is quasi-convex if and only if $\Delta \leq 0$. In this situation, we have $b_2 \geq -3\,b_1^{2/3}/2$.*

Proof. From algebra of cubic equations we have that

$$\begin{aligned}
&(I) \quad \Delta > 0 \Longrightarrow h' \text{ has three distinct real roots,}\\
&(II) \quad \Delta = 0 \Longrightarrow h' \text{ has a multiple root and all its roots are real,}\\
&(III) \quad \Delta < 0 \Longrightarrow h' \text{ has one real root and two non-real complex conjugate roots,}
\end{aligned}$$

Case (I) implies that h has a local maximum, and two local minima. This cannot hold for a quasi-convex function in view of Theorem 2. Hence it is enough to show that the other two cases imply that h satisfies the unimodality property described in Theorem 2. This is clear in case (III), since by coercivity the unique real root of h' must be a global minimum. So in case (III) h is quasi-convex by Theorem 2. In Case (II), we have two possibilities: either all three roots coincide (i.e., we have a triple root of h') or one of the real roots is double. The case in which we have a triple root of h' implies that $h'(x) = 4(x - x_0)^3$ and hence is strictly increasing. So h is convex, and hence quasi-convex. We are left only with the case of a double real root x_0 and a simple real root x_1. In this case, $h'(x) = 4(x-x_1)(x-x_0)^2$ and it is clear that $h'(x) \leq 0$ for $x \leq x_1$ and $h'(x) \geq 0$ for $x \geq x_1$. This implies directly (using mean value theorem) that h verifies the unimodality property given in Theorem 2 with $m = x_1$, and hence h is quasi-convex. We have shown that $\Delta \leq 0$ implies h quasi-convex. Conversely, assume that h is quasi-convex, and let m be as in Theorem 2. Since h is coercive, we must have $m \in \mathbb{R}$. Because h is a polynomial, h cannot be constant in any interval, so m must be the only global minimum of h. This implies that h' cannot have three different roots, so we cannot be in Case (I) and hence we must have $\Delta \leq 0$. The last statement of the lemma follows directly from the expression of the discriminant. □

Proposition 4 *Suppose that f is a monic depressed quartic polynomial. Then $\mu(\cdot, t)$ is quasi-convex if, and only if,*

$$t \geq \frac{1}{2}\sqrt{\max\left\{0, -\left(3\,a_1^{2/3} + 2\,a_2\right)\right\}}\,. \tag{41}$$

Proof. By (29), $\mu(\cdot, t)$ is a monic quartic depressed polynomial with $\mu_x(x, t) = 4\,x^3 + 2\,(a_2 + 2\,t^2)\,x + a_1$. By Lemma 7 applied to $\mu(\cdot, t)$, we have that $\mu(\cdot, t)$ is quasi-convex if, and only if, the discriminant of $\mu_x(\cdot, t)$ is non-positive, i.e.,

$$\Delta = -16\left[8\,(a_2 + 2\,t^2)^3 + 27\,a_1^2\right] \leq 0\,,$$

a re-arrangement of which yields (41). □

Remark 7 Lemmas 6 and 7 imply that the right-hand-side of the ODE in (34) cannot be discontinuous in the interior of $[0, t_0]$ when f is a quasi-convex quartic polynomial. Indeed, assume that the right-hand-side of the ODE in (34) is discontinuous in the interior of $[0, t_0]$ and f is quasi-convex. By Lemma 6, this implies that the system (36)–(37) has a solution in $\mathbb{R} \times (0, t_0)$. By part (b) of the lemma this yields $(3\,a_1^{2/3} + 2\,a_2) \leq 0$. On the other hand, by Lemma 7, f is quasi-convex if and only if $(3\,a_1^{2/3} + 2\,a_2) \geq 0$. This yields $(3\,a_1^{2/3} + 2\,a_2) = 0$. Using (37) gives $\widehat{t} = 0$ as the unique solution of the system. Since the nonnegative solution is unique, this implies that there is no solution for t in the interior of $(0, t_0)$, and hence the denominator in the right-hand side of the ODE in (34) cannot vanish for $t \in (0, t_0)$.

4.3 Well-definedness of Algorithm 2

Lemma 8 (Solutions of ODEs) *Consider a monic depressed quartic polynomial f, with $a_2 < 0$ and $a_1 \neq 0$. Let $t_0 = \sqrt{-a_2/2}$ and $x_0 = -\sqrt[3]{a_1/4}$. The following hold.*

(a) There exists $r > 0$ such that there is a unique solution $x(\cdot)$ of (34) in $(t_0 - r, t_0 + r)$.

(b) There exists a maximal interval to the left of t_0, say $(m_0, t_0]$, such that there exists a solution of (34) in $(m_0, t_0]$.

(c) Either $m_0 = -\infty$, or $m_0 \in \mathbb{R}$ and in this case we must have $\mu_{xx}(x(m_0), m_0) = 0$.

Proof. Note that $x_0^2 > 0$ because $a_1 \neq 0$. Part (a) follows from the classical Picard-Lindelöf existence and uniqueness theorem (see [2]), since the denominator $6\,x^2(t_0) + 2\,t_0^2 + a_2 = 6\,x_0^2 > 0$. By (31), this implies that $\mu_{xx}(x(t_0), t_0) > 0$, and so the right-hand side of the ODE in (34) is Lipschitz continuous in x and continuous in t in a neighbourhood of t_0. Part (b) is the classical result on maximal extension of solutions of ODEs. The option $m_0 = -\infty$ of part (c) corresponds to the case in which the right-hand side remains Lipschitz continuous in x for all $t < t_0$. The remaining option happens when the denominator

$$q(t) := 6\,x^2(t) + 2\,t^2 + a_2 \tag{42}$$

vanishes at $t = m_0$, i.e., when $\mu_{xx}(x(m_0), m_0) = 0$. This completes the proof. □

The following lemma re-formulates the initial value problem in (34).

Lemma 9 (Trajectory Along a Valley) *Consider a monic depressed quartic polynomial, with $a_2 < 0$ and $a_1 \neq 0$. Let $t_0 = \sqrt{-a_2/2}$ and $x_0 = -\sqrt[3]{a_1/4}$. With the notation of Lemma 8, let $x(\cdot)$ be the maximally extended solution of (34), and $(m_0, t_0]$ the corresponding maximal interval. Then, we have that*

$$\mu_x(x(t), t) = 0\,, \quad \mu_{xx}(x(t), t) > 0\,, \ \forall t \in [m_0, t_0]\,. \tag{43}$$

Proof. We show first that $\mu_{xx}(x(t), t) > 0\,,\ \forall t \in [m_0, t_0]\,.$ Indeed, the choices of x_0 and t_0, together with (31) give $\mu_{xx}(x(t_0), t_0) > 0$. The definition of m_0 states that IVP (34) is solvable over $(m_0, t_0]$. By Lemma 8(a), this implies that the right-hand side of the ODE is continuous on $(m_0, t_0]$. In other words, the denominator of the right-hand side of the ODE is not zero and so it does not change sign on $(m_0, t_0]$. This readily gives

$$\mu_{xx}(x(t), t) > 0\,, \tag{44}$$

for all $t \in (m_0, t_0]$, as wanted. To complete the proof, recall that the ODE in (34) is the ODE in (14) written for a quartic polynomial. Then, for all $t \in (m_0, t_0]$, the ODE in (14) is equal to the expression in (13), which is

$$\dot{x}(t)\,\mu_{xx}(x(t), t) + \mu_{tx}(x(t), t) = 0\,.$$

The above expression can in turn be expressed as

$$\frac{d}{dt}\,\mu_x(x(t), t) = 0\,. \tag{45}$$

From the first step of Algorithm 1,

$$\mu_x(x(t_0), t_0) = 0\,. \tag{46}$$

Equalities (45) and (46) imply that

$$\mu_x(x(t), t) = 0\,, \tag{47}$$

for all $t \in (m_0, t_0]$. Equality (47) holds at $t = m_0$ by continuity of μ_x and $x(\cdot)$. This completes the proof of the lemma. □

Our next step is to show that the solution $x(\cdot)$ of the initial value problem (34) has the same sign as that of x_0 over its maximal domain of definition. For proving this, we need the following auxiliary result.

Lemma 10 *Fix $a \in \mathbb{R} \cup \{-\infty\}$ and let $b > a$. Assume that the function $x(\cdot) : (a, b] \to \mathbb{R}$ is continuously differentiable and satisfies*

$$\operatorname{sgn}(\dot{x}(t)) = -\operatorname{sgn}(x(t)), \quad \forall\, t \in (a, b]. \tag{48}$$

Assume that $x(b) \neq 0$. Then, $\operatorname{sgn}(x(t)) = \operatorname{sgn}(x(b))$ for all $t \in (a, b]$. If $a \in \mathbb{R}$, then $\operatorname{sgn}(x(a)) = \operatorname{sgn}(x(b))$.

Proof. Without loss of generality, assume that $x(b) < 0$. The case $x(b) > 0$ is handled similarly, *mutatis mutandis*. We show first that $x(t) < 0$ for all $t \in (a, b]$. Suppose that, on the contrary, there exists $\tilde{t} \in (a, b]$ such that $x(\tilde{t}\,) \geq 0$. This implies that the set

$$S := \{t \in (a, b] \ : \ x(t) \geq 0\},$$

is not empty. Since S is bounded above by b, there exists $t_1 := \sup(S)$. Since $x(\cdot)$ is continuous, we have that $t_1 \in S$, or, equivalently, $x(t_1) \geq 0$. In particular, $t_1 < b$ because $x(b) < 0$ and $x(t_1) \geq 0$. The definition of t_1 implies that $x(t) < 0$ for all $t \in (t_1, b]$. Using the continuity of x we deduce that $x(t_1) \leq 0$. Altogether, we must have $x(t_1) = 0$. The Mean Value Theorem gives, for some $\theta \in (t_1, b)$:

$$0 = x(t_1) = x(b) + \dot{x}(\theta)(t_1 - b) < x(b) < 0,$$

where we used (48) for $t = \theta$ in the first inequality, and the fact that $x(\theta) < 0$. The above expression entails a contradiction and hence we must have S empty. This completes the proof of the first statement. Assume now that $a \in \mathbb{R}$. The proof of the first statement implies that $x(t) < 0$ for all $t \in (a, b]$. By continuity we deduce that $x(a) \leq 0$. We need to prove that $x(a) < 0$. Assume that, on the contrary, $x(a) = 0$. Using the Mean Value Theorem gives, for some $s \in (a, b)$:

$$0 > x(b) = x(b) - x(a) = \dot{x}(s)(b - a) > 0,$$

where we used (48) for $t = s$ and the fact that $x(s) < 0$, so $\dot{x}(s) > 0$. The above expression entails a contradiction and hence we must have $x(a) < 0$. □

Lemma 11 (Sign of a Trajectory) *Consider a monic depressed quartic polynomial, with $a_2 < 0$ and $a_1 \neq 0$. Let $t_0 = \sqrt{-a_2/2}$ and $x_0 = -\sqrt[3]{a_1/4}$. Consider the initial value problem (34). Let $x(\cdot)$ be the maximally extended solution of (34), and $(m_0, t_0]$ the corresponding maximal interval of definition of $x(\cdot)$. Then $m_0 = -\infty$ and $\operatorname{sgn}(x(t)) = -\operatorname{sgn}(a_1)$ for all $t \in (-\infty, t_0]$.*

Proof. Suppose that $a_1 > 0$, and hence $x_0 < 0$. The definition of m_0 indicates that the right-hand side of the ODE in (34) is not zero and doesn't change sign over $[m_0, t_0]$. The choice of x_0 and t_0 imply that the denominator in the right-hand side of the ODE is positive

at $t = t_0$. Hence we must have that this denominator is positive over $[m_0, t_0]$. This fact implies that property (48) holds for the ODE (34) in the interval $(a, b] := (m_0, t_0]$. Since $x(t_0) = x_0 < 0$ we can apply Lemma 10 to conclude that $x(t) < 0$ for all $t \in [m_0, t_0]$, where $m_0 \in \mathbb{R} \cup \{-\infty\}$.

Next, we prove that the solution $x(\cdot)$ can be infinitely extended to the left, in other words, $m_0 = -\infty$. Suppose that, on the contrary, $m_0 \in \mathbb{R}$. By Lemma 8(c), this can only happen if the right hand side of the ODE in (34) becomes discontinuous at $t = m_0$. This implies that

$$\mu_{xx}(x(m_0), m_0) = 0\,. \tag{49}$$

By Lemma 9, we have

$$\mu_x(x(t), t) = 0\,,$$

for all $t \in [m_0, t_0]$. Therefore,

$$\mu_x(x(m_0), m_0) = 0\,. \tag{50}$$

By Lemma 6, Equations (49)–(50) have a unique solution with $x(m_0) = \sqrt[3]{a_1}/2 > 0$. This is in contradiction with the second statement in Lemma 10, which asserts that $x(m_0) < 0$. Hence we must have $m_0 = -\infty$. The proof for the case when $a_1 < 0$ is obtained similarly. Namely, in this case we use that $x_0 > 0$ and Lemma 10 must be used for this case. □

Theorem 3 (Well-definedness Yielding Global Minimizer) *For a monic depressed quartic polynomial, with $a_1 \neq 0$ and $a_2 < 0$, Algorithm 2 is well-defined and it yields the global minimizer.*

Proof. By Step 1 of Algorithm 2, $t_0 = \sqrt{-a_2/2}$ and $x_0 = -\sqrt[3]{a_1/4}$. Since $a_2 < 0$ and $a_1 \neq 0$, by Lemma 11, Step 2 results in $\operatorname{sgn}(x(t)) = -\operatorname{sgn}(a_1)$ for all $t \in (-\infty, t_0]$. Moreover, by Lemma 9, we have that

$$\mu_x(x(t), t) = 0\,, \quad \mu_{xx}(x(t), t) > 0\,, \ \forall t \in (-\infty, t_0]\,.$$

Therefore, $x(0)$ is a local minimizer. Now, Lemma 4 and the fact that $\operatorname{sgn}(x(0)) = -\operatorname{sgn}(a_1)$ imply that $x(0)$ must be the global minimizer. □

Remark 8 By Lemma 7, if $-3\,a_1^{2/3}/2 < a_2 < 0$, then the monic depressed quartic polynomial is quasi-convex. If $a_2 > 0$ then f is convex. In either case, Algorithm 2 is not necessary. □

Corollary 1 *Algorithm 2 is well-defined and convergent for any monic quartic polynomial.*

Proof. Any quartic polynomial can be obtained from a depressed quartic polynomial through horizontal translation, or shift (and vice versa). Therefore, Lemma 2 on scale-shift invariance of Algorithm 1, furnishes the proof. □

4.4 Three types of trajectories

Clearly, a monic quartic polynomial $f(x)$ has at most three local extrema, given by the roots of $f'(x)$. If the roots of $f'(x)$ are distinct, then they correspond to two local minimizers and one local maximizer of $f(x)$. Figure 1 considers three cases in which (a) $f'(x)$ has a single real root, (b) $f'(x)$ has symmetric real roots and (c) $f'(x)$ has nonsymmetric real roots. The trajectories run "forward" from each of these roots are depicted in Figure 1, providing a full characterization, on the surface defined by (29).

The case when the root of $f'(x)$ is unique is exemplified in Figure 1(a): with $f(x) = x^4 - 0.09\,x^2 - 0.03\,x - 1$, the minimizer of $f'(x)$ is $x \approx 0.304668$. In this case, $a_1 = -0.03 \neq 0$ and $a_2 = -0.09 < 0$. It is easily checked that the conclusion of Lemma 11 is satisfied, in that $\operatorname{sgn}(x(t)) = -\operatorname{sgn}(a_1)$ for all $t \in [0, t_0]$.

The case when $f'(x)$ has more than one (distinct) real root is exemplified in Figures 1(b)-(c). In Figure 1(b), we have $f(x) = x^4 - 0.98\,x^2 + 1$, where $a_1 = 0$, so $f(x)$ has two global minimizers, $x = -0.7$ and $x = 0.7$, and the local maximizer $x = 0$. This case (when $a_1 = 0$) is trivial, for which there is no need to implement Algorithm 2.

On the other hand, the case when $f'(x)$ has more than one (distinct) real root, and these roots are nonsymmetric, is exemplified in Figure 1(c), with the polynomial $f(x) = x^4 - 4\,x^3/15 - 0.82\,x^2 + 0.168\,x + 1$. The polynomial $f(x)$ has its global minimum at $x = -0.6$ and a local minimum at $x = 0.7$. The local maximizer of $f(x)$ is $x = 0.1$. One can easily verify Lemma 11, in that $\operatorname{sgn}(x(t)) = -\operatorname{sgn}(a_1)$ for all $t \in [0, t_0]$. In this case, the system $\mu_x(x,t) = 0$ and $\mu_{xx}(x,t) = 0$ has a solution by Lemma 6, which is shown at the bottom left in Figure 1 as the point where two of the trajectories emanating from the local maximum and local minimum points merge on the surface.

5 Trajectory Methods with Quadratic Regularization

The regularization idea for polynomial optimization is not new. Such an approach, although not explicitly stated as a regularization, is employed in [17], for finding a global minimizer of a monic polynomial f of even degree. In [17], in an algorithm similar to Algorithm 1, the function

$$\varphi(x,t) = f(x) + \frac{t}{2}\,x^2\,, \tag{51}$$

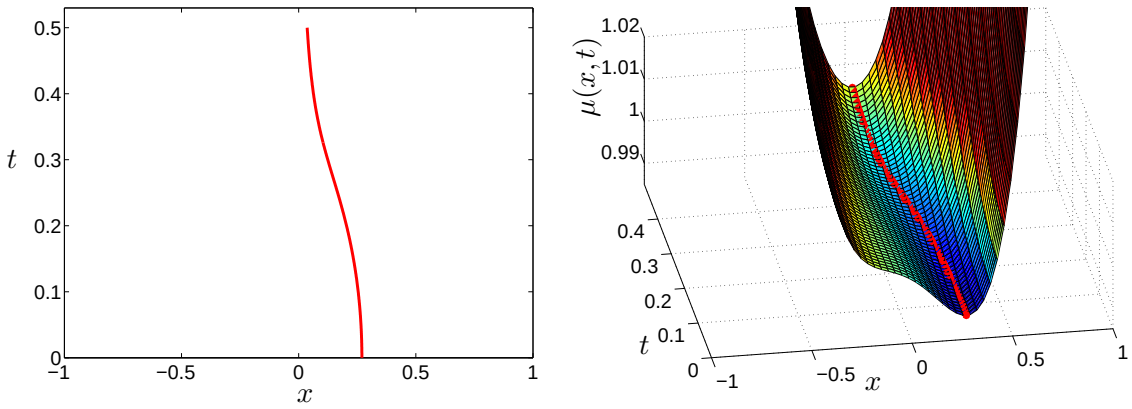
is effectively used, instead of $\mu(x,t)$. We refer to φ as the *quadratic regularization* of f. A direct computation from (51) yields the partial derivatives

$$\varphi_x(x,t) = f'(x) + t\,x\,, \quad \varphi_{xx}(x,t) = f''(x) + t\,, \quad \varphi_{tx}(x,t) = x\,. \tag{52}$$

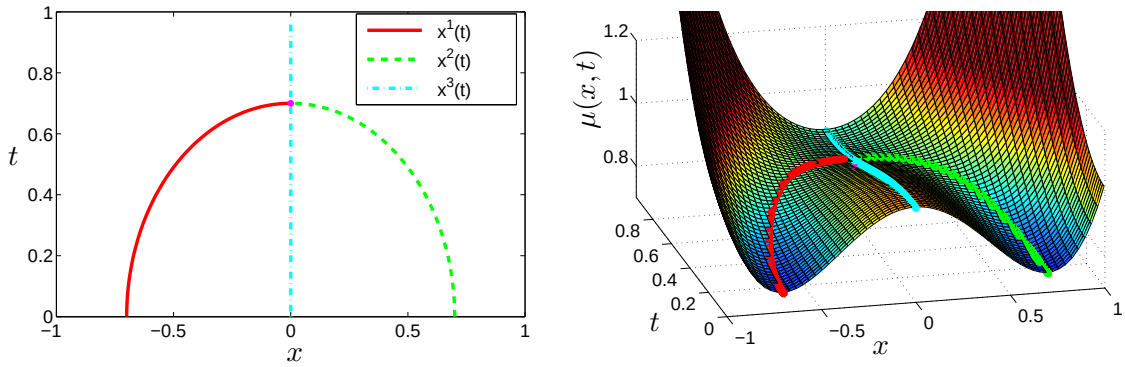
The formula for φ_{xx} directly gives a well-known convexity result analogous to Theorem 1, for the regularization $\varphi(\cdot, t)$.

Remark 9 *Assume that f is twice differentiable, and that $l_0 := \inf_{x\in\mathbb{R}} f''(x) \in \mathbb{R}$. Then, $\varphi(\cdot,t)$ is convex for $t \geq -l_0$.*

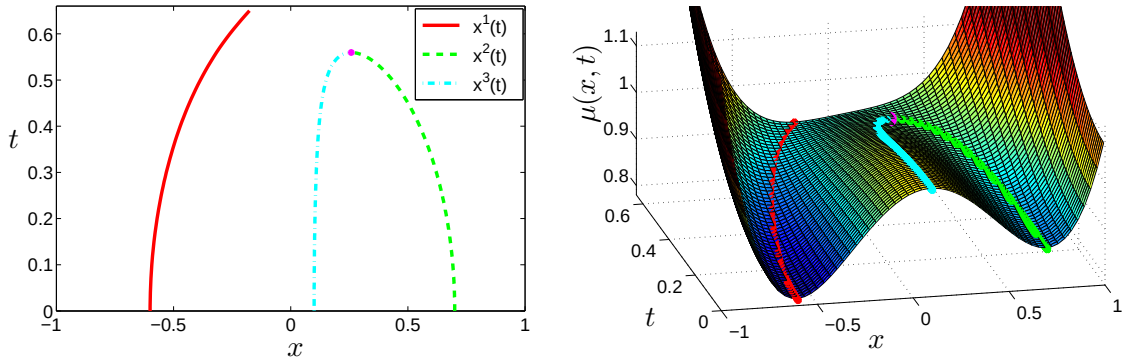
This remark can be used in Steps 1–2 to give Algorithm 3, companion of Algorithm 1, for the quadratic regularization.



(a) A quasi-convex f: $f(x) = x^4 - 0.09\,x^2 - 0.03\,x - 1$.



(b) A nonconvex f with symmetric roots: $f(x) = x^4 - 0.98\,x^2 + 1$.



(c) A nonconvex f with nonsymmetric roots: $f(x) = x^4 - 4\,x^3/15 - 0.82\,x^2 + 0.168\,x + 1$.

Figure 1: Examples for three types of polynomials and their associated trajectories.

Algorithm 3

Step 1 *Choose the parameter $t_0 > 0$ large enough so that $\varphi(\cdot, t_0)$ is convex. Find the (global) minimizer x_0 of $\varphi(\cdot, t_0)$, i.e., solve $\varphi_x(x_0, t_0) = 0$ for x_0.*

Step 2 *Solve the initial value problem*

$$\dot{x}(t) = -\frac{\varphi_{tx}(x(t), t)}{\varphi_{xx}(x(t), t)} = -\frac{x(t)}{f''(x) + t}, \quad \text{for a.e. } t \in (0, t_0], \quad \text{with } x(t_0) = x_0. \tag{53}$$

Step 3 *Report $x(0)$ as a global minimizer of $f(\cdot)$.*

The quadratic regularization defined in (51) is not scale-shift invariant, in the sense of Lemma 2. This fact has been established in [1] by means of an example. Namely, if Algorithm 3 is applied to a general quartic polynomial (not depressed) then it may not yield a global minimizer. We further illustrate this fact by means of example polynomials, including those of higher degrees, in the next section.

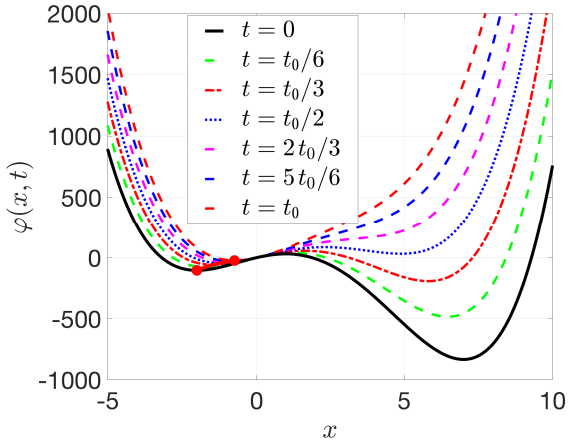
6 Numerical Experiments

In this section, via numerical experiments, we illustrate the working of our trajectory method devised utilizing Steklov regularization, i.e., Algorithm 1 (which becomes Algorithm 2 for the case in which f is a quartic polynomial), on example problems involving quartic and higher-degree polynomials, as well as an example involving a non-polynomial function. We provide comparisons with the trajectory method in [17], namely Algorithm 3, which, as pointed in Section 5, can be derived using a quadratic regularization.

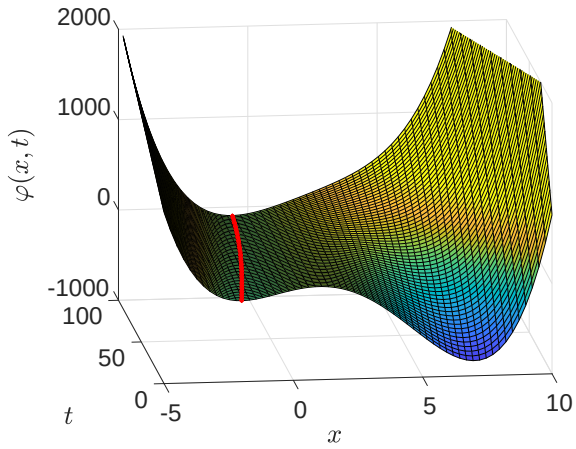
We illustrate the behaviour of the algorithms by means of graphs. In Figures 2–4 for the polynomial examples presented in this paper, the graphs in parts (a) and (c) of the figures provide the "contours of t," i.e., the graph of the regularization function (quadratic or Steklov) with a number of fixed values of t between 0 and a chosen value of t_0. In parts (b) and (d) of the figures, a surface plot of the regularizing function (quadratic or Steklov) is provided. In the figure for the non-polynomial example considered in Subsection 6.5, similar graphs are displayed.

In Subsection 6.4, we measure the performance of Algorithms 1 and 3 for randomly generated polynomials of certain degrees. Table 1 shows that Algorithm 1 is always convergent for the quartic polynomials generated randomly, in line with Theorem 3, and convergent for the great majority of the higher-degree polynomials generated randomly. It is further observed that, although Algorithm 1 does not converge for all the tested polynomials of degree greater than four, it clearly outperforms Algorithm 3.

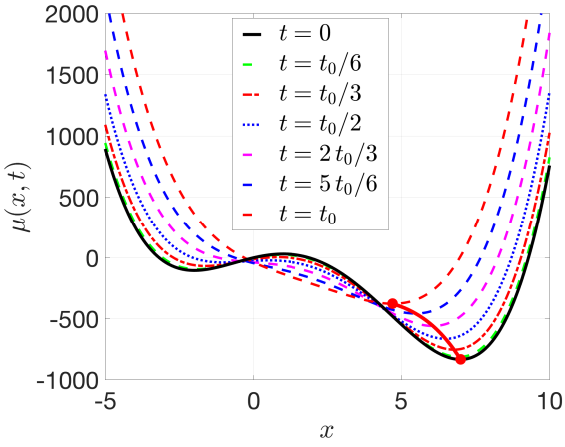
In all graphs, the trajectory, or the solution curve of an ODE, constructed by an algorithm is also depicted. A trajectory is generated by solving the pertaining initial value problem (IVP) using the MATLAB function `ode15s`, with `RelTol = 1e-08`. We have used `ode15s`, which is a choice for stiff ODEs, since only then it was possible to get a solution of the IVP or a message saying that it was not possible to get a solution, the latter being useful in obtaining the success rates of the algorithms in Subsection 6.4.



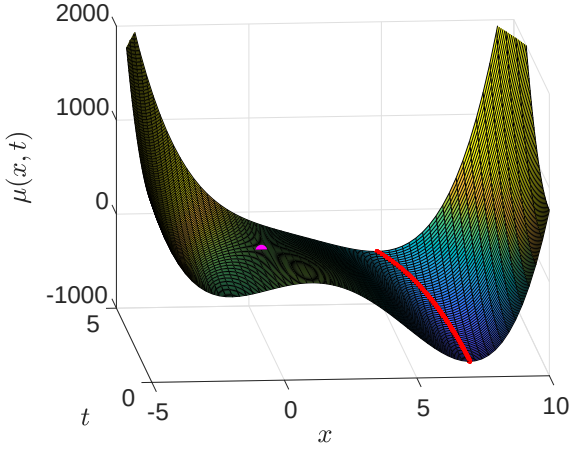
(a) Quadratic regularization – contours of t with $t_0 = 100$.



(b) Quadratic regularization – surface with $t_0 = 100$.



(c) Steklov regularization – contours of t with $t_0 = 5$.



(d) Steklov regularization – surface with $t_0 = 5$.

Figure 2: Trajectory methods for the quartic polynomial, $f(x) = x^4 - 8\,x^3 - 18\,x^2 + 56\,x$.

6.1 A quartic polynomial

Consider minimization of the polynomial

$$f(x) = x^4 - 8\,x^3 - 18\,x^2 + 56\,x\,,$$

which has local minima at $x = -2$ and $x = 7$ and a local maximum at $x = 1$. Note that $f(-2) = -104$, $f(7) = -833$ and $f(1) = 31$. Therefore, $x = 7$ is the global minimizer of $f(x)$. This polynomial is provided in [1] as a counterexample to prove that the trajectory approach in [17] using quadratic regularization, i.e., Algorithm 3 given in the present paper, does not necessarily yield to a global minimizer, as opposed to the claim in [17]. Indeed, as Figure 2(a)–(b) illustrates, the trajectory constructed by the quadratic regularization converges to the local minimizer $x = -2$ rather than the global minimizer $x = 7$. As discussed in [1], the quadratic regularization function $\varphi(\cdot, t_0)$ convexifies the given quartic polynomial with $t_0 > 84$. For visual convenience in Figure 2(a)–(b), we have used $t_0 = 100$ in Algorithm 3, as in [1]. Again, from [1], the corresponding $x_0 \approx -0.6812$.

Algorithm 1 yields the global minimizer, as expected by Theorem 3, see Figure 2(c)–(d). The polynomial $f(x)$ can be rewritten in a depressed form using the transformation $x = z+2$ as

$$f(z) = z^4 - 42\,z^2 - 80\,z - 8\,.$$

By using Lemma 5, we see that the Steklov regularization convexifies the given quartic polynomial just for $t_0 = \sqrt{21} \approx 4.5826$, for which $x_0 = 2 + \sqrt[3]{20} \approx 4.7144$. Again for visual convenience we have used $t_0 = 5$.

In fact, by the Flatness Lemma 6, the Steklov function $\mu(\cdot, t)$ is quasi-convex at $\widehat{t} \approx 2.6599$. We note that $\mu_x(\widehat{x}, \widehat{t}) = 0 = \mu_x(\widehat{x}, \widehat{t})$, with $(\widehat{x}, \widehat{t}) \approx (2.6599, -0.1544)$, which is also indicated with a (pink) mark in Figure 2(d).

One could as well have used $t_0 = \widehat{t} \approx 2.66$ for which $\mu(\cdot, t_0)$ is quasi-convex, and Algorithm 1 can be run with (x_0, t_0) and the associated initial condition $\mu_x(x_0, t_0) = 0$, in Step 1.

As will be seen also with the higher-order polynomials, the Steklov function $\mu(\cdot, t_0)$ is convex with a rather small t_0. On the other hand, the quadratic regularization function $\varphi(\cdot, t_0)$ becomes convex with a much larger t_0, which is almost 20 times the t_0 needed for the Steklov function. In the subsequent subsections, it will be observed that t_0 grows greatly with the degree of a polynomial. When solving an IVP, a big t_0 makes the time span (or time horizon) $[0, t_0]$ big and this causes ODE solvers to take a much longer time and run more often into difficulties.

6.2 A degree-6 polynomial

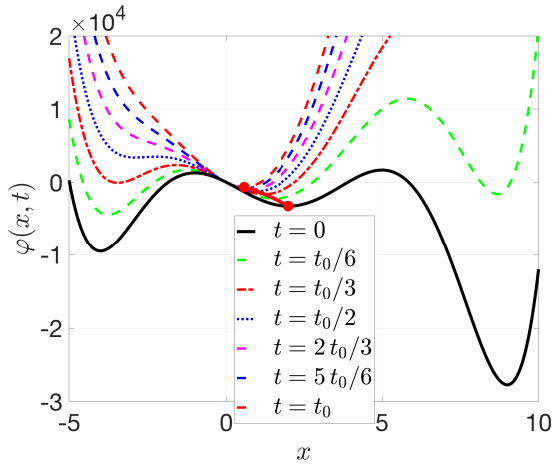
Consider minimization of the degree-6 polynomial

$$f(x) = x^6 - 66\,x^5/5 - 9\,x^4/2 + 422\,x^3 - 474\,x^2 - 2160\,x\,,$$

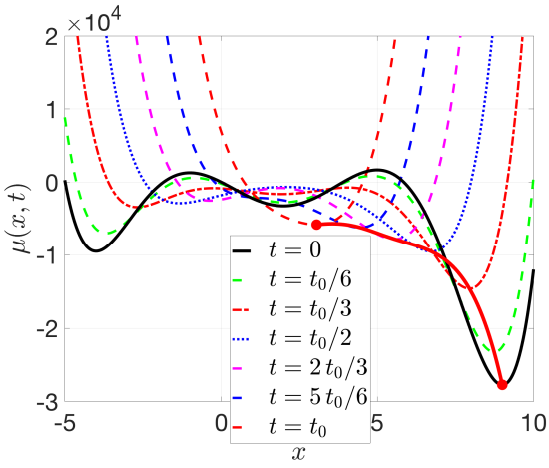
which has local minima at $x = -4, 2$ and 9 and local maxima at $x = -1$ and 5. A graph of the polynomial can be seen in Figure 3. The global minimizer of $f(x)$ is $x = 9$, with $f(9) = -27726.3$ (exactly). One has the local minima $f(-4) = -9491.2$ and $f(2) = -3270.4$, and the local maxima $f(-1) = 1273.7$ and $f(5) = 1662.5$, all exact.

We observe that $t_0 = 7$ is enough to convexify the Steklov function $\mu(\cdot, t_0)$, while the quadratic regularization function $\varphi(\cdot, t_0)$ requires $t_0 \approx 4000$ to become convex. Moreover, Algorithm 3 (using the quadratic regularization) yields the local minimizer $x = 2$, while Algorithm 1 (using the Steklov regularization) yields the global minimizer $x = 9$.

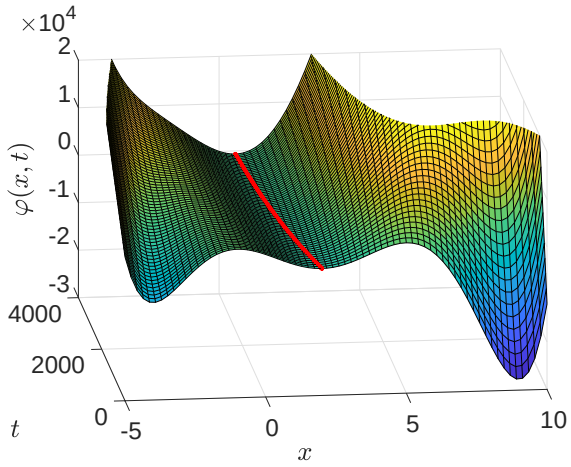
This polynomial $f(x)$ is just one degree-6 polynomial example to illustrate the working and success of Algorithm 1, as well as the working and failure of Algorithm 3. Algorithm 1 can also fail for some degree-6 polynomials, but not as often as Algorithm 3 does. As mentioned before, detailed comparisons of success rates for each of the algorithms are shown in Subsection 6.4.



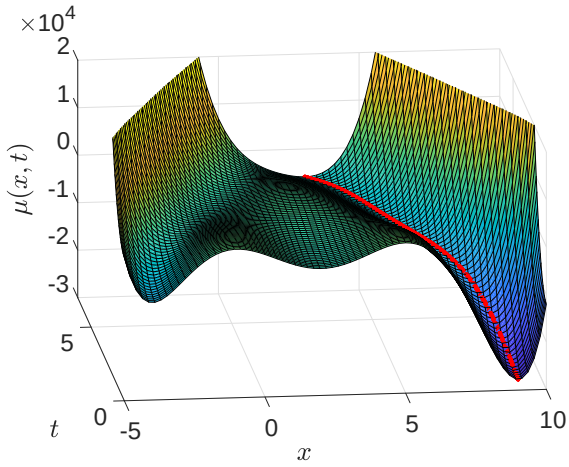
(a) Quadratic regularization – contours of t with $t_0 = 4000$.



(c) Steklov regularization – contours of t with $t_0 = 7$.

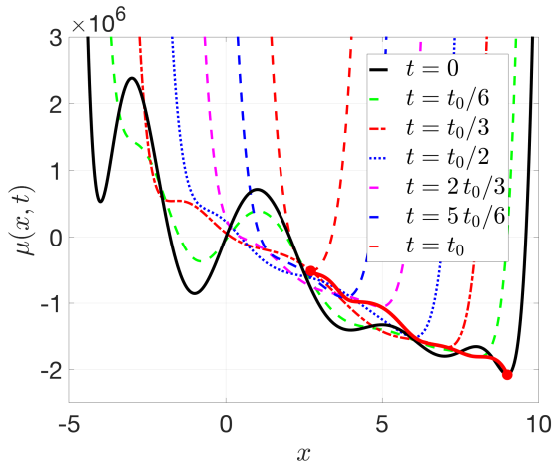


(b) Quadratic regularization – surface with $t_0 = 4000$.

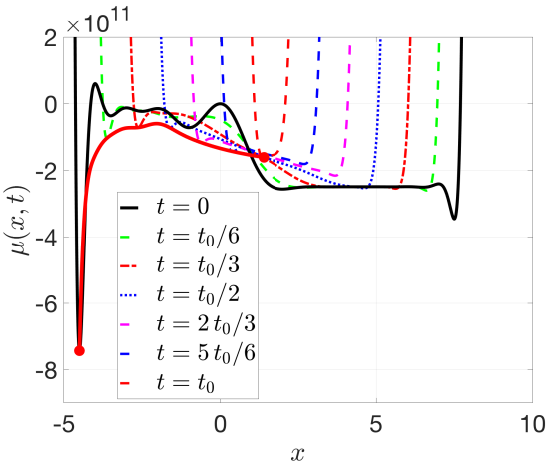


(d) Steklov regularization – surface with $t_0 = 7$.

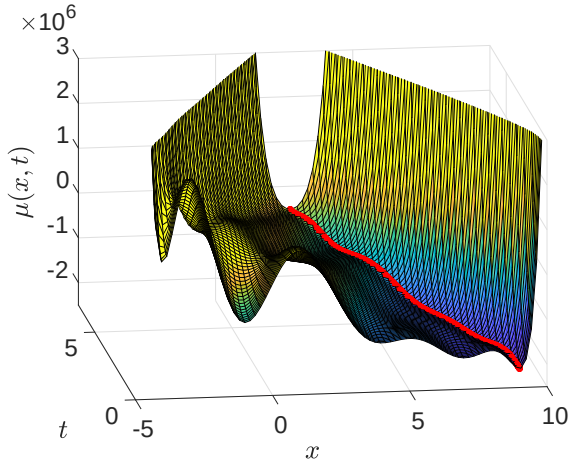
Figure 3: Trajectory methods for the degree-6 polynomial, $f(x) = x^6 - 66\,x^5/5 - 9\,x^4/2 + 422\,x^3 - 474\,x^2 - 2160\,x$.



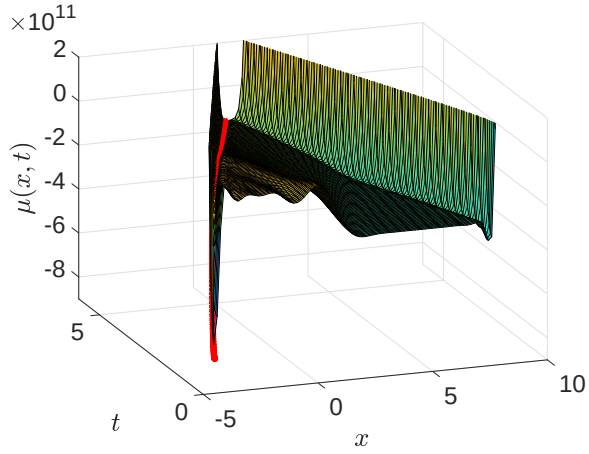
(a) Steklov regularization – contours of t with $t_0 = 7$.



(c) Steklov regularization – contours of t with $t_0 = 6$.



(b) Steklov regularization – surface with $t_0 = 7$.



(d) Steklov regularization – surface with $t_0 = 6$.

Figure 4: Algorithm 1 for some given degree-10 polynomial (parts (a) and (b)) and degree-20 polynomial (parts (c) and (d)).

6.3 Degree-10 and degree-20 polynomials

Consider minimization of the degree-10 monic polynomial $f_{10}(x)$ with the coefficients

$$[a_9, \cdots, a_0] = [260/9, 1035/4, -120, -9415, 32172, 175765/2, -1369360/3, -148560, 1209600, 0]\,.$$

The global minimizer of $f_{10}(x)$ is $x = 9$, with $f_{10}(9) = -2077224.75$ (exactly).

We also consider minimization of the degree-20 monic polynomial $f_{20}(x)$ with the coefficients

$$\begin{aligned}[a_{19}, \cdots, a_0] \;=\; & [680/19, 3935/9, -15755/17, -196105/8, 2230697/12, 20765145/112,\\ & -1351162585/208, 10221013715/768, 6382409515/64, -12625444643/32,\\ & -200463718805/288, 2498521767895/512, 465297612345/448, -2045419187205/64,\\ & 198942566751/16, 3627285358725/32, -56515087125, -201131555625, 0, 0]\,.\end{aligned}$$

The global minimizer of $f_{20}(x)$ is $x = -4.5$, with $f_{20}(-4.5) = -742786593463.8248$ (exactly).

	Algorithm 1			*Algorithm 3*	
n	t_0	Failure rate	n	t_0	Failure rate
4	6	0%	4	10^3	26%
6	7	1%	6	10^4	63%
8	7	2%	8	10^5	77%
10	7	4%	10	10^8	84%
12	7	4%	12	10^8	88%
14	7	4%	14	10^8	92%
20	7	7%	20	10^{10}	96%

Table 1: A comparison of failure rates of Algorithms 1 and 3 for degree-n polynomials (n as listed).

Algorithm 1 successfully yields a global minimizer for both polynomials as can be seen in Figure 4. Moderate sizes of t_0 (7 and 6, respectively) suffice in each case for convexification.

Algorithm 3 fails to serve the purpose for either polynomial. For the degree-10 polynomial, it only yields the local minimizer $x = -1$, with $t_0 = 2 \times 10^6$, which, although quite large, is just enough for convexification. For the degree-20 polynomial, one needs a far larger $t_0 \approx 10^{12}$ for convexification; however, the ODE solver takes an indefinite amount of time and does not provide any answer, conceivably because of the very large orders of magnitude involved in the computations. The graphs that were generated suggest that the trajectory method would yield $x = 0$, which this time is a local maximizer! For brevity, we do not provide the graphs for the quadratic regularization.

6.4 Performance comparisons between Algorithms 1 and 3

In Sections 6.1–6.3, we have applied Algorithms 1 and 3 to four selected polynomials of degrees four, six, 10 and 20, and illustrated the workings of both algorithms. Algorithm 1 was successful in finding a global minimum of each of the polynomials considered in Sections 6.1–6.3, while Algorithm 3 consistently failed. In all fairness, neither Algorithm 1 is successful in dealing with every single polynomial (computationally speaking) nor Algorithm 3 is unsuccessful for every single polynomial. To better understand how these two methods compare, we present failure rates of both algorithms for 1000 randomly generated polynomials of various degrees. We have generated the polynomials in such a way that their extremal values were uniformly distributed over the interval $[-5, 5]$.

Table 1 lists the failure rates for each algorithm as they are applied to polynomials of various degrees, where the polynomials of each degree are randomly generated 1000 times. We declare failure of the method for a given polynomial when either the algorithm did not converge, or it converged to a local minimum. For polynomials of degree higher than four, it is not trivial (if not impossible), to check convexity of $\mu(\cdot, t_0)$ or $\varphi(\cdot, t_0)$ for a given t_0. The choice we made for the value of t_0 required in each of the experiments is drastically different for each method. We observe that Algorithm 1 requires $t_0 \in \{6, 7\}$ for all cases, while Algorithm 3 requires $t_0 \in [10^3, 10^{10}]$, with values increasing with the degree of the polynomials. These large values of t_0 promote convexification of $\varphi(\cdot, t_0)$, but, at the same time, they are likely to cause numerical instabilities. In summary, perhaps not many but still some of the failures of Algorithm 3 may be attributed to (i) t_0 not being large enough for convexification, (ii) t_0 being too large, or both. It is likely that values of t_0 greater than the values of t_0 already listed in Table 1 (especially for high degree polynomials) will cause numerical instabilities. This is another reason why Algorithm 1 looks favourable, when compared with Algorithm 3.

Having made these remarks, especially for the high degree polynomials, one may consider doing a rescaling in order to avoid high orders of magnitudes in computations; however, we have not considered a rescaling of any of the polynomials in our computations. In the case of Algorithm 1, there is certainly room for choosing t_0 to be bigger in the experiments.

The failure rates for Algorithm 3 are very high, increasing sharply with degree, reaching 84–96% for polynomials of degree 10–20. Even for quartic polynomials, the failure rate of Algorithm 3 is rather high, at 26%, while Algorithm 1 has no failures for this case, as expected by Theorem 3. From the numerical experiments, we observe that Algorithm 1 can fail, even for degree-6 and degree-8 polynomials; however, the failure rate is small, at 1–2%, in practical terms. This rate is far smaller than that of Algorithm 3 for similar degree polynomials, as shown in Table 1.

6.5 A non-polynomial function

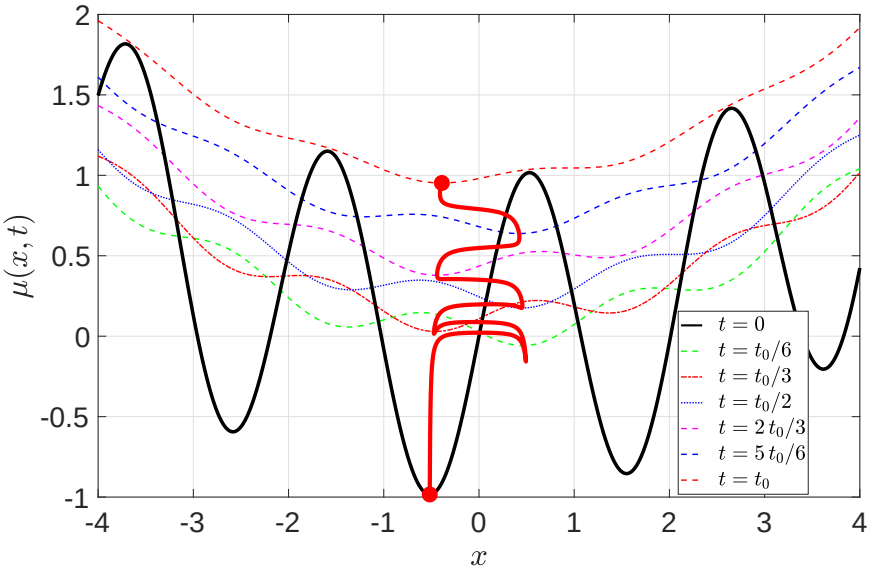
In Sections 6.1–6.4, we have tested the performance of Algorithms 1 and 3 for polynomial functions. In this section we consider the nonpolynomial coercive function

$$f(x) = 0.06\,x^2 + \sin 3x\,,$$

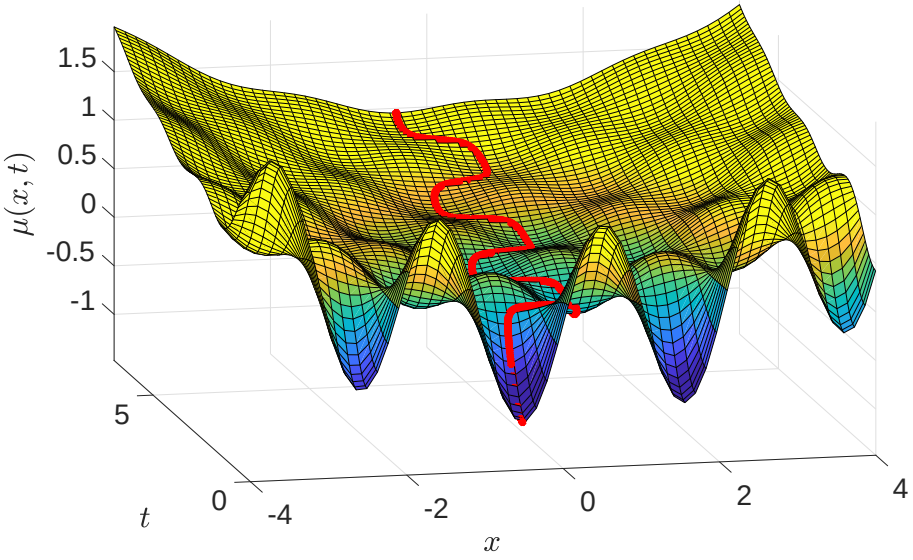
which has many local minima. Even though this function does not satisfy the assumptions of Theorem 1, it is straightforward to check that the conclusion of this theorem holds for this function. To do this, we use (3) and some elementary algebra to derive

$$\mu_{xx}(x,t) = 0.12 - 3\sin 3x\,\frac{\sin 3t}{t}.$$

Since $\lim_{t\to\infty}(\sin 3t)/t = 0$, we have that for every x there exists a $t_0 > 0$ (independent of x), such that $\mu_{xx}(x,t) > 0$ for all $t > t_0$. Thus, for those values of t, $\mu(\cdot,t)$ is strictly convex. In our experiments, however, we use a t_0 that makes $\mu(\cdot,t_0)$ quasi-convex. Namely, with the parameter value of $t_0 = 7$, the Steklov function becomes quasi-convex, and its minimizer is $x_0 = -0.3896$. This minimizer is in turn used to start solving the initial value problem. Algorithm 1 then finds the global minimizer as $x(0) = -0.5167$. Figures 5(a) and 5(b) illustrate the t-contour and surface plots, respectively, as well as the trajectory constructed by Algorithm 1, in each of parts (a) and (b).



(a) Steklov regularization – contours of t with $t_0 = 7$.



(b) Steklov regularization – surface with $t_0 = 7$.

Figure 5: Algorithm 1 for the nonpolynomial function, $f(x) = 0.06\,x^2 + \sin 3x$.

7 Conclusion

We have proposed a trajectory-based algorithm, Algorithm 1, using the Steklov regularization function $\mu(x,t)$, for finding a global minimizer of univariate coercive functions. The so-called Steklov smoothing function has been previously studied in the literature, as a smoothing tool for small values of t. Our study considers using this function as a regularization tool. Namely, we have proved that, for large enough t, μ convexifies certain univariate coercive functions. We proved convergence of Algorithm 1 for quartic polynomials. We tested it for higher-degree polynomials, as well as a non-polynomial function for illustration.

We have made comparisons with an existing trajectory-based algorithm, reformulated here as Algorithm 3, which uses a quadratic regularization instead. Using 1000 randomly generated polynomials, we found that, for degree-6 polynomials, while the failure rate of Algorithm 1 is only 1%, Algorithm 3 fails in 63% of the cases. For degree-20 polynomials, these percentages are 7 and 96, respectively, pointing to the fact that Algorithm 1 provides a better option.

Throughout the paper, we obtained auxiliary results (apart from convergence) regarding Algorithm 1, the Steklov function, and quartic polynomials, which are worthy in their on right.

In Algorithm 1, we require t_0 to be chosen so as to convexify μ; however, one may instead require t_0 to *quasi-convexify* μ, which would possibly result in an even smaller, i.e., a more desirable, t_0. One should note that most of the powerful numerical methods for minimization of convex functions are also applicable to minimization of quasi-convex functions [4].

As with any other global optimization technique, Algorithm 1 cannot find a global optimizer in every single situation. However, it provides a promising and viable option for searching global minimizers of general univariate coercive functions. A future line of investigation should concern extensions of Algorithm 1 to multi-variable coercive functions, which clearly has a much wider scope for theory and applications.

References

[1] O. Arikan, R. S. Burachik and C. Y. Kaya, "Backward differential flow" may not converge to a global minimizer of polynomials. *J. Optim. Theory Applic.*, **167**, 401–408, 2015.

[2] V. I. Arnold, *Ordinary Differential Equations*. The MIT Press, Cambridge, 1978.

[3] H, Attouch, Z. Chbani, J. Peypouquet, and P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.*, **168**(1-2), 123–175, 2018.

[4] M. S. Bazaraa, H. D. Sherali and C. M. Shetti, *Nonlinear Programming: Theory and Algorithms, 3rd edition*. Wiley InterScience, New Jersey, 2006.

[5] R. I. Boţ and E. R. Csetnek, Convergence rates for forward–backward dynamical systems associated with strongly monotone inclusions. *J. Math. Anal. Applic.*, **457**(2), 1135–1152, 2018.

[6] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization. *Math. Program., Ser. B*, **134**, 71–99, 2012.

[7] Y. M. Ermoliev, V. I. Norkin, and R. J.-B. Wets, The minimization of semicontinuous functions: mollifier subgradients. *SIAM J. Control Optim.*, **32**, 149–167, 1995.

[8] R. Garmanjani, L. N. Vicente, Smoothing and worst-case complexity for direct-search methods in nonsmooth optimization. *IMA J. Num. Anal.*, **33**, 1008–1028, 2013.

[9] A. M. Gupal, On a method for the minimization of almost-differentiable functions. *Cybernet. Syst. Anal.*, **13**, 115–117, 1977.

[10] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*. Springer-Verlag, Berlin, Heidelberg, Germany, 1996.

[11] D. Lera and Y. D. Sergeyev, Acceleration of univariate global optimization algorithms working with Lipschitz functions and Lipschitz first derivatives. *SIAM J. Optim.*, **23**(1), 508–529, 2013.

[12] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer-Verlag, Berlin, Heidelberg, Germany, 2004.

[13] D. Scholz, *Deterministic Global Optimization: Geometric Branch-and-bound Methods and Their Applications*. Springer, New York, 2012.

[14] J. A. Snyman and S. Kok, A reassessment of the Snyman–Fatti dynamic search trajectory method for unconstrained global optimization. *J. Glob. Optim.*, **43**, 67–82, 2009.

[15] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*. Springer-Verlag, Berlin-Heidelberg, 1970.

[16] X. Zhang and Y. Xiong, Impulse noise removal using directional difference based noise detector and adaptive weighted mean filter. *IEEE Signal Proc. Lett.*, **16**, 295–298, 2009.

[17] J. Zhu, S. Zhao, and G. Liu, Solution to global minimization of polynomials by backward differential flow. *J. Optim. Theory Applic.*, **161**, 828–836, 2014.