Real Root Isolation for Non-linear Functions

The approaches that we have seen so far, strongly rely on the fact that the input is a polynomial, i.e., has some algebraic structure. However, occasionally we may want to isolate the real roots of a continuos function $f : \mathbb{R} \to \mathbb{R}$, which is given to us as a black box that we can query at a value $x \in \mathbb{R}$ and get $f(x)$ in return. How do we isolate the real roots of f in this model? From the continuity of f, we know that if for $a < b$, $f(a)f(b) \le 0$ then $[a, b]$ has odd number of roots of f. Moreover, if we knew that f was monotone in the interval $[a, b]$ then there must be a unique root of f in $[a, b]$. Thus, if we can check whether f is monotone in an interval I then by looking at the signs at the endpoints of I, we can determine whether or not f has a root in I. One way to check whether f is monotone on I, is to check if I has no critical points of f, i.e., $0 \notin f'(I)$, where $f'(I) := \{f'(x) | x \in I\}$; similarly, define $f(I)$. Thus what we want, in general, is an estimate on $f(I)$, since getting the exact range may be costly. For this purpose we relax our representation of f to an **interval extension** of f: $\Box f$ takes as input an interval I and returns an interval such that $f(I) \subseteq \Box f(I)$. We additionally assume that if a sequence of intervals $I_0 \supset I_1 \supset I_2 \supset \cdots$ converges to a point α, then the sequence of intervals $\Box f(I_0) \supset \Box f(I_1) \supset \Box f(I_2)$ converges to $f(\alpha)$. Once we have such a function, the algorithm is straightforward.

INPUT: f and an input interval I_0.
OUTPUT: Isolating intervals for roots of f in I_0.
1. Initialize a queue $Q := I_0$.
2. While Q is not empty do
    Remove an interval I from Q.
    If $0 \notin \Box f(I)$ discard I.
    Else If $0 \notin \Box f'(I)$ then
        If f changes sign at endpoints of I then output I else discard I.
    Else
        Subdivide I and put the two halves onto Q.

Let $C_0 \equiv 0 \notin \Box f(I)$ and $C_1 \equiv 0 \notin \Box f'(I)$. Why does the algorithm terminate? Let's assume that f is square-free, i.e. the roots of f and f' are distinct. If the algorithm does not terminate, then there is an infinite sequence of intervals $I_0 \subset I_1 \subset I_2 \cdots$ such that $0 \in \Box f(I_k)$ and $0 \in \Box f'(I_k)$, $k = 0, 1, 2, \ldots,$. Since the intervals are converging to a point α, the assumption on the box-functions imply that $f(\alpha) = f'(\alpha) = 0$, giving us a contradiction. The algorithm above works for analytic functions, in general, but from now on we study the complexity of the algorithm for the case of polynomials. We begin with how to compute the box-function.

1 Box Functions for Polynomials

Let $f(x) = \sum_{j=0}^{n} a_j x^j$, $a_j \in \mathbb{R}$. Given an interval I, a natural way to compute $\Box f(I)$ is to compute $\sum_{j=0}^{n} a_j I^j$, but that does I^j mean, what does $a_j I^J$ mean? The correct way to interpret these terms is to do **interval arithmetic**, or the arithmetic of overestimation. Given two intervals $[a, b], [c, d] \subset \mathbb{R}$, define the following basic ring operations over intervals as follows:

1. Addition: $[a, b] + [c, d] = [a + c, b + d]$.
2. Subtraction: $[a, b] - [c, d] = [a - d, b - c]$.
3. Multiplication: $[a, b] \times [c, d] = [\min\{ac, ad, bc, bd\}, \max\{ac, ad, bc, bd\}]$. Powering $[a, b]^2$ in our setting is to be interpreted in the sense of multiplication above and not as $[\min\{0, a^2, b^2\}, \max\{a^2, b^2\}]$.

4. Negation: $-[a, b] = [-b, -a]$.
5. With any $a \in \mathbb{R}$, we associate the interval $[a, a]$.

Interval arithmetic is commutative, and associative for both addition and multiplication. However, distributivity only holds in the following weak sense: $I(J + K) \subset IJ + IK$, with equality holding when either $I \in \mathbb{R}$, or J, K have the same sign.

Now it is clear what we mean by $\sum_{j=0}^{n} a_j I^j$. Let's call this form as the **standard form**. But as is the case in polynomial evaluation, we always prefer the Horner's evaluation over the standard evaluation. The **Horner Form** of interval extension does the Horner evaluation, but with intervals instead of x. It is interesting to note that when evaluating functions over intervals, the choice of expression makes a difference. For example, $x^2 - x$, for $x = [0, 2]$, gives us using the standard form $[0, 4] - [0, 2] = [-2, 4]$ and using Horner's form $[0, 2]([0, 2] - [1, 1]) = [0, 2][-1, 1] = [-2, 2]$, which is slightly better estimate on the range. Another form is called the **mean-value form** is based upon the mean value theorem:

$$f(I) \subseteq \Box f'(I)(I - m) + f(m)$$

where m is the midpoint of I, and the interval extension of f' uses the Horner form. However, we will use a different form which uses the slope function of f w.r.t. an interval I: for all $x \in I$, we know that there exists a function $g(x)$ such that [1]

$$g(x) = \frac{f(x) - f(m)}{(x - m)}.$$

It is not hard to show that in fact

$$g(x) = \sum_{j=1}^{n} \frac{f^{(j)}(m)}{j!}(x - m)^{j-1}.$$

The **centered form** or **slope form** is given as

$$\Box f(I) := f(m) + \Box g(I)(I - m) \tag{1}$$

where the interval extension of g uses the Horner form. We can derive an upper bound on the width of the interval returned by centered form, by using the observation that the Horner form is never worse than the standard form we know that

$$\Box f(I) \subseteq f(m) + \sum_{j=0}^{n} \frac{f^{(j)}(m)}{j!}(I - m)^j.$$

There is a simpler way to write the expression above, using the fact that $(I - m) = w(I)/2[-1, 1]$, and $a[-1, 1]^j = a[-1, 1] = |a|\,[-1, 1]$:

$$\begin{aligned} \Box f(I) &= f(m) + \sum_{j=0}^{n} \left| \frac{f^{(j)}(m)}{j!} \right| \left(\frac{w(I)}{2} \right)^j [-1, 1] \\ &= f(m) + [-1, 1] \sum_{j=0}^{n} \left| \frac{f^{(j)}(m)}{j!} \right| \left(\frac{w(I)}{2} \right)^j . \end{aligned} \tag{2}$$

Therefore, the algorithm terminates when either

$$|f(m)| \geq \sum_{j=0}^{n} \left| \frac{f^{(j)}(m)}{j!} \right| \left(\frac{w(I)}{2} \right)^j \tag{3}$$

or

$$|f'(m)| \geq \sum_{j=0}^{n} \left| \frac{f^{(j+1)}(m)}{j!} \right| \left(\frac{w(I)}{2} \right)^j . \tag{4}$$

We now derive suitable converse for these conditions to hold.

[1] Note that this looks similar to the mean-value theorem. The difference is that g contains all the lower order terms as well, whereas the mean-value theorem only contains the term corresponding to the first order, namely the derivative.

2 The Integral Bound

A converse has to roughly state that if the $w(I)$ is small enough then either C_0 or C_1 holds. Our hope is that for intervals not containing a root of f, C_0 will hold and for intervals containing a root there is a sufficiently small neighbourhood where f is monotone and so C_1 will hold. Suppose there is a function $G : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that if

$$w(I)G(x) \leq 1 \tag{5}$$

for some $x \in I$ then either C_0 or C_1 holds. Call such a function a **stopping function**. Thus a stopping function looks locally and can determine if C_0 or C_1 holds. The crucial observation is that the size of the subdivision tree $\mathcal{T}$ can be bounded in terms of $G(x)$:

$$|\mathcal{T}| \leq 2 \max \left\{ 1, 2 \int_{I_0} G(x) dx \right\}. \tag{6}$$

To see this, again consider a leaf I of $\mathcal{T}$ and let J be the interval associated with its parent. Since C_0 and C_1 fail to hold at J, it follows from (5) that for all $x \in J$, $w(J)G(x) = 2w(I)G(x) \geq 1$. This implies that

$$2 \int_I G(x) dx \geq 2w(I) \min_{x \in I} G(x) \geq 1.$$

Since the leaves $I_1, I_2, \ldots, I_k$ form a partitioning of I_0, it follows that

$$2 \int_{I_0} G(x) dx = \sum_{j=1}^{k} 2 \int_{I_j} G(x) dx \geq k.$$

Since $\mathcal{T}$ is a binary tree the number of internal nodes cannot exceed k, giving us (6).

Now we have to find an appropriate stopping function. Basically, the stopping function has to capture the intuition that if there is an $x \in I$ such that $w(I)$ is smaller than $c|x - \alpha|$, for some constant, then C_0 holds, where α represents any root of f; a similar constraint, would be implied if α is replaced by α', a critical point of f. Thus a first choice for stopping function is the inverse of the distance from x to a root nearest to x. But instead of focussing on a nearest root, we will take the following amortized sum:

$$S(x) := \sum_{i=1}^{n} \frac{1}{|x - \alpha_i|}. \tag{7}$$

Why is this function interesting? Why does it yield a stopping function? If $2w(I)S(x) \leq 1$ then

$$\sum_{i=1}^{n} \frac{2}{|x - \alpha_i|} \leq \frac{1}{w(I)}$$

which implies that for all α_i, $2w(I) \leq |x - \alpha_i|$, i.e. the nearest root to x is at least $2w(I)$ distance away. Since $|m - x| \leq w(I)/2$ this implies that

$$|\alpha_i - m| \geq |\alpha_i - x| - |x - m| = |\alpha_i - x| - \frac{w}{2} \geq \frac{3}{4}|\alpha_i - x|.$$

Taking the inverse and summing it for $i = 1, \ldots, m$, we obtain that

$$S(m) \leq \frac{4}{3} S(x).$$

Since $2wS(x) \leq 1$, it follows that

$$wS(m) \leq \frac{2}{3}. \tag{8}$$

But what is the relation between $S(x)$ and the terms appearing in the RHS of (3)? Clearly, $S(x)$ is an upper bound on $|f'(x)/f(x)|$. Furthermore, we also have the following inequality

$$\left| \frac{f^{(j)}(x)}{f(x)} \right| \leq S(x)^j.$$

This is not hard to see since

$$\left|\frac{f^{(j)}(x)}{f(x)}\right| = \left|\sum_{1\le i_1<i_2<\cdots<i_j\le n} \frac{1}{(x-\alpha_{i_1})(x-\alpha_{i_2})\ldots(x-\alpha_{i_j})}\right| \le (S(x))^j .$$

Thus for any $x \in \mathbb{R}$

$$\sum_{j=1}^{n}\left|\frac{f^{(j)}(x)}{j!f(x)}\right|(w/2)^j \le \sum_{j=1}^{n}\frac{(S(x)w/2)^j}{j!}.$$

In particular, for $x = m$ we obtain from (8) that

$$\sum_{j=1}^{n}\left|\frac{f^{(j)}(m)}{j!f(m)}\right|(w/2)^j \le \sum_{j=1}^{n}\frac{1}{3^j j!} \le e^{1/3} - 1 \le 1.$$

To summarize we have shown that if there exists $x \in I$ such that $2S(x)w(I) \le 1$ then C_0 holds. Define $S'(x)$ as in (7) but w.r.t. the roots of f'; then we can similarly show that if $2S'(x)w(I) \le 1$ then C_1 holds. Therefore,

$$G(x) := 2\min\{S(x), S'(x)\}$$

is a stopping function.

2.1 Size of Subdivision Tree

The beauty of the integral bound, (6), is that instead of looking at the partitions corresponding to the intervals at the leaves, which is what we did in our earlier analyses of Sturm's method and the Descartes method, we have expressed the bound w.r.t. I_0, therefore, for analysing the bound we can now break I_0 into a partition covenient to us.

In this section, we we will show:

LEMMA B

$$\int_{I_0} G(x)dx = O(n(L + r + \log d)) \tag{9}$$

where r is the number of real roots in I_0.

Towards proving Lemma B, we first bound the integral on $G(x)$ by the sum of two integrals on $S(x)$ and $S'(x)$, respectively. Suppose $\{I_1, I_1'\}$ is a partition of I_0 into two sets. Clearly, we have the inequality

$$\int_{I_0} G(x)dx \le \int_{I_1} S(x)dx + \int_{I_1'} S'(x)dx. \tag{10}$$

This inequality is trivial if any of the integrals on the right hand side is infinite. Finiteness of the integrals on the right hand side is equivalent to ensuring that $I_1 \cap V$ and $I_I' \cap V'$ are both empty sets. We will ensure this and some additional properties in forming the partition $\{I_1, I_1'\}$.

Assume $I_0 \cap V = \{\alpha_1, \ldots, \alpha_r\}$ and

$$a < \alpha_1 < \alpha_2 < \cdots < \alpha_r < b$$

where $I_0 = [a, b]$. We may also define $\alpha_0 := a$ and $\alpha_{r+1} := b$.

For each root $\alpha \in I_0$, we define the interval I_α as the intersection of real axis with the disc centered at α and radius equal to half the distance from α to the nearest critical point; note that two such intervals do not overlap, since by Rolle's theorem we have a critical point between any two roots in I_0. Finally, we define the sets I_1 and I_1':

$$I_1' := \bigcup_{\alpha \in I_0 \cap V} I_\alpha$$

and I_1 is just the closure of $I_0 \setminus I_1'$. It is easy to see that $I_1' \cap V' = \emptyset, \quad I_1 \cap V = \emptyset$ and thus the right hand side of (10) is finite.

¶1. Bounds on two basic integrals We will reduce our integrals to one of the two forms here:

LEMMA 1. *Let $\alpha \in \mathbb{C}$ and $J = [r, s] \subseteq I_0$. Assume $\alpha \notin J$.*
(Re) *If α is real, then*

$$\int_J \frac{dx}{|\alpha - x|} = \ln \left| \frac{\alpha - s}{\alpha - r} \right|^{\delta(J>\alpha)} \leq L + 1 - \ln \min \{|\alpha - r|, |\alpha - s|\} \tag{11}$$

where $\delta(P) \in \{+1, -1\}$ is the Kronecker symbol: for any predicate P, $\delta(P) = +1$ if P holds, and $\delta(P) = -1$ otherwise.
(Im) *If α is not real, $\alpha = \mathtt{Re}(\alpha) + \mathbf{i}\mathtt{Im}(\alpha)$, then*

$$\begin{aligned} \int_J \frac{dx}{|\alpha - x|} &= \ln \left(\frac{(s - \mathtt{Re}(\alpha)) + |\alpha - s|}{(r - \mathtt{Re}(\alpha)) + |\alpha - r|} \right) \\ &\leq \ln 4 \left| \frac{(\alpha - s)(\alpha - r)}{\mathtt{Im}(\alpha)^2} \right| \\ &\leq 2(2 + L - \ln |\mathtt{Im}(\alpha)|). \end{aligned} \tag{12}$$

Proof. **(Re)** From basic calculus we verify that (see [1, p. 162])

$$\int_r^s \frac{dx}{|\alpha - x|} = \ln \left| \frac{\alpha - s}{\alpha - r} \right|^{\delta(J>\alpha)}.$$

If $J > \alpha$ then $\int_r^s dx/|\alpha - x| = \ln(s - \alpha) - \ln(r - \alpha)$. If $J < \alpha$, we reverse the roles of r and s. But $\ln \max\{|\alpha - s|, |\alpha - r|\} \leq 1 + L$, which gives us the desired upper bound in (11).
(Im) Writing $\alpha = \mathtt{Re}(\alpha) + \mathbf{i}\mathtt{Im}(\alpha) = R + \mathbf{i}I$, we have [1, p. 162]:

$$\int_r^s \frac{dx}{|\alpha - x|} = \operatorname{arcsinh}\left(\frac{s - R}{|I|} \right) - \operatorname{arcsinh}\left(\frac{r - R}{|I|} \right).$$

Since $\operatorname{arcsinh}(x) = \ln(x + \sqrt{1 + x^2})$ we conclude that

$$\int_r^s dx/|\alpha - x| = \ln \left(\frac{(s - R) + |\alpha - s|}{(r - R) + |\alpha - r|} \right)$$

where $|\alpha - s| = \sqrt{(R - s)^2 + I^2}$. The numerator $\ln((s - R) + |s - \alpha|) \leq \ln(2|\alpha - s|)$, and the denominator

$$(r - R) + |\alpha - r| = \frac{|I|^2}{|\alpha - r| - (r - R)} \geq \frac{|I|^2}{2|\alpha - r|}.$$

Thus

$$\ln \left(\frac{(s - R) + |\alpha - s|}{(r - R) + |\alpha - r|} \right) \leq \ln 4 \left| \frac{(\alpha - s)(\alpha - r)}{I^2} \right|.$$

Since $|\alpha - s|, |\alpha - r| \leq 2^{L+1}$, we obtain

$$\ln \left(\frac{(s - R) + |\alpha - s|}{(r - R) + |\alpha - r|} \right) \leq 2((2 + L) - \ln |I|)$$

as claimed in (12). **Q.E.D.**

¶2. Bounding the integral over I_1 We bound the first integral on the RHS of (10) as follows:

$$\int_{I_1} S(x)dx = O(n(L + \log n)). \tag{13}$$

To show this, we express the integral as a sum over all roots α in V:

$$\int_{I_1} S(x)dx = \sum_{\alpha\in V}\int_{I_1}\frac{dx}{|x-\alpha|}. \tag{14}$$

The summand corresponding to a particular α can be bounded using one of the two cases in Lemma 1:
(Re) Suppose $\alpha\in\mathbb{R}$. Let $I_\alpha=[\alpha^-,\alpha^+]$ be the interval associated with α. Thus $(\alpha-\alpha^-)=(\alpha^+-\alpha)=|\alpha-\alpha^*|/2$ where α^* is a critical point nearest to α. Writing $I_0=[a,b]$, we can bound the summand with the help of Lemma 1(Re):

$$\begin{aligned}
\int_{I_1}\frac{dx}{|x-\alpha|} &\le \int_{I_0\setminus I_\alpha}\frac{dx}{|x-\alpha|}\\
&= \int_a^{\alpha^-}\frac{dx}{\alpha-x}+\int_{\alpha^+}^b\frac{dx}{x-\alpha}\\
&\le (L+1-\ln|\alpha-\alpha^-|)+(L+1-\ln|\alpha-\alpha^+|)\\
&= 2(L+1)-2\ln|\alpha-\alpha^-|\\
&= 2(L+1)-2\ln\frac{|\alpha-\alpha^*|}{2}.
\end{aligned}$$

Summing over all real roots $\alpha\in V$, yields $2d(L+1)-2\ln\prod_{\alpha\in V}|\alpha-\alpha^*|/2$, which is equal to $O(n(L+\log n))$ from Mahler-Davenport.
(Im) Suppose $\alpha\notin\mathbb{R}$. Then Lemma 1(Im) says

$$\int_{I_1}\frac{dx}{|x-\alpha|}\le\int_{I_0}\frac{dx}{|x-\alpha|}\le 2(2+L-\ln|\mathtt{Im}(\alpha)|).$$

Again, summing over all non-real $\alpha\in V$ and using the Mahler-Davenport bound we get that (14) is bounded by $O(n(L+\log n))$.

Cases **(Re)** and **(Im)** imply the desired bound in (13).

¶3. Bounding the integral over I_1' It remains to bound the second integral on the RHS of (10) as follows:

$$\int_{I_1'} S'(x)dx = O(nr). \tag{15}$$

This integral is written as a double summation, summing over all critical points $\alpha'\in V'$, and summing over all $\alpha\in V\cap I_0$:

$$\begin{aligned}
\int_{I_1'} S'(x)dx &= \sum_{\alpha'\in V'}\int_{I_1'}\frac{1}{|x-\alpha'|}dx\\
&= \sum_{\alpha'\in V'}\sum_{\alpha\in V\cap I_0}\int_{I_\alpha}\frac{1}{|x-\alpha'|}dx.
\end{aligned} \tag{16}$$

Fix a particular root α and critical point α'. Write $I_\alpha=[\alpha^-,\alpha^+]$, and let α^* be a critical point nearest to α; since α is equidistant from α^+ and α^-, we express this distance as $|\alpha-\alpha^\pm|$. There are again two cases to consider.
(Re') Suppose α' is real. Then Lemma 1(Re) yields

$$\int_{I_\alpha}\frac{1}{|x-\alpha'|}dx = \ln\left|\frac{\alpha'-\alpha^+}{\alpha'-\alpha^-}\right|^{\delta(I_\alpha>\alpha')}. \tag{17}$$

By the triangular inequality

$$|\alpha'-\alpha^\pm|\ge|\alpha'-\alpha|-|\alpha-\alpha^\pm| = |\alpha'-\alpha|-\frac{|\alpha-\alpha^*|}{2}. \tag{18}$$

Since α^* is a critical point nearest to α it further follows that

$$|\alpha' - \alpha^{\pm}| \geq |\alpha' - \alpha| - \frac{|\alpha - \alpha'|}{2} = |\alpha' - \alpha|/2. \tag{19}$$

Similarly, we can show $|\alpha' - \alpha^{\pm}| \leq 2|\alpha' - \alpha|$. Thus

$$\frac{|\alpha' - \alpha|}{2} \leq |\alpha' - \alpha^{\pm}| < 2|\alpha' - \alpha|. \tag{20}$$

Note that these inequalities are independent of the fact that $\alpha' \in \mathbb{R}$. Applying these inequalities to the RHS of (17) we obtain that the integral on the LHS is at most $\ln 4$.

(Im') Suppose $\alpha' \notin \mathbb{R}$. Here, we recognize three subcases (i) $I_\alpha < \mathtt{Re}(\alpha')$, (ii) $I_\alpha > \mathtt{Re}(\alpha')$, and (iii) $\mathtt{Re}(\alpha') \in I_\alpha$. For the first two subcases, we know from [1] that

$$\int_{I_\alpha} \frac{dx}{|x - \alpha'|} \leq \ln 2 \left| \frac{\alpha' - \alpha^+}{\alpha' - \alpha^-} \right|^{\delta(I_\alpha > \mathtt{Re}(\alpha))}.$$

Furthermore, the bounds from (20) imply that the integral above is bounded by $\ln 8$. In the third subcase, Lemma 1(Im) yields

$$\int_{I_\alpha} \frac{dx}{|x - \alpha'|} \leq \ln 4 \left| \frac{(\alpha' - \alpha^+)(\alpha' - \alpha^-)}{\mathtt{Im}(\alpha')^2} \right|.$$

Applying the upper bound from (20), we further get

$$\int_{I_\alpha} \frac{dx}{|x - \alpha'|} < \ln 16 \frac{|\alpha' - \alpha|^2}{|\mathtt{Im}(\alpha')|^2}.$$

From the triangle inequality it follows that $|\mathtt{Im}(\alpha')| \geq |\alpha' - \alpha| - |\alpha - \mathtt{Re}(\alpha')|$. Since $\mathtt{Re}(\alpha') \in I_\alpha$, we further have $|\mathtt{Im}(\alpha')| \geq |\alpha' - \alpha| - |\alpha - \alpha^{\pm}|$, which we know from (18) and (19) is greater than $|\alpha' - \alpha|/2$. Thus we have $|\mathtt{Im}(\alpha')| \geq |\alpha' - \alpha|/2$.

Therefore the integral $\int_{I_\alpha} dx/|x - \alpha'|$ in subcase (iii), hence in all subcases, is at most $\ln 64$.

Thus cases **(Re')** and **(Im')** imply that each integral in the RHS of (16) is at most $\ln 64$. If r is the number of real roots in I_0 then we have

$$\int_{I_1'} S'(x)dx < (n-1)r \ln 64.$$

This proves (15).

References

[1] M. Burr and F. Krahmer. SqFreeEVAL: An (almost) optimal real-root isolation algorithm. *J. Symbolic Computation*, 47(2):153–166, 2012.