This is supplementary material for the article "Efficient Real Root Approximation" by Michael Kerber and Michael Sagraloff, printed in the *Proceedings of the 2011 International Symposium on Symbolic and Algebraic Computation*. All references, equation numbers and theorem numbers refer to that article.

A. BOUNDING THE SEPARATION

We give a complete proof of Theorem 19. First, we repeat the Davenport-Mahler theorem as given in [9]; see also Eigenwillig's PhD thesis ("Real Root Isolation for Exact and Approximate Polynomials Using Descartes' Rule of Signs", Saarland University 2008) for a more general version.

Theorem 23 (Davenport-Mahler bound). *Let $g \in \mathbb{C}[t]$ be square-free of degree d and let $G = (V, E)$ be a directed graph on the roots of g such that:*

- *G is acyclic,*
- *for every edge $(\alpha, \beta) \in E$, it holds $|\alpha| \le |\beta|$, and*
- *the in-degree of any node is at most* 1.

In this situation

$$\prod_{(\alpha,\beta)\in E} |\alpha-\beta| \ge \frac{\sqrt{|\mathrm{res}(g,g')|}}{\sqrt{|\mathrm{lcf}(g)|}\mathrm{Mea}(g)^{d-1}} \cdot \left(\frac{\sqrt{3}}{d}\right)^{\#E} \cdot \left(\frac{1}{d}\right)^{d/2}.$$

Theorem 19.

$$\Sigma_f \in O(d(\tau + \log d) + R)$$

PROOF. As before, let $z_1, \ldots, z_n$ denote the roots of f, and let $B \ge 1$ denote a bound for the maximal absolute value of a root. Observe that, when the z's are considered as vertices in the complex plane, each σ_i is given by the length of an edge connecting z_i to its nearest neighbor. This induces a directed graph on the vertices, which is known as the *nearest neighbor graph* [8] (if a root has more than one nearest neighbor, we pick the one with highest index). Let E_0 denote the edge set of this nearest neighbor graph. We can rewrite:

$$\prod_{i=1}^{d} \sigma_i = \prod_{(z_i,z_j)\in E_0} |z_j - z_i|$$

Our goal is to apply the Davenport-Mahler bound on this product. However, the nearest-neighbor graph does not satisfy any of the required properties in general. We will transform the edge set E_0 into another edge set E_3 that satisfies the requirements of the Davenport-Mahler theorem, and we will relate the root product of E_0 with the root product of E_3.

Note that a direct property of nearest neighbor graphs is that all cycles have length 2 [8]. In the first step, we remove one edge of every cycle:

$$E_1 := \{(z_i, z_j) \in E_0 \mid i < j \vee (z_j, z_i) \notin E_0\}$$

This removes at most every second edge, and for every removed edge, there is some edge in E_1 with the same length. Since every root difference is bounded by $2B$ from above, we can bound

$$(2B)^d \prod_{(z_i,z_j)\in E_0} |z_j - z_i| \ge \prod_{(z_i,z_j)\in E_1} |z_j - z_i|^2.$$

In the next step, we re-direct the edges in E_1 in order to satisfy the second condition of the Davenport-Mahler bound

$$E_2 := \{(z_i, z_j) \mid ((z_i, z_j) \in E_1 \vee (z_j, z_i) \in E_1) \wedge (|z_i| < |z_j| \vee (|z_i| = |z_j| \wedge i < j))\}$$

In simple words, every edge points to the root with greater absolute value. Note that E_2 does not contain any cycles, because the absolute value of a root is non-decreasing on any path, and if it remains the same, the index increases, thus no vertex can be visited twice on such a path. Since the only difference between E_1 and E_2 is the orientation of edges, we have

$$\prod_{(z_i,z_j)\in E_1} |z_j - z_i| = \prod_{(z_i,z_j)\in E_2} |z_j - z_i|$$

Finally, we need to ensure the last condition of the Davenport-Mahler bound, namely that each vertex has in-degree at most 1. For that, if several edges point to some z_j, we throw away all of them except the shortest one (in the definition, if the shortest edge is not unique, we keep the one with the maximal index):

$$E_3 := \{(z_i, z_j) \in E_2 \mid \forall (z_k, z_j) \in E_2 : |z_k - z_j| > |z_i - z_j| \vee (|z_k - z_j| = |z_i - z_j| \wedge k \le i)\}$$

Another basic property of the nearest neighbor graph is that two edges that meet in a vertex must form an angle of at least 60°. It follows that the degree of every vertex is bounded by 6. Since E_2 is a subgraph of the nearest neighbor graph, possibly with some edges flipped, the degree of every vertex is still bounded by 6. Since all edges in E_2 point to the root with greater absolute value, it can be easily seen that the in-degree of z_j is even bounded by 3. So, E_3 contains at least $\frac{E_2}{3}$ many edges. Since we always keep a smallest edge pointing to a z_j, we can bound

$$(2B)^{2d} \prod_{(z_i,z_j)\in E_2} |z_j - z_i| \ge \prod_{(z_i,z_j)\in E_3} |z_j - z_i|^3.$$

Putting everything together, we have that

$$\prod_{(z_i,z_j)\in E_0} |z_j - z_i| \ge (2B)^{-5d} \left(\prod_{(z_i,z_j)\in E_3} |z_j - z_i|\right)^6.$$

E_3 meets all prerequisites of the Davenport-Mahler bound and we can thus bound

$$\begin{aligned}
\prod_{i=1}^{d} \sigma_i &= \prod_{(z_i,z_j)\in E_0} |z_j - z_i| \\
&\ge (2B)^{-5d} \left(\prod_{(z_i,z_j)\in E_3} |z_j - z_i|\right)^6 \\
&\ge (2B)^{-5d} \left(\frac{\sqrt{|\mathrm{res}(f,f')|}}{\sqrt{|\mathrm{lcf}(f)|}\mathrm{Mea}(f)^{d-1}} \cdot \left(\frac{\sqrt{3}}{g}\right)^{\#E_3} \cdot \left(\frac{1}{d}\right)^{d/2}\right)^6 \\
&\ge (2B)^{-5d} \left(\frac{\sqrt{|\mathrm{res}(f,f')|}}{\sqrt{|\mathrm{lcf}(f)|}\mathrm{Mea}(f)^{d-1}} \cdot \left(\frac{1}{d}\right)^{2d}\right)^6
\end{aligned}$$

Taking the inverse on both sides and applying the logarithm, we get

$$\begin{aligned}
\Sigma_f &\le 5d\log 2B + 3\log \mathrm{lcf}(f) + 6(d-1)\log \mathrm{Mea}(f) + 12d\log d + 6\cdot R \\
&= O(d(\tau + \log d) + R),
\end{aligned}$$

exploiting the fact that $B \in O(2^\tau)$ and $\log \mathrm{Mea}(f) \in O(\tau + \log n)$. $\square$

B. ERROR ANALYSIS OF INTERVAL ARITHMETIC

We restate Lemma 3 and provide its proof.

Lemma 3. *Let f be a polynomial as in (1), $c \in \mathbb{R}$ with $|c| \le 2^\tau$, and $\rho \in \mathbb{N}$. Then,*

$$|f(c) - \mathrm{down}(f(c),\rho)| \le 2^{-\rho+1}(d+1)^2 2^{\tau d} \quad (1)$$

$$|f(c) - \mathrm{up}(f(c),\rho)| \le 2^{-\rho+1}(d+1)^2 2^{\tau d} \quad (2)$$

In particular, $\mathfrak{B}(f(c),\rho)$ has a width of at most $2^{-\rho+2}(d+1)^2 2^{\tau d}$.

PROOF. We do induction on d. The statement is clearly true for $d = 0$. For $d > 0$, we write $f(c) = a_0 + cg(c)$ with $a_0 \in \mathbb{R}$ the constant coefficient of f and g of degree $d-1$. Note that, for any real value x, $|\mathrm{down}(x,\rho) - x| < 2^{-\rho}$, same for up. Therefore, we can bound as follows (again, leaving ρ out for simplicity):

$$\begin{aligned}|f(c) - \mathrm{down}(f(c))| &= |a_0 + cg(c) - \mathrm{down}(a_0 + cg(c))| \\ &= |a_0 + cg(c) - \mathrm{down}(a_0) - \mathrm{down}(cg(c))| \\ &\le |cg(c) - \mathrm{down}(cg(c))| + 2^{-\rho}\end{aligned}$$

Note that $\mathrm{down}(c \cdot g(c)) = \mathrm{down}(H_1(c) \cdot H_2(g(c)))$ where $H_{1,2} =$ down or $H_{1,2} =$ up. Moreover, we can write $H_1(c) = c - \varepsilon$ with $|\varepsilon| < 2^{-\rho}$. Therefore, we can rearrange

$$\begin{aligned}& |cg(c) - \mathrm{down}(cg(c))| + 2^{-\rho} \\ \le\ & |cg(c) - (c-\varepsilon)\cdot H_2(g(c))| + 2^{-\rho+1} \\ \le\ & |cg(c) - c\cdot H_2(g(c))| + |\varepsilon|\cdot|H_2(g(c))| + 2^{-\rho+1} \\ \le\ & |c|\cdot|g(c) - H_2(g(c))| + 2^{-\rho}|H_2(g(c))| + 2^{-\rho+1}\end{aligned}$$

By a simple inductive proof on the degree, we can show that both $|\mathrm{up}(g(c))|$ and $|\mathrm{down}(g(c))|$ are bounded by $d2^{d\tau}$. Using that and the induction hypothesis yields

$$\begin{aligned}& |c|\cdot|g(c) - h(g(c))| + 2^{-\rho}|H_2(g(c))| + 2^{-\rho+1} \\ <\ & 2^\tau 2^{-\rho+1} d^2 2^{\tau(d-1)} + 2^{-\rho} d 2^{\tau d} + 2^{-\rho+1} \\ \le\ & 2^{-\rho+1}(d^2+d+1)2^{\tau d} \le 2^{-\rho+1}(d+1)^2 2^{\tau d}\end{aligned}$$

The bound for $|f(c) - \mathrm{up}(f(c))|$ follows in the same way. □

C. DETAILS ON QUADRATIC CONVERGENCE

For convenience, we repeat some proofs of [12] adapted to our notation. Recall from Definition 12 that

$$C_\xi := \frac{|f'(\xi)|}{8ed^3 2^\tau \max\{|\xi|,1\}^{d-1}}.$$

We need one additional lemma to prove some properties of C_ξ.

Lemma 24. *Let $\xi \in \mathbb{C}$ be a root of f.*

1. $0 < C_\xi \le \frac{1}{d}$
2. *Let $\mu \in \mathbb{C}$ be such that $|\xi - \mu| < C_\xi$. Then*

$$C_\xi \cdot |f''(\mu)| < \frac{|f'(\xi)|}{8}.$$

PROOF. By a straight-forward estimation, we can bound $|f'(\xi)|$ from above by $d^2 2^\tau \max\{|\xi|,1\}^{d-1}$. which proves the first claim.

For the second claim, we bound $|f''(\mu)|$ from above:

$$\begin{aligned}|f''(\mu)| &= |\sum_{i=2}^{d} i(i-1)a_i\mu^{i-2}| \le d^2 2^\tau \sum_{i=0}^{d-2}\max\{|\mu|,1\}^i \\ &\le d^2 2^\tau \sum_{i=0}^{d-2}\Big((1+C_\xi)\max\{|\xi|,1\}\Big)^i \\ &\le d^3 2^\tau (1+C_\xi)^{d-2}\max\{|\xi|,1\}^{d-2} \\ &< d^3 2^\tau \underbrace{(1+\frac{1}{d})^d}_{<e}\max\{|\xi|,1\}^{d-1} \quad \square\end{aligned}$$

Lemma 13. *Let (a,b) an isolating interval for ξ of width $\delta < C_\xi$. Then $|m - \xi| < \frac{\delta^2}{8C_\xi}$.*

PROOF. We consider the Taylor expansion of f at ξ. For a given $x \in (a,b)$, we have

$$f(x) = f'(\xi)(x-\xi) + \frac{1}{2}f''(\tilde{\xi})(x-\xi)^2$$

with some $\tilde{\xi} \in [x,\xi]$ or $[\xi,x]$. Thus, we can simplify

$$\begin{aligned}|m-\xi| &= \left|\frac{f(b)(a-\xi) - f(a)(b-\xi)}{f(b)-f(a)}\right| \\ &= \left|\frac{\frac{1}{2}(f''(\tilde{\xi}_1)(b-\xi)^2(a-\xi) - f''(\tilde{\xi}_2)(a-\xi)^2(b-\xi))}{f(b)-f(a)}\right| \\ &\le \frac{1}{2}|b-\xi||a-\xi|\cdot\frac{|f''(\tilde{\xi}_1)|(b-\xi) + |f''(\tilde{\xi}_2)|(\xi-a)}{|f(b)-f(a)|} \\ &\le \frac{\delta^2\max\{|f''(\tilde{\xi}_1)|,|f''(\tilde{\xi}_2)|\}}{2|f'(\nu)|}\end{aligned}$$

for some $\nu \in (a,b)$. The Taylor expansion of f' yields $f'(\nu) = f'(\xi) + f''(\tilde{\nu})(\nu - \xi)$ with $\tilde{\nu} \in (a,b)$. Since $\delta \le C_\xi$, it follows from Lemma 24

$$|f''(\tilde{\nu})(\nu-\xi)| \le |f''(\tilde{\nu})|C_\xi \le \frac{1}{8}|f'(\xi)|.$$

Therefore $|f'(\nu)| > \frac{7}{8}|f'(\xi)| > \frac{1}{2}|f'(\xi)|$, and it follows again with Lemma 24 that

$$\begin{aligned}|m-\xi| &\le \frac{\delta^2\max\{|f''(\tilde{\xi}_1)|,|f''(\tilde{\xi}_2)|\}}{|f'(\xi)|} \\ &\le \frac{\delta^2}{8\frac{|f'(\xi)|}{8\max\{|f''(\tilde{\xi}_1)|,|f''(\tilde{\xi}_2)|\}}} < \frac{\delta^2}{8C_\xi} \quad \square\end{aligned}$$

Corollary 15. *In the quadratic sequence, there is at most one failing* AQIR *call.*

PROOF. Let $(I_i,N_i) \overset{\text{AQIR}}{\to} (I_{i+1},N_{i+1})$ be the first failing AQIR call in the quadratic sequence. Since the quadratic sequence starts with a successful AQIR call, the predecessor $(I_{i-1},N_{i-1}) \overset{\text{AQIR}}{\to} (I_i,N_i)$ is also part of quadratic sequence, and succeeds. Thus we have the sequence

$$(I_{i-1},N_{i-1}) \overset{\substack{\textit{Sucess}\\ \text{AQIR}}}{\to} (I_i,N_i) \overset{\substack{\textit{Fail}\\ \text{AQIR}}}{\to} (I_{i+1},N_{i+1})$$

One observes easily that $w(I_{i+1}) = w(I_i) = \frac{w(I_{i-1})}{N_{i-1}} \le \frac{C_\xi}{N_{i-1}}$, and $N_{i+1} = \sqrt{N_i} = \sqrt{N_{i-1}^2} = N_{i-1}$. By Corollary 14, all further AQIR calls succeed. □