Error-Bounds for a Zero of a Polynomial

§1 Abstract & Introduction

A computed approximation x to a zero z of a polynomial P of degree n is given along with values of $P(x)$ and its first two derivatives $P'(x)$ and $P''(x)$. How well does x approximate z ? Exhibited below are two upper bounds for the error $|x - z|$. One comes as a classical inequality found by E. Laguerre late in the 19th century. The other inequality is tighter, sometimes much tighter, but costs more to compute. Both will be proved and then compared numerically.

§2 Proofs of Two Inequalities

Let the zeros of P be $z_1, z_2, z_3, \ldots$ and z_n. They may be complex, and some may be repeated. Evidently $P(x) = \alpha\cdot\Pi_k (x - z_k)$ for some nonzero constant α. Though the zeros z_k may be unknown, from given values $P'(x)$ and $P(x)$ we can compute

$$Q(x) := P'(x)/P(x) = \Sigma_k 1/(x - z_k) .$$

This figures in Newton's iteration, which would replace approximation x by $x - 1/Q(x)$, a better approximation if x is close enough to a zero of P. Let z denote a zero nearest x and let Σ'_k sum over all the other zeros z_k (including any others that may coincide with z). Then

$$|(x - z)\cdot Q(x) - 1| = |\Sigma_k (x - z)/(x - z_k) - 1| = |\Sigma'_k (x - z)/(x - z_k)| \le n - 1$$

because $|x - z| \le |x - z_k|$. Dividing the previous inequality by $|Q(x)|$ turns it into …

Laguerre's Inequality: $|x - 1/Q(x) - z| \le (n-1)/|Q(x)|$ (*Oeuvres* p. 60)

This inequality becomes equality just when x is equidistant from two zeros of P one of which has multiplicity $n-1$, or when P has one zero of multiplicity n. A slightly weaker version,

$$|x - z| \le n/|Q(x)| = n\cdot|P(x)|/|P'(x)| \qquad [£]$$

is the version of Laguerre's inequality that is usually remembered. It becomes equality just when all n zeros of P are coincident. But the inequality is useless if $Q(x) \approx 0$ when x is too close to a zero of P', which has $n - 1$ zeros distributed within the convex hull of the zeros z_k of P. In particular, Laguerre's inequality [£] can be excessively pessimistic when some of those zeros of P' are too near clusters of zeros of P though different from them. We seek a remedy, if one exists, for this shortcoming.

Construe $Q(x)/n = (\Sigma_k 1/(x - z_k))/n$ as the *mean* or average of all reciprocals $1/(x - z_k)$. Their complex *variance* is $\delta^2(x)/n^2$, computable from n, $P(x)$, $P'(x)$ and $P''(x)$ thus:

$$\delta^2(x) := (n-1)\cdot Q(x)^2 - n\cdot P''(x)/P(x) = n\cdot\Sigma_k \left(1/(x - z_k) - Q(x)/n \right)^2 .$$

Laguerre noticed that the polynomial $P(x)^2\cdot\delta^2(x)$ has degree at most $2n-4$, not $2n-2$, and must be nonnegative when all the zeros z_k are real, which was the case that interested him. The same polynomial, but only for $n = 3$, is buried in Curt McMullen's Proposition 1.2 where he found in 1987 that Newton's iteration applied to solve " $P(z)/(P(z)^2\cdot\delta^2(z)) = 0$ " converges cubically to a zero of any given cubic P from almost every starting point in the complex plane.

Now an identity familiar to statisticians, namely that

$$\delta^2(x)/n = \Sigma_k \left(1/(x - z_k) - Q(x)/n \right)^2 = \Sigma_k 1/(x - z_k)^2 - 2(Q(x)/n)\cdot\Sigma_k 1/(x - z_k) + n\cdot(Q(x)/n)^2$$
$$= \Sigma_k 1/(x - z_k)^2 - Q(x)^2/n \ ,$$

will be revisited to over-estimate $|\delta^2(x)|$ thus:

$$|\delta^2(x)| = \left| \Sigma_k \left(n/(x - z_k) - Q(x) \right)^2 \right|/n \le \Sigma_k | n/(x - z_k) - Q(x) |^2/n$$
$$= n\cdot\Sigma_k 1/| x - z_k |^2 - 2\mathrm{Re}\{\overline{Q(x)}\cdot\Sigma_k 1/(x - z_k)\} + |Q(x)|^2$$
$$\le n^2/|x - z|^2 - |Q(x)|^2 \quad \text{since} \quad |x - z| \le | x - z_k | \ .$$

This inequality was buried in my 1967 paper cited below. When rearranged it tells us that

$$|x - z| \le n/\sqrt{(|Q(x)|^2 + |\delta^2(x)|)}$$
$$= n\cdot|P(x)|/\sqrt{(|P'(x)|^2 + \left|(n-1)\cdot P'(x)^2 - n\cdot P(x)\cdot P''(x)\right|)} \ . \qquad [\mathcal{K}]$$

This inequality becomes equality just when x is equidistant from two zeros of P of the same multiplicity $n/2$, or when P has only one zero of multiplicity n. The inequality's right-hand side becomes ∞ only when x is a multiple zero of P' but not of P; it hardly ever happens.

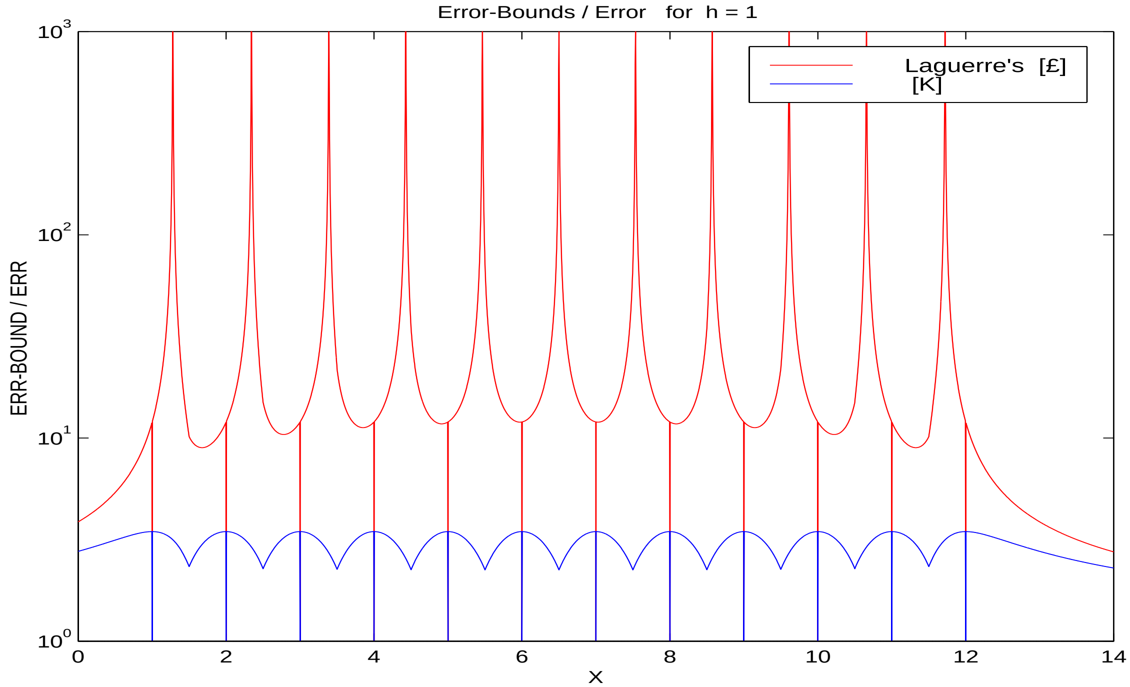
§3 How do Inequalities [£] and [$\mathcal{K}$] Compare?

As an error-bound, [$\mathcal{K}$] is always no bigger than Laguerre's [£] but costs more to compute. Typically, in the punctured neighborhood of a simple zero z separated well from all others, $[\mathcal{K}]/[£] \approx 1/\sqrt{n}$, which is not a great improvement, as error-bounds go, unless degree n is huge. We should be concerned only with situations wherein an error-bound is at least a few orders of magnitude bigger than a very small error. If such a situation arises too often, that error-bound is likely to be disregarded like *The Little Boy Who Cried* "Wolf**!**" Only if [£] encounters such situations far more often than [$\mathcal{K}$] does can [$\mathcal{K}$] be worth its extra cost.

Aside from being zero at zeros z_k of P, error-bounds [$\mathcal{K}$] and [£] are the same when x is one of the $2n-4$ zeros of $P(x)^2\cdot\delta^2(x) = (n-1)\cdot P'(x)^2 - n\cdot P(x)\cdot P''(x)$. And $[£]/[\mathcal{K}] \to 1$ as $x \to \infty$. Therefore an undesirable situation with $[£] >> [\mathcal{K}]$ can arise only in a region surrounding a zero ζ of P' that is not a zero z of P but is near one of a cluster. This follows from the equation $Q(\zeta) = 0$, which implies $\zeta - z = -1/\left(\Sigma'_k 1/(\zeta - z_k) \right)$, whence follows that $|\zeta - z|$ can be very small only if at least one other $|\zeta - z_k|$ is very small too (or else n is enormous). Among these undesirable situations are sometimes some with [$\mathcal{K}$] also excessively big. These rare situations can arise only in a region surrounding a *multiple* zero ζ of P' that is not a zero z of P but is near one of a cluster, as follows from the equations $Q(\zeta) = 0 = \delta^2(\zeta)$.

To compare [£] with [$\mathcal{K}$], let us plot Error-Bound/|Error| for both bounds and for each of three polynomials P chosen to exhibit undesirable situations. The plots' downward spikes occur at zeros of P, one or two at $z = \pm h$. Upward spikes would go to ∞ at zeros of Q but for cut-offs at the tops of the graphs. Clusters of zeros can be unobvious; for instance, the twelve zeros of $b(x) := x^{12} - 1$ on the unit circle are not clustered, though separated by less than 0.51784, but some of the twelve integer zeros of $e(x) := \prod_{1\le k\le 12} (x - k)$ are clustered, as we shall see.

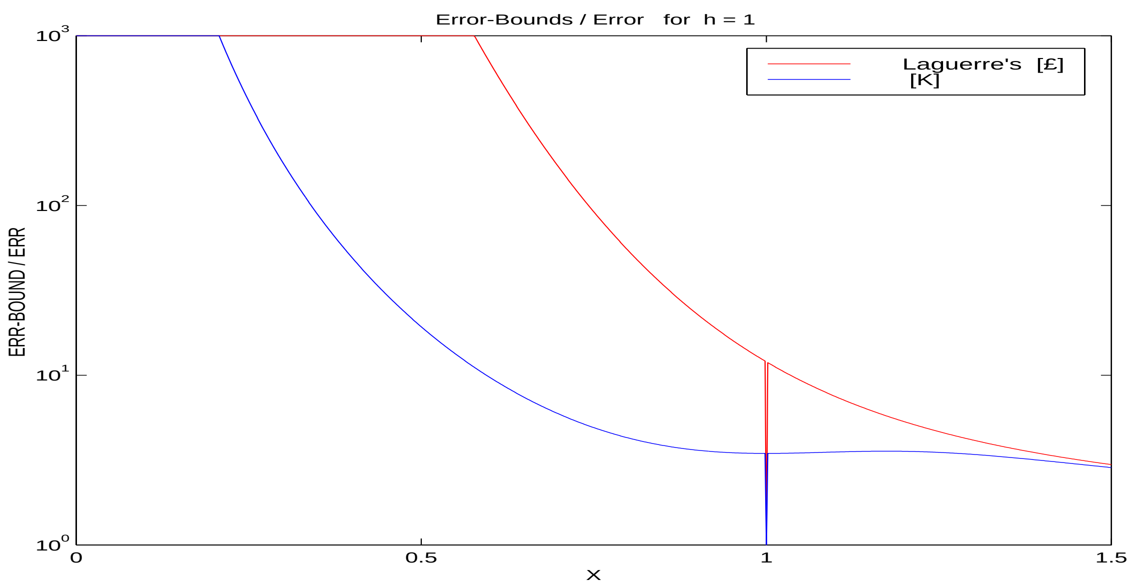
Error-Bounds/|Error| for Polynomial $e(x) := \prod_{1\le k\le 12} (x - k)$



Here $[\mathcal{K}]/|x-z|$ never rises above $\sqrt{12} \approx 3.4641$; but $[£]/|x-z|$ peaks up to ∞ between each pair of $e(x)$'s adjacent zeros. This intimation of clustering is corroborated by the sensitivity of some of the zeros to tiny perturbations of the polynomial's coefficients. For instance, nearby polynomial $\hat{e}(x) := e(x) - e(-x)\cdot\tau$ with tiny $\tau \approx 5.600278/10^{10}$ has coefficients differing from those of $e(x)$ in the tenth sig.dec.; this perturbation changes zeros 8 and 9 of e into a double zero of $\hat{e}$ near 8.4835138 . Generally, a polynomial's "clustered zeros" are less well defined than the smallest perturbation that causes some of them to coalesce, though it is hard to compute.

• • • • • • • • • • • • • • • • •

Error-Bounds/|Error| for Polynomial $b(x) := x^{12} - 1$

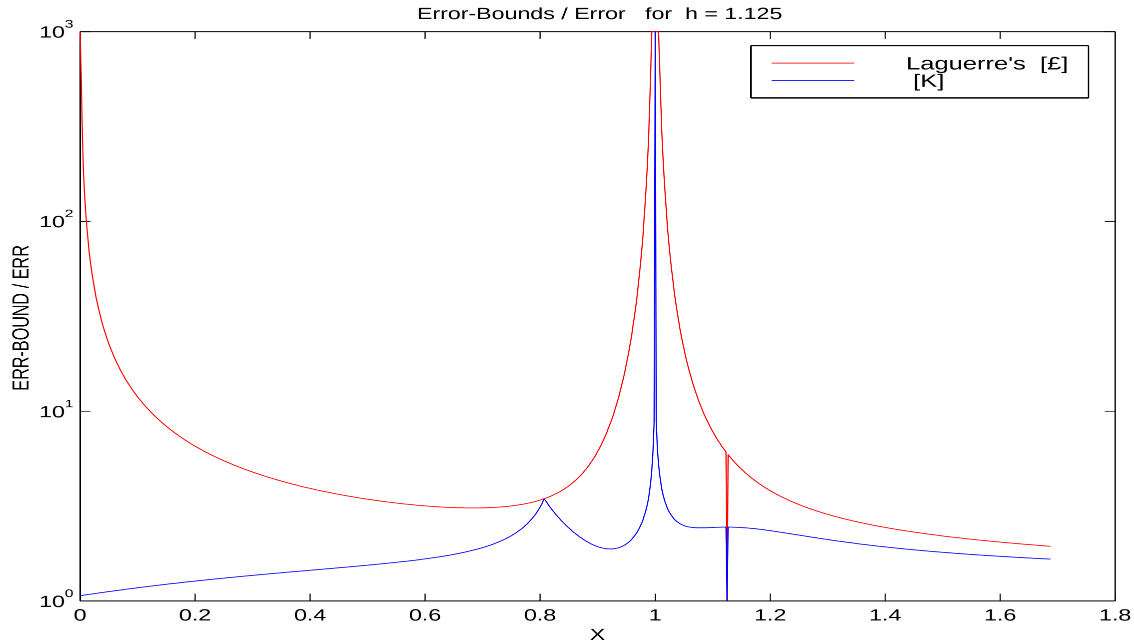


The zeros of $b(x) := x^{12} - 1$ are not clustered at all. Their multiplicities can be increased by a perturbation of nonzero coefficients only if the constant term “ 1 ” is replaced by “ 0 ”. Now both $[£]/|x-z|$ and $[\mathcal{K}]/|x-z|$ spike up to ∞ as $x \to 0$, but $[£]/|x-z|$ has a much wider spike: Roughly, for any big exponent M, and for z the zero of b nearest x,

$$[£]/|x-z| > 10^M \text{ when } |x| < 1/10^{M/11} \ ; \qquad [\mathcal{K}]/|x-z| > 10^M \text{ when } |x| < 1/(11^{1/10} \cdot 10^{M/5}) \ .$$

• • • • • • • • • • • • • • • • • •

Error-Bounds/Error for Polynomial $g(x) := (x^2-h^2)(x^4 + (h^2-3)(x^2+h^2) + 3$ @ $h = 9/8$



The zeros of g are $z = \pm 1.125$ and $z \approx \pm(0.939246 + 0.122459 \cdot \imath$ and its complex conjugate) . The derivative g' has a simple zero $\zeta = 0$, and two double zeros $\zeta = \pm 1$ where both $[£]$ and $[\mathcal{K}]$ spike up to ∞. Here $[\mathcal{K}]$’s spike is so much narrower than $[£]$’s because, as $x \to 1$, asymptotically $[£] \approx 0.0046854/|x-1|^2 >> [\mathcal{K}] \approx 0.048401/\sqrt{|x-1|}$. A pimple on the graph of $[\mathcal{K}]$/Error occurs at $x \approx 0.8071784$ where $\delta^2(x) = 0$ whence $[\mathcal{K}]/|x-z| = [£]/|x-z| \approx 3.4586$.

This example’s spikes are typical in so far as $[£]$ often spikes up to ∞ near clustered zeros of a polynomial. $[\mathcal{K}]$ hardly ever spikes up to ∞; and when it does its spike is so narrow that x is unlikely to fall onto it during the search for a zero by some iterative method. That this behavior is typical will be confirmed by the following asymptotic estimates.

§4 Summary of Asymptotic Behaviors

Here x is an approximation to a zero z of a polynomial P of degree n , and $[\pounds]$ and $[\mathcal{K}]$ are error-bounds computed at x from formulas exhibited above.

First let z be an n-tuple zero of P ; then $[\pounds] = [\mathcal{K}] = |x - z|$. This does not happen often.

Next let z be a zero of P of positive multiplicity $m < n$. As $x \to z$,

$$[\pounds]/|x-z| \to n/m \qquad > \qquad [\mathcal{K}]/|x-z| \to \sqrt{n/m} \; .$$

This case, with m = 1 , occurs most often.

Next let ζ be a simple zero of P' , but not a zero of P . As $x \to \zeta$,

$$[\pounds] \approx n\cdot|P(\zeta)/P''(\zeta)|/|x-\zeta| \to \infty \quad \gg \quad [\mathcal{K}] \to \sqrt{n\cdot|P(\zeta)/P''(\zeta)|} \; .$$

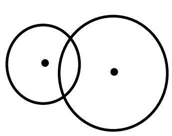
Finally let ζ be a zero of P' of multiplicity $m \ge 2$, but not a zero of P . As $x \to \zeta$,

$$[\pounds] \approx n\cdot m!\cdot|P(\zeta)/P^{[m+1]}(\zeta)|/|x-\zeta|^m \quad \gg \quad [\mathcal{K}] \approx \sqrt{\left(n\cdot(m-1)!\cdot|P(\zeta)/P^{[m+1]}(\zeta)|\,/|x-\zeta|^{m-1} \right)} \; .$$

In this rare case, as both $[\pounds]$ and $[\mathcal{K}]$ spike up to ∞ , $|x-\zeta|\cdot[\pounds]/[\mathcal{K}]^2 \to m$ in the spike. This is why $[\mathcal{K}]$’s spike is so much narrower than $[\pounds]$’s and why, if $[\mathcal{K}]$ has a spike, x is so unlikely to fall onto it.

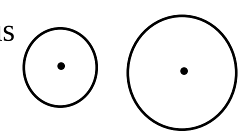
§5 Resolution of Ambiguity

Suppose two estimates x_1 and x_2 of zeros of P have been found. Do x_1 and x_2 approximate different zeros, or the same one? Sometimes this question can be answered by error-bounds:



Ambiguous
overlapping
error-bounds

Unambiguous
disjoint
error-bounds



Error-bounds $[\pounds]_1$ for x_1 and $[\pounds]_2$ for x_2 can be computed. They answer the question above unambiguously if $[\pounds]_1 + [\pounds]_2 < |x_2 - x_1|$, in which case x_1 and x_2 aproximate different zeros.

Otherwise $[\mathcal{K}]_1$ for x_1 and $[\mathcal{K}]_2$ for x_2 can be computed. They answer the question above unambiguously if $[\mathcal{K}]_1 + [\mathcal{K}]_2 < |x_2 - x_1|$, in which case x_1 and x_2 approximate different zeros.

Otherwise something else must be done to answer the question. If the computed values of P" , P' and especially P are accurate enough, they can be used to improve the estimates x_1 and x_2 by using, say, Laguerre’s iteration formula cited in my 1967 paper. On very rare occasions, however, attempts to improve clustered estimates actually worsen them unless extra-precise arithmetic is employed.

§6 Conclusions

[£] can spike up to ∞ in the neighborhood of clusters of zeros. When both [£] and [$\mathcal{K}$] spike up, which can happen only very rarely, [$\mathcal{K}$]'s spike is so much narrower than [£]'s that [$\mathcal{K}$]'s is unlikely to be encountered.

Though it costs more than Laguerre's [£], error-bound [$\mathcal{K}$] is worth computing when it is likely to be far smaller than [£], as is likely during the numerical computation of a polynomial's zero that belongs to a cluster of zeros hypersensitive to small perturbations. But a cluster is no more obvious in advance than an unmarked minefield, so it makes sense to compute [£] first and then, if it seems too big, compute [$\mathcal{K}$] .

§7 Citations

In chronological order:

*Oeuvres de Laguerre, Tome **1*** (1898) ed. by Ch. Hermite, H. Poincaré & E. Rouché, Gauthier-Villars, Paris. Originally, Laguerre published his inequality on p. 60 in 1878,

W. Kahan (1967) "Laguerre's Method and a Circle which Contains At Least One Zero of a Polynomial" pp. 474-482 in *SIAM J. Numer. Anal.* **4**.

Curt McMullen (1987) "Families of rational maps and iterative root-finding algorithms" pp. 467-493 in *Annals of Mathematics* **125**.