Lagrange and the Solution of Numerical Equations

Reinhard Laubenbacher
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
reinhard@nmsu.edu

Gary McGrath
Department of Mathematics
Pittsburg State University
Pittsburg, KS 66762, USA
gmcgrath@mail.pittstate.edu

David Pengelley
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
davidp@nmsu.edu

Abstract

In 1798 J.-L. Lagrange published an extensive book on the solution of numerical equations. Lagrange had developed a general systematic algorithm for detecting, isolating, and approximating all real and complex roots of a polynomial equation with real coefficients, with arbitrary precision. In contrast to Newton's Method, Lagrange's algorithm is guaranteed to converge. We discuss some lesser known aspects of Lagrange's work. In particular, some of his powerful ideas and techniques adumbrated methods developed much later in geometry and abstract algebra, such as Möbius transformations and quotient rings of polynomial rings. We also show that his techniques included accelerating both the convergence and calculation of his continued fraction expansions of the roots.

1 Introduction.

Joseph–Louis Lagrange's work on the roots of algebraic equations extended over a significant period of his professional life [3]. He saw the subject as divided into two distinct parts, one concerned with numerical, the other with algebraic, solutions:

> The solution of every determinate problem, in the final analysis, reduces to the solution of one or more equations, whose coefficients are given in numbers, and which one can call *numerical equations*. ...
>
> One should distinguish the solution of numerical equations from that which one calls in Algebra the general solution of equations. The first is, properly speaking, an arithmetic operation, justified, indeed, on the general principles of the theory of equations, but whose results are only numbers, where one no longer recognizes the first numbers that served as the rudiments, and which retain no trace of the particular different operations which produced them. The extraction of square and cube roots is the simplest operation of this sort: it is the solution of numerical equations of the second and third degree, in which all the intermediate terms are lacking. ... [7, pp. 13ff]

In 1769/1770 Lagrange published two large papers (see below) on numerical solutions of equations, presenting theoretically complete algorithmic methods for finding all solutions of real numerical polynomial equations. With regard to algebraic solutions, on the other hand, in 1770 Lagrange published a long memoir entitled *Réflexions sur la Résolution Algébrique des Equations*, in which he surveyed and tried to extend existing work on the old problem of solving algebraic equations by formulas involving radicals, and expressed and supported his view that the situation beyond degree four looked unpromising. This latter manuscript represents an important milestone in setting the stage for the later work of Abel and Galois, whose new methods and discoveries confirmed the truth of Lagrange's pessimistic assessment about algebraic solutions, thus finishing a long story in the history of mathematics.

Lagrange's work on solutions of numerical equations, however, continued strongly for decades longer. In 1795 he lectured extensively on this topic at the Ecole Polytechnique, and published his lecture notes [6]. In 1798 he published the *Traité de la Résolution des Equations Numériques de Tous les Degrees* (revised in 1808 [7]), which essentially republished his large earlier papers from 1769/1770 in six chapters, and added voluminous additional Notes. This book presents an exhaustive collection of methods for classifying, isolating, and approximating real and complex roots of equations. The *Traité* has been discussed in some detail in [5], as well as in [3, 4, 9]. But his basic approximation algorithm using continued fractions is now all but forgotten, even though it was accorded a separate section in H. Weber's *Lehrbuch der Algebra* [10, pp. 445–447], first published in 1894.

There are many ideas and techniques contained in the *Traité* that have not been discussed in the secondary literature, and which we consider very significant for historical as well as mathematical reasons. We will explore here the significance of three particular features of the *Traité*. In Section 2 we outline Lagrange's principal algorithm. In Section 3 we present his first two methods for finding a lower bound on the differences of the real roots of a given real polynomial by computing the coefficients of an auxiliary equation, the "equation of differences", whose roots are the differences of all the distinct roots of the original equation. This is a significant first step in any approximation algorithm that allows the isolation of roots. Lagrange gives a very interesting algorithm using Newton's formulas for the power sums of the roots. In Section 4 we analyze the other two methods Lagrange developed much later, published in 1795/1798, to find such a lower bound more efficiently, without actually computing the coefficients of the equation of differences. Our interest is particularly in the historical significance of the final version of his technique. From a modern point of view he clearly decides to work, to great practical advantage, in the quotient ring of $\mathbf{R}[x]$ by the ideal generated by the original polynomial, one of the earliest instances of this idea. In Section 5 we discuss two improvements Lagrange made to his approximation algorithm for the roots by continued fractions. One accelerates convergence by allowing non-simple continued fractions. The other explicitly uses Möbius transformations (but long before Möbius) to track the iterative process, thereby speeding calculation by eliminating the need for successive expansions of equations by substitutions similar to those in Newton's Method, and profiting from the particular form of iterated Möbius transformations. Throughout the paper we use Lagrange's notation as much as possible.

2 The Algorithm.

In Chapters I–IV of the *Traité* Lagrange presents his algorithm for computing all roots of a polynomial equation

$$f(x) = x^m - Ax^{m-1} + Bx^{m-2} - Cx^{m-3} + \cdots = 0, \tag{1}$$

where the coefficients $A, B, C, \ldots$ are real numbers. We now outline this algorithm as it pertains to the real roots. To begin with, Lagrange assumes that (1) has only simple roots, which can be achieved by dividing f by $\gcd(f, f')$. Furthermore, his algorithm focuses on the positive roots, since the negative roots are just the positive roots of $f(-x)$.

The essence of the algorithm divides into three key steps:

1. Find a lower bound $\Delta > 0$ on the unsigned differences between the real roots of (1). Lagrange proposes four methods for finding Δ, with and without explicit computation of the auxiliary *equation of differences* whose roots are the differences between all ordered pairs of distinct roots of (1). (Lagrange's algorithm simultaneously detects and finds multiple roots, so we assume all roots are simple.)

2. Use Δ, along with possible rescaling of f, and techniques (e.g., of Newton/Maclaurin) that bound the magnitude of possible roots, to identify integers p such that each interval $(p, p+1)$ contains a unique real root, and all real roots are so captured. (Lagrange detects and dispenses here with integer roots, and discusses what to do if there is more than one root between consecutive integers, e.g., by rescaling f.)

3. Use an alternative to Newton's algorithm to approximate the root in $(p, p+1)$. Lagrange's approximation is by continued fractions, which has crucial advantages over Newton's method: it always converges to a root, and it provides an error estimate.

3 Isolating the Real Roots.

The first step in Lagrange's algorithm is to find a lower bound on the differences of the real roots of (1). This is done by constructing an auxiliary equation, the equation of differences, whose roots are the differences of all ordered pairs of distinct roots of (1). Any lower bound on the positive real roots of this equation will provide a Δ such that any interval of length Δ

will contain at most one real root of (1). Obviously, it is desirable to find the largest possible lower bound, since it reduces the number of intervals to be checked for roots. Lagrange presented two methods for finding the equation of differences in his papers of 1769/1770, republished in [7, Ch. 1]. The second method, building on the first, uses a clever double application of symmetric functions and Newton's formulas.

In order to derive the difference equation, let x be any fixed root of (1), x' any root, and $u = x' - x$. Substituting $x' = x + u$ into (1), Lagrange obtains an equation

$$X + Yu + Zu^2 + Vu^3 + \cdots + u^m = 0$$

in u, whose coefficients, functions of x, are $X = f$, $Y = f'$, $Z = \frac{1}{2}f''$, etc. Since $f(x) = 0$, the constant term vanishes, and dividing by u we obtain

$$Y + Zu + Vu^2 + \cdots + u^{m-1} = 0. \qquad (2)$$

The roots of this equation are precisely all the differences between x and the other roots of (1).

If we now apply Bezout's elimination method [1, 11] to eliminate x from Equations (1) and (2), we obtain an equation in u whose roots are the differences of all ordered pairs of distinct roots of (1). Since for every root $\alpha - \beta$ of this equation, $\beta - \alpha$ is also a root, we see that this is an equation in u^2. Substituting $v = u^2$, we obtain the desired equation of differences, which has degree $n = m(m-1)/2$:

$$D(v) = v^n - av^{n-1} + bv^{n-2} - cv^{n-3} + \cdots = 0. \qquad (3)$$

As Lagrange says in the *Traité* [7, p. 25], this method can sometimes be very laborious. He goes on to propose another way of deriving the same equation of differences. For this second method, Lagrange uses Newton's formulas for the sums of powers of the roots of (1), in terms of its coefficients. Let $\alpha, \beta, \gamma, \ldots$ be the roots of (1). Let

$$\begin{aligned} A_1 &= \alpha + \beta + \gamma + \cdots, \\ A_2 &= \alpha^2 + \beta^2 + \gamma^2 + \cdots, \\ A_3 &= \alpha^3 + \beta^3 + \gamma^3 + \cdots, \\ &\cdots . \end{aligned}$$

Then Newton's formulas are

$$\begin{aligned}
A_1 &= A, \\
A_2 &= AA_1 - 2B, \\
A_3 &= AA_2 - BA_1 + 3C, \\
A_4 &= AA_3 - BA_2 + CA_1 - 4D \\
&\cdots .
\end{aligned}$$

Recall that the roots of equation (3) are the squares of differences of roots of (1). Let

$$\begin{aligned}
a_1 &= (\alpha - \beta)^2 + (\alpha - \gamma)^2 + (\beta - \gamma)^2 + \cdots, \\
a_2 &= (\alpha - \beta)^4 + (\alpha - \gamma)^4 + \cdots, \\
a_3 &= (\alpha - \beta)^6 + \cdots \\
&\cdots .
\end{aligned}$$

Lagrange observes (and elaborates in Note III) that one can express the a_i in terms of the A_i rather simply, using binomial coefficients:

$$\begin{aligned}
a_1 &= mA_2 - 2\frac{A_1^2}{2}, \\
a_2 &= mA_4 - 4A_1A_3 + 6\frac{A_2^2}{2}, \\
a_3 &= mA_6 - 6A_1A_5 + 15A_2A_4 - 20\frac{A_3^2}{2}, \\
&\cdots .
\end{aligned}$$

And finally, he finds the coefficients of the equation of differences (3) in terms of the a_i, by reversing Newton's formulas:

$$\begin{aligned}
a &= a_1, \\
b &= \frac{aa_1 - a_2}{2}, \\
c &= \frac{ba_1 - aa_2 + a_3}{3}, \\
&\cdots .
\end{aligned}$$

To illustrate this procedure, consider the example

$$f(x) = x^3 - 2x - 5 = 0,$$

the equation Newton used to illustrate his method for approximating roots, and which Lagrange uses as well [7, p. 51f]. We obtain

$$A_1 = 0,\ A_2 = 4,\ A_3 = 15,\ A_4 = 8,\ A_5 = 50,\ A_6 = 91.$$

In turn, we compute

$$a_1 = 12,\ a_2 = 72,\ a_3 = -1497.$$

Finally, we find the equation of differences:

$$D(v) = v^3 - 12v^2 + 36v + 643 = 0.$$

Recall that in general our goal is to find a lower bound Δ on the real roots of this equation of differences. Replacing v by $1/v$ we are reduced to finding an upper bound on the roots of the resulting equation. One way to proceed is to use Maclaurin's method, which gives such an upper bound as the largest absolute value of the negative coefficients, increased by one [7, pp. 31–32].

4 An Abstract Algebra Method for Finding Δ.

As Lagrange observes in Note IV of the *Traité* [7, p. 146], either method above for finding the equation $D(v) = 0$ is quite laborious, since in general its degree is quite high relative to the degree of (1). In this section we discuss two methods he developed for finding Δ without explicitly computing $D(v)$. The second builds from the first, as happened with the previous two methods, and the final version is impressive, allowing him to work with polynomials of degree less than the degree of f, and to find Δ much more effectively. Lagrange was pleased with this final approach, as one can tell from his comment, added to the Introduction of the second edition of the *Traité* in 1808:

> Since the first edition of this work in 1798, different methods for resolving numerical equations have appeared; but the rigorous solution of the problem has remained at the same place to which I carried it, and up to now one has found nothing that could dispense with the search for a bound less than the smallest difference between the roots, or which would be preferable to the means given in Note IV for facilitating this search. [7, p. 18]

A first version of Lagrange's new method for finding Δ appears in his lectures from 1795 [6, pp. 114ff], and is augmented to its final form in Note IV of the *Traité* of 1798 [7, pp. 146ff]. The idea behind his final approach deserves our attention, since it amounts to working in the quotient ring obtained from the polynomial ring $\mathbf{R}[x]$ by dividing out the ideal generated by f. Even though the techniques of modular arithmetic were in the air at the time Note IV was published, this method appears to be one of the earliest examples of implicit use of such algebraic congruences. Algebraic congruences were first treated by A. Cauchy in his 1847 memoir *Memoir on the Theory of Algebraic Equivalences* [2], according to [8, p. 244]. We describe Lagrange's method largely using his notation in the *Traité*.

As with the earlier methods, from a fixed root x and arbitrary root $x' = x + u$, Lagrange derives an equation of degree $m - 1$ in u:

$$Y + Zu + Vu^2 + \cdots + u^{m-1} = 0,$$

where $X = f$, $Y = f'$, $Z = \frac{1}{2}f''$, $V = \frac{1}{3}f'''$, etc., all evaluated at x. Its roots are the differences between x and all the other roots of (1). Setting

$$u = \frac{1}{i},$$

and rewriting the resulting equation, he obtains an equation

$$i^{m-1} + \frac{Z}{Y}i^{m-2} + \frac{V}{Y}i^{m-3} + \cdots + \frac{1}{Y} = 0. \tag{4}$$

Note that the coefficients of this equation in i depend on the root x. Our problem is now reduced to finding an upper bound $L > 0$ for the roots of all the equations we obtain by substituting the roots of (1) into (4). The desired lower bound for the differences between the roots of (1) is then $1/L$. Lagrange recognized that a big difficulty in carrying out this program arises from the appearance of x in the denominator Y of the coefficients, since he has no *a priori* means of finding lower bounds for the magnitude of Y at the roots of (1). In the lectures published in 1795 he proposed achieving this by an elimination approach, eliminating x between the equation for Y in terms of x and the original equation $X = 0$. This would provide an equation satisfied by different values of Y at the roots of $X = 0$, from which one could obtain a lower bound on these values.

In Note IV of the *Traité*, however, Lagrange admits that this is very laborious [7, pp. 148–149], and says

> Since then, I have given thought that one could always eliminate the unknown x of the polynomial Y by multiplying it by a suitable polynomial of the same degree $m-1$, and make all the powers of x higher than x^{m-1} disappear, by means of the equation $X = 0$. [7, p. 149]

Thus begins Lagrange's excursion into computing in the ring $\mathbf{R}[x]/(f)$, in order to get around the problem Y poses in his denominators. Since we are going to substitute roots of equation (1) into the coefficients of (4), Lagrange reasons that we may as well use equation (1) to express x^m in terms of the other monomials in f, likewise for higher powers of x^m. That is, he proposes to consider Equation (4) as an equation in the ring $\mathbf{R}[x]/(f)$ rather than $\mathbf{R}[x]$.

Lagrange explains how to solve a system of linear equations to find the polynomial ξ of degree $m-1$ such that $\xi Y \equiv K = \gcd(X, Y)$ modulo the relation $X = 0$, i.e., in $\mathbf{R}[x]/(f)$. In the language of abstract algebra, since $\mathbf{R}[x]$ is a Euclidean domain, we can find polynomials ξ and g in $\mathbf{R}[x]$ such that $\xi Y + gX = \gcd(X, Y)$, yielding $\xi Y \equiv K = \gcd(X, Y)$ in $\mathbf{R}[x]/(f)$. If f has only simple roots, then $K = \gcd(X, Y)$ is a nonzero constant, and equation (4) becomes

$$i^{m-1} + \frac{\xi Z}{K} i^{m-2} + \frac{\xi V}{K} i^{m-3} + \cdots + \frac{\xi}{K} = 0 \tag{5}$$

in $\mathbf{R}[x]/(f)$. Then Lagrange can appeal to familiar techniques, based on methods of Newton and Maclaurin, to obtain bounds on the original roots of (1), and thereby on the roots of (5), whose coefficients, essentially the derivatives of f, are known polynomials evaluated at these roots, and thus finally the desired bound L. He illustrates the entire procedure on an example, calculated in the quotient ring $\mathbf{R}[x]/(f)$.

Lagrange ends Note IV by contrasting the algorithmic efficiency of his approaches, noting that the number of operations required for this new technique grows only as the degree of the original equation, whereas the number of operations required to actually calculate the coefficients of the equation of differences grows as the square of the original degree.

5 An Improved Continued Fraction Approximation and Möbius Transformations.

The basic idea of Lagrange's approximation method for the roots of (1) is quite easy to explain [7, Ch. 3]. Suppose we have located an interval $(p, p+1)$, p an integer, that contains a unique root. In Newton's Method, x then gets replaced by $p + y$ in (1), whereas Lagrange substitutes $p + \frac{1}{y}$ and then expands the resulting equation to obtain a new equation in y. Since $y > 1$, this new equation has a root that is greater than 1. Now, using the same procedure, he finds an integer $q \geq 1$ closest to, while not exceeding, that root and substitutes $q + \frac{1}{z}$ for y. Similarly, $z = r + \frac{1}{u}$, and so on. In this way he obtains a sequence of integers $p, q, r, \ldots \geq 1$ such that

$$x = p + \frac{1}{y},\ y = q + \frac{1}{z},\ z = r + \frac{1}{u}, \ldots \quad .$$

Successive back substitution gives a continued fraction

$$x = p + \cfrac{1}{q + \cfrac{1}{r + \cfrac{1}{s + \frac{1}{\cdots}}}},$$

whose convergents approximate the root x arbitrarily closely. Since it is a simple continued fraction, it also provides an error estimate at each step, because consecutive convergents lie on opposite sides of the limit, bracketing it.

In an effort to improve the speed of convergence of the continued fraction to the root [7, Ch. 6, Art. III, p. 101f], Lagrange introduces more general continued fractions. The basic idea is that, if the root of (1) is closer to $p + 1$ than to p, he wants to substitute $(p + 1) - \frac{1}{y}$ instead of $p + \frac{1}{y}$. This results in non-simple continued fractions, that allow negative denominators. The denominators can now always be chosen strictly greater than one in absolute value, thus producing faster convergence. They do not, however, have the bracketing property of the simple continued fractions.

Lagrange embarks on a lengthy investigation of their convergence properties. In the process he introduces what we now call Möbius transformations of the various convergents in the algorithm. Specifically, after converting the convergents into fractions, he states that

> We have found in general [...] that, if $\frac{\bar{\omega}}{\bar{\omega}'}$ and $\frac{\rho}{\rho'}$ are two consec-

utive fractions converging toward the value of x, one has

$$x = \frac{\rho t + \bar{\omega}}{\rho' t + \bar{\omega}'}.$$

Hence, if one substitutes this expression for x into the equation in x whose root one seeks, then one has a transformed equation in t, which is necessarily the same as that one obtained by making the successive substitutions of $p + \frac{1}{y}$ in the place of x, of $q + \frac{1}{z}$ in place of y, ...; and to get the following fraction $\frac{\sigma}{\sigma'}$, it is enough to find the integer closest to t, which we call k, so that

$$\sigma = k\rho + \bar{\omega},\ \sigma' = k\rho' + \bar{\omega}'. \qquad \text{[7, Ch. 6, Art. IV, p. 115]}$$

These latter observations represent a significant improvement for implementing his approximation algorithm, allowing one to completely avoid the laborious simplification of the successive equations in y, z, ... obtained by his substitutions. All approximations can be found using the original equation (1). Specifically, let p be the closest integer to the root (actually, the integer on either side of it may be chosen, if allowing non-simple continued fractions), found by looking for a sign change in the value of f at consecutive nonzero integers p, $p+1$. Of course the sign change does not indicate which of p or $p+1$ to select for fastest convergence (Lagrange explained earlier [7, p. 109] how to make sure one's choice always produces fastest convergence), but either choice will produce a solution (Lagrange also developed other much more sophisticated means of finding p than looking for sign changes [7, Ch. 6, Art. IV]). Let $x = p + \frac{1}{y}$, $|y| \geq 1$, or choose $|y| > 1$ for fastest convergence. Lagrange's original algorithm now makes the substitution to obtain an equation

$$f_1 = y^m + ay^{m-1} + \cdots = 0.$$

Then it finds an integer closest to the root y by again looking for sign changes of f_1 at consecutive integers.

Instead, Lagrange's improved method proceeds as follows. Rewrite the substitution as a Möbius transformation

$$x = p + \frac{1}{y} = \frac{py + 1}{y + 0} = \phi(y).$$

Now, rather than calculating f_1 and looking for sign changes at successive integers, one can search for sign changes of $f \circ \phi$ at successive integers. This is much easier in practice, since it is just the original polynomial evaluated at the Möbius transformation of integers, which is not hard to simplify sufficiently to look for sign changes. Having found the value of q, the original algorithm would now substitute $y = q + \frac{1}{z}$ into f_1 and iterate the process to obtain f_2. Instead, having found q directly from f, we first express this substitution as

$$y = \frac{qz+1}{z+0} = \psi(z),$$

and compose Möbius transformations to obtain

$$x = \chi(z) = \phi\psi(z) = \frac{(pq+1)z+p}{qz+1}.$$

Note that $\frac{p}{1}$ and $\frac{pq+1}{q}$ are the first two convergents of the continued fraction expansion of x.

Instead of calculating and searching $f_2(z)$ for a sign change at consecutive integers, we can simply search $f \circ \chi$ for a sign change at successive integers, a similarly easy task as for $f \circ \phi$, since the composition of Möbius transformations is again of the same type, and trivially computed from the previous convergents. It appears that this improvement of Lagrange's algorithm has not been investigated before, even though it is analogous to the improvement that Raphson made [4, p. 175f] to Newton's Method.

6 Conclusion.

We have outlined several aspects of Lagrange's work on the solution of numerical equations that have not received much attention in the secondary literature. Yet they are significant for historical as well as mathematical reasons.

Lagrange uses two powerful techniques that adumbrate much later developments in geometry and algebra, namely Möbius transformations and algebra in quotient rings. It is telling that even in his numerical investigations, Lagrange's thinking was principally informed by new theoretical algebraic ideas. To our knowledge no study has been made of the influence of these ideas on later developments.

Lagrange's algorithmic method for solving numerical equations is not discussed in the current numerical analysis literature, and we have been unable to find a detailed analysis of its efficiency and a comparison with other commonly used methods. One significant advantage that Lagrange's continued fractions method has over, for instance, Newton's method is that it is certain to converge to a root. Lagrange also provided ways of accelerating convergence and speeding calculation of the convergents. In a future paper we shall investigate the rate of convergence of Lagrange's algorithm.

References

[1] É. Bezout, Sur Plusiers Classes d'Equations de Tous les Degrés qui Admettent une Solution Algébriques, in *Mémoires de l'Academie Royale des Sciences* (1762), 17–52.

[2] A. Cauchy, Théorie des Équivalences Algébriques, in *Œuvres Complètes*, Académie des Sciences, 1882–1981, vol. 14, 93–120.

[3] F. Cajori, *A History of the Arithmetical Methods of Approximation to the Roots of Numerical Equations of One Unknown Quantity*, Colorado College Publications, General Series, nos. 51 and 52, Colorado Springs, CO, 1910.

[4] J.-L. Chabert, *A History of Algorithms*, Springer Verlag, New York, 1999.

[5] R. R. Hamburg, The Theory of Equations in the 18th Century: The Work of Joseph Lagrange, *Arch. Hist. Exact Sci.* **16** (1976) 17–36.

[6] J.-L. Lagrange, *Lectures on Elementary Mathematics*, (Thomas McCormack, transl.), Open Court Publ., Chicago, 1898.

[7] J.-L. Lagrange, *Traité de la Résolution des Equations Numériques de Tous Les Degrés*, in *Œuvres de Lagrange*, vol. VIII, J.-A. Serret (ed.), Gauthier-Villars, Paris, 1879.

[8] W. F. Meyer, *Encyklopädie der Mathematischen Wissenschaften*, Teubner, Leipzig, 1898–1904, vol. 1, part 1.

[9] D. Struik, *A Source Book in Mathematics, 1200–1800*, Princeton Univ. Press, Princeton, NJ, 1986.

[10] H. Weber, *Lehrbuch der Algebra, I*, Chelsea Publ., New York.

[11] Richard Zippel, *Effective Polynomial Computation*, Kluwer, Boston, 1993.