On the Geometry of Halley’s Method

T. R. Scavo                   J. B. Thoo
trscavo@mailbox.syr.edu      jb2@math.ucdavis.edu

May 5, 1994
(To appear in The American Mathematical Monthly)

According to Traub [Tra64], Halley’s iteration function (I.F.) “must share with the secant I.F. the distinction of being the most frequently re-discovered I.F. in the literature.” Halley’s method is a close relative of Newton’s method, an iterative technique depicted as a sequence of tangent lines with zeros converging to a root of a function. The usual derivation of Halley’s method, however, lacks any obvious geometric interpretation. We present a derivation of Halley’s method having such an interpretation, and give a brief history of Halley’s work and the method that bears his name.

1 Historical Background

Edmond Halley (1656–1742), well-known astronomer and mathematician, was impressed by the work of “an ingenious professor of mathematics,” Thomas Fautet de Lagny, who, in a book published in Paris in 1692, presented some formulas for “extracting roots of pure powers, especially the cubic.” Halley sought to understand the origin of these formulas, and in the process came to generalize them.

The result of de Lagny that impressed Halley is that \( \sqrt[3]{a^3 + b} \) lies between

\[ a + \frac{ab}{3a^3 + b} \quad \text{and} \quad \frac{a}{2} + \sqrt[3]{\frac{a^2}{4} + \frac{b}{3a}} \]  

for \( a^3 \gg b > 0 \). Halley called these the rational formula and the irrational formula, respectively [Hal1694]. Each is a special case of more

---

1 See [Bat38] for a complete reference.

2 de Lagny also gave a fifth-order formula that Halley found even more impressive.
general iteration functions derived in Section 3.\textsuperscript{3}

It is ironic that Halley preferred the irrational formula over the rational formula, for it is the latter that bears his name. Indeed, virtually all of Halley’s calculations employed the irrational formula, of which he wrote [Hal1694]

And this formula is deservedly preferred before the rational one, which, on account of its large divisor, cannot be used without much trouble, in comparison of the irrational one, as manifold experience has informed me.

Apparently, extracting roots was relatively easy for Halley who claimed, for example, to have calculated eighteen significant digits of the cube root of 231 “in an hour’s time” using the irrational formula.

Another reason Halley preferred the irrational formula was his belief that it generally gives better approximations than the rational formula. Speaking of the methods in (1) he said

And between these two limits always lies the true root, being rather nearer to the irrational than to the rational formula.

While true of each example given in [Hal1694], this is not true in general, however. A counterexample is provided in Section 4.

Halley also admired the work of the 16th century French mathematician François Viète, popularizer of what later became known as “Horner’s method,” an approximation technique pioneered by several Chinese mathematicians in the 13th century [Boy68]. Viète’s method, as it is sometimes called, is a linearly converging algorithm akin to bisection and may be applied to any polynomial with at least one real root [Ypm93]. It may also be used to produce starting values for Newton’s method and other higher-order iterative procedures, something that Halley himself might have done. More importantly, it appears that Viète’s method was an important precursor of Halley’s method, and of root-finding methods in general.

Although Halley was almost certainly aware of the fledgling calculus when he wrote his paper in 1694,\textsuperscript{4} he apparently did not realize that his method involved derivatives or fluxions, as he would have called them. In

\textsuperscript{3}The formulas in (1) may be obtained by setting \( f(x) := x^3 - (a^3 + b) \) in Equations (11) and (8), respectively, and evaluating at \( x := a \).

\textsuperscript{4}Newton published his Principia in 1687, but “only after intense coaxing” by Halley [Boy68]. In fact, the well-to-do Halley had the Principia published at his own expense.
hindsight, this is not surprising considering it was Simpson, in 1740, who first realized the connection between derivatives and Newton’s method [Kol92]. What is surprising is that Brook Taylor recognized the derivatives in Halley’s method as early as 1712 [Fei85]:

[Taylor] noticed what Halley had failed to realize before him: that the coefficients in [Halley’s examples] are directly related to the successive derivatives of the original polynomial.

Moreover, applying Halley’s techniques to Kepler’s problem—an outstanding problem in astronomy with which Halley was no doubt familiar—led Taylor to a remarkable discovery [Bai89, Ypm93]. In a letter to Machin in 1712, Taylor proclaimed [Fei85]

While I was thinking of these things, I fell into a general method of applying Dr. Halley’s Extraction of roots to all Problems . . .
And it is comprehended in this Theorem . . .

which turns out to be Taylor’s Theorem!

The reason that Kepler’s problem went unsolved for so long is that it involves a transcendental equation. In a superb summary of Halley’s work, Bateman [Bat38] suggests that Halley might have preceded Taylor in the discovery of Taylor’s formula had he only “applied his methods in a general way to transcendental equations.” While not important in and of itself—after all, Gregory knew of “Taylor’s theorem” around 1668—it is noteworthy that it was Halley’s method that prompted these developments, whereas Newton’s method languished in ignorance until the time of Simpson.

Despite Taylor’s achievements, he was unable to provide a general formula for Halley’s method. It remained for Schröder [Sch1870], more than one-and-a-half centuries later, to derive Halley’s iteration function as we now know it. But Schröder made no reference to Halley. Indeed, Schröder was primarily interested in higher-order iteration functions and mentioned Halley’s formula almost in passing.

Kobald [Kob1891] derived Halley’s formula in a brief paper published in 1891, but unfortunately his derivation is unclear. Frame [Fra44], on the other hand, was the first to derive Halley’s iteration function via a second-degree Taylor polynomial (see Section 3 and also [Wal48, Ste51, Gan85, Bai89]). Some textbooks also employ this method (e.g., [Mcc67]), while others simply mention it.

Some authors have used determinants and Cramer’s rule to derive Halley’s formula and other higher-order iteration functions (see [Ham50, Ste51,
2 Preliminaries

Given a function $F: X \to X$ and a point $x_0 \in X$, one may iterate $F$ to generate the sequence of points $x_0, x_1 = F(x_0), x_2 = F(x_1)$, and so forth. The sequence thus obtained,

$$x_{n+1} = F(x_n) \quad \text{for } n = 0, 1, 2, \ldots,$$

is called the orbit of $x_0$ under iteration of $F$. A point $\alpha$ is called a fixed point of $F$ if $F(\alpha) = \alpha$. When $X$ is a subset of the real numbers, the graph of $F$ intersects the line $y = x$ at each fixed point (see Figure 1). A fixed point $\alpha$ is said to be attracting if there exists a neighborhood $U$ of $\alpha$ such that the orbit of every point $x_0 \in U$ converges to $\alpha$ under iteration of $F$. Finally, if $|F'(\alpha)| < 1$, then $\alpha$ is an attracting fixed point, an important result known over a century ago [Sch1870].

Now consider the problem of finding a root $\alpha$ of the equation

$$f(x) = 0.$$  

Figure 1: An orbit converging to a fixed point.
We assume throughout that the root in question is simple, that is, \( f'(\alpha) \neq 0 \). One way to approximate \( \alpha \) is to find another function \( F \), called an iteration function (I.F.) for \( f \), for which \( \alpha \) is an attracting fixed point. Then, for a suitably chosen initial value \( x_0 \), the iteration (2) converges to \( \alpha \). Note that the choice of I.F. is not unique (e.g., [Bur89, page 42]).

One well-known iterative root-finding method is Newton’s method,

\[
x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},
\]

a special case of (2) with \( F(x) = x - f(x)/f'(x) \). Evidently \( \alpha \) is a fixed point of \( F \) if \( \alpha \) is a simple root of (3); furthermore, this fixed point is attracting (see below). We call \( F \) the **Newton I.F.** for \( f \), and denote it by \( N_f \).

To derive (4), we approximate the given function \( f \) at \( x = x_n \) by a linear function \( y \) of the form

\[
y(x) = a(x - x_n) + b.
\]

Then the requirement that both \( f \) and \( y \), and their first derivatives, agree at \( x = x_n \) leads to

\[
y(x) = f'(x_n)(x - x_n) + f(x_n).
\]

Finally, solving \( y(x_{n+1}) = 0 \) for \( x_{n+1} \) yields (4). Since (5) is the equation of the line tangent to \( f \) at \( x = x_n \), it is clear that Newton’s method applied to \( f \) may be interpreted as a sequence of tangent lines with zeros converging to a root of the function. (See Figure 2.)

Newton’s method is a **quadratically** converging root-finding algorithm. Loosely speaking, this means that the number of significant digits eventually *doubles* with each iteration. Such a method gives rise to a **second-order algorithm**. It can be shown that the first derivative of a second-order I.F. vanishes at the corresponding fixed point. In the case of Newton’s I.F., the first derivative is

\[
N'_f(x) = \frac{f(x)f''(x)}{f'(x)^2},
\]

which clearly vanishes at a simple root \( \alpha \). Hence \( \alpha \) is an attracting fixed point. And since \( N''_f(\alpha) = f''(\alpha)/f'(\alpha) \) is nonzero in general, there exists a neighborhood for which (4) converges quadratically to \( \alpha \).

Whereas Newton’s method is second order, we show in Section 3 that
Halley’s method,
\[ x_{n+1} = x_n - \frac{2f(x_n)f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)}, \]
is a **third-order algorithm**. Such an algorithm converges *cubically* insofar as the number of significant digits eventually *triples* with each iteration. And not only does the first derivative of a third-order I.F. vanish at a fixed point, but so does the second derivative.

### 3 Halley’s Method

In Section 2 we derived Newton’s method using a linear function \( y \), the first-degree Taylor polynomial of \( f \) at \( x_n \). Let’s see what happens if we instead use a second-degree Taylor polynomial,
\[
y(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(x_n)}{2}(x - x_n)^2,
\]
where \( x_n \) is again an approximate root of \( f(x) = 0 \). As with Newton’s method, the goal is to determine a point \( x_{n+1} \) where the graph of \( y \) intersects the \( x \)-axis, that is, to solve the equation
\[
0 = f(x_n) + f'(x_n)(x_{n+1} - x_n) + \frac{f''(x_n)}{2}(x_{n+1} - x_n)^2 \tag{6}
\]
for $x_{n+1}$. Following Frame [Fra44] and others, we factor $x_{n+1} - x_n$ from the last two terms of (6) to obtain

$$0 = f(x_n) + (x_{n+1} - x_n) \left( f'(x_n) + \frac{f''(x_n)}{2} (x_{n+1} - x_n) \right),$$

from which it follows that

$$x_{n+1} - x_n = -\frac{f(x_n)}{f'(x_n) + \frac{f''(x_n)}{2} (x_{n+1} - x_n)}. \quad (9)$$

Approximating the difference $x_{n+1} - x_n$ remaining on the right-hand side of (9) by Newton’s correction $-f(x_n)/f'(x_n)$ given in (4), we obtain

$$x_{n+1} = x_n - \frac{2f(x_n)f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)}, \quad (10)$$

widely known as Halley’s method [Bat38, Ste51, Fra53, Kis54, Sny55, Tra61b, Tra64, Dav75, Bro77, Han77, Pop80, Ale81, Gan85].

Unfortunately, the preceding derivation lacks any clear geometric interpretation analogous to the tangent lines of Newton’s method. What we seek is an osculating curve to $f$ at $x_n$ (that is, a curve agreeing with $f$ at $x_n$ up through second derivative) that interpolates the points $(x_n, f(x_n))$ and $(x_{n+1}, 0)$ where $x_{n+1}$ is given in (10). If the curve crosses the $x$-axis but once, so much the better. This brings us to the so-called “method of tangent hyperbolas,” but first we make a few remarks concerning Halley’s method.

---

One might be tempted to apply the quadratic formula to (6), obtaining

$$x_{n+1} - x_n = \frac{-f'(x_n) \pm \sqrt{f'(x_n)^2 - 2f(x_n)f''(x_n)}}{f''(x_n)}. \quad (7)$$

A judicious choice of sign in (7) (see [Tra64, Gor90]) leads to the I.F.

$$C_f(x) = x - \frac{1 - \sqrt{1 - 2f(x)f''(x)/f'(x)^2}}{f''(x)/f'(x)}, \quad (8)$$

a general form of Halley’s irrational formula [Hal1694, Bat38, Gan85]. But neither (7) nor (8) is what is known as Halley’s method. We remark that rationalizing the numerator of (7) yields a special case of Laguerre’s method [Ost73, Han77] sometimes attributed to Cauchy [Tra64, Pop80].
Denote the **Halley I.F.** for $f$ by

$$H_f(x) = x - \frac{2f(x)f'(x)}{2f'(x)^2 - f(x)f''(x)}. \quad (11)$$

If $\alpha$ is a simple zero of $f$, then we see immediately that $H_f(\alpha) = \alpha$. Further, a straightforward calculation shows that $H_f'(\alpha) = H_f''(\alpha) = 0$ while $H_f'''(\alpha) \neq 0$. Thus, Halley’s I.F. is third order for simple roots. In fact, a direct computation shows that

$$H_f'''(\alpha) = -\left( \frac{f'''(\alpha)}{f'(\alpha)} - \frac{3}{2} \left( \frac{f''(\alpha)}{f'(\alpha)} \right)^2 \right) = -Sf(\alpha),$$

where $Sf(x)$ denotes the Schwarzian derivative of $f$ at $x$, a most curious result.\(^6\)

Bateman [Bat38] was the first to point out that Halley’s method may be obtained by applying Newton’s method to $f/\sqrt{f'}$, that is,

$$N_{f/\sqrt{f'}} = x - \frac{f/\sqrt{f'}}{(f/\sqrt{f')'} } = H_f.$$ And despite the uncanny similarity, Halley’s method is not to be confused with a second-order method for multiple roots discovered by Schröder,

$$x_{n+1} = x_n - \frac{f(x_n)f'(x_n)}{f'(x_n)^2 - f(x_n)f''(x_n)},$$

obtained by applying Newton’s method to $f/f'$ [Sch1870, Bod49, Tra64, Bur89].

The following special case of Halley’s method is also worth investigating. Let $g(x) = x^d - r$. Then, by (11), the Halley I.F. for $g$ is

$$H_g(x) = \frac{(d-1)x^d + (d+1)r}{(d+1)x^d + (d-1)r} x, \quad (12)$$

a result often ascribed to Bailey [Bai41, Fra45, Tra61a], but actually due to Lambert in 1770 [Kis54, Tra61b]. Traub [Tra64] remarks that some early authors called (12) “Hutton’s method” without reference. Indeed, a footnote

\(^6\)The Schwarzian derivative is an important tool in the study of discrete dynamical systems [Dev92].
in the English translation of Halley’s paper [Hal1694, page 644] specifically attributes (12) to Hutton in 1786, but this clearly postdates Lambert’s work. (See [Bat38, Wal48] and especially [Bai89] for more information on Lambert’s method.)

We close this section with a graphical example comparing the methods of Newton and Halley. Let \( f(x) = x^2 - 2 \). Then

\[
N_f(x) = \frac{x^2 + 2}{2x} \quad \text{and} \quad H_f(x) = \frac{x^3 + 6x}{3x^2 + 2}.
\]

Since \( f(\sqrt{2}) = 0 \), it follows that \( N_f(\sqrt{2}) = H_f(\sqrt{2}) = \sqrt{2} \) (see Figure 3a). Moreover, this fixed point is attracting since \( N_f'(\sqrt{2}) = H_f'(\sqrt{2}) = 0 \). And because the second derivative of \( H_f \) also vanishes, whereas the second derivative of \( N_f \) does not, the graph of \( H_f \) is flatter than that of \( N_f \) near the fixed point (see Figure 3b). This accounts for the difference in speed at which the two algorithms converge (see [Wal48, Bod49] for details). In general, the higher the order, the flatter the graph and, hence, the faster the convergence.
4 The Method of Tangent Hyperbolas

Salehov [Sal52] was apparently the first to suggest that Halley’s I.F. could be derived using an osculating rational function of the form

\[ y(x) = \frac{x + c}{ax + b}. \]  

(Recall from page 7 that an osculating curve to \( f \) at \( x_n \) is one that satisfies the equations

\[ y^{(k)}(x_n) = f^{(k)}(x_n) \]  

for \( k = 0, 1, 2 \).) For convenience, we use an equivalent form of (13),

\[ y(x) = \frac{(x - x_n) + c}{a(x - x_n) + b}. \]  

Equations (14) and (15) taken together lead to the system of equations

\[
\begin{align*}
\frac{c}{b} &= f(x_n) \\
\frac{b - ac}{b^2} &= f'(x_n) \\
\frac{2a(ac - b)}{b^3} &= f''(x_n)
\end{align*}
\]

having solution

\[
\begin{align*}
a &= \frac{-f''(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)} \\
b &= \frac{2f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)} \\
c &= \frac{2f(x_n)f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)}
\end{align*}
\]

It follows from (15) that if \( y(x_{n+1}) = 0 \), then \( x_{n+1} = x_n - c \) where \( c \) is given in (16), that is,

\[ x_{n+1} = x_n - \frac{2f(x_n)f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)}. \]
But this is precisely Halley’s method given in (10). In other words, Halley’s formula can be derived using an osculating hyperbola. Indeed, Halley’s method is sometimes called the method of tangent hyperbolas [Sal52, Saf63, Tra64].

As an example, consider the function \( f(x) = e^x - 2 \) which has a unique zero at \( \alpha = \log 2 \). The Halley I.F. for \( f \) is

\[
H_f(x) = x - \frac{2(e^x - 2)e^x}{2e^{2x} - (e^x - 2)e^x} = \frac{(x - 2)e^x + 2(x + 2)}{e^x + 2}.
\]

Observe that the graph of \( H_f \) is asymptotic to the diagonal lines \( y = x \pm 2 \) as \( x \to \pm\infty \), respectively (see Figure 4a). Indeed, a direct calculation shows that

\[
0 \leq H_f'(x) = \left( \frac{e^x - 2}{e^x + 2} \right)^2 < 1
\]

for all \( x \), making the fixed point globally attracting. Consequently, we may choose any initial value we please. For instance, using the starting value \( x_0 = 10 \), system (16) yields approximately

\[
\begin{align*}
&\begin{cases}
a = -4.539580784 \times 10^{-5} \\
b = 9.079161568 \times 10^{-5} \\
c = 1.999636834
\end{cases} \\
\text{(17)}
\end{align*}
\]

from which we obtain \( x_1 \approx x_0 - c = 8.000363166 \). Similarly, substituting \( x_1 = 8.000363166 \) in (16) gives approximately

\[
\begin{align*}
&\begin{cases}
a = -3.351160651 \times 10^{-4} \\
b = 6.702321302 \times 10^{-4} \\
c = 1.997319071
\end{cases} \quad \text{(18)}
\end{align*}
\]

Thus, \( x_2 \approx x_1 - c = 6.003044095 \).

The osculating hyperbolas (15) corresponding to (17) and (18) are plotted alongside the graph of \( f \) in Figure 4b. Notice that upon successive applications of Halley’s method, the zeros of the tangent hyperbolas tend to the zero of \( f \). This is the soughtafter geometric interpretation of Halley’s method. Incidentally, continuing this process numerically, we find that \( x_7 \) agrees with \( \alpha \) to ten decimal places, and thereafter the number of significant
digits roughly triples with each iteration.

The function $f(x) = e^x - 2$ also provides a counterexample to Halley’s claim that the irrational formula is generally better than the rational formula. Observe that the first two points on the orbit of $x_0 = 1.3$ under iteration of $C_f$ given in (8) are

$$1.3 \mapsto 0.60021187 \mapsto 0.69327247 \ldots,$$

whereas Halley’s method gives

$$1.3 \mapsto 0.71110978 \mapsto 0.69314766 \ldots.$$

Since $\log 2 = 0.69314718 \ldots$, we see that Halley’s method gives better approximations in this case. Thus, contrary to Halley’s claim, the irrational formula does not always give better approximations than the rational formula.

**Acknowledgments**

We thank M. Iklé for comments on an early draft of this paper, and G. D. Chakerian
and B. N. Parlett for critical readings of later versions. Chakerian also pointed out a faulty geometric interpretation in [Fra44]. A. Zorich graciously provided a translation of [Sal52], while G. W. Stewart made available a detailed translation of Schröder’s seminal paper [Sch1870]. Last, but not least, we thank the referees for their helpful suggestions, and for alerting us to an error in the original manuscript.

**Postscript** After this paper was accepted for publication, the authors learned from W. Gander that he gave a geometric interpretation of Halley’s method a decade earlier which was deleted from the published version of his manuscript [Gan85].
References


