Chaotic Convergence of Newton's Method

Jont B. Allen, *Life Fellow, IEEE*

***Abstract*—In 1680 Newton proposed an algorithm for finding roots of polynomials. His method has since evolved but the core concept remains intact. The convergence of Newton's Method has been widely challenged to be unstable or even chaotic. Here we briefly review this evolution, and consider the question of stable convergence. Newton's method may be applied to any complex analytic function, such as polynomials. Its derivation is based on a Taylor series expansion in the Laplace frequency $s = \sigma + \jmath\omega$. The convergence of Newton's method depends on the *R*egion of Convergence (RoC). Under certain conditions, nonlinear (NL) limit-cycles appear, resulting in a reduced rate of convergence to a root. Since Newton's method is inherently complex analytic (that is, linear and convergent), it is important to establish the source of this NL divergence, which we show is due to violations of the Nyquist Sampling theorem, also known as aliasing. Here the conditions and method for uniform convergence are explored. The source of the nonlinear limit-cycle is explained in terms of aliasing. We numerically demonstrate that reducing the step-size always results in a more stable convergence. The down side is that it always results in a sub-optimal convergence. It follows that a dynamic step-size would be ideal, by slowly increasing the step-size until it fails, and then decreasing it in small steps until it converges. Finding the optimal step-size is a reasonable solution.**

***Index Terms*—Aliasing, analytic-roots, convergence, limit-cycles, Nyquist-sampling, regions of convergence (RoC).**

I. Introduction

NEWTON'S method (NM) is a venerable method for finding the roots of polynomials. However its utility has been questioned. First, and most important, does his method always converge? From numerical experiments, it does converge for most initial conditions. Thus the key important question is "Does the convergence depend on this initial condition?" This question was carefully evaluated in 1963 by Willkinson, who studied conditions of sever divergence.[1]

Thus the question becomes "What are the necessary conditions for convergence?" In the following discussion we assume a monic polynomial of degree N. The fundamental theorem of algebra states that every polynomial $P_N(s)$ of degree N has N roots $s_n \in \mathbb{C}$.

Since Newton's algorithm converges within the RoC for any complex analytic function, it converges when the nearest root s_r ($s \in \mathbb{C}$) remains inside the RoC, out to the nearest pole. This follows because every complex-analytic point on the complex plan has a region of convergence [1]. We show that on the boundaries of the RoC regions, the method becomes hypersensitive to the initial condition (i.e., initial condition s_0), and becomes fractal.

We propose a complex *adaptive step-size* $\eta = ae^{\jmath\phi} \in \mathbb{C}$ which we adaptively adjusted, greatly reducing, even removing the nonlinear effects of aliasing as $\eta \to 0$. Historically the introduction of η is known as the *damped Newton's method* [4, p. 25].

In this report we shall investigate why such a controversy developed, and discuss how to assure convergence. In our experience, given some care, the method always converges to a root. Under some special conditions, a minor modification in the initial condition s_0 ($n = 0$) can result in the $n + 1$ estimate of the root (s_{n+1}) to cross an RoC boundary, resulting in NM to divert its initial path to a alternate root. When this happens, the change in the step $\delta = s_{n+1} - s_n$ is unpredictable, and possibly even chaotic. It is this condition that is the source of a convergence instability due to aliasing, possibly leading to a limit cycle. Such chaotic behavior is a main topic of this document.

These contiguous naturally existing *regions of convergence* are defined for all $s \in \mathbb{C} < \infty$. That is, every possible s_0 belongs to only one of the N RoC regions. Convergence naturally happens as $\eta \to 0$, since Newton's method is complex analytic (the step-size is the ratio of two polynomials with different roots). The magnitude of $|\eta| = a$ may also be manually reduced to avoid crossing the boundary between two RoCs. As a is reduced, the trajectory naturally moves away from the poles. The smaller η, the smoother the path. For this reason, adaptively adjusting the optimal η, can minimize the computation, while avoiding NL aliasing.

As an alternate to reducing $|\eta|$, one can modify its angle ϕ, redirecting the trajectory away from any RoC boundary, so as to avoid crossing it. We have not yet implemented this approach.

We show that when $\eta = 1$, depending on s_0, the solution can cross an RoC boundary (i.e., diverge). In such cases the target root will change, resulting in a chaotic trajectory. Examples are provided. Depending critically on s_0, as long as the RoC region remains the same, every iteration rapidly converges as it approaches the root.

Newton's method a is a venerable algorithm for finding roots of any complex-analytic function. Thus NM applies to polynomial $P_N(s)$, where N is the degree and $s = \sigma + \omega\jmath \in \mathbb{C}$ is the Laplace frequency. However the convergence properties of NM are controversial[2] [8, p. 347].

Manuscript received 29 June 2023; revised 15 October 2023 and 2 March 2024; accepted 4 March 2024. Date of publication 11 March 2024; date of current version 5 April 2024. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gabriel Falcao.

The author is with the Electrical and Computer Engineering Department, University of Illinois, Urbana, IL 61801 USA (e-mail: jontallen@ieee.org).

Digital Object Identifier 10.1109/TSP.2024.3376088

[1]https://en.wikipedia.org/wiki/Wilkinson's_polynomial

[2]https://en.wikipedia.org/wiki/Newton's_method#Failure_of_the_method_to_converge_to_the_root

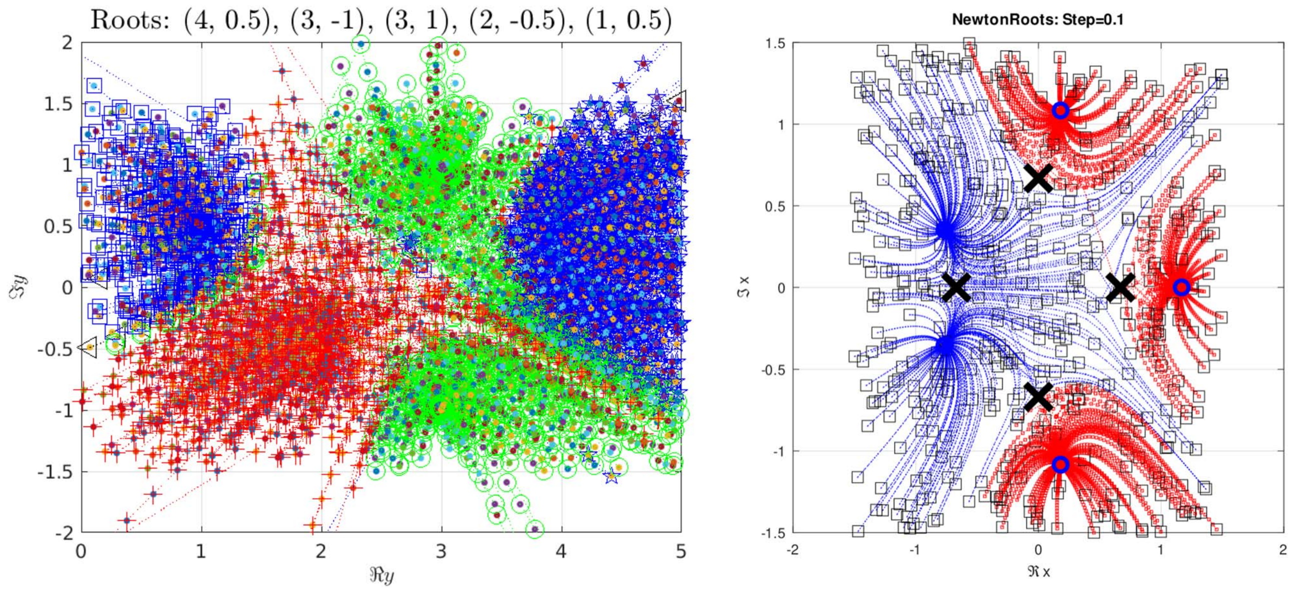

Fig. 1. Left shows a plot of a thousand trajectories for a polynomial having $N = 5$ complex roots, as summarized in the title, starting from a random initial condition between [0,5] and $\pm \jmath 1.5$. Right: (b) shows the poles and zeros of the polynomial having coefficients $\boldsymbol{C} = [1, 0, 0, 0, -1, -1]$ with random starting points, for 200 iterations of Newton's Method. In this case the roots (**o**) and poles $(\times)$ are superimposed on top of the trajectories of Newton's method. An adaptive step-size of $\eta = 0.1$ is used to reduce the NL aliasing.

II. Methods

Every initial condition s_0 on the plane of a complex analytic function, is uniquely associated with one of the N roots of that function, which is associated with a unique region of convergence (RoC). This follows from the complex analytic property of a function (those that may be expanded in a complex-analytic Taylor series).

When the trajectory jumps to a different RoC, corresponding to a different root, it has been interpreted as a failure to convergence, when in reality the target root has changed. This can only happen when the present s_n is on or near an RoC boundary. In such cases Newton's method develops properties that are similar to *dynamic analysis*, a mathematical science first introduced by Poincaré.[3]

This question of the convergence was recently explored in [1], where no instability or limit-cycles were observed. An explanation is due: Newton's method was modified by applying an *adaptive step-size* η, [4, p. 25], a widely recognized contemporary technique in the engineering numerical analysis literature.[4]

A properly chosen adaptive step-size stabilizes the convergence, by forcing the path to remain in the target RoC. Detecting the divergence of the step is easy because it must monotonically decrease, and may be stabilized by reducing $|\eta|$.

In this report we show that random jumps and limit-cycles are more likely when $\eta = 1$. When the adaptive step-size η is sufficiently small, we show that the iteration always converges. The strategy employed here is to adaptively modify $|\eta|$, thereby constraining the trajectory to the initial RoC.

[3]https://en.wikipedia.org/wiki/Dynamical_systems_theory
[4]https://en.wikipedia.org/wiki/Adaptive_step_size

In [1, Fig. 3.2], two examples were provided using a fixed adaptive step-size ($\eta = 0.5$) and a random initial condition s_0. The details of the adaptive step-size used by [1] were not discussed. One of these figures is presented in Fig. 1 (left).

While most of the curves seem to converge to a root, there is a small percentage of cases (e.g., $< 1\%$) where the trajectories take huge jumps to random locations in the complex plane. We shall show that these jumps occur when the trajectory approaches any of the poles of Newton's method, that is, at the roots of $P_N'(s) = \frac{d}{ds} P_N(s)$. Near a pole the step can be arbitrarily large, depending on how close the step comes to the pole [3]. We shall show that the poles, the cause of the NL limit cycles, are easily detected.

In Fig. 1 (left), the five RoC regions are color coded, with each RoC region associated with one of the N roots. Due to the complex analytic nature of an RoCs, every point in the RoC is a valid initial condition. However this is limited by the numerical accuracy of the computer. Also the convergence depends on the size of the steps, defined as $s_{n+1} - s_n, s \in \mathbb{C}, n \in \mathbb{N}$, which typically decreases in magnitude with n. An exception occurs if s_{n+1} approaches one of the $N - 1$ poles of NM, causing the step to abruptly diverge. The properties of this small subset of initial conditions depends critically on η, which is the main topic of this article.

For most initial conditions ($s_0 \in \mathbb{C}$) the iteration simply converges to a root, independent of $|\eta| \leq 1$. In fact for most starting values the solution converges for $|\eta| = 1$. However for s_0 values near the RoC boundary between two roots, the dependence is highly dependent on $|\eta|$, and can even be chaotic. This happens when s_0 defines a path that heads directly at a pole. In these cases the trajectory will be hypersensitive to both s_0 and $|\eta|$. The RoC regions are well defined non-overlapping

TABLE I
PROPERTIES OF THE POLYNOMIALS FOR THE LEFT AND RIGHT PANELS

Figure	$P_N(s) =$	$\Re s_r$	$\Im s_r$
Fig. (1)a, **LEFT**	$s^5 - (13+0.5\jmath)s^4 + (66.25+5\jmath)s^3$ $-(164.25+20.125\jmath)s^2$ $+(195+38.75\jmath)s - 87.5 - 31.25\jmath$	[4, 3, 3, 2, 1]	[1, -2, 2, -1, 1]/2
Fig. (1)b, **RIGHT**	$s^5 - s - 1 \leftrightarrow$ [1, 0, 0, 0, -1, -1]	[1.17, 0.18, 0.18, -0.76, -0.76]	[0, 1.08 -1.08, 0.35, -0.35]

complex-valued analytic regions. When s_0 is close to the RoC boundary, the convergence of NM critically depends on the magnitude and angle of the complex adaptive step-size η. Even when $|\eta| \ll 1$, the convergence can become NL, resulting in a chaotic path. These observations are supported by several detailed numerical examples.

In Fig. 1 (left), the red region, corresponding to the root at $(2.0 - 0.5\jmath)$, has a long narrow "RoC stream" for initial condition s_0 east of $(4.5 - 1\jmath)$. There is a second green narrow neighboring related parallel stream, just north of the red stream, for initial values s_0 north of $(5 - 1\jmath)$, which is in the RoC of root $(3 + 1\jmath)$.

While it may seem obvious given Fig. 1, I am not aware of any discussion of such natural distortion of the RoC's. The conditions for Fig. 1 are provided in Table I.

Fig. 1 (right) is a second example having different poles and zeros, which more clearly demonstrates the effect of the poles and zeros on the trajectories, which are indicated by $\times$ and **o**.

A. Convergence of Newton's Method

Given the monic polynomial of degree $N \in \mathbb{N}$ $P_N(s) = s^N + c_{N-1}s^{N-1} + c_{N-2}s^{N-2} + \cdots + c_0$, and its derivative $P'_N(s) \equiv dP(s)/ds$ of degree $N-1$, we may express Newton's method as the ratio of the two monics (the details are in the Appendix)

$$s_{n+1} - s_n = -\frac{\eta}{N}\frac{P_N(s_n)}{P'_N(s_n)} \tag{II.1}$$

Next we define the properties of the ratio of two monic polynomials, in terms of the *step-size* $S_N(s_n) \equiv P_n/P'_N$.

Equation 1 may be rewritten as

$$\frac{s_{n+1} - s_n}{\eta} = -\frac{1}{N}S_N(s_n).$$

Scaling $P'_N(s)$ as a monic does not alter its roots.

Taking the limit $\eta \to 0$ results in the complex-analytic expression for NM

$$\frac{ds}{d\eta} \equiv \lim_{\eta\to 0}\left(\frac{s_{n+1} - s_n}{\eta}\right) = -\frac{1}{N}S_N(s_n). \tag{II.2}$$

The right hand side $S_N(s_n)$ is the reciprocal of the log-derivative of $P_n(s)$ expressed as monics. The left hand side is the slope of the Laplace frequency $(s = \sigma + \jmath\omega)$ w.r.t. η.

B. What Is Going On?

In the limit as η goes to zero, close to the RoC boundaries are well defined analytic regions. But for small $\eta \neq 0$, no matter how small, the boundaries are fractal, becoming smooth only in the limit as η goes to zero. This fractal structure is always present even for the smallest nonzero values of η. Insight into how this happens is explained by the examples in Figs. 2 and 3. While the concept of a fractal is not difficult, the source of its behavior verges on the mysterious. The source of chaos is precisely explained by Figs. 2 and 3.

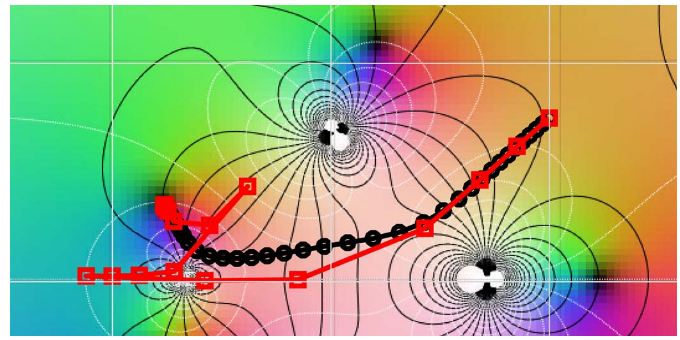

Fig. 2. This is a zoomed-in chart of Fig. 1(b) (right), presented as a colorized plot [1, p. 168] of $S_N(s)$, for $P_5(s) = s^5 - s - 1 \leftrightarrow [1, 0, 0, 0, -1, -1]$. The magnitude of $S_N(s)$ is coded by the brightness, and the phase ($\angle(L_N(s))$, Eq. (III.4)) by the color (hue). The dark regions are the zeros (roots of $P_n(s)$) while the white regions are the poles of $S_N(s)$ (roots of $P'(s)$). Two trajectories of Newton's method are shown, as the black circles and red squares. The initial value for both cases is $s_0 = 1 + 0.75\jmath$. The black circles correspond to $\eta = 0.1$, while the red squares ($\eta = 0.5$) form a brief limit cycle. The vertical white lines are at {-1.0, 0,1.0} and the horizontal white lines are at {0, 1.0}. The polynomial coefficients are $P_5(s) = [1, 0, 0, 0, -1, -1]$, with roots $s_r =$[1.167, $0.181 \pm 1.084\jmath$], $-0.765 \mp 0.3525\jmath$. The real poles (roots of $P'_5(s_r) = 0$) are $\pm 1/5^{0.25}$, while the imaginary poles are at $\pm 0.6687\jmath$.

III. EXAMPLES OF NEWTON'S METHOD

a) Example 1: We start with the monic polynomial of Example 1b, Fig. 1 (right), for $N = 5$,

$$P_5(s) = s^5 - s - 1. \tag{III.1}$$

In this case, monic $\frac{1}{5}P'(s) = s^4 - 1/5$ has four poles, shown as black-bold $\times$ symbols. The five zeros are the black-bold **O** symbols.

As shown in Example 1b, Fig. 1 (right), given any initial condition s_0 and adaptive step-size η, and $n \to \infty$, s_n approaches a unique root s_r. Complex s_{n+1} is the $n+1$ estimate of the root, given the n estimate s_n, as defined by Eq. (II.1).

The example shown in Fig. 2 is a zoomed-in version of Example 1b, Fig. 1 (right). To study the convergence and limit-cycles it is helpful to vary both η and s_0.

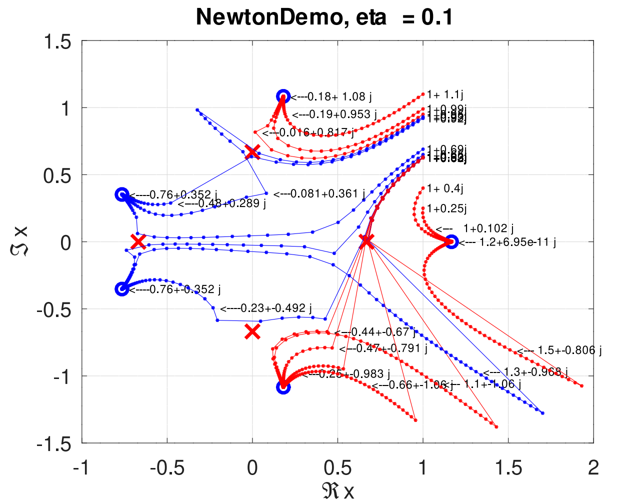

Fig. 3. This numerical experiment for polynomial coefficients $[1, 0, 0, 0, -1, -1]$ (the same polynomial as shown on the right panel of Fig. 1) having an adaptive step-size of 0.1), reveals the inner workings of Newton's method. We number the roots counter-clockwise from 1-5, with $s_1 = 1.2$, $s_2 = 0.18123 + 1.08395\jmath$ and its conjugate $s_5 = s_2^*$. Seventeen different starting values have been carefully chosen, to determine the root the path converges to. All the starting values are of the form $s_0 = 1 + \jmath\beta$. Each β and its converged root are indexed in Table II. The scattering angle is determined by the residue of the scattering pole. Each curve is labeled twice, once at the starting point and at a second point on the path. This carefully evaluated case is for starting points between $1 + 0.62\jmath$ and $1 + 0.5999\jmath$, which converge to dramatically different RoCs, due to the trajectory squarely hitting the positive real pole at $s_0 = 1 \pm 0.001 + \jmath\ 0.6$.

TABLE II
TABLE OF STARTING VALUES $s_0 = 1 + \beta\jmath$ USE IN FIG. 3, ALONG WITH THE ROC TARGETED ROOT INDEX, DEFINED AS #1 FOR THE REAL ROOT AT $0.2^{1/4}$. ROOT #1 CONVERGES FROM $s_0 = 0 + 1.25\jmath$, ROOT #2 IS DEFINED BY COUNTING COUNTER-CLOCKWISE FROM #1, AT $0.18 + 1.08\jmath$, STARTING FROM $s_0 = 1 + 0.69\jmath$. ROOT #3 ALSO CONVERGES FROM THREE VALUES OF β. ROOT #5 IS THE MOST CAREFULLY EXPLORED, STARTING FROM $1 + \beta\jmath$. IT IS SHOWN TO CONVERGE TO ROOTS 1, 3, 4, 5, BUT NOT 2, WHICH IS REACHABLE FROM VERY SELECTIVE VALUES OF β. FOR OTHER CHOICES OF β_0, ALL 5 ROOTS CAN BE REACHED, AS SHOWN IN FIG. 3 FOR $\eta = 1$

β	root
0.25, 0.4	#1
0.95, 0.99, 1.1	#2
0.69, 0.92, 0.93	#3
0.65, 0.632	#4
0.63, 0.631	#5

A. Discussion of Fig. 2

Fig. 2 shows two paths with the same initial condition $s_0 = 1 + 0.75\jmath$ with different adaptive step-sizes. The utility of the reduced adaptive step-size is clear from the figure.

The black circles show a smooth analytic trajectory, while the red-squares are chaotic. For the larger step-size of 0.5 it takes additional steps to limit-cycle, recover and finally converge. Once near a zero, fewer than 10 steps typically give double-precision floating-point machine accuracy. In a small neighborhood around any pole, every point in $\mathbb{C}$ is present [3].

As discussed in the figure caption, if s_n is close to a root s_r of P'_N (i.e., a pole), the recursion dramatically fails, because the step becomes arbitrarily large, forcing the next trial to a random location in the s plane, denoting $\tilde{s}_r$. In such cases the solution typically converges to a different root (RoC). It is not difficult to detect these large random steps by monitoring $|s_{n+1} - s_n|$, which must monotonically decrease.

As shown by the black circles, as the adaptive step-size η is reduced from 0.5 to 0.1, the path approaches the pole, it moves away, avoiding the limit-cycle. With step-sizes of 0.2 and 0.3 it also becomes captured by the pole. With the adaptive step-size of 0.9 (not shown), the trajectory is similar to that of 0.5. After 5 steps it is well within a different RoC, corresponding to the zero at $-.8 - .3\jmath$, where it quickly converts to that root.

In summary, given a larger step-size it still converges, but much more slowly, since the NL becomes greater. Thus the convergence time seems a crude metric of the NL. The smoothness of the trajectory may be more appropriate. This NL result is due to the reduced sample step-size, also known as *aliasing*.

a) Example 2: This example (not shown) is for $P_3(x) = x^3 - x + 1$ ($C = [1, 0, -1, 1]$), an example where Newton's method appears to fail. This example has two imaginary roots, $0.66236 \pm 0.56228\jmath$, and a real root -1.32472. If the initial condition is taken to be $s_0 = 1$, the recursion proceeds using real arithmetic (Matlab and Octave). Due to the restriction that the computation is real, the solution is forced to the real line, where it limit cycles between 1.155 and 0.694. The iteration cannot converge if $s_r \in \mathbb{R}$ and $s_0 \in \mathbb{C}$ [1].

If $x_0 = \jmath$, the solution converges in 3 steps to the upper complex root. If one starts the iteration with an imaginary component at $1 + \jmath 10^{-6}$, the iteration converges to the imaginary root in 13 steps.

Roots $s_r \in \mathbb{C}$ may be found by a recursion that denotes a sequence $s_n \to s_r \in \mathbb{C}, n \in \mathbb{N}$, such that $P_N(s_n) \to 0$ as $n \to \infty$. As shown in Fig. 2, solving for s_{n+1} using Eq. (II.1) always gives one of the roots, due to the analytic behavior of the complex *logarithmic derivative* $P'_N/P_N = d\ln(P_n(s))$.

b) In Summary: When there are no limit cycles, each step (s_{n+1}) is always closer to the root. As s_n approaches the root, the linearity assumption becomes more accurate, resulting in a faster convergence.

Even for cases where fractional derivatives are involved, Newton's method will converge, since the log-derivative linearizes the equation [1, p. 197, #5].

B. Discussion of Fig. 3

In Fig. 3 an infinitesimal change in s_0 leads to large jumps into a different RoC. This is best shown by two starting values at $s_0 = 1 + \beta\jmath$ for β just above 0.5 and again just below 1.0. The effect occurs only when the trajectory heads directly at a pole. Any starting value s_0 that goes directly at one of the poles can jump to a random RoC.

Given this mapping, infinitesimal changes in the starting points which head directly at a pole, are reassigned to a random RoC, due to this analytic mapping.

We have defined η as the *adaptive step-size*, because we can set η to modify the *step-size* S_N. This result follows from a mathematical property cited by [3], that the entire plane may be found in the neighborhood of every pole.

a) In Summary: Determining the RoCs for NM by analytic methods seems difficult, since the function $S_N(s_n)$ has poles, confounding the locations of the RoC boundaries. Based on Fig. 1, the RoC are complicated. If s_n approaches one of these the poles, the update can become arbitrarily large, depending on how close s_n is to the pole. If the adaptive step-size is within the RoC, this will not occur. When the value of s_{n+1} falls outside the RoC there can be an arbitrary increase in step-size. Normally this will not happen, since when s_n approaches a pole s_{n+1} is naturally "pushed" away from the pole, as may be seen in Fig. 2 (black circles).

If we start the iteration with the larger step-size, the path develops into a NL limit-cycle near a pole. It is a combination of the large steps and the proximity to the real pole that results in the nonlinear limit-cycle. On the 10th step s_n comes out of the limit cycle, and after 10 more steps, has converged to the root.

C. Newton's Method Applied to Functions Other Than Polynomials

a) Example 2: Example of Plank's formula for Black Body radiation.

Planks famous BB radiation formula is [1], [6]

$$S(\nu) = \frac{\nu^3}{e^{h\nu/kT} - 1}. \tag{III.2}$$

In this historically important example, because the function is real (it is not complex analytic), the spectrum only has one pole, at $\nu = 0$. This formula is known to match the experimental data of the smoothed (non-analytic) black-body power spectrum [5].

If we replace the real frequency ν with the negative Laplace frequency $-s = -\sigma - \omega\jmath$, Eq. (III.2) becomes

$$S(-s) = \frac{-s^3}{e^{-\hbar s/kT} - 1}, \tag{III.3}$$

which is complex analytic, thus has a causal inverse Laplace transform. To use Newton's method we must compute a NM update $L(s_n)$, defined as the reciprocal of the logarithmic derivative (see derivation in the Appendix). Taking the log followed by its derivative w.r.t. $s \in \mathbb{C}$, gives

$$\begin{aligned}\frac{1}{L(s)} &\equiv \frac{d}{ds}\ln S(-s)\\ &= \frac{d}{ds}[-3\ln s + \ln(e^{-\hbar s/kT} - 1)],\\ &= -\frac{3}{s} - \frac{\hbar}{kT}\cdot\frac{e^{-\hbar s/kT}}{e^{-\hbar s/kT} - 1}.\end{aligned} \tag{III.4}$$

Thus there is a first order pole at $s = 0$ and poles at $h\nu_n/kT = 2\pi n$ for $n \in \mathbb{N}$. The discrete frequencies account for the eigen-modes in the black-body radiation, as discussed by Kuhn, Plank and Einstein [5]. Eq. (III.3) and thus Eq. (III.4) are causal, since it has the causal inverse \$LT\$ [1, p. 321]

$$-\frac{1}{L(s)} \leftrightarrow 3u(t) + \frac{\hbar}{kT}\sum_{n=1}^{\infty}\delta\left(t - n\frac{\hbar}{kT}\right) \tag{III.5}$$

The application of NM to Plank's famous formula can be used to make it complex analytic, by replace ν with the Laplace frequency $s = 2\pi\nu_n\jmath$ and h by $\hbar$. It is well established that complex analytic functions of the Laplace frequency s are *causal* (zero for negative time) [1, p. 158]. In the case of Eq. (III.2), $S(-s)$ is causal, due to the Laplace transform relation of the exponent

$$\delta(t - \tau_o) \leftrightarrow e^{-s\tau_o}.$$

Here the time delay $\tau_o = \hbar/kT^\circ = (6.63/2\pi k)\cdot 10^{-11}$ s, (6,280 GHz), $\lambda \approx \frac{\pi}{2}10^{-11}$ m, or $\frac{\pi}{20}$ Å, and T°K is the temperature.

Newton's method uses the reciprocal of $L(s)$ (Eq. (III.4)) to find s_r ($S(s_r) = \infty$), given by

$$Nd(s_r) = 1 - e^{-\hbar s_r/kT} = 0. \tag{III.6}$$

There are an infinite number of such roots, since the roots are $\hbar s_r/kT \approx 2\pi\jmath$. These poles are the missing discrete spectral lines (atomic resonances), required by quantum mechanics.

Applying Newton's method gives

$$x_{n+1} = x_n - \frac{e_n^x - 2}{e_n^x} = s_n - (1 - 2e^{-s_n}).$$

Since e^x is entire, there are no convergence issues.[5] Since $x \in \mathbb{C}$, the imaginary part quickly decays to zero, and depending on the starting condition, approaches one of the infinite number of solutions, within a few steps.

D. Example 3

The impact of s_0 is shown in greater detail in Example 3, as shown in Fig. 3. When the value to s_0 is finely tuned, such that the trajectory intercepts a pole, a host of NL limit-cycles are exposed.

The Gauss-Lucas theorem[6] comes into play at this point [1, p. 81]. This theorem says that the convex hull of the roots of a polynomial bound the roots of its derivative. This theorem is relevant to the convergence of Newton's method. [4] has 75 relevant citations, many citing the same problems addressed here. The key to avoiding the troublesome limit-cycles is to detect them, and then reduce the adaptive step-size.

The following quote is from [4, p. 39]:

The possibility that a small change in s_0 can cause a drastic change in convergence indicates the nasty nature of the convergence problem. The set of divergence points of the Newton method is best described for real polynomials.

As demonstrated in Fig. 3, we agree with Galántai's first point. His second seems vague: Is a "real polynomial" one with real coefficients, real roots, or both?

For the example in Fig. 1, the red "stream" corresponding to the root near $(2 - 0.5\jmath)$ has a long narrow "RoC-stream,"

[5]https://www.quantamagazine.org/how-mathematicians-make-sense-of-chaos-20220302/

[6]https://en.wikipedia.org/wiki/Gauss-Lucas_theorem

converging from the lower-right quadrant, first seen at $(4.5 - 1\jmath)$. There is a second green RoC-stream just north of the red stream, first seen near $(4.5 - 0.9\jmath)$. Thus a small change in the starting value s_0 robustly converges to a totally different root.

I am not aware of any discussion in the literature of this distortion of the RoC regions, bound to Newton's method. Presently I know of no way to predict the conformal remapping of the RoC regions for NM, other than tracking them, as done here. It seems likely that methods for doing must exist using modern analysis techniques (see Appendix).

In the example of Fig. 3,

$$s_{n+1} = s_n - \frac{0.1}{5} \cdot \frac{s_n^5 - s_n - 1}{s_n^4 - 1/5},$$

for 17 carefully chosen initial condition $s_0 \in \mathbb{C}$. For readability, each trajectory is color-coded either red or blue.

a) Nonlinear Limit Cycles: It is well documented that limit cycles are nonlinear. Newton's method on the other hand is a linear recursion equation, with poles and zeros in the complex plane. The obvious research question is *"Why does the complex-analytic linear equation become nonlinear?"* We show how the these NL limit-cycles may be easily avoided by removing (linearizing) the NL recursion once it is detected.

The suggested procedure will result in a net convergence speed-up, because the NL limit-cycle adds meandering NL steps to the recursion. If you experience a slowdown, try changing the adaptive step-size angle. This may be a panacea, since this is a 'local' modification that deals directly with the main problem of being on a RoC boundary. If you find an angle that reduces the chaos, then your moving in the right direction, away from the RoC boundary. This method seems obvious, yet unexplored. On the fractal boundary, movement can lead to chaos.

b) Ratios of Monics as NM: It can be notationally useful to define the *adaptive step-size* $S_N(s)$ as the ratio of monic polynomials

$$\frac{s_n^N + c_{N-1}s_n^{N-1} + \cdots + c_0}{s_n^{N-1} + \frac{N-1}{N}c_{N-1}s_n^{N-2} \ldots + \frac{1}{N}c_1} = \frac{1}{N}\frac{P_N(s_n)}{P'_N(s_n)}. \quad \text{(III.7)}$$

Using this trick we can absorb the factor of N into the definition of $\eta \equiv \frac{1}{N}e^{\phi\jmath}$. Increasing N from 1 to 0.1 dramatically improves the convergence, while the poles (and zeros) of $S_N(s)$ are unmodified.

Fig. 4 quantifies the effect of reducing the step by up to $1/N$ $(|\eta| = [1, 1/2, 1/5, 1/10])$. For the largest step-size, the trajectory of red squares in Fig. 2 limit cycle. This natural reduction in step-size by N, due to expressing the step-size as the ratio of monics, is dramatic. Given s_n, everything on the right is known; thus when s_n is within the RoC, s_{n+1} will converge to a unique root of $P_N(s)$ as $n \to \infty$. For sufficiently small step-size, the roots of Eq. (III.7) are the solution to a linear difference equation, the simplest example being [4]

$$s_{n+1} = s_n - \frac{\eta}{N}S_N(s_n). \quad \text{(III.8)}$$

Introducing the *adaptive step-size* $(|\eta| < 1 \in \mathbb{C})$ linearizes the iteration when s_n is in the neighborhood of a pole.

The step-size $|S_N(s_n)|$ can become arbitrary large near any pole, introducing aliasing (the source of the nonlinear) into the iteration.

IV. Summary and Discussion

a) The Role of the Adaptive Step-Size: In the derivation of NM we modified Eq. (II.1) with the adaptive step-size $\eta < 1$, to obtain Eq. (II.1). The effect of the reduced adaptive step-size is to force the trajectory to be more sensitive to the influence of the poles, rather than stepping over them. The modification of the step-size S_N by η is an important modification to Newton's method. The smaller adaptive step-size can eliminate the nonlinear limit-cycles, as seen in the example of Fig. 4.

When the initial value for the iteration s_0 is close to the cross-over of two RoCs, $s_n \to s_{n+1}$ can cross over an RoC boundary, changing the limit point (root it converges to). A limit cycle can happen when s_n comes close to one of the poles of $S_N(s_n)$. At a pole, the value of S_N can become arbitrary large, causing the unmodified $(\eta = 1)$ update $S_N = s_{n+1} - s_n$ to fail to satisfy the required RoC convergence condition (Eq. (A.6)).

One strategy for detecting the pole is to look at the magnitude of the step $(|\eta|)$. If $|\hat{s}_{n+1} - s_n| > 1$, the RoC condition has failed. The step must then be reverted back to s_n, and the adaptive step-size reduced, and s_{n+1} recomputed. This then repeated until the RoC condition $(|s_n| > |s_{n+1}|)$, thus avoiding a possible limit cycle.

Based on our numerical results, the addition of the convergence factor η seems unnecessary when the initial value is well within the RoC, as required by Eq. (A.6). The main question is when (and why) the limit-cycles are created with Newton's method. This question is at least partial explored in the example of Fig. 2. As long as the RoC condition is maintained, each step will progress closer to a root, and in the limit, as $n \to \infty$,

$$\frac{P_N(s_n)}{P'_N(s_n)} \to 0, \quad \text{(IV.1)}$$

since $s_n \to s_r$ as $n \to \infty$.

We don't understand many observations in science (math and physics). But with some basic analysis, they are eventually explained. Einstein's 1905 analysis is the best known example. It is the *reductionist* method in science, and explains the success of the scientific method. This might be viewed as a form of evolution: success begets more success, while failure eventually dies off, perhaps slowly.

The process of systematically exploring these seemingly tiny discrepancy, almost always leads to new knowledge. Seeking out these idiosyncratic inconsistencies and trying to explain them is at the heart of the scientific method. When a problem is longstanding and considered fundamental, its resolution can even lead to a paradigm shift. Not surprisingly such deep insights are rarely welcomed by the scientific community, rather they are viewed with great skepticism. This can be good when if doesn't go on for 50 years.

The problem of finding roots using Newton's method is an excellent example. It is a case that can be explained only after careful thought and iterative analysis. I feel we are either close

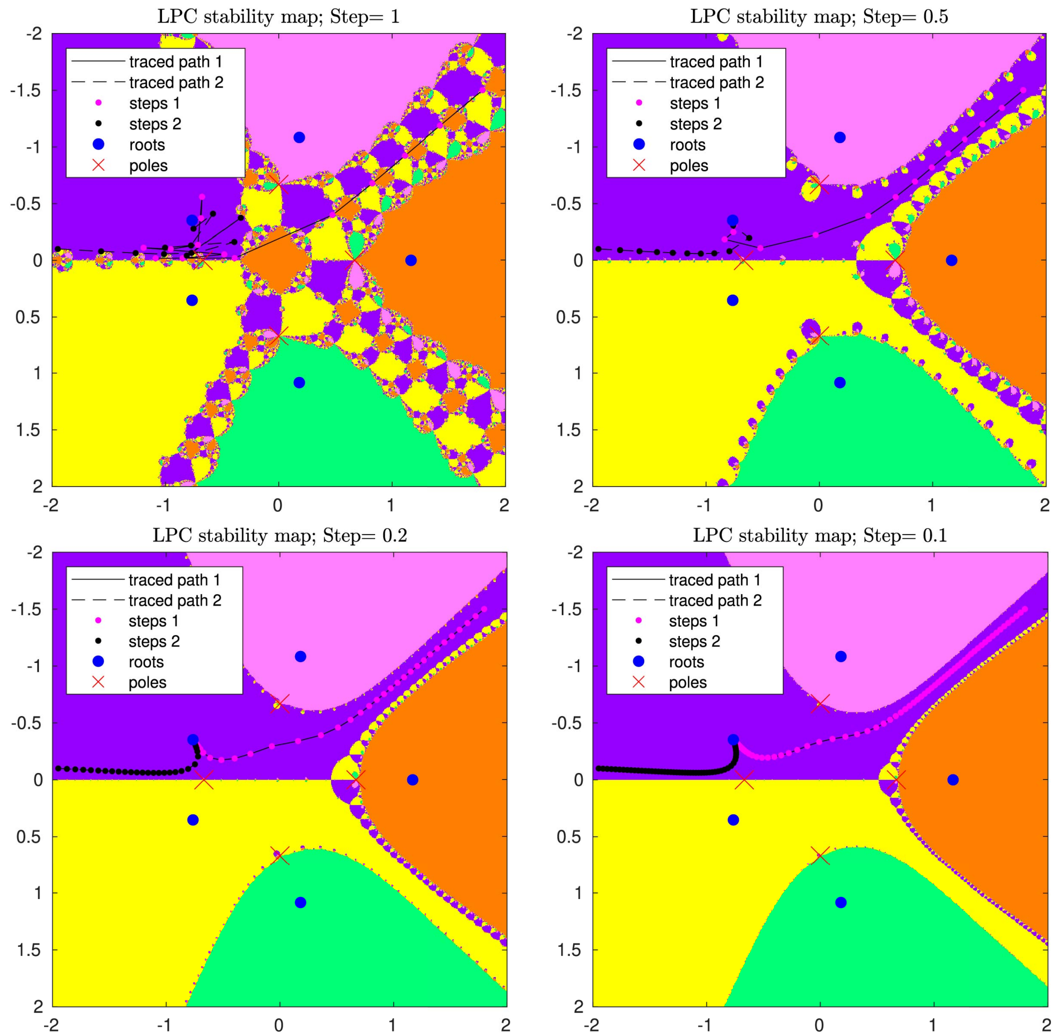

Fig. 4. Four colorized plots for $P_N = [1, 0, 0, 0, -1, -1]$ showing the $N = 5$ regions of convergence and two trajectories, for $s_0 = 1.8 - 1.5\jmath$ and $-1.95 - 0.1\jmath$. The four adaptive step-sizes are $\eta = \{1.0, 0.5, 0.2, 0.1\}$ (note that the imaginary axis is reversed). The fractal regions reside on the RoC boundaries, the sizes of which depend on the adaptive step-size, with the adaptive step-size of $\eta = 1.0$ (Upper-Left) resulting in large fractal regions. Reducing the adaptive step-size to $\eta = 1/2$ dramatically reduces the fractal regions. For $\eta = 0.1$ they almost disappear, except at $0.5 + 0\jmath$. In the dark RoC (purple) corresponding to root $-0.76 - 0.352\jmath$, two trajectories are shown. For the adaptive step-size of 1, a limit cycle is seen, for both trajectories. For the other adaptive step-sizes [0.5, 0.2, 0.1], there are no limit cycles. As the trajectories approach the negative real pole, labeled as the red $\times$, they head for the root at $-0.76 - .352\jmath$. In summary: 1) limit cycles are wasted steps, easily fixed by reducing the adaptive step-size. 2) Given a smaller adaptive step-size, the fractal regions shrink, but never totally disappear. 3) Detecting a limit cycle is easy because the path reverses (oscillates). An obvious method for avoiding limit cycles is to detect that the boundary has been crossed, corresponding to a different root, and restart with a reduced adaptive step-size, at step s_n or s_{n-1}.

to that understanding, or it has been explained clearly enough that the debate can be stopped, and final conclusions may be reached. However, realize that there is no "final."

Limit cycles do exist in Newton's method, but in my view, they are due to under-sampling the complex plane. This is an example of aliasing, in the Nyquist sense, [1, p. 153, 262]. An under-sampled process becomes nonlinear when the "high frequencies" alias into the "base-band" frequencies. This nonlinear effect is easily removed by increasing the sampling rate above the Nyquist sampling frequency, defined as twice the highest frequency in the signal. While that concept is not clear in the context of Newton's method, it can explain limit-cycles, and slightly (2x-3x) increasing the computation, by decreasing the adaptive step-size (η), the aliasing may be brought under control, and the problem becomes linear and well behaved. The onset of aliasing is easily detected. This leads to a well know method in signal processing called the *adaptive step-size*, which has been successfully applied in many engineering problems. It is, I believe, well understood and characterized in terms of aliasing [2], [7, Sec. V, p. 126].

b) The Linear Prediction Algorithm: An interesting alternative to stabilize NM is to use the *linear prediction* method, a causal recursion method invented in the 1940's [9]. It seems likely to me that the use of Linear Prediction (LP) could greatly improve the convergence properties of NM. The down side is that the LP method assume the step-size only has poles, which in our case is clearly not true. The zeros of $P_N(s)$ bias the estimate in a negative manner. However when the trajectory steps near a pole, the LP algorithm should fit the data extremely well, thus removing the influence of the pole. This approach could be especially effective if there are several poles in proximity.

Appendix

Consider the monic polynomial $P_N(s)$, with $s, s_r, c_n \in \mathbb{C}$ and $n, k, N \in \mathbb{N}$:

$$P_N(s) = (s - s_r)^N + \sum_{k=1}^{N} c_{N-k}(s - s_r)^{N-k}, \tag{A.1}$$

where Taylor's formula is used to determine the coefficient vector $\boldsymbol{C} = [c_N, c_{N-1}, \cdots c_0]^T_{N\times 1}$

$$c_k = \frac{1}{k!}\frac{d^k}{ds^k}P_N(s)\bigg|_{s=s_r}. \tag{A.2}$$

Here $s = \sigma + \jmath\omega$ is called the *Laplace frequency*, as defined by the Laplace transform [1]. Depending on physical considerations, the coefficients c_k may be real or complex.

Assuming our initial estimate for the root s_0 is within the RoC (close to root s_r, we replace s_r with s_1 and s with s_0, since $|(s_1 - s_0)^k| \ll |(s_r - s_0)|$ for $k \geq 2 \in \mathbb{N}$. Here we have assumed that within the RoC, the higher order terms may be ignored.

Iterating we increase n by 1. Thus $s_0 \to s_1$ and $s_1 \to s_2$, so the truncated Taylor series becomes

$$P_N(s_2) \approx (s_2 - s_1)\,\frac{d}{ds}P_N(s)\bigg|_{s_2} + P_N(s_2). \tag{A.3}$$

Generalizing this for $n \gg 1$ we find replace find $|(s_{n+1} - s_n)^k| \ll |(s_1 - s_0)|$ for $k \geq n \in \mathbb{N}$, (i.e., $\epsilon_n = s_1 - s_0$ is within its RoC), thus we may truncate Eq. (A.1) to its linear term $n = 1$, resulting in the approximation Thus for large $n \to \infty$, $s_{n+1} \to s_r$, resulting in

$$\cancelto{0}{P_N(s_{n+1})} = (s_{n+1} - s_n)^N + \sum_{k=1}^{N} c'_k(s_{n+1})^{N-k} \tag{A.4}$$

Here c'_{n+1} is shorthand for $dP_N(s_{n+1})/ds$.

Solving for s_{n+1} gives Newton's method:

$$s_{n+1} = s_n - \frac{P_N(s_n)}{P'_N(s_n)}. \tag{A.5}$$

Importantly, if s_n approaches a root of $P'(s)$, the denominator can become arbitrarily large, resulting in a restart of the entire procedure.

On the other hand, if any estimate of the root s_n is close to a root of (i.e., $P_N(s_r \pm \epsilon) \approx 0$) then for $n \geq 2 \in \mathbb{N}$, $\epsilon = s_n - s_r$ is within the RoC. Namely for all $k \in \mathbb{N} + 1$

$$|(s_n - s_r)^k| \ll |(s_n - s_r)|. \tag{A.6}$$

This *complex analytic linearization step* is the key to Newton's method. *It will only be true if the difference equation remains linear, which requires Eq. (A.6).*

In summary: Newton's method is a linear approximation that critically depends on the RoC condition (Eq. (A.6)).

References

[1] J. B. Allen, *An Invitation to Mathematical Physics and Its History*. New York, NY, USA: Springer, 2020.

[2] J. B. Allen and M. M. Sondhi, "Cochlear macromechanics: Time-domain solutions," *J. Acoustical Soc. Am.*, vol. 66, no. 1, pp. 120–132, 1979.

[3] R. Boas, *Invitation to Complex Analysis*. New York, NY, USA: Random House, 1987.

[4] A. Galántai, "The theory of Newton's method," *J. Comput. Appl. Math.*, vol. 124, no. 1–2, pp. 25–44, 2000.

[5] D. Haar, *The Old Quantum Theory*. Amsterdam, The Netherlands: Elsevier, 2016.

[6] T. Kuhn, *Black-Body Theory and the Quantum Discontinuity, 1894–1912*. Oxford, U.K.: Oxford Univ. Press, 1978; Chicago, IL, USA: University of Chicago Press, 1987.

[7] J. Rinzel and R. N. Miller, "Numerical calculation of stable and unstable periodic solutions to the Hodgkin-Huxley equations," *Math. Biosci.*, vol. 49, no. 1, pp. 27–59, 1980.

[8] J. Stewart, *Essential Calculus: Early Transcendentals*. Boston, MA, USA: Cengage Learning, 2012. Accessed: Mar. 10, 2024. [Online]. Available: https://books.google.com/books?id=AcQJAAAAQBAJ

[9] P. P. Vaidyanathan, "The theory of linear prediction," *Synth. Lect. Signal Process.*, vol. 2, no. 1, pp. 1–184, 2007.

Jont B. Allen (Life Fellow, IEEE) received the Ph.D. degree in EE from the University of Pennsylvania, Philadelphia, PA, USA, in 1970. He had two employments: first was with the Acoustics Research Department, Bell Labs, from 1971 to 2003, where he learned signal process, and measured the properties of auditory neurons in the cat brain (Columbia University, New York), and developed a model of the inner ear (cochlea). During this period, he also did animal experiments on live cats, recording from the auditory nerve. This research resulted in more than 15 publications, some on modeling efforts, and others on the results derived from the neural responses (tuning curves and distortion products generated by the cochlea). In 2003, he retired from AT&T and moved with the ECE Department, University of IL, where he began the teaching career. The courses he taught are in a range of topics from auditory science to mathematics, which are detailed on his web page: https://auditorymodels.org. During this period, he wrote several books and many publications on speech features encoded in the acoustic speech signal. This work complemented the previous work at Bell Labs, covering the more practical side of human language processing.