NO PERIOD TWO IMPLIES CONVERGENCE,

OR

WHY USE TANGENTS WHEN SECANTS WILL DO?[†]

W. Kahan

Department of Mathematics, and
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley

October 1979

[†]Research sponsored by Office of Naval Research Contract N00014-76-C-0013.

NO PERIOD TWO IMPLIES CONVERGENCE,
OR
WHY USE TANGENTS WHEN SECANTS WILL DO?†

W. Kahan

Abstract: A familiar task is to solve $f(x) = 0$ given a continuously differentiable real function f. Newton's iteration could be tried; so could the Secant iteration. Except when the derivative f' costs appreciably less to evaluate than does f, the Secant iteration tends in practice to converge ultimately more efficiently than Newton's whenever both iterations converge to the desired root. When will they both converge? We find roughly that whenever Newton's iteration converges from every starting point in an interval I, so must the Secant iteration converge from every pair of starting points in I provided only that f actually reverses sign in I. This is an unexpected way for the Secant iteration to dominate Newton's.

The foregoing discovery was uncovered by techniques like those which produced the following one. Suppose $\phi(x)$ maps the closed finite interval I continuously into itself, and hence has at least one fixed point $x = \phi(x)$ in I, perhaps several. The iteration $x_{n+1} = \phi(x_n)$ need not necessarily converge, but it will converge in I from every starting point x_0 in I if and only if no two distinct points in I are exchanged by ϕ. Therefore the iteration converges only if it cannot be trapped in a cycle of period 2; on the other hand it is known that the existence of a cycle of "Period Three Implies Chaos" (T.-Y. Li & J. A. Yorke, Amer. Math. Monthly 82 1975). These last results and more can be found in a little-known paper by A. N. Šarkovskiĭ (1964) "Co-existence of cycles of a continuous mapping of the line into itself" (Russian) Ukrain. Mat. Ž 16 #1 (Math. Rev. 28 (1964) #3121).

§0. Introduction

This work is presented in two parts. Part I deals with the iteration $x_{n+1} = \phi(x_n)$ when $\phi(x)$ maps an interval I continuously into itself without exchanging two distinct points of I. The theorem that then the iteration converges is proved here not because the theorem is new (a proof by Bashurov and Ogibin (1966) has been translated into English) but because that proof contains ideas that will be needed in part II and, besides, deserve to be better known.

Part II deals with Newton's and the Secant iterations for finding where a continuously differentiable real function $f(x)$ vanishes. The results presented appear to be new and to provide further incentive, if any be needed, for preferring the Secant iteration over Newton's.

This introduction includes a summary of all subsequent sections' contents in order to help the browsing reader locate more easily whatever interests him. All proofs are terminated by the □ sign in order to help the casual reader skip over them.

Part I, §1 The No Swap Theorem

The main result is stated here together with some of its history, but not proved until §4. The result is due principally to Šarkovskiĭ. It is valid both on finite intervals and on infinite intervals regarded as line segments, but not valid on the projectively closed real axis (regarded as a circle) except in special cases.

§2 Two Conditions Equivalent to the No Swap Condition

These are technical details used only in §3.

§3 The One-Sided Condition

This condition, first articulated by Šarkovskiĭ (1965), is the most potent equivalent to the No Swap Condition.

§4 Proof of the No Swap Theorem and some Applications

Among the applications are validations of familiar conditions sufficient for convergence, and characterizations of the regions from which convergence is assured. Newton's iteration is seen to be in one sense a special case, in another sense more general than the iteration discussed above. The No Swap theorem is applied to show that Newton's iteration always converges when used to find a zero of a rational function whose poles and zeros are all real, simple, and interlace.

Part II, §5 Newton's and the Secant Iteration

This is where significantly new results begin. The two iterations are described, and first mention is made of a pathological discontinuity that must complicate matters (it partially invalidates some of the claims in the abstract above) even if we begin with an infinitely differentiable function $f(x)$. The Secant iteration's sparse literature and history are sketched briefly, and then five examples are presented to give the reader some feeling for the possibilities with which our theory must cope.

§6 Projective Invariance of Newton's and the Secant Iterations

Old but unfamiliar results culminate in a Mean Value lemma which binds the two iterations together more firmly than can any hand-waving about the Secant iteration being a "discretization" of Newton's.

§7 Inferences from $N(I) \subseteq I$

The hypothesis that $N(x) \equiv x - f(x)/f'(x)$ maps an interval I into itself, a prelude to the assumption that Newton's iteration converges, has profound consequences ranging from the monotonicity of f to the Darboux continuity of N, all of them needed in the next section's proofs.

§8 The No Swap Theorem for Newton's Iteration

Here are the necessary and sufficient conditions for Newton's iteration to generate from every starting point a sequence of iterates of which some subsequence converges to a zero of $f(x)$. An example shows that the subsequence complication is theoretically unavoidable though practically ignorable. One application of these conditions is a generalization of the familiar convexity conditions sufficient for convergence; we can allow at least one inflexion. Consequently certain financial calculations can be accomplished via Newton's (or the Secant) iteration with no need first to obtain safe starting values. Another application provides for rapid computation of all the real zeros of a sufficiently differentiable function (e.g. a polynomial) with no recourse to Sturm sequences.

§9 The Secant Iteration

Finally Part II's main result, which tells when the Secant iteration works as well as Newton's, is stated accurately and proved via a long sequence of ten propositions which exploit almost all that has gone before. A final example shows once again that subsequence complications are theoretically unavoidable though practically ignorable.

§10 Bibliography

PART I

§1. The No Swap Theorem

$\phi(x)$ is a function which maps a closed finite interval I continuously into itself. Since $x - \phi(x)$ cannot have the same non-zero sign at both ends of I it must vanish at least once in I; consequently ϕ must have at least one and possibly several fixed points $x = \phi(x)$ in I. A natural way to seek a fixed point is to iterate,

$$x_{n+1} = \phi(x_n) \quad \text{for} \quad n = 0,1,2,3,\ldots ;$$

but the iteration cannot always be expected to converge. For example, $x_{n+1} = \sin(2\pi x_n)$ almost never converges, the exceptions being a countable set of starting values x_0 from which one of the three fixed points $x = 0$ or $x = \pm.429368\ldots$ is reached after finitely many iterations.

Theorem: The condition necessary and sufficient for the iteration $x_{n+1} = \phi(x_n)$ to converge from every x_0 in I turns out to be

The No Swap Condition: No two distinct points in I are exchanged by ϕ; i.e. if $x = \phi(\phi(x))$ in I then $x = \phi(x)$ too.

This theorem will be vindicated below in stages designed to exhibit intermediate results which will be useful in Part II. But first we digress to discuss the condition's history and generality.

The No Swap theorem is equivalent to the assertion that $x_{n+1} = \phi(x_n)$ converges from every x_0 in I if and only if no x_0 in I leads to a sequence of iterates cycling on two points $x_{2n} = x_0$ and $x_{2n+1} = x_1 \neq x_0$. This theorem appears (not altogether correctly) in A.N. Sarkovskii (1960,1961) and is (correctly) elaborated upon in subsequent papers (1964,1965) wherein Šarkovskiĭ proves, among other things, that the integers can be re-ordered

thus

$$3,5,7,9,\ldots,2i+1,\ldots,$$

$$6,10,14,18,\ldots,4i+2,\ldots,$$

$$\ldots,$$

$$2^k3,2^k5,2^k7,2^k9,\ldots,2^k(2i+1),\ldots,$$

$$\ldots, \text{ and finally}$$

$$\ldots,2^j,2^{j-1},\ldots,8,4,2,1$$

in such a way that if from some $x_0 = x_0^{(m)}$ the iteration $x_{n+1} = \phi(x_n)$ cycles on m distinct points then there are other starting values $x_0 = x_0^{(m')}$ in I from which follow cycles of every length m' subsequent to m in the re-ordering. I am indebted to Prof. Rufus Bowen for references to Šarkovskiĭ's work, which seems to go beyond what has appeared recently in the English language; cf. Li and Yorke (1975), Stepleman (1975), and Bashurov and Ogibin (1966).

In this paper only those parts of Šarkovskiĭ's work that bear upon Part II will be repeated.

The No Swap theorem is stated above for a continuous map ϕ of a closed finite (i.e. compact) interval I to itself. Must I be both closed and finite? No; it could be neither. But a more general form of the No Swap theorem involves complications which obscure proofs already complicated enough. For instance, because the No Swap theorem neglects to mention that the iteration $x_{n+1} = \phi(x_n)$, when it converges, converges to a fixed point of ϕ, the theorem remains valid whether or not I includes its end-points; the example $\phi(x) \equiv x^2$ on the open interval $I \equiv \{0 < x < 1\}$ illustrates the possibility of convergence to an end-point not in I. The reader who wishes to prove the No Swap theorem for non-closed intervals I may do so by first adjoining to I any end at which $\lim(\phi(x) - x) = 0$ and then modifying in routine ways every reference to an end of I in §§2-4; fortunately no end

of I to which ϕ cannot be continued continuously figures significantly in the theorem.

I does not have to be finite, but when I is infinite ambiguities can arise concerning the meanings of "continuous" and "convergent", and whether ∞ can be a fixed point (e.g. of $\phi(x) = x+1$), and whether $+\infty$ differs from $-\infty$. To avoid these ambiguities and other circumlocutions in subsequent sections of this paper we have assumed without loss of generality that I is a finite interval. Only in this section, §1, do we digress to justify that assumption by discussing infinite intervals.

The No Swap theorem remains valid when ϕ is a continuous map of an infinite or semi-infinite interval I to itself provided the word "converge" be understood to include possible "convergence" to ∞. This is so because a suitable change of variables will transform the infinite interval I into a finite one. For instance, suppose $\phi(x)$ maps $I \equiv \{0 \le x \le +\infty\}$ continuously to itself. The new variables $y \equiv (x-1)/(x+1)$ and $\psi(y) \equiv (\phi(x)-1)/(\phi(x)+1)$ @ $x = (1+y)/(1-y)$ exhibit the corresponding continuous map ψ of the transformed interval $J \equiv \{-1 \le y < 1\}$ to itself. The example $\phi(x) \equiv x+1$ with an attractive fixed point at $x = +\infty$ corresponds to $\psi(y) = (1+y)/(3-y)$ with an attractive fixed point at $y = 1$; just as $y_{n+1} = \psi(y_n)$ converges to 1, so must $x_{n+1} = \phi(x_n)$ "converge" to $+\infty$. The foregoing change of variables is an instance of a bilinear rational transformation appropriate when $\phi(I)$'s closure is not the whole real axis. Otherwise, when $\phi(I)$'s closure is the whole real axis, non-rational changes of variables are more appropriate. One example is $y = \tanh x$ which maps the affinely closed real axis $I \equiv \{-\infty \le x \le +\infty\}$ onto the closed finite interval $J \equiv \{-1 \le y \le 1\}$, and transforms a function $\phi(x)$ continuous at all real x into $\psi(y) \equiv \tanh(\phi(\operatorname{arctanh}(y)))$ which is continuous at least in J's interior. Hence the following ostensibly more general form of the No Swap theorem is true:

If I is a line segment, including or not including its ends, finite or infinite, and if ϕ maps I continuously to itself, then the iteration $x_{n+1} = \phi(x_n)$ converges from every x_0 in I if and only if ϕ exchanges no two distinct points of I.

A line segment, which has two ends that may or may not be regarded as part of that line segment, is quite different from a circle which has no end at all. The No Swap theorem is not generally applicable on a circle, and consequently not generally applicable to real functions $\phi(x)$ which, like rational functions, are continuous in an extended sense on the projectively closed real axis $I \equiv \{\text{all real numbers } x\} \cup \{\infty\}$. The usual change of variables is the Stereographic Projection (Fig. 1) $y \equiv 2 \arctan(x) \bmod 2\pi$ which maps the projectively closed real axis I onto the circle $C \equiv \{-\pi \leq (y \bmod 2\pi) < \pi\}$, and transforms $\phi(x)$ into $\psi(y) \equiv 2 \arctan(\phi(\tan y/2)) \bmod 2\pi$ which must map C continuously to itself whenever ϕ is a rational function or, more generally, whenever either $\phi(x)$ is continuous or $1/\phi(x)$ is continuous at every real x, and either $\phi(1/\omega)$ or $1/\phi(1/\omega)$ is continuous at $\omega = 0$. But ψ may lack a fixed point in C; take for example $\psi(y) \equiv y+1 \bmod 2\pi$, the transform of the rational function $\phi(x) \equiv \tan(\frac{1}{2} + \arctan x)$ $= (x + \tan\frac{1}{2})/(1 - x\tan\frac{1}{2})$, which has neither cycle nor fixed point. Another violator of the No Swap theorem is $\psi(y) \equiv -2y \bmod 2\pi$, the transform of $\phi(x) \equiv 2x/(x^2-1)$, which also swaps no two distinct points but does have three repulsive fixed points to which iteration converges only from a countable set of starting points. A final example $\phi(x) \equiv x(x+2)/((1-x)(x^2+2x+2))$ never swaps two distinct points and has just one fixed point to which the iteration $x_{n+1} = \phi(x_n)$ converges from most starting points, but the iteration defies the No Swap theorem by cycling through one set of eleven points starting at $x_0 = -3.2956364...$ and through another eleven starting at $x_0 = -4.2078536...$ (there are no shorter cycles). See Fig. 2.

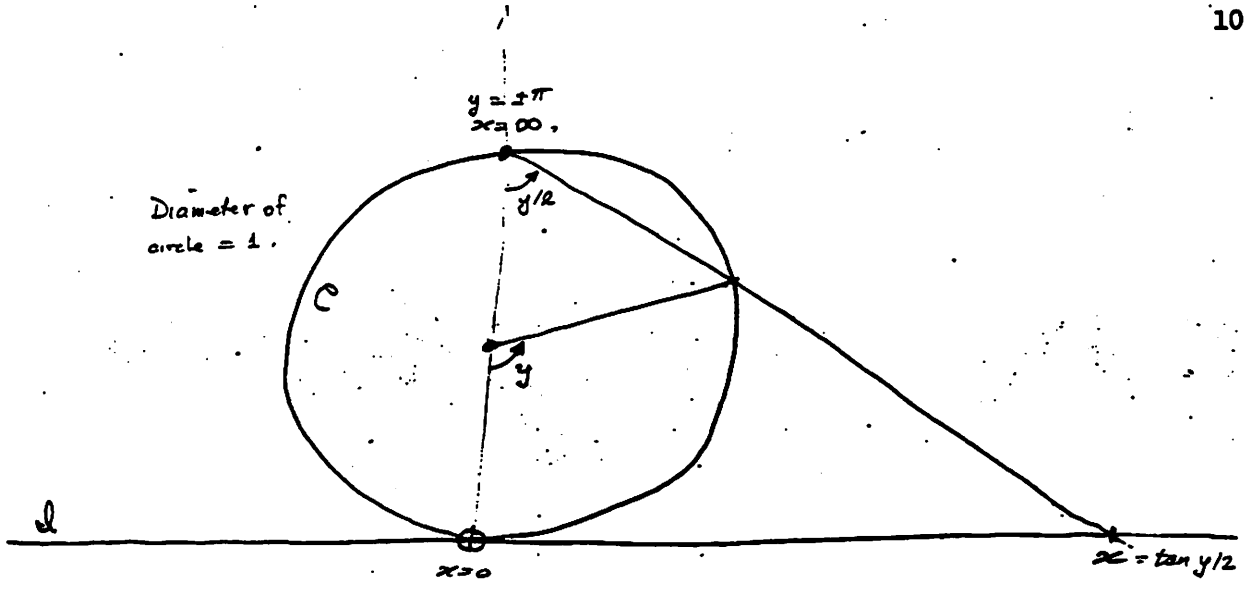


Fig 1: Stereographic projection of projectively closed real axis ℓ onto circle C.

$x = \tan y/2 \qquad y = 2 \arctan x \quad , -\pi < y \bmod 2\pi \leq \pi$

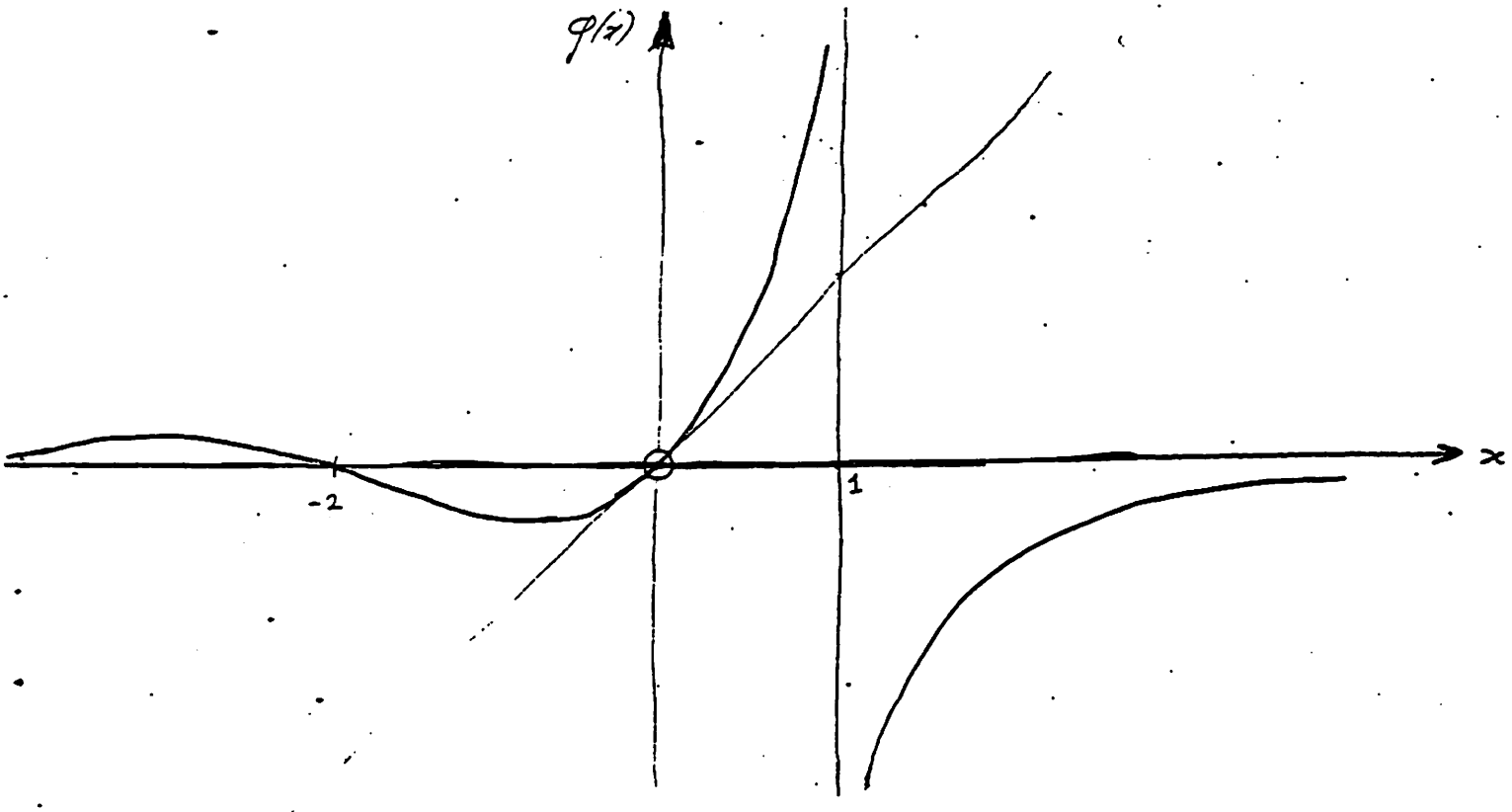


Fig 2: $\varphi(x) = x(x+2)/((1-x)(x^2+2x+2))$

Two useful special circumstances are known when the No Swap theorem may be applied to a continuous map ψ of the circle C to itself. One we have already discussed is the case when $\psi(C)$ is a proper sub-arc of C in which case C may be transformed into a line segment on which ψ is continuous by cutting C at any point not in $\psi(C)$. Such a case is exemplified by a rational function $\phi(x) \equiv (x^{2m}-1)/(x^{2m+1}-1)$ with $m \geq 1$ that maps the projectively closed real axis into the interval $-1 < \phi(x) \leq 1$ without swapping two distinct points, whence the No Swap theorem implies that $x_{n+1} = \phi(x_n)$ converges from any x_0. The second special circumstance arises when ψ has in C at least one fixed point $\eta = \psi(\eta)$ at which cutting C yields a line segment on which ψ remains continuous and therefore entitled to the No Swap theorem. Such a circumstance arises whenever ψ never winds past η, i.e. whenever the equation $\psi(y) = \eta$ has no solution y other than perhaps $y = \eta$ across which the expression $\psi(y) - \eta$ changes sign. One example is the rational function $\phi(x) \equiv (x^3+1)/(x-1)^2$ whose unique fixed point ∞ is approached from just one side as $x \to 1$; consequently the projectively closed real axis may be cut at ∞ to produce the affinely closed real axis $\{-\infty \leq x \leq +\infty\}$ which ϕ maps "continuously" to itself in a way which justifies the inference from the No Swap theorem that $x_{n+1} = \phi(x_n)$ $\to +\infty$ from every real x_0. Another rational example is $\phi(x) \equiv (x^{2m-1}-1)/(x^{2m}-1)$ with $m \geq 1$ which can be shown to map the projectively closed real axis one-to-one onto itself without swapping two points; moreover ϕ has two fixed points, an attractive one between $2^{-1/(2m-1)}$ and $2^{-1/(2m-1/2)}$ and a repulsive one between -2 and $-2^{1/(m+1/2)}$. After the circle is cut at the latter fixed point the No Swap theorem implies that $x_{n+1} = \phi(x_n)$ will converge through the extended reals (possibly including one $x_n = \infty$ and the next $x_{n+1} = 0$) to the attractive fixed point from every starting point x_0 except the repulsive fixed point. This example and the one before last will reappear in §5.

Finally, note that no generality is gained by allowing the closed interval I mapped continuously to itself by ϕ to be a subset of the real axis instead of all of it, because ϕ could be defined outside I by extending ϕ's graph horizontally from its ends. I is significant only in so far as it represents that part of ϕ's domain in which the No Swap condition is satisfied, outside which the condition might be violated. I's significance will become clearer in §4 during the discussion of catchment basins.

Here ends the digression concerning infinite intervals. Henceforth until §5 assume I is a closed finite interval mapped continuously to itself by ϕ.

§2. Two Conditions Equivalent to the No Swap Condition

The first such equivalent condition to be considered is

The No Separation Condition: No z in I can strictly separate $\phi(z)$ from $\phi(\phi(z))$; i.e. either $\phi(\phi(z)) \leq z \leq \phi(z)$ or $\phi(z) \leq z \leq \phi(\phi(z))$ in I implies $\phi(\phi(z)) = z = \phi(z)$.

Since the No Swap condition is an obvious implication of the No Separation condition, our task is to verify that the former condition implies the latter, so assume that the No Separation condition is violated by, say, $\phi(\phi(z)) < z < \phi(z)$ in I and we shall exhibit a consequent violator v of the No Swap condition. See Fig. 3.

Since $\phi(x) - x$ takes opposite signs at $x = \phi(z)$ and at $x = z$, $\phi(x)$ must have a fixed point $y = \phi(y)$ strictly between z and $\phi(z)$. Similarly, as x runs down from z to I's left-hand end-point, $\phi(\phi(x)) - x$ runs from a negative value at $x = z$ to a non-negative value ($\phi(\phi(x))$ maps I's left-hand end-point into I), so $\phi(\phi(x))$ must have a first fixed point $v = \phi(\phi(v)) < z$; by "first" is meant that $\phi(\phi(x)) - x < 0$ for $v < x \leq z$.

Can $v = \phi(v)$? No; otherwise $\phi(x) - y$ would be positive at $x = z$ and negative at $x = v$, in which case we should have $\phi(u) - y = 0$ at some u strictly between v and z and then $\phi(\phi(u)) - u = \phi(y) - u = y - u > 0$ contradicting our choice of v as the first value of $x < z$ for which $\phi(\phi(x)) - x \geq 0$. Therefore $v = \phi(\phi(v)) \neq \phi(v)$ violates the No Swap condition. □

The second condition equivalent to the No Swap condition will be called

The No Crossover Condition: If $\phi(v) \leq u \leq v \leq \phi(u)$ in I then $\phi(v) = u = v = \phi(u)$ too.

Since this condition obviously implies the previous two, our task now is to verify that they imply this one, which we shall do by inferring from a violation $\phi(v) < u < v < \phi(u)$ of the No Crossover condition that there exists a violation z of the No Separation condition. See Fig. 4.

Consider $\phi(\phi(v))$. If $\phi(\phi(v)) \geq v$ then v violates the No Separation condition. Otherwise, if $\phi(\phi(v)) < v$, we plot $\phi(x) - v$ as x runs through $\phi(v) \leq x \leq u$. Since $\phi(x) - v < 0$ at $x = \phi(v)$ and $\phi(x) - v > 0$ at $x = u$ there must exist some z in $\phi(v) < z < u$ with $\phi(z) = v$, and this z violates the No Separation condition because

$$\phi(\phi(z)) = \phi(v) < z < u < v = \phi(z) .$$ □

We shall use the No Crossover condition in lieu of the No Swap condition to prove inferences from the latter.

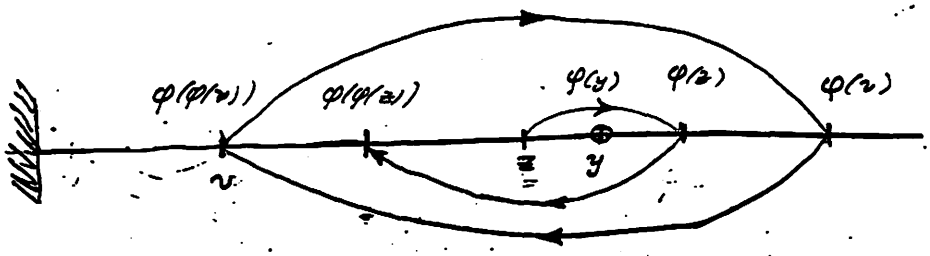


Fig. 3 : If z violates the No Separation condition then v must violate the no-swap condition.

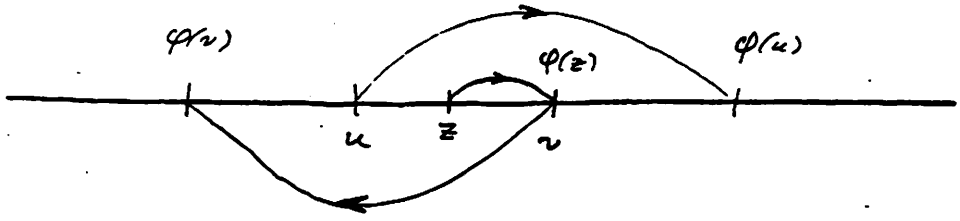


Fig. 4 : If u and v violate the No Crossover condition then z must violate the No Separation condition.

§3. The One-Sided Condition

We return now to the iteration $x_{n+1} = \phi(x_n)$ and show that the No Swap condition is equivalent to another condition first articulated by Šarkovskiĭ (1965) and used also by Bashurov and Ogibin (1966, Lemma 2).

The One-Sided Condition: Whenever $x_1 = \phi(x_0) \neq x_0$ in I all subsequent iterates $x_{n+1} = \phi(x_n)$ also differ from x_0 and lie on the same side of x_0 as does x_1.

Since the No Separation condition is an obvious inference from the One-Sided condition, our task is to infer the converse by assuming that x_0 violates the One-Sided condition and deducing that some subsequent iterates violate the No Crossover condition.

Assume for definiteness that $x_1 = \phi(x_0) > x_0$ but that some subsequent iterate $x_m \leq x_0$. We may take x_m to be the first such iterate and then have $x_1, x_2, \ldots, x_{m-1}$ all greater than x_0. Next let x_k be the first of these iterates to satisfy $x_k \geq x_{m-1}$; now $x_m \leq x_0 \leq x_{k-1} < x_{m-1} \leq x_k$, which exhibits x_{k-1} and x_{m-1} as violators of the No Crossover condition. □

The One-Sided condition has been interpreted above as a property which an iterating function $\phi(x)$ can possess if and only if ϕ also satisfies the No Swap condition. But the One-Sided condition can also be regarded as a property of sequences irrespective of their genesis:

The One-Sided Condition is satisfied by the sequence $\{x_0, x_1, x_2, \ldots\}$ whenever each member x_n of the sequence lies on the same side of all subsequent members x_{n+m}, $m > 0$; i.e. each x_n satisfies

$$x_n < x_{n+m} \quad \text{for all } m > 0,$$

$$\text{or} \quad x_n > x_{n+m} \quad \text{for all } m > 0,$$

$$\text{or else } x_n = x_{n+m} \quad \text{for all } m > 0.$$

In particular, if the sequence $x_{n+1} = \phi(x_n)$ is generated by a One-Sided iterating function $\phi(x)$ then the sequence must be One-sided too. Such a sequence is the subject of the following lemma.

No-Man's Land Lemma: If the sequence $\{x_0,x_1,x_2,\ldots\}$ satisfies the One-Sided condition, and if it is not ultimately monotonic*, then the sequence can be partitioned into two disjoint infinite subsequences one of which increases strictly monotonically to a limit $\check{x}$ while the other decreases strictly monotonically to a limit $\hat{x}$, and $\hat{x} \geq \check{x}$.

Proof: The increasing subsequence consists of those x_n which satisfy $x_n < x_{n+1}$ and the decreasing subsequence consists of those x_n which satisfy $x_n > x_{n+1}$. For instance, if x_m is a local maximum and x_l the subsequent local minimum in the sequence, so that

$$\cdots x_{m-1} < x_m > x_{m+1} > \cdots > x_{l-1} > x_l < x_{l+1} \cdots \qquad (m < l)$$

then x_{m-1} and x_l are consecutive members of the ascending subsequence (note that $x_{m-1} < x_l$ because of the One-Sided condition) while $x_m,x_{m+1},\ldots,x_{l-1}$ are consecutive members of the descending subsequence. Evidently each subsequence is strictly monotonic and bounded by the other, and consequently each subsequence converges to a limit which separates it from the other. If the limits $\hat{x}$ and $\check{x}$ are different they are separated by a no-man's land which no member x_n of the sequence may enter. □

*"ultimately monotonic" means that either $x_{n+1} \geq x_n$ for all sufficiently large n or else $x_{n+1} \leq x_n$ for all sufficiently large n.

§4. Proof of the No Swap Theorem and some Applications

Theorem: Suppose I is a closed finite interval mapped continuously to itself by ϕ. Then the iteration $x_{n+1} = \phi(x_n)$ converges in I from every x_0 in I if and only if any of the following equivalent conditions is satisfied.

The No Swap Condition: If $x = \phi(\phi(x))$ in I then $x = \phi(x)$.

The No Separation Condition: Either $\phi(\phi(z)) \le z \le \phi(z)$ or $\phi(z) \le z \le \phi(\phi(z))$ in I implies $\phi(\phi(z)) = z = \phi(z)$.

The No Crossover Condition: If $\phi(v) \le u \le v \le \phi(u)$ in I then $\phi(v) = u = v = \phi(u)$.

The One-Sided Condition: Whenever $x_1 = \phi(x_0) \neq x_0$ in I all subsequent iterates $x_{n+1} = \phi(x_n)$ also differ from x_0 and lie on the same side of x_0 as does x_1.

Proof: The equivalence of these four conditions has already been established, and the necessity of the No Swap condition is evident since otherwise the iteration could cycle on two distinct points of I. All that is left is to show that these conditions suffice to ensure convergence.

The sequence of iterates x_n must be a One-Sided sequence. Therefore, according to the No-Man's Land lemma, it is either ultimately monotonic and therefore convergent in I, or else it can be partitioned into two disjoint infinite subsequences, one ascending to a limit $\check{x}$ and the other descending to a limit $\hat{x} \ge \check{x}$. Is $\hat{x} > \check{x}$? No. Those iterates x_n which belong to the descending subsequence that converges to $\hat{x}$ are followed by iterates $x_{n+1} = \phi(x_n)$ which must, because ϕ is continuous, converge to $\phi(\hat{x})$; and yet at least some of those iterates x_{n+1} belong to the ascending

subsequence and must converge also to $\check{x}$. Therefore $\check{x} = \phi(\hat{x})$, and similarly $\hat{x} = \phi(\check{x})$. Finally the No Swap condition implies $\check{x} = \hat{x}$ and the theorem is proved; $x_n \to \check{x} = \hat{x}$. □

The No Swap theorem may be applied in several ways to decide whether an iteration $x_{n+1} = \phi(x_n)$ converges. One way is graphical; the graphs of $y = \phi(x)$ and $x = \phi(y)$ are mirror images by reflection in the mirror $y = x$, and only if those graphs intersect nowhere but on the mirror is the No Swap condition satisfied by ϕ. This technique was used to settle a conjecture by Stepleman (1975, p. 894) that

$$x_{n+1} = x_n(\sin(1/x_n) - 1/8) \qquad (\sin(\text{radians}))$$

would converge to zero from all x_0, or at least from all sufficiently small x_0. The transformation $X_n = 180/(\pi x_n)$ converts the iteration into an equivalent form

$$X_{n+1} = X_n/(\sin(X_n) - 1/8) \qquad (\sin(\text{degrees}))$$

which is easier to deal with both numerically and graphically. Then we find cycles $X_2 = X_0 \neq X_1$ in abundance, for instance $X_0 = 1355.5094^0\ldots$ and $X_1 = -1209.865^0\ldots$, or $X_0 = 1723.154^0\ldots$ and $X_1 = -1568.269^0\ldots$, so Stepleman's conjecture is false. But his iteration appears always to converge in the presence of roundoff.

Another way to apply the No Swap theorem is algebraic, applicable when $\phi(x)$ is a rational function. Then $1 + (\phi(\phi(x)) - \phi(x))/(\phi(x) - x)$ is also a rational function, and only if it has no zeros in I which are not also zeros of $\phi(x) - x$ is the No Swap condition satisfied by ϕ. Therefore the No Swap condition can be tested by removing some common divisors from certain polynomials and then invoking Sturm sequences to decide whether the

polynomials change sign in I. The condition $\phi(I) \subseteq I$ can also be tested by using Sturm sequences to see whether certain polynomials change sign in I. The details have been worked out by R.J. Fateman (1977), who has written a computer program that runs on M.I.T.'s MACSYMA system and realizes the following assertion:

> When $\phi(x)$ is a rational function the question, whether $x_{n+1} = \phi(x_n)$ converges in I from every x_0 in I, can be decided by a finite number of rational arithmetic operations without solving any polynomial equations.

Other applications of the No Swap theorem include easy validation of conditions sufficient for convergence in I from every x_0, three examples of which are these:

i) $|\phi(u) - \phi(v)| < |u-v|$ for all distinct u and v in I separated by the (it turns out to be unique) fixed point of ϕ in I's interior.

ii) $-1 < (\phi(u) - \phi(v))/(u-v)$ for all distinct u and v in I separated by one of ϕ 's fixed points.

iii) ϕ has in I just one fixed point that divides I into at most two sub-intervals at least one of which is mapped to itself by ϕ.

In all three cases I is presumed to be mapped continuously to itself by ϕ.

The reader is asked to verify that each of the foregoing three conditions implies the No Swap condition, keeping in mind that ϕ must have a fixed point between any two points of I that ϕ swaps. Although the three conditions refer to ϕ's fixed point(s) the conditions do not require that any fixed point's location be known; for example, if ϕ is differentiable one of the inequalities $|\phi'| < 1$ or $\phi' > -1$ or $\phi' > 0$ respectively would suffice.

A troublesome problem encountered frequently in practice arises when conditions somewhat like those above are known to be satisfied by ϕ in some interval J which is *not* known to be mapped to itself by ϕ. This problem, locating a suitable sub-interval mapped into itself, is so important that we digress here to discuss how the No Swap theorem can shed light on it.

First some terminology. The interval J is the domain of a continuous real function $\phi(x)$. A sub-interval X of J is called *an interval attracted to* ξ whenever from every x_0 in X the iteration $x_{n+1} = \phi(x_n)$ converges to the fixed point $\xi = \phi(\xi)$ even though perhaps some iterates may lie outside X (but all lie in J); we do not insist that $\phi(X) \subseteq X$. The fixed point $\xi = \phi(\xi)$ is called *attractive* whenever it belongs to some *non-degenerate* interval (one that contains interior points) attracted to ξ. The *catchment basin* $X(\xi)$ belonging to an attractive fixed point ξ is the largest interval X containing ξ and attracted to ξ. (When ξ is an end-point of its catchment basin most other writers would call ξ a *one-sided* attractive fixed point and reserve the unmodified term "attractive fixed point" for one which lies in its catchment basin's interior.) One property of $X(\xi)$ is that

$$\xi \in \phi(X(\xi)) \subseteq X(\xi) \ ;$$

this is true because $X(\xi)$ is just that connected component containing ξ of the union of all intervals attracted to ξ, while $\phi(X(\xi))$ is one of the intervals attracted to and containing ξ in that union. That property does not characterize $X(\xi)$ because other larger and smaller intervals possess the same property; given any sub-interval $K \subset X(\xi)$, the smallest sub-interval $X \supseteq K$ with $\xi \in \phi(X) \subseteq X$ turns out to be

$$X = (\text{convex hull}(\{\xi\} \cup K) \cup \phi(\text{convex hull}(\{\xi\} \cup K)) \subset X(\xi)$$

as can be verified easily with the aid of the One-Sided condition valid throughout $X(\xi)$. (Cf. Theorem 2 of Bashurov and Ogibin (1966).) Consequently the problem of locating a suitable subinterval X mapped into itself in $X(\xi)$ is reduced to the problem of deciding when a sub-interval K lies between the ends of $X(\xi)$; the crucial question is just

Where in J are the ends of ξ's catchment basin $X(\xi)$?

Lemma: The closure of $X(\xi)$ is the largest closed interval containing ξ whose every interior point x satisfies

1) x lies inside J, and

2) $\phi(x)$ lies in J, and

3) $(\phi(\phi(\phi(x))) - \xi)/(\phi(x) - \xi) < 1$ if $\phi(x) \neq \xi$.

Consequently the ends of $X(\xi)$ lie among those points x which satisfy

-1) x is an end of J, or

-2) $\phi(x)$ is an end of J, or

-3) $\phi(\phi(\phi(x))) = \phi(x)$.

Proof: $X(\xi)$ can contain no fixed point other than ξ of $\phi(x)$ nor of $\phi(\phi(x))$ nor of $\phi(\phi(\phi(x)))$ ...; on the contrary the One-Sided condition satisfied by ϕ throughout $X(\xi)$ implies that both of $\phi(\phi(x))$ and $\phi(\phi(\phi(x)))$ must lie on strictly the same side of $\phi(x)$ as does ξ provided $\phi(x) \neq \xi$, and hence 3) is satisfied. Moreover each end of $X(\xi)$ which is not an end of J cannot be mapped by ϕ into $X(\xi)$'s interior, but must be mapped onto either itself or the other end of $X(\xi)$, and hence must satisfy -2) or -3).

This is a convenient place to tabulate the six ways by which the ends of ξ's catchment basin $X(\xi)$ may be recognized. We shall denote $X(\xi)$'s

ends by η and ζ, not necessarily both different from ξ.

Cases 1-3: $\zeta = \phi(\zeta)$ is a fixed point at one end of $X(\xi)$ whose other end η is

Case 1: an end η of J with $\phi(\eta)$ in $X(\xi)$'s interior, or

Case 2: another fixed point $\eta = \phi(\eta) \neq \zeta$, or

Case 3: mapped onto $\zeta = \phi(\eta)$.

These three cases are the only ones which allow ξ to lie at one end (ζ) of $X(\xi)$.

Cases 4-5: ξ lies strictly inside the closed interval $X(\xi)$ one of whose ends, ζ, is an end of J with $\phi(\zeta)$ in $X(\xi)$'s interior and the other end η is

Case 4: also an end η of J with $\phi(\eta)$ in $X(\xi)$'s interior, or

Case 5: mapped onto the first end $\zeta = \phi(\eta)$.

Case 6: ξ lies strictly inside the open interval $X(\xi)$ whose ends $\eta = \phi(\zeta)$ and $\zeta = \phi(\eta)$ are swapped by ϕ.

The lemma's proof will conclude with demonstrations appropriate to the six cases that $X(\xi)$'s closure cannot be a proper sub-interval of a larger interval X whose every interior point satisfies 1), 2) and 3). Each case will be ruled out either on the grounds that any end of $X(\xi)$ interior to X would then violate 1), 2) or 3) or else because some open neighborhood of an end of $X(\xi)$ inside X would then be attracted to ξ contrary to $X(\xi)$'s definition as the largest interval containing ξ attracted to ξ.

Suppose, then, that $X(\xi)$'s closure is a proper sub-interval of a larger interval X whose every interior point x (including at least one of $X(\xi)$'s ends) satisfies 1), 2) and 3). Case 4 is ruled out by 1), case 6 by 3), and case 5 by the observation that any sufficiently small open neighborhood of η inside X that satisfies 2) must be mapped by ϕ

into $X(\xi)$ and therefore ought to belong to $X(\xi)$. Case 1 is ruled out either by 1) (so η is not inside X) or 3) unless $\zeta = \xi$; and case 2 persists only if one of the fixed points at the end of $X(\xi)$, say η, is an end of X too while the other end $\zeta = \xi$ is inside X; and case 3 persists only if $\zeta = \xi$ is inside X too. In summary, the persistent cases are now these (see Fig. 5):

Cases 1' & 2': $\xi = \phi(\xi)$ lies at one end of $X(\xi)$ and inside X, but the other end η of $X(\xi)$ is an end of X too.

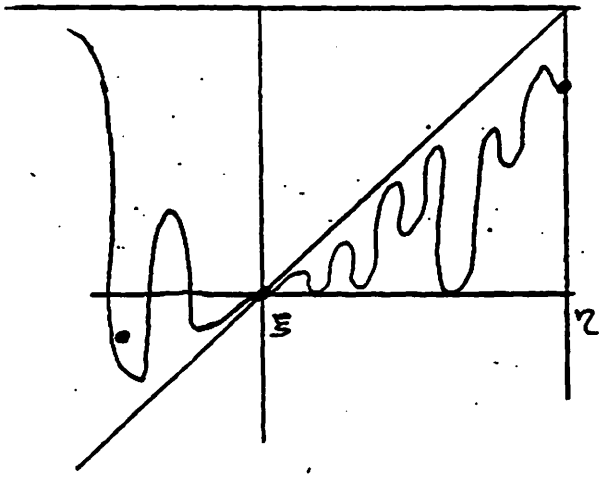
Case 3': $\xi = \phi(\xi)$ lies at one end of $X(\xi)$, the other end η is mapped onto the first $\xi = \phi(\eta)$, and at least one of these ends lies inside X.

To dispose of these remaining cases introduce $Y \equiv \phi(X) \cup X(\xi)$. Evidently $Y \not\subseteq X(\xi)$'s closure; otherwise an open neighborhood around one of $X(\xi)$'s ends would be mapped into $X(\xi)$ and would consequently be attracted to ξ contradicting $X(\xi)$'s maximality. In fact, every open neighborhood around one of $X(\xi)$'s ends inside X must be mapped by ϕ onto an interval part of which lies outside $X(\xi)$'s closure, and since each end of $X(\xi)$ inside X is mapped onto ξ we conclude that Y contains some open neighborhood N around ξ. Moreover, every $y \neq \xi$ inside that neighborhood N must satisfy

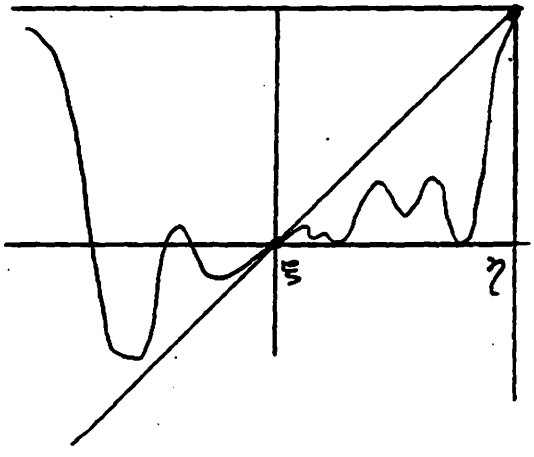
$$3') \qquad (\phi(\phi(y)) - \xi)/(y-\xi) < 1$$

either because y lies in $X(\xi)$ where ϕ is One-Sided or because $y = \phi(x)$ for some x in X's interior where 3) is satisfied. If N be chosen small enough to keep $\phi(\phi(y))$ no farther from $\xi = \phi(\phi(\xi))$ than η, inequality 3') above will force the iteration $x_{n+2} = \phi(\phi(x_n))$ to converge to ξ from any x_0 in N either monotonically through N outside $X(\xi)$

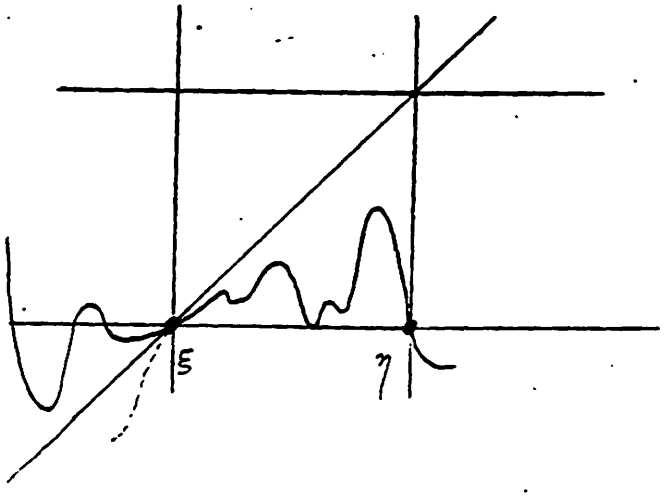
Fig. 5: Illustrations for the proof of the Lemma



Case 1'



Case 2'



Case 3'

or ultimately monotonically within $X(\xi)$. (Note that $0 \leq (\phi(x) - \xi)/(x-\xi) < 1$ for all x inside $X(\xi)$ in all three cases 1', 2' and 3' because then ϕ has no fixed point inside $X(\xi)$.) Since ϕ is continuous at ξ the iteration $x_{n+1} = \phi(x_n)$ must converge to ξ too, and from any x_0 in N, contradicting $X(\xi)$'s maximality again. □

The reader may well wonder why the lemma's condition 3) was not written in the simpler form

3') $$(\phi(\phi(x)) - \xi)/(x-\xi) < 1 \quad \text{if} \quad x \neq \xi \ .$$

One reason is that the lemma so modified could be contradicted by a counter-example falling into case 3' in the proof; take $\xi = 0$, $\phi(x) = x(1-x)$, $\eta = 1$, and the positive real axis for X in which 3') is satisfied. Another reason is that conditions like 3) have already appeared in the literature; one instance is

Theorem 3.2 of Stepleman (1975, p. 891): Suppose $\phi(x)$ is continuous throughout some interval containing in its interior a fixed point $\xi = \phi(\xi)$ but no other solution x of $\phi(x) = \xi$; then ξ is a point of (two-sided) attraction for the iteration $x_{n+1} = \phi(x_n)$ if and only if

$$|\phi(\phi(\phi(x))) - \xi| / |\phi(x) - \xi| < 1 \quad \text{whenever} \quad \phi(x) \neq \xi$$

throughout some open neighborhood around ξ.

(The hypothesis that $\phi(x) \neq \xi$ whenever $x \neq \xi$ in the interval cannot be dropped without abandoning the theorem to counter-examples which answer negatively Stepleman's "open question" following his Example 3.3 on p. 891. Incidentally, that example is wrong.)

None the less, the lemma is defective for practical purposes in so far as it sacrifices Applicability to Elegance. Equation -3) could better by replaced by the pair

-3') $$\phi(\phi(x)) = x \quad \text{or} \quad \phi(\phi(x)) = \phi(x) \ ,$$

and inequality 3) by the pair

3") $$\frac{\phi(\phi(x)) - \xi}{x - \xi} < 1 \quad \text{and} \quad \frac{\phi(\phi(x)) - \xi}{\phi(x) - \xi} < 1 \quad \text{if} \quad \phi(x) \neq \xi$$

that attest to the One-Sidedness of ϕ in $X(\xi)$. The reader is asked to verify, by retracing its proof, the truth of the lemma when 3) and -3) are replaced respectively by 3") and -3'), noting that when 3") is satisfied throughout a non-degenerate interval containing ξ so too must $(\phi(x) - \xi)/(x-\xi) < 1$ at every $x \neq \xi$ in that interval. Replacing -3) by -3') improves the lemma because the latter's solution-set is never larger and often smaller than the former's, and therefore -3') usually costs less to solve than -3). And inequalities 3") cost less to test than 3) partly because $\phi(\phi(\phi(x)))$ is such a mess. Finally, the modified lemma is accompanied by convergence theorems, analogous to Stepleman's but simpler, of which the following is an example.

Corollary: $\xi = \phi(\xi)$ is an attractive fixed point of the continuous real function ϕ if and only if 3") is satisfied throughout some non-degenerate interval containing ξ; and ξ is attractive if it lies strictly inside an interval throughout which 3")'s first inequality is satisfied.

Proof: 3") is necessary because ϕ must be One-Sided throughout $X(\xi)$. That 3") is sufficient too will follow after we find an interval I with

$\xi \in \phi(I) \subseteq I$ throughout which $(\phi(x) - \xi)/(x-\xi) < 1$ if $x \neq \xi$. Then ϕ will have no fixed point in I except ξ and will satisfy the No Swap condition, and the iteration $x_{n+1} = \phi(x_n)$ will have to converge to ξ from every x_0 inside I.

Let X be the non-degenerate interval containing ξ mentioned in the corollary and partition X into two sub-intervals which have only ξ in common; say $X = X_L \cup X_R$ where $X_L \equiv \{x \leq \xi \text{ in } X\}$ and $X_R \equiv \{x \geq \xi \text{ in } X\}$ and at least one of X_L or X_R has interior points. 3")'s first inequality, valid inside X_L and X_R, prevents any point of X except ξ from being a fixed point of ϕ; consequently $(\phi(x) - \xi)/(x-\xi) - 1$ cannot change sign inside X_L or X_R and must be negative to avoid contradicting 3")'s first inequality in some neighborhood of ξ. Thus, if $X \supseteq \phi(X)$, or if non-degenerate $X_L \supseteq \phi(X_L)$ or $X_R \supseteq \phi(X_R)$, the choice of I that completes the proof is obvious. Otherwise there are two cases to consider.

The boundary case: Either $X_L = \{\xi\}$ or $X_R = \{\xi\}$. Say the former; then ξ lies at one end of the non-degenerate interval X_R but $\phi(X_R) \subseteq X_R$. However, $\phi(X_R) \leq (\sup x \text{ in } X_R)$ because $(\phi(x) - \xi)/(x-\xi) < 1$ inside X_R, and also $\phi(\phi(X_R)) \leq (\sup x \text{ in } X_R)$ because of 3")'s first inequality. On the other hand, $\phi(\phi(X_R)) \geq (\inf x \text{ in } \phi(X_R))$ because of 3")'s second inequality, which implies $(\phi(x) - \xi)/(x-\xi) < 1$ inside $\phi(X_R)$. Choose $I = \phi(X_R) \cup X_R$ to complete the proof.

The interior case: ξ lies inside X, and only the first of 3")'s inequalities is assumed by hypothesis to hold inside X_L and X_R. We have already dealt with the possibilities $\phi(X) \subseteq X$, $\phi(X_L) \subseteq X_L$ or $\phi(X_R) \subseteq X_R$; the only possibility left that is compatible with the inequality $(\phi(x) - \xi)/(x-\xi) < 1$ valid inside X_L and X_R is either $X_L \subset \phi(X_R)$ or $X_R \subset \phi(X_L)$. Say the former; then let the non-degenerate

interval Y be that component of $\phi^{-1}(X_L) \cap X_R$ containing ξ, and let $I \equiv X_L \cup Y = \phi(Y) \cup Y \subset X$, observing that $\phi(I) \subseteq I$ because of 3")'s first inequality valid inside Y. □

Here ends the discussion of the No Swap theroem for arbitrary continuous iterating functions ϕ. The rest of the paper concerns Newton's and the Secant iterations for solving an equation $f(x) = 0$. Contrary appearances notwithstanding, Newton's iteration $x_{n+1} = x_n - f(x_n)/f'(x_n)$ is *not* just a special case of the previously studied iteration $x_{n+1} = \phi(x_n)$.

> Proposition: Newton's iteration is ubiquitous; if $\phi(x)$ maps the finite closed interval I continuously into itself, and if ϕ has just one fixed point $\zeta = \phi(\zeta)$ in I, then $\phi(x) = x - f(x)/f'(x)$ for some function f which is continuous in I, vanishes only at ζ in I, and is continuously differentiable in I except possibly at ζ.

In fact $f(x) = c \exp \int^x dw/(w - \phi(w))$ where the constants $c \neq 0$ and the lower limit of integration are assigned different values for x on one side of ζ than on the other. The hypotheses concerning ϕ ensure that $x - \phi(x)$ has always the same sign as $x - \zeta$, so the integral is properly defined for all $x \neq \zeta$ in I and $f(x) \to 0$ as $x \to \zeta$. If f has non-zero one-sided derivatives at $x = \zeta$ the constants may be altered if necessary to make f continuously differentiable at $x = \zeta$ too. □

More generally, the fixed points of $\phi(x) = x - f(x)/f'(x)$ that are not zeros of f turn out to be places where $f'(x) = \infty$. Rather than digress into generalities, let us consider the following useful application of the No Swap theorem to Newton's iteration.

Suppose $f(x)$ is a rational function whose poles and zeros are all real, simple, and interlace, with one pole at $x = \infty$. Such a function has the form $f(x) = c\big(x - \beta - \sum_1^n \omega_i/(x-\pi_i)\big)$ with $c \neq 0$ and all $\omega_i > 0$ and $\pi_1 < \pi_2 < \cdots < \pi_n$; or it may have the form

$$f(x) = \det(x\mathbf{I}-A)/\det(x\mathbf{I}-\tilde{A})$$

where A is an hermitian matrix, $\tilde{A}$ is obtained by striking off A's last row and column, and the $\mathbf{I}$'s are identity matrices. Such functions f play important rôles during, for example, the calculation of eigenvalues of hermitian matrices A; cf. Y. Saad (1974). Now we shall see why Newton's iteration almost always converges to a zero of f; we shall find that convergence to a zero can be precluded only if x_0 is one of a countable sparse set of starting points from which the iteration $x_{n+1} = x_n - f(x_n)/f'(x_n)$ terminates at a finite pole π_i of f after finitely many steps.

Proof: Write $f(x) \equiv c \prod_1^{n+1} (x-\zeta_j)/\prod_1^{n}(x-\pi_j)$ where $\zeta_1 < \pi_1 < \zeta_2 < \pi_2 < \cdots < \zeta_n < \pi_n < \zeta_{n+1}$ displays the interlacing poles π_j and zeros ζ_i of f. Useful equivalent forms for f are

$$f(x) = c\big(x - \beta - \sum_1^n \omega_i/(x-\pi_i)\big) = c/\sum_1^{n+1} v_i/(x-\zeta_i)$$

where $\beta = \sum_1^n \pi_j - \sum_1^{n+1} \zeta_j$, $\omega_j = \prod_{i=1}^{j=1} \frac{\pi_j-\zeta_i}{\pi_j-\pi_i}(\pi_j-\zeta_j)(\zeta_{j+1}-\pi_j) \prod_{j+1}^{n} \frac{\zeta_{i+1}-\pi_j}{\pi_i-\pi_j} > 0$,

$v_j = \prod_{i=1}^{j-1} \frac{\zeta_j-\pi_i}{\zeta_j-\zeta_i} \prod_{j+1}^{n+1} \frac{\pi_{i-1}-\zeta_j}{\zeta_i-\zeta_j} > 0$ and $\sum_1^{n+1} v_j = 1$. Corresponding forms for the iterating function $\phi(x) \equiv x - f(x)/f'(x)$ are

$$\phi(x) = \left(\beta + \sum_1^n \frac{\omega_i(2x-\pi_i)}{(x-\pi_i)^2}\right)/\left(1 + \sum_1^n \frac{\omega_i}{(x-\pi_i)^2}\right) = \left(\sum_1^{n+1} \frac{v_j\zeta_j}{(x-\zeta_j)^2}\right)/\left(\sum_1^{n+1} \frac{v_j}{(x-\zeta_j)^2}\right)$$

from which follow immediately the conclusions that the rational function $\phi(x)$ is continuous for all real x, has fixed points $\zeta_j = \phi(\zeta_j)$ and $\pi_j = \phi(\pi_j)$, and maps the whole real axis onto the interval $\zeta_1 \le \phi(x) \le \zeta_{n+1}$. Consequently Newton's iteration $x_{n+1} = \phi(x_n)$ will converge from every real x_0 if and only if ϕ swaps no two distinct values x and y. But if ϕ did swap them we could rearrange the equations $x = \phi(y)$ and $y = \phi(x)$, subtract, divide out $(x-y)$, and infer that $\sum_1^{n+1} v_j\left((x-\zeta_j)^{-2} + (x-\zeta_j)^{-1}(y-\zeta_j)^{-1} + (y-\zeta_j)^{-2}\right) = 0$ when in fact it must be positive. Therefore the iteration must converge to one of ϕ's fixed points. Since each zero ζ_j is a strongly attractive fixed point ($\phi'(\zeta_j) = 0$) but each pole π_j is strongly repulsive ($\phi'(\pi_j) = 2$), convergence to a pole can occur only "by accident" after finitely many iterations, and even then rounding errors are likely to intervene in our favour and deflect the iteration to converge to an attractive fixed point, a zero of f. □

PART II

§5. Newton's and the Secant Iterations

I is again a closed finite interval in which we now seek a zero ζ of a real function $f(x)$ that is continuous in I and continuously once differentiable too except possibly at its zero ζ. The search for ζ begins at one or two starting approximations x_0 and x_1 in I and attempts to improve them via one of the following iterations;

Newton's $\quad x_{n+1} = N(x_n) \quad$ for $n = 0,1,2,3,\ldots$ or

Secant $\quad x_{n+1} = S(x_n, x_{n-1}) \quad$ for $n = 1,2,3,4,\ldots$

where, as illustrated in Figs. 6 and 7,

$$N(x) \equiv x - f(x)/f'(x) \quad \text{if} \quad f(x) \neq 0 \ ,$$
$$\equiv x \quad \text{if} \quad f(x) = 0 \quad \text{no matter what happens to} \quad f'(x) \ ,$$

$$S(x,y) \equiv x - f(x)(x-y)/(f(x)-f(y)) \equiv S(y,x) \quad \text{if} \quad y \neq x \quad \text{and} \quad f(x) \neq 0 \ ,$$
$$\equiv N(x) \quad \text{if} \quad y = x \quad \text{or} \quad f(x) = 0 \ .$$

Whether either iteration converges, and which iteration is the better, are important questions without simple answers; but the following theorem sheds some light upon them.

> Theorem: If $N(x)$ is continuous in I, and if Newton's iteration converges in I from every starting point x_0 in I, and provided f not merely vanishes but actually reverses sign across its zero ζ in I, then the Secant iteration also converges in I from every pair of starting points x_0 and x_1 in I.

Before embarking upon the theorem's proof we shall digress first to discuss the almost superfluous continuity requirement upon N, second to expose

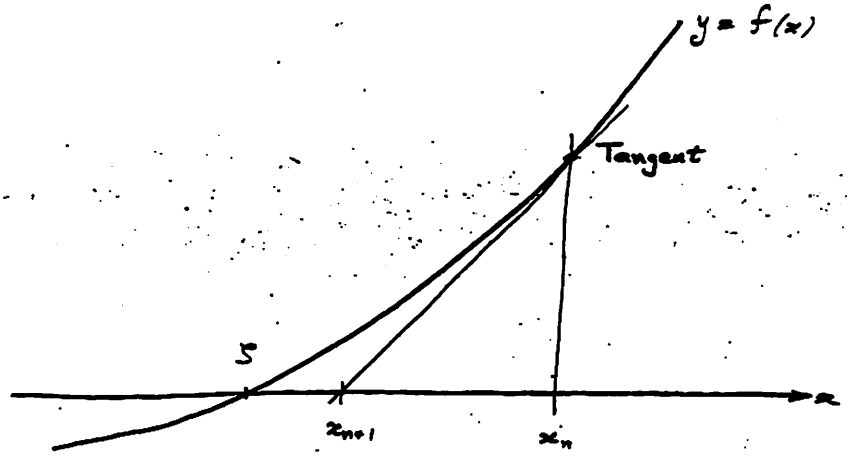


Fig. 6: Newton's iteration

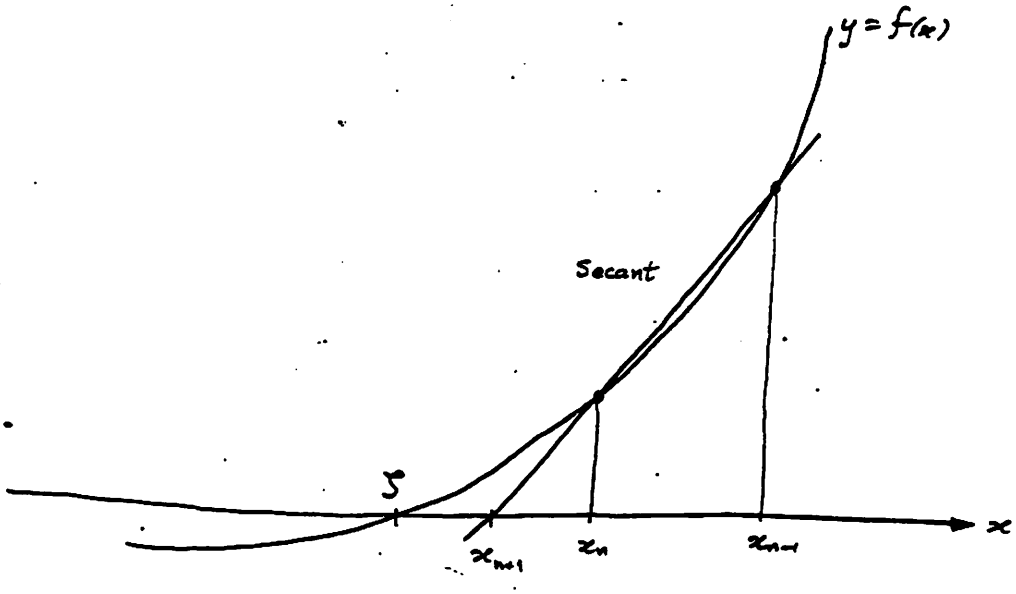


Fig. 7: Secant iteration.

some of the Secant iteration's history, and third to explore some examples.

When N is continuous in I then, for reasons exposed in §7 and §9, the theorem's hypotheses imply that S is continuous too in $I \times I$. This is the normal state of affairs and arises, for example, when f is twice differentiable in I and f'' reverses sign therein only finitely often (see §8's corollary's proof). However, even if f is infinitely differentiable in I, N can be discontinuous in I; an example is given in §7. N can be discontinuous only at ζ, and then only if f' takes values arbitrarily close to zero in the neighborhood of ζ. Despite this discontinuity of N, if it occurs, the theorem above perseveres nearly unchanged as follows.

> Theorem: If Newton's iteration generates from every x_0 in I a sequence of iterates $\{x_n\}$ of which some subsequence converges to ζ, and provided f not merely vanishes but actually reverses sign across its zero ζ in I, then the Secant iteration also generates from every x_0 and x_1 in I a sequence of iterates of which some subsequence converges to ζ.

This statement of the Theorem includes the previous version above for reasons exposed in §8 where conditions necessary and sufficient for convergence of all Newton iterates or a subsequence of them are exhibited.

The possible discontinuity of N complicates proofs but can have no practical consequences if, as is customary in well-designed computer programs, two criteria are used to decide when to terminate an iteration designed to calculate a zero ζ of f. Either stop when $f(x_n)$ is negligible, and then accept x_n as the approximation to ζ; or stop when several successive iterates $\ldots, x_{n-2}, x_{n-1}, x_n$ differ from each other negligibly and the values $\ldots, f(x_{n-2}), f(x_{n-1}), f(x_n)$ are not all of the same sign, and then accept x_n.

These criteria are equally applicable when only a subsequence of the iterates converges to ζ. Hence, for practical purposes the theorem says roughly that whenever Newton's iteration must succeed in finding a zero ζ at which f reverses sign, so must the Secant iteration succeed.

Compared with Newton's iteration, the Secant iteration has a sparse literature. For a long time the Secant iteration $x_{n+1} = S(x_n, x_{n-1})$ was not distinguished from the REGULA FALSI $x_{n+1} = S(x_n, x_0)$, a far slower procedure. Consequently numerical analysis texts used to give it short shrift, favouring Newton's iteration instead; for instance take the volte-face between the first and second editions of "Modern Computing Methods" (1957 and 1961). The Secant iteration was first used on the earliest electronic computers because their users calculated that a simpler (no need to compute a derivative) but possibly slow method executed on an electronic computer will usually yield a correct answer sooner than a complicated but faster (fewer iterations) method executed by hand. The first person to realize that both Newton's and the Secant iterations run at comparable speeds when both are executed on the same computer appears to have been David Wheeler who (according to Wilkes, 1966) modified the Secant iteration cleverly to serve as a fast general-purpose zero-finder on one of the first electronic computers, EDSAC I at Cambridge; see program F2 in Wilkes, Wheeler and Gill (1951). We shall not digress into Wheeler's method beyond listing its order of convergence $3^{1/3} = 1.442...$ published by Wilkinson (1967) and by Dowell and Jarratt (1971 — what they call "the Illinois Algorithm" is the ILLIAC I program transcribed from Wheeler's after he visited the University of Illinois in the early 1950s). See also Dahlquist, Björck and Anderson (1974, pp. 231-3). Wheeler's program is still widely used, for example as program STD14B distributed with the Hewlett-Packard shirt-pocket calculator HP-65 (1974). Better programs, faster, more reliable and, alas, more complicated, have been

devised recently by Brent (1973) and by Bus and Dekker (1974).

Most of the Secant iteration's literature dwells upon its local convergence properties. For instance, any iteration has an order of convergence defined as

$$\liminf_{x_m \to \zeta} (-\ln|x_m - \zeta|)^{1/m} \geq 1 \ ;$$

the greater its order the faster its convergence. In the usual case when f is a smooth non-linear function with a simple zero ζ, i.e. $f(\zeta) = 0 \neq f'(\zeta)$ and $f''(\zeta) \neq 0$, Regula Falsi has order 1, Newton's iteration 2, and the Secant iteration $(1+\sqrt{5})/2 = 1.618...$. This last number, first derived by Bachmann (1954), does not imply that the Secant iteration is slower than Newton's; on the contrary, as pointed out by Ostrowski (1960 et seq.), by Traub (1964), and (briefly) by Dahlquist, Björck and Anderson (1974), the Secant iteration is usually the faster unless the time consumed computing $f'(x)$ adds less than half to $f(x)$'s computation.

Conditions sufficient to ensure the Secant iteration's convergence are found in survey texts like Ostrowski's, Traub's or Ortega and Rheinboldt's, and summarized in Householder (1970) or Dahlquist, Björck and Anderson (1974). Characteristic of all such sufficient conditions in the literature known to this writer is that they also suffice to ensure Newton's iteration's convergence. This characteristic might suggest that Newton's iteration converges whenever the Secant iteration does, but the facts are contrariwise as stated in the Theorem above and illustrated by our first example A below.

The following five examples A to E all have only $f(0) = 0$, so $\zeta = 0$, and either $f(-x) \equiv -f(x)$ or else $f(-x) \equiv f(x)$. Moreover, f is everywhere at least once continuously differentiable.

Example A: $f(x) \equiv 23x - 10x^3 + 3x^5$.

Despite that $f(x)$ is strictly monotone increasing for all x, and despite that the Secant iteration converges to ζ from every real x_0 and x_1, Newton's iteration will converge from x_0 if $|x_0| < \sqrt{23/27} = .922958207...$ but alternates with $x_n = (-1)^n x_0$ if $x_0 = \pm\sqrt{23/27}$. Worse, if $\sqrt{23/27} < x_0 < 1.06977829...$ then $(-1)^n x_n \to 1$.

Example B: $f(x) \equiv x\big(4+\sqrt{5}+|x|\big)/\big(1+(4+\sqrt{5})|x|\big)$.

Despite that $f(x)$ is strictly monotone increasing for all x, the Secant iteration fails to converge but cycles instead on four points, $x_{n+2m} = (-1)^m x_n$, from $x_0 = 2+\sqrt{5} = 4.23606797...$, $x_1 = 1$, $x_2 = -x_0$, $x_3 = -x_1$, ... ; otherwise I conjecture that the Secant iteration converges from almost all starting points. Newton's iteration fares worse; it converges provided $|x_0| < \check{x} \equiv \big(16+7\sqrt{5}-2\sqrt{95+56\sqrt{5}}\big)/11 = .179351475...$, but otherwise diverges to a cycle on two points, with $(-1)^n x_n \to \pm\big(16+7\sqrt{5}+2\sqrt{95+56\sqrt{5}}\big)/11$ $= 5.57564413...$ unless $(-1)^n x_n = \pm\check{x}$.

The next three examples illustrate what can happen when f vanishes at ζ but does not reverse sign, thereby vindicating the proviso in the Theorem above. The change of sign can be essential for the Secant iteration's convergence though irrelevant for Newton's because the latter is unchanged when $f(x)$ is replaced by $|f(x)|$.

Example C: $f(x) \equiv x^{m+1}$ for integer $m \geq 1$.

This example's analysis is facilitated by the observation that first

$$x_{n+1} = N(x_n) \text{ is tantamount to } x_{n+1}/x_n = m/(m+1)$$

and secondly

$$x_{n+1} = S(x_n, x_{n-1}) \text{ is tantamount to } x_{n+1}/x_n = \phi(x_n/x_{n-1})$$

where $\phi(y) \equiv (y^m-1)/(y^{m+1}-1)$ has been discussed at the end of §1 above.

Newton's iteration converges to $\zeta = 0$ from every x_0. So does the Secant iteration when m is even, but when m is odd (then f does not reverse sign) the iteration converges from almost all x_0 and x_1. The exceptions are first when x_1/x_0 coincides with the negative fixed point of ϕ, in which case $(-1)^n x_n$ diverges monotonically to infinity, and secondly when x_1/x_0 coincides with one of a countable set of values from which will follow an $x_n = \infty$, $x_{n+1} = -x_{n-1}$, $x_{n+2} = mx_{n-1}/(m+1)$ and subsequently $x_{n+j} \to 0$ as $j \to \infty$.

Example D: $f(x) \equiv x^2\left(7 - 2\sqrt{5} + |x|\right)/\left(1 + (7 - 2\sqrt{5})x^2\right)$.

Newton's iteration converges from every x_0. The Secant iteration converges if $x_1x_0 \geq 0$ and $|x_1| < .365966339...$ but otherwise is likely to tend to cycle. One cycle on four points is

$$x_0 = 2 + \sqrt{5} = 4.236067977..., \; x_1 = -1, \; x_2 = -x_0, \; x_3 = -x_1, \; x_{n+2m} = (-1)^m x_n \; .$$

Another is

$$x_0 = 7.7019690..., \; x_1 = -1.81818824..., \; x_2 = -x_0, \; x_3 = -x_1, \; x_{n+2m} = (-1)^m x_n \; ,$$

and this cycle is stable and attractive with a contraction factor per cycle near .278 .

Example E: $f(x) \equiv 1.0 - 3/(3+x^2)$.

Newton's iteration converges to $\zeta = 0$ from any x_0 with $x_0^2 < 9$, oscillates with $x_n = (-1)^n x_0$ if $x_0 = \pm 3$, and diverges alternatingly to $\pm\infty$ if $x_0^2 > 9$. The Secant iteration converges if $x_0x_1 \geq 0$ and both $x_0^2 < 3$ and $x_1^2 < 3$ but frequently diverges otherwise, and certainly diverges when $x_0^2 \geq 9$ and $x_1^2 \geq 9$; it can cycle on four points $x_0 = \sqrt{15 + 6\sqrt{5}} = 5.33070425...$, $x_1 = -\sqrt{15 - 6\sqrt{5}} = -1.25840857...$, $x_2 = -x_0$, $x_3 = -x_1$, $x_{n+2m} = (-1)^m x_n$.

Changing the constant 1.0 in the last example to, say, 1.000···0001 illustrates how difficult in practice is the problem of determining where a function $f(x)$ vanishes when it does not reverse sign. The difficulty is not caused entirely by roundoff. For example, even though the values of

$$f(x) \equiv (x - (5 - (x - (5-x))))^2 \quad \text{(Do not remove the parentheses!)}$$

and its derivative must be calculated, on any North American or Western European elctronic computer, precisely (i.e. with no rounding errors) for every x close enough to $3\frac{1}{3}$, none of those values of $f(x)$ vanishes because the value $x = 3\frac{1}{3} = 3.333333...$ is never represented precisely in floating point. Consequently the Theorem's proviso might as well be taken for granted in practice.

§6. Projective Invariance of Newton's and the Secant Iteration

Projective transformations are those which transform straight lines into straight lines. They are pertinent to Newton's and the Secant iterations because tangents are transformed into tangents, secants into secants. The particular projective transformations useful here map the pair $\{x,f(x)\}$ onto a pair $\{\xi,\phi(\xi)\}$ in such a way that either both or neither of $f(x)$ and $\phi(\xi)$ are linear functions of their respective arguments, and yet the mapping is independent of f and ϕ. Well-known results from Projective Geometry lead to the following formulas.

Let $\xi = \rho(x) \equiv (\alpha x+\beta)/(\gamma x+\delta)$ with $\alpha\delta - \beta\gamma = 1$. Hence $\rho(x)$ is invertible; $x = \rho^{-1}(\xi) = (\beta-\delta\xi)/(\gamma\xi-\alpha)$ or, more symmetrically, $(\gamma x+\delta)(\alpha-\gamma\xi) = 1 = (\alpha+\beta/x)(\delta - \beta/\xi)$. Having transformed the variable x into ξ we further construct

$$\phi(\xi) \equiv f(x)/(\gamma x+\delta) \quad @ \quad x = \rho^{-1}(\xi)$$
$$= (\alpha-\gamma\xi)f((\beta-\delta\xi)/(\gamma\xi-\alpha))$$

as the corresponding transform of f. Now calculation suffices to verify that the transforms of $N(x)$ and $S(x,y)$ are respectively

$$H(\xi) \equiv \xi - \phi(\xi)/\phi'(\xi) = \rho(N(\rho^{-1}(\xi))) \quad \text{and}$$
$$\Sigma(\xi,\eta) \equiv \xi - \phi(\xi)(\xi-\eta)/(\phi(\xi)-\phi(\eta)) = \rho(S(\rho^{-1}(\xi),\rho^{-1}(\eta))) \ ,$$

whence Newton's iteration applied to ϕ takes the form $\xi_{n+1} = H(\xi_n)$, the Secant iteration is $\xi_{n+1} = \Sigma(\xi_n,\xi_{n-1})$. The projective invariance implied by these equations can be stated in words as follows.

Projective Invariance: Let $\{x_n\}$ be the sequence of iterates generated when either Newton's or the Secant iteration is applied to $f(x)$. Then the projective transformation

$$\xi = \rho(x) = (\alpha x+\beta)/(\gamma x+\delta) \quad \text{and} \quad \phi(\xi) = (\alpha-\gamma\xi)f(\rho^{-1}(\xi)) \quad \text{with} \quad \alpha\delta-\beta\gamma = 1$$

maps the iterates $\{x_n\}$ upon the sequence $\{\xi_n = \rho(x_n)\}$ generated respectively by either Newton's iteration from $\xi_0 = \rho(x_0)$ or the Secant iteration from $\xi_0 = \rho(x_0)$ and $\xi_1 = \rho(x_1)$ applied to $\phi(\xi)$.

One application of projective invariance is that whatever convergence theory pertains to the iterations in finite intervals I may be extended with few changes to semi-infinite intervals $\rho(I)$ by virtue of an appropriate choice for ρ. One of those few changes is a nuisance illustrated by the example $f(x) \equiv 1+\exp(-x)$ on the semi-infinite interval $0 \le x \le +\infty$; both iterations converge to $+\infty$ but $f(+\infty) \neq 0$. Consequently semi-infinite intervals will not be mentioned again in this paper.

A second implication is that natural hypotheses implying the iterations' global convergence should also be projectively invariant. For instance, since the second derivative

$$\phi''(\xi) = (\gamma x+\delta)^3 f''(x) \quad @ \quad x = \rho^{-1}(\xi) \ ,$$

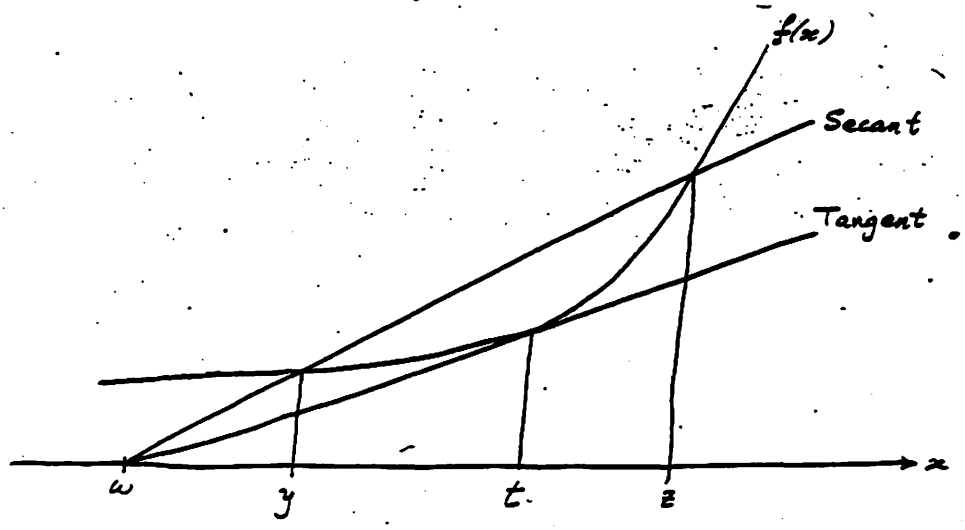
any hypotheses about the number of zeros f'' has in I, or about $\text{sign}(ff'') = \text{sign}(\phi\phi'')$, are invariant provided $\rho(I)$ remains an interval. Just such hypotheses abound in the literature (cf. Ostrowski (1960 et seq., ch. 9 and 10) or Dahlquist, Björck and Anderson (1974, p. 225)) and are tantamount to the observation that the convexity of f's graph is a projective invariant implying ultimately steady convergence of both iterations provided they do not first escape from I. More about this in §8.

A third implication of projective invariance, the one most pertinent to our proof, is a kind of mean value theorem which relates the two iterating functions $N(x)$ and $S(x,y)$ in a way more general than $S(x,x) = N(x)$.

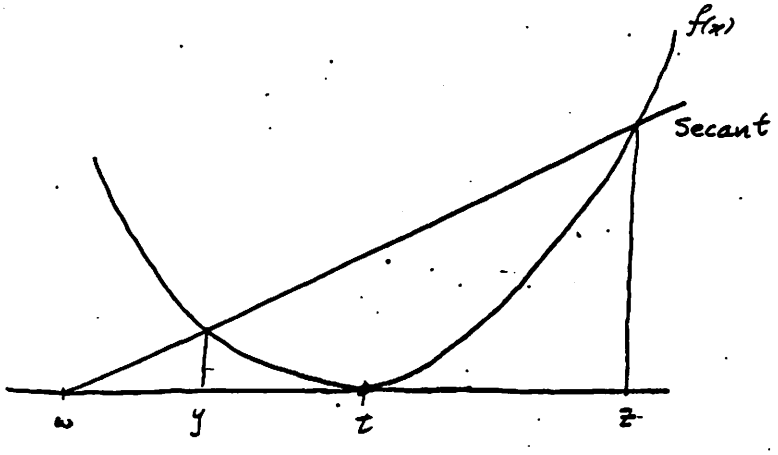
Mean Value Lemma: If $S(y,z)$ does not lie between y and z, (i.e. if $f(y)f(z) > 0$), if $y \neq z$, and if f is differentiable between y and z, then strictly between y and z must lie some t for which either $S(y,z) = N(t)$ or $f(t) = f'(t) = 0$.

See Fig. 8.

Proof: A projective transformation could be applied to push $w = S(y,z)$ to ∞ while preserving the interval between y and z, but the resulting calculations would boil down to what follows. Let $\psi(x) \equiv f(x)/(x-w)$, and observe that the equation $w = S(y,z)$ is equivalent to $\psi(y) = \psi(z) = (f(y)-f(z))/(y-z)$. Since $\psi(x)$, like $f(x)$, is differentiable between y and z, Rolle's theorem (cf. Apostol (1967)) provides that $\psi'(t) = 0$ at some t strictly between y and z. That t turns out to satisfy either $f(t) = f'(t) = 0$ or $w = N(t)$. □



$\omega = S(y,z) = N(t)$.



$\omega = S(y,z)$, $f(t) = f'(t) = 0$.

Fig. 8: The Mean Value Lemma.

In other words, just as $N(x) = S(x,x)$ implies that Newton's iteration cannot escape from an interval from which the Secant iteration cannot escape, the Mean Value lemma implies that the Secant iteration cannot escape from an interval from which Newton's iteration cannot escape unless f vanishes in that interval without changing sign. This implication will become clearer later.

§7. Inferences from $N(I) \subseteq I$

Henceforth we activate two of the hypotheses of the Theorem of §5. First, I is a closed finite interval in which $f(x)$ is continuous, and continuously once differentiable too except possibly where f vanishes. Second, $N(x)$ maps I into itself. We do not yet assume that $x_{n+1} = N(x_n)$ converges. What do these hypotheses tell us about f and N?

Lemma: f has just one zero ζ in I, and ζ divides I into at most two sub-intervals inside each of which $f(x)$ is strongly monotonic (i.e. $f'(x)$ cannot vanish inside either sub-interval).

Proof: At the outset we beg the reader to put up with an abuse of language; in the unlikely event that f vanishes throughout a non-degenerate sub-interval of I we shall count that sub-interval as a single zero ζ of f in I and write "$x = \zeta$" when we mean "x belongs to the interval ζ". This perversion avoids circumlocution in dealings with functions that are not analytic but merely differentiable.

The essential hypothesis is that $N(I) \subseteq I$. Now f cannot have two or more distinct zeros in I because otherwise Rolle's theorem would supply between two adjacent distinct zeros of f at least one x_0 where $f'(x_0) = 0 \neq f(x_0)$ whence $N(x_0) = \infty$ would escape from I. Neither can f fail to vanish in I since otherwise f would take non-zero values

with like signs at I's ends, whereupon the Secant iteration started from I's ends would escape from I to a place whither Newton's iteration, according to §6's mean value lemma, could escape too. Therefore f does have just one zero ζ in I. If f were not strongly monotonic strictly between ζ and either end of I, I would contain some x_0 where $f'(x_0) = 0 \neq f(x_0)$ so again $N(x_0) = \infty$ would escape from I. □

Corollary: No $x \neq \zeta$ in I can separate ζ from $N(x)$; i.e. $(N(x) - x)/(\zeta - x) > 0$ for every x in I except $x = \zeta$. And if Newton's iteration $x_{n+1} = N(x_n)$ converges from some x_0 in I it must converge to ζ, the zero of f in I.

Proof: $(N(x) - x)/(\zeta - x) = -f(x)/\big((\zeta - x)f'(x)\big)$ is continuous and non-vanishing in I strictly between I's ends and ζ, and therefore conserves the positive sign it enjoys at I's end(s) different from ζ. And if $x_\infty \equiv \lim_{n\to\infty} x_n$ exists and if N is continuous at x_∞ (else $x_\infty = \zeta$) then $f(x_\infty) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty}(x_n - x_{n+1})f'(x_n) = 0 \cdot f'(x_\infty) = 0$ so $x_\infty = \zeta$. □

The corollary foreshadows some technical arguments, the burden of the rest of this section, §7, concerning the continuity of $N(x)$ at $x = \zeta$ where $f'(\zeta)$ has not been assumed to exist. Elsewhere $N(x) = x - f(x)/f'(x)$ is, like $f(x)$ and $f'(x)$, continuous. But at $x = \zeta$ we shall infer scarcely more than that $N(x)$ is Darboux continuous, which means that N assumes in every sub-interval around ζ all values between those N takes at that sub-interval's ends. Darboux continuity is an intermediate-value property possessed not only by continuous functions but also, for instance, by derivatives even when they are discontinuous. For more details see Bruckner and Ceder (1965).

The trouble with N is not necessarily caused by our failure to assume that f' exists and is continuous at ζ. For example consider

$$f(x) \equiv (1+x)\exp(\sin(1+1/x) - 1/x) \quad \text{for } 0 < x \leq 1 ,$$
$$\equiv -\exp(1/x) \quad \text{for } -1 \leq x < 0 ,$$
$$\equiv 0 \text{ at } x = 0 .$$

This $f(x)$ is infinitely differentiable in the interval $I \equiv \{-1 \leq x \leq 1\}$, vanishes only at $x = 0$ and is elsewhere in I strongly monotonic increasing. However

$$N(x) = x\big(1 - (x+1)\cos(1+1/x)\big)/\big(x^2 + x + 1 - (1+x)\cos(1+1/x)\big) \quad \text{for } 0 < x \leq 1 ,$$
$$= x(1+x) \quad \text{for } -1 \leq x < 0 ,$$
$$= 0 \text{ at } x = 0 ,$$

behaves discontinuously as $x \to 0+$. None the less we may infer, after verifying first that $-1 \leq N(x) < x$ for $0 < x \leq 1$ and second that $x < N(x) \leq 0$ for $-1 \leq x < 0$, that Newton's iteration converges to 0 (slowly!) from every x_0 in I.

Proposition: The functions $N(x)$, $N(x) - x$, $N(N(x))$ and $N(N(x)) - x$ are Darboux continuous everywhere in I including at ζ, and the first two are continuous everywhere in I except possibly at ζ; this means that each of these functions assumes in every sub-interval of I all values between the ones taken at that sub-interval's ends.

Proof: The proposition will be proved for $N(x)$ first, for which we need only be concerned with sub-intervals of I that contain ζ. And if such a sub-interval is divided at ζ into two parts for each of which the proposition is verified separately, the proposition will be verified for the whole sub-interval.

For definiteness choose any $\eta < \zeta$ in I and let us verify the proposition for the closure of the unclosed interval $J \equiv \{\eta \le x < \zeta\}$. Since $N(x)$ is continuous in J its image $N(J)$ is also an interval, and to verify the proposition we need only show that its closure contains $\zeta = N(\zeta)$. Now there are only two cases to rule out, namely $\zeta <$ closure$(N(J))$ and $\zeta >$ closure$(N(J))$.

If for some $\varepsilon > 0$ we found $\zeta < \zeta + 2\varepsilon < N(x)$ throughout J, i.e. for $\eta \le x < \zeta$, we should have to find that

$$\psi(x) \equiv 2f(x)/f(\eta) - (\zeta+\varepsilon-x)/(\zeta+\varepsilon-\eta)$$

takes values of opposite sign at J's ends;

$$\psi(\eta) = 1 > 0 > -\varepsilon/(\zeta+\varepsilon-\eta) = \psi(\zeta) \ ,$$

so ψ must first change sign somewhere in J, say at ξ where $\psi(\xi) = 0 \ge \psi'(\xi)$. But this would imply $2f(\xi)/f(\eta) = (\zeta+\varepsilon-\xi)/(\zeta+\varepsilon-\eta) > 0$ and $-2f'(\xi)/f(\eta) \ge 1/(\zeta+\varepsilon-\eta) > 0$ whence $N(\xi) = \xi + \big(f(\xi)/f(\eta)\big)/\big(-f'(\xi)/f(\eta)\big)$ $\le \xi + (\zeta+\varepsilon-\xi) = \zeta + \varepsilon < \zeta + 2\varepsilon$! Therefore $\zeta \notin$ closure$(N(J))$.

On the other hand, if for some $\varepsilon > 0$ we found $N(x) < \zeta - 2\varepsilon < \zeta$ for all $\eta \le x < \zeta$ we should have to violate the foregoing corollary's inequality $0 < (N(x) - x)/(\zeta - x)$ at $x = \max\{\zeta-\varepsilon, \eta\}$. Therefore $\zeta \notin$ closure$(N(J))$.

Thus we conclude that $N(x)$ is Darboux continuous. Consequently $N(N(x))$ is Darboux continuous too. Moreover, we shall now find that $N(x)$, and then $N(N(x))$, belong to the first Baire class of functions, the pointwise limits of continuous functions. This is evident because we can approximate $N(x)$ by a continuous function differing from N only in an open deleted neighborhood of ζ, though both functions match at ζ and at the

boundaries of the neighborhood, and then let the neighborhood shrink down onto ζ. The first Baire class is significant because the sum of a Darboux continuous function in that class with a continuous function is known to be another Darboux continuous function in that class; see Theorem 7.5 of Bruckner and Ceder (1965, p. 109). Therefore $N(x) - x$ and $N(N(x)) - x$ are Darboux continuous and the proposition is proved. □

This proposition is crucial; it allows Part I's No Swap Theorem to be generalized enough to cover $N(x)$. One contrary implication of the No Swap condition that can be obtained almost immediately is the following, which shows that the last example above is typical of the kind of discontinuity that can befall N without precluding convergence. As is customary, we write $x \to \zeta-$ to mean that x increases to the limit ζ, and $x \to \zeta+$ when x decreases to ζ. Can N be discontinuous both as $x \to \zeta+$ and as $x \to \zeta-$?

Aside: If $N(x)$ swaps no two distinct points in I, $N(x)$ cannot be discontinuous on both sides of ζ but at most one; and as $x \to \zeta$ from the side opposite the discontinuity $N(x)$ must ultimately lie between ζ and x.

Specifically, if $N(x) \not\to \zeta$ as $x \to \zeta-$ set $\eta \equiv \limsup_{x \to \zeta-} N(x) > \zeta$; then $\zeta \leq N(x) < x$ whenever $\zeta < x \leq \eta$.

Proof: Suppose on the contrary that $N(v) < \zeta$ and $\zeta < v < \eta$. Since N is Darboux continuous at $\zeta = N(\zeta)$, N must, for any $y < \zeta$ in I, assume all values between ζ and η as x runs from y up to ζ; therefore $N(u) = v$ for at least one u in $y < u < \zeta$. Therefore $N(N(u)) - u = N(v) - u < N(v) - y < 0$ provided y be first selected in $N(v) < y < \zeta$; on the other hand $N(N(x)) - x > 0$ for x close enough to I's left-hand

end since $N(I) \subseteq I$. Consequently, because $N(N(x)) - x$ is Darboux continuous too, $N(N(x_0)) - x_0 = 0$ for at least one $x_0 < u < \zeta$ in I, and then $x_1 = N(x_0) > x_0$ because of the Corollary above. So N would swap the distinct points x_0 and x_1 contrary to hypothesis. □

§8. The No Swap Theorem for Newton's Iteration

Theorem: Suppose I is a closed finite interval in which $f(x)$ is continuous, and continuously once differentiable too except possibly where f vanishes. Let $N(x) \equiv x - f(x)/f'(x)$ except $N(x) \equiv x$ when $f(x) = 0$. Then Newton's iteration $x_{n+1} = N(x_n)$ generates from every x_0 in I a sequence $\{x_n\}$ of which some subsequence converges to a zero ζ of f, i.e. a subsequence of $\{f(x_n)\}$ converges to zero, if and only if $N(I) \subseteq I$ (and therefore ζ is unique) and N satisfies any of the following conditions (they are equivalent).

The No Swap Condition: If $x = N(N(x))$ in I then $x = N(x)$ $(= \zeta)$.

The No Separation Condition: Either $N(N(z)) \leq z \leq N(z)$ or $N(z) \leq z \leq N(N(z))$ in I implies $N(N(z)) = z = N(z)$ $(= \zeta)$.

The No Crossover Condition: If $N(v) \leq u \leq v \leq N(u)$ in I then $N(v) = u = v = N(u)$ $(= \zeta)$.

The One-Sided Condition: Whenever $x_1 = N(x_0) \neq x_0$ in I all subsequent iterates $x_{n+1} = N(x_n)$ also differ from x_0 and lie on the same side of x_0 as does x_1 (and ζ).

If also N is continuous, or if also either

$$N(\limsup_{x \to \zeta-} N(x)) \neq \zeta \quad \text{or} \quad N(\liminf_{x \to \zeta+} N(x)) \neq \zeta$$

then any of the foregoing conditions implies that $x_{n+1} = N(x_n) \to \zeta$ from every x_0 in I.

Proof: Only when N is not continuous need the proof involve a little more work than merely writing $N(\cdots)$ in place of $\phi(\cdots)$ in part I. Having established in §7 that $N(x)$, $N(x)-x$, $N(N(x))$ and $N(N(x))-x$ are Darboux continuous, we may infer as before that certain expressions must vanish somewhere between any two points at which they change sign, so most of Part I's arguments will not need revision. But arguments that formerly depended upon $\lim \phi(x) = \phi(\lim x)$ must be revised lest $N(x) \not\to \zeta$ as $x \to \zeta$. The first such revision is needed to validate the No Separation condition; please turn back to and re-read §2 in conjunction with what follows.

To show that No Swap implies No Separation assume the latter condition violated by, say, $N(N(z)) < z < N(z)$ in $\mathcal{I}$ and seek a consequent violator v of the former. As before, N must have a fixed point y strictly between z and $N(z)$, but this time $N(\mathcal{I}) \subseteq \mathcal{I}$ implies $y = \zeta$ is the unique fixed point of N, the zero of f in $\mathcal{I}$ (cf. §7). Also as before $N(N(x))$ has a fixed point $v' < z$ in $\mathcal{I}$, but this time we don't care whether v is a "first" fixed point or not because $v < z < \zeta$ so v cannot be N's unique fixed point ζ. Therefore N swaps v and $N(v) \neq v$.

The No Crossover and One-Sided conditions' proofs survive unchanged. The next revision is needed to adapt §4's proof that One-Sidedness following from the No Swap condition implies convergence, because now that implication is invalid.

As before, the sequence of iterates $x_{n+1} = N(x_n)$ is a One-Sided sequence which is either ultimately monotonic and therefore convergent to ζ (cf. §7's Corollary) or else can be partitioned into two disjoint infinite subsequences, one ascending to a limit $\check{x}$ and the other descending to a limit $\hat{x} \geq \check{x}$. It is not possible for both limits to differ from ζ

because then N would be continuous at both limits and §4's argument would make them both equal ζ. Therefore at least one of $\hat{x}$ and $\check{x}$ equals ζ, as was claimed above.

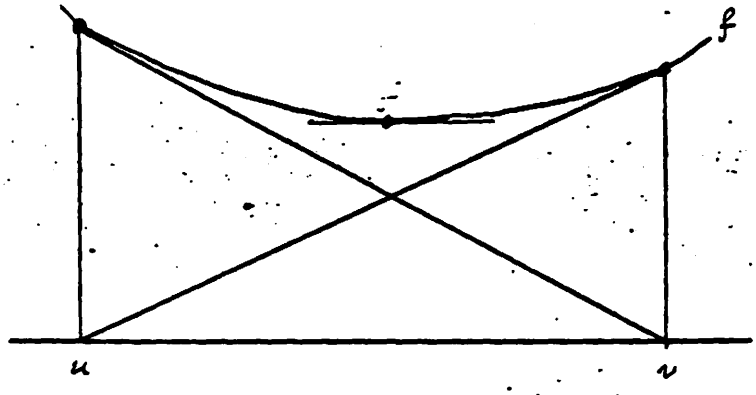
Moreover, suppose for definiteness that $\hat{x} > \check{x} = \zeta$; we conclude the proof by showing that $\eta \equiv \limsup_{x \to \zeta-} N(x) \geq \hat{x} > \zeta = N(\eta)$. Since some of those ascending $x_n \to \check{x} = \zeta$ are followed by $x_{n+1} = N(x_n) \to \hat{x}+ > \zeta$ we must find $\zeta < \hat{x} \leq \eta$. Moreover, since N is continuous at $\hat{x}$, all those descending $x_n \to \hat{x}+$ close enough to $\hat{x}$ comply with the No-Man's Land lemma (§3) by having $x_{n+1} = N(x_n) \to \zeta-$, so $N(\hat{x}) = \zeta$. But if any such x_n lay in $\hat{x} < x_n \leq \eta$ we could infer from §7's last Aside that $\zeta \leq x_{n+1} = N(x_n) < x_n$ contrary to the last sentence; therefore no descending x_n lies in $\hat{x} < x_n \leq \eta$ and hence $\hat{x} = \eta$ and $N(\eta) = \zeta$ as claimed.

Here is an example to show that the phenomenon $\check{x} < \hat{x}$ analyzed in the last paragraph is possible albeit unlikely. Let $I = \{-1 \leq x \leq 3\}$ and therein let

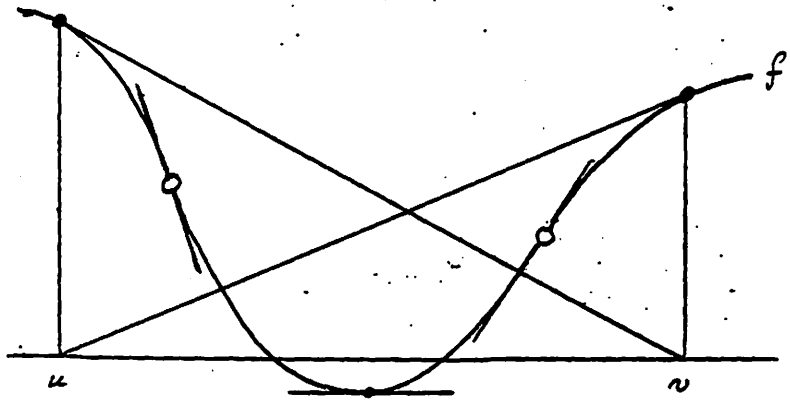
$$\begin{aligned} f(x) &\equiv (3x-2)^{2/3} && \text{for } 1 \leq x \leq 3 , \\ &\equiv x^2 && \text{for } 0 \leq x \leq 1 , \\ &\equiv -\sqrt{1-x}\,\exp(\sin(-1/x)+1/x) && \text{for } -1 \leq x < 0 . \end{aligned}$$

This $f(x)$ is once continuously differentiable (it could be modified on $0 < x < 2$ to be made infinitely differentiable and substantially more complicated), it vanishes only at $\zeta \equiv 0$ in I and is elsewhere strongly monotonic. $N(x) = x - f(x)/f'(x)$ maps I into I and swaps no two distinct points, and is continuous too except as $x \to \zeta-$. If Newton's iteration $x_{n+1} = N(x_n)$ is started at $x_0 = -1/(2\pi)$ then $x_{2n} = -1/(2^{n+1}\pi) \to \zeta-$ but $x_{2n+1} = 2 - x_{2n} \to 2+$. □

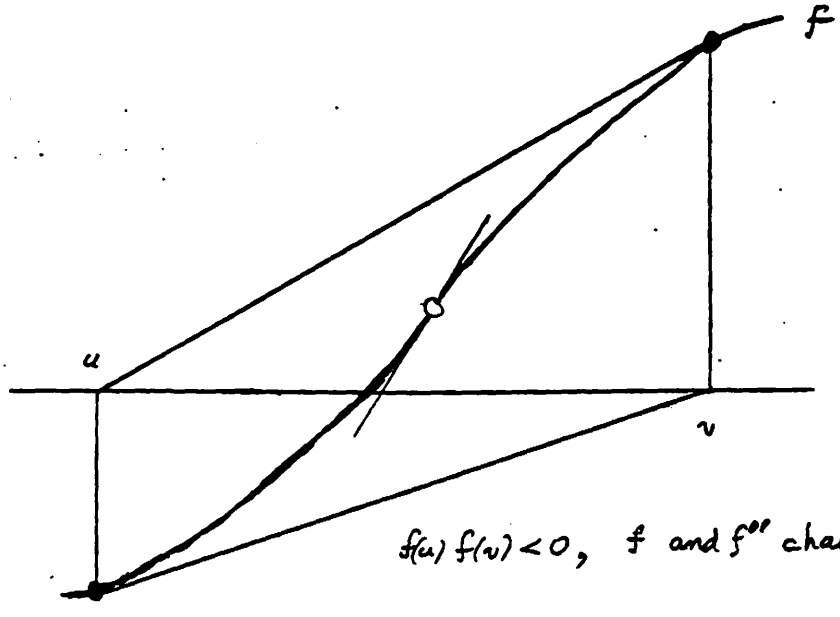
Fig. 9: Some situations which arise when N swaps u and v



$f > 0$, f' changes sign, $f'' > 0$.



$f(u)\,f(v) > 0$, f' changes sign once, f'' changes sign twice



$f(u)\,f(v) < 0$, f and f'' change sign once.

How may a function f eligible for the application of the No Swap theorem be recognized in practice? One way is to apply the contrapositive of the next lemma which exhibits some of the properties possessed by f when N violates the No Swap condition. Any f which lacks those properties satisfies that condition. Fig. 9 illustrates these properties.

Lemma: If $N(x) = x - f(x)/f'(x)$ swaps the ends u and v of an interval J in which f is twice differentiable, then one of the following situations must arise.

1. $f(u)f(v) > 0$, and then $f'(u)f'(v) < 0$ so f' must reverse sign somewhere inside J, and f'' must take the same sign as $f(u)$ somewhere in J; if also f vanishes somewhere in J then ff'' must take negative values in two open sub-intervals of J between which f'' reverses sign at least twice.
2. $f(u)f(v) < 0$ so f reverses sign at least once inside J, and then ff'' takes negative values at places inside J where f takes both positive and negative values; therefore f'' reverses sign at least once, and at least once more if f' ever vanishes in J.

Proof: This is a tedious exercize in curve-tracing whose object is to describe the ups and downs of f' in J. For definiteness assume $u < v$ and $f(v) > 0$, and re-write $u = N(v)$ and $v = N(u) > u$ in the forms

$$f'(v) = f(v)/(v-u) > 0 \quad \text{and} \quad f'(u) = f(u)/(u-v) .$$

Case 1: $f(u) > 0$. Now $f'(u) < 0 < f'(v)$ and consequently f'' must take positive values somewhere between u and v. If also at some η between u and v we find $f(\eta) = \min_{u<x<v} f(x) \leq 0$, then $f'(\eta) = 0$ and $f'' > 0$ somewhere in that neighborhood of η wherein $f'' \geq 0$; we find

further that, as x increases from u to η to v, $f'(x)$ moves from $f'(u) < 0$ through some lesser value $(f(\eta)-f(u))/(\eta-u) \leq -f(u)/(\eta-u)$ $< -f(u)/(v-u) = f'(u)$ and then moves to a higher value $f'(\eta) = 0$ and on up through $(f(v)-f(\eta))/(v-\eta) \geq f(v)/(v-\eta) > f(v)/(v-u) = f'(v)$ and back down to $f'(v)$. Hence f' has a negative minimum and positive maximum inside J, which means that f'' reverses sign at least twice in J. As x increases from u to v, $f'(x)$ has to decrease before $f(x)$ can vanish, and therefore $ff'' < 0$ in some sub-interval before f vanishes; similarly $ff'' < 0$ somewhere between the last zero of f and v.

Case 2: $f(u) < 0$. Now $f(\zeta) = 0$ at some first ζ between u and v. As x increases from u to ζ, $f'(x)$ moves from a positive value $f'(u)$ through a larger value $(f(\zeta)-f(u))/(\zeta-u) > -f(u)/(v-u) = f'(u)$, so $f''(x) > 0$ somewhere between u and ζ. Similarly $f'' < 0$ somewhere between f's last zero and v. More precisely, f' has a positive maximum inside J at which f'' reverses sign. If f' ever vanishes inside J then f' has a non-positive minimum at which again f'' reverses sign at least once more.

□

The next corollary is an advance beyond what was previously known because it deals with functions f whose graphs are not convex but may have at most one or two inflexions. As long as Newton's iteration cannot escape from I via the ends or the inflexion points of f's graph, the outcome turns out to be the same as if that graph were convex (cf. Dahlquist, Björck and Anderson, 1974, p. 125).

Corollary: Suppose w and z are the ends of a closed (possibly infinite) interval I in which f is twice differentiable, f' never vanishes except possibly if and where f vanishes, f'' reverses sign at most once except possibly again if and where both f and f' vanish

simultaneously, and at least one of $f(w)f(z)$, $f(w)f''(w)$ or $f(z)f''(z)$ is positive. Suppose too that $N(x) \equiv x - f(x)/f'(x)$ maps into I's interior both I's ends and the places if any where f'' reverses sign. f'' may vanish arbitrarily often without changing sign. Then f has in I just one zero ζ and Newton's iteration $x_{n+1} = N(x_n)$ converges to it from every x_0 in I.

Proof: An argument similar to §7's lemma shows that f must have just one zero ζ in I; more would violate a hypothesis by providing some x in I at which $f'(x) = 0 \neq f'(x)$, and fewer would either do the same or violate a different hypothesis by forcing N to map at least one of I's ends outside I.

A second argument proves N continuous in I; this is obvious from the formula for N when $f' \neq 0$, so only the possibility $f'(\zeta) = 0$ needs further explanation. When x is confined to a small neighborhood of ζ in which f'' never reverses sign except possibly at ζ, $f'(x)$ must be monotonic separately on each side of ζ and consequently $0 < f'(y)/f'(x) \leq 1$ for every y strictly between ζ and x in that small neighborhood. Since $f(x) = \int_{\zeta}^{x} f'(y)dy$,

$$N(x) = \zeta + \int_{\zeta}^{x} (1 - f'(y)/f'(x))dy \ ;$$

consequently $N(x)$ always lies between ζ and x in that small neighborhood. Therefore N is continuous at $\zeta = N(\zeta)$ and hence throughout I even when $f'(\zeta) = 0$.

The next task is to infer $N(I) \subseteq I$. Were this not so despite that N maps I's ends inside I, N would have to achieve either a maximum or a minimum value $N(\hat{x})$ outside I at some $\hat{x}$ inside I, and that $\hat{x} \neq \zeta$

because $N(\zeta) = \zeta$ lies inside I. Moreover, N' would have to reverse sign at $\hat{x}$ or at the ends of that sub-interval of I containing $\hat{x}$ on which $N = N(\hat{x})$. But $N' = ff''/(f')^2$ is prevented by hypothesis from reversing its sign in I except possibly at ζ and at most one other place also, by hypothesis, carried by N into I's interior. Therefore $N(\hat{x})$ lies inside I after all, and hence $N(I) \subseteq I$.

Now let us verify that N satisfies the No Swap condition. If not, N would swap the ends of some sub-interval J and the lemma above would imply that f'' changes sign at least once inside J, at least twice inside J if $f'(\zeta) = 0$, and at least once outside J between its end and that end of I where $ff'' > 0$. Thus f'' would vanish more often than allowed by the hypotheses. Therefore N does satisfy the No Swap condition, and Newton's iteration does converge to ζ. □

Application 1: Suppose $g(x)$ and $h(x)$ are thrice differentiable for all $x \geq 0$ and $g(0) = h(0) = 0$ but $g' > 0$, $g'' \geq 0$, $g''' \geq 0$, $h' > 0$, $h'' \geq 0$ and $h''' \geq 0$ for all $x > 0$; and let $f(x) \equiv g(x) - xh(1/x)$ be the function whose zero ζ is sought. Such a computation is encountered in certain financial transactions in which $x = 1+i$ is related to the interest rate i, $g(x)$ represents the present value of various past investments, $xh(1/x)$ represents the present value of anticipated returns from those investments, and $f(x)$ is the net present value of the transaction. ζ, where $f(\zeta) = 0$, determines the putative rate of return on money invested in the transaction.

Newton's iteration can be shown to converge to f's sole positive zero ζ from every positive starting iterate x_0 by invoking the corollary above. Note that for all $x > 0$

$$(xh(1/x))' = h(1/x) - h'(1/x)/x = \int_0^{1/x} (h'(w)-h'(1/x))\,dw \leq 0$$

whence $f' > 0$. And $f'''(x) = g'''(x) + 3x^{-4}h''(1/x) + x^{-5}h'''(1/x) \geq 0$ so f'' can reverse sign at most once for $x > 0$. Moreover, for all $x > 0$

$$N(x) \equiv x - f(x)/f'(x) = \big(xg'(x) - g(x) + h'(1/x)\big)/\big(g'(x) - (xh(1/x))'\big) > 0$$

because $xg'(x) - g(x) = \int_0^x (g'(x)-g'(w))dw \geq 0$; therefore N maps the positive real axis to itself. Finally observe that $xh(1/x) \to h'(0) > 0$ as $x \to +\infty$, while $\lim_{x\to 0^+} (xh(1/x)) = \lim_{y\to+\infty} h(y)/y =$ (either $+\infty$ or $\lim_{y\to+\infty} h'(y)$) > 0, and consequently $f(x)$ and $x - N(x)$ must take negative values in the neighborhood of $x = 0+$, positive values in the neighborhood of $x = +\infty$; and since $f''' \geq 0$ either $f'' \equiv 0$ for all $x > 0$ or else ff'' must take positive values in at least one of those neighborhoods. Therefore f satisfies the corollary's conditions in some closed sub-interval I of the positive real axis. □

Application 2: Suppose ξ and η are two consecutive distinct zeros of f' between which f is twice differentiable, and suppose $f(\xi)f(\eta) < 0$. Then strictly between ξ and η lies just one zero ζ of f, and ζ's catchment basin for Newton's iteration $x_{n+1} = x_n - f(x_n)/f'(x_n)$ includes in its interior at least one place, also strictly between ξ and η, where f'' reverses sign.

Proof: The term "catchment basin" was defined near the middle of §4. Let $N(x) \equiv x - f(x)/f'(x)$; it is continuous at ζ because $f'(\zeta) \neq 0$. Therefore by restricting x and y to a sufficiently small neighborhood around ζ we may ensure that $|1 - f'(y)/f'(x)|$ is as small as we please. Consequently $(N(x)-\zeta)/(x-\zeta) = \int_\zeta^x (1 - f'(y)/f'(x))dy/(x-\zeta)$ may be made as small as we please for all x close enough to ζ, and hence ζ is a strongly attractive fixed point of N, which has no other fixed point

between ξ and η. By invoking §4's lemma, or otherwise, we deduce that the ends of ζ's catchment basin lie strictly between ξ and η, straddle ζ, and are swapped by N. The lemma above implies now that f'' reverses sign at least once inside the catchment basin as claimed. □

This result is significant because it permits all the real zeros of a function f in any interval to be calculated quickly via Newton's iteration provided first all the zeros of two consecutive derivatives $f^{(n)}$ and $f^{(n+1)}$ ($n \geq 1$) in that interval are known. Having straddled a zero of $f^{(n-1)}$ with two consecutive (approximate) zeros of $f^{(n)}$, compute $f^{(n-1)}$ at the enclosed zero(s) of $f^{(n+1)}$ to straddle the desired zero of $f^{(n-1)}$ more closely, and then start Newton's iteration from one of the straddling zeros of $f^{(n+1)}$. The iteration (not just a subsequence of it) must converge to the straddled zero of $f^{(n-1)}$, and must do so One-Sidedly and rapidly (faster than any geometric progression), when started from the right one of those straddling zeros of $f^{(n+1)}$. The right one can be distinguished from the wrong one when iterates started from the wrong one straddle the right one. The rapidity of convergence, the fact that the precision of the computation need not much exceed whatever is required to separate adjacent zeros (multiple zeros announce themselves first as simple zeros of a higher derivative), and the freedom from Sturm sequences are three reasons for considering the foregoing vaguely described algorithm as a potential replacement for others that have appeared elsewhere; cf. Heindel (1971), Collins (1974) and Verbaeten (1975).

§9. The Secant Iteration

Theorem: Suppose I is a closed finite interval in which f is continuous, and continuously once differentiable too except possibly where f vanishes in I. Suppose too that $N(x)$, defined by

$$N(x) \equiv x - f(x)/f'(x) \quad \text{except} \quad N(x) \equiv x \quad \text{when} \quad f(x) = 0 \ ,$$

maps I into itself and satisfies the No Swap condition or one of its equivalents (see §8), as must be the case if Newton's iteration $x_{n+1} = N(x_n)$ converges in I from every x_0 in I. Finally suppose f reverses sign across its (necessarily unique) zero ζ in I. Then the Secant iteration $x_{n+1} = S(x_n, x_{n-1})$, where

$$\begin{aligned} S(x,y) \equiv S(y,x) &\equiv x - f(x)(x-y)/\big(f(x)-f(y)\big) && \text{if} \quad y \neq x \quad \text{and} \quad f(x) \neq 0 \\ &\equiv N(x) && \text{if} \quad y = x \quad \text{or} \quad f(x) = 0 \ , \end{aligned}$$

generates from every x_0 and x_1 in I a sequence $\{x_n\}$ of which some subsequence converges to ζ; i.e. a subsequence of $\{f(x_n)\}$ converges to 0. If also N is continuous then $x_n \to \zeta$.

Proof: Our strategy is to identify certain subsequences of $\{x_n\}$ which converge monotonically to ζ. The Mean Value lemma of §6 and the properties of N and f exposed in §7 and §8 will be exploited heavily. For instance, §7's lemma implies that in I f is monotonic, has just one zero ζ, and is strongly monotonic except possibly at ζ where f' may vanish or fail to exist but cannot reverse sign. Moreover $N(x)$ is continuous in I except possibly at $x = \zeta$, and $S(x,y)$ is continuous in $I \times I$ except possibly at $x = y = \zeta$. These inferences are but the first of a long chain which has been organized into a list of propositions numbered for easier reference. Some of the propositions, like the first, have a proof so

straightforward that it is omitted.

Proposition 1: $x = S(x,y)$ for some x and y in I if and only if $x = \zeta$; consequently the Secant iteration $x_{n+1} = S(x_n, x_{n-1})$ has every $x_{n+1} - x_n \neq 0$ for $n \geq 1$ unless $x_n = \zeta$.

To avoid trivial nuisances we assume henceforth in the theorem's proof that all $x_n \neq \zeta$.

Proposition 2: If x_n and x_{n-1} both lie in I so does $x_{n+1} = S(x_n, x_{n-1})$, and therefore all Secant iterates x_n lie in I.

Proof: Recall $N(I) \subseteq I$ and invoke §6's Mean Value lemma after observing that if ζ lies between x_n and x_{n-1} so must x_{n+1} since $f(x_n)f(x_{n-1}) \leq 0$.

Proposition 3: If some subsequence of $\{x_{n+1}-x_n\}$ converges to zero then the corresponding subsequences of $\{x_n\}$ and $\{x_{n+1}\}$ converge to ζ.

Proof: Since $S(x,y) - x$ is continuous throughout the compact square $I \times I$ except possibly at the one point $x = y = \zeta$, and since $S(x,y) - x = 0$ only on the line $x = \zeta$ according to proposition 1, $x_{n+1} - x_n = S(x_n, x_{n-1}) - x_n \to 0$ implies $x_n - \zeta \to 0$ and hence $x_{n+1} \to \zeta$ as claimed.

Definitions: An iterate $x_n = S(x_{n-1}, x_{n-2})$ is called a Variance whenever $f(x_{n-1})/f(x_n) < 0$, and then $x_{n+1} = S(x_n, x_{n-1})$ and ζ must both lie strictly between x_n and x_{n-1}.

An iterate $x_n = S(x_{n-1}, x_{n-2})$ is called a Permanence whenever $f(x_{n-1})/f(x_n) > 1$, and then $x_{n+1} = S(x_n, x_{n-1})$ and ζ must both lie strictly on the same side of both x_n and x_{n-1}.

Proposition 4: Every iterate $x_n = S(x_{n-1}, x_{n-2})$ with $n \geq 2$ is either a Permanence or a Variance.

See Fig. 10.

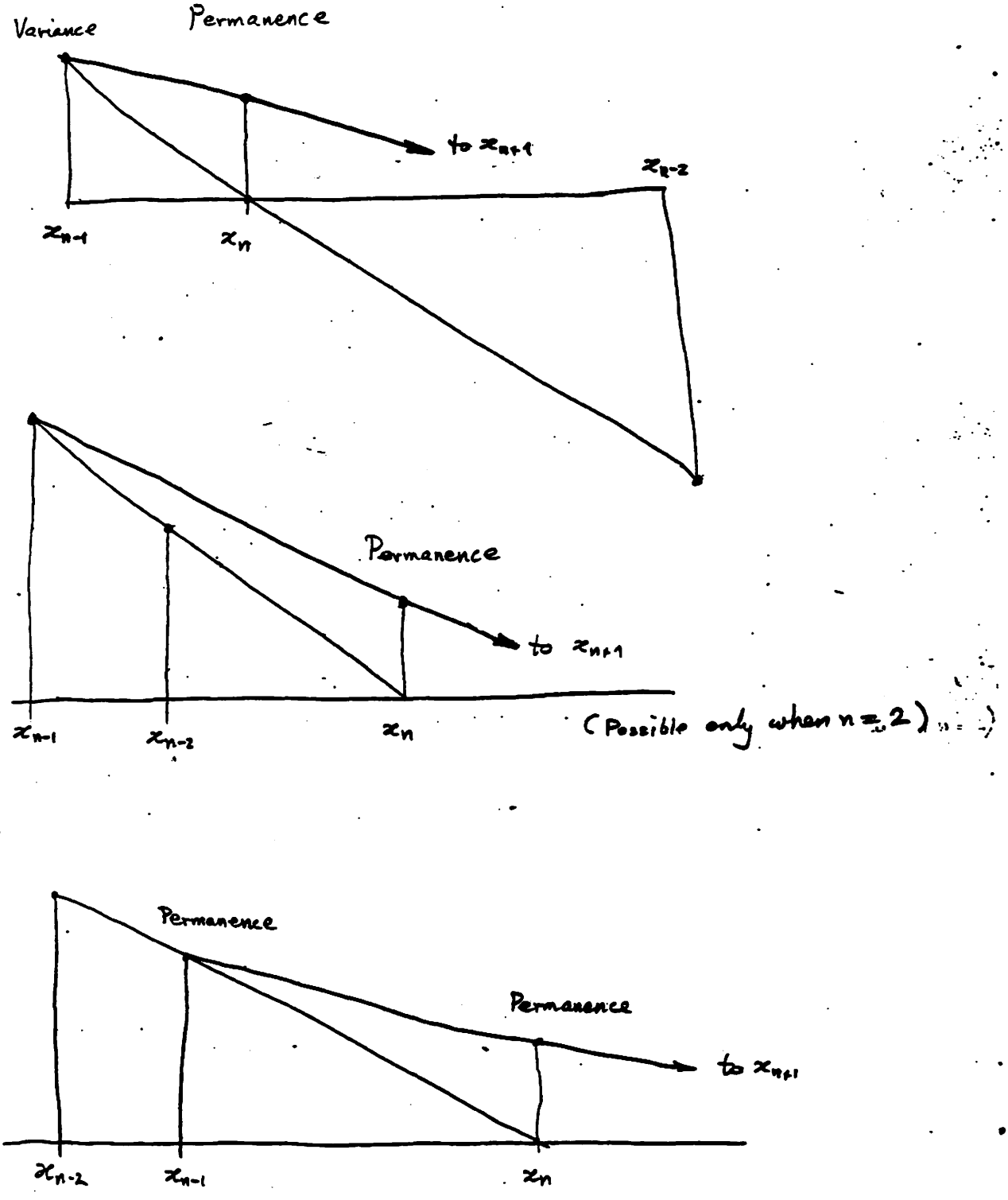


Fig. 10: Three illustrations of Proposition 4.

Proof: Only the proof that when $n \geq 2$ and $f(x_{n-1})/f(x_n) \neq 0$ then $f(x_{n-1})/f(x_n) \geq 1$ requires as much effort as invoking the monotonicity of f in two cases, $f(x_{n-2})/f(x_{n-1}) < 0$ and $f(x_{n-2})/f(x_{n-1}) > 0$.

Proposition 5: If two consecutive iterates $x_n = S(x_{n-1},x_{n-2})$ and $x_{n+1} = S(x_n,x_{n-1})$ are both Variances, then x_{n+1} lies strictly between x_{n-1} and x_n, and x_{n+2} and ζ both lie strictly between x_{n+1} and x_n, and

$$(x_n-x_{n-1})/(x_{n+2}-x_{n+1}) > 4 .$$

See Fig. 11.

Proof: Only the last inequality is an unobvious inference. By hypothesis both $f(x_{n-1})/f(x_n) < 0$ and $f(x_n)/f(x_{n+1}) < 0$; and because x_{n+1} lies between x_{n-1} and ζ, and f is strongly monotonic, $f(x_{n-1})/f(x_{n+1}) > 1$. Therefore

$$\begin{aligned}
\frac{x_n - x_{n-1}}{x_{n+2} - x_{n+1}} &= \frac{(x_n-x_{n-1})(f(x_{n+1}) - f(x_n))}{-f(x_{n+1})(x_{n+1}-x_n)} \\
&= \frac{(f(x_{n+1}) - f(x_n))(f(x_n) - f(x_{n-1}))}{f(x_{n+1})f(x_n)} \\
&= 1 - \frac{f(x_n)}{f(x_{n+1})} - \frac{f(x_{n-1})}{f(x_n)} + \frac{f(x_{n-1})}{f(x_{n+1})} \\
&\geq 1 + 2\sqrt{f(x_{n-1})/f(x_{n+1})} + f(x_{n-1})/f(x_{n+1}) \\
&> 4 .
\end{aligned}$$

Proposition 6: If the Secant iterates x_n are ultimately (i.e. for all sufficiently large n) all Permanences, or ultimately all Variances, they converge to ζ.

Proof: If ultimately all x_n are Permanences they form a sequence which is ultimately monotonic and bounded (by ζ) in I; therefore the sequence converges and, by proposition 3, converges to ζ. If ultimately

Fig. 11: To illustrate Proposition 5

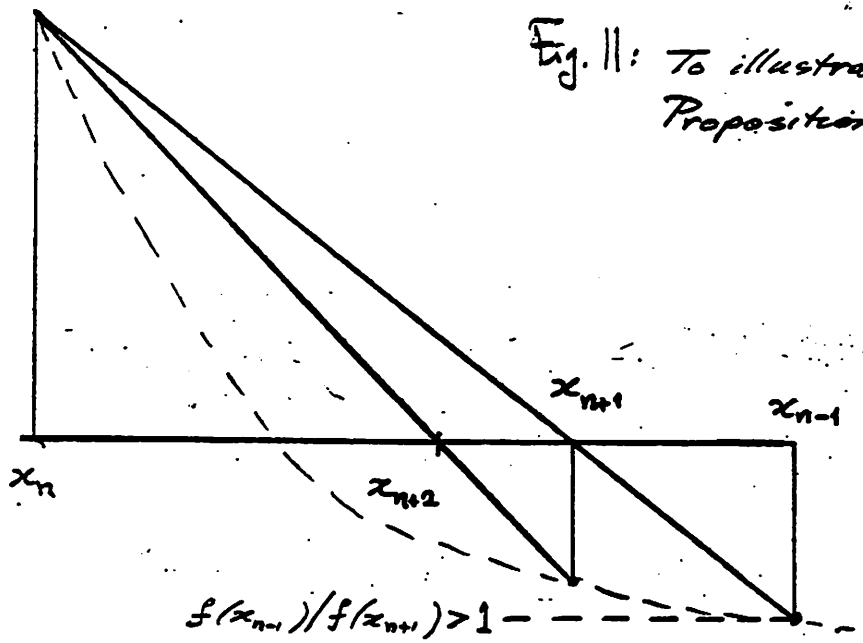
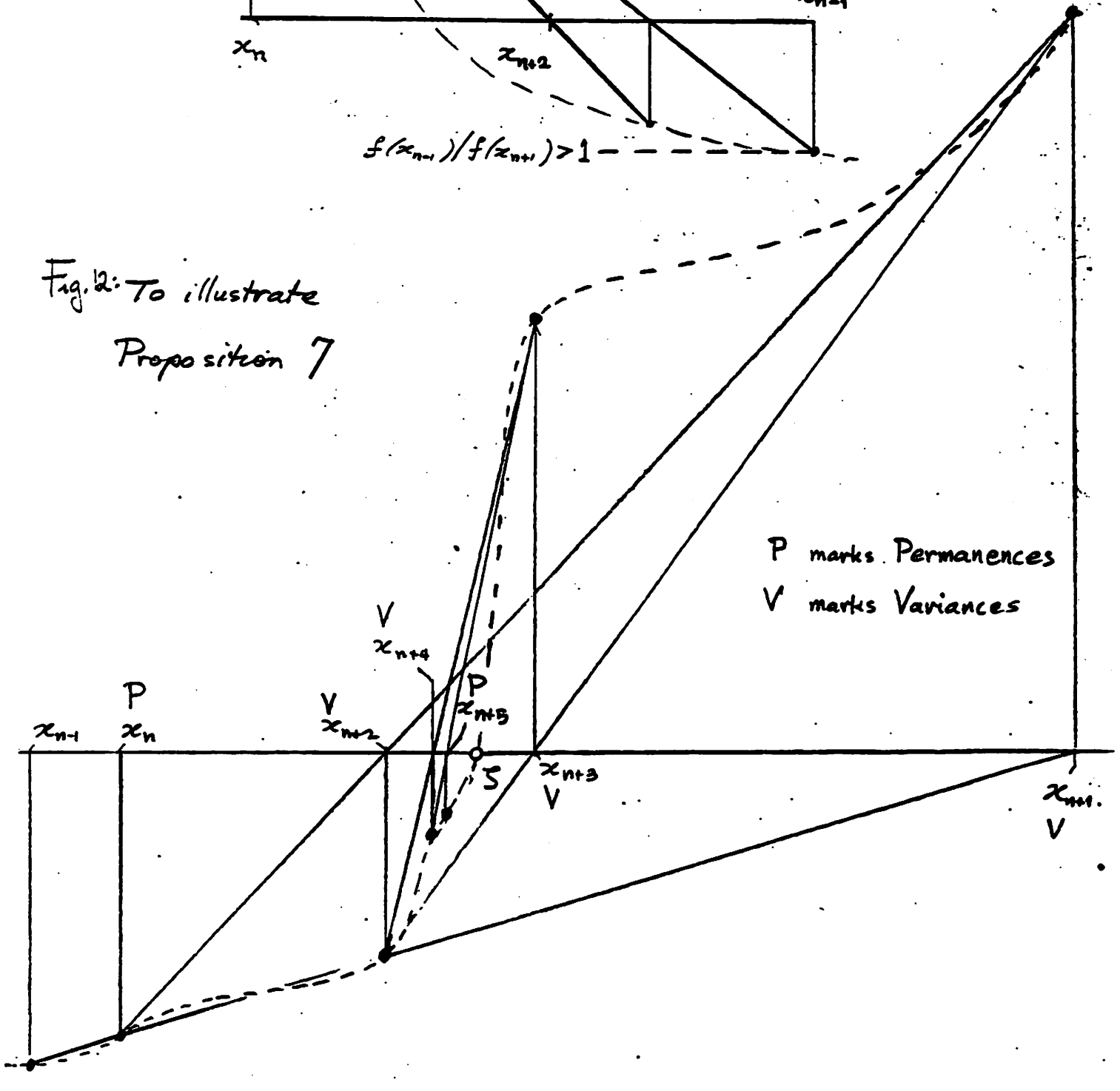


Fig. 12: To illustrate Proposition 7



all x_n are Variances then the subsequences $\{x_{2n}\}$ and $\{x_{2n+1}\}$ are ultimately monotonic and convergent; moreover $|x_{2n+1} - x_{2n}| \to 0$ at least as fast as some multiple of 4^{-n} because of proposition 5, so $x_n \to \zeta$ as claimed.

Permanences are best thought of as punctuation marks separating strings of consecutive Variances, and the only significant property of each string is whether its length is even or odd. The significance of even length is suggested by the next proposition, whose straightforward proof is omitted.

Proposition 7: If a Permanence x_n is followed by an even number $2k \geq 0$ of consecutive Variances $x_{n+1}, x_{n+2}, \ldots, x_{n+2k}$ before the next Permanence x_{n+2k+1}, then the numbers $x_{n-1}, x_n, x_{n+2}, x_{n+4}, \ldots, x_{n+2k}, x_{n+2k+1}, \zeta, x_{n+2k-1}, \ldots, x_{n+3}, x_{n+1}$ are exhibited here in strictly monotonic order. (If $2k = 0$ or 2 delete the appropriate right-hand-most x's.) See Fig. 12.

If at most finitely many strings of Variances have odd length the convergence properties of the sequence $\{x_n\}$ are relatively transparent, as the next proposition shows.

Proposition 8: If ultimately no two Permanences are separated by an odd number of Variances then the Permanences converge to ζ; if also N is continuous at ζ the Variances converge to ζ too.

Proof: By re-numbering the iterates to discard some early ones if necessary, we may assume no two Permanences are separated by an odd number of Variances, in which case the previous proposition implies that the Permanences and their antecedent iterates constitute a monotonic bounded (by ζ) subsequence of the iterates. In other words, if the successive Permanences are $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$ then the numbers

$$x_{n_1-1}, x_{n_1}, x_{n_2-1}, x_{n_2}, x_{n_3-1}, x_{n_3}, \ldots, \zeta$$

are displayed here in strictly monotonic order. Obviously this subsequence of x's must converge and, by proposition 3, it must converge to ζ. Before we find out what happens to the rest of the iterates x_n let us invoke §6's Mean Value lemma to define y_{n_j} as a solution of $x_{n_j+1} = N(y_{n_j})$ between x_{n_j-1} and the Permanence x_{n_j} for $j = 1,2,3,\ldots$. Evidently $y_{n_j} \to \zeta$ too. If N is continuous at ζ we may infer first that $x_{n_j+1} = N(y_{n_j}) \to N(\zeta) = \zeta$ and then, from proposition 7, that all $x_n \to \zeta$ as $n \to \infty$. If N is discontinuous at ζ there is some risk that $x_{n_j+1} = N(y_{n_j}) \not\to N(\zeta)$; this situation arises with examples f one of which is exhibited at the proof's end.

To complete the theorem's proof we need only deal with the possibility that infinitely many pairs of consecutive Permanences are separated by odd numbers of Variances. This possibility is awkward only because the notation required to deal with it is complicated.

Let us invoke §6's Mean Value lemma again to define for every Permanence x_n the solution y_n of $N(y_n) = x_{n+1}$ between x_n and x_{n-1} and closest to x_n. These solutions y_n were useful already during the previous proposition's proof, but they are crucial below because they provide the sole entrée for N's No-Swap condition. Note that y_n is so far defined not at every n but only at those n for which x_n is a Permanence.

Next define a Scout to be a Permanence x_n followed by an odd number of Variances, say $x_{n+1}, x_{n+2}, x_{n+3}, \ldots, x_{n+2k_n+1}$, and then another Permanence x_{n+2k_n+2}. That last Permanence might be a Scout too or it might not. Also define x_{n+1} to be a Guard whenever x_n is a Scout. We shall think of each y_n as a part of a convoy with a Scout ranging ahead of it and a Guard bringing up the rear, and our last task will be to show that alternate convoys converge monotonically to ζ from opposite sides. See Fig. 13.

Fig. 13: Examples of Convoys with Scouts and Guards

P marks Permanences S marks Scouts
V marks Variances G marks Guards
$N(y_j) = x_{j+1}$ whenever x_j is a Permanence.

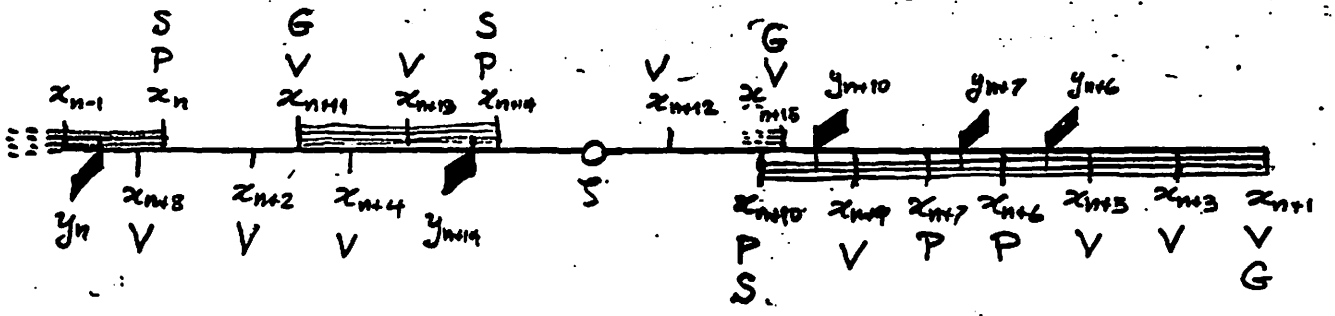
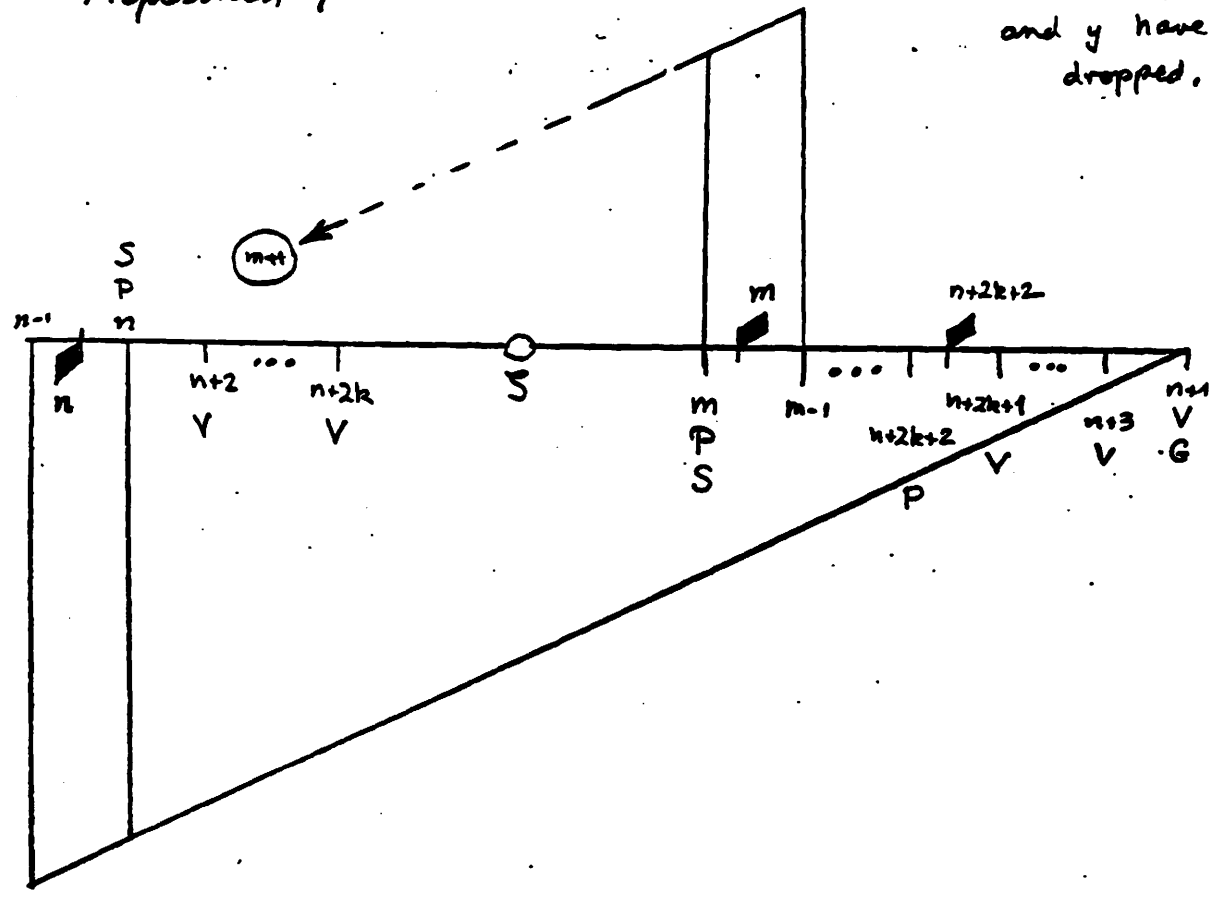


Fig 14: To illustrate Proposition 9

Same notation as above except x and y have been dropped.



Proposition 9: Suppose x_n and x_m are consecutive Scouts with $m > n$; then $m \geq n+2$ and x_{n-1}, y_n, x_{m+1}, ζ, y_m, x_{n+1} are displayed here in strictly monotonic order.

See Fig. 14.

Proof: For definiteness assume $x_{n-1} < x_n$, whereupon it follows that $x_{n-1} < y_n \leq x_n < \zeta < x_{n+1}$. Moreover we know that the Scout x_n is followed by an odd number of Variances $x_{n+1}, x_{n+2}, \ldots, x_{n+2k+1}$ and then the next Permanence x_{n+2k+2}, so $m \geq n+2k+2 \geq n+2$ and $x_{n-1} < y_n \leq x_n < x_{n+2} < x_{n+4} < \cdots < x_{n+2k} < \zeta < x_{n+2k+2} < x_{n+2k+1} < \cdots < x_{n+3} < x_{n+1}$. We also know that an even number of Variances separates every two consecutive Permanences between x_{n+2k+2} and x_m, so proposition 7 implies

$$\zeta < x_m < x_{m-1} \leq x_{n+2k+1} \leq x_{n+1} ;$$

and then $x_{m+1} < \zeta < x_m \leq y_m < x_{m-1} \leq x_{n+1}$. The only question left is whether or not $y_n < x_{m+1}$. If not, if instead $x_{m+1} = N(y_m) \leq y_n < \zeta < y_m < x_{n+1} = N(y_n)$, then the No Crossover condition satisfied by N would be violated contrary to the theorem's hypotheses. Therefore proposition 9 is proved and more;

$$x_{n-1} < y_n \leq x_n < \zeta \quad \text{and} \quad y_n < x_{m+1} < \zeta < x_m \leq y_m < x_{m-1} \leq x_{n+1} .$$

Proposition 10: If the sequence of Secant iterates x_n contains infinitely many Scouts then a subsequence of Guards converges to ζ, and all $x_n \to \zeta$ if N is continuous.

Proof: Let the integer sequence $m(1) < m(2) < m(3) < \cdots$ characterize consecutive Scouts $x_{m(j)}$, Guards $x_{m(j)+1}$ and convoy's contents $y_{m(j)}$. Assuming for definiteness that $x_{m(1)} < \zeta$, we infer from proposition 9 et seq. that for $j = 1,2,3,\ldots$

$$y_{m(2j-1)} < x_{m(2j)+1} < \zeta < x_{m(2j)} \leq y_{m(2j)} < x_{m(2j)-1} \leq x_{m(2j-1)+1} \quad \text{and}$$
$$x_{m(2j)+1} \leq x_{m(2j+1)-1} < y_{m(2j+1)} \leq x_{m(2j+1)} < \zeta < x_{m(2j+1)+1} < y_{m(2j)} .$$

By induction it follows that $y_{m(2j+1)}$ increases to a limit $\check{y}$ and $y_{m(2j)}$ decreases to a limit $\hat{y} \geq \zeta \geq \check{y}$, and the same is true respectively for the Guards $x_{m(2j)+1} \to \check{y}$ and $x_{m(2j+1)+1} \to \hat{y}$, as $j \to \infty$. (The Scouts $x_{m(2j)}$ and $x_{m(2j+1)}$ need not form monotonic subsequences.)

The possibility that $\check{y} < \zeta < \hat{y}$ can be ruled out because $\hat{y} = \lim y_{m(2j)} \neq \zeta$ would imply $\check{y} = \lim x_{m(2j)+1} = \lim N(y_{m(2j)}) = N(\hat{y})$ and similarly $\hat{y} = N(\check{y})$ in violation of the No Swap condition satisfied by N. And if N is continuous at ζ a similar argument shows $\check{y} = \zeta = \hat{y}$; but in this case we soon infer that all iterates, Scouts included, are squeezed towards ζ by the Guards and hence converge to ζ.

Propositions 6, 8 and 10 exhaust all possible ways for the sequence of Secant iterates to behave, and hence prove the theorem. □

Example: In this example $f(x)$ is infinitely differentiable throughout $I \equiv \{-1 \leq x \leq -1/(1 - 1/\ln 2) = 2.25889...\}$, vanishes only at $x = \zeta \equiv 0$, and is elsewhere in I strongly monotonic. Newton's iteration converges from every x_0 in I, but the Secant iteration suffers from an oscillation in which every third iterate $x_{3n+2} = -1$ though the remaining iterates $\{x_{3n}$ and $x_{3n+1}\}$ converge to ζ. The construction of f is complicated enough that only an outline can be presented here.

Start by defining for $n = 0,1,2,...$ the descending sequences $\xi_{3n} \equiv 1/\ln(n+2)$ and $\xi_{3n+1} \equiv (\xi_{3n} + \xi_{3n+3})/2$, while $\xi_{3n+2} \equiv -\infty$ for all $n \geq 0$. Next define

$$\phi(\xi) \equiv \xi \exp(1/\xi) \quad \text{for} \quad -\infty \le \xi < 0 ,$$
$$\equiv 0 \quad \text{at} \quad \xi = 0 ,$$
$$\equiv (\xi_{3n} - \xi_{3n+3})/2 \quad \text{for} \quad \xi_{3n+1} \le \xi \le \xi_{3n} ,$$
$$\equiv \alpha_n + \beta_n \sigma\left((2\xi - \xi_{3n+1} - \xi_{3n})/(\xi_{3n+1} - \xi_{3n+3})\right) \quad \text{for} \quad \xi_{3n+3} \le \xi \le \xi_{3n+1} ,$$

where $\alpha_n \equiv (\phi(\xi_{3n+1}) + \phi(\xi_{3n+3}))/2$,

$\beta_n \equiv (\phi(\xi_{3n+1}) - \phi(\xi_{3n+3}))/2$, and

$\sigma(\theta) \equiv \tanh(\tan(\pi\theta/2))$ if $-1 \le \theta \le 1$,

$\equiv \text{sign}(\theta)$ otherwise .

Finally $f(x) \equiv (1+x)\phi(x/(1+x))$ and $x_n \equiv \xi_n/(1-\xi_n)$. The tedious verification of the claims made above for f are left to the reader.

The theorem that has just been proved does not yet render Newton's iteration obsolete. Rather it supplies a powerful incentive for replacing Newton's by the Secant iteration in those cases where, as in §8's Application 1, the calculation of a derivative appears to confer no advantage. Many cases remain to be analyzed; for instance, we cannot yet say whether the Secant iteration works acceptably well on that rational function in §4 with interlacing poles and zeros on which Newton's iteration works so well.

A final warning; computer programs based upon Newton's or the Secant iteration rarely use an iteration in its pristine form. Programs usually incorporate extra "features" which modify the iterations in ways that their designers hope will effect some improvements. Sometimes those features do yield an improvement, sometimes not; they almost always undermine the foregoing analysis.

Acknowledgment: I am indebted to Professor B.N. Parlett for helpful discussions without which this paper would have been more nearly impossible to read.

§10. Bibliography

anonymous. "Modern Computing Methods--Notes on Applied Science No. 16" National Physical Laboratory, 1st ed. (1957) p. 23, 2nd ed. (1961) p. 56. Her Majesty's Stationery Office, London.

anonymous (1974). "HP-65 Standard Pac" 00065-90207 Rev. 6/74, pp. 42-45 and 73. Hewlett-Packard, Cupertino, Calif.

T.M. Apostol (1967). "Calculus" 2nd ed. vol. I pp. 184-7. Wiley, N.Y.

K.-H. Bachmann (1954). "The Order of Convergence of inverse interpolation iterations for solving equations" (German) Zeits. angew. Math. Mech. 34 pp. 282-3.

V.V. Bashurov and V.N. Ogibin (1966). "Conditions for the Convergence of Iterative Processes on the Real Axis" (Russian). Zh. vychisl. Mat. mat. Fiz. 6 (1966) pp. 913-916; translated by H.F. Cleaves in U.S.S.R. Computational Mathematics and Mathematical Physics 6 #5 pp. 178-184 (1968) (Math. Rev. 34 (1968) #1472).

R.P. Brent (1973). "Algorithms for Minimization without Derivatives" pp. 58-60. Prentice-Hall, Englewood Cliffs, N.J.

A.M. Bruckner and J.G. Ceder (1965). "Darboux Continuity" Jahresbericht der Deutschen Mathem.-Verein. 67 I. Abt. pp. 93-117 (Math. Rev. 32 (1966) #4217).

J.C.P. Bus and T.J. Dekker (1974). "Two efficient algorithms with guaranteed convergence for finding a zero of a function" Stichting Mathematisch Centrum, Amsterdam, Afdeling Numerieke Wiskunde NW 13/74, or ACM Transactions on Math. Software 1 (1975) pp. 330-345.

G.E. Collins (1974). "High-Precision Calculation of Real Algebraic Numbers" (Abstract) ACM SIGSAM Bulletin 8 p. 2.

G. Dahlquist, Åke Björck and Ned Anderson (1974). "Numerical Analysis" pp. 227-232. Prentice-Hall, Englewood Cliffs, N.J.

M. Dowell and P. Jarratt (1971). "A modified Regula Falsi method for computing the root of an equation" BIT 11 pp. 168-174.

R.J. Fateman (1977). "An Algorithm for Deciding the Convergence of the Rational Iteration $x_{n+1} = f(x_n)$" ACM Trans. Math. Software 3 pp. 272-8.

L.E. Heindel (1971). "Integer Arithmetic Algorithms for Polynomial Real Zero Determination" Jl. ACM 18 pp. 533-548.

A.S. Householder (1970). "The Numerical Treatment of a Single Nonlinear Equation" ch. 4. McGraw-Hill, N.Y.

T.-Y. Li and J.A. Yorke (1975). "Period three implies chaos" Amer. Math. Monthly 82 pp. 985-992.

J.M. Ortega and W.C. Rheinboldt (1970). "Iterative Solution of Nonlinear Equations in Several Variables". Academic Press, New York.

A. Ostrowski (1960). "Solution of Equations and Systems of Equations" ch. 3. Also 2nd ed. (1966) and 3rd ed. (1973). Academic Press, New York.

Y. Saad (1974). "Shifts of Origin for the QR Algorithm" in "Information Processing 74 -- Proceedings of the IFIP Congress of Aug. 5-10, 1974 in Stockholm" pp. 527-531. North-Holland Publ., Amsterdam.

A.N. Šarkovskiĭ (1960). "Necessary and sufficient conditions for convergence of one-dimensional iterative processes" (Russian) Ukrain. Mat. Ž 12 No. 4, pp. 484-9 (Math. Rev. 25 (1963) #353).

_____ (1961). "The reducibility of a continuous function of a real variable and the structure of the stationary points of the corresponding iteration process" (Russian) Dokl. Akad. Nauk SSSR 139, pp. 1067-1070 (Math. Rev. 25 (1963) #352).

_____ (1964). "Co-existence of cycles of a continuous mapping of the line into itself" (Russian, with English summary) Ukrain. Mat. Ž 16 No. 1, pp. 61-71 (Math. Rev. 28 (1964) #3121).

_____ (1965). "On cycles and the structure of a continuous mapping" (Russian) ibid. 17 No. 3, pp. 104-111 (Math. Rev. 32 (1966) #4213).

R.S. Stepleman (1975). "A Characterization of Local Convergence for Fixed-point Iterations in R^1" SIAM J. Numer. Anal. 12 pp. 887-894.

J.F. Traub (1964). "Iterative Methods for the Solution of Equations" Prentice-Hall, Englewood Cliffs, N.J.

P. Verbaeten (1975). "Computing Real Zeros of Polynomials with SAC-1" ACM SIGSAM Bulletin 9 pp. 8-10 & 24.

M.V. Wilkes, D.J. Wheeler and S. Gill (1951). "The Preparation of Programs for an Electronic Digital Computer" pp. 84-85 and 130-132. Addison-Wesley, Cambridge, Mass.

M.V. Wilkes (1966). "A Short Introduction to Numerical Analysis" p. 12 Cambridge University Press.

J.H. Wilkinson (1967). "Two Algorithms based on successive linear interpolation" Computer Science Department Report CS 60, Stanford University.