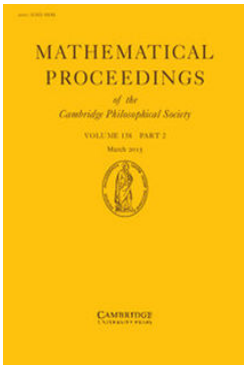
Mathematical Proceedings of the Cambridge Philosophical Society

http://journals.cambridge.org/PSP

Additional services for ***Mathematical Proceedings of the Cambridge Philosophical Society:***

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here



The solution of equations by iteration

W. A. Coppel

Mathematical Proceedings of the Cambridge Philosophical Society / Volume 51 / Issue 01 / January 1955, pp 41 - 43
DOI: 10.1017/S030500410002990X, Published online: 24 October 2008

Link to this article: http://journals.cambridge.org/abstract_S030500410002990X

How to cite this article:
W. A. Coppel (1955). The solution of equations by iteration. Mathematical Proceedings of the Cambridge Philosophical Society, 51, pp 41-43 doi:10.1017/S030500410002990X

Request Permissions : Click here

CAMBRIDGE JOURNALS

Downloaded from http://journals.cambridge.org/PSP, IP address: 138.251.14.35 on 18 Mar 2015

THE SOLUTION OF EQUATIONS BY ITERATION

By W. A. COPPEL

Communicated by P. Hall

Received 29 March 1954

The method of solving equations by iteration is very old and is discussed in many well-known books*. But conditions for its validity have never been properly formulated. In the first place, it is necessary to know that the method will not carry us outside the domain of definition of our functions, and that the 'approximations' will not converge to something which is not a solution of the equations. These difficulties are easily forestalled; it is more difficult to ensure that the process really will converge. Our object will be to find the most general conditions under which we can set off with the certainty that we will ultimately arrive at a root, in the case of one equation with one unknown.

We may suppose that the equation to be solved is written in the form $x = f(x)$ and that, starting from some point x_1, our successive approximations are $x_2 = f(x_1)$, $x_3 = f(x_2)$, etc. To prevent the difficulties first mentioned from occurring we make, once and for all, the following restrictions on the function f:

$f(x)$ is defined as a continuous function in an interval $a \leqslant x \leqslant b$, and $a \leqslant f(x) \leqslant b$.

Thus, if x lies in the interval (a, b), so does $x' = f(x)$. And if $x_n \to x$, then since f is assumed continuous, $x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(x)$.

From $f(c) = c$ it follows that $f(f(c)) = f(c) = c$. Suppose the equation $f(f(x)) = x$ has a solution c which is not also a solution of the equation $f(x) = x$. If we take $x_1 = c$ we will have $x_2 = f(c) \neq c$, $x_3 = c$, and in general $x_{2n-1} = c$, $x_{2n} = f(c)$. Thus the sequence (x_n) will oscillate. Therefore, if the iteration sequence is to converge whatever initial point is chosen, it is necessary that the equation $f(f(x)) = x$ have no roots beside the roots of the equation $f(x) = x$. We intend to show that, conversely, if this condition is satisfied every iteration sequence converges.

To assist us in this task we use the iterates of $f(x)$. These are the functions defined by the recurrence relations $f_1(x) \equiv f(x)$, $f_{n+1}(x) \equiv f(f_n(x))$ $(n = 1, 2, 3, \ldots)$. They are continuous for $a \leqslant x \leqslant b$ and satisfy the same inequality as f: $a \leqslant f_n(x) \leqslant b$. It is easily proved by a double induction that $f_n(f_m(x)) \equiv f_{m+n}(x) \equiv f_m(f_n(x))$. If $f(x) = x$ for a particular x, then $f_n(x) = x$ for all n. The proof of the main theorem depends on the following

Preliminary theorem. *If every root of $f_2(x) = x$ is a root of $f_1(x) = x$ then for any c and any n, $f_n(c) >, =, < c$ according as $f(c) >, =, < c$ $(a \leqslant c \leqslant b)$.*

(I) Suppose it is known that for $n \leqslant m$ the theorem is true and that $f_{m+1}(x) = x$ implies $f(x) = x$. This is certainly known already for $m = 1$.

* E.g. E. T. Whittaker and G. Robinson, *The calculus of observations* (4th ed. Glasgow, 1944), chap. vi or R. Courant, *Differential and integral calculus* (Glasgow, 1934), vol. I, p. 358.

Let $f_{m+1}(c) < c$. Then $f(c) \neq c$. We will assume $f(c) > c$ and deduce a contradiction. Since $f_{m+1}(a) \geqslant a$ and $f_{m+1}(c) < c$ we have $c > a$ and there is a point d, $a \leqslant d < c$, such that $f_{m+1}(d) = d$ and $f_{m+1}(x) < x$ for $d < x \leqslant c$. By supposition this implies $f(d) = d$. If at some point between d and c we had $f(x) = x$ we would also have $f_{m+1}(x) = x$, contrary to the definition of d. Therefore $f(x) \neq x$ for $d < x < c$ and from $f(c) > c$ it follows that

$$f(x) > x \quad \text{for} \quad d < x < c. \tag{1}$$

Hence, by our supposition,

$$f_m(d) = d \quad \text{and} \quad f_m(x) > x > d \quad \text{for} \quad d < x < c. \tag{2}$$

Choose e, between d and c, so near d that $d < f_m(e) < c$. Then by (1) and (2),

$$f_{m+1}(e) = f(f_m(e)) > f_m(e) > e,$$

which contradicts the definition of d. We must therefore have $f(c) < c$.

Similarly, using values of x between c and b, we show that if $f_{m+1}(c) > c$ then $f(c) > c$. On the one hand this proves the theorem for $n = 2$. On the other hand, it shows that if we can prove that $f_{m+1}(x) = x$ implies $f(x) = x$, on the supposition that the theorem is true for $n \leqslant m$, then the theorem will be completely proved by induction.

(II) Suppose, then, that the theorem is true for $n \leqslant m$ $(m \geqslant 2)$ and let $f_{m+1}(c) = c$.

Assume that if possible $f(c) > c$. Then by supposition $d = f_m(c) > c$. If we had $d = f_{m-1}(f(c)) \geqslant f(c)$ our supposition would give first $f(f(c)) \geqslant f(c)$ and then $f_m(f(c)) \geqslant f(c)$; that is, $f_{m+1}(c) = c \geqslant f(c)$, which is contrary to assumption. Therefore $d < f(c)$. Also, $f(d) = f_{m+1}(c) = c$.

Thus we have $f(d) = c < d < f(c)$. Therefore, at some point between c and d, $f(x)$ takes the value d. Let e be the point nearest to c at which this happens. Then $c < e < d$, $f(e) = d$, and $f(x) > d > e$ for $c \leqslant x < e$. Hence $f_2(e) = f(d) = c < e$; but $f_2(c) > c$, because $f(c) > c$. Therefore at some point x between c and e we must have $f_2(x) = x$, and hence $f(x) = x$. But this is impossible because $f(x) > e$ for $c \leqslant x < e$.

Similarly, the assumption that $f(c) < c$ leads to a contradiction. We conclude that $f(c) = c$. This completes the proof of the theorem.

It is now easy to prove our

MAIN THEOREM. *A necessary and sufficient condition that the iteration sequence converge, whatever initial point is chosen, is that the equation $f(f(x)) = x$ have no roots except the roots of the equation $f(x) = x$.*

The necessity of the condition has already been pointed out. Suppose, then, that the condition is satisfied and let (x_n) be any iteration sequence. If, for some m, $x_{m+1} = x_m$ then $x_n = x_m$ for all $n > m$ and the sequence certainly converges. We may therefore assume that $x_{n+1} \neq x_n$ for every n. Also, since the sequence is confined to the interval (a, b), it will converge if from some point on it is monotonic. We may therefore assume that $x_{n+1} > x_n$ for infinitely many n and $x_{n+1} < x_n$ for infinitely many n.

Denote by p suffixes of the first type, for which $f(x_p) > x_p$. Then, by the preceding theorem, $x_n = f_{n-p}(x_p) > x_p$ for all $n > p$. Hence the subsequence (x_p) increases to a limit l and $l = \lim x_n$. Similarly, if we denote by q suffixes of the second type, for which $f(x_q) < x_q$, we see that the subsequence (x_q) decreases to a limit k and $k = \lim x_n$. But every x_n belongs to one or other of these subsequences, and so for infinitely many

n we must have x_n in the first subsequence and x_{n+1} in the second. Proceeding to the limit through these values of n we get $x_n \to l$, $x_{n+1} = f(x_n) \to k$. Hence $k = f(l)$. For infinitely many n also we must have x_n in the second subsequence and x_{n+1} in the first, from which it follows in the same way that $l = f(k)$. Thus $f(f(l)) = l$. Hence, by the hypothesis of the theorem, $f(l) = l$; that is, $k = l$ or $\overline{\lim}\, x_n = \underline{\lim}\, x_n$. Thus the theorem is proved.

Since the function $f(x) - x$ is non-negative for $x = a$ and non-positive for $x = b$, we know without the hypothesis of the main theorem that the equation $f(x) = x$ has at least one root. Nothing has been said so far to exclude the possibility that it has more than one root. However, for the iteration process to be of real value we must know to which root we are approximating. We therefore desire only one root in the interval. Again, the mere fact of convergence is not enough for the practical computation of a root; if possible, we would like the approximation to improve at every step. Just when these desires are satisfied is told in our

CONCLUDING THEOREM. *A necessary and sufficient condition that the equation $f(x) = x$ has exactly one root, that every iteration sequence converge to it, and that the approximations always improve at each step, is that there exist a number r such that $|r - f(x)| < |r - x|$ for $x \neq r$.*

The necessity of the conditions is immediate. For let r be the root of the equation $f(x) = x$; then the inequality is just the requirement that for the initial point x the second approximation be better than the first.

Conversely, suppose that $|r - f(x)| < |r - x|$ for $x \neq r$. If the equation $f(x) = x$ had a root $s \neq r$ we would have $|r - s| < |r - s|$, which is impossible. Since the equation does have a root, r must be the one and only root. If for two numbers s, t distinct from r we had $t = f(s)$, $s = f(t)$ we would have both $|r - t| < |r - s|$ and $|r - s| < |r - t|$, which is impossible. Therefore, by the main theorem, every iteration sequence converges. Since r is the unique root of our equation, the given inequality states that the approximations improve at each step.

A sufficient condition for the conclusions of this theorem, which can also be proved independently of the main theorem, is that for any two distinct points x and y we have $|f(x) - f(y)| < |x - y|$*. This may be compared with the condition usually given (without considering if the iteration process is well-defined); namely, that there is a constant θ, $0 < \theta < 1$, such that $|f(x) - f(y)| < \theta |x - y|$.

* Take r to be a root of the equation $f(y) = y$.

TRINITY COLLEGE
CAMBRIDGE