An Extension of a Method For Solving the Equation $x=\phi(x)$

by

J. R. Chang

Associate Professor, National Chioa-Tung University

and

J. T. Chu

New York University, New York, U. S. A

Of all methods for solving an equation of the form $x=\phi(x)$ by digital computers, the most powerful is the one based on recursive or iterative technique. That is: assume an initial value of $x=x_0$, compute $x_1=\phi(x_0)$, and in general, compute $x_{i+1}=\phi(x_i)$, $i=0,1,2,\cdots$ The sequence $x_0, x_1, x_2,\cdots$ will approach the solution x_t if the condition, $\underset{x\epsilon\omega}{\text{Max}} |\phi'(x)| < 1$, is satisfied, where ω is the domain under consideration.

If this condition is not satisfied, there is no guarantee that the sequence $x_0, x_1, x_2,\cdots$ will converge to x_t. In this paper, we shall extend the method to a case when this condition is not satisfied.

Before continuing the discussion, we restate the problem as follows: To solve a general single-variable equation of the form $x=\phi(x)$ by using iterative or recursive method, where $\phi(x)$ is a continuous function of x in the domain ω under consideration.

I THEORY OF CONTRACTION MAPPINGS

Consider an arbitrary metric space M. A continuous mapping or transformation T of the metric space M into itself is said to be a contraction if there exists a real number $\alpha<1$ such that:

Note: This work was done in the Moore school of Electrical Engineering, Univ. of Pennsylvania, Phila., Pa U. S. A. from September 1964 to October 1967

$$D(Tx, Ty) \leq \alpha D(x, y),$$

where D(x, y) is a metric distance defined between two points x and y in the metric space.

Two basic terms in a metric space are needed in order to understand Theorem I on contraction mapping, they are,

(1) Fundamental Sequence: A sequence $\{x_n\}$ of points of a metric space M is fundamental if it satisfies the Cauchy criterion, i. e., if for arbitrary $\epsilon > 0$ there exists a N_ϵ such that $D(x_i, x_j) < \epsilon$ for all i, $j > N_\epsilon$.

(2) Complete Metric Space: A metric space M is said to be complete if every fundamental sequence in the space M converges to a point in M.

The following theorem can be found in Ref. (1).

Theorem I: Every contraction mapping T defined on a complete metric space M has one and only one fixed point with respect to T, i. e., the equation $x = Tx$ has one and only one solution.

Proof: Let x_0 be an arbitrary point in the metric space M. Set $x_1 = Tx_0$, $x_2 = Tx_1 = T^2x_0$, and in general let $x_n = Tx_{n-1} = T^n x_0$. we shall show that the sequence $\{x_n\}$ is fundamental. Consider two points x_n and x_m in M, where $m \geq n$, the distance between them is

$$D(x_n, x_m) = D(T^n x_0, T^m x_0)$$
$$\leqq \alpha^n D(x_0, T^{m-n} x_0),$$

but $D(x_0, T^{m-n}x_0) = D(x_0, x_{m-n})$, therefore one has:

$$D(x_n, x_m) \leqq \alpha^n D(x_0, x_{m-n})$$
$$\leqq \alpha^n [D(x_0, x_1) + D(x_1, x_{m-n})] \cdot$$

In general one has:

$$D(x_n, x_m) \leqq \alpha^n [D(x_0, x_1) + D(x_1, x_2) + \cdots + D(x_{m-n-1}, x_{m-n})]$$
$$\leqq \frac{\alpha^n D(x_0, x_1)}{1-\alpha}.$$

Since $\alpha < 1$, this distance is arbitrary small for sufficiently large value of n. Since M is complete, $\lim_{n \to \infty} x_n$ exists. Now let $x_t = \lim_{n \to \infty} x_n$. Then by virtue of the continuity of the mapping T,

$$Tx_t = T \lim_{n\to\infty} x_n$$

$$= \lim_{n\to\infty} Tx_n = \lim_{n\to\infty} x_{n+1} = x_t.$$

Hence the existence of a fixed point in the metric apace M is proved. The uniqueness of this point is proved as follows:

If $Tx=x$ and $Ty=y$, then $D(x, y)=D(Tx, Ty)\leqq\alpha\, D(x, y)$ since $\alpha<1$ and α is real, this implies $D(x, y)=0$, or $x=y$.

This completes the proof.

The above theorem may be used to prove that under certain conditions, the equation $x=\phi(x)$ has an unique solution, and that the solution may be found by iterative methed.

Consider a closed interval R of real numbers. Define distance as the absolute value of the difference between two real numbers, or:

$$D(x, y)= |x-y| \cdot$$

Then this set R is a metric space because:

1. $D(x, y)>0$ if $x\neq y$

   $D(x, y)=0$ if $x=y$

2. $D(x, y)+D(y, z)= |x-y| + |y-z|$

   $\geq |x-z| =d(x,z)$

In addition, this metric space is complete. From now on, we shall consider only such a metric space R.

Theorem II: If $\phi(x)$ is a continuous function defined on a closed interval ω, $\phi'(x)$ is continuous and $|\phi'(x)| \leq\alpha<1$ for all $x\epsilon\omega$, and ϕ is a mapping from the interval ω to itself, then the equation $x=\phi(x)$ has an unique solution which can be obtained by the iterative procedure $x_{i+1}=\phi(x_i)$.

Proof: Since

1. $\phi(x)$ is a mapping from the interval ω to ω, and

2. $1>\alpha=\max_{x\epsilon\omega} |\phi'(x)| \geq |\phi'(x_j)|\ x_1\leq x_j\leq x_2$

$$\geq \left|\frac{\phi(x_2)-\phi(x_1)}{x_2-x_1}\right| = \frac{|\phi(x_2)-\phi(x_1)|}{|x_2-x_1|} \text{ for } x_1, x_2 \in \omega \text{ and } x_1 \leq x_2,$$

$$\therefore\ |\phi(x_2)-\phi(x_1)| \leq \alpha |x_2-x_1| .$$

Therefore the mapping ϕ is a contraction mapping. From Thm. I, we conclude that the equation $x=\phi(x)$ has an unique solution and this solution may be obtained by the iterative process $x_{i+1}=\phi(x_i)$, $i=0, 1, 2, \cdots\cdots$

It is clear that if the condition $1>\alpha\geq |\phi'(x)|$ is not satisfied in the interval ω, then the sequence $x_0, x_1, x_2, \cdots$ may not converge to the solution of the equation $x=\phi(x)$.

Corollary I-I

If $\phi(x)$ is a continuous function defined on a closed interval ω, $\omega=[a, \infty]$, $\phi'(x)$ is continuous and $0<\phi'(x)\leq\alpha<1$ for all $x\epsilon\omega$, and if there exists a solution x_t of the equation $x=\phi(x)$, then the solution can be obtained by the iterative procedure $x_{i+1}=\phi(x_i)$.

Proof: The only thing we have to show is that ϕ is a mapping from the interval ω to itself. Since

$$1>\alpha\geq\phi'(x)>0$$

$$\therefore\quad 1\geq \frac{\phi(x_t)-\phi(x)}{x_t-x}$$

Therefore, we have

$$\phi(x)\geq x\geq a$$

and

$$a\leq\phi(x)<\infty$$

or $\phi(x)$ is a mapping from ω to ω. From Theorem II, we conclude that the solution x_t can be obtained by the iterative procedure, $x_{i+1}=\phi(x_i)$.

II MODIFIED ITERATIVE PROCEDURE

From Theorem II, the unique solution of $x=\phi(x)$ can be found if $1>\alpha\geq |\phi'(x)|$ for all x in the closed interval ω. In this section, we investigate an iterative procedure for the case where $1>\alpha\geq |\phi'(x)|$ does not hold in the interval ω. Theorem III is given here to provide a theoretical background.

Theorem III: If $\phi(x)$ is a continuous function defined on a closed interval ω, $\phi'(x)$ is continuous, and $1<m\leqq\phi'(x)\leqq M$ for all $x\epsilon\omega$, and if there exists a solution to the equation, $x=\phi(x)$, in the same interval, then there exists an $A\neq 0$ such that the sequence of x_i provided iteratively by

$$x_{i+1}=x_i+\frac{\phi(x_i)-x_i}{A} \qquad (1)$$

converges to the solution x_t of the equation $x=\phi(x)$.

Proof: Since $\min_{x\epsilon\omega} |\phi'(x)|>1$ and $\phi'(x)$ is continuous on ω, $\phi'(x)$ will not change its sign on ω, otherwise $\min_{x\epsilon\omega} |\phi'(x)|=0$, which is contrary to the hypothesis. Let us assume that $\phi'(x)$ is positive on ω. Then

$$M\geqq\phi'(x)\geqq m>1 \qquad \text{for all } x\epsilon\omega.$$

First, we shall show that there exists an A such that

$$\max_{x\epsilon\omega}\left|\frac{d}{dx}\left(x+\frac{\phi(x)-x}{A}\right)\right|<1,$$

but
$$\frac{d}{dx}\left(x+\frac{\phi(x)-x}{A}\right)=1+\frac{\phi'(x)-1}{A}.$$

Let $A=-(M-1)$, then

$$\max_{x\epsilon\omega}\left|1+\frac{\phi'(x)-1}{A}\right|=\max_{x\epsilon\omega}\left(1-\frac{\phi'(x)-1}{M-1}\right)=1-\min_{x\epsilon\omega}\left(\frac{\phi'(x)-1}{M-1}\right)$$

$$=1-\frac{m-1}{M-1}=\frac{M-m}{M-1}<1, \text{ for } m>1$$

Secondly, we shall show that the function

$$\phi^*(x)=x+\frac{\phi(x)-x}{A}=x-\frac{\phi(x)-x}{M-1} \qquad (2)$$

is a mapping from the interval ω to itself. Let the solution of $x=\phi(x)$ be x_t, and $a\leq x_t\leq b$, where ω is [a, b]. The minimum value of the derivative of the mapping (2) is greater than or equal to zero, therefore the derivative of the function (2) is positive and less than one. And we have

(1) $a\leq x<\phi^*(x)\leq\phi^*(x_t)$ for $a\leq x\leq x_t$,

(2) $\phi^*(x_t) \leq \phi^*(x) < x \leq b$ for $x_t \leq x \leq b$.

we shall prove the first inequality, the second inequality can be proved in the similar way. For $a \leq x \leq x_t$,

$$\frac{\phi(x_t)-\phi(x)}{x_t-x}=\phi'(x_j)>1, \text{ for } x \leq x_j \leq x_t.$$

$$\therefore \quad x>\phi(x), \text{ because } x_t=\phi(x_t).$$

Therefore, we have

$$\phi^*(x)=x-\frac{\phi(x)-x}{M-1}=\frac{Mx-\phi(x)}{M-1}>\frac{Mx-x}{M-1}=x, \text{ or } \phi^*(x)>x. \quad (3)$$

In addition, we have

$$M \geq \frac{\phi(x_t)-\phi(x)}{x_t-x}=\phi'(x_k), \text{ for } x \leq x_k < x_t,$$

or

$$Mx-\phi(x) \leq Mx_t-\phi(x_t).$$

Therefore, we have

$$\phi^*(x)=\frac{Mx-\phi(x)}{M-1} \leq \frac{Mx_t-\phi(x_t)}{M-1} \text{ or } \phi^*(x) \leq \phi^*(x_t). \quad (4)$$

From Eqs. (3) and (4), we conclude that

$$a \leq x < \phi^*(x) \leq \phi^*(x_t) \text{ for } a \leq x \leq x_t.$$

Consequently, from the above two inequalities and from the assumption that the function $\phi(x)$ is continuous in the interval ω, the function (2) is a mapping from ω to ω itself.

Consequently by using Theorem II, the iterative procedure (1) provides a sequence that converges to the solution x_t of $x=\phi(x)$. For the other case where $\phi'(x)$ is negative, the theorem may be proved in a similar way.

The geometrical meaning of the transformation $\phi^*(x)=x+\frac{\phi(x)-x}{A}$ will reveal some intuitive notion about Theorem V-3. Referring to Fig. 1, two lines, $y=x$ and

$$-1=\frac{y-\phi_1(x_t)}{x-x_t},$$

are shown. There are two curves $y=\phi_1(x)$ and $y=\phi_2(x)$, having the same root x_t, $|\phi_1'(x)|<1$ and $|\phi_2'(x)|<1$ in the region of consideration, i. e., region I as shown in the same diagram. Suppose that x_{10} and x_{20} are two initial values taken as approximate solutions of the equations $x=\phi_1(x)$ and $x=\phi_2(x)$ respectively. The theory of contraction mapping states that if the absolute values of $\phi'_1(x)$ and $\phi'_2(x)$ are less than one in the region under consideration, then both of the iterative procedures

$$x_{1,i+1}=\phi_1(x_{1,i}),$$

$$x_{2,i+1}=\phi_2(x_{2,i}),$$

converge to the solution point x_t of the equations $x=\phi_1(x)$ and $x=\phi_2(x)$ respectively.

The solution sequences x_{10}, x_{11}, $x_{12}, \cdots$ and x_{20}, x_{21}, $x_{22}, \cdots$ are shown in the same diagram, demonstrating the convergence.

On the other hand, if a function $\phi(x)$ is defined in the region II instead of in the region I, then the iterative procedure

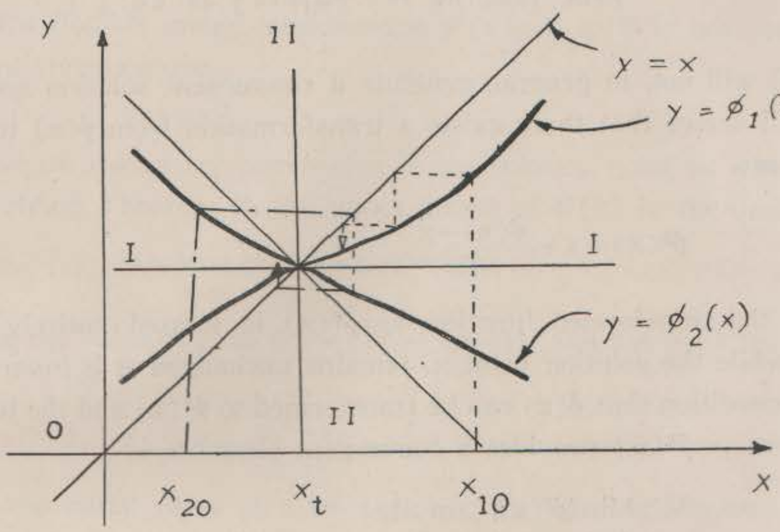

Figure 1

Plot showing the convergent processes.

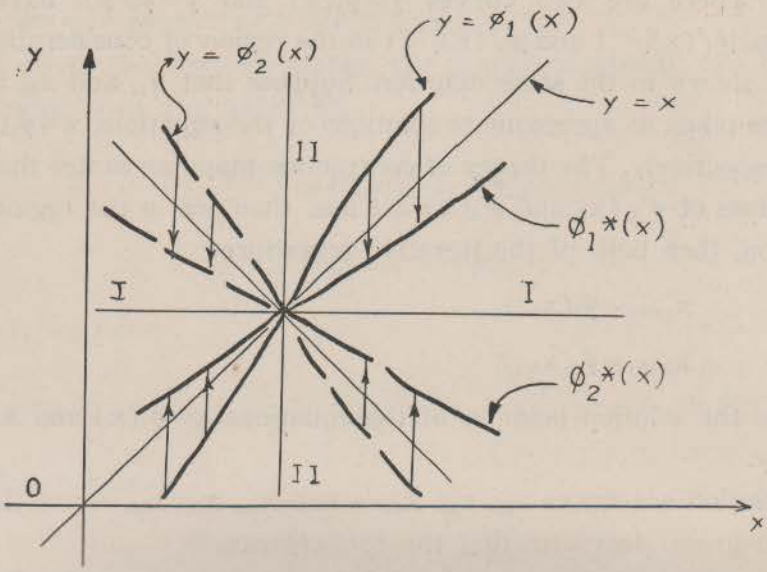

Figure 2

Plot showing the mapping of Eq. (1.)

$x_{i+1}=\phi(x_i)$ will not, in general, generate a convergent solution sequence. Theorem II shows that there exists a transformation from $\phi(x)$ to $\phi^*(x)$ of the form

$$\phi^*(x)=x+\frac{\phi(x)-x}{A}$$

such that the transformed function $y=\phi^*(x)$ is located entirely in the region I, while the solution point x_t remains unchanged or is invariant. A sufficient condition that $\phi(x)$ can be transformed to $\phi^*(x)$ and the iterative procedure $x_{i+1}=\phi^*(x_i)$ provides a convergent sequence is

$$M\geq\underset{x\epsilon\omega}{\mathrm{Min}}\,|\phi'(x)|\geq m>1,$$

where M and m are the upper bound and the lower bound of $|\phi'(x)|$ in the region under consideration.

Figure 2 shows the two pairs of functions, $\phi_1(x)$ and $\phi_1^*(x)$, and $\phi_2(x)$ and $\phi_2^*(x)$. The arrows show the direction of the transformations.

If the function $\phi(x)$ does not satisfy either conditions, $\max_{x\epsilon\omega} |\phi'(x)| < 1$ or $M \geq \min_{x\epsilon\omega} |\phi'(x)| \geq m > 1$, but $|\phi'(x)|$ is still bounded above, then it is necessary to confine the domain just in the neighborhood of x_t as ω_s, such that either $\max_{x\epsilon\omega_s} |\phi'(x)| < 1$ or $M \geq \min_{x\epsilon\omega_s} |\phi'(x)| \geq m > 1$ is satisfied. And the initial value $x = x_0$ must be chosen inside the domain ω_s instead of the original domain ω.

III EXAMPLES

In this section, two examples are given to illustrate the theorem developed in the last section. The first example is an algebraic equation, and the second example involves the logarithmic function.

(a) The equation is $x = x^3$. The region under consideration is ω: $1 \leq x \leq 2$. Compare this equation with the standard form given in the theorem, we see that

$$\phi(x) = x^3,$$

and

$$\phi'(x) = 3x^2.$$

In the domain under consideration $\phi'(x) > 1$, so it is not certain that the iterative procedure

$$x_{i+1} = x_i^3$$

will generate a sequence converging to the solution point x_t which is one. Now applying Theorem II, the upper bound of $\phi'(x)$ in the domain is

$$M = \max_{x\epsilon\omega} |\phi'(x)| = 3 \cdot 2^2 = 12,$$

so choose the value of A as $A = -(12-1) = -11$, and the modified iterative process will be

$$x_{i+1} = x_i + \frac{x_i^3 - x_i}{-11}.$$

Let the initial value of x be $x = x_0 = 2$, the solution sequence is found as follows

$$x_1 = 2 + \frac{8-2}{-11} = 1.454$$

$$x_2 = 1.454 + \frac{(1.454)^3 - 1.454}{-11} = 1.307.$$

The following table shows some data:

i	0,	1,	2,	3,	4,	5,	6,⋯
x_i	2.0,	1.454,	1.307,	1.223,	1.167,	1.129,	1.101,⋯

which converges to the true solution point 1.

(b) The equation is $x=-\log(x)$. The region under consideration is ω: $0<\delta<x<1$, where δ is a small positive number, Comparing this equation with the standard form given in the theorem, we have

$$\phi(x)=-\log(x),$$

and

$$\phi'(x)=-1/x.$$

So in the domain under consideration $|\phi'(x)|>1$. So again it is not certain that the iterative procedure

$$x_{i+1}=-\log(x_i),$$

will generate a sequence converging to the solution point x_t, which is located botween 0.56 and 0.57. Now the upper bound of $\phi'(x)$ in the domain is

$$M=\underset{x\epsilon\omega}{\text{Max}}\,|\phi'(x)|=\frac{1}{\delta}.$$

For $\delta=0.\ 1$, choose $A=12$, and use the iterative procedure

$$x_{i+1}=x_i+\frac{-\log(x_i)-x_i}{12}$$

Let the initial value of x be $x=x_0=1$, the solution sequence is found as follows:

$$x_1=1-\frac{1}{12}=0.91,$$

$$x_2=0.91+\frac{0.90-0.91}{12}=0.84.$$

The following table shows some data:

i	1,	2,	3,	4,	5,	6,	7,⋯
x_i	1.0,	0.917,	0.847,	0.790,	0.744,	0.707,	0.677,⋯

which converges to the true solution x_t, which is located between 0.56 and 0.57.