ARTICLES

A Generic Approach to Iterative Methods

Variations on a theme by Stefan Banach

CLIFFORD H. WAGNER
The Pennsylvania State University
Middletown, PA 17057

Iterative processes usually appear in the undergraduate curriculum as unrelated topics, each having a specific purpose and independent existence. Yet such topics as Newton's method, Picard's method for showing the existence and uniqueness of solutions to initial value problems, the Economic Cobweb Theorem for describing supply and demand equilibria, the convergence theorem for the long-run behavior of regular Markov chains, as well as many other iterative processes, are simply different manifestations of the general theme set forth in the **Contraction Mapping Principle**. The first abstract setting for this principle is credited to Stefan Banach **[2]**, **[17]**, **[27]**, who showed that, under a very general hypothesis, all sequences generated by the repeated evaluation of a distance-decreasing function must converge to a unique fixed point. This convergence is the essence of an iterative technique which can be used in a variety of applications to find an approximate solution, to assert that a unique solution must exist, or to show that a given sequence converges to a known solution. Such applications of the Contraction Mapping Principle are the substance of this article.

A good example of a simple iterative process was shown to me by my five-year-old son Donald: Turn on a scientific calculator and repeatedly press the COS button. On a typical calculator, this algorithm computes iterated values of the calculator function COS(x), degree mode (a discrete rational function that approximates the continuous real function $f(x)=\cos(2\pi x/360)$), and displays successive terms of the sequence

$$\begin{aligned}
x_0 &= 0,\\
x_1 &= 1,\\
x_2 &= .9998476952,\\
x_3 &= .9998477415,\\
x_4 &= .9998477415,\\
&\vdots\\
x_n &= .9998477415 \qquad \text{for } n \geqslant 3.
\end{aligned}$$

This iterative sequence, starting at $x_0=0$ and generated by the function COS(x), converges to a fixed point $x^*=.9998477415$ for which COS(x^*) $=x^*$. In fact, you can choose any starting value you like and, *regardless of your choice for* x_0, the repeated evaluation of the COS function always produces the same fixed point x^* after at most four iterations. The Elementary Contraction Mapping Principle explains this phenomenon.

The Elementary Contraction Mapping Principle and simple fixed points

In general, a **fixed point** of the function $f(x)$ is a value x^* (in both the domain and range of the function) for which $f(x^*) = x^*$. A **contraction** (also called a **contraction mapping**) is a real-valued function $f(x)$ for which there is a constant K (called a **Lipschitz constant**) such that

$$0 \leqslant K < 1 \text{ and } |f(x) - f(y)| \leqslant K|x - y| \tag{1}$$

for all x and y in the domain.

For contractions defined on a closed interval I, the existence of a fixed point implies that the range is contained in I since (1) implies

$$|f(x) - x^*| = |f(x) - f(x^*)| \leqslant |x - x^*|$$

for each x in the interval. The converse of this proposition is not only valid but also more noteworthy.

THEOREM 1 (Elementary Contraction Mapping Principle). *If $f(x)$ is a contraction which maps a closed interval $[a, b]$ into itself, then*

(i) *there is a unique fixed point x^* in the interval $[a, b]$,*
(ii) *every iterative sequence generated by $f(x)$ converges to x^* (that is, x^* is the limit of every sequence $x_0, x_1, x_2, \ldots$ for which x_0 is in $[a, b]$ and $x_n = f(x_{n-1})$ for $n \geqslant 1$), and*
(iii) *if $\{x_n\}$ is any iterative sequence generated by $f(x)$, then*

$$|x_n - x^*| \leqslant \frac{K}{1-K}|x_n - x_{n-1}| \leqslant \frac{K^n}{1-K}|x_1 - x_0|, \tag{2}$$

where K is a Lipschitz constant for the contraction $f(x)$ and n is a positive integer.

The inequalities given in (2) can be used to estimate the error in approximating the fixed point x^* by the terms x_n of an iterative sequence generated by $f(x)$. The Elementary Contraction Mapping Principle (ECMP) is a special case of the General Contraction Mapping Principle (GCMP) given below. Many applications of the ECMP are facilitated by the following corollary of the Mean Value Theorem.

THEOREM 2 (Bounded Derivative Condition). *If the function $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b), and if there is a constant K such that $|f'(x)| \leqslant K < 1$ for $a < x < b$, then $f(x)$ is a contraction with Lipschitz constant K.*

We can apply Theorems 1 and 2 to see why every iterative sequence generated by the function $f(x) = \cos(2\pi x/360)$ discussed earlier converges very rapidly to the unique fixed point $x^* \doteq .9998477415$. We take the interval $I = [-1, 1]$, and the constant $K = 4 \times 10^{-4}$. This K is an acceptable Lipschitz constant for $f(x)$ because for $-1 \leqslant x \leqslant 1$, the maximum magnitude of $f'(x)$ is

$$|f'(1)| = \frac{\pi}{180}\sin\left(\frac{\pi}{180}\right) \doteq .000305.$$

The graph of $f(x)$ is symmetric about the y-axis and monotonically decreasing for $0 \leqslant x \leqslant 1$; therefore, regardless of the starting value x_0 in I, the first three terms of the sequence are bounded by the indicated constants:

$$\begin{gathered} -1 \leqslant x_0 \leqslant 1, \\ f(1) \leqslant x_1 \leqslant f(0), \\ f^2(0) \leqslant x_2 \leqslant f^2(1), \text{ and} \\ f^3(1) \leqslant x_3 \leqslant f^3(0). \end{gathered}$$

In this example, inequality (2) leads to

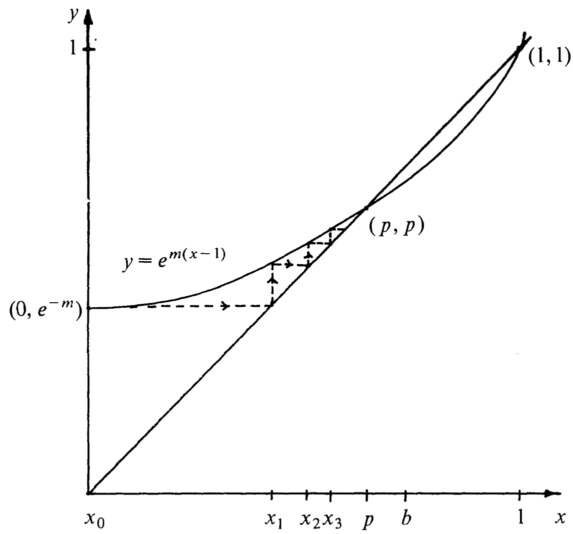

FIGURE 1. **The iterative sequence generated by $f(x) = e^{m(x-1)}$ when $m > 1$.**

$$|x_3 - x^*| \leqslant \frac{.0004}{.9996}|x_3 - x_2| \leqslant \frac{1}{2499}\left[f^3(0) - f^2(0)\right]$$
$$\doteq \frac{.9998477415 - .9998476952}{2499} \doteq 1.85 \times 10^{-11}.$$

Thus x^* is approximated with an accuracy of at least ten significant places by the third term of every iterative sequence which starts in $I = [-1, 1]$ and is generated by $\cos\left(\frac{2\pi x}{360}\right)$. Iterative sequences which start with an x_0 of magnitude greater than 1 must have x_1 in I, and then x_4 approximates x^* with the same accuracy. Similarly, the ten-place calculator function COS(x), degree mode, always reaches its fixed point by the fourth iteration.

Another simple fixed-point problem arises in R. A. Fisher's model [7] for the progeny of a single mutant gene in a cross-pollinated cereal plant. In Fisher's model, the probability for the ultimate extinction of the mutant genes is the smallest positive solution of

$$e^{m(x-1)} = x, \tag{3}$$

where m is the *average* number of offspring per mutant gene [**18**]. When $m \leqslant 1$, ultimate extinction is certain and $x^* = 1$ is the smallest solution of (3). However when $m > 1$, the extinction probability is less than 1 and may be approximated by the iterative method of this section.

When successive generations are increasing in size ($m > 1$), the function $f(x) = e^{m(x-1)}$ has two fixed points in the interval $[0, 1]$. The extinction probability p is the smaller fixed point and the iterative sequence generated by $f(x)$ and starting at $x_0 = 0$ must converge to this p because there is a number b between p and 1 for which $f'(b) < 1$, and then $f(x)$ is a contraction mapping of $[0, b]$ into itself with Lipschitz constant $K = f'(b)$ (see FIGURE 1). For example, Fisher considered the case $m = 1.01$, a rate of increase which is only 1% greater than the rate for exact replacement (on average) of each generation, and estimated the extinction probability to be .9803 by computing the iterative sequence generated by $f(x)$.

Fixed-point problems also occur very naturally in calculus. For example, the critical points of the function $(\sin x)/x$, where $x > 0$, are found by solving the equation $(x \cos x - \sin x)/x^2 = 0$, and this is equivalent to solving $\tan x = x$.

For positive values of x, the fixed points of $f(x) = \tan x$ occur near the discontinuities at $x = 3\pi/2$, $5\pi/2$, etc. (See FIGURE 2). But the tangent function is not contractive in any interval,

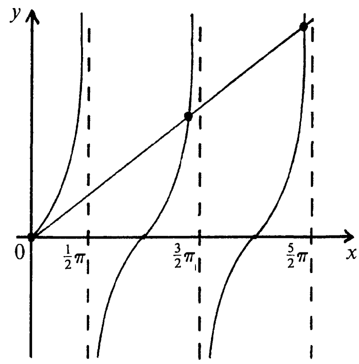

FIGURE 2. **Solutions of $x = \tan x$, with $x > 0$.**

and therefore one cannot apply the Contraction Mapping Principle with $f(x) = \tan x$. This problem is solved more easily by finding the roots of $\tan x - x = 0$. In the next section we describe techniques for finding roots of such equations; these too use the Contraction Mapping Principle in a special way.

Finding roots

A general method for approximating a root of $g(x) = 0$ is to find an auxiliary function $A(x)$ which never vanishes and a closed interval I in such a way that

$$f(x) = x - A(x)g(x)$$

is a contraction mapping from I into itself. Then all iterative sequences generated by $f(x)$ will converge to a value x^* that is both a fixed point of $f(x)$ and a zero of $g(x)$. Newton's method [**5**], [**10**] uses the auxiliary function $A(x) = 1/g'(x)$.

THEOREM 3 (Newton's Method). *If the equation $g(x) = 0$ has a root x^* somewhere in an open interval J where $g'(x)$ and $g''(x)$ are continuous, and where $g'(x)$ never vanishes, then J contains a closed subinterval $I = [a, b]$ such that*

(i) $a < x^* < b$,

(ii) *the function* $f(x) = x - \left[\dfrac{1}{g'(x)}\right] g(x)$ *is a contraction mapping from I into itself, and*

(iii) *the root x^* is the limit of every iterative sequence $x_0, x_1, x_2, \ldots$ which starts in I and is determined recursively by*

$$x_{n+1} = x_n - \left[\frac{1}{g'(x_n)}\right] g(x_n),$$

for $n \geqslant 0$.

Proof. The key here is that $f(x)$ satisfies the Bounded Derivative Condition (THEOREM 2) for some interval containing x^*.

The function $f(x)$ is defined for all x in J and its derivative,

$$f'(x) = \frac{g(x)g''(x)}{[g'(x)]^2},$$

is continuous throughout J with $f'(x^*) = 0$. Consequently, for *any* choice of K, where $0 < K < 1$, there is a closed subinterval $I = [a, b]$ for which $a < x^* < b$ and

$$|f'(x) - f'(x^*)| = |f'(x)| \leqslant K < 1 \text{ for all } x \text{ in } I.$$

Thus $f(x)$ is a contraction when its domain is restricted to I, and since I contains the fixed point, f must map I into itself. The ECMP (Theorem 1) is used to complete the proof.

Although every convergent iterative sequence based on the Contraction Mapping Principle must have at least linear convergence in the sense that a known bound for $|x_{n+1} - x^*|$ is proportional to $|x_n - x^*|$, it is well known that Newton's method actually generates quadratic convergence where the bound for each successive error is proportional to the *square* of the previous error. (Quadratic convergence occurs because $f'(x^*) = 0$ and a Taylor expansion of $f(x)$ about $x = x^*$ gives

$$x_{n+1} = f(x_n) = x^* + 0(x_n - x^*) + \frac{f''(c)}{2}(x_n - x^*)^2$$

for some c in the interval I.)

The next theorem illustrates a simpler iterative method using a constant auxiliary function, $A(x) = 1/M$. This method is useful when linear convergence is sufficient or when $g'(x)$ is not easily evaluated.

THEOREM 4. *If the equation $g(x) = 0$ has a root x^* in a closed interval $I = [a, b]$ where $g(x)$ is continuous on I and where $0 < m \leqslant g'(x) \leqslant M$ for x in (a, b), then $f(x) = x - \frac{1}{M}g(x)$ is a contraction mapping of I into itself and x^* is the limit of every iterative sequence generated by $f(x)$ and starting in I.*

Proof. The proof is similar to that of Theorem 3. In this case, a suitable Lipschitz constant is $K = 1 - m/M$ since

$$0 \leqslant f'(x) = 1 - \frac{g'(x)}{M} \leqslant K.$$

COROLLARY. *In the context of* Theorem 4, *if $g'(x)$ is always negative with $0 < m \leqslant |g'(x)| \leqslant M$ for x in (a, b), then x^* is the limit of every iterative sequence starting in I and generated by $f(x) = x + \frac{1}{M}g(x)$.*

Proof. Apply Theorem 4 to the function $[-g(x)]$.

We can now complete the solution of the fixed-point problem for the function $g(x) = \tan x - x$ and the interval $I = [4.4, 4.5]$ which contains the smallest positive root of $g(x) = 0$. The hypothesis of Theorem 4 is satisfied with $m = 9.504$ and $M = 21.6$ because $g'(x) = \tan^2(x)$ and thus

$$\min\{g'(x) : x \in I\} = \tan^2(4.4) \doteq 9.587 \geqslant m,$$

and

$$\max\{g'(x) : x \in I\} = \tan^2(4.5) \doteq 21.505 \leqslant M.$$

A Lipschitz constant for $f(x) = x - \frac{g(x)}{M}$ is

$$K = 1 - \frac{m}{M} = .56$$

Starting with $x_0 = 4.5$, the iterative sequence has the values

$$x_1 = f(4.5) \doteq 4.49364, \; x_2 = f(x_1) \doteq 4.49342, \text{ etc.}$$

One might suspect that x_2 has an accuracy of at least three significant decimal places; this is in fact so, since

$$|x_2 - x^*| \leqslant \frac{K}{1-K}|x_2 - x_1| \doteq 2.8 \times 10^{-4} < 5 \times 10^{-4}.$$

Thus 4.49342 approximates the smallest positive zero of $\tan x - x$ which is also the smallest positive critical value of $(\sin x)/x$.

By comparison, Newton's method starting with $x_0 = 4.5$ gives $x_1 = 4.49361390$ and $x_2 = 4.49340966$, where x_2 is actually correct to six significant decimal places.

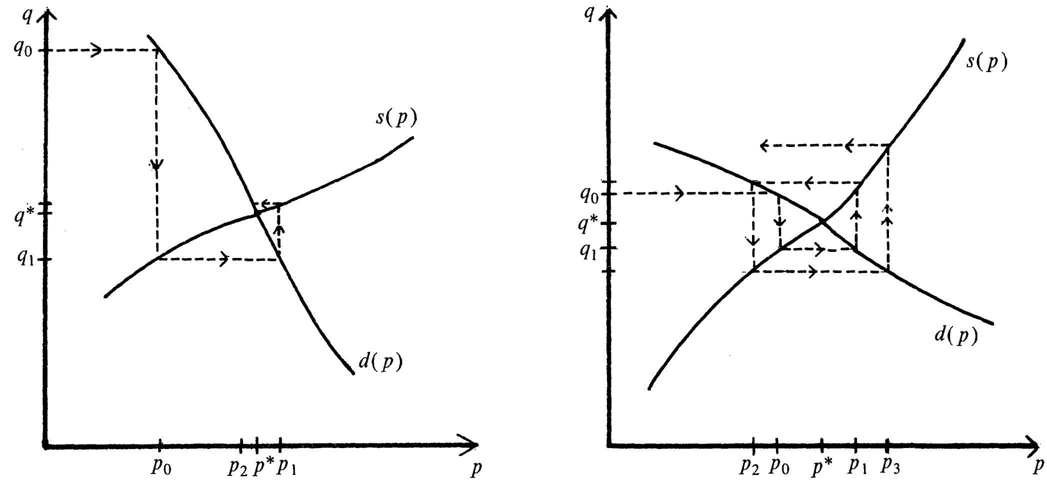

FIGURE 3. **Two economic cobwebs.**

The Cobweb Theorem

In mathematical economics, the market for a particular product often is modelled using a supply function $s(p)$ to represent the total quantity of the product that sellers are willing to supply at a given price level p, and a demand function $d(p)$ to represent the total quantity of the product that buyers are willing to purchase at a given price level p. It is customary to assume that these functions are strictly monotonic and intersect at some equilibrium point (p^*, q^*) where $q^* = d(p^*) = s(p^*)$.

The **cobweb model** [1], [21], [22] concerns a qualitative analysis of markets in which supply adjustments have a time lag and demand adjustments occur with no delay. For example, if an agricultural product is harvested once a year, then supply adjustments which producers make now in reaction to current price levels will affect the quantity to be offered for sale in next year's market. After this lag of one year, competitive market behavior will lead immediately to a new price level consistent with the buyer's demand curve.

Starting with price and quantity levels p_0 and $q_0 = d(p_0)$ the cobweb model determines subsequent levels by the relations

$$q_n = s(p_{n-1}),$$

and

$$p_n = d^{-1}(q_n)$$

for $n = 1, 2, 3, \ldots$. A relevant economics problem is to find conditions for the supply and demand functions to guarantee long term market stability in the sense that the sequence (p_n, q_n) converges to the equilibrium (p^*, q^*). The cobweb graphs of FIGURE 3 illustrate two possible situations.

Convergence occurs if, at each price level p, suppliers' reactions to small price changes, as measured by $|s'(p)|$, are smaller than buyers' reactions as measured by $|d'(p)|$.

THEOREM 5 (The Cobweb Theorem). *If the functions $d(p)$ and $s(p)$ have continuous nonvanishing derivatives throughout a closed interval I, if I contains a value p^* such that $d(p^*) = s(p^*)$ and if*

$$0 < |s'(p)| < |d'(p)|$$

for all p in I, then the cobweb sequences $p_n = d^{-1}(q_n)$ and $q_n = s(p_{n-1})$ converge to p^* and $q^* = d(p^*)$, respectively, for every initial price, p_0, in I.

Proof. Define $f(q) = s[d^{-1}(q)]$ for q in the closed interval $J = d(I)$. The derivative,

$$f'(q) = s'[d^{-1}(q)] \cdot \frac{1}{d'[d^{-1}(q)]},$$

is a continuous function in J, and because $|s'(p)|$ is always less than $|d'(p)|$ there is a constant K such that $0 < |f'(q)| \leqslant K < 1$ for all q in J. This means that $f(q)$ must be a contraction. As observed previously, a contraction mapping with a fixed point maps the domain into itself.

By the ECMP (Theorem 1), the iterative sequence generated by f, starting with $q_0 = d(p_0)$ and satisfying

$$q_n = f(q_{n-1}) = s[d^{-1}(q_{n-1})] = s(p_{n-1})$$

for $n \geqslant 1$, converges to the unique fixed point q^*. And since d^{-1} must be continuous, the sequence, $p_0 = d^{-1}(q_0)$, $p_1 = d^{-1}(q_1)$, etc., must converge to $p^* = d^{-1}(q^*)$.

The General Contraction Mapping Principle

So far, the Elementary Contraction Mapping Principle has been applied to contractions defined on closed intervals. In the more general setting of contractions on complete metric spaces, the General Contraction Mapping Principle (GCMP) brings unity to an even greater variety of iterative methods.

Before stating the GCMP, we recall several definitions. Suppose X is a metric space (with metric function d) and $x_0, x_1, x_2, \ldots$ is a sequence in X; then $\{x_n\colon n = 0, 1, 2, \ldots\}$ is a **Cauchy sequence** if for each $\varepsilon > 0$ there is an integer N such that

$$d(x_n, x_m) < \varepsilon \text{ whenever } n > N \text{ and } m > N.$$

The metric space X is **complete** (with respect to the metric d) if every Cauchy sequence in X converges to a limit (which must be unique). (For example, a closed interval in $\mathbb{R}$ with the usual metric, $d(x, y) = |x - y|$, is a complete metric space.) If the function F maps X into itself and if there is a Lipschitz constant K such that

$$0 \leqslant K < 1 \text{ and } d(F(x), F(y)) \leqslant K d(x, y) \text{ for all } x \text{ and } y \text{ in } X,$$

then F is a **contraction.**

The General Contraction Mapping Principle asserts that a contraction from a complete metric space into itself has a unique fixed point which is the limit of every iterative sequence generated by the contraction. The metric is needed for a general definition of contraction and the completeness property assures the convergence of the iterative sequences.

THEOREM 6 (General Contraction Mapping Principle). *If X is a complete metric space and F is a contraction from X into itself, then*

(i) *there is a unique fixed point x^*, such that $F(x^*) = x^*$, and*
(ii) *for every starting value x_0 in X, the iterative sequence $x_0, x_1, x_2, \ldots$, defined by $x_n = F(x_{n-1})$ for $n \geqslant 1$, converges to x^* with*

$$d(x_n, x^*) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}) \leqslant \frac{K^n}{1-K} d(x_1, x_0),$$

for $n = 1, 2, 3, \ldots$.

A useful corollary of the GCMP concerns functions $F(x)$ for which some iterated power $F^{(k)}(x)$ is a contraction, where $F^{(1)}(x) = F(x)$, and $F^{(n+1)}(x) = F[F^{(n)}(x)]$ for $n = 1, 2, \ldots$.

COROLLARY. *If X is a complete metric space, if F is a function from X into X, and if $F^{(k)}$ is a contraction for some positive integer k, then F has a unique fixed point x^* and every sequence $x_0, x_1, x_2, \ldots$, defined by $x_n = F(x_{n-1})$ for $n \geqslant 1$, converges to x^*.*

Since the proofs of the GCMP and its corollary are somewhat technical, we omit them here and include them in the last section for interested readers. The ECMP (Theorem 1) is a specific interpretation of the GCMP for the case where X is a closed interval of $\mathbb{R}$.

The iterative methods of the next three sections involve contractions defined on: (1) the set of m-dimensional probability vectors with the box metric, (2) the set $\mathbb{R}^m$ of m-dimensional real vectors with the box metric, and (3) the set of continuous functions from one closed interval to another with the sup metric. That each of these is a complete metric space can be proven directly (using the completeness property of the real numbers) or indirectly within the general theory of metric spaces [**12**].

Fixed points for Markov chains

A finite Markov chain with m states is determined by specifying an $m \times m$ transition matrix $\mathbf{P} = (p_{ij})$ where p_{ij} is the probability of passing from state i to state j during any transition period (and where $\sum_{j=1}^{m} p_{ij} = 1$ for each integer i from 1 to m).

To illustrate, suppose the annual marketing goal for the automobile insurance division of Farm States Insurance Company is to retain 80% of its policyholders and to capture 10% of the automobile policies of other companies. The auto insurance market can be modelled as a Markov chain with two states: State 1, the customers of Farm States, and State 2, all other customers. Let the initial probability vector $\mathbf{x}_0 = (.15\ .85)$ indicate that Farm States currently has a 15% market share. Let the transition matrix be

$$\mathbf{P} = \begin{bmatrix} .80 & .20 \\ .10 & .90 \end{bmatrix},$$

and let the auto insurance market after n years be determined by the probability vector $\mathbf{x}_n = \mathbf{x}_{n-1}\mathbf{P} = \mathbf{x}_0\mathbf{P}^n$. For example, the market after one year is given by $\mathbf{x}_1 = (.205\ .795)$ and after two years by $\mathbf{x}_2 = (.2435\ .7565)$. What is the long-term market share projection for Farm States?

A remarkable property of this hypothetical auto insurance example, as well as other so-called finite regular Markov chains, is that the ultimate convergence of the chain (that is, the limit of $\mathbf{x}_0\mathbf{P}^n$ as $n \to \infty$) is totally independent of the initial probability vector $\mathbf{x}_0$. This claim is the essence of Theorem 7, for which we need a few definitions. A Markov chain is **regular** if, for some integer k, every state is reachable from every other state in exactly k transitions, or equivalently if the kth power of the transition matrix has only positive entries [**13**], [**20**]. For any finite Markov chain having m states, the associated **set of probability vectors** is the set $\mathbf{X}$ defined by

$$\mathbf{X} = \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_m) \colon \sum_{i=1}^{m} x_i = 1 \text{ and } x_i \geqslant 0 \text{ for each } i \right\}.$$

The **distance** between probability vectors $\mathbf{x}$ and $\mathbf{y}$ may be given by the **box metric**.

$$d(\mathbf{x},\mathbf{y}) = \max\{|x_i - y_i| \colon i = 1, 2, \ldots, m\}.$$

The set of probability vectors is a complete metric space with respect to the box metric.

THEOREM 7. *Every finite regular Markov chain has a unique probability vector* $\mathbf{x}^*$ *which is a fixed point of its transition matrix* $\mathbf{P}$, *and all iterative sequences* $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots,$ *determined by* $\mathbf{x}_n = \mathbf{x}_{n-1}\mathbf{P}$ *for* $n \geqslant 1$, *converge to* $\mathbf{x}^*$.

Proof. Let $\mathbf{X}$ denote the associated set of probability vectors for the regular Markov chain with transition matrix $\mathbf{P}$. Define $f\colon \mathbf{X} \to \mathbf{X}$ by $f(\mathbf{x}) = \mathbf{x}\mathbf{P}$.

First suppose that $\mathbf{P}$ has no zero entries. A suitable Lipschitz constant is $K = 1 - \varepsilon$, where ε is the smallest entry of $\mathbf{P}$.

For arbitrary probability vectors $\mathbf{x}$ and $\mathbf{y}$ we have $m_0 = \min(x_i - y_i) \leqslant 0$, $M_0 = \max(x_i - y_i) \geqslant 0$, and $d(\mathbf{x},\mathbf{y}) = \max\{-m_0, M_0\}$. And letting z_i represent the ith coordinate of $\mathbf{xP} - \mathbf{yP}$, we have

$$d(\mathbf{xP},\mathbf{yP}) = \max|z_i|. \tag{4}$$

As shown by Kemeny and Snell [**13**, pp. 69–70],

$$z_i \leqslant M_0 - \varepsilon(M_0 - m_0) = M_0(1-\varepsilon) + \varepsilon m_0 = KM_0 + \varepsilon m_0,$$

and similarly,

$$z_i \geqslant Km_0 + \varepsilon M_0$$

for each i. But $\varepsilon m_0 \leqslant 0$ and $\varepsilon M_0 \geqslant 0$, so that

$$Km_0 \leqslant z_i \leqslant KM_0$$

for each i. From (4) then,

$$d(\mathbf{xP},\mathbf{yP}) \leqslant K\max\{-m_0, M_0\} = Kd(\mathbf{x},\mathbf{y}).$$

For this case, the desired result now follows from the General Contraction Mapping Principle.

For the case where $\mathbf{P}$ has zero entries but $\mathbf{P}^k$ does not, the function $g(\mathbf{x}) = f^{(k)}(\mathbf{x}) = \mathbf{xP}^k$ is a contraction, and the desired result follows from the Corollary of the General Contraction Mapping Principle.

Let's determine the ultimate market share for Farm States Insurance Company in the preceding example. Since the Markov chain is finite and regular, all iterative sequences $\mathbf{x}_0, \mathbf{x}_0\mathbf{P}$, $\mathbf{x}_0\mathbf{P}^2, \ldots$ converge to a unique probability vector $\mathbf{x}^*$. Because $\mathbf{x}^*$ must satisfy $\mathbf{x} = \mathbf{xP}$, it can be found directly by solving the following redundant system of equations:

$$\begin{aligned} .8x_1 + .1x_2 &= x_1, \\ .2x_1 + .9x_2 &= x_2, \\ x_1 + x_2 &= 1. \end{aligned}$$

The unique solution is $\mathbf{x}^* = (1/3 \;\; 2/3)$, and the Farm States market share converges to $33\frac{1}{3}$ percent, regardless of the initial vector $\mathbf{x}_0$. Also note that $\mathbf{x}^*$ can be approximated using any iterative sequence $\mathbf{x}_0, \mathbf{x}_0\mathbf{P}, \mathbf{x}_0\mathbf{P}^2, \ldots, \mathbf{x}_0\mathbf{P}^n$.

Systems of linear equations

In standard matrix form, a system of m linear equations with m unknowns is given by $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is an $m \times m$ matrix of constants, and $\mathbf{x}$ and $\mathbf{b}$ are $m \times 1$ column vectors of variables and constants. For example, the system

$$\begin{bmatrix} 4 & -1 & 0 & 0 & 0 & -1 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & -1 & 0 & 0 \\ 0 & 0 & -1 & 4 & -1 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 \\ -1 & 0 & 0 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \\ 20 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

corresponds to a particular finite difference approximation for the following heat distribution problem. A long metal bar with rectangular cross sections of dimensions 3 by 4 is half submerged in water. The water has a constant temperature of 0° and the atmosphere above the water has a constant temperature of 10°. When the distribution of heat within the bar reaches equilibrium, what are the equilibrium temperatures at various points near the middle of the bar?

The finite difference method which leads to the system of equations (5) uses a mesh of sixteen evenly spaced points in a cross section near the middle, with the assumption that cross sections nearby have essentially the same temperature distribution at equilibrium.

Each of the six unknown equilibrium temperatures may be approximated by the average of the equilibrium temperatures at the four closest mesh points in FIGURE 4. For example,

$$x_1 = (x_2 + x_6 + 10 + 10)/4$$

is equivalent to the first equation represented in (5).

The given system of equations is analogous to a continuous Dirichlet problem involving an

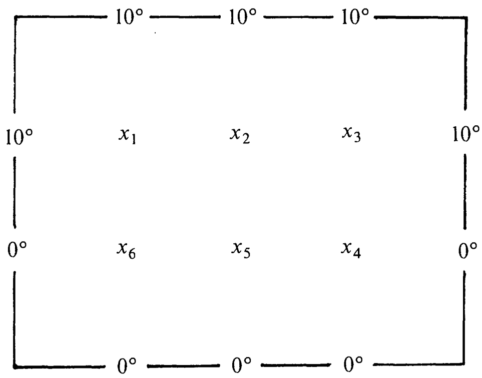

FIGURE 4. **Equilibrium temperatures in a cross section.**

elliptic partial differential equation with boundary conditions [**8**], [**25**]. Although the corresponding continuous solution is obtainable in the case of rectangular regions [**3**], finite difference solutions have wider application.

A common class of linear systems $\mathbf{Ax} = \mathbf{b}$ are those, such as the given system (5), for which the coefficient matrix $\mathbf{A}$ is **strictly diagonally dominant,** which means that

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}}^{m} |a_{ij}|$$

for $i = 1, 2, \ldots, m$. For such systems, there is always exactly one solution. One way of approximating that solution is to generate an iterative sequence using $f(\mathbf{x}) = \mathbf{D}^{-1}[\mathbf{b} - (\mathbf{A} - \mathbf{D})\mathbf{x}]$, where $\mathbf{D}$ is the diagonal matrix

$$\mathbf{D} = \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{mm} \end{bmatrix},$$

and where $\mathbf{x}$ is a column vector in $\mathbb{R}^m$. If we measure the distance between two vectors by the box metric, then the set $\mathbb{R}^m$ becomes a complete metric space.

The function $f(\mathbf{x})$ is not a magical discovery, for the matrix equation $\mathbf{Ax} = \mathbf{b}$ is equivalent to $[(\mathbf{A} - \mathbf{D}) + \mathbf{D}]\mathbf{x} = \mathbf{b}$, from which it follows that a solution $\mathbf{x}$ must satisfy

$$\mathbf{Dx} = \mathbf{b} - (\mathbf{A} - \mathbf{D})\mathbf{x}, \text{ or } \mathbf{x} = \mathbf{D}^{-1}[\mathbf{b} - (\mathbf{A} - \mathbf{D})\mathbf{x}] = f(\mathbf{x}).$$

THEOREM 8 (Jacobi Method). *If* $\mathbf{A}$ *is an* $m \times m$ *strictly diagonally dominant matrix,* $\mathbf{D}$ *is the diagonal matrix* (a_{ii}), *and* $\mathbf{b}$ *is a fixed vector in* $\mathbb{R}^m$, *then*

(i) *the function* $f(\mathbf{x}) = \mathbf{D}^{-1}[\mathbf{b} - \mathbf{Ax} + \mathbf{Dx}]$ *is a contraction from* $\mathbb{R}^m$ *into itself with Lipschitz constant*

$$K = \max\left\{ \frac{1}{|a_{ii}|} \sum_{\substack{j=1\\ j\neq i}}^{m} |a_{ij}| : i = 1, 2, \ldots, m \right\},$$

(ii) *the unique fixed point* $\mathbf{x}^*$ *of* $f(\mathbf{x})$ *is also a unique solution of the system* $\mathbf{Ax} = \mathbf{b}$, *and*

(iii) *every iterative sequence generated by* $f(\mathbf{x})$ *must converge to* $\mathbf{x}^*$.

Proof. The ith coordinate of $f(\mathbf{x}) - f(\mathbf{y})$ has norm

$$\left|\frac{1}{a_{ii}}\left(b_i-\sum_j a_{ij}x_j+a_{ii}x_i\right)-\frac{1}{a_{ii}}\left(b_i-\sum_j a_{ij}y_j+a_{ii}y_i\right)\right|$$
$$=\frac{1}{|a_{ii}|}\left|\sum_{\substack{j=1\\ j\neq i}}^{m} a_{ij}(x_j-y_j)\right|\leqslant\frac{1}{|a_{ii}|}\sum_{\substack{j=1\\ j\neq i}}^{m}|a_{ij}|\,d(\mathbf{x},\mathbf{y})\leqslant Kd(\mathbf{x},\mathbf{y}).$$

Therefore $d[f(\mathbf{x}),f(\mathbf{y})]\leqslant Kd(\mathbf{x},\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$, confirming part (i). The other conclusions follow from the Contraction Mapping Principle and our earlier observation that fixed points of $f(\mathbf{x})$ coincide with solutions of $\mathbf{Ax}=\mathbf{b}$.

A similar, but generally more efficient, method for solving many systems of equations is the Gauss-Seidel method based upon iteratively evaluating the function

$$g(\mathbf{x})=(\mathbf{D}+\mathbf{L})^{-1}(\mathbf{b}-\mathbf{Ux}), \tag{6}$$

where $\mathbf{L}=(\ell_{ij})$ is the **lower part** of $\mathbf{A}$, defined by

$$\ell_{ij}=\begin{cases}a_{ij} & \text{if } i>j,\\ 0 & \text{if } i\leqslant j,\end{cases}$$

and $\mathbf{U}$ is the **upper part** of $\mathbf{A}$, defined by $\mathbf{U}=\mathbf{A}-(\mathbf{D}+\mathbf{L})$. Since $\mathbf{A}=(\mathbf{D}+\mathbf{L})+\mathbf{U}$, the system $\mathbf{Ax}=\mathbf{b}$ is also equivalent to $(\mathbf{D}+\mathbf{L})\mathbf{x}=\mathbf{b}-\mathbf{Ux}$ or $\mathbf{x}=(\mathbf{D}+\mathbf{L})^{-1}(\mathbf{b}-\mathbf{Ux})=g(\mathbf{x})$.

THEOREM 9 (Gauss-Seidel Method). *If* $\mathbf{A}$ *is strictly diagonally dominant and* $g(\mathbf{x})$ *is the mapping* (6), *then* $g(\mathbf{x})$ *is a contraction from* $\mathbb{R}^m$ *into itself with the same Lipschitz constant as the Jacobi function* $f(\mathbf{x})$ (Theorem 8, (i)). *Moreover, every iterative sequence generated by* $g(\mathbf{x})$ *converges to the unique solution of* $\mathbf{Ax}=\mathbf{b}$.

Proof. Let z_i be the ith coordinate of $g(\mathbf{x})-g(\mathbf{y})$. We only need to show that each $|z_i|$ is bounded above by $Kd(\mathbf{x},\mathbf{y})$, since then we have $d[g(\mathbf{x}),g(\mathbf{y})]=\max|z_i|\leqslant Kd(\mathbf{x},\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$. To show that $|z_i|\leqslant Kd(\mathbf{x},\mathbf{y})$, we note that the individual scalar equations represented by the matrix equation

$$(\mathbf{D}+\mathbf{L})[g(\mathbf{x})-g(\mathbf{y})]=\mathbf{U}(\mathbf{y}-\mathbf{x})$$

are each of the form

$$\sum_{j=1}^{i} a_{ij}z_j=\sum_{j=i+1}^{m} a_{ij}(y_j-x_j),\ \text{where } 1\leqslant i\leqslant m.$$

Solving for the term with z_i gives

$$a_{ii}z_i=\sum_{j=1}^{i-1}-a_{ij}z_j+\sum_{j=i+1}^{m}a_{ij}(y_j-x_j). \tag{7}$$

We now proceed by induction on i. For $i=1$, equation (7) leads to

$$|z_1|\leqslant\frac{1}{|a_{11}|}\sum_{j=2}^{m}|a_{1j}||y_j-x_j|\leqslant Kd(\mathbf{x},\mathbf{y}).$$

If we assume that $|z_i|\leqslant Kd(\mathbf{x},\mathbf{y})\leqslant d(\mathbf{x},\mathbf{y})$ for $i=1,2,\ldots,k-1$, then (7) leads to

$$|a_{kk}||z_k|\leqslant\sum_{j=1}^{k-1}|a_{kj}|Kd(\mathbf{x},\mathbf{y})+\sum_{j=k+1}^{m}|a_{kj}|d(\mathbf{x},\mathbf{y})$$
$$\leqslant d(\mathbf{x},\mathbf{y})\sum_{\substack{j=1\\ j\neq k}}^{m}|a_{kj}|.$$

From the definition of K, it follows that $|z_k|\leqslant Kd(\mathbf{x},\mathbf{y})$.

The actual two-place solution of (5) is $x_1 = x_3 = 7.39$, $x_2 = 6.96$, $x_4 = x_6 = 2.61$, and $x_5 = 3.04$. If the iterative sequences start with an initial vector having $x_i = 0$ for each i, then the Jacobi sequence attains an accuracy of two significant decimal places in fifteen steps while the Gauss-Seidel requires only eight steps. Indeed the latter method usually converges more quickly, although this is not indicated by Theorems 8 and 9 where the same Lipschitz constant is given for both $f(\mathbf{x})$ and $g(\mathbf{x})$.

The methods just outlined, as well as the SOR method which generalizes the Gauss-Seidel method, are described more fully in many linear algebra and numerical analysis texts (e.g., **[11]**, **[16]**, **[19]**, **[24]**).

Ordinary differential equations

The standard method for proving the existence and uniqueness of solutions to ordinary differential equations is due to Émile Picard **[3]**, **[14]**, **[23]**. Picard's theorem involves a contraction on the set X of all continuous functions from I into J, where I and J are specified closed intervals of $\mathbb{R}$. If y and z are functions in X, their distance can be given by the **sup metric**

$$d(y, z) = \max\{|y(x) - z(x)|: x \in I\}.$$

The set X is a complete metric space with respect to this metric.

THEOREM 10 (Picard's Existence-Uniqueness Theorem). *If the function $f(x, y)$ and its partial derivative $\partial f/\partial y$ are continuous at all points in a nontrivial rectangle $R = \{(x, y): |x - x_0| \leqslant a$ and $|y - y_0| \leqslant b\}$, then for some nontrivial interval $I = [x_0 - h, x_0 + h]$, there exists a unique real valued function y^* which has domain I and solves the initial value problem*

$$\frac{dy}{dx} = f(x, y) \text{ for all } x \text{ in } I, \text{ and}$$
$$y(x_0) = y_0.$$

Proof. Pick any K in $(0, 1)$; K will be a Lipschitz constant for a contraction which we construct. Let

$$M_1 = \max\{|f(x, y)|: (x, y) \text{ is in } R\},$$

and

$$M_2 = \max\{|\partial f/\partial y|: (x, y) \text{ is in } R\}.$$

Choose an $h > 0$ small enough to satisfy

$$h \leqslant a, \quad hM_1 \leqslant b, \quad \text{and} \quad hM_2 \leqslant K.$$

This "pick and choose" process determines the intervals

$$I = [x_0 - h, x_0 + h], \text{ and } J = [y_0 - b, y_0 + b],$$

and the set of functions $X = \{y: y \text{ is a continuous function from } I \text{ into } J\}$.

Define a mapping from X into itself by $F(y) = F_y$ where for each function y in X, F_y is the function defined by

$$F_y(x) = y_0 + \int_{x_0}^{x} f[t, y(t)]\, dt, \; x \in I.$$

For each y in X, the restrictions on h imply that $F(y)$ is also a continuous function from I into J, and therefore F does indeed map X into itself. Moreover, we can show that the function F is a contraction with respect to the metric for X. If y and z are functions in X, then for each t in I there is some c between $y(t)$ and $z(t)$ for which

$$f[t, y(t)] - f[t, z(t)] = \frac{\partial f(t, c)}{\partial y}[y(t) - z(t)],$$

and therefore,

$$|f[t, y(t)] - f[t, z(t)]| \leqslant M_2|y(t) - z(t)| \leqslant M_2 d(y, z).$$

Bounds for the distance between F_y and F_z involve maxima with respect to all x in I, and

$$\begin{aligned} d[F(y), F(z)] &= \max_x |F_y(x) - F_z(x)| \\ &= \max_x \left| \int_{x_0}^{x} (f[t, y(t)] - f[t, z(t)])\, dt \right| \\ &\leqslant \max_x \left| \int_{x_0}^{x} M_2 d(y, z)\, dt \right| \\ &= M_2 d(y, z) h \leqslant K d(y, z). \end{aligned}$$

Consequently, $F(y)$ has a unique fixed point y^* and every iterative sequence generated by $F(y)$ converges to y^*. Since $F(y)$ is defined in such a way that its fixed points coincide exactly with solutions of the given initial value problem, the theorem follows (use the Fundamental Theorem of Calculus and its converse).

The restriction in Theorem 10 of the solution function $y^*(x)$ to some interval $I = [x_0 - h, x_0 + h]$ is not always binding. For if $h < a$ and $|y^*(x_0 + h) - y_0| < b$, then $y^*(x)$ can be extended (again uniquely) to a larger interval by solving the given differential equation with the new initial condition $y(x_0 + h) = y^*(x_0 + h)$, and then joining y^* together with the new right-hand solution. Repeating this extension process as needed on both the right- and left-hand sides leads to a unique function which not only solves the original initial value problem but also possesses a graph that passes from one boundary edge of R to another.

Conclusion

If the fixed point of a contraction is approximated by an iterative sequence and you wish to improve the accuracy of the approximation, follow the advice of a popular rock and roll lyric: [repeat the iterative process] *one more time* [**6**]. If you wish to view iterative techniques with a wide-angle lens, use the Contraction Mapping Principle. And if you want to learn how two versions of this principle are proved, read the following two proofs.

THEOREM 6 (General Contraction Mapping Theorem [**9**],[**14**],[**17**],[**26**]). *If X is a complete metric space and F is a contraction mapping from X into itself with Lipschitz constant K such that $0 \leqslant K < 1$ and*

$$d(F(x), F(y)) \leqslant K d(x, y) \textit{ for all } x \textit{ and } y \textit{ in } X,$$

then

(i) *there is a unique fixed point x^*, such that $F(x^*) = x^*$, and*
(ii) *for every starting value x_0 in X, the iterative sequence $x_0, x_1, x_2, \ldots$, defined by $x_n = F(x_{n-1})$ for $n \geqslant 1$, converges to x^* with*

$$d(x_n, x^*) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}) \leqslant \frac{K^n}{1-K} d(x_1, x_0),$$

for $n = 1, 2, 3, \ldots$.

Proof. First consider a single iterative sequence.

(a) For all pairs of positive integers i and j with $i < j$,

$$\begin{aligned} d(x_j, x_{j-1}) &= d[F(x_{j-1}), F(x_{j-2})] \leqslant K d(x_{j-1}, x_{j-2}) \\ &\leqslant K^2 d(x_{j-2}, x_{j-3}) \leqslant \cdots \leqslant K^{j-i} d(x_i, x_{i-1}), \end{aligned}$$

that is,

$$d(x_j, x_{j-1}) \leqslant K^{j-i} d(x_i, x_{i-1}). \tag{8}$$

Repeated use of the triangle inequality and inequality (8) leads to a generalization of (8): for positive integers n and m with $n < m$,

$$\begin{aligned} d(x_m, x_n) &\leqslant d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+2}, x_{n+1}) + d(x_{n+1}, x_n) \\ &\leqslant K^{m-n} d(x_n, x_{n-1}) + K^{m-n-1} d(x_n, x_{n-1}) \\ &\quad + \cdots + K^2 d(x_n, x_{n-1}) + K d(x_n, x_{n-1}) \\ &= \frac{K - K^{m-n+1}}{1-K} d(x_n, x_{n-1}) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}), \end{aligned}$$

that is,

$$d(x_m, x_n) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}). \tag{9}$$

Combining (8) and (9), we have

$$d(x_m, x_n) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}) \leqslant \frac{K^n}{1-K} d(x_1, x_0). \tag{10}$$

For n sufficiently large, the bound $\frac{K^n}{1-K} d(x_1, x_0)$ is arbitrarily close to 0, so the iterative sequence is Cauchy and has a unique limit x^*.

(b) The limit x^* must be a fixed point of F because all contractions are continuous, and therefore

$$F(x^*) = F\Big(\lim_{n\to\infty} x_n\Big) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = x^*.$$

(c) For any positive integers n and m with $n < m$, it follows from (10) that

$$d(x^*, x_n) \leqslant d(x^*, x_m) + d(x_m, x_n) \leqslant d(x^*, x_m) + \frac{K}{1-K} d(x_n, x_{n-1}). \tag{11}$$

But $\lim_{m\to\infty} d(x^*, x_m) = 0$, so (11) implies that

$$d(x^*, x_n) \leqslant \frac{K}{1-K} d(x_n, x_{n-1}). \tag{12}$$

The desired bounds for $d(x^*, x_n)$ are given by (12) and (10).

Now consider all possible iterative sequences. As just shown, each such sequence converges to a fixed point. But a contraction can have at most one fixed point since $x^* = F(x^*)$ and $y^* = F(y^*)$ imply that $d(x^*, y^*) \leqslant K d(x^*, y^*)$. Consequently, the unique fixed point x^* is the limit of every iterative sequence.

COROLLARY. (See [4], [**15**], [**17**].) *If X is a complete metric space, if F is a function from X into X, and if $F^{(k)}(x)$ is a contraction for some positive integer k, then F has a unique fixed point x^* and every sequence $x_0, x_1, x_2, \ldots$, defined by $x_n = F(x_{n-1})$ for $n \geqslant 1$, converges to x^*.*

Proof. By Theorem 6 the function $G(x) = F^{(k)}(x)$ has a unique fixed point x^* which is the limit of each of the following sequences:

$$\begin{gathered} x_0, x_k = G(x_0), x_{2k} = G(x_k), x_{3k}, \ldots \\ x_1, x_{k+1} = G(x_1), x_{2k+1} = G(x_{k+1}), x_{3k+1}, \ldots \\ x_2, x_{k+2}, x_{2k+2}, x_{3k+2}, \ldots \\ \vdots \\ x_{k-1}, x_{2k-1}, x_{3k-1}, \ldots . \end{gathered}$$

These sequences are subsequences of $x_0, x_1, x_2, \ldots$ which must then also converge to x^*. That is, every iterative sequence generated by F converges to the unique fixed point of G.

The functions G and F share the same set of fixed points, for if $y^* = F(y^*)$, then $G(y^*) = F^{(k)}(y^*) = y^*$; and if $x^* = G(x^*)$, then

$$F(x^*) = F[G(x^*)] = F^{(k+1)}(x^*) = G[F(x^*)].$$

But G has exactly one fixed point and therefore $F(x^*) = x^*$, and so x^* is also a unique fixed point of F.

Bounds to indicate the rate of convergence in the corollary are given by the inequalities

$$d(x_{nk}, x^*) \leqslant \frac{K}{1-K} d(x_{nk}, x_{nk-k}) \leqslant \frac{K^n}{1-K} d(x_k, x_0).$$

References

[1] R. G. D. Allen, Mathematical Economics, 2nd ed., St. Martin's Press, New York, 1966.

[2] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fund. Math., 3 (1922) 133–181.

[3] M. Braun, Differential Equations and Their Applications, 2nd ed., Springer-Verlag, New York, 1978.

[4] V. W. Bryant, A remark on a fixed-point theorem for iterated mappings, Amer. Math. Monthly, 75 (1968) 399–400.

[5] R. C. Buck, Advanced Calculus, 3rd ed., McGraw-Hill, New York, 1978.

[6] R. Charles, What'd I Say, Atlantic Recording No. 2031, New York, 1959.

[7] R. A. Fisher, The Genetical Theory of Natural Selection, 2nd revised ed., Dover, New York, 1958.

[8] G. E. Forsythe and C. B. Moler, Computer Solution of Linear Algebraic Systems, Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

[9] C. Goffman, Preliminaries to functional analysis, Studies in Modern Analysis, R. C. Buck, ed., Math. Assoc. Amer., Washington, D.C., 1962, pp. 138–180.

[10] P. Henrici, Elements of Numerical Analysis, Wiley, New York, 1964.

[11] E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.

[12] I. Kaplansky, Set Theory and Metric Spaces, 2nd ed., Chelsea, New York, 1977.

[13] J. G. Kemeny and J. L. Snell, Finite Markov Chains, Springer-Verlag, New York, 1976.

[14] A. N. Kolmogorov and S. V. Fomin, Elements of the Theory of Functions and Functional Analysis, vol. 1, Graylock, Rochester, New York, 1957.

[15] I. I. Kolodner, On the proof of the contractive mapping theorem, Amer. Math. Monthly, 74 (1967) 1212–1213.

[16] B. Noble and J. W. Daniel, Applied Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1977.

[17] J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[18] J. H. Pollard, Mathematical Models for the Growth of Human Populations, Cambridge Univ. Press, New York, 1973.

[19] A. Ralston and P. Rabinowitz, A First Course in Numerical Analysis, 2nd ed., McGraw-Hill, New York, 1978.

[20] F. S. Roberts, Discrete Mathematical Models, Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[21] P. A. Samuelson, Foundations of Economic Analysis, Atheneum, New York, 1965.

[22] P. A. Samuelson, Economics, 11th ed., McGraw-Hill, New York, 1980.

[23] G. F. Simmons, Differential Equations with Applications and Historical Notes, McGraw-Hill, New York, 1972.

[24] G. Strang, Linear Algebra and Its Applications, 2nd ed., Academic Press, New York, 1980.

[25] R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[26] C. O. Wilde, The contraction mapping principle, (UMAP Unit 326), Birkhäuser Boston, Cambridge, Massachusetts, 1978.

[27] A. Wouk, Direct iteration, existence and uniqueness, Nonlinear Integral Equations, P. M. Anselone, ed., Univ. of Wisconsin Press, Madison, Wisconsin, 1964, pp. 3–34.