1-1-1986

The contraction mapping principle

Elaine Rosenbloom

THE CONTRACTION MAPPING PRINCIPLE

Elaine Rosenbloom
Honors Capstone
May 9, 1986

THE CONTRACTION MAPPING PRINCIPLE

In the world of mathematics, it is often necessary to approximate the solution to a problem by using an iterative method. One such method, not totally dependant on the initial guess, is the contraction mapping method. This paper will explore the contraction mapping principle for the real line, some extensions of it, and the principle in other contexts.

Before we begin the discussion of the contraction mapping principle for the real line, let's first define the term contraction.

<u>Definition</u> Let I be a real interval $[a,b]$, and let f be a function defined on I into I. Then f is called a <u>contraction on I</u> if there exists a constant r such that for all x,t in I we have $/f(x)-f(t)/\leq r/x-t/$ with $0<r<1$.

Some examples of contraction mappings are:

1) $f(x)=\cos x$ on $[0, \pi/2]$
2) $f(x)=\frac{1}{4}x+3$ on the whole real line.

<u>Remark</u> A contraction mapping is continuous at each point in I. To see this, we apply the definition of continuity to the contraction: For $\varepsilon>0$, we must produce a $\delta>0$ such that $/f(x)-f(t)/<\varepsilon$ when $/x-t/<\delta$ and x,t in I. Now, since f is a contraction, $/f(x)-f(t)/\leq r/x-t/$, $0<r<1$ and x,t in I, which is $<\varepsilon$ when $\delta=\varepsilon/r$. So f~~(x)~~ is continuous at all points in I when f is a contraction.

The contraction mapping theorem now follows:

<u>Theorem</u> Let I be an interval defined on the real numbers and let f be a real-valued function defined on I into I. If f is a contraction on I, then the equation $f(x)=x$ has a unique solution in I. Furthermore, this solution may be obtained by choosing any point x_0 in I and forming the sequence $\{x_n\}=\{f(x_{n-1})\}$ for n=1,2,... This sequence converges to the solution x.

Proof Suppose I= $[a,b]$ for f defined on I into I, and that f is a contraction. Since f is a contraction, there exists a number r with $0<r<1$ such that $|f(x)-f(t)|\leq r|x-t|$ for all x,t in I. Let x_0 be a point in I and define a sequence $\{x_n\}$ such that $x_n=f(x_{n-1})$. Note that x_n is an element of I for all n.

$|x_2-x_1|=|f(x_1)-f(x_0)|\leq r|x_1-x_0|$

$|x_3-x_2|=|f(x_2)-f(x_1)|\leq r|x_2-x_1|\leq r^2|x_1-x_0|$

$|x_4-x_3|=|f(x_3)-f(x_2)|\leq r|x_3-x_2|\leq r^2|x_2-x_1|\leq r^3|x_1-x_0|$ and so on.

By induction, $|x_{n+1}-x_n|\leq r^n|x_1-x_0|$. So for all positive integers n and k,

$|x_{n+k}-x_n|=|x_{n+k}-x_{n+k-1}+x_{n+k-1}-x_{n+k-2}+x_{n+k-2}-\dots-x_{n+2}+x_{n+2}-x_{n+1}+x_{n+1}-x_n|$

$\leq|x_{n+k}-x_{n+k-1}|+|x_{n+k-1}-x_{n+k-2}|+\dots+|x_{n+2}-x_{n+1}|+|x_{n+1}-x_n|$ (by triangle inequality)

$\leq r^{n+k-1}|x_1-x_0|+r^{n+k-2}|x_1-x_0|+\dots+r^{n+1}|x_1-x_0|+r^n|x_1-x_0|$

$=(r^{n+k-1}+r^{n+k-2}+\dots+r^{n+1}+r^n)|x_1-x_0|=\sum_{i=n}^{n+k-1} r^i|x_1-x_0|=|x_1-x_0|(S_{n+k-1}-S_{n-1})$

where $S_m=\sum_{i=1}^{m} r^i$=partial sums for a geometric series since $r<1$. Since $\{S_m\}$ is Cauchy (since the geometric series converges), the sequence $\{x_n\}$ is also Cauchy and hence converges to say x. Since f is a contraction, it is continuous. So $f(x)=f(\lim_{n\to\infty} x_n)=\lim_{n\to\infty} f(x_n)=\lim_{n\to\infty} x_{n+1}=x$. Since f(x)=x, x is a fixed point of f. To show uniqueness, let x' be another fixed point of f in I. Then, $|x-x'|=|f(x)-f(x')|\leq r|x-x'|$. So $(1-r)|x-x'|\leq 0$. But $(1-r)>0$. Therefore, $0\leq(1-r)|x-x'|\leq 0$. So by the squeeze law, $(1-r)|x-x'|=0$, so $|x-x'|=0$ (since r cannot equal one) and x=x'. Hence the fixed point is unique.

Now let's apply the theorem to the examples given:

Example 1 f(x)=cos x on $[0,\pi/2]$ Beginning with $x_0=0$, the contraction mapping principle gives the following values:

x	formula for f(x)	approx. x
$x_0=0$	cos (0)	$1=x_1$
$x_1=1$	cos (1)	$.5403=x_2$
$x_2=.5403$	cos (.5403)	.8576
.8576	cos (.8576)	.6543
and so on until		
$x_{23}=.7391$	cos (.7391)	.7391

Hence, to four decimal places, the fixed point solution to cos x=x is x=.7391

Example 2 $f(x)=\frac{1}{4}x+3$ on the whole real line. Taking $x_0=-5$ yields these values:

x	f(x)
-5	1.75
1.75	3.4375
3.4375	3.859375
3.859375	3.9648438
3.9648438	3.9912109
3.9912109	3.9978027
3.9978027	3.9994507
3.9994507	3.9998627
3.9998627	3.9999657
3.9999657	3.9999914
3.9999914	3.9999979
3.9999979	3.9999995
3.9999995	3.9999999
3.9999999	4.0000000
4.0000000	4.0000000

So after 14 iterations, the fixed point solution to $\frac{1}{4}x+3=x$ is x=4

Two extensions of the contraction mapping principle follow.

First, let's define the term expansion.

Definition A function f defined on $I=[a,b]$ is an expansion if for all x,t in I, $|f(x)-f(t)| \geq k|x-t|$ for $k>1$.

An example of an expansion can be seen by taking $f(x)=2x-1$, defined on the whole real line. Note that f(x) is not a contraction since $|f(x)-f(t)|=|2x-1-(2t-1)|=2|x-t|$, and 2 is not between 0 and 1. However, it is an expansion with k=2 which is greater than one.

Theorem For f defined on $I=[a,b]$, such that the range of f includes I, if f is an expansion, then the following hold:

i) f is 1 to 1 so f^{-1} is defined on R

ii) f^{-1} maps R into I and I is contained in R so f^{-1} maps R into R

iii) f^{-1} is a contraction on R with $r=1/k$.

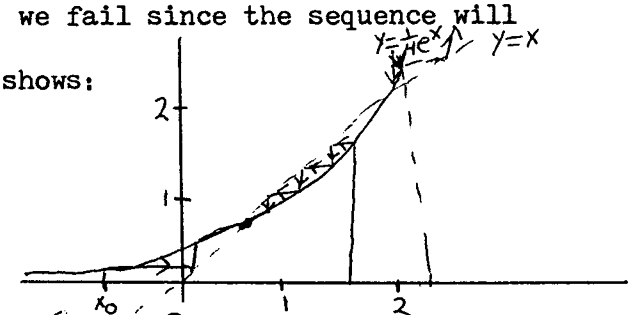
Hence there's a unique x in R such that $f^{-1}(x)=x$. Since the graph of f and f^{-1} are symmetric in the line y=x, x must be in I and f(x)=x.

Using this theorem with the example given, the inverse of f is also defined on the whole real line and is given by $f^{-1}(x)= \frac{1}{2}x+\frac{1}{2}$. $f^{-1}(x)$ is a contraction on I since $/f^{-1}(x)-f^{-1}(t)/=/\frac{1}{2}x+\frac{1}{2}-(\frac{1}{2}t+\frac{1}{2})/=\frac{1}{2}/x-t/$ for x,t in I and $r=\frac{1}{2}=1/k$. We can now apply the contraction mapping principle to $f^{-1}(x)$ and get x=1 as the solution to the original problem f(x)=2x-1.

A second extension of the contraction mapping principle is found using Picard's iteration.

<u>Theorem</u> For f defined on I into I, where I= a,b , let $f_1(x)=f(x)$, $f_2(x)=f(f_1(x))$, $f_3(x)=f(f_2(x))=f(f(f_1(x)))$, etc. I there exists a positive integer n such that f_n is a contraction on I, then the fixed point of f_n is also the fixed point of f in I, and the fixed point can be found by choosing a point x_0 and iterating the function f, even though f is not a contraction.

An example of a function which is not a contraction, but for some iterate is, is given by $f(x)=\frac{1}{4}e^x$. This has two fixed point solutions: one between 0 and 1, and one greater than 2. If we take our initial x_0 between the solutions in an attempt to reach the larger solution, we fail since the sequence will not converge to the solution as this graph shows:



But by iterating f, f_4 is a contraction on $I=\{x: x\leq 2\}$ with r=.9.

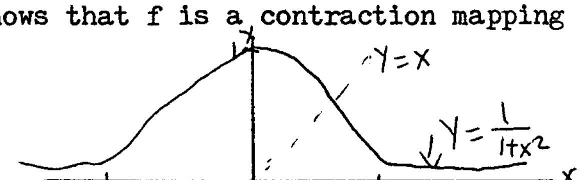
Two practical problems of the contraction mapping principle are now given[1].

<u>Example 3</u> The function f defined on $[0,3\pi]$ by f(x)=tan x. When a beam of light passes through a narrow slit, it spreads out in the shadow regions. This effect is called diffraction. By an application of basic principles of optics, the light intensities on the screen can be expressed in the form

$I=A_0^2 \sin^2 B/B^2$ where B is a suitably chosen spatial variable. The quantity $A=A_0 \sin B/B$ is called the amplitude of the vibration. A problem of interest in optics is to determine the location of the maximum intensities, thus we wish to optimize the function $A(B)=A_0 \sin B/B$. By taking the derivative and finding the critical points, we have $A_0(B\cos B-\sin B)/B^2=0$. The critical values can be obtained by solving the equation $\tan B=B$ or $f(x)=\tan x=x$. The first maximum intensity is at x=0. For larger values of x the fixed points will be near odd multiples of pi/2. We find the location of the first non zero intensity using the Picard algorithm applied to the inverse function y=pi+arctan x, starting at $x_0=0$. The resulting sequence of approximations is 0, 3.1416, 4.4042, 4.4891, 4.4934, 4.4934, ... The first fixed point of tan x=x to the right of zero is at x=4.4934.

Example 4 f defined on the whole real line where $\frac{d}{dx}(f(x))^2=4x^3+4x-4$

In submarine location problems it is often important to find the submarine's closest point of approach (CPA) to a sonobuoy in the water. Suppose the sonobuoy is located at $(2,-\frac{1}{2})$ on a rectangular system and that a submarine travels on a parabolic path along the curve $y=x^2$. For any point (x,x^2) on the parabola, the distance to the sonobuoy is $D(x)=(x^4+2x^2-4x+17/4)^{\frac{1}{2}}$. We want to find the critical x that minimizes D, which also minimizes D^2. We proceed and $\frac{d}{dx}D^2=4x^3+4x-4$. So we wish to solve the equation $x^3+x-1=0$, which can be rewritten as a fixed point problem: $1/(1+x^2)=x$. A rough sketch of the graph shows that f is a contraction mapping on the whole line.



So we implement the Picard algorithm, starting at $x_0=0$ and obtain a sequence 0, 1, .5, .8, .6098, .7290, .6530, .7011, .6705, .6899, .6775, .6854, .6804, ..., .6823, .6823,... So the CPA of the submarine to the sonobuoy is approximately the point $(.6823,(.6823)^2)$ or $(.6823,.4656)$.

Now let's turn our attention to the contraction mapping principle in other contexts: E_n, metric spaces, and uses.

In n-dimensional real space E_n, the principle takes on the following form.

<u>Theorem</u> Let F be a function defined in E_n and let D be a closed region such that F maps D into D and F is a contraction (defined below). Then F has a unique fixed point x in D and $x=\lim_{n\to\infty} x_n$ where $x_n=F(x_{n-1})$ for any initial guess x_0 in D.

<u>Definition</u> F defined on E_n into E_n is a <u>contraction</u> if $//F(x)-F(t)// \leq r//x-t//$ where $0<r<1$ and $//a//$ is the length of a.

<u>Proof</u> Instead of doing a proof similar to the one for the one-dimensional case, the proof is as follows.[2] Define f on C into R by $f(x)=//x-T(x)//$, where T(x) is a contraction on C into C and C is a subset of E_n. Note that a zero for f is a fixed point for T. It is easy to show that $/f(x)-f(y)/\leq (1+a)//x-y//$ so f is continuous. If C is bounded, the max/min principle implies the existence of p in C such that f(p) is a minimum. Then $f(p)\leq f(T(p))\leq af(p)$. Since $f(p)\geq 0$ and $a<1$, we have $f(p)=0$.

If C is not bounded, choose q in C and set $C'=\{x \text{ in } C : f(x)\leq f(q)\}$

If x is in C', then $//x-q//\leq //x-T(x)//+//T(x)-T(q)//+//T(q)-q//$ $\leq 2f(q)+a//x-q//$. Hence $//x-q//\leq 2f(q)/(1-a)$, so C' is closed and bounded. From $f(T(p))\leq af(p)$, it follows that T preserves C', and we may proceed as above. Finally, if p,q are in C and are both fixed points of T, then $//p-q//=//T(p)-T(q)//\leq a//p-q//$, and so $//p-q//=0$ and the fixed point is unique.

The contraction mapping theorem may also be formulated in a metric space.[3]

<u>Definition</u> Let X be a non-empty set. A function d defined on XxX into R, i.e. taking pairs of elements of X into real numbers, is called a <u>metric</u> on the set X iff the following conditions hold:

i) $d(x,y) \geq 0$ for every x,y in X

ii) $d(x,y)=0$ iff $x=y$

iii) $d(x,y)=d(y,x)$ for every x,y in X

iv) $d(x,w) \leq d(x,y)+d(y,w)$ for every x,y,w in X.

Definition Let X be a nonempty set, d be a metric on X, and B_d be the metric topology generated by d. Then the topological space (X,B_d) is called a metric space. We shall use the slightly simpler notation (X,d) to denote the metric space.

Definition A metric space (X,d) is said to be complete if every Cauchy sequence in (X,d) is a convergent sequence in (X,d).

Definition Let (X,d) be a metric space and F be defined on X into X. Then F is a contraction if there exists a real number K such that for all x,y in X, $d(F(x),F(y)) \leq Kd(x,y)$ where $0 \leq K < 1$.

An example of a metric space involves the distance between points in the Euclidean plane. Here X=RxR and the distance between points $x=(x_1,x_2)$ and $y=(y_1,y_2)$ can be defined by $d(x,y)=(|x_1-y_1|^2 + |x_2-y_2|^2)^{\frac{1}{2}}$. Clearly, d(x,y) is symmetric and strictly positive. That it also obeys the triangle inequality is the Pythagorean theorem of Euclid's geometry.

Now, having defined the terms needed, let us look at the contraction mapping principle for metric spaces.

Theorem Let (X,d) be a complete metric space and let F be defined on X into X. If F is a contraction mapping, F has a unique fixed point. The proof for this follows along the same lines as in the real line case and is therefore omitted.

Examples of the contraction mapping principle for metric spaces can be seen by taking X=R and $d(x,y)=|x-y|$ for any x,y in X. Now apply the contraction mapping principle for the real line.

Examples of further use for the contraction mapping principle are linear systems, differential equations, integral equations, and non-linear equations in higher dimensions.

The contraction mapping principle can be used to solve a system of linear equations.[4] Consider an arithemetical n-dimensional space. If $x=(x_1,x_2,\dots x_n)$ and $y=(y_1,y_2,\dots y_n)$ we can let $p(x,y)=\max_i/x_i-y_i/$. This metric space M_n is complete. Also consider the operator $y=A(x)$ which is defined by the equations $y_i=\sum_{j=1}^{n} a_{ij}x_j+b_i$ for $i=1,2,\dots n$

So $p(y_1,y_2)=p(A(x_1),A(x_2))=\max_i/\sum_{j=1}^{n} a_{ij}(x_j^{(1)}-x_j^{(2)})/\leqslant\max_i\sum_{j=1}^{n}/a_{ij}//x_j^{(1)}-x_j^{(2)}/\leqslant \max_j/x_j^{(1)}-x_j^{(2)}/ * \max_i\sum_{j=1}^{n}/a_{ij}/=p(x_1,x_2) * \max_i\sum_{j=1}^{n}/a_{ij}/$. Now if we suppose that $\sum_{j=1}^{n}/a_{ij}/\leqslant r<1$ for all i, we prove the applicability of the contraction mapping principle and so the operator $A(x)$ possesses exactly one fixed point. The system of equations $x_i-\sum_{j=1}^{n}a_{ij}x_j=b_i$ for $i=1,2,\dots n$ has exactly one solution $x^{(0)}=(x_1^{(0)},x_2^{(0)},\dots x_n^{(0)})$ for an arbitrary $b=(b_1,b_2,\dots b_n)$. The solution can be found using an iteration method beginning from an arbitrary vector x.

<u>Example 5</u> Let $(a_{ij})=\begin{pmatrix}\frac{1}{4} & \frac{1}{4}\\ .2 & .2\end{pmatrix}$ and let $b=(0,1)$.

$\sum_{j=1}^{n}a_{ij}\leqslant .5<1$, so we may apply the contraction mapping principle to the system to obtain the solution. Starting with $x_0=(.5,1.5)$, we get the following sequence of vectors: (.5,1.5), (.5,1.4), (.475,1.38), (.46375,1.371), (.4586875,1.36695), (.45640938,1.3651275), (.45538422,1.3643074),...,(.45454857,1.36363886)=x_{13}. After 13 iterations, the approximate solution is accurate to 5 decimal places. The actual solution is (.45454545,1.36363636).

There is a use for the contraction mapping principle in differential equations of the type $dy/dx=f(y,x)$ where f is a real-valued continuous function defined on R into R.[5] We want a solution $y(x)$ which satisfies the initial condition $y(x_0)=y_0$. Since f is continuous, $y(x)=y_0+\int_{x_0}^{x}f(y(s),s)ds$ (1.1)

This has a unique solution as follows: Let $w=F(y)$ where $w(t)=y_0+\int_{x_0}^{t} f(y(s),s)ds$. If y is continuous, w is continuous. So F maps D into D where D is the space of real valued continuous functions defined on some interval I containing x_0. So $y(x)$ is a solution of (1.1) if and only if $y=F(y)$, i.e. y is a fixed point of F.

The use for the contraction mapping principle in the integral equations is as follows.[6] The integral equation $u(x)-k\int_a^b K(x,y,u(y))dy=f(x)$ (1.2) has a unique solution u in $L_2[a,b]$ (L_2 is the squared norm) provided that

i) f is in $L_2[a,b]$

ii) K satisfies a Lipschitz conditions with respect to its third argument, $|K(x,y,w_1)-K(x,y,w_2)|\leq N(x,y)|w_1-w_2|$ for all w_1,w_2 where N is square-integrable with $\int_a^b\int_a^b |N(x,y)|^2 dx\,dy=P^2$ say

iii) $K(x,y,0)$ is continuous for x,y in $[a,b]$

iv) $|k|<1/P$

The unique solution has the form $u(x)=f(x)+k\int_a^b K(x,y,u(y))dy$

A use for the contraction mapping principle for non-linear equations in higher dimensions is as follows.[7] Consider the system of equations

$$f''(x)+u(x,f(x))=0 \qquad (1.3)$$
$$f(0)=f(1)=0$$

Assume u is real valued, continuous, and locally Lipschitz in the sense that given any interval $[a,b]$ there is a number m (which may depend on a,b) such that $|u(x,w_1)-u(x,w_2)|\leq m|w_1-w_2|$ for x in $[0,1]$ and w_1,w_2 are in $[a,b]$

<u>Definition</u> A function g in $C^2([0,1])$ is called a lower solution of (1.3) iff

$-g''(x)\leq u(x,g(x)) \quad (0\leq x\leq 1)$

$g(0),g(1)\leq 0$. An <u>upper solution</u> is defined by reversing the inequalities. Suppose there are lower and upper solutions g,v respectively with $g\leq v$, and set $a=\inf_{x \in [0,1]} g(x)$ and $b=\sup_{x \in [0,1]} v(x)$. Then $[a,b]$ is a finite interval in R. Since u is locally Lipschitz there is a real number w such that for each x in $[0,1]$ $u(x,p)+w^2p$ is non-decreasing in p for p in $[a,b]$. Since the differential equation may be rearranged as $-f''(x)+w^2f(x)=u(x,f(x))+w^2f(x)$, the system (1.3) is equivalent to the integral equation f=Af where $Af(x)=\int_0^1 k(x,y)(u(y,f(y))+w^2f(y))dy$ and k is the Green's function for $-d^2/dx^2+w^2$ with the stated boundary conditions. A is monotone, so there is a theorem wich can be applied which guarantees the existence of a fixed point solution of the integral equation in $[g,v]$. Further, if $f_0=g$ and f_n is the solution of the linear equation $-f_n''(x)+w^2f_n(x)=u(x,f_{n-1}(x))+w^2f_{n-1}(x)$ under the boundary condibions $f_n(0)=f_n(1)=0$, then $\{f_n\}$ is a montone increasing sequence which tends to the solution of (1.3).

As a final point, I include some comments as to how to find x for which x=f(x). In the real line, besides equating x_n to $f(x_{n-1})$, there are methods such as the bisection method, the secant method, and Newton's method for calculating the sequence which converges to the solution. Newton's method of calculating $\{x_n\}$ is to choose a starting point x_0 and then apply the algorithm $x_{n+1}=x_n- g(x_n)/g'(x_n)$, where g(x)=0. This converges faster than the iteration procedure would converge to the solution of g(x)+x=x. However, Newton's method requires the extra step of finding g'(x), and the iteration method is easier to implement on a calculator.

In metric spaces, we let $\{x^{(k)}\}$ be defined by $x_i^{(k)}=\sum_{j=1}^{n} a_{ij}x_j^{(k-1)}+b_i$ for i=1,2,...n , as in the case when solving linear systems.
In E_n, we let $\{\vec{x}_k\}$ be defined by $\vec{x}_k=\overrightarrow{F(\vec{x}_{k-1})}$.

In conclusion, the contraction mapping principle has many variations and many uses. It is useful in that, provided the conditions of the theorem are met, it guarantees a unique fixed point in the interval, and also provides a method for finding such a solution.

REFERENCES

[1] UMAP Module # 326 The Contraction Mapping Principle, Carroll O. Wilde

[2] The Contraction Mapping Lemma, Lance D. Drager and Robert L. Foote, The American Math. Monthly, Volume 93 (1976), pp. 52-54

[3] Applied Functional Analysis, J. Tinsley Oden, Prentic-Hall, Inc. New Jersey 1979

[4] Elements of Functional Analysis, L. A. Lusternik and V. J. Sobolev, Hindustan Publishing, Co., India, 1961

[5] Linear Operator Theory in Engineering and Science, A. W. Naylor and G. Sell, Holt Rinehart and Winston Inc., New York, 1971, pp. 126-130.

[6] Applications of Functional Analysis and Operator Theory, V. Hutson and J. S. Pym, Academic Press, London, 1980