

Chapter 3

Contraction Mappings

BANACH'S CONTRACTION-MAPPING PRINCIPLE

Overview. In this chapter we'll study the best-known of all fixed-point theorems: the Banach Contraction-Mapping Principle, which we'll apply to Newton's Method, initial-value problems, and stochastic matrices.

Prerequisites. Undergraduate-level real analysis and linear algebra. The basics of metric spaces: continuity and completeness.

3.1 Contraction Mappings

The theorem we're going to apply to Newton's Method, Initial-Value Problems, and the Internet was proved by the Polish mathematician Stefan Banach as part of his 1922 doctoral dissertation. Although the setting of Banach's theorem is far more general than that of Brouwer's, the restricted nature of the mappings involved makes its proof a lot simpler.

Banach's theorem is set in a *metric space*: a pair (S, d) where S is a set and d is a "metric" on S , i.e., a function $d: S \times S \rightarrow \mathbb{R}_+$ such that for all $x, y, z \in S$

- (m1) $d(x, y) = 0$ iff $x = y$,
- (m2) $d(x, y) = d(y, x)$, and
- (m3) $d(x, z) \leq d(x, y) + d(y, z)$.

The last property is called, for obvious reasons, "the triangle inequality."

Example. Let S be \mathbb{R}^N , or a subset thereof, and take $d(x, y)$ to be the Euclidean distance between x and y : $d(x, y) = \|x - y\|$. Alternatively d could be the distance on \mathbb{R}^N induced in the same way by the *one-norm* introduced in the proof of Theorem 1.6. As we pointed out there, the two metrics are equivalent in that they have the same convergent sequences.

The mappings addressed by Banach's Principle are called *strict contractions*.¹ To say $F: S \rightarrow S$ is one of these means that there is a positive "contraction constant" $c < 1$ for which

$$d(F(x), F(y)) \leq cd(x, y) \quad \forall x, y \in S. \quad (3.1)$$

Clearly every strict contraction is continuous on S .

Definition 3.1. A *Cauchy sequence* in a space with metric d is a sequence (x_n) such that: For each $\varepsilon > 0$ there is a positive integer $N = N(\varepsilon)$ such that $d(x_n, x_m) < \varepsilon$ whenever the indices m and n are larger than N . A *complete metric space* is one in which every Cauchy sequence converges.

Theorem 3.2 (The Banach Contraction-Mapping Principle). *Suppose (S, d) is a complete metric space and $F: S \rightarrow S$ is a strict contraction. Then F has a unique fixed point, and every iterate sequence converges to this point.*

We'll prove this shortly; first, a few comments.

Iterate sequence. Recall that, for a mapping F taking a set S into itself, the *iterate sequence* starting at $x_0 \in S$ is (x_n) where $x_{n+1} = F(x_n)$ for $n = 0, 1, 2, \dots$

Uniqueness. If (S, d) is a metric space on which F is a strict contraction and $p \in S$ is a fixed point of F , then *there can be no other fixed point*.

Proof. If $q \in S$ is also a fixed point of f then

$$d(p, q) = d(F(p), F(q)) \leq cd(p, q).$$

Since $0 < c < 1$ we must have $d(p, q) = 0$, whereupon condition (m1) in the definition of "metric" guarantees that $p = q$. \square

"Non-strict" contractions. If in Theorem (3.1) we merely assume that the contraction constant c is 1, then:

- *Existence* can fail. Example: $F(x) = x + 1$ defined on the real line.
- *Uniqueness* can also fail. Example: the identity map on a metric space with more than one point.

Exercise 3.1 (Necessity of completeness). Give an example of an *incomplete* metric space on which there is a strict contraction with no fixed point.

Fixed points and iterate sequences. We contended in Sect. 1.2 (page 4) that if the iteration of Newton's method for an appropriate function f were to converge, then that limit had to be a root of f (i.e., a fixed point of the Newton function of f). The next result justifies this contention in a far more general setting.

Proposition 3.3. *If (S, d) is a metric space, $F: S \rightarrow S$ is continuous, and x_0 is a point of S for which the iterate sequence $\{x_0, F(x_0), F(F(x_0)), \dots\}$ converges, then the limit of that sequence has to be a fixed point of F .*

¹ These are often just called "contractions"; the terminology here is more in keeping with conventions used in (linear) operator theory.

Proof. Suppose the iterate sequence (x_n) of x_0 converges to $p \in S$, i.e., $\lim_n d(x_n, p) = 0$. Then the continuity of F insures that $x_{n+1} = F(x_n) \rightarrow F(p)$. Also $\lim_n d(x_{n+1}, p) = 0$, i.e., $x_{n+1} \rightarrow p$, so (because limits in metric spaces are unique) $p = F(p)$. \square

If we assume further that F is a strict contraction, then there results a very strong converse.

Proposition 3.4. *Suppose F is a strict contraction on a metric space. If p is a fixed point of F then every iterate sequence converges to p .*

Proof. Let c denote the contraction constant of the mapping F , so $0 < c < 1$ and F satisfies (3.1) above. Fix $x_0 \in S$ and define the iterate sequence (x_n) in the usual way: $x_1 = F(x_0), \dots, x_n = F(x_{n-1}), \dots$. Then

$$d(x_n, p) = d(F(x_{n-1}), F(p)) \leq cd(x_{n-1}, p) \leq \dots \leq c^n d(x_0, p),$$

so $d(x_n, p) \rightarrow 0$ as $n \rightarrow \infty$, i.e., (x_n) converges to p . \square

Exercise 3.2 (Lessons from a simple initial-value problem). For the initial-value problem (IVP) $y' = y$, $y(0) = 1$, write down the integral operator T on $C(\mathbb{R})$ defined on page 4 by Eq. (IE), and compute explicitly the iterate sequence that has $y_0 \equiv 1$ as its initial function. On which intervals $[-a, a]$ does this iterate sequence converge uniformly to a solution of the IVP? For which of these intervals does the Contraction-Mapping Principle guarantee such convergence?

Proof of the Contraction-Mapping Principle. In view of Proposition 3.4 only one strategy will work: fix a point $x_0 \in S$ and prove that its iterate sequence (x_n) converges. By Proposition 3.3 this limit must be a fixed point.

Since our metric space is complete it's enough to show that (x_n) is a Cauchy sequence. To this end, consider a pair of indices $m < n$ and use the triangle inequality to observe that

$$d(x_n, x_m) \leq \sum_{j=m}^{n-1} d(x_{j+1}, x_j).$$

From the strict contractiveness of F :

$$d(x_{j+1}, x_j) = d(F(x_j), F(x_{j-1})) \leq cd(x_j, x_{j-1}) \leq \dots \leq c^j d(x_1, x_0),$$

whereupon (since $c < 1$)

$$d(x_m, x_n) \leq \sum_{j=m}^{n-1} c^j d(x_1, x_0) = d(x_1, x_0) \sum_{j=m}^{\infty} c^j = \frac{d(x_1, x_0)}{1-c} c^m.$$

Now given $\varepsilon > 0$, we may choose N so that $\frac{d(x_1, x_0)}{1-c} c^N < \varepsilon$, which insures, by the above chain of inequalities, that $N \leq m < n \Rightarrow d(x_m, x_n) < \varepsilon$, hence our iterate sequence (x_n) is indeed Cauchy. \square

The Contraction-Mapping Principle seems to be a perfect theorem: easy to prove and widely applicable. However there is a catch: proving a given mapping to be a strict contraction usually requires some work—as you'll see in the next few sections.

3.2 Application: Stochastic Matrices/Google

In Sect. 1.9 we introduced the “Google matrix” G , a stochastic matrix with entries all positive, and observed with the help of the Brouwer Fixed-Point Theorem (a key step in our proof of Perron’s Theorem) that G has an essentially unique positive fixed point whose coordinates rank internet web pages.

There remains, however, the problem of proposing an algorithm for actually *finding* this fixed point. Recall that the application of Brouwer/Perron to the Google matrix ultimately rested on the stochasticity of that matrix, which implied that G (indeed each $N \times N$ stochastic matrix) maps the standard N -simplex Π_N continuously into itself. The positivity of G then guaranteed the uniqueness of its fixed point.

The generalization to \mathbb{R}^N of the “walking-through-rooms” proof of Brouwer’s theorem set out for $N = 2$ in Sect. 2.3 could provide the basis for an algorithm that approximates the desired fixed point. On the other hand, Banach’s theorem has built into it a scheme that—at least in theory—is easily implemented: *Use iterate sequences to approximate fixed points*. However to be certain that this will work we need each positive stochastic matrix to induce a *strict* contraction on its standard simplex (so far we’ve established only “non-strict” contractivity: Exercise 1.5).

Does stochasticity imply strict contractivity? Is this too much to ask? Read on!

Theorem 3.5. *Every $N \times N$, positive, stochastic matrix induces a strict contraction on the standard simplex Π_N , taken in the metric induced by the one-norm.*

Proof. Suppose A is a positive, stochastic, $N \times N$ matrix. We already know (proof of Theorem 1.6) that A takes Π_N into itself. We’re claiming that there exists a positive number c strictly less than 1 such that

$$\|Ax - Ay\|_1 \leq c\|x - y\|_1 \quad (x, y \in \Pi_N). \quad (3.2)$$

Let $a_{i,j}$ denote the matrix A ’s entry in the i -th row and j -th column. Since each of these numbers is positive we may choose a positive number ε that is strictly less than all of them. Since each column of A sums to 1 we know that $N\varepsilon < 1$ (Proof: for j an index, $1 = \sum_i a_{i,j} > N\varepsilon$). Thus we may form the new $N \times N$ matrix B , whose (i, j) -entry is

$$b_{i,j} = \frac{a_{i,j} - \varepsilon}{1 - N\varepsilon}.$$

Clearly B is a positive matrix, and it’s easy to check that B is stochastic. Now

$$A = (1 - N\varepsilon)B + \varepsilon E$$

where E is the $N \times N$ matrix, all of whose entries are 1.

Claim. A satisfies (3.2) with $c = (1 - N\varepsilon)$.

Proof of Claim. Since $N\varepsilon$ lies strictly between 0 and 1, so does c . What makes this argument work is the fact that if $x \in \Pi_N$ then Ex is the vector in \mathbb{R}^N , each of whose

coordinates is the sum of the coordinates of x , namely 1. In particular if x and y belong to Π_N then $Ex = Ey$, whereupon

$$Ax - Ay = c(Bx - By) + \varepsilon(Ex - Ey) = c(Bx - By).$$

By Exercise 1.5, every $N \times N$ stochastic matrix induces, in the 1-norm, a (possibly non-strict) contraction on \mathbb{R}^N , so from the last equation and the linearity of B :

$$\|Ax - Ay\|_1 = \|c(Bx - By)\|_1 = c\|B(x - y)\|_1 \leq c\|x - y\|_1 \quad (x, y \in \Pi_N),$$

which proves the Claim, and with it the theorem. \square

Corollary 3.6. *If A is an $N \times N$ positive stochastic matrix, then its (unique) Perron eigenvector is the limit of the iterate sequence of each initial point $x_0 \in \Pi_N$.*

In particular, the unique ranking of web pages produced by the Google matrix can be computed by iteration. For $x_0 \in \Pi_N$ the iterate sequence of Corollary 3.6 is (x_n) , where

$$x_n = Ax_{n-1} = A^2x_{n-2} = \dots = A^n x_0 \quad (n = 1, 2, \dots).$$

For this reason the approximation scheme of the Corollary is called *power iteration*; it is used widely in numerical linear algebra for eigen-value/vector approximation.

3.3 Application: Newton's Method

Suppose f is a real-valued function defined on a finite, closed interval $[a, b]$ of the real line, and that we know f has a root somewhere in the open interval (a, b) . We're going to use the Contraction-Mapping Principle to show that, under suitable hypotheses on f , Newton's method for each appropriate starting point converges to this root.

More precisely, suppose $f \in C^2(I)$ with f' never zero on I , and suppose f has different signs at the endpoints of I ; say (without loss of generality) $f(a) < 0$ and $f(b) > 0$. Then f has a unique root x^* in the interior (a, b) of I . Under these hypotheses we have

Theorem 3.7. *There exists $\delta > 0$ such that for every x_0 in $[x^* - \delta, x^* + \delta]$, Newton's method with starting point x_0 converges to x^* .*

In other words, under reasonable hypotheses on f : for starting points close enough to a root of f the iterate sequence for the Newton function

$$F(x) = x - \frac{f(x)}{f'(x)} \quad (x \in I)$$

will converge to that root.

Proof. Let M denote the maximum of $|f''(x)|$ as x ranges through I , and let m denote the corresponding minimum of $|f'(x)|$. By the continuity of f'' , and the hypothesis that f' never vanishes on I , we know that M is finite and $m > 0$.

Differentiation of F via the quotient rule yields

$$F'(x) = \frac{f(x)f''(x)}{f'(x)^2} \quad (x \in I)$$

which, along with our bounds on f' and f'' , provides the estimate

$$|F'(x)| \leq \frac{M}{m^2}|f(x)| \quad (x \in I_\delta).$$

Thus, upon shrinking δ enough to insure that

$$|f(x)| \leq \frac{m^2}{2M} \quad \text{for} \quad |x - x^*| < \delta$$

(possible because f is continuous at x^* and takes the value zero there) we see that $|F'(x)| \leq 1/2$ for each $x \in I_\delta = [x^* - \delta, x^* + \delta]$. This estimate on F' does the trick! For starters, if $x, y \in I_\delta$ then, along with the Mean-Value Theorem of differential calculus, it shows that

$$|F(x) - F(y)| = |F'(\bar{x})(x - y)| \leq \frac{1}{2}|x - y| \quad \forall x, y \in I_\delta$$

where on the right-hand side of the equality, \bar{x} lies between x and y . Thus F is a strict contraction on I_δ —once we know F maps that interval into itself. But it does, since the same inequality shows that for each $x \in I_\delta$ (upon recalling that the root x^* of f is a fixed point of F):

$$|F(x) - x^*| = |F(x) - F(x^*)| \leq \frac{1}{2}|x - x^*| \leq \frac{1}{2}\delta < \delta$$

so $F(x) \in I_\delta$, as desired.

Thus Banach's Contraction-Mapping Principle applies to the strict contraction F acting on the complete metric space $I_\delta = [x^* - \delta, x^* + \delta]$, and guarantees that for every starting point in I_δ the corresponding F -iteration sequence converges to the fixed point of F , which must necessarily be the unique root of f in I_δ . \square

In the course of this proof we had to overcome a problem that occurs frequently when one seeks to apply Banach's Principle:

The metric space for which the problem is originally defined is often *not* the one to which you apply Banach's Principle!

For example, the hypotheses of Theorem 3.7 refer to the Newton function F defined on the compact interval (i.e., the complete metric space) I , but the theorem's *proof* depended on cutting this space down to the smaller one I_δ on which F acted as a strict contraction.

We'll see this scenario play out again in the next section, where we'll have to shrink an entire metric space of continuous functions!

3.4 Application: Initial-Value Problems

It's time for a careful treatment of the initial-value problem (IVP) of Sect. 1.3. Recall its form: There is a differential equation plus initial condition

$$y' = f(x, y), \quad y(x_0) = y_0 \quad (\text{IVP})$$

with $(x_0, y_0) \in \mathbb{R}^2$ and f assumed initially to be continuous on all of \mathbb{R}^2 . Here we'll just assume that f is continuous on a closed rectangle $R = I \times H$, where I and H are compact intervals of the real line, I having radius r and center x_0 and H having radius h and center y_0 . Thus R is a compact “ r by h ” rectangle in the plane, centered at the point (x_0, y_0) .

We'll operate in the metric space $C(I)$ consisting of real-valued functions that are continuous on I . In the course of our work we'll need to shrink the radius r of I . To keep the notation simple we'll re-assign the original symbols I , r , and R to the newly shrunken objects, taking care to be sure that what we've accomplished in one setting transfers intact to the new one.

Since continuous functions are bounded on compact sets and attain their maxima thereon, we can define on $C(I)$ the “max-norm”

$$\|u\| = \max_{x \in I} |u(x)| \quad (f \in C(I))$$

and use this to define a metric d by:

$$d(u, v) = \|u - v\| \quad (u, v \in C(I)).$$

In this metric a sequence converges (resp. is Cauchy) if and only if it converges (resp. is Cauchy) *uniformly* on I . A fundamental property of uniform convergence is that every sequence of functions in $C(I)$ that is uniformly Cauchy on I converges uniformly on I to a function in $C(I)$.² Thus the metric space $(C(I), d)$ is *complete*. As in our treatment of Newton's Method, we'll have to find an appropriate subset of $C(I)$ in which to apply Banach's Theorem. We'll break this quest into several steps.

STEP I. $C(I)$ is too large. For (IVP) to make sense for a prospective solution $y = u(x)$ we have to make sure that for every $x \in I$ the point $(x, u(x))$ lies in the domain of the function f on the right-hand side of the differential equation in (IVP). We must therefore restrict attention to functions $u \in C(I)$ having graph $y = u(x)$ contained in R , i.e., for which $|u(x) - y_0| \leq h$ for every $x \in I$. In metric-space language this means that in order for (IVP) to make sense, our prospective solutions must lie in

² See [101, Theorems 7.14 and 7.15, pp. 150–151], for example.

$$\bar{B} = \bar{B}(y_0, h) = \{u \in C(I) : \|u - y_0\| \leq h\},$$

the closed ball in $C(I)$ of radius h , centered at the constant function y_0 .

STEP II. The integral equation. As we observed in Sect. 1.3, a real-valued function y defined on the interval I satisfies IVP if and only if it satisfies the integral equation

$$y(x) = y_0 + \int_{t=x_0}^x f(t, y(t)) dt \quad (x \in I). \quad (\text{IE})$$

The right-hand side of this equation makes sense for every $u \in \bar{B}$, and defines an integral transformation T on $C(I)$ by

$$(Tu)(x) = y_0 + \int_{t=x_0}^x f(t, u(t)) dt \quad (u \in \bar{B}, x \in \mathbb{R}). \quad (\text{IT})$$

By an argument entirely similar to the one used in Sect. 1.3 (pages 4 and 5) to prove that (IVP) is equivalent to the problem of finding a fixed point for (IT), we have

Lemma 3.8. *If $u \in \bar{B}$ then Tu is differentiable on I and $(Tu)'(x) = f(x, u(x))$ for every $x \in I$.*

In particular, T maps \bar{B} into $C(I)$.

STEP III. Insuring that $T(\bar{B}) \subset \bar{B}$. To use the Banach Contraction-Mapping Principle we must at the very least insure that T maps \bar{B} into itself. For the moment, let's continue to assume only that f is continuous on the rectangle R , and set

$$M = \max\{|f(x, y)| : (x, y) \in R\}.$$

Fix this value of M for the rest of the proof. Although we'll allow ourselves to shrink the horizontal dimension of the rectangle R , we won't be changing the value of M .

Lemma 3.9. *For M as above: if we redefine the interval I to have radius $r \leq h/M$ then $T(\bar{B}) \subset \bar{B}$.*

Proof. For $|x - x_0| \leq h/M$ we have for each $u \in \bar{B}$:

$$|Tu(x) - y_0| = \left| \int_{t=x_0}^x f(t, u(t)) dt \right| \leq M|x - x_0| \leq Mh/M = h.$$

Thus redefining I to have radius $\leq h/M$ insures that $\|Tu - y_0\| \leq h$ for each $u \in \bar{B}$, i.e., that T maps \bar{B} into itself. \square

STEP IV. Strict contractivity. So far we've found how to shrink the original interval I so that the closed ball \bar{B} of radius h in $C(I)$ is mapped into itself by the integral operator T . This ball, being a closed subset of the complete metric space $C(I)$, is itself complete in the metric inherited from $C(I)$. However to apply Banach's Principle we need to know that T is a *strict contraction* on \bar{B} . For this we'll assume that the function f , in addition to being continuous on the rectangle R , is also differentiable

there with respect to its second variable, and that this partial derivative (call it f_2) is continuous on R .

Our goal now is to show that T is a strict contraction mapping on \overline{B} for some positive $r \leq h/M$. Then Banach's Contraction-Mapping Principle will guarantee a fixed point for T in \overline{B} , hence a unique solution therein to the integral equation (IE), and therefore to the initial-value problem (IVP) on the interval $I = [x_0 - r, x_0 + r]$. Once done we'll have proved

Theorem 3.10 (The Picard–Lindelöf Theorem). *Suppose $(x_0, y_0) \in \mathbb{R}^2$, U is an open subset of \mathbb{R}^2 that contains (x_0, y_0) , and f is a real-valued function that is continuous on U and has thereon a continuous partial derivative with respect to the second variable. Then the initial-value problem (IVP) has a unique solution on some nontrivial interval centered at x_0 .*

Proof. By the work above we may choose a compact rectangle $R = I \times H$ in U , centered at (x_0, y_0) , such that $T(\overline{B}) \subset \overline{B}$ whenever the length of I is sufficiently small. It remains to see how much further we must shrink I in order to achieve strict contractivity for T on \overline{B} . To this end let $M' := \max\{|f_2(x, y)| : (x, y) \in R\}$, where the compactness of R and the continuity of f_2 on R guarantee that the maximum exists. Note first that if y_1 and y_2 belong to the interval H with $y_1 \leq y_2$ then the Mean-Value Theorem of differential calculus guarantees for each $x \in I$ that

$$|f(x, y_2) - f(x, y_1)| = |f_2(x, \eta)(y_2 - y_1)| \leq M'|y_2 - y_1| \quad (3.3)$$

where on the right-hand side of the equality, η lies between y_1 and y_2 . Thus if u and v are functions in \overline{B} and $x \in I$, we have upon letting $J(x)$ denote the closed interval between x and x_0 :

$$\begin{aligned} |Tu(x) - Tv(x)| &= \left| \int_{J(x)} [f(t, u(t)) - f(t, v(t))] dt \right| \leq \int_{J(x)} |f(t, u(t)) - f(t, v(t))| dt \\ &\leq M' \int_{J(x)} |u(t) - v(t)| dt \leq M' \|u - v\| \cdot \text{length of } J(x) \\ &= M' \|u - v\| \cdot |x - x_0| \leq M' r \|u - v\| \end{aligned}$$

where the second inequality follows from estimate (3.3). Thus

$$\|Tu - Tv\| \leq M' r \|u - v\| \quad (u, v \in \overline{B}),$$

so we can insure that T is a strictly contractive self-map of \overline{B} simply by demanding that, in addition to the restriction $r \leq h/M$ already placed on the radius of I , we insure that r be $< 1/M'$. \square

Note that the proof given above will still work if the differentiability of f in the second variable is replaced by a "Lipschitz condition"

$$|f(x, y_1) - f(x, y_2)| \leq M'|y_2 - y_1| \quad ((x, y_1), (x, y_2) \in R).$$

For initial-value problems, the interval of existence/uniqueness promised us by Banach's Principle could be very small (see Exercise 3.2 for an example of this). There is, however, always a *maximal* such interval, and this interval has the property that *the solution's graph over this interval continues out to the boundary of the region on which the function f is defined and satisfies the Picard–Lindelöf hypotheses*. For details see, e.g., [93, Sect. 2.4].

As an illustration of this phenomenon, consider the simple initial-value problem $y' = a(1 + y^2)$, $y(0) = 0$, where $a > 0$. One checks easily that $y = \tan(ax)$ is a solution for which the maximal interval of existence is $(-\frac{\pi}{2a}, \frac{\pi}{2a})$, and a separation-of-variables argument shows that this is the *only* solution. Thus, even though the right-hand side $f(x, y) = a(1 + y^2)$ of this IVP's differential equation is infinitely differentiable (even real-analytic) *on the entire plane*, the solution exists only on a finite interval, which for large a is very small.

Conclusion: In nonlinear situations, singularities can arise “unexpectedly.”

Notes

Banach's doctoral dissertation. This is [5]; the Contraction-Mapping Principle is Theorem 6 on page 160 of that paper.

Stochastic matrices. The proof that every positive stochastic matrix induces a strict contraction on its standard simplex (Theorem 3.5) is from [66], where the result is attributed to Krylov and Bogoliubov. The same proof is in [20, Sect. 4, pp. 578–9].

We mentioned that the “power iteration” method of Corollary 3.6 works in more generality. For more on this, see, e.g., [117, Lecture 27].

For the Google matrix G , revisited in Sect. 3.2, there is still the issue of its enormous size. A preliminary discussion of how to handle this can be found in [17].

Initial-value problems. The Picard–Lindelöf Theorem originates in Lindelöf's 1894 paper [70], in which he generalizes earlier work of Picard. In our special case the iteration associated with Banach's Principle is often called “Picard Iteration.”

Higher orders, higher dimensions. The restriction of our discussion of initial-value problems to first order differential equations is not as severe as it seems. Consider, for example, the second order problem for an open interval I containing the point x_0 :

$$y'' = f(x, y, y'), \quad y(x_0) = y_0, \quad y'(x_0) = y_1 \quad (x \in I).$$

This problem can be rewritten as: $Y' = F(x, Y)$, $Y(x_0) = Y_0$ for $x \in I$, where $Y = (y, y')$ is a function taking I into \mathbb{R}^2 , $Y_0 = (y_0, y_1)$ is a vector in \mathbb{R}^2 (now thought of as a space of row vectors), and $F(x, Y) = (y', f(x, y, y'))$ maps the original domain of f (a subset of \mathbb{R}^3) into \mathbb{R}^2 .

It's not difficult to check that the proof given above for our original "scalar-valued" IVP works almost *verbatim* in the new setting, with the absolute-value norm on the real line replaced in the higher dimensions by the Euclidean one, thus producing a unique solution for the second order IVP. Of course the idea generalizes readily to initial-value problems of order larger than 2.

Newton's Method again. In a similar vein, our analysis of Newton's Method can be generalized to higher dimensions. Suppose the function f maps some open subset G of \mathbb{R}^N into itself, and that $f(p) = 0$ for some point $p \in G$. If we assume that all first and second order partial derivatives of the components of f are continuous, and that the derivative f' , which is now a linear transformation on \mathbb{R}^N , is nonsingular at every point of G , then, just as in the single-variable case, we can form the "Newton function" $F(x) = x - f'(x)^{-1} f(x)$, where on the right-hand side we see the inverse of the linear transformation $f'(x)$ acting on the vector $f(x)$. A bit more work than before shows that, when restricted to a suitable closed rectangle centered at p , the function F is a strict contraction, so for every point in that rectangle the Newton iteration converges to p .