Chapter 1
From Newton to Google

WHAT ARE FIXED POINTS? WHAT ARE THEY GOOD FOR?

Overview. After setting out the definition of "fixed point" we'll give examples of their role in finding solutions: to equations (Newton's Method), to initial-value problems, and to the problem of ranking internet web pages. After this we'll show how the notion of fixed point arises in set theory, where it provides an easy proof of the Schröder–Bernstein theorem. We'll introduce the famous Brouwer Fixed-Point Theorem, show how it applies to the study of matrices with positive entries, and discuss the application of these results to the internet page-ranking problem.

Prerequisites. Calculus (continuity and fundamental theorem of integral calculus), differential equations (initial-value problems), basic linear algebra, some familiarity with sets and the operations on them.

1.1 What Is a Fixed Point?

Definition (Fixed Point)**.** Suppose f is a map that takes a set S into itself. A *fixed point* of f is just a point $x \in S$ with $f(x) = x$.

A map f can have many fixed points (example: the identity map on a set with many elements) or no fixed points (example: the mapping of "translation-by-one," $x \to x+1$ on the real line).

Exercise 1.1. The fixed points of a function mapping a real interval into itself can be visualized as the x-coordinates of the points at which the function's graph intersects the line $y = x$. Use this idea to help in determining the fixed points possessed by each of the functions $f\colon \mathbb{R} \to \mathbb{R}$ defined below.

(a) $f(x) = \sin x$

(b) $f(x) = x + \sin x$

(c) $f(x) = 2\sin x$

J.H. Shapiro, *A Fixed-Point Farrago*, Universitext,
DOI 10.1007/978-3-319-27978-7_1

1.2 Example: Newton's Method

Suppose for simplicity that f is a differentiable function $\mathbb{R} \to \mathbb{R}$, with derivative f' continuous and never vanishing on $\mathbb{R}$. Consider for f its "Newton function" F, defined by

$$F(x) = x - \frac{f(x)}{f'(x)} \qquad (x \in \mathbb{R}). \tag{1.1}$$

One can think of $F(x)$ as the horizontal coordinate of the point at which the line tangent to the graph of f at the point $(x, f(x))$ intersects the horizontal axis. Since f' doesn't vanish, F is a continuous mapping taking $\mathbb{R}$ into itself. The roots of f (those points $x \in \mathbb{R}$ such that $f(x) = 0$) are precisely the fixed points of F.

Newton's method involves iterating the Newton function in the hope of generating approximations to the roots of f. One starts with an initial guess x_0, sets $x_1 = F(x_0)$, $x_2 = F(x_1)$ ..., and hopes that the resulting sequence of "Newton iterates" converges to a fixed point of F. Geometrically it seems clear that if the Newton iterate sequence converges then it must converge to a root of f. We'll see later, as a consequence of something far more general (Proposition 3.3, page 28), that this indeed the case.

1.3 Example: Initial-Value Problems

From a continuous function $f\colon \mathbb{R}^2 \to \mathbb{R}$ and a point $(x_0, y_0) \in \mathbb{R}^2$ we can create an *initial-value problem*

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

Geometrically, (IVP) asks for a differentiable function y whose graph is a smooth "solution curve" in the plane that has the following properties:

(a) The curve passes through the point (x_0, y_0), and
(b) at each of its points (x, y) the curve has slope $f(x, y)$.

As a first attempt to solve the differential equation $y' = f(x, y)$ one might try integrating both sides with respect to x. If by "integrate both sides" we mean "take the definite integral from x_0 to x," then there results the *integral equation*

$$y(x) = y_0 + \int_{t=x_0}^{x} f(t, y(t))\, dt \tag{IE}$$

which is implied by (IVP) in the sense that each function y satisfying (IVP) for some interval of x's containing x_0, also satisfies (IE) for that same interval.

Conversely, suppose $y \in C(\mathbb{R})$ satisfies (IE) on some open interval I. Fix $x \in I$. Then for $h \in \mathbb{R} \backslash \{0\}$ small enough that $x + h \in I$, the Mean-Value Theorem of integral calculus provides a point ξ between x and $x + h$ such that

$$\frac{y(x+h)-y(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t,y(t))\,dt = f(\xi,y(\xi)).$$

Thanks to the continuity of f, as $h \to 0$ the expression on the right, and therefore the difference quotient on the left, converges to $f(x,y(x))$. Thus y is differentiable at x and $y'(x) = f(x,y(x))$, i.e., the function y satisfies the differential equation in (IVP) on the interval I. That it satisfies the initial condition is trivial.

Conclusion: (IVP) $\equiv$ (IE).

To make the connection with fixed points, let $C(\mathbb{R})$ denote the vector space of continuous, real-valued functions on $\mathbb{R}$, and consider the *integral transform* $T: C(\mathbb{R}) \to C(\mathbb{R})$ defined by

$$(Ty)(x) = y_0 + \int_{t=x_0}^{x} f(t,y(t))\,dt \qquad (x \in \mathbb{R}). \tag{1.2}$$

Thus equation (IE) can thus be rewritten $Ty = y$, so to say $y \in C(\mathbb{R})$ satisfies (IVP) turns out to be the same as saying: *y is a fixed point of the mapping T*. In Chap. 3 we'll discuss the existence and uniqueness of such fixed points.

1.4 Example: The Internet

At each instant of time the publicly accessible Internet consists of a collection of N web pages[1] each of which can have links coming in from, and going out to, other pages. To be effective, search engines such as Google must seek to determine the importance of each individual page. Here's a first attempt to do this.

To each web page P_i $(1 \le i \le N)$ we'll assign a non-negative real number $\mathrm{imp}(P_i)$ that measures the "importance" of that page. The page P_i will derive its importance from all the pages that link into it: if P_j has a total of λ_j outgoing links then we decree that it bestow importance of $\mathrm{imp}(P_j)/\lambda_j$ to each of the pages into which it has links. In other words, if we think of P_j as having $\mathrm{imp}(P_j)$ "votes" then our rule is that it must distribute these votes evenly among the pages into which it links. The importance of a given web page is then defined to be the sum of the importances it receives from each of the web pages that link to it (self-links are allowed).

This definition of "importance" for a web page may seem at first glance to be circular, but it's not! To make matters precise, define L_i to be the set consisting of all indices j for which P_j links into P_i. Then $\mathrm{imp}(P_i)$ is given by the equation

$$\mathrm{imp}(P_i) = \sum_{j \in L_i} \frac{1}{\lambda_j}\mathrm{imp}(P_j) \qquad (1 \le i \le N) \tag{1.3}$$

[1] According to `www.worldwidewebsize.com`, N was $\ge 4.74 \times 10^9$ on August 4, 2015.

This is a set of N linear equations in the N unknowns $\mathrm{imp}(P_i)$; for these equations to provide a reasonable ranking of web pages there needs to be a solution not identically zero, all coordinates of which are non-negative.

To see how fixed points enter into this discussion, let v to be the column vector with $\mathrm{imp}(P_i)$ in the i-th position, and define the "hyperlink matrix" H to be the $N \times N$ matrix whose j-th column has $1/\lambda_j$ in the i-th entry if P_j links to P_i, and zero otherwise. With these definitions Eq. (1.3) can be rewritten in matrix form

$$v = Hv,$$

so the "importance vector" v we seek is a fixed point of the transformation that H induces on the set of vectors in $\mathbb{R}^N \backslash \{0\}$, all of whose coordinates are non-negative.

In the language of linear algebra: we demand that 1 be an eigenvalue of H and v be a corresponding eigenvector *with non-negative entries.* We'll see in Sect. 1.7 that such a vector actually exists. In Sect. 1.8 we'll also take up the crucial question of uniqueness (at least up to positive scalar multiples); to be effective our method needs to produce an *unequivocal* ranking of web pages.

Mini-example. Consider the fictional mini-internet pictured in Fig. 1.1 below which consists of six web pages, the label on each link denoting the proportion of the donor page's importance being granted to the recipient page.

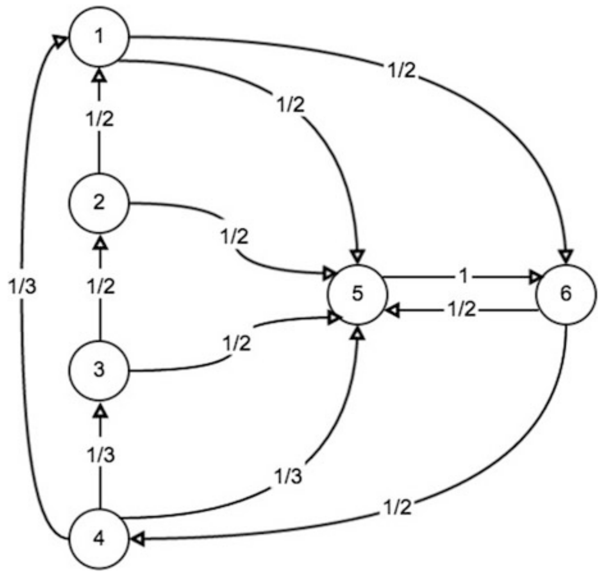

Fig. 1.1 An imaginary six-page internet

The hyperlink matrix for our mini-web is

$$H_0 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

for which your favorite matrix calculation program will verify that 1 is an eigenvalue, so the equation $H_0 v = v$ does have a non-zero solution in $\mathbb{R}^6$. Furthermore, the calculation will show that this solution is unique up to scalar multiples, and all of its entries have the same sign. When normalized to have positive entries and Euclidean norm 1, this vector is, to two significant digits,

$$v = [0.14,\ 0.057,\ 0.11,\ 0.34,\ 0.61,\ 0.69]^t$$

where the superscript "t" denotes "transpose." This gives the following ranking, from most to least important: $(P_6,\ P_5,\ P_4,\ P_1,\ P_3,\ P_2)$.

This example illustrates a critical fact about the process of ranking pages: the page with the most incoming links need not be the most important! In particular, Page 6, with only two incoming links, is more important than Page 5, which has five incoming links. Similarly, Page 4, with just one incoming link, is more important than Page 1, which has two such links (Exercise: Can you explain in just a few words what's making this happen?).

Exercise 1.2. "Importance vectors" need not be unique. Consider the following mini-web, still with six pages. but now with links that look like this:

$$1 \to 2 \to 3 \to 1 \qquad \text{and} \qquad 4 \to 5 \to 6 \to 4.$$

Write out the hyperlink matrix for this mini-web and show that it has several independent importance vectors (some of which contain zeros).

The hyperlink matrix H for the full internet, although huge, consists mostly of zeros; each web page links to a relatively tiny number of others.[2] Furthermore each column of H corresponding to a page with outlinks will sum to 1, but each column corresponding to a page with *no* outlinks is identically zero. This latter kind of page (a particularly annoying one) is called a "dangling node." Were the internet to have no such pages, its hyperlink matrix would be *stochastic*: each entry non-negative and each column summing to one. This is the case for the six-page example worked out above, as well as the mini-web of Exercise 1.2.

[2] According to [12], on average somewhere in the hundreds.

Stochastic matrices have particularly nice properties; we'll show in Sect. 1.7 that their associated linear transformations possess nontrivial fixed points (i.e., 1 is an eigenvalue) *with non-negative entries.* Thus, were our internet to have no dangling nodes, the hyperlink matrix would have a fixed point that would provide a ranking of websites. For this reason we'd like to find a modification of H that achieves "stochasticity" without compromising the intuition behind our definition of "importance."

One way to do this is to think of a dangling node as, rather than linking to *no* other web pages, actually linking to *every* web page (including itself), contributing $1/N$ of its importance to every web page. This models the behavior of a web surfer who, stuck at a page with no outlinks, decides to skip directly to a random page, thus establishing to that page a "link" of weight $1/N$. Our new $N \times N$ hyperlink matrix, call it H_1, is now stochastic; the columns previously identically zero are now identically $1/N$. We'll return to this matrix in Sect. 1.7.

1.5 Example: The Schröder–Bernstein Theorem

The famous Schröder–Bernstein theorem of set theory asserts:

> If X and Y are sets for which there is a one-to-one mapping taking X into Y and a one-to-one mapping taking Y into X, then there is a one-to-one mapping taking X *onto* Y.

This result follows from a more general one depicted in Fig. 1.2 below:

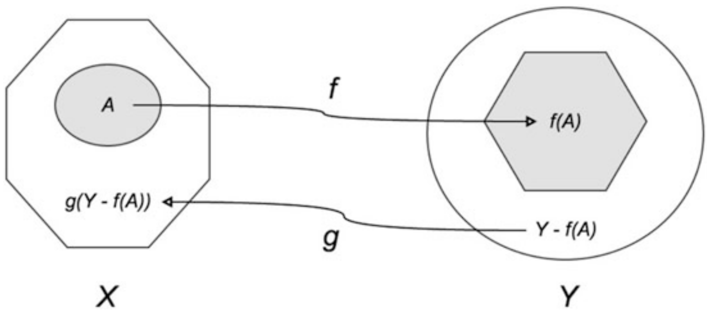

Fig. 1.2 The Banach Mapping Theorem

Theorem 1.1 (The Banach Mapping Theorem)**.** *Given sets X and Y and functions $f: X \to Y$ and $g: Y \to X$, there is a subset A of X whose complement is the g-image of the complement of $f(A)$.*

To see how the Banach Mapping Theorem implies the Schröder–Bernstein Theorem, suppose in the statement of the Banach theorem that the maps f and g are one-to-one. Then the map $h\colon X \to Y$ defined by setting $h = f$ on A and $h = g^{-1}$ on $X \backslash A$ is the one promised by Schröder–Bernstein. □

The Banach Mapping Theorem is, in fact, a fixed-point theorem! Its conclusion is that there is a subset A of X for which $X \backslash A = g(Y \backslash f(A))$. This equation is equivalent, upon complementing both sides in X, to

$$A = X \backslash g(Y \backslash f(A)). \tag{1.4}$$

For a set S, let's write $\mathscr{P}(S)$ for the collection of all subsets of S. Define the function $\Phi\colon \mathscr{P}(X) \to \mathscr{P}(X)$ by

$$\Phi(E) = X \backslash g(Y \backslash f(E)) \qquad (E \in \mathscr{P}(X)).$$

With these definitions, Eq. (1.4) asserts that the set A is a fixed point of Φ.

That such a fixed point exists is not difficult to prove. The mapping Φ defined above is best understood as the composition of four simple set-mappings:

$$\mathscr{P}(X) \xrightarrow{f} \mathscr{P}(Y) \xrightarrow{C_Y} \mathscr{P}(Y) \xrightarrow{g} \mathscr{P}(X) \xrightarrow{C_X} \mathscr{P}(X)$$

where C_X denotes "complement in X" and similarly for C_Y, while f and g now denote the "set functions" induced in the obvious way by the original "point functions" f and g. Since f and g preserve set-containment (i.e., $E \subset F \implies f(E) \subset f(F)$) while C_X and C_Y reverse it, the composite mapping Φ *preserves* set-containment.

With these observations, the theorems of Banach and Schröder–Bernstein follow from:

Theorem 1.2 (The Knaster–Tarski Theorem)**.** *If X is a set and $\Phi\colon \mathscr{P}(X) \to \mathscr{P}(X)$ is a mapping that preserves set-containment, then Φ has a fixed point.*

Proof. Let $\mathscr{E}$ be the collection of subsets E of X for which $E \subset \Phi(E)$. Since $\mathscr{E}$ contains the empty subset of X, it is nonempty. Let A denote the union of all the sets in $\mathscr{E}$.

Claim. $\Phi(A) = A$.

Proof of Claim. Suppose $E \in \mathscr{E}$. Then $E \subset \Phi(E)$ by the definition of $\mathscr{E}$, and $E \subset A$ by the definition of A. Thus $\Phi(E) \subset \Phi(A)$ by the containment-preserving nature of Φ, hence $E \subset \Phi(A)$. Consequently $A \subset \Phi(A)$, whereupon $\Phi(A) \subset \Phi(\Phi(A))$, which places $\Phi(A)$ in $\mathscr{E}$. Conclusion: $\Phi(A) \subset A$, hence $\Phi(A) = A$. □

1.6 The Brouwer Fixed-Point Theorem

The most easily stated—and deepest—of the fixed-point theorems we'll discuss in this book was proved in 1912 by the Dutch mathematician L.E.J. Brouwer. Its initial setting is the closed unit ball B of Euclidean space $\mathbb{R}^N$.

Theorem 1.3 (The Brouwer Fixed-Point Theorem)**.** *Every continuous mapping of B into itself has a fixed point.*

It's easy to see that the result remains true if B is replaced by a homeomorphic image (i.e., a set $G = f(B)$ where f is continuous, one-to-one, with $f^{-1}\colon G \to B$ also continuous).

For $N = 1$ the proof of Brouwer's Theorem is straightforward. In this case f is a continuous function mapping the real interval $[-1,1]$ into itself. We may suppose f doesn't fix either endpoint (otherwise we're done), so $f(-1) > -1$ and $f(1) < 1$. In other words, the value of the continuous function $g(x) = f(x) - x$ is positive at $x = -1$ and negative at $x = 1$. By the Intermediate Value Theorem, g must take the value zero at some point of the interval $(-1,1)$; that point is a fixed point for f.

The proof for $N > 1$ is much more difficult, and there are many different versions. For $N = 2$ we'll prove in the next chapter a famous combinatorial lemma due to Sperner which yields Brouwer's Theorem for that case,[3] and in Chap. 4 we'll prove the full result using methods of "advanced calculus."

The Brouwer Theorem for convex sets. To say a subset of $\mathbb{R}^N$, or more generally of a vector space over the real field, is *convex* means that if two points belong to the set, then so does the entire line segment joining those points. More precisely:

Definition 1.4. To say a subset C of a real vector space is *convex* means that: whenever x and y belong to C then so does $tx + (1-t)y$ for every real t with $0 \le t \le 1$.

In the course of proving the Brouwer Theorem for $\mathbb{R}^N$ we'll develop enough machinery to obtain it for all closed, bounded convex sets therein (Theorem 4.5). Officially:

Theorem 1.5 (The "Convex" Brouwer Fixed-Point Theorem)**.** *Suppose N is a positive integer and C is a closed, bounded, convex subset of $\mathbb{R}^N$. Then every continuous mapping taking C into itself has a fixed point.*

1.7 Application: Stochastic Matrices

Recall from Sect. 1.4 that a *stochastic matrix* is a square matrix that is *non-negative* (all entries ≥ 0), all of whose columns sum to 1. In that section (see page 8) we offered as an example the modified internet hyperlink matrix H_1: the original

[3] Our version of Sperner's Lemma generalizes to dimension $N > 2$, where it also implies Brouwer's Theorem. However we will not pursue this direction.

internet hyperlink matrix with the zero-columns replaced by "$1/N$-columns." H_1 is an $N \times N$ stochastic matrix with N on the order of several billion. Our proposed method for ranking internet websites depended on finding a vector $v \in \mathbb{R}^N \backslash \{0\}$ with non-negative entries such that $H_1 v = v$. Now there's no secret that $H_1 v = v$ for *some* vector $v \in \mathbb{R}^N \backslash \{0\}$, i.e., that 1 is an eigenvalue for H_1; in fact this is true of *every* stochastic matrix.

To see why, let e denote the (column) vector in $\mathbb{R}^N$, all of whose entries are 1. Let A be an $N \times N$ stochastic matrix. Since all the columns of A sum to 1 we have $A^t e = e$, where the superscript "t" denotes "transpose." Thus 1 is an eigenvalue of A^t. Since each square matrix has the same eigenvalues as its transpose (the determinant of a square matrix is the same as that of its transpose, hence both matrix and transpose have the same characteristic polynomial) we see that 1 is an eigenvalue of A, i.e., there exists $x \in \mathbb{R}^N \backslash \{0\}$ with $Ax = x$. However, to be meaningful for the internet our eigenvector must have all coordinates non-negative and—up to positive scalar multiples—be unique.

Uniqueness is a special problem. For example, the $N \times N$ identity matrix is stochastic, but (if $N > 1$) has lots of essentially different non-negative fixed points. We'll return to this question in the next chapter. Right now let's see how the Brouwer Fixed-Point Theorem proves that, questions of uniqueness aside:

Theorem 1.6. *Every stochastic matrix has a fixed point, all of whose entries are non-negative, and at least one of which is positive.*

In particular, the modified hyperlink matrix H_1 of Sect. 1.4 has a fixed point that produces at least one ranking of web pages.

For the proof of Theorem 1.6 we'll view $\mathbb{R}^N$ as a space of column vectors, but with distances measured in the metric arising from the "one-norm:"

$$\|x\|_1 = |\xi_1| + |\xi_2| + \ldots + |\xi_N| \qquad (x \in \mathbb{R}^N), \tag{1.5}$$

where ξ_j is the j-th coordinate of the vector x.

Exercise 1.3. Check that $\|\cdot\|_1$ is a norm[4] on $\mathbb{R}^N$, and that

$$\frac{1}{\sqrt{N}}\|x\| \leq \|x\|_1 \leq \sqrt{N}\|x\| \qquad (x \in \mathbb{R}^N)$$

where $\|x\| = (\sum_j \xi_j^2)^{1/2}$, the Euclidean norm of the vector x. Show that this implies that the distance d_1 defined on $\mathbb{R}^N$ by $d_1(x,y) = \|x-y\|_1$ is *equivalent* to the one induced by the Euclidean norm, in that both distances give rise to the same convergent sequences.

Definition 1.7 (Standard Simplex)**.** The *standard N-simplex* Π_N is the set of non-negative vectors in the closed $\|\cdot\|_1$-unit "sphere" of $\mathbb{R}^N$, i.e.,

$$\Pi_N = \{x \in \mathbb{R}^N_+ : \|x\|_1 = 1\},$$

where $\mathbb{R}^N_+$ denotes the set of vectors in $\mathbb{R}^N$ with all coordinates non-negative.

[4] For the definition of "norm" see Appendix C.2, page 194.

For example, Π_2 is the line segment in $\mathbb{R}^2$ joining the points $(0,1)$ and $(1,0)$, while Π_3 is the triangle in $\mathbb{R}^3$ with vertices $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. In general Π_N is the *convex hull* of the standard unit vector basis in $\mathbb{R}^N$, the smallest convex subset of $\mathbb{R}^N$ that contains those vectors (Proposition C.4 of Appendix C).

Exercise 1.4. Show that Π_N is closed and bounded in $\mathbb{R}^N$, hence compact.

Proof of Theorem 1.6. Our goal is to show that A (more accurately: the linear transformation that A induces on $\mathbb{R}^N$) has a fixed point in Π_N. For this it's enough to show that $A(\Pi_N) \subset \Pi_N$, after which the continuity of A (see Exercises 1.5 and 1.6 below), the compactness of Π_N (Exercise 1.4 above), and Theorem 1.5 (the "Convex" Brouwer Fixed-Point Theorem) will combine to produce the desired fixed point.

To see that the matrix A takes Π_N into itself, let's denote by $a_{i,j}$ the element of A in row i and column j. Fix $x \in \Pi_N$, and let ξ_j denote its j-th coordinate. Then, since all matrix elements and coordinates are non-negative:

$$\|Ax\|_1 = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} a_{i,j}\xi_j\right) = \sum_{j=1}^{N}\left(\sum_{i=1}^{N} a_{i,j}\right)\xi_j = \sum_{j=1}^{N}\xi_j = 1,$$

the third equality reflecting the fact that, to its left, the sum in parentheses is the j-th column-sum of A, which by "stochasticity" equals 1. Thus $x \in \Pi_N$, as desired. □

Regarding the Internet. Theorem 1.6 establishes that the modified hyperlink matrix H_1 has fixed points in Π_N, and so yields (possibly many) rankings of web pages. This issue of non-uniqueness arose in Exercise 1.2; we'll resolve it in the next section.

Exercise 1.5. Modify the argument above to show that if A is a stochastic matrix then $\|Ax\|_1 \leq \|x\|_1$ for every $x \in \mathbb{R}^N$ (and even for every $x \in \mathbb{C}^N$).

Exercise 1.6. Use the inequality of the previous exercise to show that A, or more accurately the linear transformation A induces on $\mathbb{R}^N$ (and even on $\mathbb{C}^N$), is continuous in the distances induced on $\mathbb{R}^N$ (and even on $\mathbb{C}^N$) by both the one-norm and the more familiar Euclidean norm.

Exercise 1.7. Modify the ideas in the previous two exercises to establish the continuity of the linear transformation induced on $\mathbb{R}^N$ (and even on $\mathbb{C}^N$) by *any* $N \times N$ real matrix.

Exercise 1.8. Theorem 1.6 shows that every stochastic matrix has 1 as an eigenvalue. Use Exercise 1.5 to show that no eigenvalue, real or complex, has larger modulus. In matrix-theory language: Every stochastic matrix has *spectral radius* 1.

Exercise 1.9. Show that the collection of $N \times N$ stochastic matrices is a convex subset of the real vector space of all $N \times N$ matrices.

1.8 Perron's Theorem

To call a real matrix A of any dimensions (square, row, column ...) "non-negative" (written "$A \geq 0$") means that all its entries are non-negative, and to call it "positive" (written "$A > 0$") means that all its entries are strictly positive.[5] Our first result picks up where Theorem 1.6 above left off, and forms the core of Perron's famous 1907 theorem on eigenvalues of positive matrices.

Theorem 1.8. *Every positive square matrix has a positive eigenvalue, to which corresponds a positive eigenvector.*

We'll prove this by modifying the argument used for Theorem 1.6. The difficulty to be overcome is that, without the hypothesis of stochasticity, our matrices need not take the standard N-simplex Π_N into itself. This is easy to fix.

Lemma 1.9. *Suppose A is an $M \times N$ matrix that is > 0. Then $Ax > 0$ for every vector $x \in \mathbb{R}^N_+ \backslash \{0\}$.*

In words: If every entry of A is strictly positive and every entry of $x \in \mathbb{R}^N \backslash \{0\}$ is non-negative, then every entry of Ax is strictly positive.

Proof of Lemma. Suppose $x \in \mathbb{R}^N_+ \backslash \{0\}$. Fix an index j and note that the j-th coordinate of Ax is the dot product of the j-th row of A with the (transpose of the) column vector x. Since the entries of A are all strictly positive, and the entries of x are non-negative and not all zero, this dot product is strictly positive. □

Proof of Theorem. Suppose A is an $N \times N$ positive matrix. Lemma 1.9 insures that $Ax > 0$ for every $x \in \Pi_N$, hence the equation

$$F(x) = \frac{Ax}{\|Ax\|_1} \qquad (x \in \Pi_N)$$

defines a map F that's continuous on Π_N and takes that simplex into itself. Theorem 1.5 then guarantees for F a fixed point $x_0 \in \Pi_N$. Thus x_0 is a vector with non-negative coordinates, $\|x_0\|_1 = 1$, and $Ax_0 = \lambda x_0$, where $\lambda = \|Ax_0\|_1 > 0$, hence $Ax_0 > 0$ by Lemma 1.9. Conclusion: $x_0 = \lambda^{-1} Ax_0 > 0$. □

Perron Eigenpairs. Theorem 1.8 guarantees that every positive matrix A has what we might call a *Perron eigenpair* (λ, x): a positive ("Perron") eigenvalue λ that has a positive ("Perron") eigenvector x with $\|x\|_1 = 1$. Now we've seen in Theorem 1.6 that every *stochastic* matrix has a "weak" Perron eigenpair $(1, x)$ ("weak" because some coordinates of x may be zero), and in Exercise 1.8 that for stochastic matrices, no eigenvalue (real or complex) has modulus larger than 1. Our next result derives a stronger conclusion from a weaker hypothesis.

Theorem 1.10 (Perron's Theorem)**.** *Each positive square matrix A possesses exactly one Perron eigenpair. Among all the (possibly complex) eigenvalues of A, the Perron eigenvalue has the largest modulus.*

---

[5] *Warning:* This is not to be confused with the notion of "positive-definite," which is something completely different.

Proof. Suppose A is a positive $N \times N$ matrix and (λ, x) is a Perron eigenpair. To prove that λ is the *only* Perron eigenvalue, observe that since A is a positive matrix, so is its transpose A^t. Thus Theorem 1.8 applies to A^t as well, and produces what we might call a "left-Perron[6] eigenpair" (μ, y), where μ is a positive eigenvalue for A^t and y a positive eigenvector for μ.

From the associative property of matrix multiplication:

$$\mu(y^t x) = (y^t A)x = y^t(Ax) = y^t(\lambda x) = \lambda(y^t x).$$

Now $y^t x$, being the dot product of the positive column vectors y and x, is > 0, thus $\mu = \lambda$. This establishes the uniqueness of Perron eigenvalues, since if λ' is another Perron eigenvalue for A then $\lambda' = \mu = \lambda$.

To show that the Perron eigenvalue of A is the *largest* eigenvalue, let $r(A)$ denote the *spectral radius* of A, i.e.,

$$r(A) = \max\{|\gamma| : \gamma \text{ is an eigenvalue of } A\},$$

where on the right-hand side we allow *all* eigenvalues of A, even the complex ones!

Claim. *$r(A)$ is the Perron eigenvalue of A.*

Proof of Claim. We wish to show that if λ is the unique Perron eigenvalue of A and μ is an eigenvalue (real or complex) of A, then $|\mu| \leq \lambda$. To this end, let x be a Perron eigenvector for λ. Suppose μ is an eigenvalue of A and w a corresponding eigenvector, so $w \in \mathbb{C}^N \backslash \{0\}$ and $Aw = \mu w$. Let $|w|$ denote the column vector whose j-th entry is the absolute value of the corresponding entry of w. Then since the entries of A are non-negative:

$$A|w| \geq |Aw| = |\mu w| = |\mu|\,|w|, \tag{1.6}$$

where the inequality is coordinatewise. Now let y be a Perron eigenvector for A^t, so $y > 0$ and, since the Perron eigenvalues for A and A^t coincide, $y^t A = \lambda y^t$. Upon left-multiplying both sides of (1.6) by the positive vector y^t we obtain

$$\lambda y^t|w| = (y^t A)|w| \geq y^t|Aw| = |\mu| y^t|w| \tag{1.7}$$

which implies (since the scalar $y^t|w|$ is > 0) that $\lambda \geq |\mu|$. Thus $\lambda = r(A)$, as desired.

So far we know that the positive matrix A has exactly one Perron eigenvalue, namely the spectral radius $r(A)$. Now we want to show that there is just one Perron eigenvector for this eigenvalue.

Suppose to the contrary that there are two distinct Perron eigenvectors x and y for $r(A)$, so that the pair x, y is linearly independent in $\mathbb{R}^N$.

Claim: There exists a real number $a > 0$ such that the vector $w = ax - y$ is ≥ 0 but not > 0 (i.e., it has at least one coordinate equal to zero).

[6] The terminology "left-Perron" comes from the fact that $y^t A = \mu y^t$, i.e., the row vector y^t is a "left eigenvector" for A with "left eigenvalue" μ.

Granting this Claim: By linear independence, $w \neq 0$, and clearly $Aw = r(A)w$. Every entry of the matrix A is positive, and those of w are non-negative and not all zero, so Lemma 1.9 assures us that $r(A)w = Aw > 0$. But $r(A) > 0$ by Theorem 1.8, so $w > 0$: a contradiction.

Proof of Claim. It remains to find the positive constant a. For this, let ξ_j denote the j-th coordinate of the Perron vector x, and η_j the j-th coordinate of y. We're looking for $a \in \mathbb{R}$ such that $a\xi_j \geq \eta_j$ for all j, and $a\xi_k = \eta_k$ for *some* k. Since no coordinate of x is zero, we can rewrite our criteria as: $a \geq \eta_j/\xi_j$ for all j, and $a = \eta_k/\xi_k$ for some k; in other words the positive real number $a = \max_j \eta_j/\xi_j$ does the job. □

Corollary 1.11. *If A is a positive $N \times N$ stochastic matrix, then there is a positive vector $x \in \Pi_N$ such that $Ax = x$. The vector x is, up to scalar multiples, the unique non-zero fixed point of A.*

Proof. Since A is stochastic, Theorem 1.6 supplies a vector $x_0 \in \mathbb{R}^N_+ \backslash \{0\}$ with $Ax_0 = x_0$. Thus $x = x_0/\|x_0\|_1$ is a fixed point of A that lies in Π_N. Since A is also positive, Perron's Theorem (Theorem 1.10) guarantees that x is positive and is (up to scalar multiples) the unique eigenvector of A for the eigenvalue 1. □

Exercise 1.10 (Uniqueness of the Perron Eigenvector). Extend the argument above to show that if A is a positive $N \times N$ matrix and x is a Perron vector for $r(A)$, then the real eigenspace $\{w \in \mathbb{R}^N : Aw = r(A)w\}$ is one dimensional. Then show that the corresponding *complex* eigenspace is also one dimensional.

Exercise 1.11 (Loneliness of the Perron Eigenvalue). Show that if A is an $N \times N$ positive matrix then its Perron eigenvalue is the only eigenvalue on the circle $\{z \in \mathbb{C} : |z| = r(A)\}$.

Suggestion: Suppose μ is an eigenvalue of A (real or complex) with $|\mu| = r(A)$. Let $w \in \mathbb{C}^N \backslash \{0\}$ be a μ-eigenvector of A. Without loss of generality we may assume that some coordinate of w is positive. Our assumption that $|\mu| = r(A)$ implies that there is equality in (1.7), and this implies that $\sum_j a_{i,j}|w_j| = \left|\sum_j a_{i,j}w_j\right|$ for all indices i. Conclude that $w_j \geq 0$ for all j, hence $\mu > 0$.

1.9 The Google Matrix

We'd like to apply Corollary 1.11 to the problem of ranking internet pages. Unfortunately, the modified hyperlink matrix H_1 we created at the end of Sect. 1.4 (page 8), while stochastic, is far from positive; in fact we noted that "almost all" of its entries are zero. But all is not lost: a simple modification of H_1 shows that a reasonable model of web-surfing can arise from a positive stochastic matrix.

Let E denote the $N \times N$ matrix, each of whose entries is 1. Fix a "damping factor" d with $0 < d < 1$, and let G denote the "Google Matrix"

$$G = dH_1 + \frac{(1-d)}{N}E\,.$$

Since the matrices H_1 and $\frac{1}{N}E$ are both stochastic, so is G (cf., Exercise 1.9, page 12). Furthermore $G > 0$, so Corollary 1.11 guarantees a fixed point $w > 0$ that is unique up to scalar multiples. Thus G provides a unique ranking of web pages.

Why does G provide a reasonable model for web-surfing? Recall that we've already noted how the modified hyperlink matrix H_1 represents what might be termed a "semi-deterministic" model for web-surfing, wherein the surfer *at a given page* chooses randomly among its outlinks with uniform probability and, if there are no outlinks, chooses randomly, again with uniform probability, from all possible web pages. In this vein, the matrix $\frac{1}{N}E$ represents a purely random surfing strategy, wherein our surfer at a given page ignores all links and moves to another page (or stays put) with probability $1/N$. Thus the matrix G models the behavior of a surfer who, at a given page, chooses the next one using the semi-deterministic method with probability d, and the purely random one with probability $1-d$. Google's early experiments indicated that $d = 0.85$ could provide a reasonable start on a web-surfing model [17].

The Elephant in the Room. Let's not forget that G is a huge matrix: $N \times N$ with N in the billions! Thanks to Brouwer and Perron we now know that G produces a unique ranking of web pages, but it's still not clear how to effectively compute this ranking. The fixed-point theorem of Chap. 3 will show us a way that is simple—at least in principle—to do this.

Notes

The Banach Mapping Theorem. Theorem 1.1 first appeared (for one-to-one mappings) in Banach's paper [7]. I learned the result from John Erdman, who presented it, along with its application to the Schröder–Bernstein Theorem, in a seminar at Portland State. The same proof has recently been found independently by Ming-Chia Li [69]. We'll encounter this result again in Chap. 11 when we take up the remarkable subject of paradoxical decompositions.

The Knaster–Tarski Theorem. In [116, page 286] Tarski points out that in the 1920s he and Knaster discovered Theorem 1.2, with Knaster publishing the result in [63]. Tarski goes on to say that he found a generalization to "complete lattices" and lectured on it and its applications during the late 1930s and early 1940s before finally publishing his results in [116].

The Brouwer Fixed-Point Theorem. This result (Theorem 1.3) appeared in [18, 1912], where it was proved using topological methods developed by Brouwer. It is one of the most famous and widely applied theorems in mathematics; see [91] for an exhaustive survey of the legacy of this result, and [21, Chap. 1] for a popular exposition.

Positive matrices. The arguments used here to prove Theorem 1.10, the famous theorem of Perron (1907), follow those of [30, Chap. 2]. In 1912 Frobenius extended Perron's results to certain matrices with non-negative entries. The resulting "Perron–Frobenius" theory is the subject of ongoing research, with an enormous literature spanning many scientific areas. For more on this see, e.g., [30, 72] or[76].