

1 Introduction

In 1920, a young Polish mathematician named Stefan Banach (aged 28) was working on integral equations in Lwów (now Lviv, Ukraine). In seeking to obtain existence and uniqueness of solutions, he proved a contraction theorem for normed spaces. In 1930, independently in Naples, Renato Caccioppoli (aged 26) formulated the result directly in metric space language and with a distinct approximation flavor.

Meanwhile, in 1998, two graduate students at Stanford University, Larry Page and Sergey Brin (both aged 25), were developing an algorithm to rank web pages. Unbeknownst to many, the mathematical principle underlying their revolutionary *PageRank algorithm* was precisely the *Banach-Caccioppoli Contraction Theorem*.

This lecture explores this remarkable connection, demonstrating how abstract mathematical theory developed between the wars became the foundation for a technology that would change how humanity accesses information—generating extraordinary economic value in the process.

2 Part I: The Birth of a Theorem (1920s-1930s)

2.1 Historical Context: The Polish School of Mathematics

The 1920s were a golden age for Polish mathematics. Following Poland's independence in 1918 after over a century of partition, Polish mathematicians established world-class centers of research, particularly in Lwów and Warsaw. The Lwów School, led by figures such as Hugo Steinhaus and Stefan Banach, became especially renowned for functional analysis.

Legend has it that Banach's mathematical career began with a chance encounter. In 1916, Steinhaus overheard two young men discussing the Lebesgue integral on a park bench in Kraków—one of them was Banach, then an engineer with no formal mathematical training. This encounter led to one of the most productive collaborations in mathematical history.

2.2 The Contraction Mapping Theorem

2.2.1 The Motivating Problem

Banach was investigating integral equations of the form:

$$x(t) = f(t) + \lambda \int_a^b K(t, s)x(s) ds$$

where x is the unknown function, $K(t, s)$ is a kernel function, $f(t)$ is given, and λ is a parameter. Such equations arise naturally in physics, particularly in problems involving heat conduction, elasticity, and electromagnetic theory.

The traditional approach to such equations involved complex analytical techniques. Banach sought a more general, abstract method that would guarantee both existence and uniqueness of solutions while simultaneously providing a constructive algorithm for finding them.

2.2.2 The Banach Contraction Theorem

Theorem 1 *Let (X, d) be a complete metric space and let $T : X \rightarrow X$ be a contraction mapping of order α , i.e., there exists $\alpha \in [0, 1)$ such that*

$$d(T(x), T(y)) \leq \alpha d(x, y) \quad \text{for all } x, y \in X.$$

Then:

1. T has a unique fixed point $x^* \in X$, i.e., $T(x^*) = x^*$.
2. For any starting point $x_0 \in X$, the sequence defined by $x_{n+1} = T(x_n)$ converges to x^* .
3. The rate of convergence is geometric:

$$d(x_n, x^*) \leq \frac{d(x_1, x_0)}{1 - \alpha} \alpha^n.$$

Lemma 2 (contractive sequences) *Let $(x_0, x_1, \dots)$ be a sequence in a metric space (X, d) . If there exists $\alpha \in [0, 1)$ such that*

$$d(x_{k+1}, x_k) \leq \alpha d(x_k, x_{k-1}) \tag{1}$$

for all $k \geq 1$, then

$$d(x_m, x_n) \leq \frac{d(x_1, x_0)}{1 - \alpha} \alpha^n = C \alpha^n, \tag{2}$$

for all $m \geq n \geq 0$. In particular, $\{x_n\}$ is Cauchy. Moreover, if $\{x_n\}$ has limit x^* , then $d(x_n, x^*) \leq C \alpha^n$ for all $n \geq 0$.

Proof. Iterating (1), for each $k \geq 1$, we have:

$$d(x_{k+1}, x_k) \leq \alpha d(x_k, x_{k-1}) \leq \alpha^2 d(x_{k-1}, x_{k-2}) \leq \dots \leq \alpha^k d(x_1, x_0),$$

thus $d(x_{k+1}, x_k) \leq \alpha^k d(x_1, x_0)$.

Let $m \geq n \geq 0$. If $m = n$, then $d(x_m, x_n) = 0$ and (2) is satisfied. Else, $m = n + p$ with $p \geq 1$, and:

$$\begin{aligned} d(x_m, x_n) &= d(x_{n+p}, x_n) \leq d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + \cdots + d(x_{n+p-p+1}, x_{n+p-p}) \\ &= d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \\ &\leq \alpha^{m-1} d(x_1, x_0) + \alpha^{m-2} d(x_1, x_0) + \cdots + \alpha^n d(x_1, x_0) \\ &= d(x_1, x_0) \sum_{j=n}^{m-1} \alpha^j \leq d(x_1, x_0) \sum_{j=n}^{\infty} \alpha^j = d(x_1, x_0) \frac{\alpha^n}{1-\alpha}, \end{aligned}$$

again (2) is satisfied.

For each $\varepsilon > 0$, there exists $N \geq 0$ such that $C\alpha^l \leq \varepsilon$ for all $l \geq N$, then for all $m, n \geq N$ (assuming w.l.o.g. $m \geq n$) we have $d(x_m, x_n) \leq C\alpha^n \leq \varepsilon$ and $\{x_n\}$ is Cauchy.

If the sequence has limit x^* , then for each fixed n , $d(x_m, x_n) \rightarrow d(x^*, x_n)$ as $m \rightarrow \infty$.¹ But (2) implies that, for all $m \geq n$, $d(x_m, x_n) \leq C\alpha^n$, and passing to the limit $d(x^*, x_n) \leq C\alpha^n$.

Proof of Theorem 1. Arbitrarily choose $x_0 \in X$ and define $x_{n+1} = T(x_n)$ for $n \geq 0$. Since T is a contraction with constant $\alpha \in [0, 1)$, we have:

$$d(x_{k+1}, x_k) = d(T(x_k), T(x_{k-1})) \leq \alpha d(x_k, x_{k-1})$$

all $k \geq 1$. By Lemma 2, $\{x_n\}$ is Cauchy, by completeness it has a limit x^* , and

$$d(x_n, x^*) \leq \frac{d(x_1, x_0)}{1-\alpha} \alpha^n = C\alpha^n$$

for all $n \geq 0$. It remains to be shown that x^* is the unique fixed point of T .

For each $n \geq 0$,

$$d(T(x^*), x^*) \leq d(T(x^*), T(x_n)) + d(T(x_n), x^*) = d(T(x^*), T(x_n)) + d(x_{n+1}, x^*) \leq \alpha d(x^*, x_n) + d(x_{n+1}, x^*),$$

and passing to the limit as $n \rightarrow \infty$ yields $d(T(x^*), x^*) \leq 0$ (because $x_n \rightarrow x^*$). Thus x^* is a fixed point.

Finally, if x^* and y^* are both fixed points. Then:

$$d(x^*, y^*) = d(T(x^*), T(y^*)) \leq \alpha d(x^*, y^*).$$

and since $\alpha < 1$, this is only possible if $d(x^*, y^*) = 0$, i.e., $x^* = y^*$. □ ■

2.2.3 The Power of Abstraction

The beauty of Banach's theorem lies in its generality. Rather than solving specific integral equations, Banach abstracted the essential property—contraction—and showed that any operator with this property must have a unique fixed point. Moreover, the theorem is constructive: it provides an algorithm (repeated application of T) that is guaranteed to converge.

This represents a paradigm shift in mathematics: from solving particular problems to identifying general structural properties that guarantee solutions exist (and in this case also being able to approximate them).

2.3 Renato Caccioppoli and the Italian Contribution (1930)

2.3.1 Caccioppoli's Life and Work

Renato Caccioppoli (1904-1959) was one of the most brilliant and tragic figures in 20th-century Italian mathematics. Working in Naples, he made fundamental contributions to functional analysis, measure theory, and the calculus of variations.

In 1930, while studying elliptic partial differential equations, Caccioppoli independently developed a fixed-point theorem similar to Banach's, though in a more abstract setting. His formulation emphasized the geometric aspects of the contraction property and extended the results to more general spaces.

In the literature, especially in Italian and some European traditions, the theorem is called the *Banach-Caccioppoli Theorem*.

3 Part II: The Digital Revolution (1990s-2000s)

3.1 The Web Search Problem

By the mid-1990s, the World Wide Web was growing explosively, but finding relevant information was increasingly difficult. Early search engines like *AltaVista* and *Lycos* relied primarily on keyword matching and simple heuristics. They had no principled way to assess which pages were most important or authoritative.

The fundamental challenge: *Given a query, how do we rank billions of web pages to present the most relevant and authoritative results first?*

¹Because $d(\cdot, x_n)$ is a continuous function.

3.2 The Web as a Graph

The World Wide Web can be modeled as a directed graph (W, L) where:

- $W = \{1, 2, \dots, w\}$ is set of web pages.
- $L : W \rightarrow 2^W$ is the link function that associates to each page j the set $L(j)$ set of pages that j links to i (the **out-neighbors**).

With this, for each page i the set of pages j that links to i (the in-neighbors) is

$$I(i) = \{j \in W : i \in L(j)\}.$$

Thus $j \in I(i) \iff i \in L(j)$. The number $d_j = |L(j)|$ is called the **out-degree** of page j .

Note that

$$d_j = |\{i \in W : i \in L(j)\}| = \sum_{i=1}^w \mathbf{1}(i \in L(j)) = \sum_{i=1}^w \delta_i(L(j)) \quad (3)$$

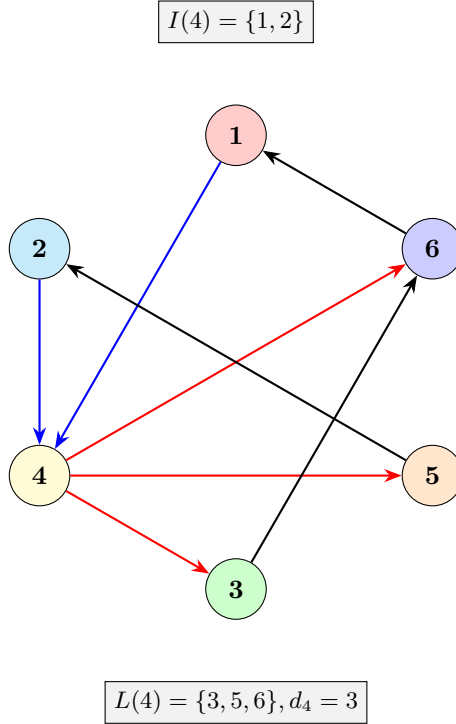


Figure 1: Directed graph of the web with 6 pages, showing hyperlinks and sets $I(4)$ and $O(4)$ for page 4.

3.3 The Random Surfers Model

Larry Page and Sergey Brin modeled web surfing as follows:

There is a huge population of surfers, at every turn:

- A fraction $\alpha \in (0, 1)$ (e.g., 85%), follow a random outgoing link.
- A fraction $1 - \alpha$, jump to a random page (teleportation).

Assume that at time t , the population is distributed as $x = (x_1, \dots, x_w)^T$, with x_j the fraction of surfers at page j ($j = 1, 2, \dots, w$). At time $t + 1$, the fraction of surfers jumping to page i from page j , is

$$\begin{cases} \alpha \frac{x_j}{d_j} + (1 - \alpha) \frac{x_j}{w} & \text{if } i \in L(j) \\ (1 - \alpha) \frac{x_j}{w} & \text{if } i \notin L(j) \end{cases} = \left(\alpha \frac{\mathbf{1}(i \in L(j))}{d_j} + (1 - \alpha) \frac{1}{w} \right) x_j.$$

Call this fraction G_{ij} . The **Google matrix** is the $w \times w$ matrix $G = [G_{ij}]$. With this the total fraction the population of surfers at page i , at time $t + 1$, will be

$$y_i = \sum_{j=1}^w G_{ij} x_j,$$

and the population will be distributed as $y = Gx$.

3.4 PageRank as a Contraction Mapping

Theorem 3 (PageRank Convergence) *The Google matrix G defines, via $x \mapsto Gx$, a contraction mapping of order α on the simplex $\Delta = \{x \in \mathbb{R}^w : x_i \geq 0, \sum x_i = 1\}$ of possible population distributions, with respect to the distance*

$$\|x - y\|_1 = \sum_{i=1}^w |x_i - y_i|,$$

induced by the $\|\cdot\|_1$ norm on $\mathbb{R}^w$.²

Proof. Set

$$P_{ij} = \frac{\mathbf{1}(i \in L(j))}{d_j}$$

for all pages i, j .

Step 1: G maps Δ to Δ

For any $x \in \Delta$:

$$\begin{aligned} (Gx)_i &= \sum_{j=1}^w G_{ij}x_j = \sum_{j=1}^w \left(\alpha P_{ij} + (1-\alpha) \frac{1}{w} \right) x_j \\ &= \alpha \sum_{j=1}^w P_{ij}x_j + (1-\alpha) \frac{1}{w} \sum_{j=1}^w x_j \\ &= \alpha \sum_{j=1}^w P_{ij}x_j + \frac{1-\alpha}{w}. \end{aligned}$$

Clearly $(Gx)_i \geq 0$. Also:

$$\begin{aligned} \sum_{i=1}^w (Gx)_i &= \sum_{i=1}^w \left(\alpha \sum_{j=1}^w P_{ij}x_j + \frac{1-\alpha}{w} \right) = \sum_{i=1}^w \alpha \sum_{j=1}^w P_{ij}x_j + \sum_{i=1}^w \frac{1-\alpha}{w} \\ &= \alpha \sum_{j=1}^w \sum_{i=1}^w P_{ij}x_j + 1 - \alpha = \alpha \sum_{j=1}^w \left(x_j \sum_{i=1}^w P_{ij} \right) + 1 - \alpha \end{aligned}$$

but

$$\sum_{i=1}^w P_{ij} = \sum_{i=1}^w \frac{\mathbf{1}(i \in L(j))}{d_j} = \frac{1}{d_j} \sum_{i=1}^w \mathbf{1}(i \in L(j)) = 1 \quad \forall j \in W, \quad (4)$$

because $d_j = \sum_{i=1}^w \mathbf{1}(i \in L(j))$, and so:

$$\sum_{i=1}^w (Gx)_i = \alpha \sum_{j=1}^w \left(x_j \sum_{i=1}^w P_{ij} \right) + 1 - \alpha = \alpha \sum_{j=1}^w x_j + 1 - \alpha = 1,$$

that is, $Gx \in \Delta$.

Step 2: G is a contraction

For $x, y \in \Delta$, we want to show that

$$\|Gx - Gy\|_1 \leq \alpha \|x - y\|_1$$

where $\alpha < 1$ is the damping factor.

Substep 2.1: Decompose G and analyze the difference

Recall that

$$G_{ij} = \alpha P_{ij} + (1-\alpha) \frac{1}{w}.$$

We can write:

$$G = \alpha P + (1-\alpha)E,$$

where E is the $w \times w$ matrix with all entries equal to $1/w$.

Now compute:

$$Gx - Gy = G(x - y) = \alpha P(x - y) + (1-\alpha)E(x - y)$$

Substep 2.2: Evaluate $E(x - y)$

The matrix E has all entries equal to $1/w$, so for any vector z :

$$(Ez)_i = \sum_{j=1}^w E_{ij}z_j = \sum_{j=1}^w \frac{1}{w} z_j = \frac{1}{w} \sum_{j=1}^w z_j$$

² $\|v\|_1 = \sum_{i=1}^w |v_i|$ for all $v \in \mathbb{R}^w$.

For $z = x - y$ where both $x, y \in \Delta$:

$$\sum_{j=1}^w (x - y)_j = \sum_{j=1}^w x_j - \sum_{j=1}^w y_j = 1 - 1 = 0.$$

Therefore:

$$E(x - y) = 0.$$

and:

$$Gx - Gy = \alpha P(x - y).$$

Substep 2.3: Normed vector space calculus

Taking norm of both sides:

$$\|Gx - Gy\|_1 = \|\alpha P(x - y)\|_1 = \alpha \|P(x - y)\|_1 \leq \alpha \|P\|_{1,1} \|x - y\|_1$$

where $\|P\|_{1,1}$ is the $(1, 1)$ operator norm of P , that is,

$$\|P\|_{1,1} = \max_{v \neq 0} \frac{\|Pv\|_1}{\|v\|_1} = \max_{1 \leq j \leq n} \|P_{\cdot j}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m P_{ij} = \max_{1 \leq j \leq n} 1 = 1$$

where the first equality is the definition the $\|\cdot\|_{1,1}$ operator norm, the second equality is a classical theorem ($P_{\cdot j}$ is the j -th column of P), the third descends from the definition of $\|\cdot\|_1$ (and the fact that all elements of P are positive), the fourth from (4). Summig up:

$$\|Gx - Gy\|_1 \leq \alpha \|P\|_{1,1} \|x - y\|_1 = \alpha \|x - y\|_1,$$

as desired. ■

But Δ is a closed subset of the complete metric space $(\mathbb{R}^w, \|\cdot\|_1)$, therefore it is complete too. and so the Banach-Caccioppoli Theorem applies, and starting from any initial distribution x , the sequence $(x, Gx, G^2x, \dots)$ converges to a fixed point $r = Gr$.

Definition 4 (PageRank vector) *The PageRank vector is the unique $r \in \Delta$ satisfying*

$$r = Gr. \tag{5}$$

Proposition 5 *For $r \in \Delta$, the fixed-point equation (5) is equivalent to the component-wise system*

$$r_i = \alpha \sum_{j \in I(i)} \frac{r_j}{d_j} + \frac{1 - \alpha}{w} \quad \text{for all } i \in W.$$

Proof. By definition, $(Gr)_i = \sum_{j=1}^w G_{ij} r_j$ with

$$G_{ij} = \alpha \mathbf{1}(i \in L(j)) / d_j + (1 - \alpha) / w. = \alpha \mathbf{1}(j \in I(i)) / d_j + (1 - \alpha) / w.$$

Thus

$$(Gr)_i = \alpha \sum_{j=1}^w \frac{\mathbf{1}(j \in I(i))}{d_j} r_j + \frac{1 - \alpha}{w} \sum_{j=1}^w r_j = \alpha \sum_{j \in I(i)} \frac{r_j}{d_j} + \frac{1 - \alpha}{w},$$

since $\sum_j r_j = 1$. Hence $r_i = (Gr)_i$ is exactly the claimed identity. Conversely, if the component-wise identities hold, stacking them yields $G\mathbf{r} = \mathbf{r}$. ■

The term $\sum_{j \in I(i)} r_j / d_j$ aggregates *endorsements* from in-neighbors: a page j passes its importance r_j equally across its d_j outlinks, so endorsements from high-importance (high r_j), low-out-degree (low d_j , that is selective) pages carry more weight.

A page is influential if linked by influential pages.³

References

1. Banach, S. (1922). “Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales.” *Fundamenta Mathematicae* **3**: 133–181.
2. Caccioppoli, R. (1930). “Un teorema generale sull’esistenza di elementi uniti in una trasformazione funzionale.” *Rendiconti dell’Accademia Nazionale dei Lincei* **11**: 794–799.
3. Brin, S. and Page, L. (1998). “The anatomy of a large-scale hypertextual Web search engine.” *Computer Networks and ISDN Systems* **30**: 107–117.

³Did anyone did it for math/econ/cs papers?

A Matrix norms

Definition 6 The induced $\|A\|_{1,1}$ norm of a matrix A is defined as:

$$\|A\|_{1,1} = \max_{\mathbf{v} \neq \mathbf{0}} \frac{\|A\mathbf{v}\|_1}{\|\mathbf{v}\|_1}.$$

This is the smallest constant C such that $\|A\mathbf{v}\|_1 \leq C\|\mathbf{v}\|_1$ for all vectors $\mathbf{v}$. Clearly:

$$\|A\|_{1,1} = \max_{\|\mathbf{u}\|_1=1} \|A\mathbf{u}\|_1$$

Theorem 7 For any $m \times n$ matrix A :

$$\|A\|_{1,1} = \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}| = \max_{1 \leq j \leq n} \|A_{.j}\|_1$$

(i.e., the maximum absolute column sum).

Proof. Let $\mathbf{e}_j$ be the j -th standard basis vector. Then:

$$\|A\mathbf{e}_j\|_1 = \left\| \sum_{i=1}^m A_{ij} \mathbf{e}_i \right\|_1 = \sum_{i=1}^m |A_{ij}|.$$

Since $\|\mathbf{e}_j\|_1 = 1$, we have:

$$\|A\|_{1,1} \geq \max_j \|A\mathbf{e}_j\|_1 = \max_j \sum_{i=1}^m |A_{ij}|$$

For the reverse inequality, consider any $\mathbf{v}$ with $\|\mathbf{v}\|_1 = 1$:

$$\begin{aligned} \|A\mathbf{v}\|_1 &= \sum_{i=1}^m |(A\mathbf{v})_i| = \sum_{i=1}^m \left| \sum_{j=1}^n A_{ij} v_j \right| \\ &\leq \sum_{i=1}^m \sum_{j=1}^n |A_{ij}| |v_j| \quad (\text{triangle inequality}) \\ &= \sum_{j=1}^n \sum_{i=1}^m |A_{ij}| |v_j| = \sum_{j=1}^n |v_j| \sum_{i=1}^m |A_{ij}| \\ &\leq \sum_{j=1}^n |v_j| \cdot \max_k \sum_{i=1}^m |A_{ik}| \\ &= \max_k \sum_{i=1}^m |A_{ik}| \cdot \sum_{j=1}^n |v_j| \\ &= \max_k \sum_{i=1}^m |A_{ik}| \quad (\text{since } \|\mathbf{v}\|_1 = 1). \end{aligned}$$

■