

From 1920s Lviv to the Cyberspace

(BAMS inaugural lecture)

Fabio Maccheroni

Bocconi, October 27, 2025

Definition (metric space)

A **metric space** is a pair (X, d) with $d : X \times X \rightarrow \mathbb{R}$ such that:
for all $x, y, z \in X$,

- ① $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$,
- ② $d(x, y) = d(y, x)$,
- ③ $d(x, y) \leq d(x, z) + d(z, y)$.

Vocabulary: $d(x, y)$ is the **distance** between x and y

Examples

Pythagorean distance: $X = \mathbb{R}^2$, for $x = (x_1, x_2)$ and $y = (y_1, y_2)$,

$$d_2(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Manhattan distance: $X = \mathbb{R}^w$, for $x = (x_1, x_2, \dots, x_w)$ and $y = (y_1, y_2, \dots, y_w)$,

$$d_1(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_w - y_w|$$

Uniform distance: $X = C[0, 1]$, for $x : t \mapsto x(t)$ and $y : t \mapsto y(t)$,

$$d_\infty(x, y) = \max_{t \in [0, 1]} |x(t) - y(t)|$$

Convergent sequences

Definition (convergent sequence)

Let (X, d) be a metric space. A sequence $(x_0, x_1, x_2, \dots)$ in X **converges** to a point x in X (written $x_n \rightarrow x$ as $n \rightarrow \infty$) if and only if:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ such that whenever } n \geq N, d(x_n, x) < \varepsilon.$$

Intuition: for any prescribed “error level” ε , the points x_n eventually approximate x within ε distance

Vocabulary: the point x is the **limit** of the sequence

Cauchy sequences

Definition (Cauchy sequence)

Let (X, d) be a metric space. A sequence $(x_0, x_1, x_2, \dots)$ in X is a **Cauchy sequence** if and only if:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ such that whenever } m, n \geq N, d(x_m, x_n) < \varepsilon.$$

Intuition: as the sequence progresses, its elements become arbitrarily close to each other

Fact: every convergent sequence is Cauchy, the converse is in general false.
Think of $X = (0, 1)$ with the usual distance and the sequence $(1/2, 1/3, 1/4, \dots)$

Definition (complete metric space)

A metric space (X, d) is **complete** if every Cauchy sequence in X converges to a point in X .

Fact: $\mathbb{R}^w$ with d_p is complete, for $p = 1, 2, \dots, \infty$

Fact: A subset A of a complete metric space (X, d) is complete wrt the inherited distance $d|_{A \times A}$ iff any sequence of points of A that converges in X has limit in A (i.e. A is **closed** in X)

Banach Contraction Theorem

Theorem (Banach, 1920, page 28/49)

Let (X, d) be a complete metric space. If $T : X \rightarrow X$ is a map for which there exists a constant $0 \leq \alpha < 1$ such that

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

then T is called a **contraction (of order α)**, and:

- 1 there exists a unique fixed point $x^* \in X$ with $Tx^* = x^*$;
- 2 for each $x_0 \in X$, the sequence $x_{n+1} = T(x_n)$ converges to x^* ;
- 3 convergence is geometric:

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

Hint: Prove that the sequence $(x_0, Tx_0, T^2x_0, \dots)$ is Cauchy, then completeness guarantees limit x^* , and use contractivity of T again to conclude

The equations that started it all

In the early 20th century analysts sought general methods to solve **integral equations** arising in physics

Banach's original problem:

$$x(t) = g(t) + \lambda \int_0^1 K(t, s) x(s) ds$$

Banach realized these equations could be written as $x = Tx$ where

$$\begin{aligned} T : C[0, 1] &\longrightarrow C[0, 1] \\ x(t) &\longmapsto (Tx)(t) = g(t) + \lambda \int_0^1 K(t, s) x(s) ds \end{aligned}$$

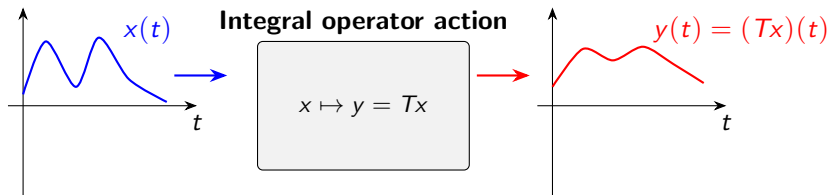
and that the key to understanding existence and uniqueness of solutions was to study **when** the operator T is a **contraction**

Functional Analysis: functions treated as points in a metric space

Kernels

For each $x \in C[0, 1]$, $y = Tx$ is defined by

$$y(t) = g(t) + \lambda \int_0^1 K(t, s) x(s) ds$$



where $K : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ is an integrable function, called **kernel**

Modern relevance: large networks

When $[0, 1]$ models a continuum of agents, $K \geq 0$, and $\int_0^1 K(t, s) ds = 1$, the integral part of $y(t)$ is the average, weighted by $K(t, \cdot)$, of the values of x , so $K(t, s)$ is the **per-unit-influence** of the value of x in s on the value of y in t .

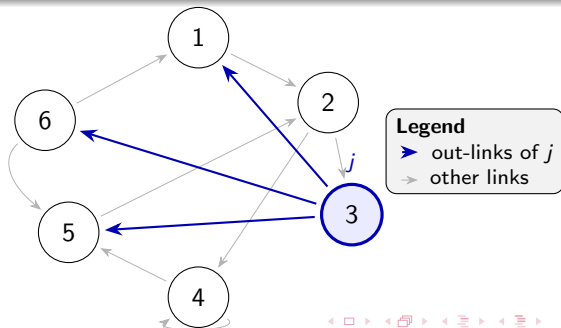
The WWW

Model (what *Theory* is about)

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

- $W = \{1, 2, \dots, w\}$ is the set of all web pages;
- $L : W \rightarrow 2^W$ describes for each page j the set $L(j)$ of pages i that j links to, the **out-links**;
- $d_j = |L(j)|$ is the **out-degree** of j .

$$L(3) = \{1, 5, 6\}$$
$$d_3 = 3$$



Random surfers (Brin and Page, 1998)

Random surfer dynamics

At each time, each of the surfers:

- with probability $\alpha \in (0, 1)$ follows a *random out-link*;
- with probability $1 - \alpha$ *teleports* randomly to any page in W .

The Google matrix G

IF at time n , the population is distributed as $x = (x_1, \dots, x_w)$,
with x_j the fraction of surfers at page j

THEN at time $n + 1$, *the fraction of surfers jumping to i from j , is*

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

DEFINE by $G = [G_{ij}]$ the **Google matrix**

SUMMING UP at time $n + 1$, *total fraction of the population at page i*

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

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

G is a contraction

Theorem (PageRank convergence)

The map $x \mapsto Gx$ is a contraction of order α on the set

$$\Delta = \left\{ x \in \mathbb{R}^w : x_i \geq 0, \sum_{i=1}^w x_i = 1 \right\}$$

of all population distributions, with respect to the distance

$$d_1(x, y) = \sum_{i=1}^w |x_i - y_i|.$$

Now: Δ is closed in the complete $(\mathbb{R}^w, d_1)$, so it is complete too.

By Banach's theorem, starting from any initial distribution x , the sequence $(x, Gx, G^2x, \dots)$ converges to a **limit distribution**, the fixed point $r = Gr$

A page is influential if linked by influential pages

Definition (PageRank vector)

The **PageRank vector** is the only $r \in \Delta$ solving $r = Gr$.

Proposition (Brin and Page, 1998)

The PageRank vector is the unique $r \in \Delta$ satisfying

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

where $I(i) = \{j \in W : i \in L(j)\}$ is the set of pages j that link to i .

Backflip: $\sum_{j \in I(i)} r_j / d_j$ aggregates *endorsements* from in-pages:
page j passes its importance r_j uniformly across its d_j out-links,
so endorsements from important (and selective) pages carry more weight

The life of a theorem

- **1920** — Stefan Banach (aged 28) lays, in his PhD dissertation, the foundations of *Functional Analysis* in just 49 pages. This *extreme abstraction* is fueled by his attempt to solve equations arising from physics
- **1930** — Renato Caccioppoli (aged 26) “o’ Genio” extends Banach’s results to nonlinear PDEs
- **1957** — Richard Bellman (aged 37) uses the Banach Contraction Theorem as the engine of his brainchild, *Dynamic Programming*, changing the history of *Economic Theory*
- **1988** — Richard S. Sutton (aged 32) introduces *Temporal Difference Learning*, a recursive, convergent learning rule (based on Dynamic Programming) that became the mathematical core of modern *Reinforcement Learning* and *Artificial Intelligence*
- **1998** — Sergey Brin and Larry Page (both aged 25) apply the Banach Contraction Theorem to design the PageRank algorithm (Google now makes roughly one billion USD a day)

The power of abstraction

Banach was addressing problems in pure mathematics — integral equations arising from physics. He could not have imagined that his abstract theorem would, in a century, become the foundation for dynamic optimization and for organizing human knowledge in the digital age (not to speak of quantum mechanics)

Stanislaw Ulam, reports him saying: *A mathematician is a person who can find analogies between theorems, a better mathematician is one who can see analogies between proofs, and the best mathematician can notice analogies between theories*

theories are models