The statistics of continued fractions I

Past and present

Anish Ghosh, TIFR

A Grecian take

(1) If stick a is at least n_1 times as long as stick b, but less than $n_1 + 1$ times as long as stick b, write down the integer n_1

(2) Form a new stick c by breaking from stick a a stick of length n_1 times the length of stick b. Now we have sticks b and c, with stick b longer than stick c, so we can repeat the procedure.

So $\text{length}(a) = n_1 \,\text{length}(b) + \text{length}(c)$,

$$\alpha := \frac{\text{length}(b)}{\text{length}(a)} = \frac{1}{n_1 + \dfrac{\text{length}(c)}{\text{length}(b)}}.$$

So if $n_1, n_2, \ldots, n_k$ are the successive positive integers produced by applying the algorithm to commensurable sticks a and b, then one easily sees (by induction) that

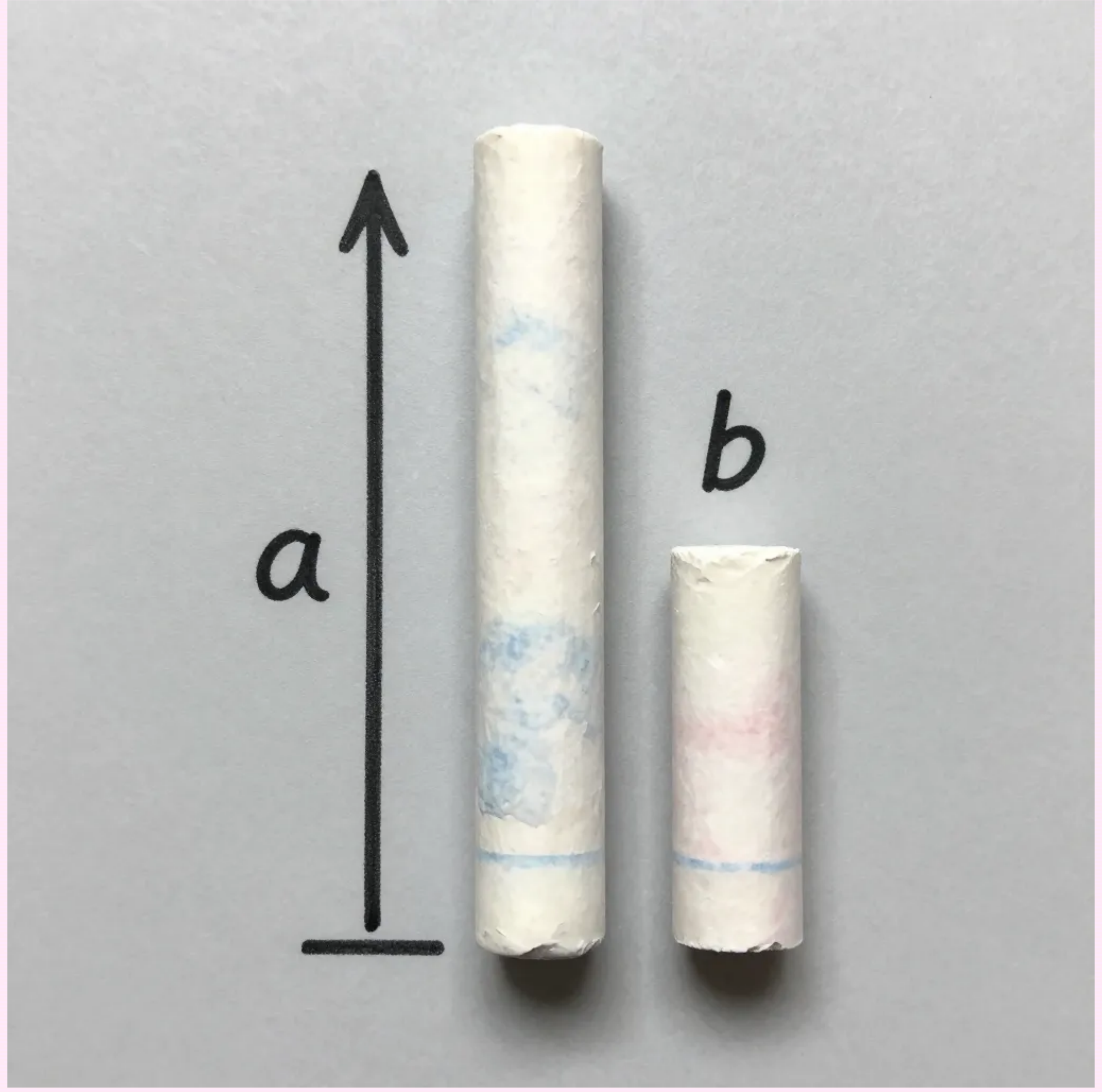
$$\alpha = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{\ddots + \cfrac{1}{n_k}}}}.$$

If sticks a and b are incommensurable, the process never terminates. In that case,

$$\alpha = \lim_{k\to\infty} \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{\ddots + \cfrac{1}{n_k}}}}$$

and we write the corresponding infinite continued fraction as

$$\alpha = \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \ddots}}}.$$



- $\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, \ldots]$
- $\frac{1+\sqrt{5}}{2} = [1; 1, 1, 1, \ldots]$
- Irrationals have a unique representation
- Rationals have two such representations
- $\theta = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{n-1} - 1, 1]$
- A continued fraction expansion of θ is *eventually periodic* if and only if θ is a solution to a quadratic equation with integer coefficients
- Example: $\theta = [0; 2, 3, 2, 3, \ldots]$
- $\theta = \frac{1}{2+\frac{1}{3+\theta}}$
- So $2\theta^2 + 6\theta - 3 = 0$
- The converse is slightly more complicated. I leave it as an exercise

Tilings and cutting sequences

Christoffel, 1875; Series, 1985

- Consider the line L with slope $\kappa = \text{slope}(L) = \beta/\alpha$. Then

  ✓ Between any two a's is at least one b.

  ✓ Between any two a's there are either $[\beta/\alpha]$ or $[\beta/\alpha]+1$ b's.

- A sequence of a's and b's satisfying 1 and 2, will be called *almost constant*, and call the exponent $[\beta/\alpha]$ or $[\alpha/\beta]$ its value.

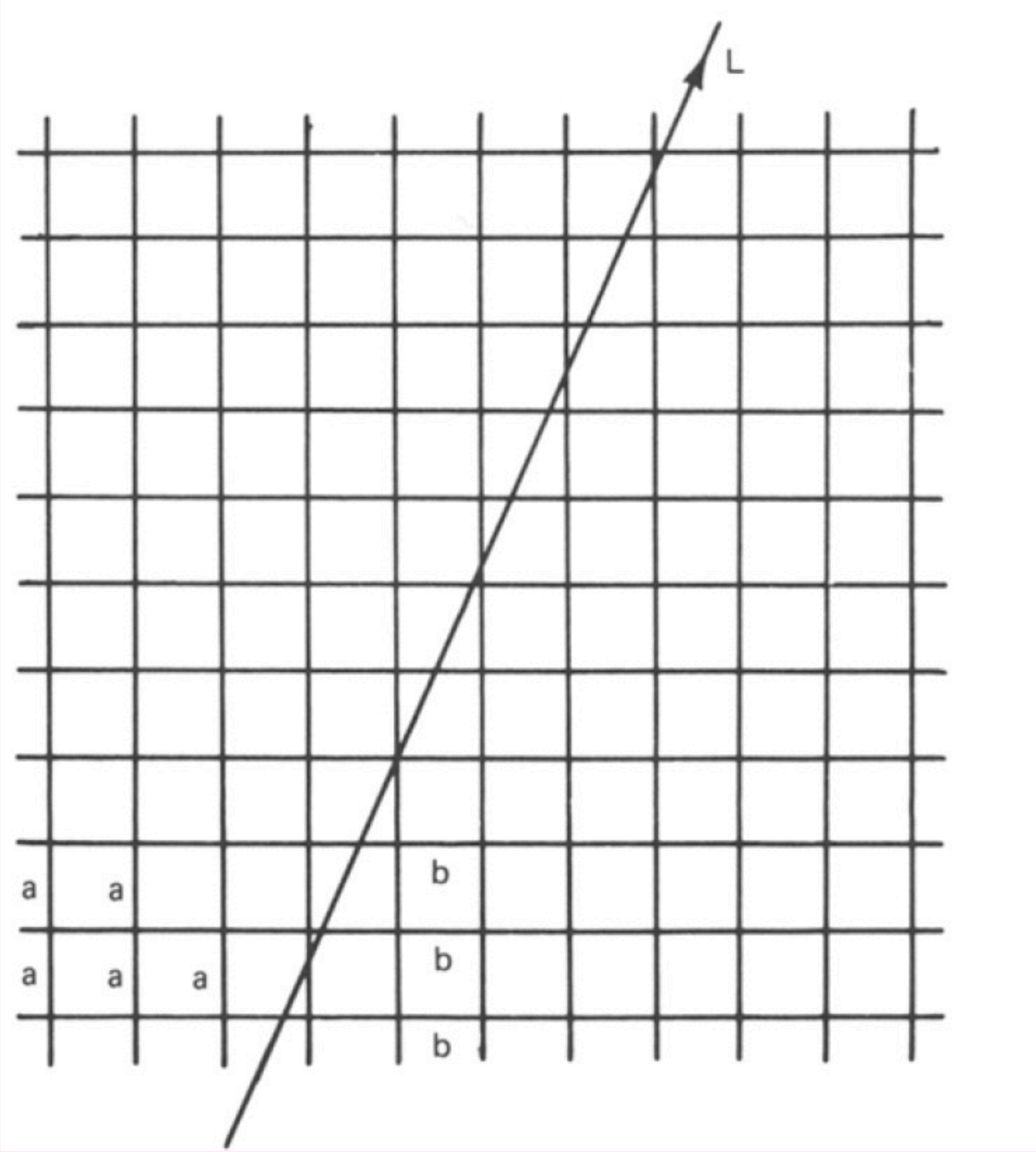
- Given any almost constant sequence s of value n, set $a' = ab^n$, $b' = b$. It is clear that we can rewrite s as a sequence s' in the symbols a', b', called the derived sequence of s. Of course, s' may itself be almost constant, in which case we may derive it again. Call a sequence which can be derived arbitrarily many times, characteristic.

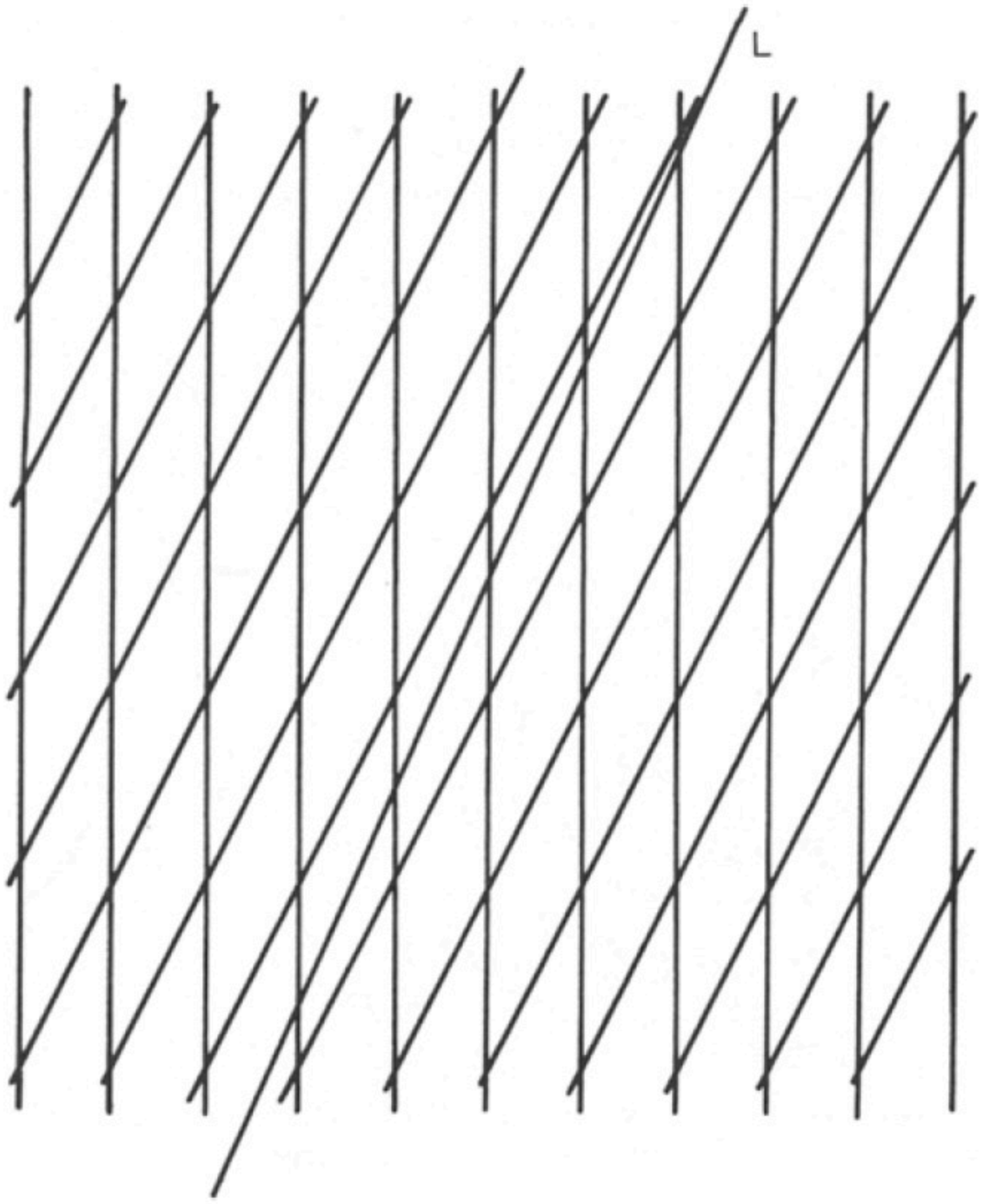
- The cutting sequence of a line L is characteristic and the values of the successive derived sequences are $n_0, n_1, n_2, \ldots$, where

$$\lambda = \text{slope}(L) = [n_0, n_1, n_2, \ldots].$$

$$a^3ba^2ba^2ba^2ba^3ba^2ba^2ba^2ba^3ba^2ba^2ba^2ba^2ba^3ba^2ba^2ba^2ba^3ba^2ba^2ba^2ba^2b$$

$$[0, 2, 4, 3, 2] = 30/67$$

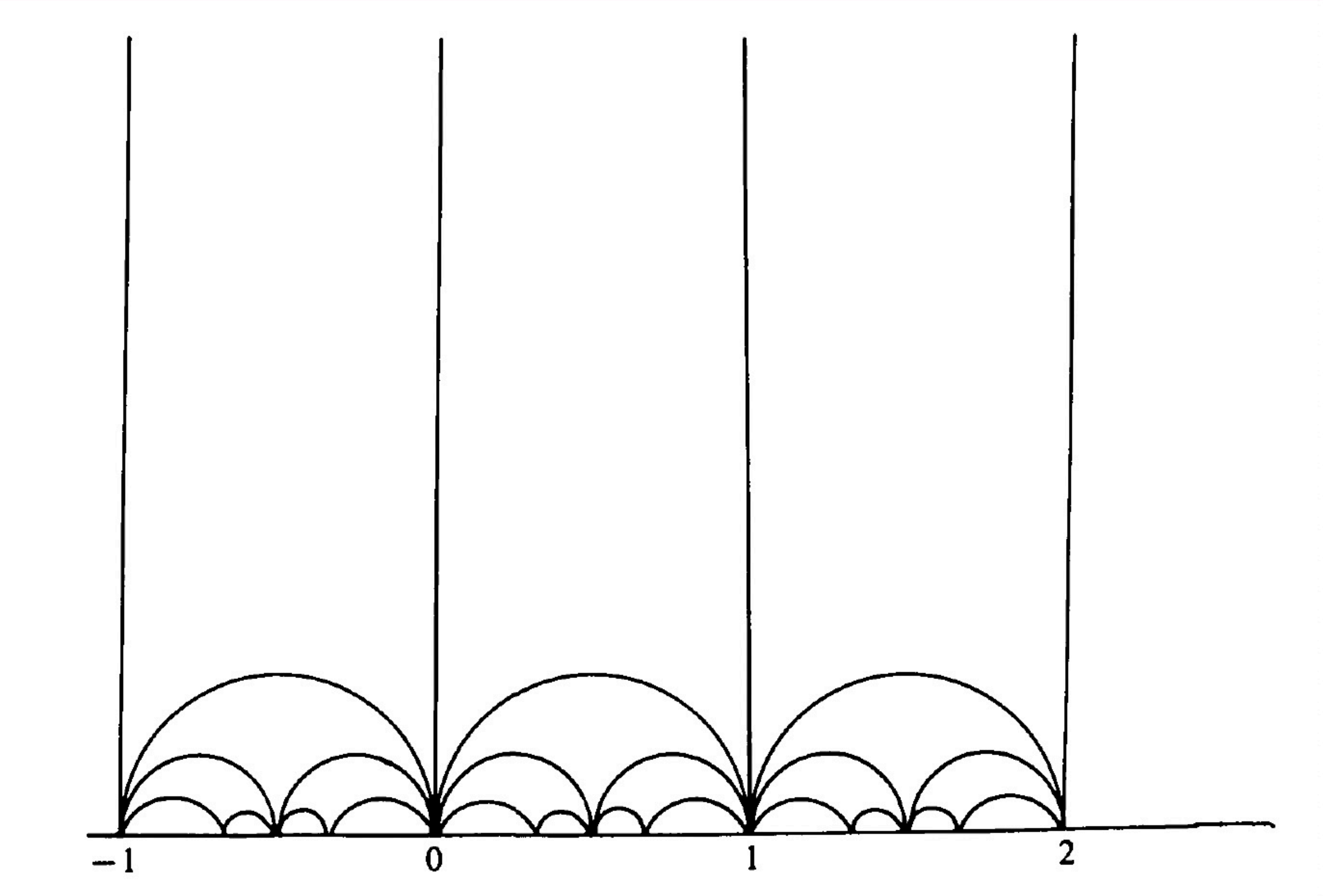


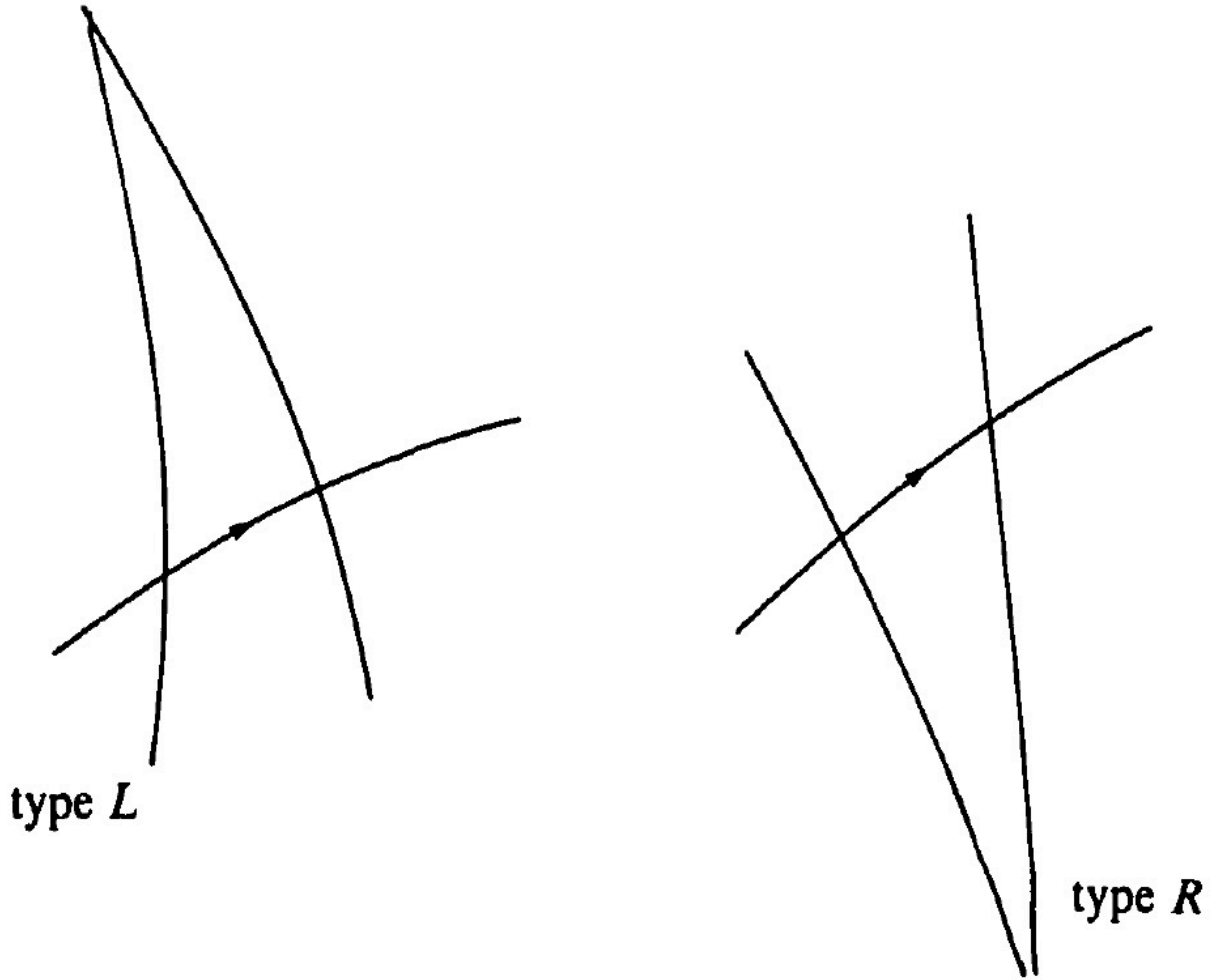

L

Tilings and cutting sequences II

The hyperbolic plane

- Geodesics are vertical lines or semicircles centred on the x axis
- The *Farey* tessellation is a tiling of the hyperbolic plane by *ideal* triangles
- The vertices are precisely the rational numbers
- The sides are exactly the images of the imaginary axis under the group SL(2, Z)



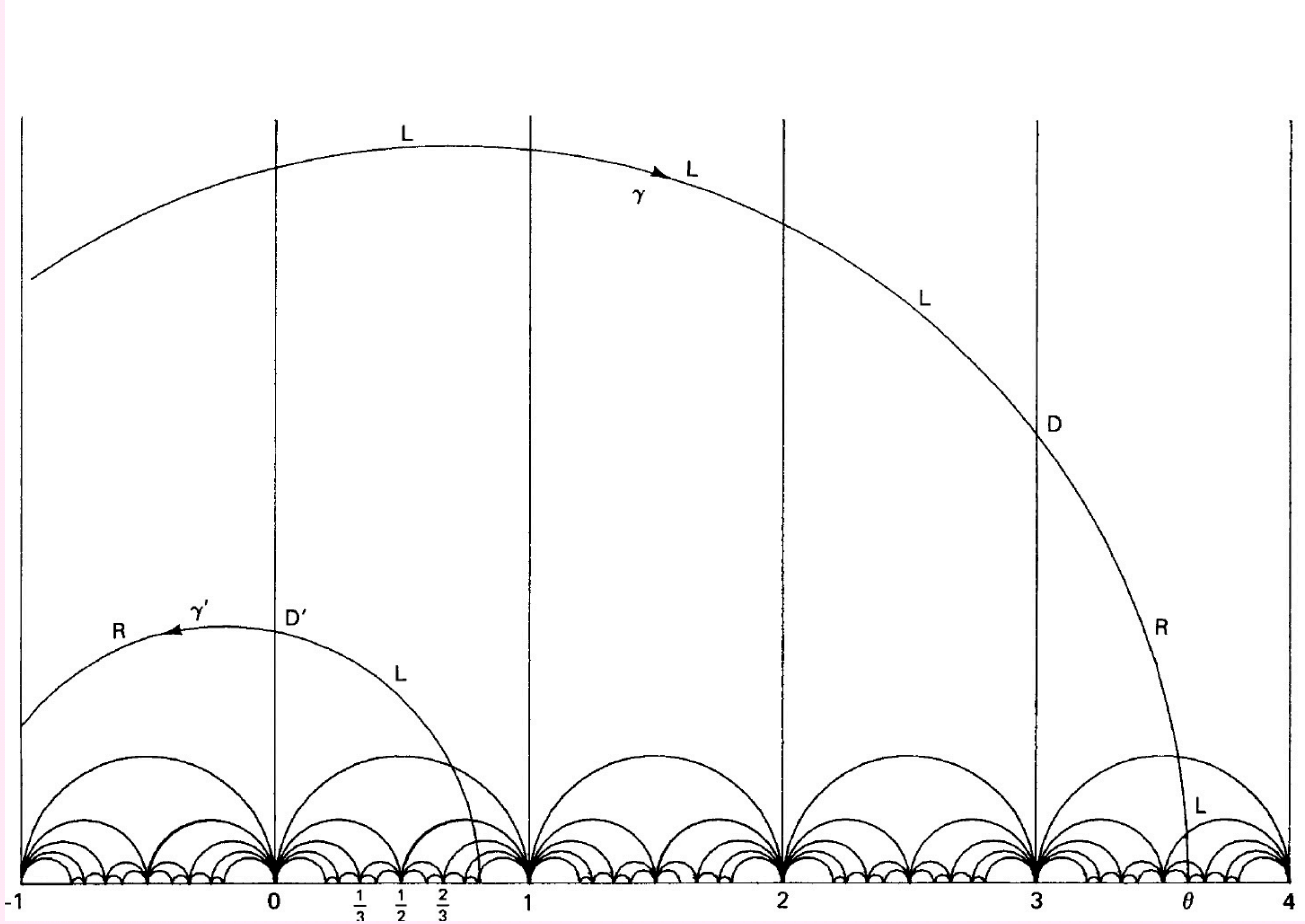

type L
type R

Reading off a cfe

- An oriented geodesic in $\mathbb{H}$ is divided into segments as it traverses the triangles composing the fundamental domain $\mathcal{F}$
- During this crossing, a segment s cuts two sides of Δ that meet at a vertex at infinity
- Label the oriented segment s with R (Right) or L (Left) depending on whether this vertex lies to the right or left of s
- This labeling is invariant under the action of Γ
- Consequently, any geodesic γ on the surface $\mathbb{H}/\Gamma$ can be associated with a *cutting sequence* of the form $\ldots R^{n_0} L^{n_1} R^{n_2} \ldots$, where $n_i \in \mathbb{N}$
- If a point $x \in \gamma$ lies at the end of a segment labeled R, there exists a unique lift $\tilde{\gamma}$ of γ such that the lift $\tilde{x}$ of x lies on the imaginary axis
- The positive and negative endpoints of $\tilde{\gamma}$ are, respectively,

$$\gamma_\infty = n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \cfrac{1}{n_4 + \cfrac{1}{n_5 + \cdots}}}}$$

$$\gamma_{-\infty} = \cfrac{-1}{n_0 + \cfrac{1}{n_{-1} + \cfrac{1}{n_{-2} + \cfrac{1}{n_{-3} + \cdots}}}}$$



A letter from Gauss to Legendre on continued fractions, translated from the French

From:
Arnoux, Pierre; Schmidt, Thomas A. Natural extensions and Gauss measures for piecewise homographic continued fractions. Bull. Soc. Math. France 147 (2019), no. 3, 515–544.

January 30, 1812

Sir,

I tell you one thousand thanks for the two memoirs you made me the honor to send me and that I received these past few days. The functions you discuss there, as well as the questions of probabilities on which you prepare a large work, have a great appeal for me, although I myself have worked little on those last. I remember nevertheless a curious problem, which I considered 12 years ago, but that I could not then solve to my satisfaction. You might interest yourself with it for a little time, in which case I am sure that you will find a more complete solution. Here is the problem. Let M be an unknown quantity between the limits 0 and 1, for which all values are, either equally probable, or more or less following a given law. We suppose it converted in a continued fraction

$$M = \cfrac{1}{a' + \cfrac{1}{a'' + \text{etc}}}$$

What is the probability that, stopping the expansion at a finite term $a^{(n)}$, the following fraction

$$\cfrac{1}{a^{(n+1)} + \cfrac{1}{a^{(n+2)} + \text{etc}}}$$

be between the limits 0 and x? I denote it by $P(n, x)$, and I have, supposing for M all values equally probable,

$$P(0, x) = x;$$

$P(1, x)$ is a transcendental function depending on the function

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{x}.$$

which Euler calls unexplainable and on which I just gave several researches in a memoir presented to our society of sciences which will soon be printed. But for the cases where n is larger, the exact value of $P(n, x)$ seems intractable. However, I found by very simple reasoning that, for infinite n, one has

$$P(n, x) = \frac{\log(1 + x)}{\log 2}$$

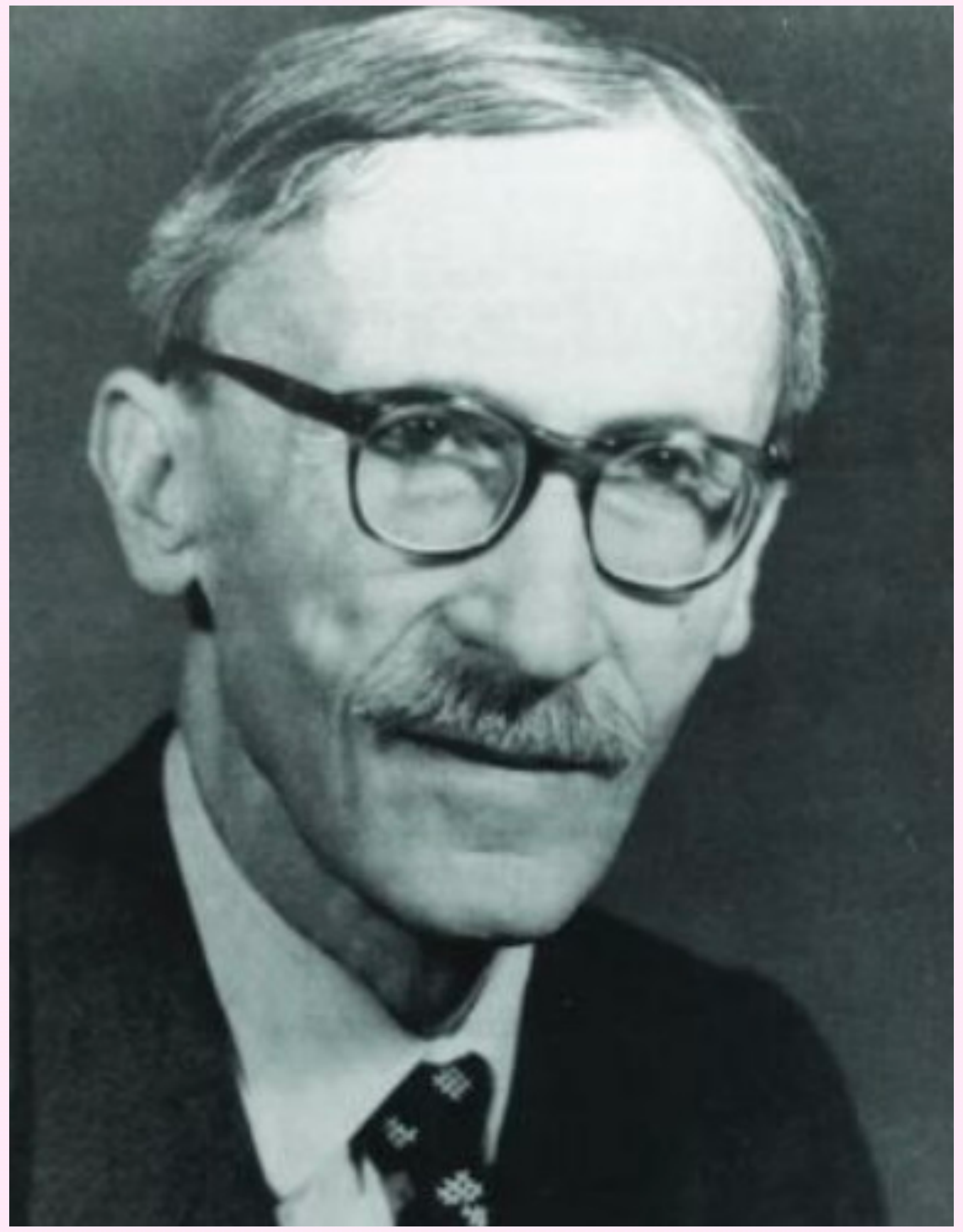
But the efforts I made during my researches to assign

$$P(n, x) - \frac{\log(1 + x)}{\log 2}$$

for a very large, but not infinite, value of n have been fruitless.

$(\cdots)$

The Levy-Khinchin theorem



- For almost every real number

$$\lim_{n\to\infty} {q_n}^{1/n} = e^{\beta}$$

where

$$e^{\beta} = e^{\pi^2/(12\ln 2)} = 3.27582291872181115978768188 2\ldots$$

The Gauss Map

- A measure-preserving system $(X; B; \mathcal{B}; T)$ is a finite measure space $(X; B; \mathcal{B})$ equipped with a (measurable) map $T : X \to X$ that is measure-preserving, i.e.$\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$

- The system is *ergodic* if whenever $A \in \mathcal{B}$ satisfies $A = T^{-1}A$, we have that $\mu(A) = 0$ or $\mu(A) = \mu(X)$

- Example: Irrational rotations of the circle

- Example: $\times 2$ on the circle

- Let $(X; B; \mathcal{B}; T)$ be a measure preserving system. For any integrable $f : X \to \mathbb{C}$, the *time average*

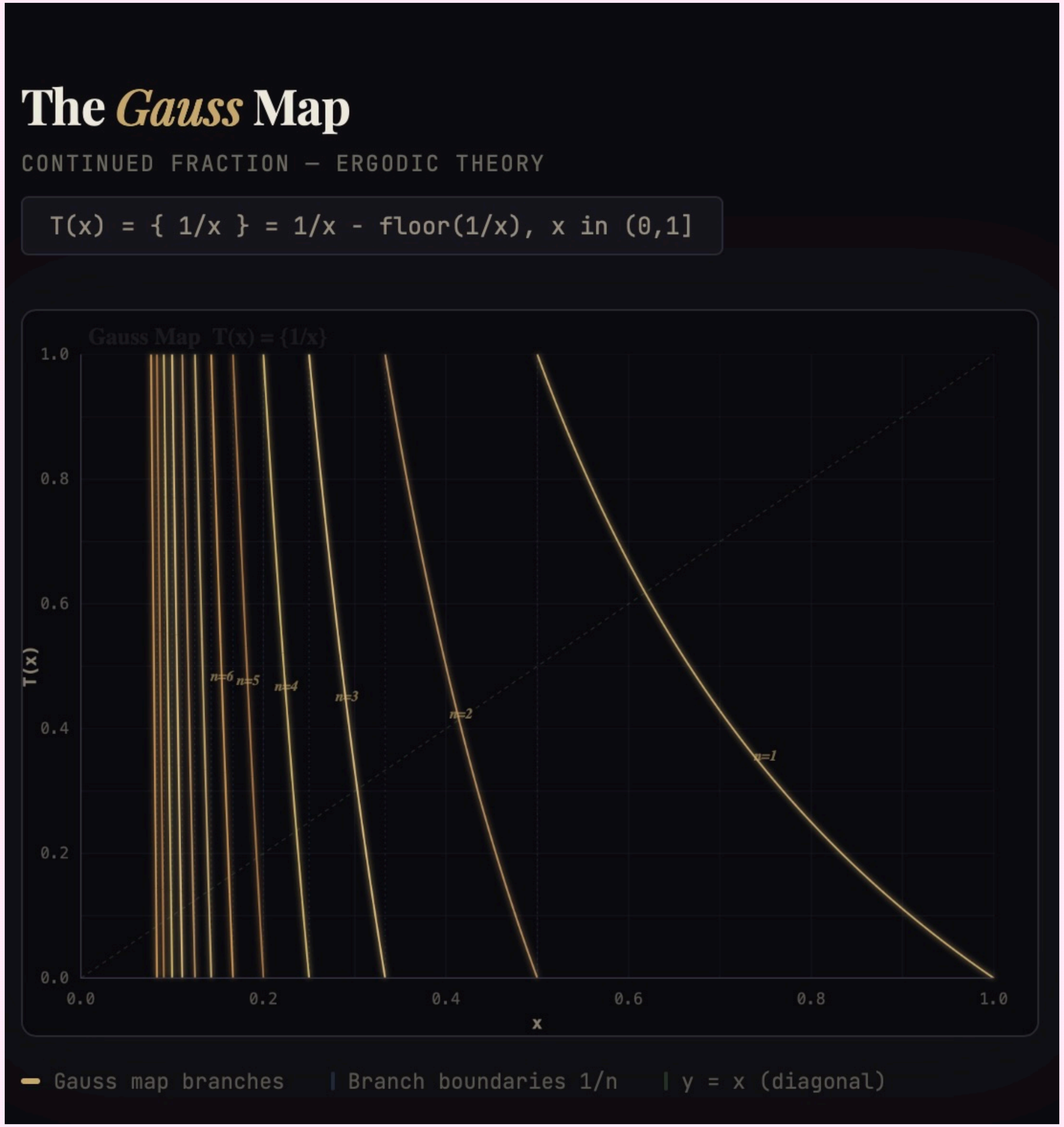
- $f^*(x) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$

- exists for a.e. $x \in X$. It is T-invariant, is in $L^1(X)$, and $\int f d\mu = \int f^* d\mu$

- If T is μ ergodic, then $f^*(x) = \frac{1}{\mu(X)} \int_X f d\mu$ for a.e. $x \in X$.

The Gauss map

- Take $X = [0,1]$ and $T(x) = \left\{\frac{1}{x}\right\} = \frac{1}{x} - \lfloor\frac{1}{x}\rfloor$
- If $x = [a_1, a_2, \dots]$ then $T(x) = [a_2, a_3, \dots]$; so T is a shift on $\mathbb{N}^{\mathbb{N}}$
- Gauss measure: $d\mu(x) = \frac{1}{\log 2}\frac{1}{1+x}$
- T is μ invariant and ergodic
- Consider the modular surface $\mathbb{H}/SL(2,\mathbb{Z})$
- The Gauss map is a factor of a cross section of the geodesic flow on the unit tangent bundle of the modular surface
- Take $A = (0, a)$, these sets generate the Borel σ algebra
- Then $\log 2\mu(T^{-1}(A)) = \mu\left(\bigsqcup_n \left(\frac{1}{n+a}, \frac{1}{n}\right)\right)$
- $= \sum_n \int_{1/n+a}^{1/n} \frac{dx}{1+x} = \sum_n \log\left(\frac{1+\frac{1}{n}}{1+\frac{1}{n+a}}\right)$
- $= \log(1+a) + \lim_{N\to\infty} \log\left(\frac{N+1}{N+a+1}\right)$
- $= \log(1+a) = \int_0^a \frac{dx}{1+x} = (\log 2)\mu(A)$



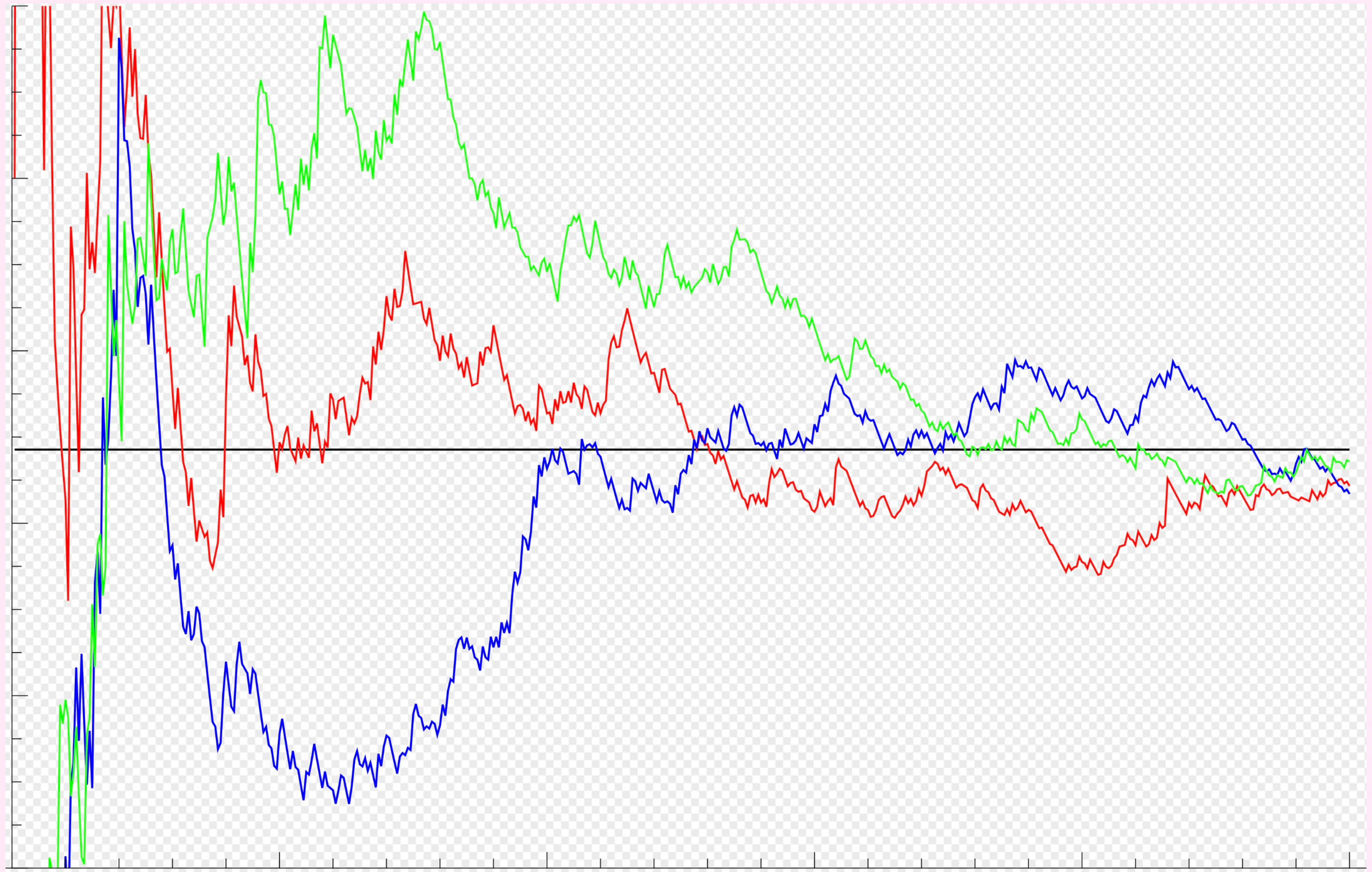
Khinchin’s constant

- For almost every real number

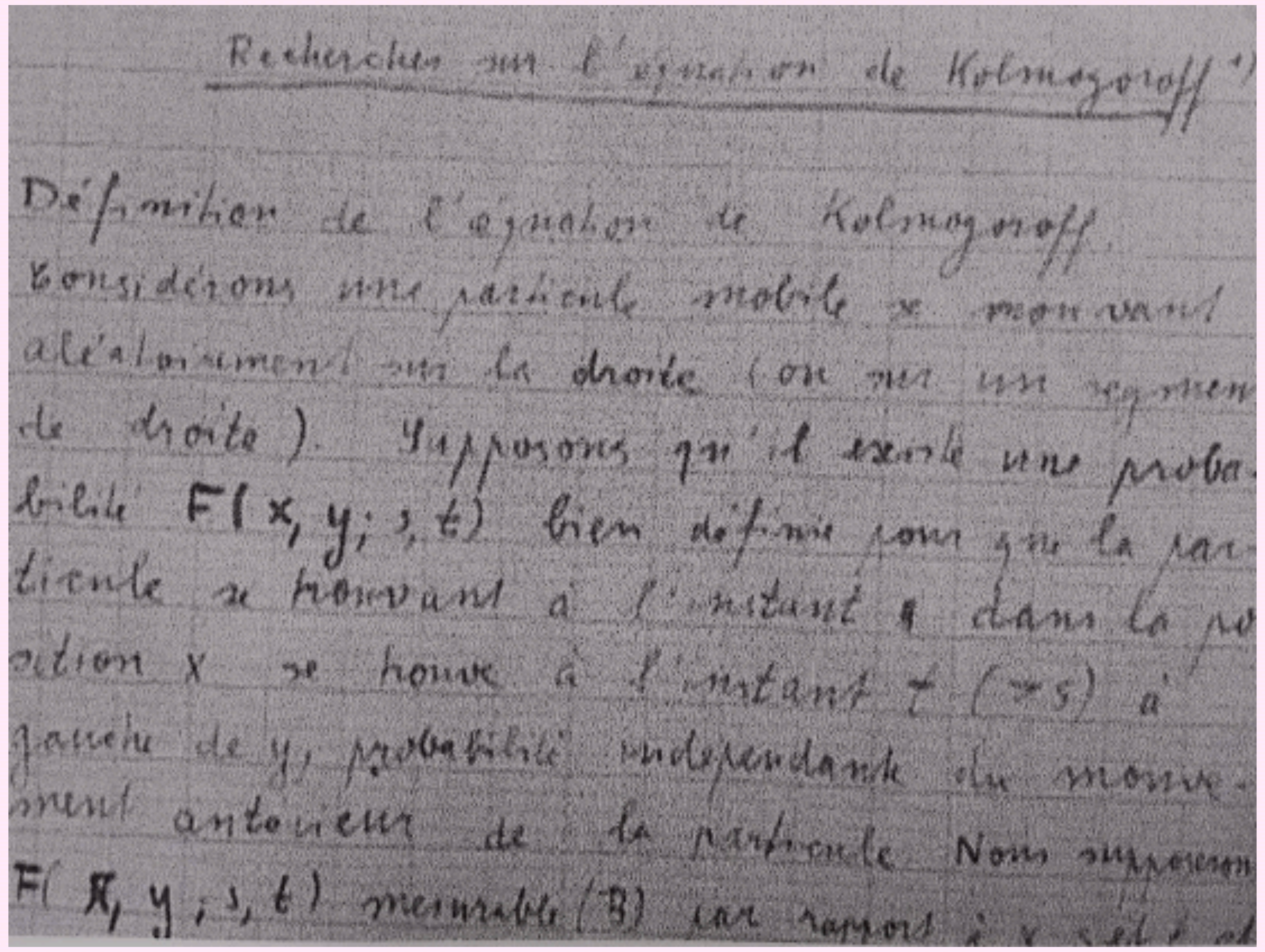
$$\lim_{n\to\infty} \left(a_1 a_2 \ldots a_n\right)^{1/n} = K_0$$

where

$$K_0 = 2.68545\,20010\,65306\,44530\ldots$$

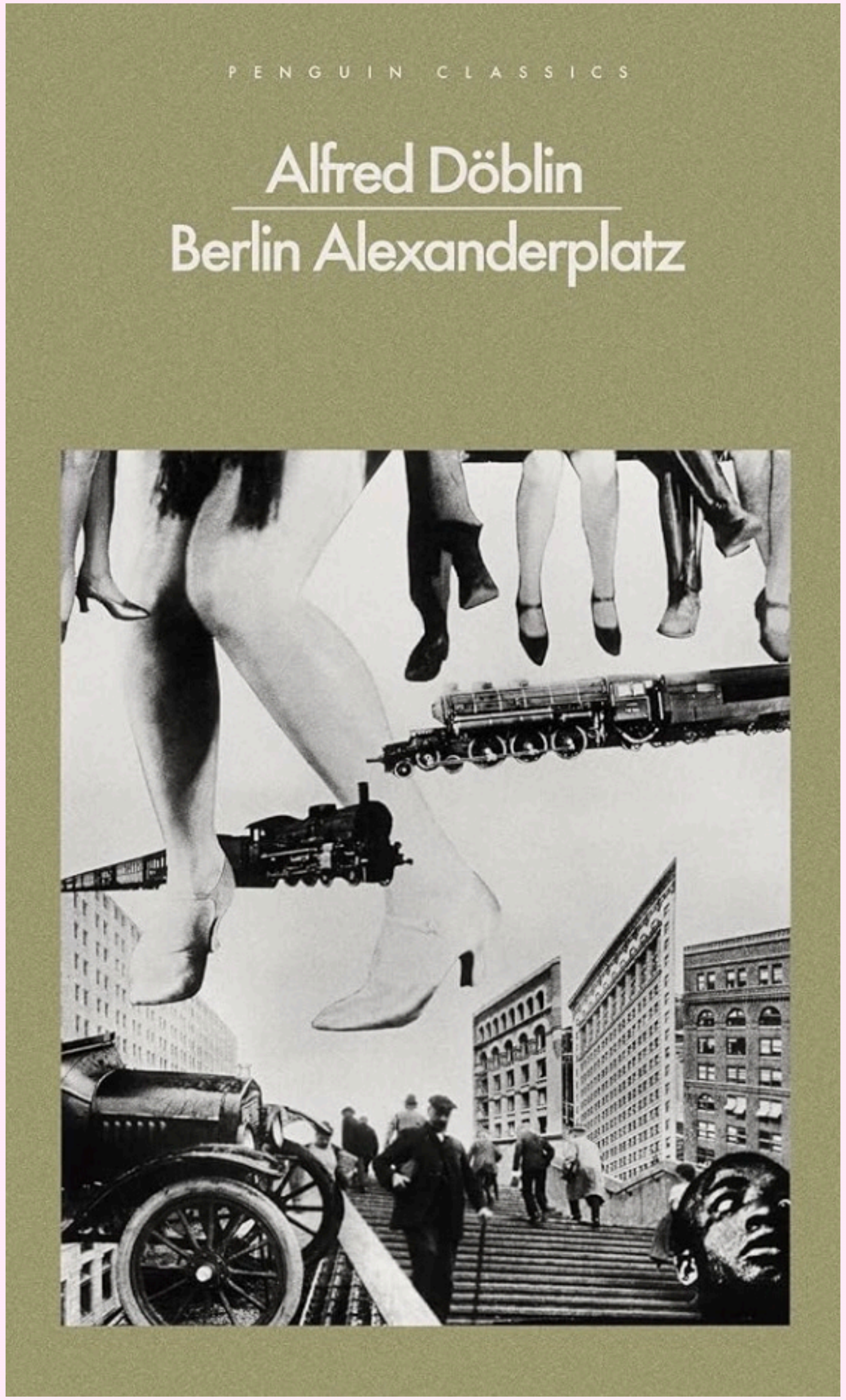


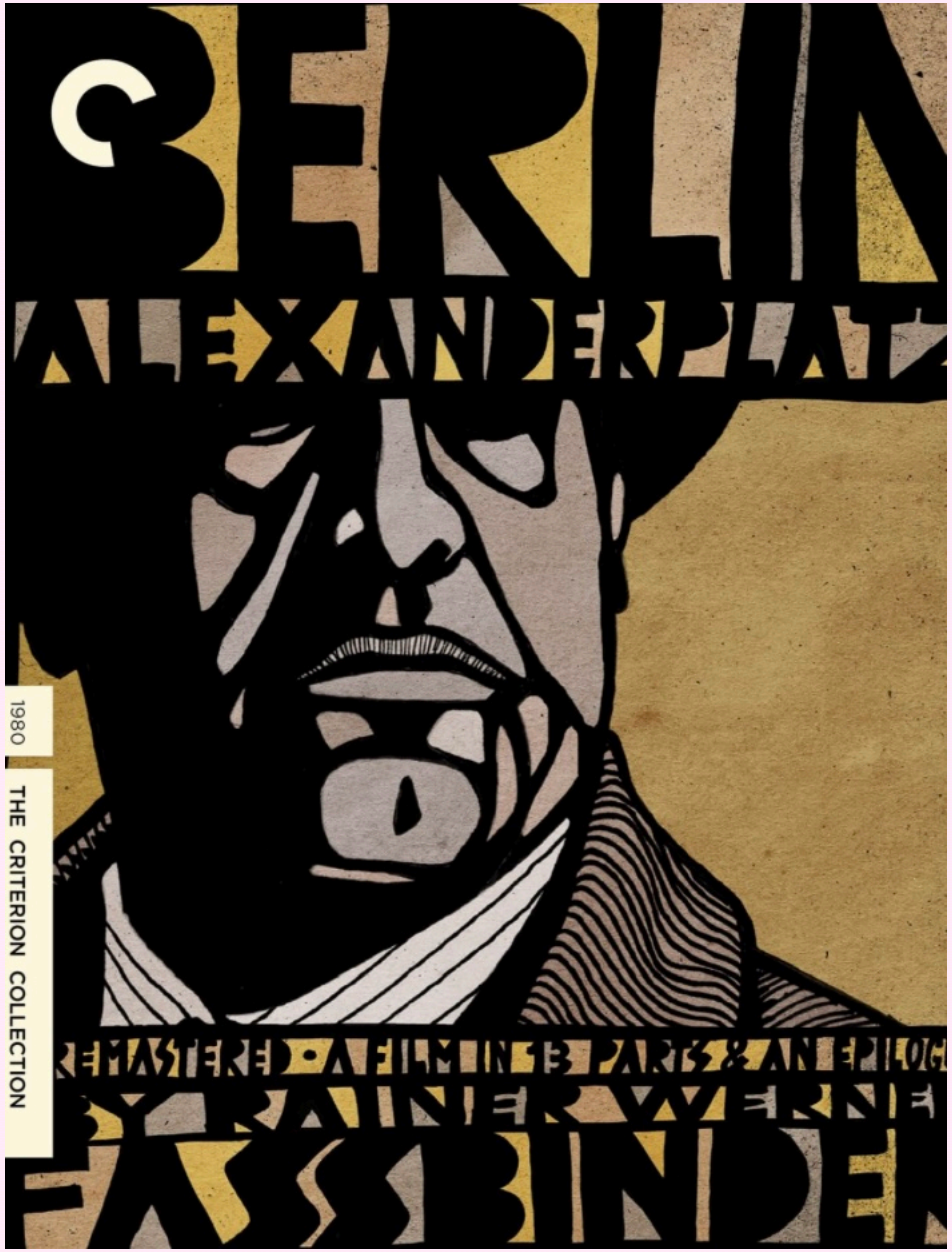
Plis Cachete no. 11668



Recherches sur l'équation de Kolmogoroff 1)

Définition de l'équation de Kolmogoroff. Considérons une particule mobile se mouvant aléatoirement sur la droite (ou sur un segment de droite). Supposons qu'il existe une probabilité $F(x, y; s, t)$ bien définie pour que la particule se trouvant à l'instant s dans la position x se trouve à l'instant t $(> s)$ à gauche de y, probabilité indépendante du mouvement antérieur de la particule. Nous supposons $F(x, y; s, t)$ mesurable (B) par rapport à x


PENGUIN CLASSICS
Alfred Döblin
Berlin Alexanderplatz


BERLIN
1980
THE CRITERION COLLECTION
REMASTERED · A FILM IN 13 PARTS & AN
BY RAINER

Revealed: the maths genius of the Maginot line

By Julian Coman
28 January 2001 • 12:00am

A YOUNG French soldier who died during the Second World War has been acclaimed as a mathematical genius, 60 years after scribbling down formulae while serving on the Maginot line.

Vincent Doblin came from a German-Jewish family that fled Berlin for France in the 1930s as anti-semitism spread through the Third Reich. His story is now being celebrated as an example of intellectual heroism in the most extreme circumstances. A promising mathematician when Hitler invaded France, Mr Doblin's career was cut short tragically in 1940. At the age of 25, he committed suicide rather than surrender to the invading German troops.

The recovery last summer of wartime papers, which he had written as he fought in the Ardennes, has established the young soldier as one of the most important figures in modern mathematics and the founder of contemporary probability theory. The discovery of Mr Doblin's work has caused a sensation in French intellectual circles. This month, in an unprecedented "special edition" of its journal, the French Academy of Sciences has published the entirety of the recovered research, in order to "pay homage to his genius".

According to one of France's most prominent historians of mathematics, Prof Bernard Bru, the Doblin papers "provide a bridge between mathematical analysis before the Second World War, and the advances in probability theory by the Japanese in the 1950s". Mr Doblin's research concerned one of the most important areas of applied mathematics - predicting the outcome of events that are subject to random disturbances, for example the spread of particles through fluids such as water.

Remarques sur la théorie métrique des fractions continues

par

W. Doeblin [1])

Paris

But de ce travail: Les propriétés métriques des développements en fractions continues furent étudiées après Gauß par plusieurs mathématiciens éminents, en particulier par MM. Borel, Kuzmin, P. Lévy, Khintchine et Denjoy. Nous nous proposons de montrer qu'une certaine partie des résultats obtenus peut être démontrée assez facilement en utilisant la théorie des chaînes à liaisons complètes. Nous appliquons donc les méthodes et théorèmes du Calcul des Probabilités; il n'est peut-être pas superflu de remarquer que cet usage n'implique aucun abandon de rigueur. Nous traduirons quelquefois le langage des probabilités dans le langage de la théorie de la mesure. La méthode employée permettra de préciser sur quelques points les résultats connus.

§ 1. La formule de Gauß. Application de la théorie des chaines à liaisons complètes.

Notations, la formule fondamentale de Borel-Lévy. Soit x un nombre irrationel compris entre 0 et 1,

$$x = \frac{1}{x_1},\ x_1 = a_1 + \frac{1}{x_2}, \ldots, x_n = a_n + \frac{1}{x_{n+1}},$$

a_n est la partie entière de x_n, x_n (a_n) est le n-ième quotient complet (incomplet). Tous les x_n sont > 1, les $a_n \geqq 1$, pour n'importe quel nombre irrationnel le développement est illimité. x est une fonction homographique de x_n de la forme

$$x = \frac{P_n x_n + P_{n-1}}{Q_n x_n + Q_{n-1}},$$

---

[1]) ,,M. Doeblin m'a remis ce travail il y a plus d'un an; probablement à la fin de 1937. C'est par suite d'un malentendu, dont je suis sans doute le principal

In 1938, Wolfgang Doeblin outlined a remarkable limiting law for these approximation coefficients. He showed that for Lebesgue-almost every $\theta \in (0,1)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} F\big(q_k|\theta q_k - p_k|\big) = \int_0^1 F(z)\, d\nu(z)$$

for any bounded continuous function $F : [0,1] \to \mathbb{R}$, where ν is the probability measure on $[0,1]$ with density

$$\frac{d\nu}{dz} = \begin{cases} \dfrac{1}{\ln 2}, & 0 \le z \le 1/2, \\ \dfrac{1 - 1/z}{\ln 2}, & 1/2 < z \le 1. \end{cases}$$

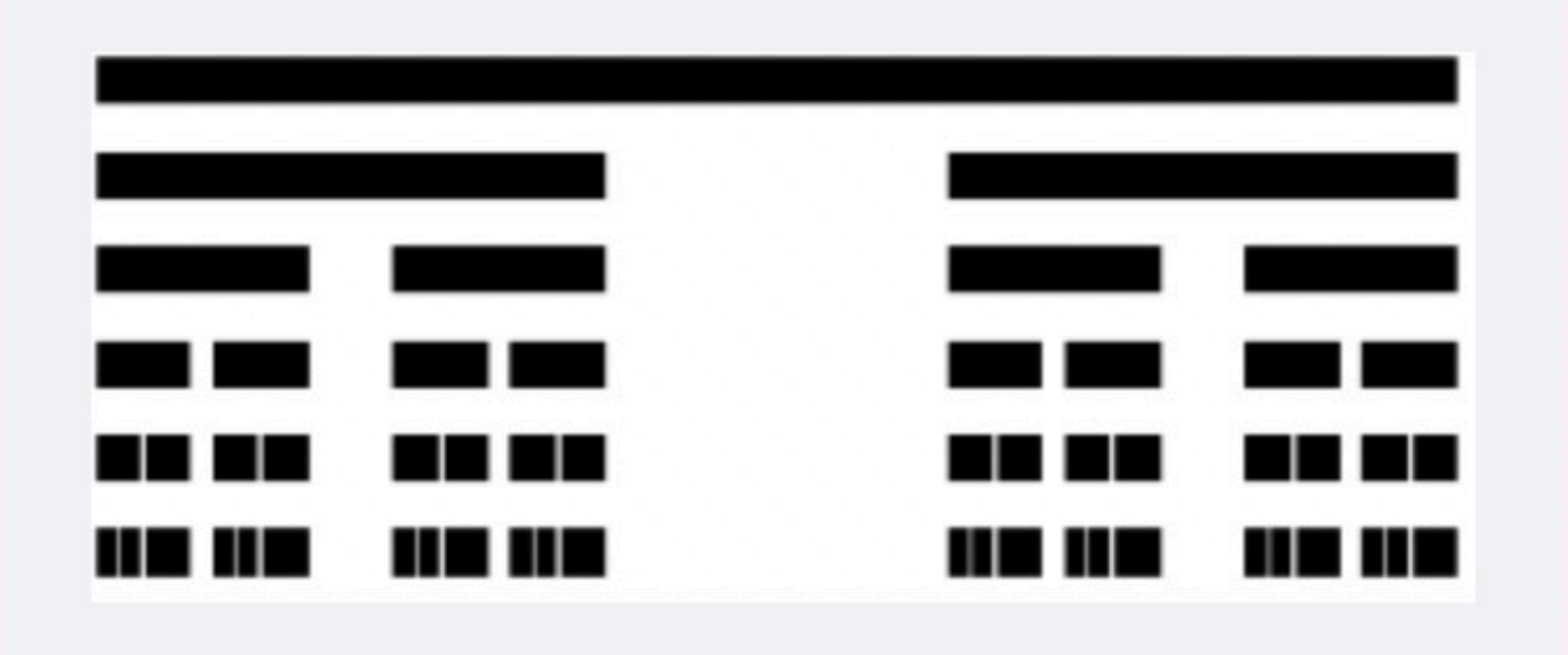
- Doeblin provided a sketch of the proof of this result
- His work remained largely unnoticed until Hendrik Lenstra independently conjectured the same result decades later
- The law is now known as the *Doeblin-Lenstra law*
- A full proof was provided in 1983 by Bosma, Jager, and Wiedijk, using ergodic theory applied to the natural extension of the Gauss map

Effective Doeblin-Lenstra

Theorem (Aggarwal-Ghosh 25): Let $F : \mathbb{R} \to \mathbb{R}$ be differentiable with bounded derivative. Then for any $\varepsilon > 0$, for Lebesgue-almost every θ,

$$\frac{1}{N}\sum_{k=1}^{N} F\big(q_k|\theta q_k - p_k|\big) = \int_0^1 F(z)\, d\nu(z) + O\big(N^{-1/2}\log^{3/2+\varepsilon} N\big).$$

Mahler's questions



Mahler asked the following question. "How close can irrational elements of Cantor's set be approximated by rational numbers

(1) in Cantor's set, and

(2) by rational numbers not in Cantor's set?"

- Does a Lévy-Khintchine theorem hold for almost every point on **middle third Cantor set** with respect to $\log 2/\log 3$-dimensional Hausdorff measure? With a rate?

- Does a Döblin-Lenstra theorem hold for almost every point on **middle third Cantor set** with respect to $\log 2/\log 3$-dimensional Hausdorff measure? With a rate?

Fractal Doeblin-Lenstra

Theorem (Aggarwal-Ghosh 25): Let μ be a non-atomic self-similar measure on $\mathbb{R}$ generated by similarities with a common contraction ratio (e.g., the Hausdorff measure on the middle-third Cantor set). Let $F : \mathbb{R} \to \mathbb{R}$ be differentiable with bounded derivative. Then for any $\varepsilon > 0$, for μ-almost every θ,

$$\frac{1}{N}\sum_{k=1}^{N} F\big(q_k|\theta q_k - p_k|\big) = \int_0^1 F(z)\,d\nu(z) + O\big(N^{-1/2}\log^{3/2+\varepsilon} N\big).$$