Strange Nested Square Roots

Nguyen Vu Hung

November 2, 2025

1 Introduction

This document explores the fascinating properties of the discrete dynamical system defined by the recurrence relation

$$x_{k+1} = 2x_k\sqrt{1 - x_k^2}$$

with initial condition $x_0 \in [0, 1]$. We investigate several fundamental questions:

- Finding initial conditions that yield rational number sequences
- Identifying and characterizing the invariant measure (a probability distribution that is preserved under the transformation) of this system
- Establishing connections to normal numbers (numbers whose digits are uniformly distributed in every base) via the dyadic map
- Demonstrating ergodicity and explaining its significance
- Exploring connections to machine learning and artificial intelligence

The study of this system reveals deep connections between number theory, dynamical systems, ergodic theory (the mathematical study of systems where time averages equal space averages), and probabilistic methods that are fundamental to modern machine learning.

2 The Problem

We consider the discrete dynamical system defined by the recurrence relation:

$$x_{k+1} = 2x_k\sqrt{1 - x_k^2}, \quad x_0 \in [0, 1]$$

The main problems of interest are:

1. Find an x_0 such that all x_k are rational numbers.
2. Find the main **invariant measure** of this dynamical system. Explain what it means, and find related systems and their main invariant measure.

3. Can there be more than one invariant measure?

4. Establish the connection to **normal numbers** via the **dyadic map** (a related dynamical system).

5. Show that these systems are **ergodic**, and explain this concept and the role that it plays.

2.1 Solution Outline

Let (p, r, q) be a **Pythagorean triple**, that is, p, q, r are integers with $p^2 + r^2 = q^2$. If $x_k = \frac{p}{q}$, then it is easy to show that

$$x_{k+1} = \frac{p'}{q'}$$

with $p' = 2p\sqrt{q^2 - p^2}$, $q' = q^2$, $r' = q^2 - 2p$, defines a new Pythagorean triple (p', q', r').

So $x_0 = \frac{3}{5}$ works, answering the first question.

The invariant measure of the mapping $x \mapsto 2x\sqrt{1 - x^2}$ on $[0, 1]$ is given by the probability density function:

$$f_X(x) = \frac{2}{\pi} \frac{1}{\sqrt{1 - x^2}}, \quad 0 \leq x \leq 1$$

If you start with almost any $x_0 \in [0, 1]$, the sequence (x_k) is aperiodic and the **empirical distribution** of the successive values has a density approaching this invariant measure as the number of terms increases. There are exceptions: for instance, if x_0 is such that $x_0 = x_{10}$ (i.e., a periodic point), the sequence will not follow the main invariant measure.

Likewise, in the dyadic map $x_{k+1} = 2x_k - \lfloor 2x_k \rfloor$, if x_0 is a rational number, the sequence is periodic and will not follow the main invariant measure: the uniform distribution on $[0, 1]$. Otherwise, x_0 is called a **normal number**.

The ergodic property is used in proving these results. It states that you can retrieve the invariant measure using one infinite sequence with a single seed x_0, or using infinitely many very short sequences (x_0, x_1), each one with a random seed x_0.

3 The Sequence 345

Starting with $x_0 = \frac{3}{5}$ from the Pythagorean triple $(3, 4, 5)$, we obtain a sequence of rational numbers.

Using the recurrence relation: $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$

For a Pythagorean triple (p, r, q) where $p^2 + r^2 = q^2$, if $x_k = \frac{p}{q}$, then:

$$x_{k+1} = \frac{2pr}{q^2}$$

3.1 Sequence Values

Starting with $x_0 = \frac{3}{5}$ (from Pythagorean triple $(3, 4, 5)$), we obtain:

$$
\begin{aligned}
x_0 &= \frac{3}{5} \quad \text{(from Pythagorean triple } (3, 4, 5))\\
x_1 &= \frac{24}{25} \quad \text{(from Pythagorean triple } (24, 7, 25))\\
x_2 &= \frac{336}{625} \quad \text{(from Pythagorean triple } (336, 527, 625))\\
x_3 &= \frac{354144}{390625} \quad \text{(from Pythagorean triple } (354144, 164833, 390625))\\
x_4 &= \frac{116749235904}{152587890625} \quad \text{(from Pythagorean triple } (116749235904, 98248054847, 152587890625))
\end{aligned}
$$

This demonstrates that the recurrence preserves the property of being a rational number when starting from a rational initial condition derived from a Pythagorean triple.

4 Closed-Form Solution

Remarkably, the recurrence relation $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$ admits a closed-form solution using trigonometric functions. This elegant representation reveals the underlying structure of the system and provides a direct method to compute any term in the sequence without iteration.

4.1 Derivation Using Trigonometric Substitution

The key insight is to use the trigonometric substitution $x_k = \sin(\theta_k)$. Under this transformation, the recurrence relation becomes:

$$\sin(\theta_{k+1}) = 2\sin(\theta_k)\sqrt{1 - \sin^2(\theta_k)} = 2\sin(\theta_k)\cos(\theta_k) = \sin(2\theta_k)$$

This immediately implies that $\theta_{k+1} = 2\theta_k$ (modulo periodicity considerations), which is the doubling map on the circle.

Note: The doubling map $\theta \mapsto 2\theta \pmod{2\pi}$ on the circle is chaotic. This chaotic behavior is fundamentally related to the fact that π is irrational. Since π is irrational, the doubling map exhibits ergodic and mixing properties: typical trajectories densely fill the circle, making long-term prediction impossible despite the deterministic nature of the system. This irrationality ensures that the map cannot settle into periodic orbits for typical initial conditions, which is precisely what makes the system chaotic.

Starting with $x_0 = \sin(\theta_0)$, where $\theta_0 = \arcsin(x_0)$, we obtain:

$$
\begin{aligned}
x_0 &= \sin(\theta_0)\\
x_1 &= \sin(2\theta_0)\\
x_2 &= \sin(4\theta_0) = \sin(2^2\theta_0)\\
x_3 &= \sin(8\theta_0) = \sin(2^3\theta_0)\\
&\vdots\\
x_k &= \sin(2^k\theta_0)
\end{aligned}
$$

4.2 Explicit Formula

The closed-form solution for the strange nested square roots recurrence is:

$$x_k = \sin(2^k \arcsin(x_0))$$

or equivalently, if we define $\theta_0 = \arcsin(x_0)$:

$$x_k = \sin(2^k \theta_0), \quad \text{where } \theta_0 = \arcsin(x_0)$$

4.3 Implications and Properties

The closed-form solution provides several important insights:

1. **Direct Computation:** Any term x_k can be computed directly from x_0 without iterating through all intermediate values, which is particularly useful for computing terms with very large indices.

2. **Geometric Interpretation:** The sequence corresponds to repeatedly doubling an angle on the unit circle and taking the sine. This geometric picture helps visualize why the system exhibits chaotic behavior and why it explores the entire interval $[0, 1]$ aperiodically.

3. **Periodicity:** If θ_0 is a rational multiple of π (i.e., $\theta_0 = \frac{p}{q}\pi$ for some integers p, q), then the sequence is periodic. Otherwise, the sequence is dense in $[0, 1]$ and exhibits chaotic behavior.

4. **Connection to Doubling Map:** The closed-form confirms that our map is conjugate to the doubling map $\theta \mapsto 2\theta$ on the circle, which explains many of the system's properties, including the invariant measure and Lyapunov exponent.

5. **Rational Initial Conditions:** For the special case $x_0 = \frac{3}{5}$ discussed earlier, we have $\theta_0 = \arcsin(3/5)$. While this yields a closed-form expression, the values remain rational due to the Pythagorean triple structure, demonstrating an interesting intersection between the algebraic and trigonometric representations.

The existence of this closed-form solution is a beautiful example of how seemingly complex nonlinear recurrences can sometimes be reduced to simple trigonometric functions through appropriate coordinate transformations. This transformation not only simplifies computation but also reveals the deep geometric and dynamical structure underlying the system.

5 The Invariant Measure

The invariant measure of the mapping $T(x) = 2x\sqrt{1-x^2}$ on the interval $[0, 1]$ is given by the density function

$$p(x) = \frac{2}{\pi\sqrt{1-x^2}}$$

with respect to the Lebesgue measure, i.e.,

$$d\mu(x) = \frac{2}{\pi\sqrt{1-x^2}}dx$$

5.1 Verification Using the Perron-Frobenius Operator

An invariant density $p(x)$ for a map $T(x)$ must satisfy the equation

$$p(y) = \sum_{x \in T^{-1}(y)} \frac{p(x)}{|T'(x)|}$$

where the sum is over all preimages of y.

5.1.1 Step 1: Find the Preimages

Preimages are the values of x that map to a given y under the transformation T. In other words, they are the solutions to $T(x) = y$.

For a given $y \in (0, 1)$, the equation $y = 2x\sqrt{1 - x^2}$ has two solutions for x, which are the preimages of y:

$$x_1 = \sqrt{\frac{1 + \sqrt{1 - y^2}}{2}} \quad \text{and} \quad x_2 = \sqrt{\frac{1 - \sqrt{1 - y^2}}{2}}$$

5.1.2 Step 2: Calculate the Derivative

The derivative of the map is:

$$T'(x) = \frac{2 - 4x^2}{\sqrt{1 - x^2}}$$

5.1.3 Step 3: Substitute into the Transfer Operator Equation

If we propose the invariant density $p(x) = \frac{C}{\sqrt{1-x^2}}$ for some constant C, we find that for the two preimages x_1 and x_2, the sum is:

$$\frac{p(x_1)}{|T'(x_1)|} + \frac{p(x_2)}{|T'(x_2)|} = \frac{C}{2\sqrt{1 - y^2}} + \frac{C}{2\sqrt{1 - y^2}} = \frac{C}{\sqrt{1 - y^2}}$$

The result of the sum is exactly the form of the proposed density, $p(y)$, which confirms that it is an invariant density.

5.1.4 Step 4: Normalize the Density

To be a valid probability measure, the integral of the density over the domain $[0, 1]$ must equal 1:

$$\int_0^1 \frac{C}{\sqrt{1 - x^2}} dx = C[\arcsin(x)]_0^1 = C(\arcsin(1) - \arcsin(0)) = C(\pi/2 - 0) = \frac{C\pi}{2}$$

Setting this equal to 1 gives $C = \frac{2}{\pi}$.
Therefore, the invariant density is

$$p(x) = \frac{2}{\pi\sqrt{1 - x^2}}$$

This map is a specific case of a family of maps known as **Chebyshev maps** (dynamical systems defined by Chebyshev polynomials, which are a family of orthogonal polynomials defined recursively as $T_n(\cos\theta) = \cos(n\theta)$ and which appear as trigonometric polynomial maps that exhibit chaotic behavior), for which this measure is a well-known result.

5.2 Numerical Verification

To verify the theoretical result, we numerically simulate the sequence starting from an initial condition $x_0 = \frac{3}{4}$ and compute the empirical density. The simulation generates 100,000 iterations using the recurrence relation $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$, and the empirical distribution of the values is computed using a histogram.

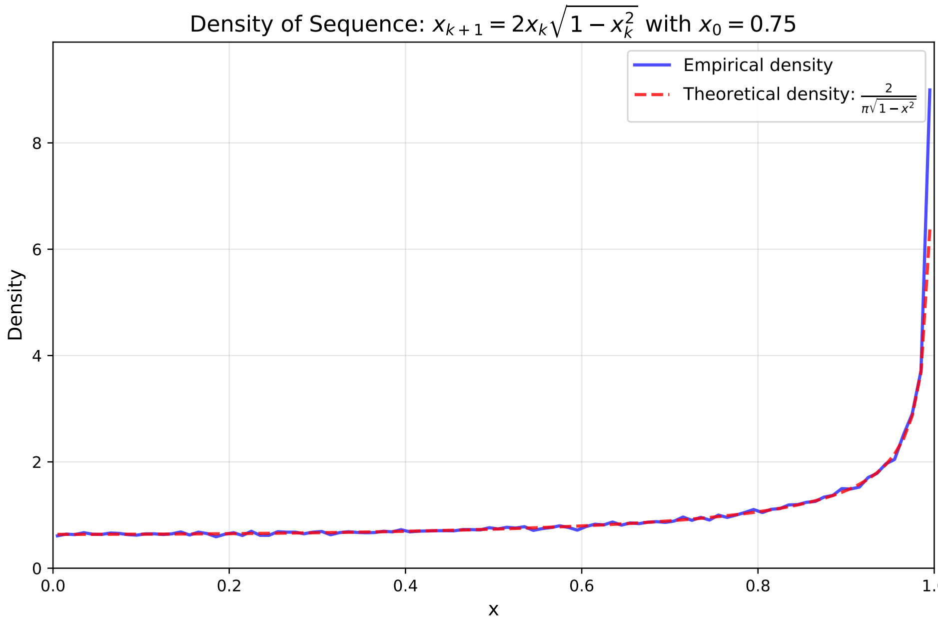

Figure 1: Empirical density of the sequence $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$ with initial condition $x_0 = \frac{3}{4}$, compared with the theoretical invariant density $p(x) = \frac{2}{\pi\sqrt{1-x^2}}$. The simulation shows excellent agreement between the empirical distribution (blue line) and the theoretical prediction (red dashed line), confirming that the invariant measure is indeed given by the stated formula.

Figure 1 shows the empirical density obtained from the simulation alongside the theoretical invariant density. The excellent agreement between the two demonstrates that for typical initial conditions, the long-term empirical distribution of the sequence converges to the theoretical invariant measure, as predicted by ergodic theory.

6 The Lyapunov Exponent

The **Lyapunov exponent** is a fundamental quantity in dynamical systems theory that measures the sensitivity to initial conditions. It quantifies how fast two nearby trajectories diverge or converge over time, providing a mathematical characterization of chaotic behavior.

6.1 Definition and Interpretation

For a discrete one-dimensional map $x_{n+1} = f(x_n)$, the Lyapunov exponent λ measures the exponential rate of separation of initially close trajectories. If we start with two close initial points x_0 and $x_0 + \varepsilon$, the distance after n steps is approximately:

$$|\delta x_n| \approx |\delta x_0| e^{n\lambda}$$

The Lyapunov exponent is defined as:

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |f'(x_i)|$$

where $f'(x_i)$ is the derivative of f at the i-th iterate.

The value of λ determines the system's behavior:

- **$\lambda > 0$:** Trajectories diverge exponentially; the system exhibits chaotic behavior and is sensitive to initial conditions.
- **$\lambda = 0$:** Neutral stability; trajectories neither converge nor diverge exponentially.
- **$\lambda < 0$:** Trajectories converge; the system is stable and insensitive to initial conditions.

6.2 Lyapunov Exponent for the Strange Nested Square Roots Map

For our map $T(x) = 2x\sqrt{1-x^2}$, we can compute the Lyapunov exponent using the invariant measure and the ergodic property. The derivative of the map is:

$$T'(x) = \frac{2-4x^2}{\sqrt{1-x^2}}$$

Since the system is ergodic with invariant density $p(x) = \frac{2}{\pi\sqrt{1-x^2}}$, the Lyapunov exponent can be computed as the space average:

$$\lambda = \int_0^1 \ln |T'(x)|\, p(x)\, dx = \frac{2}{\pi} \int_0^1 \frac{\ln |T'(x)|}{\sqrt{1-x^2}}\, dx$$

A more elegant approach uses the trigonometric substitution $x = \sin\theta$. Under this transformation, we find that $T(\sin\theta) = 2\sin\theta\cos\theta = \sin(2\theta)$, which reveals that our map is conjugate to the doubling map $\theta \mapsto 2\theta$ on the circle $[0, \pi/2]$.

The doubling map on the circle has a well-known Lyapunov exponent of $\ln 2$. Since Lyapunov exponents are preserved under conjugacy (smooth coordinate transformations), our map inherits this same Lyapunov exponent.

6.3 Result for the Strange Nested Square Roots System

For the map $x_{k+1} = 2x_k\sqrt{1-x_k^2}$, the Lyapunov exponent is:

$$\lambda = \ln 2 \approx 0.693$$

This positive Lyapunov exponent confirms that the system exhibits **chaotic behavior**. The exponential divergence of nearby trajectories means that even tiny differences in initial conditions will grow exponentially over time, making long-term prediction impossible despite the deterministic nature of the system.

The fact that $\lambda = \ln 2$ is consistent with the observation that this map is related to the Chebyshev polynomial of degree 2 and is conjugate to the doubling map. This connection explains why the system shows strong mixing properties and why typical trajectories explore the entire interval $[0, 1]$ in an aperiodic manner, leading to the invariant measure we identified earlier.

7 The Dyadic Map

A dyadic map is a simple but important mathematical function in chaos theory and dynamical systems. It's defined as

$$f(x) = 2x - \lfloor 2x \rfloor$$

where $\lfloor 2x \rfloor$ is the floor function, which gives the greatest integer less than or equal to $2x$. The map is also called the **2x mod 1** map because it is equivalent to $f(x) = 2x \pmod 1$. This map takes any number in the interval $[0, 1)$ and stretches it by a factor of two, then takes the fractional part, effectively folding the interval back onto itself.

The name "dyadic" refers to **base-2** or binary numbers. The dyadic map is closely linked to the binary representation of a number. If a number x is represented in binary as $0.b_1b_2b_3...$, applying the dyadic map, $f(x)$, is equivalent to a **left shift** of the binary digits. For example, if $x = 0.10110...$ in binary, then $2x = 1.0110...$, and $f(x) = 0.0110....$ The first digit, b_1, is "shifted" to the left of the decimal point and removed, and the rest of the digits follow.

7.1 Dyadic Map and Normal Numbers

The dyadic map provides a clear connection to the concept of **normal numbers**. A number is **normal** in base 2 if, in its binary representation, every finite sequence of digits appears with the same limiting frequency as every other sequence of the same length. The dyadic map helps prove that for almost all initial values $x_0 \in [0, 1)$, the sequence generated by iterating the map, (x_k), has a **uniform distribution**. This means that over a long period, the values of x_k will be evenly distributed across the interval $[0, 1)$.

- If x_0 is a **rational number** its binary expansion is either terminating or repeating. Iterating the dyadic map on a rational number leads to a **periodic sequence** of values, meaning it will eventually repeat itself. For example, starting with $x_0 = 1/3 = 0.010101..._2$ leads to the sequence $1/3, 2/3, 1/3, 2/3, ...$ which does not uniformly fill the interval.

- If x_0 is an **irrational number**, the sequence of values will not repeat. For most irrational numbers, the empirical distribution of the sequence (x_k) will approach the uniform distribution on $[0, 1]$. A number is normal in base 2 if and only if the sequence of its dyadic map iterations is uniformly distributed. This is a powerful result, connecting a property of a number's digits to the behavior of a dynamical system.

The **main invariant measure** for the dyadic map is the **uniform distribution on** $[0, 1)$. This is the distribution that is preserved by the map. If you take a collection of points distributed uniformly and apply the dyadic map, the new collection of points will also be uniformly distributed.

7.2 Ergodicity of the Dyadic Map

The dyadic map is an **ergodic system** (a dynamical system where time averages equal space averages, meaning a single trajectory's long-term behavior reflects the statistical properties of the entire system). In simple terms, ergodicity means that the long-term time average of a single trajectory of the system is the same as the average over all possible starting points at a single moment in time.

For the dyadic map, this means:

1. **Time average:** If you take a single initial value x_0 (as long as it's not a rational number) and iterate the dyadic map to generate a long sequence $(x_0, x_1, x_2, ...)$, the average behavior of this single sequence will statistically reflect the entire system. For instance, the frequency of values falling into any subinterval of $[0, 1)$ will be proportional to the length of that subinterval.

2. **Ensemble average:** The average over all possible initial values in $[0, 1)$ at a given time is also a uniform distribution.

The ergodicity of the dyadic map allows us to study the behavior of a typical infinite sequence starting from a single number and draw conclusions about the properties of almost all numbers. For example, it helps to prove that almost all numbers are normal, even though we can't explicitly construct many of them.

More than one invariant measure can exist. For the dyadic map, the uniform distribution is the **main** or **unique absolutely continuous invariant measure** (a measure that has a probability density function). However, there are also other invariant measures, such as **discrete invariant measures**. For instance, a measure that assigns probability 1 to a single fixed point (e.g., $x = 0$) is an invariant measure. If the system is **ergodic**, it means that any invariant measure can be decomposed into a sum of ergodic measures.

8 The Logistic Map and the Mandelbrot Set

The Mandelbrot set and the logistic family are mathematically related through a change of variable that connects the classic quadratic iteration in the Mandelbrot set to the logistic map's difference equation. This relationship reveals how quadratic complex dynamics and real population models share universal routes to chaos and fractal structures.

8.1 The Mandelbrot Set Iteration

The Mandelbrot set uses the quadratic map:

$$z_{n+1} = z_n^2 + c$$

where z and c are generally complex numbers. The Mandelbrot set consists of all values of c for which the sequence starting with $z_0 = 0$ remains bounded.

8.2 The Logistic Map

The logistic family is given by:

$$x_{n+1} = rx_n(1 - x_n)$$

where r is a real-valued parameter and x_n lies in the interval $[0, 1]$. This map is a classic example in chaos theory and population dynamics, exhibiting a rich bifurcation structure as the parameter r varies.

8.3 Mathematical Transformation

A change of variables connects these two systems. Specifically:

$$z = r\left(\frac{1}{2} - x\right), \quad c = \frac{r}{2}\left(1 - \frac{r}{2}\right)$$

If you substitute and expand, you can recast the quadratic map of the Mandelbrot set into the logistic form (or vice versa). Both systems generate bifurcation phenomena and chaotic dynamics, and the bifurcation diagram of the logistic map is essentially a cross-section (along the real axis) of the Mandelbrot set.

For certain values of c in the Mandelbrot set, the iterative behavior (fixed points, cycles, chaos) matches the logistic map's bifurcation diagram. The Mandelbrot set encodes all the interesting dynamics of the logistic map within its "middle section" — that is, along the real axis of c.

8.4 Visual Correspondence

Plotting the long-term behavior of both models reveals that the bifurcation diagram of the logistic map lines up with features on the real axis of the Mandelbrot set. This deep connection illustrates how quadratic complex dynamics (Mandelbrot) and real population models (logistic) share universal routes to chaos and fractal structures.

The correspondence between the quadratic family (Mandelbrot set) and logistic map highlights their shared mathematical structure and intertwining bifurcation features. Both systems exhibit:

- Period-doubling bifurcations leading to chaos
- Universal scaling properties
- Rich fractal structures in their parameter spaces

8.5 Connection to the Strange Nested Square Roots Map

The strange nested square roots map $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$ studied in this document is directly related to the logistic map and Mandelbrot set through its identity as a **Chebyshev map** of degree 2.

8.5.1 Chebyshev Map Structure

As established earlier, our map $T(x) = 2x\sqrt{1 - x^2}$ is a specific case of the family of Chebyshev maps. Using the trigonometric substitution $x = \sin\theta$, we find that:

$$T(\sin\theta) = 2\sin\theta\cos\theta = \sin(2\theta)$$

This reveals that our map is conjugate to the doubling map $\theta \mapsto 2\theta$ on the circle, which is a fundamental example in dynamical systems theory.

8.5.2 Relationship to Quadratic Maps

Chebyshev maps belong to the same universality class as quadratic polynomial maps. In particular, Chebyshev maps can be transformed into quadratic polynomial form through appropriate coordinate changes. This means that:

- The strange nested square roots map shares the same universal properties as the logistic map and Mandelbrot iterations
- All three systems exhibit period-doubling routes to chaos with the same Feigenbaum constants
- They share similar bifurcation structures and scaling laws

8.5.3 Fixed Parameter Interpretation

Unlike the logistic map (which depends on the parameter r) or the Mandelbrot set (which depends on the complex parameter c), our map has no free parameters. This fixed form corresponds to a specific point in the parameter space of quadratic maps, specifically in the **fully chaotic regime**.

The Lyapunov exponent $\lambda = \ln 2$ confirms that our map operates in the chaotic regime, where the system exhibits maximal sensitivity to initial conditions and ergodic behavior.

8.5.4 Unifying Framework

The connection between these systems reveals a unifying framework:

- **Mandelbrot set:** $z_{n+1} = z_n^2 + c$ (complex quadratic family with parameter c)
- **Logistic map:** $x_{n+1} = rx_n(1 - x_n)$ (real quadratic family with parameter r)
- **Strange nested square roots:** $x_{k+1} = 2x_k\sqrt{1 - x_k^2}$ (Chebyshev map, fixed parameter, fully chaotic)

All three belong to the quadratic iteration universality class. The Mandelbrot set provides a complete picture of all possible behaviors, with the logistic map representing the real-axis cross-section, and our Chebyshev map representing a specific fully-chaotic instance within this framework.

This deep connection explains why our map exhibits chaotic behavior, why it has the invariant measure $p(x) = \frac{2}{\pi\sqrt{1-x^2}}$, and why it shares universal scaling properties with other quadratic dynamical systems. The mathematical structure underlying these connections provides a rich theoretical foundation for understanding chaos, ergodicity, and the statistical properties of deterministic systems.

9 The Ergodic Theorem

The **Ergodic Theorem** is one of the most fundamental results in dynamical systems theory and statistical mechanics. It provides a bridge between the microscopic behavior of individual trajectories and the macroscopic statistical properties of a system.

9.1 Mathematical Formulation: Birkhoff's Ergodic Theorem

Theorem (Birkhoff, 1931): Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving dynamical system, where:

- X is a measurable space
- $\mathcal{B}$ is a σ-algebra
- μ is a probability measure
- $T : X \to X$ is a measure-preserving transformation

For any integrable function $f : X \to \mathbb{R}$ (i.e., $\int |f| d\mu < \infty$), the **time average**

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

exists for μ-almost every $x \in X$.

If the system is **ergodic**, then this time average equals the **space average**:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int_X f \, d\mu$$

for μ-almost every $x \in X$.

9.2 Intuitive Explanation

The Ergodic Theorem tells us that in well-behaved dynamical systems:

1. **Time Average = Space Average:** The long-term behavior of a single trajectory mirrors the overall statistical behavior of the entire system.
2. **Individual vs. Ensemble:** Instead of studying the entire ensemble of possible states, we can understand the system's properties by following one typical trajectory for a long time.
3. **Statistical Regularity:** Even though individual trajectories may appear chaotic or unpredictable, their long-term statistical properties are deterministic and predictable.

9.3 Key Concepts

9.3.1 Measure-Preserving Transformation

A transformation T is **measure-preserving** if $\mu(T^{-1}(A)) = \mu(A)$ for all measurable sets A. This means the transformation preserves the "volume" or "probability" of regions.

9.3.2 Ergodicity

A measure-preserving system is **ergodic** if it cannot be decomposed into smaller invariant components. Mathematically, if $T^{-1}(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

Intuitively, ergodicity means:

- The system is "irreducible" — you can't break it into disconnected parts
- Trajectories are "mixing" — they eventually visit all regions of the space
- The system has no non-trivial conserved quantities

9.4 The Dyadic Map and Normal Numbers

The **dyadic map** provides an elegant and powerful framework from dynamical systems theory to establish the existence of **normal numbers**. The connection is a direct consequence of the **Ergodic Theorem**.

If we write a number x in binary as $x = 0.b_1b_2b_3...$, where each b_i is a 0 or a 1, then applying the map T to x simply shifts its binary digits one position to the left, dropping the first digit.

By applying the Ergodic Theorem, we can connect time averages to space averages:

$$\text{Time Average} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) = \int_0^1 f(x)\,dx = \text{Space Average}$$

The time average represents the frequency of a digit (or block of digits) in the binary expansion of x. The space average represents the expected probability of that digit (or block) in a uniform distribution.

The theorem guarantees that for "almost all" numbers (meaning all numbers except for a set of measure zero, like the rationals), the frequency of any binary digit (0 or 1) will be exactly 1/2. This can be extended to show that the frequency of any block of binary digits of length k will be exactly $1/2^k$. This is precisely the definition of a **binary normal number**.

10 Conclusion: Connections to Machine Learning and AI

The study of dynamical systems, invariant measures, and ergodic theory has profound implications for machine learning and artificial intelligence. These connections manifest in several key areas:

10.1 Sampling and Monte Carlo Methods

The ergodic property — that time averages equal space averages — is fundamental to **Monte Carlo methods**, which are ubiquitous in machine learning. When we use Markov Chain Monte Carlo (MCMC) methods to sample from complex probability distributions, we rely on the ergodic properties of the Markov chain to ensure that the long-term behavior of a single trajectory accurately represents the target distribution. This is exactly what the Ergodic Theorem guarantees: by following one trajectory for a sufficiently long time, we can recover the invariant measure.

10.2 Normalization and Representation Learning

The connection between normal numbers and the dyadic map reveals fundamental properties about how numbers are represented in different bases. In machine learning, we often work with numerical representations in various forms (binary, decimal, floating-point, etc.). Understanding the ergodic properties of these representations helps us design better normalization schemes and understand the statistical properties of data transformations.

10.3 Invariant Representations

The concept of an **invariant measure** is directly related to learning **invariant representations** in deep learning. Just as the dynamical system preserves its invariant measure under transformation, neural networks often seek to learn representations that are invariant to certain transformations (e.g., translation, rotation, scaling). The mathematical framework of invariant measures provides a theoretical foundation for understanding why and how such representations can be learned.

10.4 Stochastic Processes and Neural Networks

Recurrent neural networks (RNNs) and their variants can be viewed as discrete dynamical systems. Understanding the ergodic properties and invariant measures of such systems helps us:

- Analyze the long-term behavior of RNNs
- Design stable architectures that avoid vanishing or exploding gradients
- Understand how networks preserve or forget information over time

10.5 Generative Models

Generative models, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), learn to approximate probability distributions. The theory of invariant measures provides insights into:

- How generators learn to produce samples from target distributions
- The convergence properties of training algorithms
- Why certain architectures or training procedures lead to stable distributions

10.6 Theoretical Guarantees

The mathematical rigor of ergodic theory provides **theoretical guarantees** about the behavior of learning algorithms. Just as the Ergodic Theorem guarantees that time averages converge to space averages for ergodic systems, we can prove convergence guarantees for machine learning algorithms that rely on iterative updates and sampling.

10.7 Future Directions

The deep connections between ergodic theory and machine learning suggest several promising research directions:

- Using ergodic theory to analyze and improve optimization algorithms
- Applying invariant measure concepts to understand generalization in deep learning
- Developing new sampling methods based on ergodic properties of neural network dynamics
- Using dynamical systems theory to design more stable and efficient architectures

In conclusion, the strange nested square roots system and its associated mathematical framework of invariant measures, ergodic theory, and normal numbers form a rich theoretical foundation that underlies many practical techniques in modern machine learning and artificial intelligence. Understanding these connections not only deepens our theoretical understanding but also provides guidance for developing more effective and theoretically grounded learning algorithms.

Website: `https://vuhung16au.github.io/`
GitHub: `https://github.com/vuhung16au/`
LinkedIn: `https://www.linkedin.com/in/nguyenvuhung/`