Machine Learning Model Discovery

Connecting Random Walks and Partial Differential Equations

Edward Berman

Advised: Professor Calina Copos

10/22/2022

Abstract

Motivated by Random walk models in biology by Codling, Plank, and Benhamou, I attempt to nurture an understanding of the relationship between Random walk models and Partial Differential Equations. Moreover, I hope to motivate the use of Machine Learning in deriving physical laws. This paper invokes examples and anecdotes from history to make the subject matter more approachable. Assumes undergraduate coursework up to statistics and ordinary partial differential equations, though the standard calculus sequence may prove to be enough to follow along. In this paper I start by developing the essentials of random walk models and partial differential equations. I then connect them through the idea of images and formatting heat data for machine learning.

Contents

Chapter 1

The Necessity of Machine Learning

1.1 Forward

The advent of Machine Learning has provided Physicists with a new tool to derive Physical Laws. Moreover, this tool allows us to derive Physical Laws that we could not before.

The issue is that, Physical Laws are unknown for many systems that are a product of stochastic processes. An analogy may be made between Kepler and Newton. While Kepler was able to describe the elliptical motion of the planets around the sun, it was Newton who derived the universal law of gravitation. In present day, we may find equations to accurately describe motion in cellular dynamics, but the underlying physical laws have alluded us. [1] What makes these physical laws so hard to derive is that we are basing our first principles on stochastic processes. While planets routinely make elliptic orbits around two foci, it is not so simple to model human cells, which may move in a random direction at any given time. In simple cases, we may use analysis to derive a Partial Differential Equation from a stochastic process, but for more complicated processes the task becomes too difficult.

Machine Learning helps us solve this issue. Admittedly, the use of Machine Learning is a bit of a last result, as simpler techniques have been insufficient. For data, we may ask a computer to generate samples of a stochastic process happening over many time steps. We then use a Neural Network to extract a Partial Differential Equation by training on that data. In this way, we are extracting a physical law from first principles for a system driven by stochastics without the need for analysis.

[1] For more on this metaphor, read (1)

Chapter 2

Random Walk Essentials and Brownian Motion

Circa 60 B.C., a Roman Poet named Titus Lucretius Carus Published *De rerum natura*, "the nature of things." In volume 2 of this 6 volume work, he wrote about the "dance of Atoms." He mystifies the nature of this motion in writing

So that we can behold the dust-motes dancing in the sun,
Although the blows that move them can't be seen by anyone.

Contrary to the name, Robert Brown was not the first to observe the issue of particles dancing as if they were alive. However, we may credit him as one of the first to do it scientifically. Robert Brown was a botanist concerned with the jiggling of pollen in water. It troubled him that there was this seemingly random and alive-like motion of pollen particles independent of the surrounding fluid.

Albert Einstein alluded to Brown when he published *On the Motion of Small Particles Suspended in a Stationary Liquid, as Required by the Molecular Kinetic Theory of Heat*, saying that "it is possible that the motions described here are identical to the so-called Brownian molecular motion."

In the following chapter, I make precise the nature of these dancing particles through the theory that Lucretius, Brown, Einstein, and many others thereafter have worked to develop.

2.1 Simple Random Walks

Definition 2.1.1: Simple (Isotropic) Random Walks

A Simple Random Walk may be thought of as a random experiment illustrated by the following properties.

- We start with some particle sitting on the origin of the real line. It's position is updated n times.
- Let us insist the the probability that the particle moves to the right is equal to the probability that the particle moves to the left.
- $X_i := \begin{cases} \delta,\ p = \frac{1}{2} \\ \text{-}\delta,\ p = \frac{1}{2} \end{cases}$, where X_i represents the i_{th} random update in the experiment.

Notice that each X_i acts similarly to an indicator variable[a], and so we may define a Binomial distribution $X_n = \sum_{i=1}^{n} X_i$.

[a] It is most common for indicator variables to take on values of 1 and 0, so I abuse terminology a slight bit to explain a concept

Claim 2.1.1 $p(x,n) = (\frac{1}{2})^n \binom{n}{\frac{n-x}{2}}$, where x and n are even

This proof is left as an exercise for the reader.

From Claim 2.1.1, we find that we have a form of the binomial distribution as we should expect. Moreover, for large n, we have convergence to the normal or Gaussian distribution.

2.2 Biased Random Walks

Definition 2.2.1: Biased Random Walks

A Biased Random Walk differs from a simple random walk in that the probability of moving right is not the same as the probability of moving left. Accordingly, a Biased Random Walk may be thought of as a random experiment illustrated by these slightly different properties.

- We have some particle sitting on the origin of the real line. It's position is updated n times.
- $X_i := \begin{cases} \delta, & r \\ \text{-}\delta, & l \\ 0, & \text{1 - r - l} \end{cases}$, where X_i represents the i_{th} random update in the experiment.

A Biased Random Walk follows a distribution given by X_n, where $X_n = \sum_{i=1}^{n} X_i$. [a]

[a] Notice that if $r, l = \frac{1}{2}$, then we return to the case of a simple random walk

Chapter 3

Partial Differential Equations Essentials

3.1 On Physical Laws

Newton never explicitly wrote $F = ma$ in *Philosophiæ Naturalis Principia Mathematica.* What he wrote was the that the net force is proportional to the change in momentum, $F = k\frac{d(m\cdot v)}{dt}$. Assuming constant mass and SI units, this reduces to $F = ma$. The point is that physical laws are expressed through the language of differential equations. The heat equation, the wave equation, Maxwell's equations, Newton's Second Law, all of these Physical Laws are expressed through the language of Partial Differential Equations. In this chapter we are going to gain an intuition into solving a Partial Differential Equation through an example. We will discuss the various Physical conditions that allow us to reach a final answer and describe the motion that our Partial Differential Equation describes.

3.2 Introduction by way of Example

Consider a heated rod with insulated ends. That is, we have *boundary conditions* $u_x(0,t) = 0 = u_x(L,t)$, $t > 0$

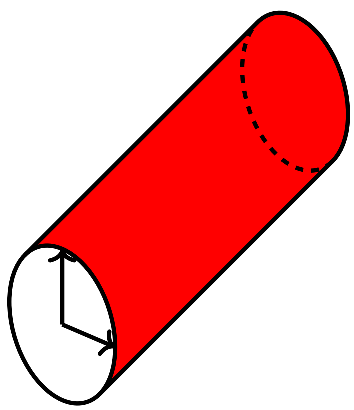

Figure 3.1: Heated Rod with Insulated Ends

The heat equation tells us that

$$\frac{\partial u}{\partial t} = c^2 \frac{\partial u}{\partial x^2}. \tag{1}$$

We guess that $u(x,t) = X(x)T(t)$. Substituting this into (1), we find

$$T'(t)X(x) = c^2 T(t)X''(x).$$

Now, we separate the variables,

$$\frac{T'}{c^2 T} = \frac{X''}{X}.$$

For this to be true, both the right and left hand sides must be equal to a constant. We call this constant the separation constant.

$$\frac{T'}{c^2 T} = \frac{X''}{X} = k$$

$$\begin{cases} X'' - kX = 0 \\ T' - kc^2 T = 0 \end{cases}$$

So far, we have assumed that our partial differential equation has a solution generated by multiplying a function of x by a function of t. This has allowed us to take our Partial Differential Equations and reduce to a series of Ordinary Differential Equations. Not only does this make solving much easier, but it tells us something about the Physics of our system. That is, a change in x does not change the heat u in the same way that a change in t does.

Now, let us revisit our boundary conditions. We are given that the ends are insulated. Thus, $X'(0) = 0 = X'(L)$. We want to choose k so that we only get nontrivial solutions for our equation.

Case 1: $k = \mu^2$

This does not yield any nontrivial solutions.

Case 2: $k = 0$

This yields the nontrivial solution $X_0 = 1$. We also have that $T_0 = a_0$, where a_0 is some constant.

Case 3: $k = -\mu^2$

We find that

$$X = c_1 \cos(\mu x) + c_2 \sin(\mu x). \tag{2}$$

Moreover,

$$X' = -c_1 \mu sin(\mu x) + c_2 \mu cos(\mu x).$$

Now, we use our boundary conditions. $X'(0) = 0$ tells us that $c_2 = 0$.Knowing that $c_2 = 0$ and $X'(L) = 0$ tells us that $\sin(\mu L) = 0$.

$\sin(\mu L) = 0$ has an infinite family of solutions,

$$\mu_n = \frac{n\pi}{L}, n = 1, 2, ...$$

We may now simplify (2),

$$X_n = \cos(\frac{n\pi x}{L}).$$

Recall that $T' - kc^2 T = 0$. Therefore,

$$T' + (\frac{cn\pi}{L}^2)T = 0.$$

So, T has the form

$$T_n = a_n e^{-(\frac{cn\pi}{L})^2 t}, n = 1, 2, ...$$

Recall that we assumed $u(x,t) = X(x)T(t)$, meaning

$$u_n(x,t) = 1(a_0) + a_n e^{-(\frac{cn\pi}{L})^2 t} \cos(\frac{n\pi x}{L}).$$

Recall from Ordinary Differential Equations that solutions to homogeneous equations are linear, [1]

$$u(x,t) = a_0 + \sum_{i=1}^{\infty} a_n e^{-(\frac{cn\pi}{L})^2 t} \cos(\frac{n\pi x}{L}).$$

In the second part of this solution, we have utilized our boundary conditions to determine the constant coefficients for one of our ordinary differential equations (ODE). We then multiplied the solutions to each ODE to create an infinite family of solutions for u(x,t). We now need to determine our coefficients a_0 and a_n. We will do this by considering an initial condition.
$f(x) := u(x,0)$

$$f(x) = a_0 + \sum_{i=1}^{\infty} a_n \cos(\frac{n\pi x}{L}).$$

Note:-

This is an example of something called a Fourier Series, which for our purposes is just an infinite sum of sine and cosine functions.

[1] It is often times advantageous to use vector spaces as a way of encapsulating all of the solutions to a differential equations

Example 3.2.1 (The Heat Equation With Insulated Ends and Initial Condition $f(x) = 5 + \cos(x)$. For this example we also have that $c = 1, L = \pi$)

We may calculate a_0 and a_n using Fourier's Trick or by inspection.

$$5 + \cos(x) = a_0 + \sum_{i=1}^{\infty} a_n \cos(\frac{n\pi x}{L}).$$

Since $L = \pi$,

$$5 + \cos(x) = a_0 + \sum_{i=1}^{\infty} a_n \cos(nx).$$

We find that

$$a_0 = 5$$

$$a_n = \begin{cases} 1, n = 1 \\ 0, n \neq 1 \end{cases}.$$

We conclude that

$$u(x,t) = 5 + e^{-t} \cos(x).$$

In the final step in solving the heat equation, we used the initial condition to uniquely determined our remaining constants.

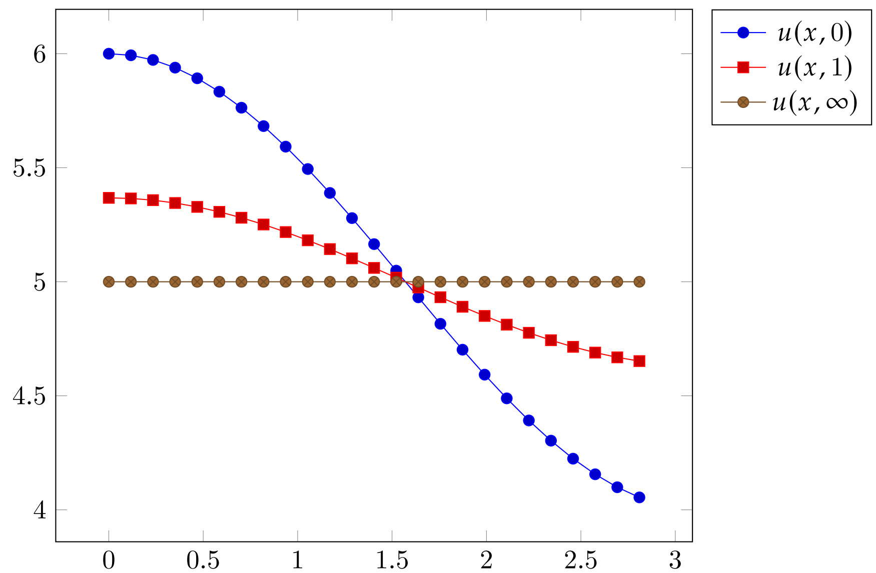

Figure 3.2: Heated Rod over Time

We have some curve, representing a heated rod, where the ends are "flat". Over time the heat distribution converges to the average temperature of $f(x)$, a_0. We see that we still maintain our sinusoidal structure but each point on the curve tends to zero exponentially as time increases.

Chapter 4

Deriving the Heat Equation

Joseph Fourier was the first to publish the heat equation in 1807 in a manuscript presented to the Institute of Paris. In this body of work, he dodged making assumptions about the nature of heat on small scales and instead used conservation principles and boundary conditions to make his argument. This emphasis on boundary conditions helped him in developing his theory on solving partial differential equations using infinite sine series. Moreover, it also accelerated the need for tools in Real Analysis. Joseph Fourier had played with the fire that is infinity and so Mathematicians had an even larger calling to make their notions of infinity rigorous. It is also intriguing how in modern physics and mathematics we've had lots of issues reconciling our macroscopic and models with new models for things on small scales. You may think on a grandiose scale, such as the dilemma with general relativity and quantum mechanics, but similar issues have come up in all sorts of areas, for instance, the chemical gradients of a neuron during an action potential. The Fitzhugh-Nagumo equations were once thought to be successful in capturing the process, only for more careful inspection to reveal that they were saying something physically impossible both according to stochastic models and measurements.

4.1 The Random Walk Derivation

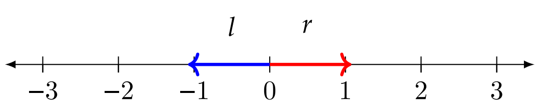

Figure 4.1: Integer Multiples of δ

Recall from definition 2.2.2 that for a given timestep n, we update our position according to a random variable

$$X_i := \begin{cases} \delta, & r \\ \text{-}\delta, & l \\ 0, & \text{1 - r - l.} \end{cases}$$

For reasons that will become clear, let us replace our dummy variable n with t. From this, we see that

$$P(x, t+\tau) = P(x-\delta, t)r + P(x+\delta, t)l + p(x,t)(1-l-r).$$

We now assume that τ and δ are small and use a Taylor Series to compute

$$\frac{\partial p}{\partial t} = -\frac{\delta(r-l)}{\tau}\frac{\partial p}{\partial x} + \frac{(l+r)\delta^2}{2\tau}\frac{\partial^2 p}{\partial x^2} + \mathcal{O}(\tau^2) + \mathcal{O}(\delta^3).$$

$\lambda := \lim_{\tau,\delta\to\ 0} \frac{\delta(r-l)}{\tau}$
$D := \lim_{\tau,\delta\to\ 0} \frac{(l+r)\delta^2}{2\tau}$
Now, let us truncate the higher order terms. Note, for D to be finite, $\lim_{\tau,\delta\to\ 0} \frac{(l+r)\delta^2}{2\tau}$ must be finite. Thus, our truncation error is $\mathcal{O}(\tau+\delta^2)$. With that, we have

$$\frac{\partial p}{\partial t} = -\lambda\frac{\partial p}{\partial x} + D\frac{\partial^2 p}{\partial x^2}. \qquad \text{(Advection Transport)}$$

Corollary 4.1.1 From Advection Transport to Diffusion Transport

If we have $r = l = \frac{1}{2}$ then $u = 0$, and we return to the case where our heat flow is only diffusing.

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}. \qquad \text{(Diffusion Transport)}$$

This is to say, a biased random walk yields the partial differential equation where heat is advecting (flowing in some net direction). Alternatively, a simple random walk yields the partial differential equation where heat is diffusing (moving randomly in no net direction).

We conclude that barring a truncation error, the motion described by a simple random walk is identical to heat diffusing through a rod. That is, the stochastic process has a direct match with a Partial Differential Equation.

4.2 The Physical Derivation

Claim 4.2.1 A Partial Differential Equation for the Heat Equation

$\frac{\partial}{\partial t}u(x,t) = D\frac{\partial}{\partial x^2}u(x,t)$

Derivation: We appeal to Conservation Principles and Fourier's Law of Heat Conduction.
The following quantities will be used:

1. $u(x,t)$, temperature
2. $e(x,t)$, thermal energy density
3. $\phi(x,t)$, heat flux
4. $Q(x,t)$, heat source
5. $s(x)$, specific heat
6. $\rho(x)$, mass density

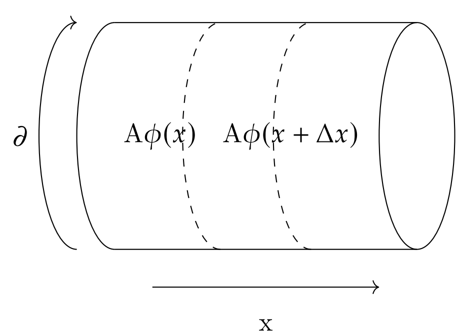

Figure 4.2: Heat Flux Through a Rod

Consider the rod in figure 4.1. We may calculate the thermal energy of the volume contained between x and Δx in two ways, as shown.

$$W = e(x,t)A\Delta x$$
$$W = s(x)u(x,t)\rho(x)A\Delta x.$$

It follows that

$$e(x,t) = s(x)u(x,t)\rho(x). \qquad (1)$$

Now we recall the conservation of heat energy. We observe that the total energy in our slice must be conserved, save for some external source $Q(x,t)$. Consequently, the difference in the energy entering one cross sectional area at x and the energy leaving another cross sectional area at x + Δx must be 0. Thus,

$$\frac{\partial}{\partial t}[e(x,t)A\Delta x] = \phi(x,t)A - \phi(x+\Delta x,t)A + Q(x,t)A\Delta x.$$

Meaning,

$$\frac{\partial}{\partial t}[e(x,t)] = \frac{\phi(x,t) - \phi(x+\Delta x,t)}{\Delta x} + Q(x,t).$$

Now we take the limit as Δ x approaches 0. We see that $\lim_{\Delta x \to 0} = \frac{\phi(x,t)-\phi(x+\Delta x,t)}{\Delta x} = -\frac{\partial \phi}{\partial x}$. Fourier's Law tells us that $\phi(x,t) = -K_0 \frac{\partial u}{\partial x}$. [1] Therefore,

$$\frac{\phi(x,t) - \phi(x+\Delta x,t)}{\Delta x} = K_0 \frac{\partial^2 u}{\partial x^2}. \tag{2}$$

By (1) and (2), $\frac{\partial}{\partial t}[e(x,t)] = \frac{\phi(x,t)-\phi(x+\Delta x,t)}{\Delta x} + Q(x,t)$ can be rewritten as follows:

$$\frac{\partial}{\partial t}[s(x)u(x,t)\rho(x)] = K_0 \frac{\partial^2 u}{\partial x^2} + Q(x,t)$$

Assuming s and p are constant and that there is no internal heat source, we obtain

$$\frac{\partial}{\partial t} u(x,t) = D \frac{\partial}{\partial x^2} u(x,t)$$

where $D = \frac{K_0}{s\rho}$.[2]

□

4.3 Generalizing Beyond 1-D

Without loss of generality, we may extend our results in one dimension to n-dimensions using the gradient and the Laplacian.

Definition 4.3.1: Gradient

The gradient is an operator that acts on a scalar function and returns the partial derivative with respect to each x_i.

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Definition 4.3.2: Laplacian

The Laplacian is also an operator defined with respect to some function f.

$$\nabla^2 f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}$$

Note:-

One should be cautious in how they intuit the Laplacian on Scalar Functions versus the Laplacian on Vector Functions. Here, we are treating f as a scalar function since heat energy is a scalar value.
Additionally, you may see the Laplacian written as Δf instead of $\nabla^2 f$. The motivation behind the $\nabla^2 f$ notation is that it is representative of taking the divergence of the gradient.

$$\frac{\partial p}{\partial t} = -u \cdot \nabla p + D\nabla^2 p \tag{Advection Transport}$$

$$\frac{\partial p}{\partial t} = D\nabla^2 p. \tag{Diffusion Transport}$$

[1] K_0 is known as the thermal conductivity.
[2] D is known as the diffusion constant. In solving the heat equation, you may see textbooks write $c^2 = \frac{K_0}{s\rho}$

4.4 The Solution to the Heat Equation

For $\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}$,

$$p(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{\frac{-x^2}{4Dt}}.$$

4.4.1 The Central Limit Theorem

Recall from Claim 2.1.1 that for a simple random walk our binomial distribution converges to a normal. This is because all updates to the position of a random walker are mutually independent and we may take an arbitrarily large amount of time steps, so the central limit theorem applies. So, without any actual solving of the partial differential equation we should expect to see an exponential function.

4.4.2 The Dirac Delta Function Initial Condition

The step by step process for solving this partial differential equation is out of the scope of this paper. Part of what makes the derivation difficult is that $p(x,t)$ doesn't factor into a function of x and a function of t. However, there is one aspect of the derivation of high importance. That is, it relies on the initial condition of the Dirac Delta function. The Dirac Delta Function has many definitions, all satisfying the properties that

$$\int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{1}$$

$$\delta(x) = \begin{cases} \infty, x = 0 \\ 0, x \neq 0 \end{cases} . \tag{2}$$

From these properties it is clear that the Dirac Delta Function isn't really a function, it's whats known as a generalized function or distribution. It is applicable to our solution to the heat equation because it gives us a tool to talk about the heat distribution of our rod in figure 4.1 at time zero. Consider infinitely many random walkers. At time zero, we will have infinitely many random walkers at the origin. As time increases, the walkers will diffuse through the rod according to the diffusion equation. The density of random walkers at a given (x_0, t_0) may be used to determine the heat energy $u(x_0, t_0)$. This idea will prove invaluable in chapter 5. We take the following definition for the Dirac Delta Function.

Definition 4.4.1: $\delta(x)$

$$\delta(x) = \lim_{\epsilon \to 0} \begin{cases} \frac{1}{2\epsilon}, |x| < \epsilon \\ 0, |x| > \epsilon \end{cases}$$

Proving (1) and (2) for our definition of the Dirac Delta Function is left as an exercise to the reader.

Chapter 5

Formatting Heat Data for Machine Learning

5.1 Random Walks

The following codes produces samples of a Simple Random Walk that may be formatted for Machine Learning.

Algorithm 1: Documenting Random Walkers

```
time_steps = 1000/* Or some other integer */
walkers = 1000/* Or some other integer */
uxt = matrix(walkers, time_steps)
p = 0.5*ones(walkers, 1)
foreach time_steps do
    r = randomColumnMatrix(walkers, 1)
    uxt[ :,i + 1] = uxt[:,i] + (r < p) + (-1*(r > (1 - p))) ;
    /* Update Position */
```

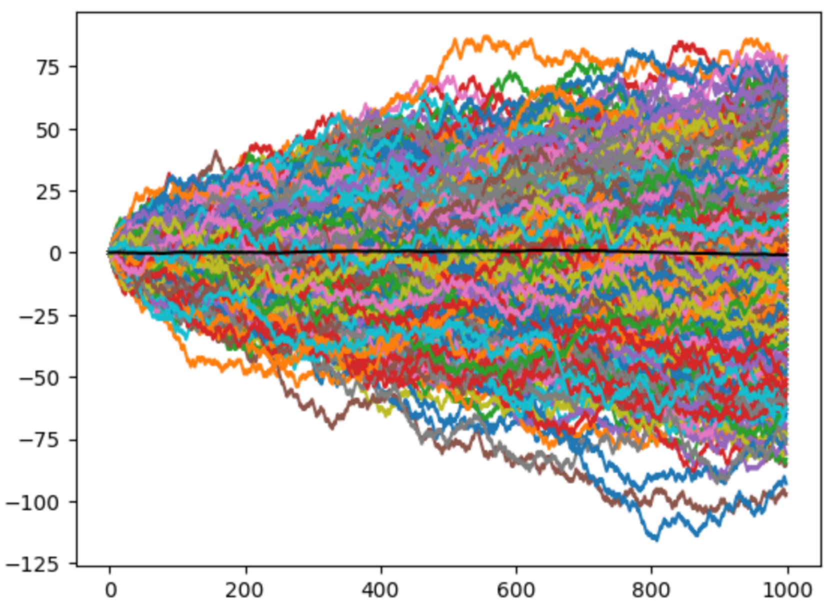

Figure 5.1: Walker Paths and Mean (Black)

5.2 Synthesizing Image

Recall what I had mentioned in chapter 4. The density of walkers may be used as a mechanism to calculate the heat energy u. We may take the $xt-plane$ and use color as a means graph the density of walkers at each point. This gives us a color graph $u(x,t)$.

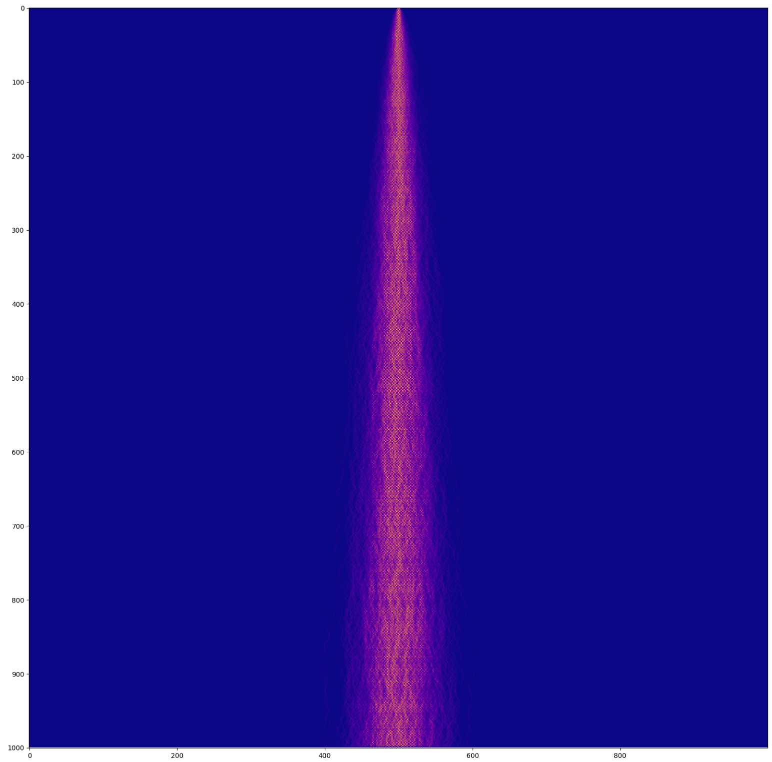

Figure 5.2: u(x,t)

5.3 The Network

In the case of the a simple random walk, we see that from the stochastic model we may derive a Partial Differential Equation, the heat equation. So, putting this image through a well designed Neural Network should successfully reproduce the heat equation. If the algorithm is successful, then we hope to generalize to more complicated stochastic processes. What we want to do is create a Neural Network that does the following:

1. Learn the Image, (i.e, the frequencies corresponding to energies u)
2. From the image, learn coefficients for all partial derivatives, up to some reasonable library of terms

5.4 Common Architectures

5.5 The Current State of Model Discovery