Analysis of the Application for Random Walk in Computer Science, Physics and Mathematical Equations

Joey Yang*

International Division of No.2 High School of East China Normal University, Shanghai, China

*Corresponding author: poneilung47397@student.napavalley.edu

Abstract. As a matter of fact, serving as one of the power methods to describe stochastic process, random walk theoretical analysis incorporating with Monte Carlo simulations are widely applied in various fields. With this in mind, this study will discuss the principle and the state-of-art applications in three specific fields. In detail, this study will demonstrate the use of random walk as a tool in different fields: computer science, physics, and mathematics. By discussing the personalized PageRank, the random walk particle tracking method as well as the random walk simulation to solve partial differential equations. According to the analysis, this study summarized and illustrates the significance of random walk and its wide applications. At the same time, the limitations for current advance processes will be demonstrated and the prospect for improvement of the random walk simulations will be proposed. Overall, these results shed light on guiding further exploration of random walk applications in different fields.

Keywords: Random walk; PageRank; heat transfer model; Laplace's equation.

1. Introduction

Random walk is a concept in probability theory. In this random process, an object moves by chance in a mathmetical space. It is commonly recognized that random walk was first proposed by Karl Pearson in 1905. Pearson published a letter in Nature seeking assistance concerning a particular problem: determining the likelihood that an individual, who commences from point O and walks l yards along a straight line, turns at any angle and repeats this process n times before arriving at a distance of r to r + δr from their initial position [1]. This problem was also called "The Drunkard's Walk". Then, in 1921, George Polya proved that in all 1 and 2-dimensional simple random walk including the one that Pearson proposed, the man is going to return to the starting point given infinite stretches [2].

With in a century, people have developed from exploring the concept of random walk to inventing random walk-based algorisms and applications. Until now, random walks is still a popular topic which allows numerous scholars to discover new theorems and models that can even be applied to the most recent technologies. The application of random walk has even extended to machine learning where it can be used as a testing solution for entities such as personnel or components, placing a limit on the number of successive failures allowed [3].

Apart from machine learning and artificial intellegence, random walks also play an important role in the field of finance and economy. Before being proved empirically that random walk is not frequently observed consistently in financial markets, the theory that the stock price follows a random walk was believed and utilized by many economists [4]. This study aims to introduce and analyze the principle of random walks and their application in computer science, physics and solving equations. In the following sections, this study will first introduce the basic principles of random walk, then focus on each of the three examples of the applications and specify the role of random walk as well as its significance. Afterwards, this paper discuss the limitations and future outlooks following a brief conclusion.

2. Basic Descriptions

There is not a single fixed definition for random walk, people have similar but slightly different definitions for it. A random walk can also be regarded as a Markov chain in which ξ_t is a random variable describing the position of a random walk after t steps [5]. In such model, all steps are independent, meaning that the previous steps do not affect the next step. Therefore, the probability of getting position i after t steps can be presented as the following:

$$P_t(i) = P(\xi_t = i) \tag{1}$$

and the t steps transition probability is defined as the following [5]:

$$p_{ij}^{(t)} = P(\xi_t = j \backslash \xi_0 = i) \tag{2}$$

Furthermore, if and only if the object in the walk returns to where it started given infinite steps, one call the random walk "recurrent", if not, one calls it "transient". In mathematical expression, a random walk is recurrent if:

$$\mathrm{P}(\xi_n = 0 \text{ i.o.} = 1) \tag{3}$$

In a theorem given by Polya in 1921, in simple random walk in $\mathbb{Z}^n$, there exists some constant C = C(d) such that for n large enough,

$$\mathrm{P}(\xi_n = 0) \sim \mathrm{C}\, \mathrm{n}^{-\frac{d}{2}} \tag{4}$$

3. Application in Computer Science

The application of random walk in the Personalized PageRank(PPR) algorithm plays a critical role in understanding and evaluating the importance of a web page. Due to the randomness and its close relationship with markov chain, it enables us to model the path a person might take in a webgraph. When applied to the PPR algorithm, random walk helps determine the relevance of web pages based on users and their preferences [5]. The standard PageRank(PR) algorism was developed by Brin and Larry Page, the cofounders of google search engine [6]. PageRank is the stationary distribution of a random walk between the nodes, which, at each step, with a certain probability p jumps to a random node, and with probability 1−p goes through a randomly chosen outgoing edge from the current node [7]. PPR is the same as PR, except that there is a certain probility α of going back to the source node. In this random walk, the PPR is used to evaluate the significance of a target node t with respect to the source node s by calculating the PPR value $\pi(s, t)$, which is the stationary distribution of the markov chain representing the network [6].

PageRank and PPR have been applied in various contexts, ranging from suggesting Twitter influencers and YouTube videos to being employed "outside of the Web" in domains like bioinformatics [8]. Take one of the most popular social media Twitter(now called X) as an example, people often receive the "Who to Follow" recommendations, and they are amazed by its ability to grab people's interests so percisely. In fact, the principle of this personalized recommendation is based on the PPR algorithm in which the user is the source node. At first, the suggestions might be random, but as the users keep spending time on certain type of influencers and posts, the PPR value for those types will increase, resulting in a larger probability of seeing the same type of content on the "Who to Follow" page. However, there is an increasing number of users of the social media platforms nowadays, and calculating the PageRank for all of them is taking longer and longer time. In recent years, people have been discovering new ways of modifying the algorithm to reduce the amout of the work needed and to estimate the PPR more quickly, such as bidirectional PPR and FAST-PPR [8, 9].

4. Application in Physics

The application of random walk in physics is highly significant, laeding to important breakthroughs in different fields of physics. It allowed the scientists to better understand and to predict the behavior of particles in air or other fluids. For example, Diffusion is the natural movement of molecules from areas of high concentration to areas of lower concentration [10]. In people's daily lives, diffusion happens frequently: the smell of the flowers, the coffee powders moving in the cup... These behaviors of the particles are also called Brownian motion, where the particles collide with the other particles near them and move randomly. The movement of those particles can be modelled by random walks. Furthermore, there are a lot of other phenomenons that can be modelled by random walks, such as pollutant transport and heat transfer [11]. The authors of a research paper on heat transfer in gas-solid systems have devised a Direct Numerical Simulation(DNS) framework, which employs the lattice Boltzmann method (LBM) to resolve hydroidynamics and the random walk particle tracking method (RWPT) to analyze temperature distribution [12]. RWPT does not directly involve in the advection-diffusion equation, instead, it monitors the trajectory of numerous Brown tracers as their movement is determined by the local velocity field and unpredictable fluctuations [12]. The workflow of RWPT method is shown in Fig. 1.

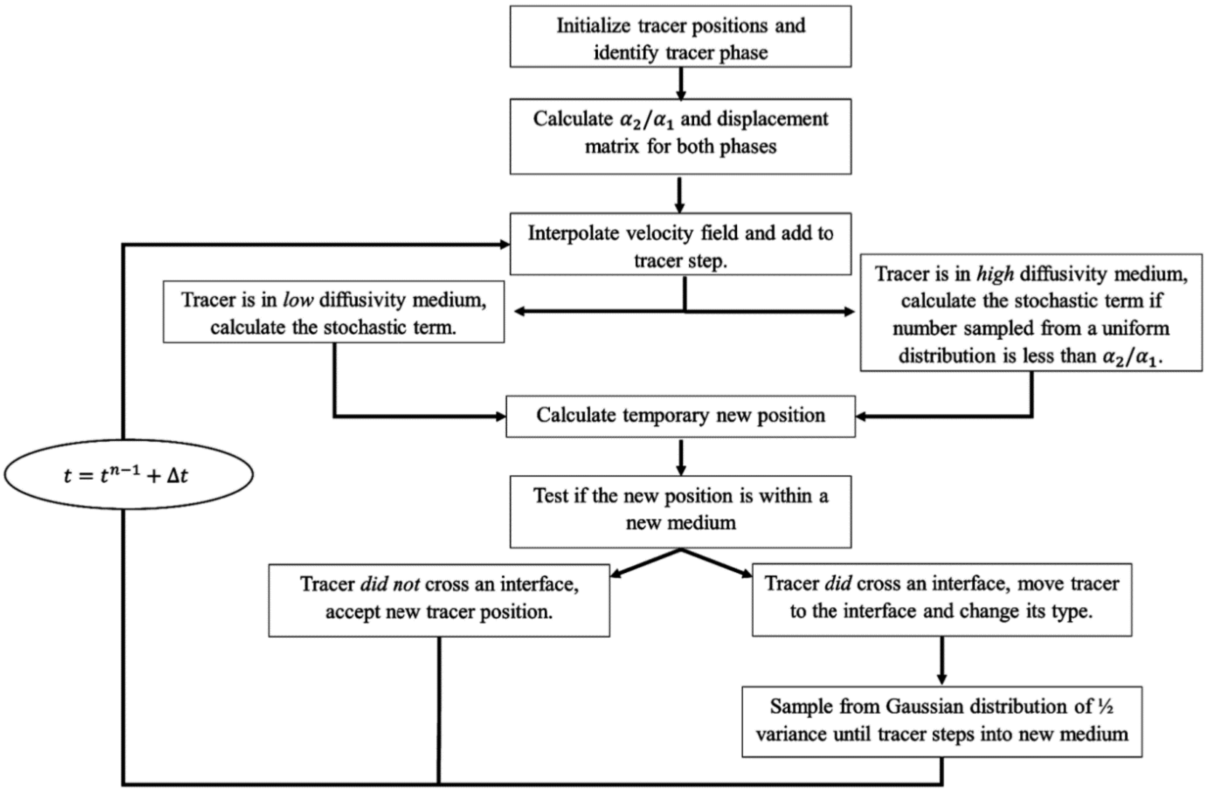

Fig. 1 The workflow of RWPT.

By experimental results, this hybrid LBM-RWPT method is shown to be more accurate in modelling heat transfer phenomenon than traditional DNS works. After carrying out 10 simulations on each method, it was revealed that the outcomes generated by the new code were consistent with prior analytical predictions and numerical bounds across the range of diffusivity ratios examined. On the other hand, the original code predicts the effective diffusivity inaccurately due to tracers exhibiting a preference for the lower diffusivity medium (phase biasing), which diminishes the αeff from its true value [12]. In sum, random walks provide a simple yet powerful way to model diffusion processes. By assuming that particles move randomly through a series of steps, one can simulate the motion of particles and calculate the resulting distribution. These models can be adjusted to reflect various factors that affect diffusion in complex systems. In addition, these models are widely applicable in various real life technologies. One example of the use of heat transfer model is to optimize the energy efficiency of electronic devices such as computer and cell phones, which will increase the overall performance of the devices (seen from Fig. 2).

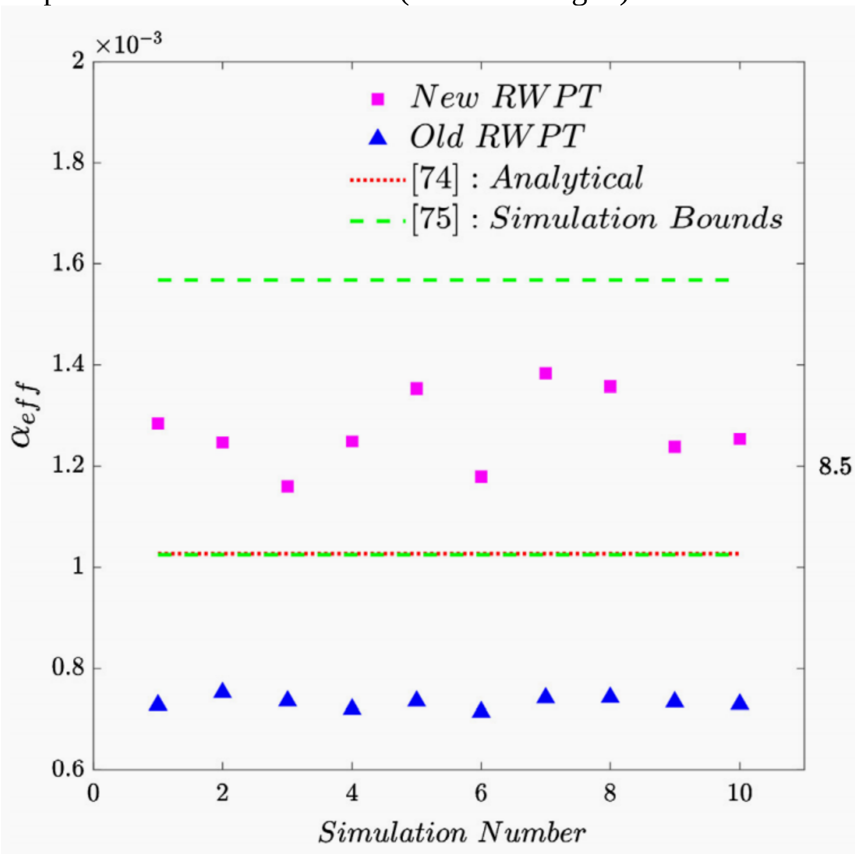

Fig. 2 The effective diffusivity obtained for each matrix configuration[.

5. Application in Mathmatics

The applications of random walk in Mathematics are more fundamental, meaning that there can be further applications in other fields like physics.

In Mathematics, random walks can provide a powerful tool for approximating solutions to otherwise intractable problems. For example, a stochastic differential equation (SDE) is a type of differential equation where at least one term consists of stochastic processes, and as a result, the solution to the equation is also a stochastic process. Whereas ordinary differential equations (ODE) describe variables which change according to a deterministic rule, SDEs describe variables whose change is governed partly by a deterministic component and partly by a stochastic component[13]. Besides that, a common type of differential equation, Partial Differential Equations(PDE), can be solved numerically by using random walk.

A simple case for the important role of random walk played in solving equations numerically is the approach to solve Laplace’s Equation, which plays a central role in the modeling of electromagnetic fields, fluid dynamics, and heat transfer. In simple terms, Laplace's equation can be expressed as the following: The total of the second-order partial derivatives of an unknown function, denoted as u, with respect to the Cartesian coordinates is equal to zero:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

Although the boundry value of Laplace's equation can be solved analytically for simple geometries and boundary conditions, the application of numerical approaches is necessary for solving problems with more complex geometries and boundary conditions in real life. For example, it can be n dimensional. As a result, people sometimes have to rely on numerical solution methods to approximate the solution.

The process to solve such an equation with random walk depends on the idea of the "exit point" of a path of the walk. One starts a random walk at $x \in D$, where D is the domain. At some point it will intersect with ∂D, thus exiting D. The point where the path of the random walk exits D after starting at x is defined as Exit(x,D). With this defintion, one has a formula for the solution u [14]:

$$u(x) = \mathbb{E}(f(\mathrm{Exit}(x, D)).$$

where f is a function defined by the corresponding value of the function on the boundry ∂D[14]. The author runned the python code below to get a numerical solution for the equation as $\mathbb{E}(f(\mathrm{Exit})) = 0.49756011454$, which is very close to the real solution of 0.5. This technique can be extended to solve many different partial differential equations, including PDEs. One such equation is the Black-Scholes PDE from finance or the Backwards Heat Equation, which has applications to deblurring imaging [14].

6. Limitations and Future Outlooks

Random walks have emerged as a powerful tool for simulating and modeling complex systems in many areas of science, ranging from materials science, from physics to economics and finance. Random walk-based methods have enabled researchers to develop mathematical models that accurately describe the statistical behavior of particles that undergo stochastic processes or diffusive behavior. However, while random walk methods have various advantages, limitations and future directions can be addressed as well. One of the key limitations of random walks is the assumption of independence and identical distribution (i.i.d) of steps. This implies that each step is independent of previous steps and follows the same statistical distribution, which may not be valid in real-world scenarios. For instance, in stock market prediction, the i.i.d assumption often fails as market trends and economic factors demonstrate dependencies over time.Random walks also assume that the transition probabilities are known and constant, which is not always the case. In many real-world scenarios, the probabilities vary over time and the changes are unknown. This limitation can lead to inaccurate modeling and predictions.

Another limitation is the difficulty in determining the right step size in random walks. Larger step sizes may be computationally efficient but can introduce errors due to approximation. In some cases, a smaller step size might be more accurate, but it leads to the problem of computational costs. As the number of particles increases, the simulation of a stochastic process becomes computationally intensive, posing a serious constraint on the feasibility of the method. Even with modern computing power, numerical using limited numbers of particles can be time-consuming, making the random walk approach unsuitable for problems that require large-scale simulations. Despite these limitations, the future of random walks in scientific and technological applications is promising. Advancements in computational power and machine learning techniques are opening new areas for improving and extending the use of random walks. It is hoped to see more sophisticated use of random walks in more advanced technologies.

7. Conclusion

To sum up, this study analyzed the principles and the use of random walk in different applications in computer science, physics, and mathematics. In computer science, a kind of algorithm is introduced, followed by attempts to optimize the algorithms to be more efficient. Next, a kind of random walk particle tracking method is discussed to model a common phenomenon in physics. Lastly, this study looks at the method to solve partial differential equations by creating a corresponding random walk. The limitations and the future outlooks are also discussed. By examining the wide applications of random walks, one has gained a deeper understanding of the significance of it as a tool. One should keep on finding new applications and new useful tools.

References

[1] Pearson K. The problem of the Random Walk. Nature News, Nature Publishing Group, 1905, 7: 1.

[2] Ubereine A, der Wahrscheinlichkeitsrechnung b. Die Irrfahrt im Strassennetz, Math. Ann., 1921, 84: 149-160.

[3] Yazidi A, Granmo O C, Oommen B J. A new tool for the modeling of AI and machine learning applications: Random walk-jump processes. SpringerLink, Springer Berlin Heidelberg, 1970, 1: 1.

[4] Fama E F. Random walks in stock market prices. Financial analysts journal, 1995, 51(1): 75-80.

[5] Xia F, Liu J, Nie H, et al. Random walks: A review of algorithms and applications. IEEE Transactions on Emerging Topics in Computational Intelligence, 2019, 4(2): 95-107.

[6] Park S, Lee W, Choe B, et al. A survey on personalized PageRank computation algorithms. IEEE Access, 2019, 7: 163049-163062.

[7] Bahmani B, Chowdhury A, Goel A. Fast incremental and personalized pagerank. arXiv preprint arXiv:1006.2880, 2010.

[8] Vial D, Subramanian V. On the role of clustering in Personalized PageRank estimation. ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS), 2019, 4(4): 1-33.

[9] Lattanzi A M, Yin X, Hrenya C M. A hybrid lattice Boltzmann–random walk method for heat transfer in gas–solids systems. Journal of Computational Physics: X, 2019, 1: 100007.

[10] Masuda N, Porter M A, Lambiotte R. Random walks and diffusion on networks. Physics reports, 2017, 716: 1-58.

[11] Wu X, Liang D. Study of pollutant transport in depth-averaged flows using random walk method. Journal of Hydrodynamics, 2019, 31: 303-316.

[12] Lattanzi A M, Yin X, Hrenya C M. A hybrid lattice Boltzmann–random walk method for heat transfer in gas–solids systems. Journal of Computational Physics: X, 2019, 1: 100007.

[13] Sauer T. Numerical solution of stochastic differential equations in finance. Handbook of computational finance. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 529-550.

[14] Scott R. On solving partial differential equations with Brownian motion in Python. Scott Rome, 2016, 4: 24.