Approach to Data Science with Multiscale Information Theory

Shahid Nawaz, Muhammad Saleem, F. V. Kusmartsev and Dalaver H. Anjum

Preprint v2

Dec 4, 2023

https://doi.org/10.32388/B225L5.2

Approach to Data Science with Multiscale Information Theory

Shahid Nawaz
18 Lake Shore Drive, Apt 2B, Watervliet, NY 12189, USA

Muhammad Saleem
Department of Physics, Bellarmine University, 2001 Newburg Road, Louisville, KY 40205, USA

F. V. Kusmartsev and Dalaver H. Anjum*
Department of Physics, Khalifa University, PO Box 127788, Abu Dhabi, United Arab Emirates

Data science is a multidisciplinary field that plays a crucial role in extracting valuable insights and knowledge from large and intricate datasets. It has the potential to drive accurate predictions and enhance decision-making capabilities across various domains, including finance, marketing, healthcare, and scientific disciplines. In this paper, we developed a multiscale entropy dynamic (MED) methodology that is applicable to the field of data science. As an example, we apply this methodology to the data science framework of a large and intricate quantum mechanical system composed of particles. Our research demonstrates that the dynamic and probabilistic nature of such systems can be effectively addressed using the proposed MED approach. Through this approach, we are able to describe the system's dynamics in a multiscale form of equation of motion which turned out to be a general form of the Nonlinear Schrödinger Equation (NSE). It becomes the conventional linear Schrödinger equation for the case of smallest size particles, namely electrons, and quite expectedly nonlinear Schrödinger equation for the cases of quasi-particles, such as plasmons, polarons, and solitons. By employing this innovative approach, we pave the way for a deeper understanding of quantum mechanical systems and their behaviors within complex materials.

1. Introduction

The fields of statistical physics, information theory, and data science share common mathematical and conceptual foundations. These are interconnected in several ways [1, 2] and deal with the analysis of complex systems and make use of similar mathematical and statistical tools to model and analyze these systems. One of the most important connections between these fields is the concept of entropy [3]. In statistical physics, entropy is the measure of disorder or randomness of a physical system and can be used as a main driving force of the Second Law of thermodynamics [4]. In information theory, entropy is a measure of the uncertainty or randomness of a message or information content of a data set. Whereas in the field of data science, entropy is often used to measure the randomness or unpredictability of data sets in classical and quantum models [5]. Another connection between these fields is the use of probability theory and statistical methods to model and analyze complex systems [6]. Statistical physics uses probability theory to model the behavior of large systems of particles, while information theory uses probability theory to quantify the uncertainty or randomness as well as a measure of information from the messages in data sets [7]. Data science uses statistical methods to model and analyze data sets and to develop algorithms for processing and interpreting data [8].

Quantum mechanics has been conventionally formulated in the Hilbert Space with the use of two conjugate variables that obey Heisenberg's uncertainty principle. For instance, the momentum p and the position coordinate q, in the form of $\Delta q \Delta p \geq \hbar/2$. In this way, Heisenberg's principle describes the statistical nature of these self-adjoint operators $\hat{q}$ and $\hat{p}$ in the Hilbert Space. The ED formulation of quantum mechanics was introduced in 2009 and has been applied to fields such as quantum measurement problem [9, 10], uncertainty relations [11] curved space-time [12], scalar fields [13–15], and finance [16]. The ED approach is an alternative formulation of QM in which the dynamics of a probability distribution are derived from the entropy [17, 18]. This approach seems to describe the discord between the dynamics and probabilistic nature of quantities in a more convenient way compared to other approaches. In the entropic dynamics (ED) formulation, the quantum nature of these operators is given a secondary role. In fact, in this approach, the uncertainty in these variables stems from the diffusion process of the Brownian motion of the particles [11, 19]. It is important to note that momentum is not real in ED formulation but is an epistemic property of wave functions and not a property of the particles. In classical mechanics, we assume that the particles' positions and their momenta are real. In the ED approach, only the positions are real because ED does not describe the dynamic of particles, but only the dynamic of their probabilities. Another important fact of ED formulation is that there is no conventional momentum that is canonically conjugated to the generalized coordinates. As we mentioned above, the generalized coordinates represent the probabilities, not the actual positions. Thus, there is no momentum of the particles in ED formulation. Of course, we can always consider translations, and, following convention, we can call the generator of translations by the name "linear momentum". But this is just a name for operators that are not properties of the particles.

In this paper, we extend ED formulation which is applicable to complex systems of scientific disciplines including data science. The ED formulation has been extended by developing a multiscale ED (MED) methodology. The term multiscale represents the hierarchical system of particles or entities participating in the dynamics of a particular system. The paper is organized as follows. In section 3 the ED formulation of QM is briefly presented. The approach to data science using MED is discussed in section 4. In section 5 we demonstrate how the MED framework can be used to derive Generalized Schrödinger Equation (GSE). In section **??**, the derived GSE is applied to a couple of well-known systems from solid-state physics that describe the dynamics of quasi-particles (e.g., solitons) in nonlinear phenomena. Furthermore, a form for the interaction of electromagnetic fields with solitons is also presented. However, this was done under the assumption that solitons have self-trapped particles like electrons and therefore are charged. The impact and ramifications of the GSE in the context of describing the dynamics of solitons in 2D materials are discussed in the section 7. In the end, the overall conclusions drawn from the presented examples of the application of the MED method to various complex systems are presented and discussed in the context of Data Science, see the section 8.

2. Foundation of Entropy Evolution Approach to Data Science

Data science is a multidisciplinary field dedicated to extracting valuable insights and knowledge from both structured and unstructured data. Using a diverse set of techniques, methods, and tools from disciplines such as statistics, mathematics, computer science, and domain-specific knowledge. The roots of data science trace back to the practice of data collection, which has become a ubiquitous aspect of nearly every human activity over time. As computer facilities have advanced, the ability to gather relevant data from various sources has resulted in the creation of extensive databases. Consequently, ensuring the quality and quantity of collected data has become paramount, prompting a detailed analysis to directly investigate their impact. Exploratory Data Analysis (EDA) plays a crucial role in this process, involving thorough examination, analysis, and visualization of data. Its objective is to understand the patterns, trends, and relationships within the data, facilitating the construction of predictive models through algorithms that learn from the data at hand.

One crucial aspect of Data Science is Feature Data Engineering, which involves predicting the evolution of data flow. This process incorporates Big Data Technologies, capable of handling and processing large volumes of data. The field of data science is rapidly evolving, which requires practitioners to stay up-to-date on new techniques and technologies. In this dynamic and interdisciplinary field, it is essential to find and implement models that can perform fast and efficient analyses, providing realistic predictions over time. Considerations such as scalability, real-time processing, and model drift are integral to the practice of Data Science. Practitioners often specialize in specific areas based on their interests and expertise. The application of data science is extensive, spanning across various industries, including healthcare, finance, marketing, and technology. In simple terms, Data Science uses data organized into numerous sets, with each mathematical set comprising data identified by their similarities. This concept of similarity is broad, encompassing scenarios such as sets containing snapshots of cars crossing red lights at road intersections or exceeding speed limits in specific locations of special importance. Consequently, we obtain multiple sets, allowing for the identification of the probability of traffic rule violations, denoted p_i, and the estimation of the information entropy describing the information content. The first probability (p_i) represents the ratio of the number of elements in a particular set to the total number of elements in all sets.

Using probability functions, the next step involves estimating the Shannon information entropy, denoted S_s. This marks the initial phase. Subsequently, in the second step, we need to consider the values of police fines established for each of these traffic rule violations. The fines for the $n-$th type of violation occur with the probability p_n, and have the value E_n and therefore these fines can be incorporated into the expression of entropy using Lagrange multipliers, e.g., a constant β. By determining the maximum of the total entropy (or information content) function equal to $S_{total} = S_s(p_1, ..., p_N) + \beta \sum_n p_n E_n$ we find the initial entropy or content information in the existing data set. Here, the function $S_s(p_1, ..., p_N)$ is expressed in the form of Shannon entropy and is a measure of the uncertainty or information content in a set of data. It is calculated using the formula:

$$S_s = -\sum_{n=1}^{N} p_n \log_2(p_n) \tag{1}$$

where p_n represents the probability of the occurrence of the $n-$th event described the a certain set of data. In the context of the provided text, this entropy measure is applied to assess the information content associated with traffic rule violations.

Here, we may reveal the dependence of each p_n on the value of the fine E_n in a manner very similar what is used in statistical mechanics at the derivation of the Boltzmann distribution. With this step, the analysis of the existing data sets is concluded and the probability distribution of police fines is derived. In the subsequent second stage, our objective is to find or predict the evolution of this complex system. Here we used an analogy between the money of the fine and the energy in statistical analysis, see for detail the Ref.[20].

It is a well-established fact that in any complex system in physics, e.g. such as classical and quantum gases, the entropy of any initial equilibrium state increases with time during interactions between particles. This fundamental observation constitutes the cornerstone of the Second Law of Thermodynamics, with an immense number of confirmations in nature. Naturally, this prompts the assumption that the evolution of entropy, particularly its growth, extends far beyond the conventional subject of physics. Thus, in the second stage of our approach, our primary objective is to derive equations for the time evolution of the entropy in any complex system. The evolution of entropy begins with the initial entropy calculated from the data sets in the first stage of our approach. Remarkably, in numerous cases considered in this paper, the equations governing entropy evolution reduce to various types of nonlinear Schrödinger equations. This is noteworthy, especially considering that traditional data science methods, such as the multidimensional version of gradient descent, have been predominantly used. While such methods prove useful in many cases, they face challenges in realistic systems due to their multidimensional nature. For instance, introducing the concept of the system's internal energy, represented as an integral function of entropy, reveals a complex energy landscape with numerous minima and maxima. The multidimensional gradient descent method may lead the system to a false minimum or maximum. An alternative and more advanced approach involves machine learning, which improves the situation by training the evolution of entropy on the initial time steps and then extrapolating it for all subsequent times. However, the approach proposed in this paper offers a comprehensive formulation that can be applied to any database.

We propose using specific equations used to describe the time evolution of the entropy, which can vary depending on the characteristics of the system under consideration. In the context of the information provided by data science, the entropy evolution may be described by nonlinear Schrödinger equations (NLSE). It is a partial differential equation that commonly appears in various areas of physics, including optics, plasma physics, and condensed matter physics. Its general form is:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi + g|\psi|2 \tag{2}$$

In this equation: ψ represents the wave function, t is time, x is the coordinate of the generalized spatial position, as in classical mechanics. $\hbar$ is the reduced Planck constant, m is the particle mass, $V(x)$ is the potential energy and the constant g_n characterizes the strength of the non-linearity. This equation describes the evolution of the wave function $\psi(t)$ over time, and its nonlinearity term $g|\psi|^2\psi$ is a key feature that distinguishes it from the linear Schrödinger equation. In the context of the provided information, the use of nonlinear Schrödinger equations suggests that the evolution of entropy in certain complex systems follows a mathematical framework akin to that of quantum mechanics. The specifics of how these equations are adapted or derived for entropy evolution in data science applications would require a more detailed understanding of the particular system and the associated mathematical modeling choices.

The presented methodology of multiscale entropic dynamic (MED) is indeed applicable to any system including data science. It requires determining the constraints of the system. The details of the methodology are listed below. The entropic dynamic represents the evolution function of information in a statistical manner. Multiscale describes the size of the information carrier or a possible variety of generalized coordinates as in classical mechanics. Its smallest size information carrier obeys the linear system for the equation of motion. Whereas the equation becomes nonlinear as size of carriers scales up.

3. The Entropy Dynamic Formulation of Quantum Mechanicss

The fundamental object in QM is the wave function Ψ, which is a complex function of coordinate and/or momenta. It may, generally, be represented with the use of two real quantities ρ and Φ: $\Psi = \sqrt{\rho}\exp(i\Phi)$. Here ρ is the probability density and Φ is the phase field. The wave function is nicely described by $U(1)$ gauge symmetry group, where the phase Φ is a free gauge parameter. The elements of the abelian gauge group $U(1)$ are associated with the different values of the phase. It turns out that these two quantities are the degrees of freedom in ED methodology, which are determined in two steps. The first step involves the entropy functional (9) that leads to the Fokker-Planck (FP) equation (25). The second step involves an energy functional (28), whose conservation leads to quantum Hamilton-Jacobi equation (32). The two equations are then combined to give the Schrödinger equation. It is pertinent to mention herein that in the past the Schrödinger equation has been derived in [21] with the use of stochastic mechanics. The starting point in that derivation was the transition probability, which was conventionally described by Green's function. In that stochastic theory-based formulation [21] it's form is given below:

$$p(x', t'|x, t) = \frac{1}{(2\pi\sigma^2\Delta t)^{d/2}} e^{-\frac{(x'-x)^2}{2\sigma^2\Delta t}} , \tag{3}$$

where d is the dimension of the space and $\sigma^2/2$ is the diffusion coefficient. The Brownian motion given by equation (3) keeps track of the future given the present is known while independent of its past. Equation (3) admits the following stochastic differential equation

$$\Delta x^i(t) = b^i(x(t), t)\Delta t + \Delta w^i(t) , \tag{4}$$

where b^i is drift velocity and w^i is the fluctuation or noise in the Brownian Motion with the following correlators

$$\langle \Delta w^i \rangle = 0 , \quad \text{and} \quad \langle \Delta w^i \Delta w^j \rangle = \sigma^2 \Delta t \delta^{ij} , \tag{5}$$

On the other hand, the transition probability derived in the ED formulation is followed by the incorporation of the Brownian motion. The ED framework differs from stochastic mechanics in ways that are delineated next. The ED formulation for a certain quantum system begins with defining the entropy functional (9) subjected to relevant constraints of the system and defining the notion of entropic time. The relevant constraints are those that lead to the desired theory. Since our main concern is to derive quantum theory using ED, the relevant constraints are the phase constraint (13) and gauge constraint (14). Once the constraints are incorporated and the entropy functional is optimized, one obtains the transition probability (17) which apparently is timeless. However, this is a common feature of Bayesian or entropic inferences, where the goal is to update the prior probability to the posterior probability when new information becomes available. The new information could be in the form of data (Bayesian inference) or in the form of constraints (entropic inference). In both cases, the two methods of inference are atemporal. It does not matter whether the posterior is obtained in the past or present, one gets the same result. Interestingly it is possible to introduce time in ED. Consider a particle that moves from the initial position x to the final position x'. Generally, both positions are unknown. This means that we are dealing with the joint probability $P(x, x')$. Then using the product rule of probability, we obtain the following:

$$P(x, x') = P(x'|x)P(x) , \tag{6}$$

where $P(x'|x)$ is the probability of x' given x. Since x is also unknown, we marginalize over x to obtain the following:

$$P(x') = \int dx P(x', x) = \int dx P(x'|x)P(x) , \tag{7}$$

where $P(x)$ is the probability of the particle being located at position x.

Whereas $P(x')$ is the particle at position x'. As x occurs at an initial instant t and x' happens at a later instant t'. Therefore we finally set the probabilities in the following manner:

$$P(x) = \rho(x, t) \quad \text{and} \quad P(x') = \rho(x', t') . \tag{8}$$

In short, in the ED formulation, time is introduced as a book-keeping device that keeps track of a change. The notion of time would be further elaborated in section 5 where the duration of time would also be obtained.

4. The Multiscale Entropy Dynamic Methodology

The emergence of new information technologies has led to the rise of data science, which involves the analysis of large and complex data sets to make predictions about the evolution of systems. Data science has become an essential tool for businesses and organizations to gain insights into their customers, products, and operations, and make data-driven decisions. To conduct a data analysis, one must first take a data set and break it down into subsets. By analyzing the multiplicity of these subsets, one can determine the probabilities for the realization of different outcomes. This approach is based on the principles of probability theory, which involves quantifying uncertainty and measuring the likelihood of different events. To make predictions using data analysis, one can use methods such as statistical mechanics, which was first introduced by Boltzmann. This method is based on the concept of entropy, which measures the amount of disorder or randomness as well as the information contained in a system. By applying statistical mechanics to a data set, one can determine the most likely outcomes and predict the future evolution of the system. What we propose to do is the following: take a data set and make an analysis with the goal of decomposing it into an arbitrary number of subsets. Then by using the multiplicity of those subsets, determine the probabilities for the realization of the different subsets. The determination of the probabilities can be accomplished using methods of statistical mechanics, which are completely compatible with the presented MED methodology.

As we deal with multiscale methodology, the additional ingredient is to sum over all scales representing the hierarchal system of particles or entities. The $Q(x', s'|x, s)$ is called the prior probability distribution, which we get from our data set as mentioned earlier. The unknown function is the transition probability $P(x', s'|x, s)$, which is analogous to the transition probability distribution for a single particle when it moves from the position x to a neighboring point x', where s, s' are scaling indices or arbitrary variables of the data set which is under consideration. Our goal is to find the transition probability. But first, we have to find the prior, which we get from the analysis of the information data set, which is under consideration.

The main idea of our approach is that in any system with time, the entropy rises. The same happens in the case of the second Law of thermodynamics, which deals with a complex system and also a large number of particles. So we consider the dynamics of entropy in MED methodology in the same way as it is done in the Second Law of thermodynamics. Thus, our goal in this paper is to extend the application of ED[18] to a complex system which can be described by some complex data set. Such a system can be also a single quantum mechanical particle embedded in different environments or scales. The mathematical tools needed for a system of such directly noninteracting particles go beyond the usual statistics and calculus. Since a system of particles or a complex system can involve several constraints, one needs to adopt a multifaceted and multiscale approach to formulate the equations that describe the system accurately [22] (see, the references therein). The proposed MED methodology has applications in both natural and social sciences. For instance, the brain is a complex system of neurons and in the same way, society is a complex system of communication networks. The universe itself is a complex system too, as it is comprised of planets, stars, and ultimately galaxies.

In this paper, we consider a quantum mechanical system as an example to extend the ED method to the MED method with the eventual goal of applying MED method to the field of Data Science. As an illustration, we apply MED to the single and non-interacting particles of a quantum mechanical system, to obtain a well-known form of the Generalized Schrödinger Equation (GSE). The obtained set of equations is identical to the non-linear Schrödinger equations (NLSE) that have been applied to various systems including superconductivity. In other words, the equations derived under MED approach describe the dynamics of non-linear systems in physics consisting of plasmons, deformons, polarons,condensons, and optical or matter solitons. The dynamics of solitons is a of great interest in describing the properties of new emerging fields of materials science, e.g. nonlinear optics, two-dimensional (2d) materials, cold and hot plasma physics, and crystal lattice dynamics. As one example the matter solitons are considered as non-relativistic quanta of matter waves and represent Bose-Einstein Condensates (BEC) of atoms and electrons [23–28]. It is likely that the first form of NLSE appeared in Landau Theory of phase transition where he first introduced the cubic term in the Shroedinger equation describing order parameter [29]. Later in 1950, this equation was applied by Ginzburg and Landau to superconductors and the NLSE got a new meaning, the famous Ginzburg-Landau equation[30]. Later in 1960s the idea was applied to BEC and the NLSE equation got the name as Gross-Pitaevskii equation. On the other hand, an electron trapping by a crystal lattice, which leads to another form of NLSE was also first described by Lev Landau in 1933 [23]. Solomon Pekar proposed the concept of the polaron in 1946[31], which was further developed by Landau and Pekar in a 1948 paper[32]. This theory suggested that polarons, not free electrons, were the charge carriers in ionic crystals. Unlike quantum electrodynamics, the polaron theory is free from divergences, and the electron energy and mass remain finite. Today, research on polarons continues to expand into new areas of 2D materials, where new forms of NLSE have been obtained[33]. In general, it would be also interesting to consider these new phenomena from the principle of MED and compare them with conventional approaches. In particular, for a description of nonlinear waves, the NSLE was originally derived by Zabusky and Kruskal [24], but using MED one may include the Physics of many non-equilibrium phenomena, dissipation, and scattering and it can be used to describe the dynamics of solitons in 2d materials or many-body soliton physics.

5. Formulation of MED Methodology and Its Application to a Quantum Mechanical Complex System of Data Science

As stated earlier, our eventual goal is to introduce MED methodology for the field of Data Science. As an illustration, we apply it to describe the dynamics of quantum particles namely solitons. Whereas a complex Data Science system may also have particles such as fractals which are self-similar structures. Fractals are found in nature, besides being constructed both experimentally and mathematically. Clouds, lightning, and coastlines are examples of natural fractals, and the Sierpinski triangle is an example of geometrical fractals [34]. Moreover, geometrical objects are also fractals and can be found in Benoit B. Mandelbrot foundational book on fractals [35]. The creation of solitons in

Fractals has been discussed in [36].

For a quantum statistical system, one must first specify the microstates, the prior probability distributions, and the constraints at the stage. Similarly, in a Data Science application, the prior probability distributions originated directly from the existing data set which is the subject of the main Data Science analysis. The most important part of this analysis is to find which constraints have been used in the collection of the existing data. The correct evolution of entropy strongly depends on these constraints. In the next step, we have to incorporate these limitations in the entropy function. With these taken into account we arrive at a traditional generalized Boltzmann-like expression of entropy which is a Quantum Mechanical representation of the entropy. To derive the linear Schrödinger equation (LSE), one considers N noninteracting particles living in a flat Euclidean space. It is assumed that particles have definite initial positions (and indefinite values of momenta) and yet unknown values that are desired to be inferred. The different definite initial positions of the particles form a data set. Such a data set can be very large, depending on how many initial positions for a single particle we will take into account. The data set can be also split into subsets associated with different scales, e.g. of the fractal. Note that the microstates at each scale are different. The devised MED methodology is given below in the following steps:

5.1 MED Functional

It is assumed that at each scale, the particle resides in an Euclidean space $\mathcal{X}_s$ with metric δ_{ab}, with $a = 1, 2, 3$ for spatial coordinates. And for all particles at that scale we have $\mathcal{X}_{N_s} = \mathcal{X}_s \times \ldots \mathcal{X}_s$, which is $3N_s$ dimensional configuration space. The positions of the particles are given by $x_i^a \in \mathcal{X}_{N_s}$, where the index $i = 1, 2, \ldots N_s$. We represent x_i^a collectively by x. The multi-scale entropic functional for a system can be written in a way described by equation (**??**) (See a review on ED in [18]):

$$S[P, Q] = -\sum_{s'} \int d^n x' P(x', s'|x, s) \log \frac{P(x', s'|x, s)}{Q(x', s'|x, s)} \,, \tag{9}$$

Equation(9) is the extension of the entropy functional in [18]. As we are dealing with multi-scale, the additional ingredient is to sum over all scales. Here $Q(x', s'|x, s)$ is the prior probability distribution, and $P(x', s'|x, s)$ is the transition probability distribution as the particle moves from x to a neighboring point x', where s, s' are scaling indices. Our goal is to find the transition probability. But first, we have to determine the prior probability distribution. For any specific data set, it can be obtained directly by classifying different snapshots of the data set. For the particular case of our many-particle system, as for the ideal gas to determine prior, we will follow the original Boltzmann approach.

5.2 Prior Multiscale Probability Distribution Functional

The prior multiscale probability distribution functional $Q(x', s'|x, s)$ codifies the relation between x and x' before the information contained in constraints has been processed, where all particle positions are equally probable. In other words, it is desired to find a prior distribution that is invariant under translation and rotation. It can be obtained by maximizing the following relative entropy:

$$S(Q) = -\sum_{s'} \int d^n x' Q(\Delta x) \log \frac{Q(\Delta x)}{\mu(\Delta x)} \,, \tag{10}$$

where $\Delta x = x' - x$ is relative to the uniform measure $\mu(\Delta x)$, subject to normalization and a constraint that concerns short steps,

$$\sum_{s'} \int d^n x' Q(x', s'|x, s) \delta_{ab} \Delta x_i^a \Delta x_i^b = \langle \Delta \ell_i^2 \rangle \,, \ (i = 1, 2, \ldots N_{s'}) \tag{11}$$

where $\langle \Delta \ell_i^2 \rangle$ are constants equal to the square of the average displacement between the points x and x'. The index i indicates that there are $N_{s'}$ constraints at each scale that are rotational invariant.

The proof of the results presented below in the equation is completed in Appendix A.

$$Q(x', s'|x, s) \propto \exp[-\frac{1}{2} \sum_{s'} \sum_{i}^{N_{s'}} \frac{1}{\sigma^2_{s',i}} \delta_{ab} \Delta x_i^a \Delta x_i^b] \tag{12}$$

where $\sigma^2_{s',i}$ is a Lagrange multiplier, which will be determined later (see equation (19).To ensure small steps, this Lagrange multiplier must be very small. The right-hand side is the product of the Gaussian function, which means the short steps are independent of each other. Equation (12) is a prior probability distribution that only takes into account the original positions of a particle before the actual constraints or information is incorporated, such as the influence of the EM field. It only describes motion in short steps as the particle moves from x to x'. We want to write down a form of equation (9) that works for describing the dynamics of particles and quasi-particles in solids, considering the case when the isotropic symmetry of the space is broken, e.g., by applying an external electric field. This symmetry-breaking requires including extra constraints that are listed below. In this way, the transition probability distribution $P(x', s'|x, s)$ will be determined. The isotropic symmetry breaking of the probability distribution can be achieved by introducing an external force per unit charge which is dependent upon space. Consequently, it results in space-dependent symmetry-breaking as well. In any complex system of Data science, it is caused by the gradient of a scalar external potential. In the case of a quantum mechanical complex system, it is the electric field generated by the gradient of electric "potential" $\phi_s(x^a)$, that satisfies the following constraint:

$$\sum_{s'} \int d^n x' P(x', s'|x, s) \Delta x^a \frac{\partial \phi_s}{\partial x^a} = \kappa_{1,s} \tag{13}$$

This constraint is called the drift potential constraint in the context of a quantum mechanical complex system. The $\kappa_{1,s}$ are constants that are related to equipotential lines. These equipotential lines are related to the cross-section which are perpendicular to the applied field.

The time-dependent symmetry-breaking of the probability distribution can also be achieved by external force per unit charge that depends upon the time. This implies that the time-dependent component of the external force must be generated by a time-varying potential as well as a vector. For example, in the case of a quantum mechanical complex system, the external electric field can also be generated by the rate change of vector magnetic potential. This symmetry-breaking constraint can be imposed in the following form:

$$\sum_{s'} \int d^n x' P(x', s'|x, s) \Delta x^a A_a = \kappa_2^s \tag{14}$$

where A_a is vector potential which is a function of both space and time. The κ_2^s are constants that represent average displacement in the direction of the vector potential. It should be noted that an arbitrary form of the vector potential can be selected that results in the meaningless form of symmetry breaking. Therefore, its gauge invariant form must be selected. The gauge invariant form of vector potential implies that it causes the symmetry breaking of probability distribution functional simultaneously in time and space. This is why, the symmetry-breaking of the probability distributional functional in space due to vector potential must be included and it has been done in the following way:

$$\sum_{s'} \int d^n x' P(x', s'|x, s) \Delta x_a \epsilon^{abc} \frac{\partial A_c}{\partial x^b} = \kappa_{1,s} \tag{15}$$

This constraint is also called the drift potential constraint in the context of a quantum mechanical complex system, but due to the drift of vector potential. Therefore, the $\kappa_{1,s}$ are the same constants that are related to vector equipotential lines in this case.

5.3 Optimization of MED Functional

The maximized multiscale MED functional of equation (9) subject to the constraints are given in equations (13), (14). We get the following result for the transition probability by combining both constraints.

$$P(x', s'|x, s) \propto \exp\left[-\frac{1}{2}\sum_{s'}\sum_{i}^{N_{s'}}\left(\frac{1}{\sigma^2_{s,i}}\delta_{ab}\Delta x_i^a \Delta x_i^b - \alpha'_{s,i}\Delta x_i^a \frac{\partial \phi_s}{\partial x_i^a} + \beta_{s',i}\Delta x_i^a A_a\right)\right], \tag{16}$$

where $\sigma^2_{s,i}$, $\alpha'_{s,i}$, and $\beta_{s',i}$ are Lagrange multipliers which will be expressed in the form of Planck's constant $\hbar$, speed of light, and charge of an electron. For the sake of completeness, we should note that the gauge invariance of equation (16) can be achieved and may be written in the following way:

$$P(x', s'|x, s) \propto \exp\left[-\frac{1}{2}\sum_{s'}\sum_{i}^{N_{s'}}\left(\frac{1}{\sigma^2_{s,i}}\delta_{ab}\Delta x_i^a \Delta x_i^b - \alpha'_{s,i}\Delta x_i^a \frac{\partial \phi_s}{\partial x_i^a} + \beta_{s',i}\Delta x_a \epsilon^{abc}\frac{\partial A_c}{\partial x^b}\right)\right], \tag{17}$$

In the MED formulation, we derive below the transition probability in the Gaussian form with time evolution. In this formulation, Brownian motion was applied explicitly to the general form transition probability as well as entropic time. Entropic time is explained and derived next.

Any notion of time must involve motion and change[37]. In MED, motion/change is described by the transition probability given by (16). It is desired to obtain small change. Large changes can be obtained by accumulating small short steps. It should be noted that any notion of time must have (a) something one might identify as an instant, (b) a sense in which these instants can be ordered, (c) a convenient concept of duration measuring the separation between instants [38]. In ED an instant is defined by the information required to generate the next instant. The point x occurs at time t and x' occurs at t'. So the probability distribution evolves according to

$$\rho(x', s', t') = \sum_s \int dt dx P(x', s'|x, s)\rho(x, s, t). \tag{18}$$

We write $\rho(x, s, t) = \rho_s(x, t)$. Having introduced the notion of time, the next step is to define a duration of time. Since our goal is to derive GSE, it suffices to construct a Newtonian interval that is independent of the position x and time t. This can be achieved by the Lagrange multiplier $\sigma^2_{s,i}$ to be constant such that

$$\frac{1}{\sigma^2_{s,i}} = \frac{m_{i,s}}{\eta_s \Delta t} \tag{19}$$

where $m_{i,s}$ are the particle masses and η_s is a constant which will be shown later that it is $\hbar$. Furthermore

$$M_{ab} = m_{s,i}\delta_{ab}\,, \tag{20}$$

where M_{ab} is effective mass matrix. We have

$$P(x', s'|x, s) \propto \exp\left[-\frac{1}{2\eta_s \Delta t} M_{ab}(\Delta x^a - \langle \Delta x^a \rangle)(\Delta x^b - \langle \Delta x^b \rangle)\right]. \tag{21}$$

Here

$$\Delta x^a = \langle \Delta x^a \rangle + \Delta w^a\,, \tag{22}$$

with

$$\langle \Delta x^a \rangle = \eta_s \Delta t M^{ab}\left(\alpha'_{s,i}\frac{\partial \phi_s}{\partial x^b} - \beta_s A_b\right), \tag{23}$$

$$\langle \Delta w^a \rangle = 0 \quad \text{and} \quad \langle \Delta w^a \Delta w^b \rangle = \eta_s \Delta t M^{ab}\,. \tag{24}$$

This is Brownian motion because the drift $\langle \Delta x^a \rangle \sim O(\Delta t)$ and the fluctuation $\Delta w^a \sim O(\Delta t^{1/2})$. The trajectory is continuous but not differentiable.

5.4 Dynamic Representation of MED Functional

The probability $\rho_s(x, t)$ evolves according to the Fokker-Planck (FP) equation and its proof is provided in Appendix B.

$$\frac{\partial \rho_s}{\partial t} = -\partial_a(v^a_s \rho_s)\,. \tag{25}$$

Note that s is the scaling index and $a = 1, 2.3$ are spatial indices. A summation over repeated indices should be understood. Here v^a is the current velocity given by

$$v^a_s = M^{ab}(\alpha'_{s,i}\partial_a \Phi_s - \beta_s A_a)\,, \text{where} \quad \Phi_s = \alpha'_{s,i}\eta_s \phi_s - \eta_s \log \rho_s^{1/2} \tag{26}$$

So far we only have one dynamical variable ρ_s which evolves according to Fokker-Planck (FP). To derive the Schrödinger equation we need two dynamical variables the probability ρ_s and phase Φ_s. To promote Φ_s to a dynamical variable, we need another constraint $H = H(\rho_s, \Phi_s)$, where H is an energy functional. By requiring that the energy is conserved, we obtain the second dynamical variable as well. The functional $H(\rho_s, \Phi_s)$ can be constructed by writing the FP equation as

$$\frac{\partial \rho_s}{\partial t} = \frac{\delta H}{\delta \Phi_s}\,. \tag{27}$$

It can easily be checked that the appropriate energy function is given by

$$H(\rho_s, \Phi_s) = \sum_s \int dx \rho_s \left(\frac{1}{2} M^{ab} (\partial_a \Phi_s - \beta_s A_a)(\partial_b \Phi_s - \beta_s A_b) + V_s(x) \right) + \sum_{ss'} g_{ss'} F(\rho_s, \rho_{s'})\,, \tag{28}$$

where $V(x)$ is a scalar potential and $F(\rho_s, \rho_{s'})$ is an integration to be determined below and $g_{ss'}$ is the complexity coefficient. This term leads to the nonlinear Schrödinger equation. We have

$$\frac{\delta H}{\delta \rho_s} = \frac{1}{2} M^{ab} (\partial_a \Phi_s - \beta_s A_a)(\partial_b \Phi_s - \beta_s A_b) + V_s(x) + \sum_{s'} g_{ss'} \frac{\delta F(\rho_s, \rho_{s'})}{\delta \rho_s} \tag{29}$$

Taking total time derivative of Eq. (28) and require it to be conserved and also incorporate Eq.(27)

$$\frac{dH}{dt} = \sum_s \int dx \left[\frac{\delta H}{\delta \Phi_s} \partial_t \Phi_s + \frac{\delta H}{\delta \rho_s} \partial_t \rho_s \right] = \sum_s \int dx \left[\partial_t \Phi_s + \frac{\delta H}{\delta \rho_s} \right] \partial_t \rho_s = 0 \tag{30}$$

It holds for all $\partial_t \rho_s$ which means that

$$\frac{\partial \Phi_s}{\partial t} = -\frac{\delta H}{\partial \rho_s}\,. \tag{31}$$

We get

$$\frac{\partial \Phi_s}{\partial t} = -\frac{1}{2} M^{ab} (\partial_a \Phi_s - \beta_s A_a)(\partial_b \Phi_s - \beta_s A_b) - V_s - \sum_{s'} g_{ss'} \frac{\delta F(\rho_s, \rho_{s'})}{\delta \rho_s}\,. \tag{32}$$

This is the quantum Hamilton-Jacobi equation. Eqs. (25) and (32) can be combined using

$$\psi_s = \rho_s^{1/2} \exp[ik\Phi_s/\eta_s]\,. \tag{33}$$

The result is

$$\begin{aligned} \frac{i\eta_s}{k}\frac{\partial \psi_s}{\partial t} &= \frac{\eta_s^2}{2k^2} M^{ab} (i\partial_a - \beta_s A_a)(i\partial_b - \beta_s A_a)\psi_s + V_s \psi_s + \frac{\eta_s^2}{2k^2} \frac{M^{ab} \partial_a \partial_b \sqrt{\rho_s}}{\sqrt{\rho_s}} \psi_s \\ &+ \sum_{s'} g_{ss'} \frac{\delta F(\rho_s, \rho_{s'})}{\delta \rho_s} \psi_s\,, \end{aligned} \tag{34}$$

5.5 Gauge Invariance of MED Derived Relations

The physical meaning of the ψ_s in equation (33) is that it represents the wavefunction of the particle that generalized Schrödinger equation. Its modulus is the probability of finding the particle in space and time. In most general situations e.g. data science, the ψ_s will be the parameter controlling the transition probabilities. Note that equation (34) is invariant under the gauge transformation given below [39]

$$\psi_s \to \psi_s' = e^{i\beta\chi(x,t)} \psi_s \quad \text{and} \quad A_a \to A_a' = A_a + \partial_a \chi. \tag{35}$$

In equation (34), the third term on the right is called 'quantum potential'. Normally this term is present in the Hamilton-Jacobi equation (32). On combining this equation with the Fokker-Planck equation (25), one obtains linear Schrödinger equation (LSE) that obeys the superposition principle. In our case, the quantum potential is implicit in $F(\rho_s, \rho_{s'})$. Since we have freedom in the choice of F, we choose it such that

$$\sum_{s'} g_{ss'} \frac{\delta F(\rho_s, \rho_{s'})}{\delta \rho_s} + \frac{\eta_s^2}{2k^2} \frac{M^{ab} \partial_a \partial_b \sqrt{\rho_s}}{\sqrt{\rho_s}} = \sum_{s'} g_{ss'} f(\rho_{s'})\,. \tag{36}$$

Note the function f on the right as a function of one variable only. If the goal is to obtain an LSE, one can set $f = 0$. However, we are interested in nonzero f for the reasons given below.

$$i\hbar \frac{\partial \psi_s}{\partial t} = \frac{\hbar^2}{2} M^{ab} (i\partial_a - \frac{e}{\hbar c} A_a)(i\partial_b - \frac{e}{\hbar c} A_a)\psi_s + V_s \psi_s + \sum_{s'} g_{ss'} f(\rho_{s'})\psi_s\,. \tag{37}$$

Here we used $\eta_s/k = \hbar$, and $\beta_s = e/\hbar c$, where e is the charge of an electron, and c is the speed of light. Equation (37) is the sought Generalized Nonlinear Schrödinger Equation (GNSE) that takes into account the EM field interaction with matter waves. The last term indicates nonlinearity. A similar last term is also reported in [36]. But here we naturally derived the general NLSE using Entropic dynamics. For solitons we can take $f(\rho_{s'}) = \rho_{s'} = |\psi_{s'}|^2$. So that

$$i\hbar\frac{\partial\psi_s}{\partial t} = \frac{\hbar^2}{2}M^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_b)\psi_s + V_s\psi_s + \sum_{s'} g_{ss'}|\psi_{s'}|^2\psi_s\,. \tag{38}$$

The existence of the nonlinear terms [40–42] and, in particular, cubic term [43] is characteristic of the solitons and other nonlinear waves (see, the seminal paper by Zakharov [44] about nonlinear stability of periodic waves in deep water). The existence of such waves and stable solitons depends on the boundary conditions and their dynamical stability depends on the spatial dimension of the system [43, 45]. Besides solitons, there is a large variety of nonlinear phenomena, including shape waves [42, 46], periodic waves in deep and shallow water [44], plasma cavitons [45, 47], Urbach and Lifshiz density of states tails [48, 49], the collapse of the plasmon, Langmuir waves [50] and many others related phenomena [51, 52]. It is very interesting if the NSE equations describing these or associated phenomena can be obtained with the principle of the maximum entropy and entropic dynamics as described above.

6. Reduction of GNSE to a Few Relevant Representations

One can note that equation (38) is a complex system of equations. It has some general form, which covers numerous physical phenomena. It may have both scalar or vector(tensor) forms [50, 53]. Below we will discuss those forms of NLS that found direct applications in different areas of physics, and it covers not only solitons but other quasi-particles too [43, 53, 54] including the phenomena such as self-trapping and polarons [52, 55, 56] as well as plasma caviton formation[47, 57, 58]. Here we consider the simplest example where two solitons coupled to each other may be created [59]. In one case, we obtain the vector nonlinear Schrödinger equation (VNSE) [56]. In the other case, the scalar nonlinear Schrödinger equation (SNSE) is obtained. The difference between VNSE and SNSE is that the former involves coupled solitons and the latter involves decoupled solitons as well as many different physical phenomena.

6.1 Scalar Form of GSE for Decoupled Solitons and Electrons

The solitons do usually exist in a one-dimensional chain or system with reduced dimensions [43]. The illustrative example is Davydov solitons [59], created in protein chains. They are associated with electron self-trapping or localization of Amide-I (or CO stretching) vibrational energy in proteins. Such localization as well as electron self-trapping arises through the interaction of the Amide-I mode with lattice distortion and plays an important role in charge transport vital for all biological systems. Our starting point is the system (38). For illustration, the EM field is set to zero ($\vec{A} = 0$). The SNSE can be obtained by setting

$$g_{ss'} = 0, \quad \text{when} \quad s \neq s' \tag{39}$$

where g_{11} and g_{22} survive. Further set $\eta_s/k = \hbar$. We obtain two decoupled SNSE's as follows

$$i\hbar\frac{\partial\psi_1}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_1 + V_1\psi_1 + g_{11}|\psi_1|^2\psi_1\,. \tag{40}$$

$$i\hbar\frac{\partial\psi_2}{\partial t} = -\frac{\hbar^2}{2m}\nabla_i^2\psi_2 + V_2\psi_2 + g_{22}|\psi_2|^2\psi_2\,, \tag{41}$$

which is the desired system of two decoupled solitons. The last two equations are known as the Gross-Pitaevski equation (GPE) [60, 61]. In Bose-Einstein condensate (BSE) $g < 0$ is referred to as the bright solitons and $g > 0$ is called dark solitons [61].

6.2 A Vector Form of the GSE for Coupled Solitons

We again recall the system (38) with $\vec{A} = 0$, for illustration. Set $g_{ss'}$ such that

$$g_{ss'} = 0, \quad \text{when} \quad s = s'\,, \tag{42}$$

where $s, s' = 1, 2$. Equation (38) simplifies to

$$i\hbar\frac{\partial\psi_1}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_1 + V_1\psi_1 + g_{12}|\psi_2|^2\psi_1 \,. \tag{43}$$

$$i\hbar\frac{\partial\psi_2}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_2 + V_2\psi_2 + g_{21}|\psi_1|^2\psi_2 \,, \tag{44}$$

which is the desired system of two solitons. Generally, we have

$$i\hbar\frac{\partial\psi_1}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_1 + V_1\psi_1 + \hbar g_{11}|\psi_1|^2\psi_1 + g_{12}|\psi_2|^2\psi_1 \,. \tag{45}$$

$$i\hbar\frac{\partial\psi_2}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_2 + V_2\psi_2 + \hbar g_{21}|\psi_1|^2\psi_2 + g_{22}|\psi_2|^2\psi_2 \,, \tag{46}$$

6.3 The Interaction of Electromagnetic Fields with Coupled and Decoupled Solitons

We can also write the full equation for solitons with EM-field. Recall equation (38)

$$i\hbar\frac{\partial\psi_s}{\partial t} = \frac{\hbar^2}{2}M^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_a)\psi_s + V_s\psi_s + \sum_{s'} g_{ss'}|\psi_{s'}|^2\psi_s \,. \tag{47}$$

where $M^{ab} = \delta^{ab}/m_i$ is the the inverse of mass matrix. For coupled solitons we have

$$i\hbar\frac{\partial\psi_1}{\partial t} = \frac{\hbar^2}{2m}\delta^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_a)\psi_1 + V_1\psi_1 + g_{12}|\psi_2|^2\psi_1 \,. \tag{48}$$

$$i\hbar\frac{\partial\psi_2}{\partial t} = \frac{\hbar^2}{2m}\delta^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_a)\psi_2 + V_2\psi_2 + g_{21}|\psi_1|^2\psi_2 \,. \tag{49}$$

Similarly, for decoupled solitons we have

$$i\hbar\frac{\partial\psi_1}{\partial t} = \frac{\hbar^2}{2m}\delta^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_a)\psi_1 + V_1\psi_1 + g_{11}|\psi_1|^2\psi_1 \,. \tag{50}$$

$$i\hbar\frac{\partial\psi_2}{\partial t} = \frac{\hbar^2}{2m}\delta^{ab}(i\partial_a - \frac{e}{\hbar c}A_a)(i\partial_b - \frac{e}{\hbar c}A_a)\psi_2 + V_2\psi_2 + g_{22}|\psi_2|^2\psi_2 \,. \tag{51}$$

In summary, equations (40) through (51) are special cases of equation (38). Generally, the scaling indices s, s' may vary as $1, 2, \ldots n$ that describes a system of n coupled equations or quasi-particles. It is also worthwhile to note that the coupling enters through the complex coefficient $g_{ss'}$. If it is set to zero, the equation (38) reduces to the linear SE that obeys the usual superposition principle for n particle system.

7. Discussion

In the 1950s, Ginsburg and Landau proposed a functional called the Ginsburg-Landau (GL) functional for free energy to describe the superconducting state in solids [30]. The GL functional was developed to introduce a non-linear term in the Schrödinger equation (SE) to describe the superconductivity property of conductors, by presenting the conduction electrons as super-fluids. The minimization of this GL functional gives rise to the nonlinear Schrödinger equation (NLSE). The NLSE describes a new state of quasi-particles, the superconducting condensate, which is similar to Bose-Einstein condensation (BEC) [24]. Gross and Pitaevskii later derived the NLSE by applying the minimum energy principle to the free energy of electrons or atoms in BEC. The use of NLSE enabled the description of the quantum dynamics of other systems in the form of solitary matter waves or solitons in coupled solitons containing multi-solitons, or having optically interacting solitons.

The NLSE has been widely used in many fields of physics, including condensed matter physics, nonlinear optics, and fluid dynamics. Solitons, which are localized wave packets that maintain their shape during propagation, have been described by the NLSE. These solitons can exist in many different systems, such as in optical fibers, plasmas, and superfluids. The maximization of MED functional not only resulted in an extension of GPE, but it also provided a natural way to include other interactions in it, such as the interaction of electromagnetic fields with quasi-particles in solids. Further, it provides the tools to deal with the dynamics of the scalar as well as vector solitons in decoupled and coupled forms. Whereas, it is to be noted that the GPE deals with scalar solitons. The implications of the GNSE may be quite far-reaching, in our opinion. Its application to 2D materials may lead to opportunities for discovering the quantized energies of solitons at the defect sites of those materials. Such quantized states may turn out to be suitable for future applications in electronics, including quantum computing.

In addition to this, the presented MED methodology is envisioned as a complementary way to statistical methods which are currently applied on existing datasets to forecast the future behavior of complex systems. The entire temporal evolution is intricately linked with a complex dataset and is delineated by a system of interconnected nonlinear Schrödinger equations. This methodology draws inspiration from entropy evolution, analogous to the Second Law of Thermodynamics—regarded as one of the most elegant laws in physics. Naturally, any complex system tends to evolve towards a state of equilibrium and stability. However, over short durations, the system might become ensnared by false minima, leading to potential discrepancies in the predictions of the developed theory. Hence, in the short term, the theory's predictions may falter. Nonetheless, machine learning can yield favorable outcomes in such instances. Conversely, in the long term, machine learning predictions may prove less reliable, while the approach grounded in the Second Law remains robust. This resilience is attributed to its foundation in the core principles of statistical mechanics.

8. Conclusions

The MED formulation was successfully applied to the derivation of GNSE which represents non-relativistic quantum mechanics in both linear and non-linear forms. The latter form describes the dynamics of quasi-particles. As an example, it describes the dynamics of matter-solitons in 2D materials for coupled and decoupled solitons. Furthermore, the development of machine learning and artificial intelligence techniques has further strengthened the connection between these fields. These techniques, which are used extensively in data science, are based on statistical and probabilistic models that are similar to those used in statistical physics and information theory. For example, deep learning models are based on neural networks that are similar to those used in statistical physics to model the behavior of physical systems.

In a recent paper by analysis of the huge database, the relationship between DNA methylation and mutability, specifically how methylation can affect the emergence of novel genetic variations in eukaryotes (organisms with cells containing a nucleus) has been identified [62]. Further analysis of somatic mutation data, particularly from cancers where specific repair pathways are compromised, is necessary to understand the underlying mechanisms of this process and the involvement of particular DNA repair pathways as well as how the impact of methylation on mutability extends beyond the methylated cytosine itself. DNA methylation, which is a common epigenetic modification in eukaryotes, can affect genetic variation in ways that are not fully understood and imply that the precise mechanisms involved are complex and require further investigation [62]. Their findings suggest that methylation has a significant impact on the emergence of novel genetic variants in eukaryotes, which may have important implications for understanding genetic diversity and disease. We hope that the application of the proposed MED methodology to the huge databases can further shed light on their evolutionary mechanisms [62]. Additionally, all three fields are concerned with the extraction of meaningful information from complex and noisy data sets. Statistical physics seeks to identify and understand the underlying patterns and structures that govern the behavior of physical systems. Information theory seeks to extract and transmit useful information from noisy or uncertain data sets. Data science seeks to extract insights and knowledge from large and complex data sets.

In summary, the fields of statistical physics, information theory, and data science are connected through their use of similar mathematical and statistical tools to model and analyze complex systems. These connections have become even stronger with the development of machine learning and artificial intelligence techniques, which rely heavily on the probabilistic and statistical models developed in these fields. The MED method can be adopted to improve the accuracy of data analysis and involve in optimizing the subset selection process to minimize the error in the prediction of the system's evolution. We expect that this method will be effective in a wide range of applications, including finance, healthcare, and social media analysis

Acknowledgement

SN would like to thank A. Caticha for many insightful discussions on entropic dynamics and especially for his comment on the role of momentum in ED. This work is partially supported by funds received under ADEK grant AARE-2019-131 and with the funds received under Khalifa University's grant ESIG-2023-003. FVK is grateful to 1000 Talents Award, China; PIFI CAS fellowship; Khalifa University of Science and Technology under Award Nos FSU-2021-030/8474000371 and the EU H2020 RISE project TERASSE (H2020-823878).

Data Availability

No data was used in this study because the manuscript is based on a theoretical or mathematical work. However, the detailed steps associated with our study have been deposited into a publicly available repository namely Quieos (https://www.qeios.com/read/B225L5).

Appendix A: Derivation of Equation (12)

Collecting the constraints, one has

$$\begin{aligned} 0 = &-\delta \sum_{s^prime} \int d^n x' Q(x', s'|x, s) \log \frac{Q(x', s\prime|x, s)}{\mu(\Delta x)} \\ &- \delta \sum_{s^prime} \int d^n \alpha_0 (Q(x', s'|x, s) - 1) \\ &- \delta \sum_{s^prime} \int d^n \sum_s \sum_i^{N_s} \frac{1}{2\sigma_{s,i}^2} Q(x'', s'))\delta_{ab} \Delta x_i^a \Delta x_i^b - \langle \Delta \ell_i^2 \rangle) \end{aligned} \tag{52}$$

Varying w.r.t. to Q and then set the integrant/summon equal to zero. One has

$$\log \frac{Q(x', s'|x, s)}{\mu(\Delta x)} = -1 - \alpha_0 - \sum_s \sum_i^{N_s} \frac{1}{2\sigma_{i,s}} \delta_{ab} \Delta x_i^a \Delta^{bi}_i \tag{53}$$

Absorbing $1 - \alpha_0$ into a new constant, one gets

$$Q(x', s'|x, s) = \frac{1}{Z} \exp \left[-\frac{1}{2} \sum_s \sum_i^{N_s} \frac{1}{\sigma_{i,s}} \delta_{ab} x_i Aa \Delta x_i^b \right] \tag{54}$$

Dropping the proportionality constant, Z, one gets (12)

$$Q(x', s'|x, s) \propto \exp \left[-\frac{1}{2} \sum_s \sum_i^{N_s} \frac{1}{\sigma_{i,s}} \delta_{ab} \Delta x_-^a \Delta x_i^b \right] \tag{55}$$

Appendix B: Derivation of Fokker-Planck Equation i.e., Equation (25)

The transition probability $P(x', s'|x, s)$ holds for short steps. Finite changes can be obtained by accumulating small changes according to

$$\rho(x, s, t) = \sum_{s_0} \int d^n x_0 \rho(x_0, s_0, t) P(x, s, t|x_0, s_0, t_0) \tag{56}$$

Write

$$P(x, s, t + \Delta t|x_0, t_0) = \sum_{s_z} \int d^n z P(z, s_z, t|x_0, s_0, t_0) P(x, s, t + \Delta t|x, s_z, t) \tag{57}$$

Multiply by a smooth a function $f(x)$, then integrate ove x, and then expand $f(x)$ about z, In a few steps one arrives at the FP equation

$$\frac{\partial \rho_s(x,t]}{\partial t} = \partial_a(\rho_s v_s^a) \tag{58}$$

where

$$v_s^a = M^{ab}(\alpha'_{s,i}\partial_b\phi_s - \beta_s A_b) \tag{59}$$

---

[1] E. T. Jaynes, Information theory and statistical mechanics. ii, Physical review **108**, 171 (1957).
[2] M. Sega, P. Faccioli, F. Pederiva, G. Garberoglio, and H. Orland, Quantitative protein dynamics from dominant folding pathways, Physical review letters **99**, 118102 (2007).
[3] A. Wehrl, General properties of entropy, Reviews of Modern Physics **50**, 221 (1978).
[4] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Principles of maximum entropy and maximum caliber in statistical physics, Reviews of Modern Physics **85**, 1115 (2013).
[5] K. Kaufmann, D. Maryanovsky, W. M. Mellor, C. Zhu, A. S. Rosengarten, T. J. Harrington, C. Oses, C. Toher, S. Curtarolo, and K. S. Vecchio, Discovery of high-entropy ceramics via machine learning, Npj Computational Materials **6**, 42 (2020).
[6] A. Caticha, Lectures on probability, entropy, and statistical physics, arXiv preprint arXiv:0808.0012 (2008).
[7] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).
[8] A. Tan, S. Shi, W.-Z. Wu, J. Li, and W. Pedrycz, Granularity and entropy of intuitionistic fuzzy information and their applications, IEEE Transactions on Cybernetics **52**, 192 (2020).
[9] D. T. Johnson and A. Caticha, Entropic dynamics and the quantum measurement problem, in *AIP Conference Proceedings 31st*, Vol. 1443 (American Institute of Physics, 2012) pp. 104–111.
[10] K. Vanslette and A. Caticha, Quantum measurement and weak values in entropic dynamics, in *AIP Conference Proceedings*, Vol. 1853 (AIP Publishing LLC, 2017) p. 090003.
[11] S. Nawaz and A. Caticha, Momentum and uncertainty relations in the entropic approach to quantum theory, in *AIP Conference Proceedings 31st*, Vol. 1443 (American Institute of Physics, 2012) pp. 112–119.
[12] S. Nawaz, M. Abedi, and A. Caticha, Entropic dynamics on curved spaces, in *AIP Conference Proceedings*, Vol. 1757 (AIP Publishing LLC, 2016) p. 030004.
[13] S. Ipek and A. Caticha, Entropic quantization of scalar fields, in *AIP Conference Proceedings*, Vol. 1641 (American Institute of Physics, 2015) pp. 345–352.
[14] S. Ipek, M. Abedi, and A. Caticha, Entropic dynamics: reconstructing quantum field theory in curved space-time, Classical and Quantum Gravity **36**, 205013 (2019).
[15] S. Ipek and A. Caticha, The entropic dynamics of quantum scalar fields coupled to gravity, Symmetry **12**, 1324 (2020).
[16] L. Bai, L. Cui, Z. Zhang, L. Xu, Y. Wang, and E. R. Hancock, Entropic dynamic time warping kernels for co-evolving financial time series analysis, IEEE Transactions on Neural Networks and Learning Systems (2020).
[17] A. Caticha, Entropic dynamics, time and quantum theory, Journal of Physics A: Mathematical and Theoretical **44**, 225303 (2011).
[18] A. Caticha, The entropic dynamics approach to quantum mechanics, Entropy **21**, 10.3390/e21100943 (2019).
[19] D. Bohm and B. J. Hiley, *The undivided universe: An ontological interpretation of quantum theory* (Routledge, 2006).
[20] F. Kusmartsev, Statistical mechanics of economics i, Physics Letters A **375**, 966 (2011).
[21] E. Nelson, Quantum fluctuations, in *Quantum Fluctuations* (Princeton University Press, 2021).
[22] B. Allen, B. C. Stacey, and Y. Bar-Yam, Multiscale information theory and the marginal utility of information, Entropy **19**, 273 (2017).
[23] L. D. Landau, Phys. Z. Sowjetunion **3**, 664 (1933).
[24] N. J. Zabusky and M. D. Kruskal, Interaction of" solitons" in a collisionless plasma and the recurrence of initial states, Physical review letters **15**, 240 (1965).
[25] J. W. Fleischer, M. Segev, N. K. Efremidis, and D. N. Christodoulides, Observation of two-dimensional discrete solitons in optically induced nonlinear photonic lattices, Nature **422**, 147 (2003).
[26] R. Zaera, J. Vila, J. Fernandez-Saez, and M. Ruzzene, Propagation of solitons in a two-dimensional nonlinear square lattice, International Journal of Non-Linear Mechanics **106**, 188 (2018).
[27] M. Liu, Z.-W. Wei, A.-P. Luo, W.-C. Xu, and Z.-C. Luo, Recent progress on applications of 2d material-decorated microfiber photonic devices in pulse shaping and all-optical signal processing, Nanophotonics **9**, 2641 (2020).
[28] T. Feng, X. Li, P. Guo, Y. Zhang, J. Liu, and H. Zhang, Mxene: two dimensional inorganic compounds, for generation of bound state soliton pulses in nonlinear optical system, Nanophotonics **9**, 2505 (2020).
[29] L. D. Landau, On the theory of phase transitions, Zh. Eksp. Teor. Fiz. **7**, 19 (1937).
[30] V. L. Ginzburg, V. L. Ginzburg, and L. Landau, *On the theory of superconductivity* (Springer, 2009).
[31] S. Pekar, Zh. Eksp. Teor. Fiz. **16**, 335 (1946).
[32] L. D. Landau and S. Pekar, Zh. Eksp. Teor. Fiz. **18**, 419 (1948).

[33] W. H. Sio and F. Giustino, Polarons in two-dimensional atomic crystals, Nature Physics , 1 (2023).
[34] L. P. Kadanoff, Fractals: Where's the physics?, Physics Today **39**, 6 (1986), https://doi.org/10.1063/1.2814878.
[35] B. B. Mandelbrot, *The fractal geometry of nature*, 3rd ed. (W. H. Freeman and Comp., New York, 1983).
[36] M. Soljacic, M. Segev, and C. R. Menyuk, Self-similarity and fractals in soliton-supporting systems, Phys. Rev. E **61**, R1048 (2000).
[37] A. Caticha, Change, time and information geometry, in *AIP Conference Proceedings*, Vol. 568 (American Institute of Physics, 2001) pp. 72–82.
[38] A. Caticha, Entropic time, in *AIP Conference Proceedings*, Vol. 1305 (American Institute of Physics, 2011) pp. 200–207.
[39] A. Caticha, Entropic physics: Probability, entropy, and the foundations of physics, Online at https://www. albany. edu/physics/faculty/ariel-caticha (2021).
[40] D. Gulevich and F. Kusmartsev, New phenomena in long josephson junctions, Superconductor Science and Technology **20**, S60 (2006).
[41] D. Gulevich and F. Kusmartsev, Perturbation theory for localized solutions of the sine-gordon equation: Decay of a breather and pinning by a microresistor, Physical Review B **74**, 214303 (2006).
[42] D. R. Gulevich, F. Kusmartsev, S. Savel'ev, V. Yampol'skii, and F. Nori, Shape waves in 2d josephson junctions: exact solutions and time dilation, Physical review letters **101**, 127002 (2008).
[43] F. Kusmartsev, On classification of solitons, Physica Scripta **29**, 7 (1984).
[44] V. E. Zakharov, Stability of periodic waves of finite amplitude on the surface of a deep fluid, Journal of Applied Mechanics and Technical Physics **9**, 190 (1968).
[45] V. Zakharov and E. Kuznetsov, On three dimensional solitons, Zhurnal Eksp. Teoret. Fiz **66**, 594 (1974).
[46] D. R. Gulevich, F. Kusmartsev, S. Savel'ev, V. Yampol'skii, and F. Nori, Shape and wobbling wave excitations in josephson junctions: Exact solutions of the (2+ 1)-dimensional sine-gordon model, Physical Review B **80**, 094509 (2009).
[47] F. Kusmartsev and E. Rashba, Self-trapping in crystals and nonlinear wave processes: Self-trapping barrier for plasma caviton, Zh. Eksp. Teor. Fiz - JETP **84**, 2064 (1983).
[48] F. Kusmartsev, Multiphonon absorption of light in nonpolar crystals, Physical Review B **43**, 1345 (1991).
[49] F. Kusmartsev and E. Rashba, Disappearing of fluctuation levels and abrupt end of density state tail, Fiz. and Tekn. Poluprovod **18**, 691 (1984).
[50] V. E. Zakharov *et al.*, Collapse of langmuir waves, Sov. Phys. JETP **35**, 908 (1972).
[51] F. Kusmartsev and E. Rashba, Symmetry breaking in the theory of self-trapping barrier and in allied problems, physica status solidi (b) **121**, K87 (1984).
[52] V. M. Fomin and E. P. Pokatilov, Non-equilibrium properties of charge carriers with arbitrary coupling to the lattice in semiconductors, Physics Reports **158**, 205 (1988).
[53] F. V. Kusmartsev, Application of catastrophe theory to molecules and solitons, Physics Reports **183**, 1 (1989).
[54] S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of solitons: the inverse scattering method* (Springer Science & Business Media, 1984).
[55] F. Kusmartsev and E. Rashba, Self trapping from degenerate bands (spin s= 1) and related phenomena, Harvard Library (1984).
[56] F. Kusmartsev and E. Rashba, Jahn-teller symmetry breaking of an autolocalized barrier, ZhETF Pisma Redaktsiiu -JETP Lett **33**, 164 (1981).
[57] V. Zakharov, S. Musher, and A. Rubenchik, Hamiltonian approach to the description of non-linear plasma phenomena, Physics reports **129**, 285 (1985).
[58] F. Kusmartsev and E. Rashba, Zh. eksper. teor. fiz., pisma 37, 106 (1983); cas soviet phys, J. exper. theor. Phys., Letters-JETP Lett **37**, 106 (1983).
[59] A. Scott, Davydov's soliton, Physics Reports **217**, 1 (1992).
[60] M. Soljacic, M. Segev, and C. R. Menyuk, Self-similarity and fractals in soliton-supporting systems, Physical Review E **61**, R1048 (2000).
[61] S. Damgaard Hansen, N. Nygaard, and K. Mølmer, Scattering of matter wave solitons on localized potentials, Applied Sciences **11**, 2294 (2021).
[62] V. Kusmartsev, M. Drożdż, B. Schuster-Böckler, and T. Warnecke, Cytosine methylation affects the mutability of neighboring nucleotides in germline and soma, Genetics **214**, 809 (2020).