

Physical Phenomenon from the Viewpoint of Information (Introduction to Quantum Information Theory)

Cheng-Yuan Liou and Jyh-Ying Peng

March 28, 2004

1 Introduction

We will introduce three theories that formulate the static description and dynamic evolution of quantum physics from the viewpoint of information. The first two theories study the possible information structures of quantum mechanics, the last one is the basis of quantum information theory. The informational viewpoint is also the viewpoint of measurement, since we gain information only by the act of measurement or observation. All refutable theories involving physical phenomenon are dependent on measurement results. We can even state that all physical phenomenon arises from the act of measurement or observation, this is the central concept of this work.

Classical physics holds the assumption that physical phenomenon have a reality independent of observation, and that any observer can perform measurements with arbitrary accuracy. Regarding the first point, we note that all useful physical theories (theories that can predict future physical states) are facts about measurement results, hence they can be verified or refuted by measurement. This means that physical theories need not have an existence independent of conscious measurement or observation, and that such an assumption is neither provable nor useful, and hence is unnecessary. Once we accept this, we realize that all theories are just models constructed by our minds to account for what we perceive in nature. That is, there are no “correct” physical theory, only theories that are more useful, or make better predictions about measurement results, the “reality” of a physical theory is meaningless.

To discuss the second point stated earlier, we have to construct a “model” for the act of observation. We take currently the most “accurate” physical

theory - quantum physics as the basis for the discussion. When we observe the position of an object, that object is no longer where we observed it, because the photon our eyes received gave it momentum. This means that there are no passive observation, all physical acts are interactions of some kind; as soon as we observe something, we change its physical state. So we say that measurements to infinite accuracy is not impossible, but rather the concept doesn't have any meaning (in our model of reality); in the physical world, there is no observation, only interaction. This is the viewpoint of quantum physics, and also the viewpoint adapted here. We will take the uncertainty of the measurement process as a starting point, and demonstrate how to derive physical theories from it.

2 Information in the Measurement Process - Fisher Information

2.1 Measurement

We will now start with a description of the classical measurement process. Measuring a physical system means estimating the value of some of its physical parameters by data obtained from the system. Let the ideal value of the parameter we are trying to estimate be θ , we obtain N data values $\mathbf{y} \equiv \{y_1, y_2, \dots, y_N\}$, the values of which are determined by conditional probability $p(\mathbf{y}|\theta)$. This conditional probability represents the intrinsic physical properties of the measured parameter.

Let the relationship between ideal value and measured data be

$$\mathbf{y} = \theta + \mathbf{x}, \tag{1}$$

where $\mathbf{x} \equiv \{x_1, x_2, \dots, x_N\}$ represents the measurement uncertainty, or data fluctuations. We can define an estimator $\hat{\theta}(\mathbf{y})$ for θ based on the obtained data \mathbf{y} . A possible estimator is the sample mean $\hat{\theta}(\mathbf{y}) = \frac{1}{N} \sum_n y_n$.

Here we step beyond classical physics, and interpret \mathbf{x} as a physical quantity intrinsic to the parameter θ , and independent of the measurement process. This means that \mathbf{x} represents the uncertainty of the parameter when it is observed, and its values are not dependent on the measurement methods used or measurement errors encountered; it is an intrinsic physical property of the measured parameter. So \mathbf{x} is the measurement uncertainty, *not* measurement error; it is a physical quantity independent of measurement, but manifested only by the measurement process (it has an "experimental reality", though it may not be "real").

If we accept that the uncertainty \mathbf{x} of a parameter θ has physical meaning, then \mathbf{y} , θ , and \mathbf{x} form a closed, physically isolated system.

2.2 The Cramer-Rao Inequality

We assume that the estimator $\hat{\theta}(\mathbf{y})$ is unbiased, that is,

$$\langle \hat{\theta}(\mathbf{y}) - \theta \rangle \equiv \int d\mathbf{y} [\hat{\theta}(\mathbf{y}) - \theta] p(\mathbf{y}|\theta) = 0, \quad (2)$$

where $d\mathbf{y} \equiv dy_1 dy_2 \dots dy_N$. Here $p(\mathbf{y}|\theta)$ is the conditional probability distribution of the parameter's fluctuation, or uncertainty, given that the value of the parameter is θ . Operating by $\frac{\partial}{\partial \theta}$ on both sides of equation (2), we get

$$\int d\mathbf{y} (\hat{\theta}(\mathbf{y}) - \theta) \frac{\partial p(\mathbf{y}|\theta)}{\partial \theta} - \int d\mathbf{y} p(\mathbf{y}|\theta) = 0. \quad (3)$$

The second term on the left is 1, using $\frac{\partial p}{\partial \theta} = p \frac{\partial \ln p}{\partial \theta}$, we have

$$\int d\mathbf{y} (\hat{\theta}(\mathbf{y}) - \theta) \frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta} p(\mathbf{y}|\theta) = 1. \quad (4)$$

Separate the integrand

$$\int d\mathbf{y} \left[\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta} \sqrt{p(\mathbf{y}|\theta)} \right] \left[(\hat{\theta}(\mathbf{y}) - \theta) \sqrt{p(\mathbf{y}|\theta)} \right] = 1 \quad (5)$$

and square both sides, from the Schwarz inequality we get

$$\left[\int d\mathbf{y} \left(\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta} \right)^2 p(\mathbf{y}|\theta) \right] \left[\int d\mathbf{y} (\hat{\theta}(\mathbf{y}) - \theta)^2 p(\mathbf{y}|\theta) \right] \geq 1. \quad (6)$$

The first term on the left of eq. (6) is defined as the Fisher information I for (the measurement of) the parameter θ ,

$$I(\theta) \equiv \int d\mathbf{y} \left(\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta} \right)^2 p(\mathbf{y}|\theta) = \int d\mathbf{y} \frac{1}{p(\mathbf{y}|\theta)} \left(\frac{\partial p(\mathbf{y}|\theta)}{\partial \theta} \right)^2, \quad (7)$$

and the second term is the mean-squared error for the estimator $\hat{\theta}(\mathbf{y})$

$$e^2 \equiv \int d\mathbf{y} (\hat{\theta}(\mathbf{y}) - \theta)^2 p(\mathbf{y}|\theta) = \langle (\hat{\theta}(\mathbf{y}) - \theta)^2 \rangle. \quad (8)$$

Thus we have the Cramer-Rao inequality

$$e^2 I \geq 1. \quad (9)$$

This inequality holds for any measurement using unbiased estimates of the measured parameter, it establishes the relationship between the Fisher information and mean-squared error of any measurement. As the estimation error increases, the Fisher information decreases, so the Fisher information can be seen as a measure of information. Moreover, because the estimation errors come only from \mathbf{x} , which is an intrinsic property of the parameter, the Fisher information represents the quality of measurement attainable when there is no measurement errors or human mistakes, its value depends only on the parameter and the measured system.

2.3 The Special Case of Shift-Invariance

Assume that we take only one data value, $N = 1$, and $p(\mathbf{y}|\theta) = p(y|\theta)$. If

$$p(y|\theta) = p(y - \theta) = p(x), \quad (10)$$

which means that the fluctuation of the data value y relative to the ideal value θ is independent of the value of θ , we call this property shift-invariance. Under this condition the fluctuation (or uncertainty) x is independent of the value of θ , hence the Fisher information is also independent of θ (in three dimensions we call this Galilean invariance, invariance of physical laws to changes of reference point). Since $\frac{\partial}{\partial \theta} = -\frac{\partial}{\partial (y-\theta)}$, the Fisher information can be written as

$$I = \int dy \frac{1}{p(y-\theta)} \left[\frac{\partial p(y-\theta)}{\partial (y-\theta)} \right]^2 = \int dx \frac{1}{p(x)} \left[\frac{dp(x)}{dx} \right]^2. \quad (11)$$

So we can calculate the uncertainty and Fisher information of parameter θ without knowing its ideal value. When a parameter satisfies shift-invariance, no matter what its ideal value, the fluctuations observed in measurements are the same. To simplify the discussion, we will assume all parameters satisfy this property.

2.4 Probability Amplitude Functions

In eq. (11) the term $\frac{1}{p(x)}$ would diverge when $p(x) \rightarrow 0$, we can define real probability amplitude functions $q(x)$ to avoid this problem:

$$p(x) = q^2(x). \quad (12)$$

Using this in eq. (11) we have

$$I = 4 \int dx \left(\frac{dq(x)}{dx} \right)^2 \quad (13)$$

2.5 Extension to Vectors and Multi-Dimensions

Under the framework of relativity, physical quantities are four-vectors such as $\mathbf{x} \equiv (x, y, z, t)$, we will define a scalar Fisher information for the measurement of multiple four-vectors.

Suppose we are measuring N four-vector parameters of a physical system, these four-vectors can represent any physical attribute of the system. We obtain the data

$$\mathbf{y}_n = \boldsymbol{\theta}_n + \mathbf{x}_n, \quad n = 1, \dots, N \quad (14)$$

where $\mathbf{y}_n \equiv (y_{n1}, y_{n2}, y_{n3}, y_{n4})$, $\boldsymbol{\theta}_n \equiv (\theta_{n1}, \theta_{n2}, \theta_{n3}, \theta_{n4})$, and $\mathbf{x}_n \equiv (x_{n1}, x_{n2}, x_{n3}, x_{n4})$ are all four-vectors, they are the obtained data, ideal value, and fluctuation of the parameters respectively. For notational simplicity, we also define grand vectors

$$\begin{aligned} \boldsymbol{\theta} &\equiv (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_N) \\ \mathbf{y} &\equiv (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \\ d\mathbf{y} &\equiv dy_1 dy_2 \dots dy_N, \end{aligned} \quad (15)$$

where $d\mathbf{y}_n \equiv dy_{n1} dy_{n2} dy_{n3} dy_{n4}$. If the measurement consists of the same physical four-vector parameter measured N times, or N particles sufficient distance apart measured for the same physical quantity, then measurement of the individual four-vectors are independent, and \mathbf{y}_n represents the data of the n -th measurement, or the n -th particle.

We assume that all estimators $\hat{\boldsymbol{\theta}}_n(\mathbf{y})$ on $\boldsymbol{\theta}_n$ are unbiased, so that

$$\langle \hat{\boldsymbol{\theta}}_n(\mathbf{y}) \rangle \equiv \int d\mathbf{y} \hat{\boldsymbol{\theta}}_n(\mathbf{y}) p(\mathbf{y}|\boldsymbol{\theta}) = \boldsymbol{\theta}_n, \quad n = 1, \dots, N \quad (16)$$

Using the mean-squared error of the four components of each four-vector, we can derive the Cramer-Rao inequality for each of them. Since the Fisher information is additive [1], we get the scalar Fisher information for the measurement

$$I \equiv \sum_n \int d\mathbf{y} p(\mathbf{y}|\boldsymbol{\theta}) \sum_\nu \left(\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_{n\nu}} \right)^2 \quad (17)$$

If we assume independence between measurements of all four-vectors, then

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_n p_n(\mathbf{y}_n|\boldsymbol{\theta}_n) = \prod_n p_n(\mathbf{y}_n|\boldsymbol{\theta}_n). \quad (18)$$

So that

$$\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_{n\nu}} = \sum_m \frac{\partial \ln p_m(\mathbf{y}_m|\boldsymbol{\theta}_m)}{\partial \theta_{m\nu}} = \frac{1}{p_n(\mathbf{y}_n|\boldsymbol{\theta}_n)} \frac{\partial p_n(\mathbf{y}_n|\boldsymbol{\theta}_n)}{\partial \theta_{n\nu}}. \quad (19)$$

Using eq. (18) and eq. (19) to simplify the Fisher information eq. (17), we have

$$I = \sum_n \int d\mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\boldsymbol{\theta}_n)} \sum_\nu \left(\frac{\partial p_n(\mathbf{y}_n|\boldsymbol{\theta}_n)}{\partial \theta_{n\nu}} \right)^2. \quad (20)$$

For shift-invariant four-vectors

$$p_n(\mathbf{y}_n|\boldsymbol{\theta}_n) = p_n(\mathbf{y}_n - \boldsymbol{\theta}_n) = p_n(\mathbf{x}_n) \quad (21)$$

the Fisher information can be further simplified to

$$I = \sum_n \int d\mathbf{x}_n \frac{1}{p_n(\mathbf{x}_n)} \sum_\nu \left(\frac{\partial p_n(\mathbf{x}_n)}{\partial x_{n\nu}} \right)^2. \quad (22)$$

Using probability amplitudes $q_n(\mathbf{x}_n)$, we have

$$I = 4 \sum_n \int d\mathbf{x}_n \sum_\nu \left(\frac{\partial q_n(\mathbf{x}_n)}{\partial x_{n\nu}} \right)^2, \quad p_n(\mathbf{x}_n) = q_n^2(\mathbf{x}_n). \quad (23)$$

Finally, if all the parameters $\boldsymbol{\theta}_n$'s represent the same physical quantity, then all fluctuations \mathbf{x}_n are equivalent, we can then drop the subscript n on \mathbf{x}_n and get

$$I = 4 \sum_n \int d\mathbf{x} \sum_\nu \left(\frac{\partial q_n(\mathbf{x})}{\partial x_\nu} \right)^2, \quad (24)$$

where $\mathbf{x} \equiv (x_1, x_2, x_3, x_4)$ is the uncertainty of each four-vector.

2.6 Total Probability Function

If the measurement in the previous section is N measurements for the same four-vector parameter of the same particle, then each \mathbf{x}_n represents the same physical quantity's n -th measurement result. Since there is only one parameter being measured, we can set

$$p_n(\mathbf{y}_n|\boldsymbol{\theta}_n) = p_n(\mathbf{y}|\boldsymbol{\theta}_n) = p_{x_n}(\mathbf{x}|\boldsymbol{\theta}_n) = q_n^2(\mathbf{x}). \quad (25)$$

The probability of measuring any of the $\boldsymbol{\theta}_n$'s is equal, so the total probability (or net probability) function of the fluctuation four-vector \mathbf{x} is

$$p(\mathbf{x}) = \sum_{n=1}^N p_{x_n}(\mathbf{x}|\boldsymbol{\theta}_n) P(\boldsymbol{\theta}_n) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}). \quad (26)$$

3 Derivation of Physical Laws by EPI

3.1 The Basics

The principle of EPI (extreme physical information) by B. Roy Frieden [1] rests on the assumption that the act of measurement (or observation) produces the physical laws.

The hierarchy of physical knowledge according to EPI consists of four layers. The topmost laws are:

1. **The Fisher I -theorem.** The Fisher information monotonically decreases with time. Like entropy, I can be transferred from one system to another.
2. **Bound information J .** There is an information bound J intrinsic to each physical phenomenon. The information J represents an upper limit to the information gained in measurement.
3. **Invariance principle.** There is an invariance, or symmetry principle governing the time evolution of each physical phenomenon.

These laws exist prior to and independent of any explicit measurements, that is, they govern all physical phenomenon. At the next level we have the three axioms describing the measurement process:

Axiom 1. The conservation of information perturbation, $\delta I = \delta J$, during a measurement.

Axiom 2. The existence of information densities $i_n(\mathbf{x})$ and $j_n(\mathbf{x})$ defined as

$$I \equiv \int d\mathbf{x} \sum_n i_n(\mathbf{x}) \text{ and } J \equiv \int d\mathbf{x} \sum_n j_n(\mathbf{x}), \quad (27)$$

where $i_n(\mathbf{x}) = 4\nabla q_n \cdot \nabla q_n = 4 \sum_\nu \left(\frac{\partial q_n}{\partial x_{n\nu}} \right)^2$.

Axiom 3. The efficiency of information transition from phenomenon to intrinsic data on the microlevel

$$i_n(\mathbf{x}) - \kappa j_n(\mathbf{x}) = 0, \forall \mathbf{x}, n. \quad (28)$$

The third level of knowledge consists of the EPI principle, which include the variational principle and the zero-condition. The variational principle states the extremization of $K[\mathbf{q}] \equiv I[\mathbf{q}] - J[\mathbf{q}]$, that is

$$\delta K = \delta(I - J) = 0. \quad (29)$$

The zero-condition is the efficiency of information transition on the macrolevel

$$I = \kappa J. \quad (30)$$

These follows either from the axioms or from the existence of a physically meaningful unitary transformation space.

The fourth and lowest level is the physical laws found by calculation of EPI, that is, the $q_n(\mathbf{x}_n)$'s found to extremize K , where the form of J is determined by the invariance principles for the particular phenomenon in question.

3.2 The Schrödinger Wave Equation

This derivation runs parallel to the fully covariant EPI derivation of the Klein-Gordon equation in [1], chapter 4. However the Schrödinger equation treats space and time coordinates differently, time coordinates are assumed to have no fluctuations, and can always be determined precisely, whereas space coordinates cannot be determined with arbitrary accuracy. Since the EPI approach is covariant while the Schrödinger equation is not, some approximations have to be made in the derivation, the end result of which is also an approximation of nature.

We will derive the one-dimensional time-independent Schrödinger equation. The position θ of a particle of mass m is measured as data $y = \theta + x$, where x is a random excursion governed by probability amplitudes $q_n(x)$, which is to be found. We ignore time t in the derivation, hence we will get a stationary solution to this problem. The particle is assumed to be in a conservative field of scalar potential $V(x)$, with total energy W conserved.

The information associated with the measurement of position is

$$I = 4 \sum_{n=1}^N \int dx \left(\frac{dq_n(x)}{dx} \right)^2 \quad (31)$$

in the one-dimensional case. We define the complex wave functions $\psi_n(x)$ as

$$\psi_n(x) \equiv \frac{1}{\sqrt{N}}(q_{2n-1}(x) + iq_{2n}(x)), \quad (32)$$

where there are $N/2$ of them. The information expressed with the $\psi_n(x)$'s becomes

$$I = 4N \sum_{n=1}^{N/2} \int dx \left(\frac{d\psi_n(x)}{dx} \right)^* \left(\frac{d\psi_n(x)}{dx} \right) = 4N \sum_{n=1}^{N/2} \int dx \left| \frac{d\psi_n(x)}{dx} \right|^2. \quad (33)$$

We next define a Fourier transform space consisting of wave functions $\phi_n(\mu)$ of momentum μ

$$\psi_n(x) = \frac{1}{\sqrt{2\pi\hbar}} \int d\mu \phi_n(\mu) e^{-\frac{i\mu x}{\hbar}}, \quad (34)$$

operating on both sides by $\frac{d}{dx}$ we get

$$\frac{d\psi_n(x)}{dx} = \frac{1}{\sqrt{2\pi\hbar}} \int d\mu \left(-\frac{i\mu}{\hbar} \phi_n(\mu) \right) e^{-\frac{i\mu x}{\hbar}}. \quad (35)$$

So we have

$$\psi_n(x) \xleftrightarrow{\mathcal{F}} \phi_n(\mu) \quad (36)$$

and

$$\frac{d\psi_n(x)}{dx} \xleftrightarrow{\mathcal{F}} -\frac{i\mu}{\hbar} \phi_n(\mu), \quad (37)$$

then by Parseval's theorem we have

$$\int dx |\psi_n(x)|^2 = \int d\mu |\phi_n(\mu)|^2 \quad (38)$$

and

$$\int dx \left| \frac{d\psi_n(x)}{dx} \right|^2 = \frac{1}{\hbar^2} \int d\mu \mu^2 |\phi_n(\mu)|^2. \quad (39)$$

Using equation (39) in the information expression we have

$$I = \frac{4N}{\hbar^2} \int d\mu \mu^2 \sum_{n=1}^{N/2} |\phi_n(\mu)|^2 \equiv J, \quad (40)$$

that is, the unitary nature of the fourier transform inherent in the nature of the measurement device gives rise to the invariance principle (40), where J is the bound information, and I is constrained to be equal to J . The bound information represents the maximum information obtainable from the measurement, with $I = J$ this means that information is transferred with maximum efficiency. Yet only the workings of the input side of the measurement device is described by the fourier transform, we have not considered the measurement device's output yet, hence the hallmark of quantum phenomenon, uncertainty in the precise state of physical systems is not included in this derivation. In other words, we are describing the situation where the Schrödinger cat experiment is completed, yet no one has opened the box and looked at the cat yet. For more detail see [1], 3.8 and chapter 10.

The total probability distribution for variable x is (by eq. (26))

$$p(x) = \frac{1}{N} \sum_{n=1}^N q_n^2(x) = \sum_{n=1}^{N/2} |\psi_n(x)|^2, \quad (41)$$

so by equation (38)

$$\int d\mu \sum_n |\phi_n(\mu)|^2 = \int dx \sum_n |\psi_n(x)|^2 = \int dx p(x) = 1, \quad (42)$$

so that

$$P(\mu) = \sum_n |\phi_n(\mu)|^2 \quad (43)$$

is a probability density on μ . Thus we have

$$J = \frac{4N}{\hbar^2} \int d\mu \mu^2 \sum_{n=1}^{N/2} |\phi_n(\mu)|^2 = \frac{4N}{\hbar^2} \int d\mu \mu^2 P(\mu) = \frac{4N}{\hbar^2} \langle \mu^2 \rangle. \quad (44)$$

We use the non-relativistic approximation that the kinetic energy of the particle is $\frac{\mu^2}{2m}$, so $W = V(x) + \frac{\mu^2}{2m}$, we have

$$\begin{aligned} J &= \frac{4N}{\hbar^2} \langle \mu^2 \rangle \\ &= \frac{8Nm}{\hbar^2} \langle W - V(x) \rangle \\ &= \frac{8Nm}{\hbar^2} \int dx (W - V(x)) p(x) \\ &= \frac{8Nm}{\hbar^2} \int dx (W - V(x)) \sum_{n=1}^{N/2} |\psi_n(x)|^2. \end{aligned} \quad (45)$$

Thus we have successfully expressed J as a functional of the ψ_n 's, $J[\psi]$ is the bound information functional for this problem, and $I[\psi] = J[\psi]$.

According to the principle of extreme physical information, $K = I - J$ is extremized, that is

$$K = I - J = 4N \sum_{n=1}^{N/2} \int dx \left[\left| \frac{d\psi_n(x)}{dx} \right|^2 - \frac{2m}{\hbar^2} [W - V(x)] |\psi_n(x)|^2 \right] = \text{Extrem.} \quad (46)$$

The Euler-Lagrange equation for the variational problem is

$$\frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial \psi_{nx}^*} \right) = \frac{\partial \mathcal{L}}{\partial \psi_n^*}, \quad n = 1, \dots, N/2, \quad \psi_{nx}^* \equiv \frac{\partial \psi_n^*}{\partial x}, \quad (47)$$

using the integrand in equation (46) as the Lagrangian \mathcal{L} , the solution to this variational problem is

$$\frac{d^2 \psi_n(x)}{dx^2} + \frac{2m}{\hbar^2} [W - V(x)] \psi_n(x) = 0, \quad n = 1, \dots, N/2, \quad (48)$$

the time-independent Schrödinger wave equation.

Since the solution (SWE) is the same for each index n , an $N = 2$ solution is permitted, that is, the SWE defines a single complex wave function $\psi(x) = \frac{1}{\sqrt{2}}(q_1(x) + iq_2(x))$ and

$$\frac{d^2\psi(x)}{dx^2} + \frac{2m}{\hbar^2}[W - V(x)]\psi(x) = 0. \quad (49)$$

3.3 Uncertainty Principles

According to the Heisenberg uncertainty principle, a particle's position and momentum intrinsically fluctuates by amounts x and μ from ideal (classical) values θ_x and θ_μ with variances ϵ_x^2 and ϵ_μ^2 obeying

$$\epsilon_x^2 \epsilon_\mu^2 \geq \left(\frac{\hbar}{2}\right)^2. \quad (50)$$

This relation is conventionally derived from the fourier transform relation eq. (36) between position and momentum spaces.

This result may also be proved using the Cramer-Rao inequality of Fisher information. The mean-square error for position (θ_x) measurements is defined as

$$e_x^2 \equiv \langle (\hat{\theta}_x(y) - \theta_x)^2 \rangle, \quad (51)$$

where $\hat{\theta}_x(y)$ is a general estimator for the ideal position θ_x based on measured data y . Suppose the probability distribution for x is $p(x)$, then the Fisher information for the variable x is

$$I_x = \int dx \frac{1}{p(x)} \left(\frac{dp(x)}{dx} \right)^2. \quad (52)$$

The Cramer-Rao inequality states that

$$e_x^2 I_x \geq 1. \quad (53)$$

The one-dimensional wave function for a quantum particle is derived in the previous section, suppose the solution is attained with $N = 2$, then there is only one $\psi_1(x) = \psi(x)$. The Fisher information (on one-dimension variable) for the quantum particle is

$$I = 8 \int dx \left| \frac{d\psi(x)}{dx} \right|^2. \quad (54)$$

Since $|\psi(x)|^2 = p(x)$ we have $|\psi(x)| = \sqrt{p(x)}$ so that

$$\begin{aligned}
I &= 8 \int dx \left| \frac{d}{dx} |\psi(x)| e^{iS(x)} \right|^2 \\
&= 8 \int dx \left| \frac{d}{dx} \sqrt{p(x)} e^{iS(x)} \right|^2 \\
&= 8 \int dx \left| \frac{1}{2\sqrt{p(x)}} \frac{dp(x)}{dx} e^{iS(x)} + \sqrt{p(x)} i \frac{dS(x)}{dx} e^{iS(x)} \right|^2 \\
&= 8 \int dx \left| \frac{1}{2\sqrt{p(x)}} \frac{dp(x)}{dx} + \sqrt{p(x)} i \frac{dS(x)}{dx} \right|^2 \\
&= 8 \int dx \left(\frac{1}{4p(x)} \left(\frac{dp(x)}{dx} \right)^2 + p(x) \left(\frac{dS(x)}{dx} \right)^2 \right) \\
&= 2 \int dx \frac{1}{p(x)} \left(\frac{dp(x)}{dx} \right)^2 + 8 \int dx p(x) \left(\frac{dS(x)}{dx} \right)^2 \\
&= 2I_x + 8 \left\langle \left(\frac{dS(x)}{dx} \right)^2 \right\rangle,
\end{aligned}$$

where $S(x) \in \Re$ is the phase of $\psi(x)$. So we have $2I_x \leq I$, from eq. (99) we have

$$I = J = \frac{8}{\hbar^2} \langle \mu^2 \rangle, \quad (55)$$

so

$$I_x \leq \frac{4}{\hbar^2} \langle \mu^2 \rangle = \frac{4}{\hbar^2} \epsilon_\mu^2, \quad (56)$$

since μ is the fluctuation in momentum. Using the Cramer-Rao inequality we have

$$e_x^2 \frac{4}{\hbar^2} \epsilon_\mu^2 \geq e_x^2 I_x \geq 1, \quad (57)$$

and finally

$$e_x^2 \epsilon_\mu^2 \geq \left(\frac{\hbar}{2} \right)^2. \quad (58)$$

There is a subtle difference between ϵ_x^2 in the Heisenberg uncertainty and e_x^2 used earlier, the former represents the variance of the position fluctuation distribution of a particle, and is independent of any measurements; while the latter is a measure of the quality of a position measurement and the subsequent position estimate, which depends on the intrinsic properties of a particle, but is only manifested by actual measurements. This difference

results in different interpretations of the meaning of the Heisenberg uncertainty and its Fisher version, the former treats the fluctuations as intrinsic and exists independent of any observation, while the latter inequality arises when a measurement of position is actually made, that is, when a position measurement is made on a particle, its momentum would exhibit a fluctuation governed by the uncertainty principle. The latter interpretation is consistent with the EPI principle in that the uncertainty is intrinsic to the phenomenon, but only by an actual observation can its effects be felt.

3.4 Boltzmann Energy Distribution

We will derive the Boltzmann energy distribution law for a perfect gas in equilibrium. The gas is composed of M identical molecules within a container, all collisions with other molecules and container walls are assumed to be elastic. The gas has temperature T .

We again start the derivation on a covariant basis and choose the fisher coordinates to be

$$x_0 \equiv ix_E, \mathbf{x}_\mu \equiv c\mu \equiv (c\mu_1, c\mu_2, c\mu_3). \quad (59)$$

We take the non-relativistic approximation and treat the energy fluctuation x_E and the momentum fluctuations \mathbf{x}_μ separately. We are only interested in deriving the law on energy, hence the subscript on x_E is dropped and the measured value of the energy is

$$E = \theta_E + x, \quad E_0 \leq E \leq \infty, \quad (60)$$

where θ_E is the ideal value of the energy.

Thus we have

$$I(E) = -4 \int dx \sum_n \left(\frac{dq_n(x)}{dx} \right)^2, \quad (61)$$

where the probability amplitudes $q_n(x)$ relate to the probability distribution function by

$$p(x) = \frac{1}{N} \sum_n q_n^2(x). \quad (62)$$

The negativity of $I(E)$ is due to the use of imaginary coordinate for energy, which is justified later. The goal of this analysis is then to solve the two EPI principles

$$I(E) - J(E) = \text{extrem.} \quad \text{and} \quad I(E) = \kappa J(E), \quad (63)$$

the extremization of physical information and the zero-condition respectively.

We find the bound information functional $J[\mathbf{q}]$ by assuming that both EPI principle yields the same solution $q_n(x)$'s. The more general form $J[\mathbf{q}, x]$ is not needed since the extra dependence on energy x explicitly only yields non-equilibrium solutions, and will be discarded after further derivations.

According to axiom 2, the existence of information densities, we can represent $J[\mathbf{q}]$ as

$$J[\mathbf{q}] = 4 \int dx \sum_n J_n(q_n(x)), \quad (64)$$

hence we have

$$K \equiv I - J = -4 \int dx \sum_n (q_n'^2(x) + J_n(q_n(x))), \quad q_n'(x) = \frac{dq_n(x)}{dx}, \quad (65)$$

and the extremum principle results in the Euler-Lagrange equation of the integrand \mathcal{L}

$$\frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial q_n'(x)} \right) = \frac{\partial \mathcal{L}}{\partial q_n(x)}, \quad (66)$$

the solution of which is

$$\frac{d^2 q_n}{dx^2} = \frac{1}{2} \frac{dJ_n}{dq_n}, \quad n = 1, \dots, N. \quad (67)$$

For the information efficiency we first change the form of I by noting that

$$\int dx \left(\frac{dq_n(x)}{dx} \right)^2 = \left. \frac{dq_n(x)}{dx} q_n(x) \right]_{E_0 - \theta_E}^{\infty} - \int dx q_n(x) \frac{d^2 q_n(x)}{dx^2}, \quad (68)$$

the first term of the result is zero since we assume that the probability of the energy to be infinity or E_0 to be zero. So we have

$$I = 4 \int dx \sum_n q_n(x) \frac{d^2 q_n(x)}{dx^2}, \quad (69)$$

by the energy efficiency zero-condition we have

$$I - \kappa J = 4 \int dx \sum_n \left(q_n(x) \frac{d^2 q_n(x)}{dx^2} - \kappa J_n(q_n(x)) \right) = 0. \quad (70)$$

By axiom 3, the efficiency on the microlevel we have

$$q_n(x) \frac{d^2 q_n(x)}{dx^2} - \kappa J_n(q_n(x)) = 0, \quad n = 1, \dots, N. \quad (71)$$

Combining eqs. (67) and (71) we find that the q_n 's obey

$$\frac{\kappa J_n}{q_n} = \frac{1}{2} \frac{dJ_n}{dq_n}, \quad (72)$$

or

$$\frac{2\kappa}{q_n} = \frac{1}{2} \frac{d}{dq_n} \ln J_n. \quad (73)$$

Intgrating both sides by q_n we have

$$J_n(q_n) = A_n q_n^{2\kappa}, \quad A_n \geq 0. \quad (74)$$

Using this equation in either eq. (67) or eq. (71) yields

$$\frac{d^2 q_n(x)}{dx^2} = \alpha_n^2 q_n^{2\kappa-1}(x), \quad \alpha_n^2 \equiv \kappa A_n \geq 0. \quad (75)$$

For the invariance principle we use the normalization of $p(E)$, which is weak in the sense that any p.d.f. is normalized. Hence for this phenomenon we have minimum prior information, that is, maximum ignorance about the independent variable in question. By classical assumptions, we set $\kappa = 1$, which represents maximum information transfer from the bound information of the phenomenon to the fisher information we gain as a result of observation. So we have

$$\frac{d^2 q_n(x)}{dx^2} = \alpha_n^2 q_n(x), \quad (76)$$

the general solution of which is

$$q_n(x) = B_n e^{-\alpha_n x} + C_n e^{\alpha_n x}, \quad \alpha_n \geq 0. \quad (77)$$

Since x is bounded below (by $E_0 - \theta_E$) yet unbounded above, the C_n 's must vanish for $p(E)$ to be normalizable. The solution now becomes

$$q_n(x) = B_n e^{-\alpha_n x}. \quad (78)$$

In retrospective if the coordinate of energy $x_0 \equiv ix_E$ is taken to be real, then the right hand side of eq. (75) would be negative, and we would have obtained the general solution

$$q_n(x) = B_n e^{-i\alpha_n x} + C_n e^{i\alpha_n x}, \quad (79)$$

which is sinusoidal. Which then causes $p(E)$ to be sinusoidal too, and cannot be normalized.

From the form of the solution the N for this problem need only be one, so we have

$$q(x) = B e^{-\alpha x}, \quad (80)$$

and

$$p(x) = B^2 e^{-2\alpha x}. \quad (81)$$

Using the change of variable $x = E - \theta_E$ we have

$$p(E) = C e^{-2\alpha E}, \quad C \equiv B^2 e^{2\alpha\theta_E}. \quad (82)$$

We find the value of constants C and α by normalization and in terms of the expectation value of E :

$$\int_{E_0}^{\infty} p(E) dE = 1, \quad (83)$$

$$\langle E \rangle = \int_{E_0}^{\infty} E p(E) dE. \quad (84)$$

We have

$$p(E) = \frac{1}{\langle E \rangle - E_0} e^{-\frac{E-E_0}{\langle E \rangle - E_0}}, \quad E \geq E_0, \quad (85)$$

and $p(E) = 0$ for other values of E . Shifting the origin of E by a constant does not change the physical law, so we subtract E_0 from E and get

$$p(E) = \langle E \rangle^{-1} e^{-\frac{E}{\langle E \rangle}}, \quad E \geq 0. \quad (86)$$

The energy of a perfect gas in equilibrium with three degrees of freedom per molecule is

$$\langle E_t \rangle = \frac{3MkT}{2}, \quad (87)$$

so

$$\langle E \rangle = \frac{\langle E_t \rangle}{M} = \frac{3kT}{2}, \quad (88)$$

and the energy distribution is

$$p(E) = \frac{2}{3kT} e^{-\frac{2E}{3kT}}. \quad (89)$$

3.5 Newton's Law of Motion

Lastly we present a mock derivation of Newton's law of motion using a pseudo EPI procedure. We assume that the energy of a particle has two forms, kinetic and conservative potential, and the total energy is constant. We

define the position perturbation $q(t)$ in terms of a function of time, then the constant energy requirement becomes

$$E = \frac{1}{2}m \left(\frac{dq(t)}{dt} \right)^2 - V(q(t)) = \text{const.} \quad (90)$$

The fisher information for this phenomenon is

$$I = 4 \int dt \left(\frac{dq(t)}{dt} \right)^2 = \frac{8}{m} \int dt (E + V(q(t))) \equiv J, \quad (91)$$

and $\kappa = 1$ here. We have perfect efficiency because time can be measured to infinite accuracy in the non-relativistic picture.

We extremize

$$K = I - J = 4 \int dt \left[\left(\frac{dq(t)}{dt} \right)^2 - \frac{2E}{m} - \frac{2}{m} V(q(t)) \right] \quad (92)$$

using the Euler-Lagrange equation and get the solution

$$m \frac{d^2 q(t)}{dt^2} = - \frac{dV(q(t))}{dq(t)}, \quad (93)$$

which is Newton's law of motion.

4 The Geometrical Representation of Physical Phenomenon

According to the view of Italian physicist E. R. Caianiello, uncertainty is inherent in all branches of science, he obtained a geometrical representation of physics, especially quantum physics, using the theories and methods of information geometry. In his formulation [2], the quantum physical uncertainty appears as a "curvature" in relativistic phase space. He also tries to combine such representation (quantum geometry) to theories of entropy and information, so as to find a theoretical foundation for such representations. Like the principle of EPI, his goal is to describe physical phenomenon from the viewpoint of information. But the theories of information geometry, or the geometrization of information theory used for its foundation are more general than the Fisher information, and much more difficult to comprehend.

4.1 Information Geometry

Information geometry is a specialization of differential geometry that deals with the geometrical structure of probability distributions. Under this formulation, probability distributions are treated as points on a manifold, and the Fisher information is the distance between different probability distributions on these manifolds.

According to Caianiello [7], the geometrical representation of a model (or theory) requires essentially the choice of a metric G and a connection Γ_μ and the identification of a reference frame; these depend upon the “universe” one wishes to model. Below we give a short introduction to these concepts of differential geometry and apply them to the special case of information geometry (see [8] for an introduction).

4.1.1 Manifolds

An N -dimensional manifold \mathcal{M}_N is a “curved” space embedded in an M -dimensional affine space \mathcal{E}_M , where $M > N$. For instance, a curved surface or the surface of a sphere in 3-dimensions are both 2-dimensional manifolds. A point on an N -dimensional manifold can be specified with N parameters (or coordinates) $\mathbf{x} \equiv \{x^1, \dots, x^N\}$, so there is a mapping from each point of the manifold to a point on an affine space \mathcal{E}_N .

Because probability distributions (probability amplitude and wave functions) are used in quantum physics, Caianiello used manifolds with parametric distributions as points, and extend the definitions so obtained to descriptions of quantum physics. The form of the most general parametric distribution is $\rho_0(\mathbf{x}|\mathbf{z})$ where $\mathbf{z} \equiv (z_1, z_2, \dots, z_m) \in R^m$ is the random output and $\mathbf{x} \equiv (x^1, x^2, \dots, x^n) \in R^n$ represent the parameters (the subscripts and superscripts are all indices). So \mathbf{x} is a point in \mathcal{E}_N , but its also a point in the manifold \mathcal{M}_N formed by the probability distributions $\rho_0(\mathbf{x}|\mathbf{z})$. The parameters \mathbf{x} provide a coordinate system for \mathcal{M}_N , each point of which represents a different probability distribution.

We will use Gaussian distributions to illustrate these concepts. The mean μ and standard deviation σ determines a Gaussian distribution

$$\rho(\mathbf{x}|z) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(z - \mu)^2}{2\sigma^2} \right], \quad (94)$$

where z is the single random output, and the parameter $\mathbf{x} \equiv (x^1 = f^1(\mu, \sigma), x^2 = f^2(\mu, \sigma))$ is determined by the mean and standard deviation. Hence Gaussian distributions form a 2-dimensional manifold \mathcal{M}_2 .

4.1.2 Metric

The metric of a manifold \mathcal{M}_N with N dimensions is a $N \times N$ symmetric matrix $G(\mathbf{x})$, with real elements $g_{ij}(\mathbf{x})$ defined on every point of the manifold, so the metric is a tensor field of some sort. The metric $G(\mathbf{x})$ serves as a standard of length or distance measure on a manifold, hence its name. We define the infinitesimal distance between two points \mathbf{x} and $\mathbf{x} + d\mathbf{x} \equiv (x^1 + dx^1, \dots, x^N + dx^N)$ by the metric $G(\mathbf{x})$ as

$$ds^2(\mathbf{x}) = g_{ij}(\mathbf{x})dx^i dx^j \quad (95)$$

(note the use of Einstein's convention of summation). In affine spaces the distance is defined as

$$ds^2 = \sum_i (dx^i)^2, \quad (96)$$

hence its metric is the identity matrix \mathbf{I} at all points of the space.

The manifold \mathcal{M}_N is thus a generalized space that has a “metric” G defined for the specification of distance between its points. From eq. (95) we can see that the coordinate axes may not be orthogonal, and that different coordinates may not have the same “weight” on the value of “distance”. But more importantly, because the metric defined on different points are not in general equal, the space of a manifold is “deformed” compared to affine spaces, so the affine spaces spanned by the coordinate axes of two different points on the manifold may have no intersection.

We will now use the entropy to define a metric for the manifolds formed by probability distributions. The Shannon entropy of a p.d.f. is defined as

$$H(\rho(\mathbf{x}|\mathbf{z})) = - \int \rho(\mathbf{x}|\mathbf{z}) \ln \rho(\mathbf{x}|\mathbf{z}) d\mathbf{z}, \quad (97)$$

with continuous p.d.f. this may diverge, that is, it may not have meaning under certain contexts, hence we use the cross entropy instead. The cross entropy (or Kullback-Leibler information) of two Gaussian distributions $\rho(\mathbf{x}_1|z)$ and $\rho(\mathbf{x}_2|z)$ is

$$\mathcal{H}_c(\mathbf{x}_1, \mathbf{x}_2) = \int \rho(\mathbf{x}_1|z) \ln \frac{\rho(\mathbf{x}_1|z)}{\rho(\mathbf{x}_2|z)} dz. \quad (98)$$

The J -divergence of the two distributions $\rho(\mathbf{x}_1|z)$ and $\rho(\mathbf{x}_2|z)$ is defined as

$$\mathcal{J}(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{H}_c(\mathbf{x}_1, \mathbf{x}_2) + \mathcal{H}_c(\mathbf{x}_2, \mathbf{x}_1). \quad (99)$$

We interpret the J -divergence as a sort of “distance” between these two distributions. If we set $\mathbf{x}_1 = \mathbf{x}$, $\mathbf{x}_2 = \mathbf{x} + d\mathbf{x} \equiv (x^1 + dx^1, x^2 + dx^2)$, then the infinitesimal distance can be defined as the J -divergence

$$\mathcal{J}(\mathbf{x}|\mathbf{x} + d\mathbf{x}) = ds^2 = g_{hk}(\mathbf{x})dx^h dx^k, \quad (100)$$

where, with the notation $\partial_h = \frac{\partial}{\partial x^h}$, we have the metric

$$g_{hk}(\mathbf{x}) = g_{kh}(\mathbf{x}) = \int \rho(\mathbf{x}|z) \partial_h \ln \rho(\mathbf{x}|z) \partial_k \ln \rho(\mathbf{x}|z) dz, \quad (101)$$

which forms a symmetric matrix G . Because the elements $g_{hk}(\mathbf{x})$ are in the form of the Fisher information (and yields the familiar form when $h = k$), the metric G is a Fisher information matrix, we refer to it as the Fisher metric in this context.

If we express the two parameters of the Gaussian distribution as

$$x^1 = \frac{\mu}{\sigma^2}, \quad x^2 = \frac{1}{2\sigma^2}, \quad (102)$$

then the Gaussian distribution becomes

$$\rho(\mathbf{x}|z) = \exp \left[zx^1 - z^2 x^2 - \frac{1}{4} \frac{(x^1)^2}{x^2} + \frac{1}{2} \ln x^2 - \frac{1}{2} \ln \pi \right], \quad (103)$$

and the Fisher metric

$$G \equiv \{g_{hk}\} = \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 4\mu^2\sigma^2 + 2\sigma^4 \end{pmatrix}. \quad (104)$$

4.1.3 Coordinate Transformations

A vector field $A(\mathbf{x})$ in the affine space \mathcal{E}_N can be specified by its N components $A^i(\mathbf{x})$, which is the components on some coordinate system $\mathbf{x} = \{x^1, \dots, x^N\}$. If there is another coordinate system $\mathbf{x}' = \{x'^1, \dots, x'^N\}$ for \mathcal{E}_N , then the components of $A(\mathbf{x})$ on the new coordinate system can be expressed as

$$A'^j(\mathbf{x}') = \frac{\partial x'^j}{\partial x^i} A^i(\mathbf{x}). \quad (105)$$

We will now extend these concepts to fields defined on manifolds.

In the discussion below, we will be using local coordinate systems that are limited to a certain neighborhood of the manifold. Since manifolds are “deformed”, there may not be a global coordinate system, but a small enough locality of the manifold may be considered “flat”. Hence we will define two coordinate systems in the neighborhood of $M \in \mathcal{M}_N$, and let the coordinates of M in the two coordinate systems be $\mathbf{x} = \{x^1, \dots, x^N\}$ and $\mathbf{x}' = \{x'^1, \dots, x'^N\}$ respectively.

A scalar field $F(\mathbf{x})$ on a manifold maps its points to a real number. Its variance under coordinate transformation is

$$F'(\mathbf{x}') = F(\mathbf{x}), \quad (106)$$

which is called invariance, since the value of a scalar field on the same point does not change under coordinate transformations.

A contravariant vector field $A(\mathbf{x})$ on a manifold is the extension of normal vector fields, the variance of its components $A^i(\mathbf{x})$ under coordinate transformation is

$$A'^j(\mathbf{x}') = \frac{\partial x'^j}{\partial x^i} A^i(\mathbf{x}), \quad (107)$$

which is the same as a normal vector field and is called contravariance.

A covariant vector field $A(\mathbf{x})$ on a manifold with components $A_i(\mathbf{x})$ is defined by

$$A_i(\mathbf{x}) = g_{ij}(\mathbf{x}) A^j(\mathbf{x}) \quad (108)$$

with respect to a contravariant field and the metric. We can also consider $A(\mathbf{x})$ as a vector field with both contravariant and covariant components. Under coordinate transformations, the covariant components change as

$$A'_j(\mathbf{x}') = \frac{\partial x^i}{\partial x'^j} A_i(\mathbf{x}), \quad (109)$$

which is called covariance.

We can further generalize these concepts to a tensor field $T(\mathbf{x})$ defined as the direct product of p contravariant fields and q covariant fields. Its $N^p \times N^q$ components $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}(\mathbf{x})$ change under coordinate transformation as

$$T_{j_1 \dots j_q}^{i_1 \dots i_p}(\mathbf{x}') = \frac{\partial x'^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial x'^{i_p}}{\partial x^{k_p}} \frac{\partial x^{l_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{l_q}}{\partial x'^{j_q}} T_{l_1 \dots l_q}^{k_1 \dots k_p}(\mathbf{x}), \quad (110)$$

and we say that $T(\mathbf{x})$ is p times contravariant and q times covariant. A scalar field $F(\mathbf{x})$ can then be considered as a tensor field 0 times contravariant and 0 times covariant.

If the metric of a manifold is 2 times covariant, that is,

$$g'_{rs}(\mathbf{x}') = \frac{\partial x^i}{\partial x'^r} \frac{\partial x^j}{\partial x'^s} g_{ij}(\mathbf{x}), \quad (111)$$

then we call this manifold a Riemannian manifold.

4.1.4 Connection

In affine space \mathcal{E}_N the derivative is $\nabla \equiv (\partial_1, \partial_2, \dots, \partial_N)$, with i -th component $\nabla_i = \partial_i = \frac{\partial}{\partial x^i}$. In manifolds, we define the i -th component of the covariant derivative ∇_i on scalar fields and the components of contravariant and covariant vector fields as

$$\nabla_i F(\mathbf{x}) = \partial_i F(\mathbf{x}), \quad (112)$$

$$\nabla_i A^k(\mathbf{x}) = \partial_i A^k(\mathbf{x}) + \Gamma_{ji}^k(\mathbf{x}) A^j(\mathbf{x}), \quad (113)$$

and

$$\nabla_i A_k(\mathbf{x}) = \partial_i A_k(\mathbf{x}) - \Gamma_{ki}^j(\mathbf{x}) A_j(\mathbf{x}) \quad (114)$$

respectively. The $\Gamma_{jk}^i(\mathbf{x})$'s are the components of the affine connection field, and are called affine connection coefficients, Christoffel symbols, or Γ -symbols. They “connect”, or provides the relationship between coordinate systems of different points on manifolds, hence is needed to describe derivatives on manifolds. The name covariant derivative comes from the fact that the fields after operation are increased in covariance by 1, so that $\nabla_i F(\mathbf{x})$ are components of a covariant field, $\nabla_i A^k(\mathbf{x})$ components of a tensor field 1 time contravariant and 1 time covariant, and $\nabla_i A_k(\mathbf{x})$ components of a 2 times covariant tensor field. Note that the connection $\Gamma(\mathbf{x})$ is not a tensor field in general, only when the coordinate transformation is linear do the components of $\Gamma(\mathbf{x})$ change like the components of a tensor field 1 time contravariant and 2 times covariant.

The covariant derivatives of general tensor fields can be found by extending eqs. (113) and (114). For example, the k -th covariant derivative of Riemannian metric component $g_{ij}(\mathbf{x})$ is

$$\nabla_k g_{ij}(\mathbf{x}) = \partial_k g_{ij}(\mathbf{x}) - \Gamma_{ik}^l(\mathbf{x}) g_{lj}(\mathbf{x}) - \Gamma_{jk}^l(\mathbf{x}) g_{il}(\mathbf{x}). \quad (115)$$

We define $\Gamma_{jik} = g_{jl} \Gamma_{ik}^l$ (this is needed since Γ is not a tensor in general), the equation above becomes

$$\nabla_k g_{ij}(\mathbf{x}) = \partial_k g_{ij}(\mathbf{x}) - \Gamma_{jik}(\mathbf{x}) - \Gamma_{ijk}(\mathbf{x}). \quad (116)$$

If $\Gamma_{ijk} = \Gamma_{ikj}$, then the connection is called a Riemannian connection. Using eq. (116), we can generate two more equations by permutation of the indices i, j, k , add two of these equations and subtract by the third, with the assumption of Riemannian connection, we can express the connection coefficients by the metric components:

$$\Gamma_{ijk}(\mathbf{x}) = \frac{1}{2} (\partial_k g_{ji}(\mathbf{x}) + \partial_j g_{ik}(\mathbf{x}) - \partial_i g_{kj}(\mathbf{x})). \quad (117)$$

We now define the connection for probability distribution manifolds. The most general connection compatible with information geometry is

$$\Gamma_{ij,k}^{(\alpha)}(\mathbf{x}) = \Gamma_{kij}(\mathbf{x}) - \frac{\alpha}{2} \int \partial_i \ln \rho(\mathbf{x}|\mathbf{z}) \partial_j \ln \rho(\mathbf{x}|\mathbf{z}) \partial_k \ln \rho(\mathbf{x}|\mathbf{z}) d\mathbf{z}. \quad (118)$$

With the metric eq. (101) we have

$$\Gamma_{kij} = \int \rho \left[\partial_{ij}^2 \ln \rho \partial_k \ln \rho + \frac{1}{2} \partial_i \ln \rho \partial_j \ln \rho \partial_k \ln \rho \right] d\mathbf{z} \quad (119)$$

4.1.5 Curvature Tensor

The curvature tensor (or Riemannian tensor) $R(\mathbf{x})$ of a manifold is a tensor field 1 time contravariant and 3 times covariant, the definition of its components is

$$R_{ikl}^j(\mathbf{x}) = \partial_k \Gamma_{il}^j(\mathbf{x}) - \partial_l \Gamma_{ik}^j(\mathbf{x}) + \Gamma_{hk}^j(\mathbf{x}) \Gamma_{il}^h(\mathbf{x}) - \Gamma_{hl}^j(\mathbf{x}) \Gamma_{ik}^h(\mathbf{x}). \quad (120)$$

On a Riemannian manifold, because $g_{ij} = g_{ji}$ and $\Gamma_{jk}^i = \Gamma_{kj}^i$, we can define

$$\begin{aligned} R_{iklm} = g_{in} R_{klm}^n &= \frac{1}{2} (\partial_k \partial_l g_{im} + \partial_i \partial_m g_{kl} - \partial_k \partial_m g_{il} - \partial_i \partial_l g_{km}) \\ &+ g_{np} (\Gamma_{kl}^n \Gamma_{im}^p - \Gamma_{km}^n \Gamma_{il}^p), \end{aligned} \quad (121)$$

which is 4 times covariant, the fully covariant form of the Riemannian tensor. By the Riemannian conditions we can further state the relations

$$R_{iklm} = -R_{kilm} \quad (122)$$

$$R_{iklm} = -R_{ikml} \quad (123)$$

$$R_{iklm} = R_{lmik} \quad (124)$$

$$R_{iilm} = R_{ikll} = 0 \quad (125)$$

We use $\Gamma_{ij,k}^{(\alpha)}$ as the connection for the Gaussian manifold, using the eqs. (104), (118) and (119) for the metric and connection, we can obtain the curvature tensor component for the Gaussian manifold

$$R_{1212}^{(\alpha)} = (1 - \alpha^2) \sigma^6. \quad (126)$$

When $\alpha = 0$, the connection becomes

$$\Gamma_{ijk}^{(0)} = \Gamma_{kij} \quad (127)$$

and the curvature tensor becomes

$$R_{1212}^{(0)} = \sigma^6. \quad (128)$$

The previous equation suggests that the curvature tensor expresses our lack of information since it vanishes only when $\sigma = 0$.

4.2 Quantum Geometry

4.2.1 The Metric and Connection

Because wave functions (complex probability amplitudes) are used instead of probability functions in quantum mechanics, we need to generalize the metric defined in information geometry to use wave functions.

We begin by changing the expression of the metric to

$$g_{hk}(\mathbf{x}) = 4 \int \partial_h \sqrt{\rho(\mathbf{x}|z)} \partial_k \sqrt{\rho(\mathbf{x}|z)} dz, \quad (129)$$

which is in terms of probability amplitude $\sqrt{\rho(\mathbf{x}|z)}$. Similarly, the infinitesimal distance (J -divergence) can be expressed as

$$ds^2 = 4 \int (\partial_h \sqrt{\rho} dx^h) (\partial_k \sqrt{\rho} dx^k) dz = 4 \int (d\sqrt{\rho})^2 dz. \quad (130)$$

If we change the variable z to a discrete variable, with $\phi_\alpha(\mathbf{x}) = \sqrt{\rho(\mathbf{x}|z = \alpha)}$, we can then use the inner product form to express the metric

$$g_{hk}(\mathbf{x}) = 4 \sum_\alpha \partial_h \phi_\alpha(\mathbf{x}) \partial_k \phi_\alpha(\mathbf{x}). \quad (131)$$

The desired generalization of the information metric is now obvious: it is (neglecting the irrelevant numerical factor)

$$g_{hk}(\mathbf{x}) = \overline{g_{kh}(\mathbf{x})} = \int \psi_h(\mathbf{x}|z) \overline{\psi_k(\mathbf{x}|z)} dz. \quad (132)$$

If $\psi_h(\mathbf{x}|z) = \partial_h \phi(\mathbf{x}|z)$ then eq. (132) yields the general holonomic case, if also $\overline{\partial_h \phi} = \partial_h \phi$, we return back to the standard information metric (101).

In quantum geometry we wish to use Riemannian manifolds, hence we must have $\alpha = 0$ in the connection $\Gamma_{ij,k}^{(\alpha)}$, since only then would the covariant derivative of the metric vanish. The Riemannian property of the metric can be considered to correspond to the Hermitian property of density operators in quantum physics.

So we set $\alpha = 0$ in eq. (118), and by eq. (119) and the following relation (which is easily proven)

$$\prod_{i=1}^k \partial_i \ln \rho = 2^k \rho^{-\frac{k}{2}} \prod_{i=1}^k \partial_i \sqrt{\rho}, \quad (133)$$

we have

$$\Gamma_{ij,k}^{(0)} = 4 \int \partial_{ij}^2 \sqrt{\rho} \partial_k \sqrt{\rho} dz, \quad (134)$$

which is also expressed with probability amplitudes.

4.2.2 The Uncertainty Relation

Since the metric is in the form of Fisher information, we can use the Cramer-Rao inequality to derive the uncertainty relation.

In the one-dimensional case (with manifold \mathcal{M}_1), let \hat{x} be an unbiased estimator for the single parameter x . With variance $(\Delta x)^2 \equiv \langle (\hat{x} - x)^2 \rangle$, the Cramer-Rao inequality is

$$(\Delta x)^2 \cdot g_{11} \geq 1, \quad (135)$$

where the only component of the metric g_{11} is the Fisher information. Using eq. (129) with $x \equiv x_1$ we have

$$(\Delta x)^2 \cdot 4 \int dz \left(\frac{\partial \sqrt{\rho}}{\partial x} \right)^2 \geq 1, \quad (136)$$

and with quantum physical identities $\sqrt{\rho} = \phi$, $p_x = -i\hbar \frac{\partial}{\partial x}$ we get the uncertainty relation

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}. \quad (137)$$

Thus the uncertainty relation can be derived from Cramer-Rao inequality, and hence is not exclusive to quantum physics.

4.2.3 The Sign of Infinitesimal Distance ds^2

From the preceding derivation, the ‘‘infinitesimal distance’’ ds^2 and ‘‘infinitesimal cross-entropy’’ $d\mathcal{H}_c$ can be considered equivalent (to within appropriate identification, see eqs. (99) and (100))

$$ds^2 \equiv d\mathcal{H}_c = g_{hk} dx^h \overline{dx^k} = \int \psi_h dx^h \overline{\psi_k dx^k} dz. \quad (138)$$

Under special relativity a particle in space-time is described by the four parameters (four-vector) $\mathbf{x} = \{x^1 = ct, x^2 = ix, x^3 = iy, x^4 = iz\}$, and the metric G is identity, we have

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 \geq 0, \quad (139)$$

which is the requirement that the speed of particles not exceed c . From this inequality we also have

$$d\mathcal{H}_c \geq 0. \quad (140)$$

Since the previous equation is derived from a basic assumption of relativity, it can be considered as a basic physical principle, similar to other basic principles such as the second law of thermodynamics or the I -theorem of EPI.

5 Quantum Information Theory

In quantum information theory there are three main goals of research: to find the fundamental unit of static resources, that is, quantum information classes; to find the basic dynamical processes, that is, quantum information processing; to quantify the resources needed for these processes, and different tradeoffs between various quantum information classes. The classical information theory can be seen as a special case of quantum information theory, which is wider in scope and much more difficult.

5.1 Quantum Information Resources

Classical information theory uses Shannon entropy to quantify the uncertainty of probability distributions or information sources, the information obtained after the value of a random variable is known, and the physical resources needed for the storage of a random variable. To extend classical information to the world of quantum physics, we need to define *quantum* information sources, or quantum random variables, which is used as the fundamental unit of quantum information resources. There are many possible definitions, each of which may give rise to a different information theory, following [9], we state two possible definitions.

One simple definition of a quantum information source is a probability distribution of quantum states. A qubit with probability $\frac{1}{2}$ to be in the state $|0\rangle$ or $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ could be treated as a quantum information source, or quantum random variable. We gain information if we could somehow measure this qubit and obtain its state, but in our example, since the two possible states are not orthogonal, they are not completely distinguishable. This differs from a classical information source, in which different values of the source can always be distinguished (in principle).

The second definition comes from quantum entanglement. That is, the information we are trying to process in quantum physics is stored in the entangled states between two systems, or that the details of entanglement is our quantum information.

5.2 The Von Neumann Entropy

Using the first definition given in the last section, a quantum information source X have n possible values $\rho_1, \rho_2, \dots, \rho_n$, each ρ_i is a density operator state occurring with probability p_i . So the density operator for this informa-

tion source is

$$\boldsymbol{\rho} = \sum_i p_i \boldsymbol{\rho}_i, \quad (141)$$

where $\sum_i p_i = 1$, so $\boldsymbol{\rho}$ must be a density operator. The Von Neumann entropy of this source X (or state $\boldsymbol{\rho}$) is defined as

$$S(\boldsymbol{\rho}) \equiv -\text{tr}(\boldsymbol{\rho} \log \boldsymbol{\rho}), \quad (142)$$

where the logarithm of a matrix is the inverse function of the exponential of a matrix $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$.

If the $\boldsymbol{\rho}_i$'s are mutually orthogonal pure states, then they can be perfectly distinguished, in this special case the Von Neumann entropy is equal to the Shannon entropy:

$$S(\boldsymbol{\rho}) = H(X) = -\sum_i p_i \log p_i. \quad (143)$$

If the eigenvalues of $\boldsymbol{\rho}$ is λ_i , then the Von Neumann entropy is

$$S(\boldsymbol{\rho}) = -\sum_i \lambda_i \log \lambda_i. \quad (144)$$

This is because the eigenvectors of density operators are always mutually orthogonal. When the state is the most "mixed" $\boldsymbol{\rho} = \frac{\mathbf{1}}{d}$, where d is the dimension of the Hilbert space of $\boldsymbol{\rho}$, the Von Neumann entropy is

$$S(\boldsymbol{\rho}) = \log d. \quad (145)$$

The same as the maximum Shannon entropy under uniform distribution.

We list some properties of the Von Neumann entropy:

1. The Von Neumann entropy is non-negative, $S(\boldsymbol{\rho}) \geq 0$, with equality only when $\boldsymbol{\rho}$ is a pure state, as can be seen from the definition.
2. In a d -dimensional Hilbert space, the maximum Von Neumann entropy is $\log d$, which occurs when $\boldsymbol{\rho} = \frac{\mathbf{1}}{d}$.
3. If the composite system AB is in a pure state, then $S(A) = S(B)$.
4. If the $\boldsymbol{\rho}_i$'s have mutually orthogonal supports, then

$$S\left(\sum_i p_i \boldsymbol{\rho}_i\right) = H(p_i) + \sum_i p_i S(\boldsymbol{\rho}_i). \quad (146)$$

5. Joint entropy. If the $|i\rangle$'s are orthogonal states in the Hilbert space of system A , and the ρ_i 's states of system B , then

$$S\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i). \quad (147)$$

This can be proved by the previous property since $|i\rangle\langle i| \otimes \rho_i$ and $|j\rangle\langle j| \otimes \rho_j$ are orthogonal for $i \neq j$. The Von Neumann entropy of joint systems with states ρ and σ is

$$S(\rho \otimes \sigma) = S(\rho) + S(\sigma). \quad (148)$$

Following Shannon entropy, we can likewise define the quantum joint entropy, quantum conditional entropy, and quantum mutual information. The quantum joint entropy of composite system AB is defined as

$$S(A, B) \equiv -\text{tr}\left(\rho^{AB} \log \rho^{AB}\right), \quad (149)$$

where ρ^{AB} is the density operator for the composite system. The conditional entropy of system A when system B is known is

$$S(A|B) \equiv S(A, B) - S(B), \quad (150)$$

and the mutual information of systems A and B is

$$S(A : B) \equiv S(A) + S(B) - S(A, B) = S(A) - S(A|B) = S(B) - S(B|A). \quad (151)$$

Under this definition, the Von Neumann conditional entropy may be negative, as when system AB is in the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, the entropies have the values

$$\begin{aligned} S(A, B) &= 0 \\ S(A) &= S\left(\frac{\mathbf{I}}{2}\right) = 1 \\ S(B|A) &= S(A, B) - S(A) = -1. \end{aligned} \quad (152)$$

Below we list more properties of the Von Neumann entropy:

6. Subadditivity: $S(A, B) \leq S(A) + S(B)$.
7. Triangle inequality: $S(A, B) \geq |S(A) - S(B)|$.
8. Concavity: $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$, with equality when all ρ_i 's are equal. This means that when we have no knowledge about how a state is prepared, as when we know only $\rho = \sum_i p_i \rho_i$, and have no knowledge about the individual states and probabilities, the entropy would be higher than if they were known.
9. Strong Subadditivity: $S(A, B, C) + S(B) \leq S(A, B) + S(B, C)$.

5.3 Distinguishing Quantum States

Since non-orthogonal quantum states cannot be perfectly distinguished by measurement, even if quantum data are sent without error, the information that we can extract from the message states may still be limited. The relative amount of information that we can extract from any quantum state encoding a classical message is called the accessible information. As an example, if we use two non-orthogonal states to encode the bit values 0 and 1, the accessible information of such a quantum message will surely be less than its classical equivalent.

Considering the possibility of noise in the classical communication channel, we can define accessible information as the closeness of the values of $H(X : Y)$ and $H(X)$, where X is the message to be send, and Y is the message sent. In the case of sending classical message with a quantum channel, since we encode with quantum states, the accessible information is not only affected by the channel noise or distortion, but also by the quantum states chosen to encode the classical values. We assume that the values of the classical message are represented by the states ρ_x , with probabilities p_x , so that the state of a letter in this quantum encoded message could be represented by the density operator

$$\rho = \sum_x p_x \rho_x. \quad (153)$$

The corresponding Shannon entropy is

$$H(X) = - \sum_x p_x \log p_x \geq S(\rho). \quad (154)$$

Without channel errors, the receiver performs a measurement on the state ρ , with probability $p(y|x)$ of obtaining result y when the received state is ρ_x . We could then use the conditional probabilities to calculate the mutual information $H(X : Y)$. The quantum accessible information is defined as the maximum of $H(X : Y)$ over all possible measurements,

$$\text{Accessible Information} = \max_{\text{all measurements}} H(X : Y). \quad (155)$$

The quantum accessible information has an upper bound called the Holevo bound. Assume that letter ρ_x occurs with probability p_x , and that the receiver performs a generalized POVM measurement \mathbf{E}_y on the received state $\rho = \sum_x p_x \rho_x$, then

$$H(X : Y) \leq S(\rho) - \sum_x p_x S(\rho_x). \quad (156)$$

For the ensemble $\{p_x, \rho_x\}$ we define the Holevo χ quantity as

$$\chi \equiv S\left(\sum_x p_x \rho_x\right) - \sum_x p_x S(\rho_x). \quad (157)$$

5.4 Data Compression

According to classical information theory, a classical information source is defined as a string of random variables X_1, X_2, \dots, X_n , with the n values as its output. We assume that they are all taken from the same set of letters with the same probability distribution. So for the ensemble $X = \{p_x, x\}$, a string of length n would have on average np_x occurrence of the letter x , hence we define such strings as typical. We can prove that the number of typical strings does not exceed about $2^{nH(X)}$, where

$$H(X) = -\sum_x p_x \log p_x \quad (158)$$

is the Shannon entropy of X . So such a string can be compressed to $nH(X)$ bits of information.

Applying the above result to quantum information, we can define a quantum source as $X = \{p_x, \rho_x\}$, that is, we use different quantum states to represent different classical letters. For a message of length n , its quantum state is the composite of the individual letter states $\rho = \sum_x p_x \rho_x$:

$$\rho^n = \underbrace{\rho \otimes \rho \otimes \dots \otimes \rho}_{n \text{ times}} \quad (159)$$

The use of density operators means that individual letters may be in a mixed state, so they may be entangled with other physical systems, thus under this definition of a quantum information source we may say that the information resource is quantum state and entanglement.

Every letter ρ is in a Hilbert space \mathcal{H} , so the entire message ρ^n is in the Hilbert space

$$\mathcal{H}^n = \underbrace{\mathcal{H} \otimes \mathcal{H} \otimes \dots \otimes \mathcal{H}}_{n \text{ times}}. \quad (160)$$

If $\mathcal{H} = \mathcal{H}^{\otimes m}$ is m -dimensional, then each letter needs m qubits to store, and the entire message needs $n \times m$ qubits. Let λ_i be the eigenvalues of ρ , then

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i. \quad (161)$$

Here λ_i is the probability of measuring eigenstate $|i\rangle$ when measuring ρ in the basis formed by its eigenstates. Hence the eigenstate $|i_1\rangle \otimes |i_2\rangle \otimes \dots \otimes |i_n\rangle$ of ρ^n

has probability $\lambda_{i_1} \times \lambda_{i_2} \times \cdots \times \lambda_{i_n}$ of being measured. Following the classical case we can prove that when n is large, there will on average be $n\lambda_i$ occurrences of eigenstate $|i\rangle$ in the joint eigenstate of the whole message, and that such typical eigenstates does not exceed $nS(\boldsymbol{\rho})$ in number. Hence assuming that only typical eigenstates appear, the entire message can be represented in a $nS(\boldsymbol{\rho})$ -dimensional Hilbert space, that is, it can be compressed to $nS(\boldsymbol{\rho})$ qubits.

5.5 Data Transmission

Theories of data transmission is an important part of information theory, it deals mainly with the capacity of noisy data channels. If we transmit a classical message $X = \{p_x, x\}$ through a classical channel, and the channel changes (distorts) letter x to y with probability $p(y|x)$, that is, $p(y|x)$ reflects the channel characteristics, then according to Shannon's noisy coding theorem the maximum transmission capacity through the channel is

$$C = H(X : Y) = H(Y) - H(Y|X), \quad (162)$$

where $Y = \{p_y, y\}$, $p_y = \sum_x p(y|x)p_x$ and $H(Y|X)$ is the conditional entropy. We define the capacity of a channel to be the maximum $H(X : Y)$ over all possible distributions p_x

$$C = \max_{p_x} H(X : Y). \quad (163)$$

We can use the classical result to find the capacity for a quantum channel transmitting classical information.

With quantum states $\boldsymbol{\rho}_i$ representing individual letters, we limit the transmitted quantum states to product states $\boldsymbol{\rho}_{i_1} \otimes \boldsymbol{\rho}_{i_2} \otimes \cdots$, that is, no entanglement is allowed between letters. The capacity under this constraint is called the product state capacity. Note that there are no restrictions on how the receiver measures the received letters, hence multiple-letter measurements can be taken on the message. Let \mathcal{E} be a trace-preserving quantum operation that describes the characteristics of the quantum channel, then the Holevo-Schumacher-Westmoreland (HSW) theorem tells us that the product state capacity $\chi(\mathcal{E})$ for this channel is

$$\chi(\mathcal{E}) = \max_{\{p_i, \boldsymbol{\rho}_i\}} \left[S \left(\mathcal{E} \left(\sum_j p_j \boldsymbol{\rho}_j \right) \right) - \sum_j p_j S(\mathcal{E}(\boldsymbol{\rho}_j)) \right]. \quad (164)$$

There are currently no theories on general quantum channel capacity, but many researchers believe it is equal to the product state capacity, which means that there are no additional advantages in sending entangled letters.

References

- [1] B. Roy Frieden. *Physics from Fisher Information: A Unification*. Cambridge University Press, 1998.
- [2] E. R. Caianiello. “Quantum and other Physics as Systems Theory”. *Rivista Del Nuovo Cimento*, 15(4):1–65, 1992.
- [3] E. R. Caianiello. “Systems and Uncertainty”. In *Lecture notes in artificial intelligence*, volume 565, pages 56–66, 1991.
- [4] E. R. Caianiello and W. Guz. “Quantum fisher metric and uncertainty relations”. *Physics letters A*, 126(4):223–225, Jan 1988.
- [5] E. R. Caianiello and Canio Noce. “A Geometrical Approach to Thermodynamics. Uncertainty Relations”. *Gazzetta Chimica Italiana*, 118:67–71, 1988.
- [6] E. R. Caianiello. “Systems and Uncertainty: A geometrical approach”. In E. R. Caianiello and M. A. Aizerman, editors, *Topics in the General Theory of Structures*, pages 199–206, 1987.
- [7] E. R. Caianiello. “Geometrical ‘Identification’ of Quantum and Information Theories”. *Lettere al nuovo cimento*, 38(16):539–543, Dec 1983.
- [8] S. Amari and Hiroshi Nagaoka. *Methods of Information Geometry*. American Mathematical Society, 2000.
- [9] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.