PHYSICAL REVIEW A
GENERAL PHYSICS

THIRD SERIES, VOLUME 36, NUMBER 5 SEPTEMBER 1, 1987

Probability space of wave functions

Silviu Guiasu
Department of Mathematics, York University, North York, Toronto, Ontario, Canada M3J 1P3
(Received 2 December 1986)

The entropy maximization and the divergence minimization, two related optimization techniques coming from standard statistical mechanics and often applied nowadays to probabilistic modeling, are used here together in order to organize the space of wave functions as a probability space when the only available information on the quantum system at the initial moment is the density matrix, which describes the mean behavior of a quantum ensemble. The Schrödinger equation determines a nonconservative flow in this probability space transforming the independent Gaussian product measure, at the initial moment, into a dependent Gaussian product measure at an ulterior moment. Such a transformation can be viewed as a generalization of the Liouville theorem from classical statistical mechanics.

I. INTRODUCTION

There is no perfect analogy between quantum statistical mechanics and classical statistical mechanics. In classical statistical mechanics it is convenient to represent the state of a system of s degrees of freedom by the position of a point in a $2s$-dimensional Euclidean space of the coordinates and momenta, and then to represent an ensemble of such systems by a set of phase points distributed with the probability density ρ. This probability density is preserved or transformed by the flow induced by the equations of motion, and the mean value, at time t, for the systems in the ensemble, of a function F of the coordinates and momenta is the integral of F with respect to the probability measure induced by the density $\rho(t)$ at time t. It has been shown by von Neumann[1] that the so-called density matrix, with components ρ_{kl}, can be introduced in quantum mechanics to play a role *somewhat* similar to that of the density ρ in classical statistical mechanics.

In quantum mechanics the state of the system is a wave function ψ, represented as a weighted sum or integral of the eigenfunctions $\{\phi_k; k \in D\}$ of a given observable U whose corresponding eigenvalues are $\{u_k; k \in D\}$. It is agreed that the Schrödinger equation determines the mean behavior of the wave function. The state of a quantum system is undergoing random fluctuations about an average that satisfies the Schrödinger equation. It is agreed that this regular *mean* behavior can be predicted to a high degree of approximation, but there is no way of predicting how the individual measurements will fluctuate from one case to the next. According to Born[2] "we describe the instantaneous state of the system by a complex quantity ψ which satisfies a differential equation and therefore changes with time in a way which is completely determined by its form at time t_0, so that its behavior is rigorously causal. Since, however, physical significance is confined to the quantity $|\psi|^2$ and to other similar quadratic expressions (matrix elements), which only partially define ψ, it follows that, even when the physically determinable quantities are completely known at time t_0, the initial state ψ is necessarily not completely definable and therefore the affair of statistics."

The objective of this paper is to follow the same strategy used in classical statistical mechanics and to construct a probability measure on the space of wave functions (the state space) using as the only available data the density matrix, which describes the mean behavior of the quantum system. The tool for performing such a construction is provided by the entropy maximization and the divergence minimization, two optimization techniques coming from standard statistical mechanics, extensively used now in probabilistic modeling. Section II of the paper presents some details about the techniques needed here. The entropy maximization technique applied in the complex domain has some surprising consequences. One of them is that the nondiagonal elements of the density matrix could be interpreted as covariances between the random coefficients of the distinct eigenfunctions in the wave function. The minimization of the divergence from independence, allowing us to construct the most unbiased product probability measure subject to the given density matrix, induces a Gaussian probability measure on the space of wave functions. This is shown in Secs. III and V. Even when the density matrix is diagonal at the initial time t_0, it will have nondiagonal elements different from zero at a later time t, due to the Schrödinger equation. The corre-

sponding Gaussian probability measure essentially takes this probabilistic interdependence into account. The transformation of the independent Gaussian product measure in the state space, at time t_0, into an essentially dependent Gaussian product measure, at time $t > t_0$, may be viewed as a new quantum analog of the Liouville theorem from classical statistical mechanics. In Sec. IV some special density matrices are taken into account and in Sec. V the amount of global dependence between different eigenvalues of the observable U is analyzed. In several papers, Wiener and Siegel[3,4] introduced statistical ensembles in a "differential space," which is a Hilbert space containing a measure for which each differential space coordinate is *postulated* to have an independent normal distribution with the mean zero and the variance 1. The connection with their approach is discussed in Sec. VI. Section VII contains conclusions.

II. ENTROPY MAXIMIZATION AND DIVERGENCE MINIMIZATION

The principle of entropy maximization (PEM) is an optimization technique according to which, when both a domain and some mean values of some random variables, having the same (unknown) probability distribution on the respective domain, are given, we choose and construct *the* probability distribution of maximum entropy, on the given domain, subject to the given constraints. Here *entropy* means Shannon's entropy,[5] a global measure of the amount of uncertainty contained by a probability distribution. Generally, the feasible space, i.e., the set of all probability distributions defined on the given domain and compatible with the given constraints, is very large. Each such probability distribution contains some amount of uncertainty. By maximizing the entropy, we select the most random, or the most unbiased, probability distribution defined on the given domain and compatible with given constraints. It treats uniformly all the alternatives allowed by the given constraints. Let us emphasize the fact that PEM is not a radical new innovation from information theory; it was introduced by Boltzmann[6] and was used by Gibbs[7] to construct his canonical ensembles, by Planck[8] to derive the distribution which bears his name, and by von Neumann[1] and Born[2] to get the canonical distribution and the geometric distribution, respectively, in quantum statistical mechanics. Shannon[5] showed that Boltzmann's H function is a good measure of the amount of uncertainty contained by an abstract finite probability space, and Jaynes[9] formulated PEM as a general principle for constructing the most uncertain, or the most unbiased, probability distribution as an extension of the principle of insufficient reason of Bernoulli[10] and Laplace[11] (take the events as being equally likely if there is no reason for discriminating them), when there are constraints expressed by mean values. There are many applications of PEM in statistical inference, pattern recognition, classification, and time-series analysis.[12–16] We mention, however, Williams's remark:[17] "It has perhaps not been sufficiently appreciated that this principle (PEM), when properly understood, affords a rule of inductive inference of the widest generality."

Let ϕ be an n-dimensional probability density on $D \subset \mathbb{R}^n$. The corresponding entropy is

$$H(\phi) = -\int_D \phi(x) \ln \phi(x) dx \ .$$

Proposition 1. The one-dimensional probability density ϕ on $\mathbb{R}^1$, maximizing H subject to the mean μ and the variance σ^2, is the normal distribution $N(\mu, \sigma^2)$, whose density is

$$\phi(x) = \sigma^{-1}(2\pi)^{-1/2} \exp[-(x-\mu)^2/(2\sigma^2)] \ ,$$

and the corresponding entropy is

$$H(\phi) = \ln[\sigma(2\pi)^{1/2}] + 0.5 \ .$$

A simple proof of this known result may be found in Guiasu.[18]

The principle of divergence minimization (PDM) is an optimization technique with basically the same philosophy behind it as PEM. According to PDM, when both a reference probability measure on a domain and some mean values of some random variables, having the same (but unknown) probability distribution on the given domain, are given, we choose and construct the probability distribution of minimum divergence from the reference probability distribution, on the given domain, subject to the given constraints. Here *divergence* means the Kullback-Leibler[19] divergence, a relative entropy, measuring how far a probability distribution is from another probability distribution. According to PDM, we select the closest probability distribution to a reference probability distribution compatible with the given constraints.

A hybrid form of PDM was also used in statistical mechanics.[20] After 1959, PDM has been formulated as a general principle in statistical inference.[21,22] What we need here is a special form of PDM, called the principle of interdependence minimization (PIM).[23] Generally, when some one-dimensional probability distributions are given, then the direct (or geometric) joint probability distribution is simply their algebraic product. It corresponds to the case when the given probability distributions are independent. When we are dealing with random entities (random variables or probability distributions) they generally are dependent and the interdependence between them is measured in terms of mixed moments. The mixed moment of first order is the covariance, frequently used for measuring the amount of linear dependence between random variables. Again, when one or several mixed moments are given there are many, an infinity in fact, joint probability distributions compatible with them. By applying PIM we construct the closest joint probability distribution to the independent product of the given one-dimensional probability distributions subject to the known mixed moments. Such a joint probability distribution is the most random one, the closest one to independence between components subject to the given constraints. PEM and PIM are in fact used together: PEM for constructing the most unbiased one-dimensional probability distributions when some simple moments are known and PIM for constructing the most random joint

probability distribution subject to the given mixed moments.

Let ϕ_j $(j=1,\ldots,n)$ be one-dimensional probability densities on $\mathbb{R}^1$ and ϕ be an n-dimensional probability density such that $\phi_1(x_1)\cdots\phi_n(x_n)=0$ on a set of positive Lebesque measure from $\mathbb{R}^n$ implies $\phi(x_1,\ldots,x_n)=0$ on this set. The Kullback-Leibler divergence of ϕ from the independent product $\phi_1\cdots\phi_n$ is

$$I(\phi:\phi_1\cdots\phi_n)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\phi(x_1,\ldots,x_n)\ln\frac{\phi(x_1,\ldots,x_n)}{\phi_1(x_1)\cdots\phi_n(x_n)}dx_1\cdots dx_n\ .$$

When $\phi_1,\ldots,\phi_n$ are just the marginal probability densities of ϕ, then $I(\phi:\phi_1\cdots\phi_n)$ becomes Watanabe's[24] global measure of interdependence between $\phi_1,\ldots,\phi_n$,

$$I(\phi:\phi_1\cdots\phi_n)=W(\phi;\phi_1,\ldots,\phi_n)=\sum_{j=1}^{n}H(\phi_j)-H(\phi)\ .$$

Proposition 2. Let ϕ_j be a probability density $N(\mu_j,\sigma_j^2)$, for $j=1,\ldots,n$. The n-dimensional probability density ϕ which minimizes $I(\phi;\phi_1\cdots\phi_n)$ subject to the constraints

$$\int_{\mathbb{R}^n}x_j\phi(x)dx=\mu_j,\quad \int_{\mathbb{R}^n}(x_j-\mu_j)^2\phi(x)dx=\sigma_j^2\ ,\qquad (j=1,\ldots,n)$$

$$\int_{\mathbb{R}^n}(x_j-\mu_j)(x_k-\mu_k)\phi(x)dx=\sigma_{jk}\ ,\qquad (j=1,\ldots,n-1;k=2,\ldots,n;j<k)\ ,$$

is the n-dimensional normal distribution

$$\phi(x)=(2\pi)^{-n/2}\,|\,C_n\,|^{-1/2}\exp[(x-\mu)'C^{-1}(x-\mu)/2]\ ,$$

where $x=(x_1,\ldots,x_n)$, C_n is the covariance symmetric matrix $C_n=[\sigma_{jk}]$, with $\sigma_{jj}=\sigma_j^2,\sigma_{jk}=\sigma_{kj}$, $|\,C_n\,|$ is the determinant of the matrix C_n, C_n^{-1} is the inverse of the matrix C_n, and $(x-\mu)'$ is the row matrix $(x_1-\mu_1,\ldots,x_n-\mu_n)$, the transpose of the column matrix $(x-\mu)$ with $\mu=(\mu_1,\ldots,\mu_n)$. The one-dimensional probability densities $\phi_1,\ldots,\phi_n$ are the marginal probability densities of ϕ and the corresponding global interdependence between $\phi_1,\ldots,\phi_n$ is equal to

$$W=W(\phi;\phi_1,\ldots,\phi_n)=0.5\ln(\sigma_1^2\cdots\sigma_n^2/|\,C_n\,|)\ .$$

Details about PIM and the proof of the above proposition may be found in Guiasu, Leblanc, and Reischer.[23]

III. THE PROBABILITY MEASURE INDUCED BY THE DENSITY MATRIX

Let $\{\phi_k;k\in D\}$ be a complete orthonormal set of eigenfunctions corresponding to the eigenvalues $\{u_k;k\in D\}$ of the observable U. Let us suppose that D is a countable set, but such a supposition is not essential for what follows, simplifying only the writing. The state of the quantum system is a wave function

$$\psi=\sum_k X_k\phi_k=\sum_k(X_{k,1}+iX_{k,2})\phi_k=\sum_k|\,X_k\,|\exp(i\theta_k)\phi_k\ ,$$

which is identified with the family of complex coefficients $\{X_k;k\in D\}$ or the family of pairs of real coefficients $\{(X_{k,1},X_{k,2});k\in D\}$, or the family of modules and arguments $\{(|\,X_k\,|,\theta_k);k\in D\}$. Generally, we do not know the state ψ of the system. We can only obtain expected or mean values at the macroscopic scale and attribute average properties to the system. Throughout this paper we suppose that X_k is a complex random variable and that both $(X_{k,1},X_{k,2})$ and $(|\,X_k\,|,\theta_k)$ are pairs of real random variables, for any k.

Let $C=[\rho_{kl}]$ be the density matrix of the system. As it is well known,[20] in an ensemble of N similar noninteracting quantum-mechanical systems having the wave functions $\psi_s=\sum_k a_{s,k}\phi_k$ $(s=1,\ldots,N)$, the components of the density matrix C are estimated by

$$\rho_{kl}=\left[\sum_{s=1}^{N}a_{s,k}^{*}a_{s,l}\right]\Big/N\ ,$$

where a^* denotes the conjugate of the complex number a. Thus, ρ_{kl} may be interpreted as an estimate of the mixed moment $E(X_k^*X_l)$, where E denotes the expected (or mean) value of a random variable. Particularly, ρ_{kk} is an estimate of the mean value $E(X_k^*X_k)=E(|\,X_k\,|^2)$. But we need the probability distributions of X_k and of $X_{k,j}$. Applying PEM and PIM, we get the following.[25]

Proposition 3. The maximum entropy probability distribution on the possible values of $|\,X_k\,|^2$ compatible with ρ_{kk} is the exponential distribution with the density $f_k(y)=\rho_{kk}^{-1}\exp(-y/\rho_{kk})$ $(y>0)$. Consequently, the values taken on by $|\,X_k\,|$ are distributed according to the Weibull distribution, having the probability density $g_k(z)=2\rho_{kk}^{-1}z\exp(-z^2/\rho_{kk})$ $(z>0)$. The maximum entropy probability distribution on the possible values of the argument θ_k is the uniform distribution with the density $h_k(u)=(2\pi)^{-1}$ $(0\le u\le 2\pi)$. The joint probability distribution of minimum interdependence on the values of $(|\,X_k\,|,\theta_k)$ is

$$\xi_k(z,u)=g_k(z)h_k(u)=(\pi\rho_{kk})^{-1}z\exp(-z^2/\rho_{kk})\quad(z>0,0\le u\le 2\pi).$$

Consequently, the pair $(X_{k,1},X_{k,2})$ is distributed according to the two-dimensional probability distribution

$$\zeta_k(x_{k,1},x_{k,2})=(\pi\rho_{kk})^{-1}\exp[-(x_{k,1}^2+x_{k,2}^2)/\rho_{kk}]\ ,$$

which shows that $X_{k,1}$ and $X_{k,2}$ are two independent normal random variables $N(0,\rho_{kk}/2)$, as a kind of random fluctuation.

Based on these results we see that the mean value of X_k is $E(X_k)=E(X_{k,1}+iX_{k,2})=E(X_{k,1})+iE(X_{k,2})=0$, and the covariance between X_k and X_l is $C(X_k,X_l)=E([X_k-E(X_k)]^*[X_l-E(X_l)])=E(X_k^*X_l)=\rho_{kl}$. Particularly, the variance of X_k is $V(X_k)=C(X_k,X_k)=E(X_k^*X_k)=E(|X_k|^2)=\rho_{kk}$. Thus we have just proved the following.

Proposition 4. If the diagonal entries ρ_{kk} of the density matrix $C=[\rho_{kl}]$ are taken as estimates of the mean values of the random variables $|X_k|^2$ and the probability distributions of $|X_k|$, θ_k, $X_{k,1}$, $X_{k,2}$, and X_k are determined according to PEM and PIM, then ρ_{kl} is an estimate of the covariance between the complex random coefficients X_k and X_l of the wave function.

From proposition 3 we know that the components $X_{k,1}$ and $X_{k,2}$ of X_k are independent and identically distributed. This allows us to take $C(X_{k,j},X_{l,s})=\rho_{kl}\delta_{j,s}/2$ $(j,s=1,2)$, where $\delta_{j,s}$ is the Kronecker symbol, equal to 1 when $j=s$ and to 0 when $j\neq s$. Now we can apply PIM for constructing a probability measure on the space of the wave functions $\psi=\{(X_{k,1},X_{k,2});k\in D\}$. Let us introduce the set $A=\{(k,1),(k,2);k\in D\}$ and the collection $\mathbb{R}^A$ of all real-valued functions $\omega(\alpha)$ $(\alpha\in A)$, defined on A. Let $p_{\alpha_1\cdots\alpha_n}$ be the projection of $\mathbb{R}^A$ onto $\mathbb{R}^n$ defined by $p_{\alpha_1\cdots\alpha_n}(\omega)=[\omega(\alpha_1),\ldots,\omega(\alpha_n)]\in\mathbb{R}^n$ and let us introduce the σ field of all Borel cylinders in $\mathbb{R}^A$ with index $(\alpha_1,\ldots,\alpha_n)$, i.e., the collection of sets $I_{\alpha_1\cdots\alpha_n}=\{p^{-1}_{\alpha_1\cdots\alpha_n}(B),\ B\in\mathcal{B}^n\}$, where $\mathcal{B}^n$ is the σ field of Borel sets in $\mathbb{R}^n$. On the field $I=\cup I_{\alpha_1\cdots\alpha_n}$, where the union is taken over all finite sequences $\alpha_1,\ldots,\alpha_n$ of elements of A, we define the set function $\eta(E)=\int_B\eta_{\alpha_1\cdots\alpha_n}(x)dx$ for any set $E\in I_{\alpha_1\cdots\alpha_n}$ having the form $E=p^{-1}_{\alpha_1\cdots\alpha_n}(B)$, $B\in\mathcal{B}^n$, where, according to propositions 2 and 3, the probability density $\eta_{\alpha_1\cdots\alpha_n}$ has the expression

$$\eta_{\alpha_1\cdots\alpha_n}(x)=(2\pi)^{-n/2}|C_n|^{-1/2}\exp(-x'C_n^{-1}x/2)\ , \tag{1}$$

where, for $\alpha_1=(k_1,j_1),\ldots,\alpha_n=(k_n,j_n)$, C_n is the covariance matrix

$$C_n=[C(X_{\alpha_s},X_{\alpha_v})]=[\rho_{k_sk_v}\delta_{j_s,j_v}/2]\ . \tag{2}$$

The set function η is well defined on I and, according to the Kolmogorov extension theorem,[26] η can be extended uniquely to be a probability measure on the σ field $\sigma(I)$ generated by I. On the probability space $(\mathbb{R}^A,\sigma(I),\eta)$ we define the system of random variables $\psi=\{X_\alpha;\alpha\in A\}$ by $X_\alpha(\omega)=\omega(\alpha)$, for $\omega\in\mathbb{R}^A$. We see that ψ is a Gaussian system of random variables because

$$\begin{aligned}&\eta(\{\omega;\omega\in\mathbb{R}^A,[X_{\alpha_1}(\omega),\ldots,X_{\alpha_n}(\omega)]\in B\})\\&=\eta(\{\omega;\omega\in\mathbb{R}^A,[\omega(\alpha_1),\ldots,\omega(\alpha_n)]\in B\})\\&=\eta(p^{-1}_{\alpha_1\cdots\alpha_n}(B))=\int_B\eta_{\alpha_1\ldots\alpha_n}(x)dx\ ,\end{aligned}$$

where $\eta_{\alpha_1\cdots\alpha_n}$ is the Gaussian probability density given by (1). Thus we have proved the following.

Proposition 5. If the only information about a quantum system is given by the density matrix $C=[\rho_{kl}]$ then, according to PEM and PIM, the state space becomes a Gaussian system of random variables with the mean function equal to zero and with $\{\rho_{kl};k,l\in D\}$ as the covariance function.

In the above construction, D may be an uncountable set as well, which means that from this point of view, basically, it does not matter whether the observable U has a discrete or continuous system of eigenvalues.

As it is well known,[20] von Neumann has introduced the density matrix in order to play a role *somewhat* similar to that of the probability density ρ from the classical statistical mechanics. We can see, however, that in fact the Gaussian measure η induced by the density matrix on the space of wave functions may be considered as being the analog of the classical probability density ρ. We should take notice here that while the diagonal elements ρ_{kk} of the density matrix are generally considered as being relevant, representing the probability of the value u_k of the observable U, the nondiagonal elements ρ_{kl} prove to be equally significant, giving the covariance between the components X_k and X_l of the wave function.

IV. THE PROBABILITY MEASURE AT THE INITIAL TIME

At the initial time $t=0$, the density matrix is taken to be a diagonal matrix, i.e., $\rho_{kl}(0)=\rho_{kk}(0)\delta_{k,l}$. As a consequence, the probability density (1) becomes

$$\begin{aligned}\eta_{\alpha_1\cdots\alpha_n}(x_1,\ldots,x_n)&=(2\pi)^{-n/2}[\rho_{k_1k_1}(0)\cdots\rho_{k_nk_n}(0)]^{-1/2}\exp\left[-\sum_{j=1}^{n}\{x_j^2/[2\rho_{k_jk_j}(0)]\}\right]\\&=\eta_{\alpha_1}(x_1)\cdots\eta_{\alpha_n}(x_n)\end{aligned} \tag{3}$$

and, consequently, $\psi=\{(X_{k,1},X_{k,2});k\in D\}$ is an independent Gaussian system of random variables. There is no correlation between the components of ψ. In such a case the joint probability measure η is completely determined by the probabilities (or the probability density if D is uncountable) $\rho_{kk}(0)$ on the set of eigenvalues of the observable U. Practically, $\{\rho_{kk}(0);k\in D\}$ may be determined in one of the following ways.

(a) We examine simultaneously several identical systems and estimate either the probability distribution $\{p_k;k\in D\}$ of the possible values of U, if D is finite or countable, and put $\rho_{kk}(0)=p_k$, or the probability density

$\{f(k);k\in D\}$ on the set D of possible values of U, if D is uncountable, and put $\rho_{kk}(0)=f(k)$. Often such an approach cannot be practically implemented.

(b) We know only the mean value $\langle U\rangle$ of the observable U. Then, applying PEM, we determine the most random, or the most unbiased, probability distribution on the set of possible values of U, which, as it is well known,[18] is the canonical distribution $p_k=[\Phi(\beta)]^{-1}\exp(-\beta u_k)$, $(k\in D)$, in the discrete case, where $\Phi(\beta)=\sum_k \exp(-\beta u_k)$, and β is the unique solution of the differential equation $d\ln\Phi(\beta)/d\beta=-\langle U\rangle$. We take $\rho_{kk}(0)=p_k$, $k\in D$. In the continuous case, when the set of possible values of U is $D=(0,+\infty)$ and we know only $\langle U\rangle$, PEM gives $f(k)=(\langle U\rangle)^{-1}\exp(-k/\langle U\rangle)$ $(0<k<+\infty)$. Also, if the set of all possible values of U is $D=(-\infty,+\infty)$, and we know both the mean $\langle U\rangle$ and the variance σ_U^2, then, according to PEM,

$$f(k)=\sigma_U^{-1}(2\pi)^{-1/2}\exp[-(k-\langle U\rangle)^2/(2\sigma_U^2)]\ ,\qquad (k\in\mathbb{R}^1)\ .$$

In such cases we take $\rho_{kk}(0)=f(k)$ $(k\in D)$. Let us notice that here PEM is applied at the level of the possible values of the observable U, and the probability distribution obtained is the available information when we apply again PEM, together with PIM, for obtaining the most random, or the most unbiased, probability distribution on the space of possible wave functions. There are two distinct degrees of randomization in such an approach.

(c) Two other cases are also special consequences of PEM.

(i) D is a finite set, containing n elements, and $\rho_{kk}(0)=1/n$ for $k\in D$. The uniform distribution maximizes the entropy when our partial information as to the value of the observable U is equally well satisfied by any one of a group of n, generally neighboring, possible values.

(ii) By analogy with the normal distribution, which is a solution of PEM, we can take, even when D is finite or countable,

$$\rho_{kk}(0)=\rho_0\exp[-(u_k-u_0)^2/(2\sigma^2)]\quad (k\in D)$$

where ρ_0 and σ are constants. It corresponds to the case when our approximate measurement of U is telling us that the value of U was almost certainly in the immediate neighborhood of a particular eigenvalue u_0, with a decreasing chance for values more and more removed there from. The last two probability distributions just given are mentioned by Tolman.[20]

Each of the expressions given in this section for $\rho_{kk}(0)$ may be introduced in (3) and the joint probability density thus obtained may be used for computing the probability of some events of interest at time $t=0$, like the probability

$$P(a_s\le X_{k_s,j_s}(0)\le b_s\ \ s=1,\ldots,n)=\prod_{s=1}^{n}[2\pi\rho_{k_sk_s}(0)]^{-1/2}\int_{a_s}^{b_s}\exp\{-x_s^2/[2\rho_{k_sk_s}(0)]\}dx_s\ ,$$

for locating some components of the wave function in some arbitrary intervals.

V. THE PROBABILITY MEASURE AS A FUNCTION OF TIME

Let us take into account the time interval $[0,t]$. Let $\mathbf{H}$ be the Hamiltonian and $S_{kl}(t)$ be the matrix elements corresponding to the transformation operator $\mathbf{S}(t)=\exp(-2\pi i\mathbf{H}t/h)$, where h is the Planck constant. If $\rho_{kl}(0)=\rho_{kk}(0)\delta_{k,l}$, then the Schrödinger equation, describing the time evolution of the wave functions, tells us that $\rho_{kl}(t)=\sum_n S^*_{ln}(t)S_{kn}(t)\rho_{nn}(0)$. Thus, even whether $\rho_{kl}(0)=0$, for $k\neq l$, in general $\rho_{kl}(t)$ would not reduce to zero for $k\neq l$. Hence, with the interpretation of the elements of a density matrix discussed in Sec. III, we must conclude that the initial independence between the components of the random wave function would in general be lost as time proceeds. In the time interval $[0,t]$, due to the evolution of the system, the initial independent Gaussian probability measure generated by the family of densities (3) is replaced by the dependent Gaussian probability measure generated by the family of densities (1), where the covariance matrix (2) is

$$\begin{aligned}C_n(t)&=[C(X_{\alpha_s}(t),X_{\alpha_v}(t))]\\&=[\rho_{k_sk_v}(t)\delta_{j_s,j_v}/2]\\&=\left[\sum_n S^*_{k_vn}(t)S_{k_sn}(t)\rho_{nn}(0)\delta_{j_s,j_v}/2\right]\ .\end{aligned}\tag{4}$$

This transformation, from (3) to (1) with the covariance matrix (4), may be considered as being the generalization to our context of the Liouville theorem from the classical statistical mechanics. But here, generally, the joint probability in the state space is not conserved, as in the classical case, but changes in time. The flow generated by the Schrödinger equation in the space of wave functions does not generally preserve the probability distribution η, which evolves in time: Its mean function continues to be identically equal to zero, but the covariance function changes in time according to (4). As a consequence, the set of independent random components $\{X_{k_1,j_1}(0),\ldots,X_{k_n,j_n}(0)\}$ of the random wave function $\psi(0)$, at the initial time $t=0$, becomes a system of dependent random components $\{X_{k_1,j_1}(t),\ldots,X_{k_n,j_n}(t)\}$ of the random wave function $\psi(t)$, at time t. According to (3), (1), and (4), we see that while, at the initial time, the components of $\psi(0)$ are not correlated, which means $W(0)=0$, later, at time t, the amount of global inter-

dependence between the normal random components $\{X_{k_1,j_1}(t),\ldots,X_{k_n,j_n}(t)\}$ of $\psi(t)$ is equal to

$$W(t)=0.5\ln\{\rho_{k_1k_1}(t)\cdots\rho_{k_nk_n}(t)/|C_n(t)|\}$$
$$=0.5\ln\{[\rho_{k_1k_1}(t)\cdots\rho_{k_nk_n}(t)]/[\lambda_{k_1}(t)\cdots\lambda_{k_n}(t)]\}, \quad (5)$$

where $|C_n(t)|$ is the determinant of the matrix (4) and $\lambda_{k_1}(t),\ldots,\lambda_{k_n}(t)$ are its eigenvalues. It is worth noticing that, generally, the Schrödinger equation is used for predicting the probability $\rho_{kk}(t)$ that the value u_k could be obtained in an observation of the observable U at time t. As we can see, the nondiagonal elements $\rho_{kl}(t)$, $(k\neq l)$, of the density matrix at time t are not less important, essentially contributing to the form of the probability measure η in the space of wave functions and to the measure of the interdependence created, due to the time evolution of the system, among the components of the random wave function. As we can see from (5), if during the time interval $[0,t]$ the density matrix $C_n(t)$ remains a diagonal matrix, then $W(t)=0$. But, generally, such a thing does not happen and the components of $\psi(t)$ do become interdependent.

Because the random variables $\{X_{k_1,j_1}(t),\ldots,X_{k_n,j_n}(t)\}$ are normal, this system is independent if and only if its components are not correlated, i.e., $\rho_{k_sk_v}(t)=0$ for $j_s\neq j_v$. Also, we can calculate, using a standard technique for Gaussian systems,[26] the expectation of any polynomial in the above random variables.

Let us take the coefficients $\{X_1(t),\ldots,X_{n/2}(t)\}$ (n even), of the random wave ψ, where $X_k(t)=X_{k,1}(t)+iX_{k,2}(t)$. Let $C_n(t)=[C(X_{k,j}(t),X_{l,s}(t))]$ be the corresponding covariance matrix. Then the probability density (1) with this covariance matrix may be used for calculating the probability of some events of interest. Particularly, if ϵ, δ, a_k, and b ($a_k<b$), are positive numbers, then

$$\eta(\{\,|\,|X_k(t)|^2-a_k|<\epsilon,\ \left|\sum_{s=1}^{n/2}|X_s(t)|^2-b\right|<\delta\})$$

gives an approximation, to within an error of $\epsilon+\delta$, of the probability that the probability that the observable U will take on the value u_k, at time t, is equal to a_k/b, relatively to the eigenvalues $u_1,\ldots,u_{n/2}$ of the observable U. We can see here the double intervention of randomness: We make a prediction on the probability of a probability of an event.

VI. WIENER-SIEGEL PROBABILITY MEASURE

Long ago, Einstein[27] pleaded in favor of associating a kind of Brownian motion to elementary particles. Later, Wiener and Siegel[3,4] *postulated* a Gaussian type of probability measure in the space of the generic coefficients of the wave function representing the state of a quantum system. In their approach a kind of postulated Brownian motion was taken as a basic description not explained as a result of impacts of finer particles. According to Bohm[28] such an idea would in effect bring in a new kind of order and measure, "If it were pursued seriously, this would imply a change of possible structures that would perhaps be as great as that implied by the change from Ptolemaic epicycles to Newtonian equation of motion. Actually, this line was not seriously pursued in classical physics." Without assigning such an exaggerated significance to it, we have to admit, however, that the Wiener-Siegel approach has not heretofore received the attention it deserves. There are similarities and basic differences between the Wiener-Siegel approach and the content of this paper. The common point is that both deal with a probability measure on the space of normed or unnormed possible wave functions. But in the Wiener-Siegel approach the Gaussian measure is postulated in a rather abrupt way, which perhaps partially explains its little impact, while here the Gaussian probability measure is a consequence of the application of the principle of maximum entropy, and of the classical viewpoint from statistical mechanics according to which we are looking for the most random, or most unbiased, probabilistic model compatible with the given data and constraints. Also, the Wiener-Siegel approach does not make use of the usual kind of available information, while here the density matrix is taken as starting point.

With the above notations, the Wiener-Siegel probability measure postulated in the space of wave functions $\psi=\{(X_{k,1},X_{k,2});k\in D\}$ is generated by the independent Gaussian probability density

$$\eta_{\alpha_1\cdots\alpha_n}(x_1,\ldots,x_n)=(2\pi)^{-n/2}\exp\left[-\sum_{j=1}^{n}x_j^2/2\right],$$

having the mean vector equal to zero and the unit matrix as the covariance matrix. It induces an exponential distribution with the parameter equal to 1 on the set of possible values of $|X_k|^2$.

VII. CONCLUSIONS

When the density matrix is known, we construct the most random, or the most unbiased, probability measure on the space of wave functions by applying the principle of maximum entropy and the principle of minimum interdependence. Such an attempt extends, at a second level of randomization (i.e., on the space of wave functions which are already used for making probabilistic predictions about the possible values of an observable), the way of constructing a probability measure on the state space in the classical statistical mechanics. The Gaussian product probability measure on the infinite dimensional space of the coefficients of possible wave functions is the analog of the probability density function from the classical statistical mechanics. Its covariance function is just the density matrix and it implies an exponential distribution on the possible values of the squared absolute values of the coefficients of the wave function. In this model a diagonal density matrix at the initial time $t=0$ induces an independent Gaussian product probability measure on the space of wave functions. Due to the time evolution of the system, described by the Schrödinger equation, this initial probability measure is transformed, in the time interval $[0,t]$, into a Gaussian product probability measure with interdependence between components. The amount of in-

terdependence between the random coefficients of the wave function may be calculated using a simple formula, essentially depending on the nondiagonal elements of the density matrix at time $t > 0$. This transformation of an independent Gaussian product probability measure on the space of wave functions, at the initial time $t = 0$, into a dependent Gaussian product probability measure, at time $t > 0$, may be viewed as a generalization to the quantum case of the Liouville theorem from the classical statistical mechanics. The model given in this paper does not make any suppositions about the eventual causes of the random fluctuations characterizing a quantum system, whether they are the basic, objective feature at the quantum level or they are produced by a sort of deterministic interactions at a subquantic level. It gives only a subjective probabilistic model for the space of the wave functions, determining the most random, or the most unbiased, probabilistic description compatible with the available data by using the maximum entropy technique which has proved so useful in the classical statistical mechanics.

ACKNOWLEDGMENT

Support from the Natural Sciences and Engineering Research Council (NSERC Canada) Grant No. A-5712 is acknowledged.

[1]J. von Neumann, Mathematische Grundlagen der Quantenmechanik (Springer-Verlag, Berlin, 1932).
[2]M. Born, Atomic Physics, 7th ed. (Hafner, New York, 1962).
[3]N. Wiener and A. Siegel, Phys. Rev. **91**, 1551 (1953).
[4]N. Wiener and A. Siegel, Nuovo Cimento, Suppl. **2**, 982 (1955).
[5]C. E. Shannon, Bell. Syst. Tech. J. **27**, 379 (1948).
[6]L. Boltzmann, Wein Ber. **76**, 373 (1877).
[7]J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University, New Haven, 1902).
[8]M. Planck, *Theorie der Warmestrahlung* (Barth, Leipzig, 1906).
[9]E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[10]J. Bernoulli, Ars Conjectandi (Bale, 1713).
[11]P. S. Laplace, *Théorie Analytique Des Probabilitiés* (Courcier, Paris, 1812).
[12]R. S. Ingarden, Fortschr. Phys. **13**, 755 (1965).
[13]*The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT, Cambridge, Mass., 1979).
[14]*Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice (Cambridge University Press, Cambridge, England, 1986).
[15]E. T. Jaynes, *Papers on Probability, Statistics, and Statistical Physics* (Reidel, Dordrecht, 1983).
[16]S. Guiasu and A. Shenitzer, Math. Intelligencer **7**, 42 (1985).
[17]P. M. Williams, British J. Phil. Sci. **31**, 131 (1980).
[18]S. Guiasu, *Information Theory with Applications* (McGraw-Hill, New York, 1977).
[19]S. Kullback and R. A. Leibler, Ann. Math. Stat. **32**, 79 (1951).
[20]R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, Oxford, 1938).
[21]S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).
[22]S. Kullback and M. A. Khairat, Ann. Math. Stat. **37**, 279 (1966).
[23]S. Guiasu, R. Leblanc, and C. Reischer, J. Inf. Optimiz. Sci. **3**, 149 (1982).
[24]S. Watanabe, *Knowing and Guessing* (Wiley, New York, 1969).
[25]S. Guiasu, Int. J. Theor. Phys. **26**, 281 (1987).
[26]J. Yeh, *Stochastic Processes and the Wiener Integral* (Dekker, New York, 1973).
[27]A. Einstein, J. Franklin Inst. **221**, 313 (1936).
[28]D. Bohm, *Wholeness and the Implicate Order* (Routledge and Kegan, London, 1980).