The Statistical Wave Function

R. D. Levine[1]

Statistical considerations are applied to quantum mechanical amplitudes. The physical motivation is the progress in the spectroscopy of highly excited states. The corresponding wave functions are "strongly mixed." In terms of a basis set of eigenfunctions of a zeroth-order Hamiltonian with good quantum numbers, such wave functions have contributions from many basis states. The vector $\mathbf{x}$ is considered whose components are the expansion coefficients in that basis. Any amplitude can be written as $\mathbf{a}^\dagger \cdot \mathbf{x}$. It is argued that the components of $\mathbf{x}$ and hence other amplitudes can be regarded as random variables. The maximum entropy formalism is applied to determine the corresponding distribution function. Two amplitudes $\mathbf{a}^\dagger \cdot \mathbf{x}$ and $\mathbf{b}^\dagger \cdot \mathbf{x}$ are independently distributed if $\mathbf{b}^\dagger \cdot \mathbf{a} = 0$. It is suggested that the theory of quantal measurements implies that, in general, one can one determine the distribution of amplitudes and not the amplitudes themselves.

1. INTRODUCTION

I have had the benefit of many stimulating discussions of the maximum entropy formalism with Prof. Howard Reiss. While his most recent interest[1] was in applications to sociotechnical systems, I am primarily concerned with molecular and nuclear physics. There is a reason behind my confining the range of applications. Suppose that we have a continuous variable x defined over the interval zero to one. Without any additional information, the probability density $f_x(x)$ of x which is of maximal entropy

---

is the uniform one. Let y be another continuous variable, $y = x^2$. The probability density $f_y(y)$ which is of maximal entropy is the uniform one. This is clearly inconsistent with familiar rule

$$f_y(y) = \sum f_x(x)/|dy/dx| \tag{1}$$

where the sum is over all such values of x that correspond to a given value of y. The problem is not limited to continuous variables. Take a variable j whose range is the nonnegative integers up to some maximal value. The maximal entropy distribution of j is uniform. An example where this will be physically incorrect is if j is the length of the angular momentum vector $\mathbf{j}$ (it is $\mathbf{j}$ itself which is uniformly distributed).

The resolution of the problem is that in the maximal entropy approach one must provide an identification of the particular variable which, in the absence of any information, has a uniform distribution.[2] This is so whether the variable is discrete or continuous. The identification of this, sometimes called "natural," variable must be provided from outside of the formalism. A paradigm was provided by Maxwell when he derived the Boltzmann velocity distribution from the requirement that the distribution remains invariant under rotation of the coordinate system and by von Neumann when he used a similar invariance argument to introduce the expression for the entropy of a quantal mixture.[3] I have always felt that the (often implicit) identification of the natural variable is the most problematic aspect in the applications of the maximum entropy formalism to problems outside the physical sciences.[4] In this article I discuss a problem in quantum mechanics where the identification of the natural variable is also not obvious. The novel aspect is that I shall discuss the entropy of the distribution of quantum mechanical *amplitudes*. In this case, too, an appeal to an invariance argument will suggest the proper variable. Before turning to the technical discussion, I begin with general considerations regarding fluctuating amplitudes followed by a specific motivation, with applications to a problem of current experimental interest.

2. FLUCTUATIONS IN QUANTAL AMPLITUDES

The essence of the problem is as follows. Consider a quantum mechanical system in a particular pure state θ. Given a complete set of orthonormal eigenstates ϕ_r of some observable R, $R\phi_r = r\phi_r$, one can expand θ in this basis

$$\theta = \sum_r \langle \phi_r | \theta \rangle \phi_r \tag{2}$$

Here, $\langle \theta | \phi_r \rangle$ is the scalar product, which we shall refer to as the

amplitude. Unless θ is one of the basis functions, there will be more than one nonzero amplitude. The familiar interpretation is that the observable R upon measurement for the system in the state θ will have a distribution of values. The probability of observing R to have the value r is the modulus squared of the relevant amplitude, i.e., $|\langle\theta|\phi_r\rangle|^2$. We are usually cautioned not to confuse this inherently quantal dispersion in the results of measuring R with a classical uncertainty. The latter corresponds to a quantum mechanical mixture.[3] In the Copenhagen interpretation, a pure quantum mechanical state is sharply defined. The dispersion in the observed values for R is a result of the filtration[5] due to the measurement process. The familiar expression for the mean value of R

$$\langle R \rangle = \sum_r r\,|\langle\theta|\phi_r\rangle|^2 \tag{3}$$

may look just like a classical average, but one must not interpret the pure state as a "mixture." The loss of phase is a result of the measurement process. My intention is not to challenge these ideas, but to take them somewhat further in the same logical direction.

There is a dispersion in the measurement of R, since θ is not an eigenstate of R. Well, θ is not an eigenstate of the operator $\delta(R-r)$ either. Why then do we assign a definite answer to the measurement of this operator? In part, this is a figure of speech. What we really know is $\langle\delta(r-R)\rangle$, i.e., the mean value of the operator. The possible conclusion is then that the amplitude $\langle\theta|\phi_r\rangle$ can fluctuate about its mean value. It is these fluctuations that we want to examine.

One can always argue that the fluctuations we speak of are not those in the expansion coefficients of the wave function θ itself but of the mixed state obtained following the measurement of R, or of $\delta(R-r)$. That is possible, but does call for certain modifications on the technical level. The reason is that the traditional answer for the mixed state ρ following a measurement (without sorting[5]) of the values of R is

$$\rho = \sum_r |\langle\theta|\phi_r\rangle|^2\,|\phi_r\rangle\langle\phi_r| \tag{4}$$

In such a state $\delta(R-r)$ has a sharp value without fluctuations.

On the other hand, we must recognize that the mixed state ρ [Eq. (4)] does not correspond to a unique wave function. Hence, if we are given the ρ_{rr} (the diagonal elements of the diagonal density operator ρ), we cannot assign a unique wave function. This is the technical modification that we regard as inevitable: If the results of quantum mechanical measurements are invariably mixed states,[3,5] then all we can measure are diagonal

elements of the density matrix.[2] Under such circumstances we must allow for the possibility that the amplitudes of any particular wave function will fluctuate.

A word about notation. In his book, Tolman[6] writes $\overline{\overline{|a|^2}}$ for the quantity I denote by $\langle |x|^2 \rangle$. Tolman introduces an ensemble such that

$$\rho_{sr} \equiv \overline{\overline{a_r^* a_s}} = \frac{1}{N} \sum_{\alpha=1}^{N} a_{r(\alpha)}^* a_{s(\alpha)} \tag{5}$$

where $a_{r(\alpha)}$ is the rth amplitude for the member α, $\alpha = 1,..., N$, of the ensemble. I reiterate that our ensemble here is the collection of wave functions which correspond to a given density matrix ρ, my not quite orthodox point of view being that if ρ is what can be measured, then it is ρ (rather than the wave function) that is what we know. Of course, one can arrange for a measurement to specify a pure state, $\rho^2 = \rho$.

3. THE SIGNATURE OF CLASSICAL CHAOS

The problem I discuss arose in the context of looking for the signature of classical chaos[7] in the optical spectra of molecules.[8,9] For high-lying excited states of molecules, an eigenfunction $|\theta\rangle$ of the true molecular Hamiltonian is a linear combination of many eigenstates $|\phi_r\rangle$ of an approximate (or "zeroth-order") molecular Hamiltonian.[10] The approximate Hamiltonian has good quantum numbers, so that the optical transition probabilities $\langle i| \mu |\phi_r\rangle$ from some low-lying state $|i\rangle$ can be readily estimated. Often, only a few states ϕ_r carry oscillator strength from $|i\rangle$. These are then known as "the bright states." The transition probability $|\langle i| \mu |\theta\rangle|^2 \equiv \langle \theta| \mu |i\rangle\langle i| \mu |\theta\rangle$ is then given in terms of the amplitudes

$$\langle i| \mu |\theta\rangle = \sum_r \langle i| \mu |\phi_r\rangle\langle \phi_r|\theta\rangle \tag{6}$$

Unless there is just one bright state, the probability will contain many cross terms,

$$|\langle i| \mu |\theta\rangle|^2 = \sum_{r,s} \langle \theta|\phi_s\rangle\langle \phi_s| \mu |i\rangle\langle i| \mu |\phi_r\rangle\langle \phi_r|\theta\rangle \tag{7}$$

An important question, which I shall address below, is how to determine the minimal number of bright states necessary to interpret the spectra. For future reference note that the absorption intensity can be written as

$$\zeta = (\mathbf{d}^\dagger \cdot \mathbf{x})^2 = \mathbf{x}^\dagger \cdot (\mathbf{dd}^\dagger) \cdot \mathbf{x} \tag{8}$$

---

[2] The density matrix need be diagonal only in the appropriate basis. It need not be diagonal in another basis. Hence, off-diagonal elements can be measured by a suitable arrangement.

where **d** is the "systematic" part of the transition dipole, i.e., a column vector whose elements are $\langle i| \mu |\phi_r \rangle$, and **x** is a column vector of the amplitudes. $\mathbf{dd}^\dagger$ is a square matrix, which is the matrix representation of the operator $\mu |i\rangle \langle i| \mu$ in the basis $\{\phi_r\}$. A Gaussian distribution for tamplitudes of strongly mixed wave functions has been proposed on the basis of semiclassical considerations[11] and verified on the basis of numerical evidence.[10]

Much earlier, the problem of fluctuating amplitudes arose in molecular collision theory.[12,13] Consider a problem with many accessible final states (denoted by r; these are eigenstates of the asymptotic Hamiltonian.) Then, with $|\theta\rangle = S|i\rangle$, where S is the scattering operator, the scalar products are the transition amplitudes $\langle \phi_r|\theta\rangle = \langle \phi_r|S|i\rangle$. The unitarity of the S operator implies that the transition amplitudes can be regarded as components of a unit vector in multidimensional space. That is, of course, always true for the scalar products themselves,

$$\sum_r |\langle \phi_r|\theta\rangle|^2 = 1 \qquad (9)$$

It was felt that when the initial state $|i\rangle$ is strongly coupled to many final states ϕ_r, the unit vector will be randomly distributed over the unit sphere. The numerical evidence is discussed in ref. 14. In those early days, the concept of classical chaos was not widespread. Instead, one spoke of "dominant coupling," that is, a wave function with many nonvanishing amplitudes with respect to a basis set of physical relevance. In scattering theory there is a natural choice for such a basis as a basis for a zeroth-order Hamiltonian corresponding to noninteracting ($t \to \infty$) projectiles. Feshbach[15] pointed out that there can be other choices (PHP in his notation). This is useful whenever there is a separation of time scales.[16]

Similar developments in nuclear physics gave rise to random matrix (RM) theory.[17,18] I shall not follow this approach for two reasons. The first is that to determine the distribution of the scalar products $\langle \phi_r|\theta\rangle$ for variable r (and fixed R), the RM approach is to hold r fixed and to introduce an ensemble of operators R. The other reason is that I shall argue that a "universal" functional form for the fluctuations is only possible in the chaotic limit, which is, I claim, what the RM method provides. One needs, however, to be able to discuss deviations (possibly quite large) from that limit. For a computational example showing such deviations and the onset of chaotic behavior for stronger coupling see ref. 19.

4. THE JOINT DISTRIBUTION OF FLUCTUATIONS

A given state θ corresponds to many (possibly an infinite number) of amplitudes. Hence one must consider in general the joint distribution of these amplitudes. For such a joint distribution there are different fluctuations that can be computed. There are also different, so-called "marginal" distributions which correspond to the distribution in one or more (but not all) amplitudes. It is seldom that we are interested in the full distribution $P_x(\mathbf{x})$ where $\mathbf{x}$ is a (column) vector whose elements are the amplitudes. There are two distributions in which we are particularly interested. The first is the (marginal) distribution of a particular amplitude x, $P_x(x)$,

$$P_x(x) = \int d\mathbf{x}' \, P_x(\mathbf{x}) \tag{10}$$

where $d\mathbf{x}'$ implies integration over all *other* amplitudes.

The second distribution of interest is that of the results of measurements. Say the operator A is measured and let $\zeta \equiv \langle \theta | A | \theta \rangle$. Then, if $\mathbf{A}$ is the matrix representation of the operator A in the basis ϕ_r, we have, using (2),

$$\zeta = \mathbf{x}^\dagger \mathbf{A} \mathbf{x} \tag{11}$$

and

$$P_\zeta(\zeta) = \int d\mathbf{x} \, \delta(\zeta - \mathbf{x}^\dagger \mathbf{A} \mathbf{x}) \, P_x(\mathbf{x}) \tag{12}$$

A special case of ζ is the variable $y_r = |x_r|^2$ corresponding to $A_{s,t} = \delta_{r,s}\delta_{r,t}$. A similar form for A is for intensities in optical spectra when $\zeta = |\langle i| \mu |f\rangle|^2$ and i and f denote the initial and final states for the transition. Note that we do not necessarily imply that the state f is one of the eigenstates r. Hence, in this case we can take the operator A to be $\mu |i\rangle\langle i| \mu$ and the amplitudes of the final state f in the basis of the ϕ_r [see Eq. (2)].

5. A SIMPLE DERIVATION

A simple derivation of the form of the distribution $P_x(x)$ can be based on an ergodiclike assumption, as follows. What we always know about the amplitudes is that they are normalized[3]

$$\sum_{r=1}^{R} x_r^2 = \langle \theta | \theta \rangle \tag{13}$$

[3] Note that if $|\theta\rangle$ is normalized, $\langle \theta | \theta \rangle = 1$. The point is that the sum (13) is a property of the state $|\theta\rangle$.

where for simplicity we write (13) as if the amplitudes are real (this restriction is removed later.) Now comes the assumption that the average of x_r^2 over all states r equals the average of x^2 over its distribution $P_x(x)$. This is the assumption that each amplitude fluctuates independently with the same distribution. Then, one can rewrite (13) as

$$\langle x^2 \rangle \equiv \int x^2 P_x(x) \, dx = \langle \theta | \theta \rangle / R \tag{14}$$

A more innocent way of getting this result is to say that we replace in (13) the summation over all R states r by a summation over all values of the amplitude so that $P_x(x) \, dx$ is the fraction of states r with an amplitude in the range x to $x + dx$.

The distribution $P_x(x)$ is normalized

$$\int P_x(x) \, dx = 1 \tag{15}$$

and is known to have a given second moment [Eq. (14)]. It is otherwise unknown, and hence we select the particular distribution which satisfies the normalization (15) and the second moment (14) constraints, and is otherwise of maximal entropy. The result is a Gaussian:

$$P_x(x) = (2\pi \langle x^2 \rangle)^{-1/2} \exp(-x^2 / 2\langle x^2 \rangle) \tag{16}$$

This is not a pleasing result for the expermentalists. What is directly measured are the probabilities $y_r = x_r^2$, and (16) implies that their distribution [see Eq. (1)]

$$P_y(y) = (2\pi \langle y \rangle)^{-1/2} y^{-1/2} \exp(-y / 2\langle y \rangle) \tag{17}$$

is strongly peaked at the origin. If, say, the y_r are optical transition probabilities, then (17) implies sensitivity to the weakest transitions, i.e., the ones most likely to be masked by noise.

At this point, the caveat mentioned in the introduction needs to be considered. The result (17) is an immediate consequence of the fact that we chose to maximize the entropy $S[x]$ of $P_x(x)$,

$$S[x] \equiv -\int dx \, P_x(x) \ln P_x(x) \tag{18}$$

If instead we had decided to maximize the entropy of $P_y(y)$, we would have obtained a distribution which is still peaked at the origin but with a more moderate decline. Our choice is based on the requirement that the dis-

tribution of amplitudes, when we know nothing about the state $|\theta\rangle$ besides its normalization $\langle\theta|\theta\rangle$, should be the same for all choices of the basis sets $\{\phi_r\}$. In particular, it should be invriant to changes in the phase of the basis functions, so that $P_x(x) = P_x(-x)$. This implies that a given value of y is doubly degenerate, i.e., it can result from two distinct "elementary" events $y = |x|^2$ and $y = |-x|^2$. A more elegant way of stating the same condition is

$$P_x\left(\frac{x_1 + x_2}{\sqrt{2}}\right) P_x\left(\frac{x_1 - x_2}{\sqrt{2}}\right) = P(x_1) P(x_2) \tag{19}$$

which leads to the same conclusion.

In general, the amplitudes will be complex and $y = x_1^2 + x_2^2$, where x_1 and x_2 are the real and imaginary parts of the amplitude respectively. Hence the degeneracy of y is higher. The normalization condition (14) now reads

$$\langle|\mathbf{x}|^2\rangle \equiv \langle x_1^2\rangle + \langle x_2^2\rangle$$

$$\equiv \int dx_1 \, dx_2 \, x_1^2 P_\mathbf{x}(x_1, x_2) + \int dx_1 \, dx_2 \, x_2^2 P_\mathbf{x}(x_1, x_2) = \langle\theta|\theta\rangle/R \tag{20}$$

Maximizing the entropy of $P_\mathbf{x}(x_1, x_2)$ subject to its normalization and to (20) leads to

$$P_\mathbf{x}(x_1, x_2) = P_x(x_1) P_x(x_2) \tag{21}$$

with

$$P_x(x_1) = (2\pi\langle|x|^2\rangle)^{-1/2} \exp(-x_1^2/2 \langle|x|^2\rangle) \tag{22}$$

and similarly for the distribution of x_2. Note that it is an implication of (21) that (1) x_1 and x_2 are independently distributed and (2) the Lagrange multiplier for x_1 and x_2 has the same value. This is an immediate consequence of the constraint (20), which only specifies the value of the sum $\langle x_1^2\rangle + \langle x_2^2\rangle$ but not the magnitude of the two individual terms. For this case we have

$$P_\mathbf{x}(x_1, x_2) = (2\pi\langle|x|^2\rangle)^{-1} \exp(-|x|^2/\langle|x|^2\rangle) \tag{23}$$

and changing variables to $|x|$ and θ, $x_1 = |x| \cos\theta$, and $x_2 = |x| \sin\theta$, we have, on integration over θ,

$$P_y(y) = \langle y\rangle^{-1} \exp(-y/\langle y\rangle) \tag{24}$$

The result that different amplitudes are distributed as independent Gaussian variables of zero mean and common variance is valid only when the sole constraint imposed on the distribution is the sum rule (20). Such a sum rule is invariably valid. Hence, the joint distribution over n basis states

$$P_{\mathbf{x}}(\mathbf{x}) = (2\pi n \langle |x|^2 \rangle)^{-n} \exp(-\mathbf{x}^\dagger \mathbf{x}/2\langle |x|^2 \rangle) \qquad (25)$$

is only valid in the absence of any additional constraints. It makes sense to suggest that such would be the case for a state θ which is "strongly mixed." It has indeed been proposed[11] that the distribution corresponds to what we mean by "quantum chaos." Our own point of view[8,9] is that it is the absence of any system-specific constraints that identifies the most entropic distribution (25) as the limiting case of being most chaotic.

6. SIMPLE APPLICATIONS

In general, we need the distribution of $\zeta = \mathbf{x}^\dagger \mathbf{A} \mathbf{x}$, [see Eq. (12)] when the variables $\mathbf{x}$ are independent Gaussian variables. $\mathbf{A}$ is the matrix representation of an observable and hence can always be brought to diagonal form

$$\mathbf{A} = \mathbf{U}^{-1} \mathbf{a} \mathbf{U} \qquad (26)$$

by a unitary transformation. In (26), $\mathbf{a}$ is a diagonal matrix made up of the eigenvalues of $\mathbf{A}\mathbf{z}_i = a_i \mathbf{z}_i$. Some of these may be zero (in which case the rank of $\mathbf{A}$ is lower than its dimension), and others may be degenerate. Hence, ζ can be written, using the new amplitudes $\mathbf{x}'$,

$$\mathbf{x}' = \mathbf{U}\mathbf{x} \qquad (27)$$

as

$$\zeta = R^{-1} \mathbf{x}'^\dagger \mathbf{a} \mathbf{x}' = R^{-1} \sum_{i=1}^{n} a_i x_i'^2 = R^{-1} \sum_{i=1}^{n} a_i (\mathbf{z}_i \cdot \mathbf{x})^2 \qquad (28)$$

Here, n is the rank of $\mathbf{A}$, R is the dimension of $\mathbf{A}$, and we have rescaled the amplitudes $\mathbf{x}'$ so that $\langle |x'|^2 \rangle = 1$. Since $P_{\mathbf{x}}$ is invariant under a unitary distribution, ζ is the weighted sum of n independent Gaussian variables with zero mean and unit variance. This is a central conclusion and hence I reiterate that in the limit when the state in question is strongly mixed, any expectation value is the sum of n independent Gaussian variables. It is also to be noted that n can be quite small, e.g., $n = 1$ if A is a projection operator. Indeed, in general, by the spectral theorem

$$A = \sum_{i=1}^{n} a_i P_i \qquad (29)$$

where the P_i are projection operators on orthogonal subspaces, $P_i P_j = \delta_{i,j} P_i$. I shall restate this result below.

The moments of the distribution $P_\zeta(\zeta)$ are given in the appendix to ref. 7. They can be computed via the characteristic function $\phi_\zeta(t)$ of $P_\zeta(\zeta)$,

$$\phi_\zeta(t) = \int \exp(it\zeta) \, P_\zeta(\zeta) \, d\zeta \tag{30}$$

as follows[4]: Let $\phi_{x'}(t_1, t_2, ...)$ be the characteristic function of $P_{x'}(x'_1, x'_2, ...)$. Then, by the Cramer–Wold theorem,[20]

$$\phi_\zeta(t) = \phi_{x'}(R^{-1}a_1 t, R^{-1}a_2 t, ...)$$

where the a_i are as in Eq. (29).

The characteristic function of a normal variable of zero mean and unit variance is

$$\phi_{x'}(t) = (1 - 2it)^{-1/2}$$

Hence,

$$\phi_\zeta(t) = \prod_{i=1}^{n} (1 - 2iR^{-1}a_i t)^{-1/2} = |\mathbf{I} - 2itR^{-1}\mathbf{A}|^{-1/2} \tag{31}$$

where $\det \mathbf{A} \equiv |\mathbf{A}|$ and $\mathbf{I}$ is the unit matrix.

In general, the Fourier transform of $\phi_\zeta(t)$ cannot be evaluated analytically. An exception is when all nonzero eigenvalues of $\mathbf{A}$ are equal, so that

$$\zeta = (a/R) \sum_{i=1}^{n} x_i'^2, \qquad \langle \zeta \rangle = (a/R) \left\langle \sum_i x_i'^2 \right\rangle = n(a/R) \tag{32}$$

Then,

$$P_\zeta(\zeta) = \zeta^{n/2 - 1} \exp(-n\zeta/2\langle\zeta\rangle)/(2\langle\zeta\rangle/n)^{n/2} \, \Gamma(n/2) \tag{33}$$

which is known as a "chi-square distribution with n degrees of freedom." Otherwise, one can use $\log \phi_\zeta(t)$ as the cumulant generating function[20] to obtain for the mth cumulant

$$\kappa_m = 2^{m-1}(m-1)! \sum_{i=1}^{n} (a_i/R)^m$$
$$= 2^{m-1}(m-1)! \, R^{-m} \operatorname{Tr} \mathbf{A}^m \tag{34}$$

---

[4] Prof. W. H. Miller informs me that he has independently obtained this result.

When all nonzero eigenvalues of $\mathbf{A}$ are equal, this reduces to

$$\kappa_m = 2^{m-1}(m-1)! \, n(a/R)^m \tag{35}$$

which is the correct result for the chi-square distribution (33), as $\langle \zeta \rangle = (a/R) \, n$.

As a check of the manipulations that led to (35), we compute κ_2 directly. For this purpose note that since the x_r are independent Gaussian variables with a variance $\langle |x_r|^2 \rangle = 1/R$ independent of r; then

$$\langle |x_r|^4 \rangle = 3/R^2$$
$$\langle |x_r|^2 \, |x_s|^2 \rangle = 1/R^2 \tag{36}$$
$$\langle x_r^* x_s x_{r'}^* x_{s'} \rangle = 0, \qquad \text{otherwise}$$

Hence

$$\langle (\zeta - \langle \zeta \rangle)^2 \rangle = \frac{2}{R^2} \sum_{r=1}^{n} A_{rs} A_{sr} = 2\langle \zeta \rangle^2 \, \mathrm{Tr}(\mathbf{A}^2)/[\mathrm{Tr}(\mathbf{A})]^2 \tag{37}$$

We return now to the central result. Consider two observables

$$\zeta_1 = \mathbf{x}^\dagger \mathbf{A}_1 \mathbf{x} \qquad \zeta_2 = \mathbf{x}^\dagger \mathbf{A}_2 \mathbf{x} \tag{38}$$

A necessary and sufficient condition that these observables be independent is that[20]

$$\mathbf{A}_1 \mathbf{A}_2 = 0 \tag{39}$$

It follows that given a Hermitian matrix $\mathbf{A}$, it can always be resolved, using the spectral theorem, as a sum of matrices $\mathbf{A}_i$ which satisfy (39). Hence, one always regards the variables $\mathbf{x}'$ as providing a set of independently fluctuating components of the variable ζ. Moreover, the number of such independent components is the rank of the matrix $\mathbf{A}$. The rank is, of course, independent of the basis used to initially compute the matrix $\mathbf{A}$. This conclusion is of sufficient importance that I discuss the physical implications below. In terms of the amplitudes themselves, the two amplitudes $\mathbf{d}_1^\dagger \cdot \mathbf{x}$ and $\mathbf{d}_2^\dagger \cdot \mathbf{x}$ are independently distributed if $\mathbf{d}_1^\dagger \cdot \mathbf{d}_2 = 0$. The proof is an immediate consequence of (39) by taking $\mathbf{A}_1 \equiv \mathbf{d}_1 \cdot \mathbf{d}_1^\dagger$ and similarly for $\mathbf{A}_2$. We reiterate that all these conclusions are valid for the distribution (25), which has been derived in the simple case where the normalization is the only constraint.

7. THE INVARIANCE ARGUMENT

From the beginning I have used the notation $P_x(\mathbf{x})$ because it is the *amplitudes* that turn out to be the natural variables. At first glance, this offends our classical sensibilities, which regard the probabilities $y_r = x_r^2$ as the real observables. However, the transformation law in quantum mechanics is linear in the amplitudes. That is, the amplitudes $\mathbf{x}'$ in the new basis ϕ' are related to the amplitudes $\mathbf{x}$ by

$$\mathbf{x}' = \mathbf{U}\mathbf{x} \tag{40}$$

where $\mathbf{U}$ is a unitary (the so-called "overlap") matrix. An observable A transforms as

$$\mathbf{A}' = \mathbf{U}\mathbf{A}\mathbf{U}^{-1} \tag{41}$$

so that $\zeta = \mathbf{x}^\dagger \mathbf{A}\mathbf{x}$ is invariant to the choice of basis. Indeed, while we have not been able to compute $P_\zeta(\zeta)$ in general, note that its cumulants, Eq. (34), are independent of the basis used to represent $\mathbf{A}$, since the trace of $\mathbf{A}^m$ has the same value in any basis.

8. GATEWAY STATES

In general, as long as the amplitudes $\mathbf{x}$ are distributed as independent Gaussian variables with a common mean, we are able to write any expectation value as a sum of n, $n \leqslant R$, independent variables

$$\langle \theta | A | \theta \rangle = \sum_{i=1}^{n} a_i(\mathbf{z}_i \cdot \mathbf{x}) \tag{42}$$

To discuss the physical implication, consider a special form of A, namely

$$A = T^\dagger P T \tag{43}$$

where T is a transition operator and P is a projection. For example, if $P \equiv \delta(E - H)$ then, but for kinematic factors, $\langle \theta | A | \theta \rangle$ is the total rate of transitions out of the state $|\theta\rangle$. If T is the dipole operator and P is the projector on the initial state, then we have the intensity of the optical transition into $|\theta\rangle$, etc.

The eigenvectors $\mathbf{z}_i$, $i = 1,..., n$, provide an orthonormal basis for $\mathbf{A}$

$$\mathbf{A} = \sum_i a_i \mathbf{z}_i^\dagger \cdot \mathbf{z}_i, \qquad \mathbf{z}_i \mathbf{z}_j^\dagger = \delta_{ij} \tag{44}$$

such that the transition probability is a sum of n *independent* components. I argue that each such component represents what is often known as a

"doorway" or a "gateway" for the transition, that is, an independent component. If all the a_i are similar in magnitude, one can determine n from the magnitude of the width $\kappa_2 = \sigma^2$ of the distribution of amplitudes. Thus, using (37), with $(a/R) = \langle \zeta \rangle / n$, one has

$$\langle (\zeta - \langle \zeta \rangle)^2 \rangle = (2/n) \langle \zeta \rangle^2 \tag{45}$$

In general, using (34), one has

$$\langle (\zeta - \langle \zeta \rangle)^2 \rangle = 2R^{-2} \sum_{i=1}^{n} a_i^2 \geqslant \left(\frac{2}{n} \right) \langle \zeta \rangle^2 \tag{46}$$

Hence, if we define n_{eff} by

$$\langle (\zeta - \langle \zeta \rangle)^2 \rangle = (2/n_{\text{eff}}) \langle \zeta \rangle^2 \tag{47}$$

it follows that $n_{\text{eff}} \leqslant n$. The variance of the distributions of ζ provides a lower bound for the number of doorway states.

9. THE GENERAL CASE

Thus far I have discussed the simplest case when the only information about the state $|\theta\rangle$ was its normalization. This served to fix the sum of the variances

$$\sum_r \langle |x_r|^2 \rangle = \langle \theta | \theta \rangle \tag{48}$$

and with the invariance condition $\langle x_i \rangle = 0$ implied a normal distribution with a common variance and zero covariance for all the amplitudes. In the most general case what we know about a state is its density matrix ρ. With the interpretation (5), we have

$$\rho_{sr} = \langle x_r^* x_s \rangle = \langle (x_r - \langle x_r \rangle)^* (x_s - \langle x_s \rangle) \rangle \tag{49}$$

in other words, *at most*, the covariance matrix ρ of the amplitudes. I emphasize that this conclusion is not, as far as I can establish, in conflict with any standard result of quantum mechanics and in particular is consistent with the primary postulate

$$\langle A \rangle = \text{Tr}(A\rho) \tag{50}$$

To prove (50), note that

$$\langle A \rangle \equiv \sum_{rs} A_{rs} \langle x_r^* x_s \rangle = \sum_r (A\rho)_{rr} \tag{51}$$

On the other hand, the variance of $\zeta = \mathbf{x}^\dagger \mathbf{A} \mathbf{x}$ is *not* the same as the variance of the operator A {i.e., $\mathrm{Tr}[(A - \langle A \rangle)^2 \rho]$} unless ρ is a pure state.

The distribution of the amplitudes which is of maximal entropy subject to a given density matrix ρ is a multivariate normal distribution

$$P_\mathbf{x}(\mathbf{x}) = (2\pi)^{-R} |\rho^{-1}| \exp\left(-\frac{1}{2} \mathbf{x}^\dagger \rho^{-1} \mathbf{x} \right) \qquad (52)$$

provided ρ^{-1} exists. To prove the normalization, consider the characteristic function

$$\phi_t(\mathbf{t}) = \int \exp\left(i\mathbf{t}^\dagger \mathbf{x} - \frac{1}{2} \mathbf{x}^\dagger \rho^{-1} \mathbf{x} \right) d\mathbf{x} \qquad (53)$$

which can be evaluated by changing variables to the basis that diagonalizes ρ. Then,

$$\phi_t(\mathbf{t}) = \exp\left(-\frac{1}{2} \mathbf{t}^\dagger \rho^{-1} \mathbf{t} \right) \qquad (54)$$

The details are standard (e.g., ref. 20, p. 272), except for an extra factor of $(2\pi)^{-1/2}$ per basis function which is contributed by the integration over the phase of the amplitudes in the basis that diagonalizes ρ.

The characteristic function $\phi_\zeta(t)$ of $\zeta = \langle A \rangle = \mathbf{x}^\dagger \mathbf{A} \mathbf{x}$ can also be evaluated

$$\phi_\zeta(t) = |\mathbf{I} - 2it\mathbf{A}\rho|^{-1/2} \qquad (55)$$

The result (55) is an obvious generalization of (31), since (31) corresponds to $\langle x_r^* x_s \rangle = \delta_{rs} \langle |x_r|^2 \rangle = \delta_{rs}/R$, or $\rho = \mathbf{I}/R$, where $\mathbf{I}$ is the identity matrix. As before, the Fourier transform of (55) needs to be numerically evaluated. The moments can be evaluated using (34). As noted earlier, $\langle \zeta \rangle = \mathrm{Tr}(\mathbf{A}\rho)$, but the variance κ_2 of $\langle \zeta \rangle$, $\kappa_2 = 2\,\mathrm{Tr}(\mathbf{A}\rho)^2$, is *not* the variance of A. The generalization of the condition $\mathbf{d}_1^\dagger \cdot \mathbf{d}_2 = 0$ for the independence of two amplitudes is, when the covariance matrix of the distribution is ρ

$$\mathbf{d}_1^\dagger \rho \mathbf{d}_2 = 0 \qquad (56)$$

and for $\rho = \mathbf{I}/R$ we recover the simple case above.

The special case of a pure state corresponds to ρ being idempotent, $\rho^2 = \rho$. The density matrix is then singular. Furthermore, it is no longer accurate enough to impose normalization "on the average" $\sum_r \langle |x_r|^2 \rangle = 1$, but we must have $\sum_r |x_r|^2 = 1$.[5] This will make $P_\mathbf{x}$ proportional to $\delta(1 - \sum x_r^2)$. If R is large and $|\theta\rangle$ is not one of the basis states ϕ_r, $P_\mathbf{x}$ will still be a Gaussian.

---

[5] We can also impose this condition in the general case. It will mean that the rank of ρ is $R - 1$ and not R.

10. THE INTERMEDIATE CASE

In the general case we need to specify the complete density matrix ρ. In the simple case all that we specified is $\text{Tr}(\rho I)$, that is, the normalization of ρ. It is often the case[21] that we know several expectation values for ρ, but not ρ itself. The maximum entropy formalism seeks to determine that density matrix consistent with the given expectation values which is otherwise of maximal entropy. Our considerations would seem to suggest that this prescription is possibly incomplete. For each ρ there is a distribution of amplitudes, i.e., a distribution over states of the ensemble giving rise to ρ.[6] For each ρ there is thus an *a priori* weight associated with the number of states in the ensemble. Should we not include this in our expression for the entropy, which should not be $-\text{Tr}(\rho \ln \rho)$, but rather

$$\text{entropy} = -\text{Tr} \, \rho \ln \rho + \text{Tr} \, \rho S(\rho) \tag{57}$$

where $S(\rho)$ is the entropy of the ensemble of states which is consistent with a given ρ?

I argue that while (57) is strictly speaking a more correct expression, it turns out not to make a difference. That is, that $S(\rho) = -(1/2)\ln \rho + \text{const}$ and hence the very same ρ would result from the maximum entropy formalism whether we use $-\text{Tr}(\rho \ln \rho)$ or (57) for the entropy. The Lagrangian of the problem is

$$L = \text{entropy} - \lambda_0 \, \text{Tr} \, \rho I - \sum_j \lambda_j \, \text{Tr}(\rho A_j) \tag{58}$$

where the A_j are the observables whose expectation values are given and the λ's are Lagrange multipliers whose values are determined by the given expectation values. The constant term in $S(\rho)$ in (57) can be absorbed by the normalization constraint. The entropy of the density matrix of maximal entropy is thus the same function of the expectation values $\langle A_j \rangle$, irrespective of which scheme we use. An appeal to thermodynamics therefore cannot resolve the issue.

To avoid the complexities of taking most general variations in ρ which still leaves it Hermitian, we take the case of one observable A which is diagonal in the basis $\{\phi_r\}$. Given is $\langle A \rangle$,

$$\langle A \rangle = \sum_r \langle x_r^2 \rangle A_{rr} \tag{59}$$

and we further simplify by taking x to be real. $\langle x_r^2 \rangle$ itself is an average

$$\langle x_r^2 \rangle = \int dx_r \, x_r^2 P_{x_r}(x_r) \tag{60}$$

and

$$S_r = -\int dx_r \, P_{x_r} \ln P_{x_r} \tag{61}$$

The entropy to be maximized subject to the given value of $\langle A \rangle$ and to

$$\sum_r \langle x_r^2 \rangle = 1 \tag{62}$$

is

$$\text{entropy} = -\sum_r p_r \ln p_r + \sum_r p_r S_r \tag{63}$$

Taking first the variation of P_{x_r} at a given p_r, we get as before

$$P_{x_r} = (2\pi \langle x_r^2 \rangle)^{-1/2} \exp(-x_r^2/2\langle x_r^2 \rangle) \tag{64}$$

so that

$$S_r = (1/2)[\ln(\langle x_r^2 \rangle) + \ln(2\pi e)] \tag{65}$$

and the values of the p_r, $p_r = \langle x_r^2 \rangle$, can now be determined by a second variation, leading to the familiar final result

$$p_r = \exp(-\lambda_0 - \lambda A_{rr}) \tag{66}$$

11. CONCLUDING REMARKS

For the strongly mixed wave functions of highly excited states it makes physical sense to regard the amplitudes $\mathbf{d}^\dagger \cdot \mathbf{x}$ as fluctuating random variables. Such a point of view is indeed inherent in the statistical theories in nuclear and molecular physics. Even the Gaussian distribution that was derived here using the maximum entropy formalism has been previously proposed on the basis of semiclassical considerations of strong mixing,[19] computational evidence,[10] and random matrix theory.[17] I have, however, further suggested that a correlated Gaussian distribution may be valid under much more general circumstances. That the physics is not unreasonable is shown, for example, by the following argument. Consider the spectroscopy of low-lying states, where wave functions are not strongly mixed. As we scan the frequency we go through a sequence of final states with the optical amplitudes $\mathbf{d}^\dagger \cdot \mathbf{x}_{\mathbf{n}_1}$, $\mathbf{d}^\dagger \cdot \mathbf{x}_{\mathbf{n}_2}$,..., where the $\mathbf{n}_j$ are the sets of quantum numbers of the final states. Adjacent states can have quite

different quantum numbers.[22] The values of the amplitudes will therefore vary rapidly. Hence, in the regular regime we expect the amplitudes to fluctuate even more than at higher energies. This is equally the case if we consider a set of different amplitudes $\mathbf{a}^\dagger \cdot \mathbf{x}, \mathbf{b}^\dagger \cdot \mathbf{x}, \ldots$ for a given state. Fluctuations tend to diminish as the states become more strongly mixed.

The Gaussian fluctuations of the amplitudes are reminiscent of the thermodynamic theory of fluctuations. Here, however, it is the amplitudes which are the variables. One can, however, derive our results for the distribution of observables, e.g., Eq. (33), from a thermodynamiclike approach.[23] It then turns out that in the limit $n \to \infty$ (where n, the number of degrees of freedom, is the rank of $\mathbf{A}$, $\zeta = \mathbf{x}^\dagger \mathbf{A} \mathbf{x}$), the distribution of ζ is indeed Gaussian. The very concept of an amplitude can arise naturally in an attempt to endow statistics with a geometrical structure.[24] The inherently "quantal" aspects enter only through the representation of observables by matrices which may fail to commute. The degree of this failure is measured by Planck's constant h, while the scale of amplitude fluctuations with which we have been concerned is determined by Boltzmann's constant k. It is very tempting to speculate that these two constants are dimensionally more similar than appears at first glance. So far, however, I am not aware that such speculations have given rise to new physical insights.

ACKNOWLEDGMENTS

REFERENCES

1. H. Reiss, A. D. Hammerich, and E. W. Montroll, *J. Stat. Phys.* **42**:647 (1986).
2. U. Dinur and R. D. Levine, *Chem. Phys.* **5**:17 (1975).
3. J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955).
4. E. W. Montroll, *Proc. Natl. Acad. Sci. USA* **78**:7839 (1981).
5. J. Schwinger, *Proc. Natl. Acad. Sci. USA* **45**:1552 (1959).
6. R. C. Tolman, *The Principles of Statistical Mechanics* (Clarendon Press, Oxford, 1938).
7. E. J. Heller, *J. Chem. Phys.* **72**:1337 (1980); E. B. Stechel and E. J. Heller, *Annu. Rev. Phys. Chem.* **35**:563 (1984).
8. Y. Alhassid and R. D. Levine, *Phys. Rev. Lett.* **57**:2879 (1986).
9. R. D. Levine, *Adv. Chem. Phys.* **70**:53 (1988).
10. R. B. Gerber, V. Buch, and M. A. Ratner, *Chem. Phys. Lett.* **89**:171 (1982).
11. M. V. Berry, *J. Phys. A* **17**:2083 (1977).

12. R. B. Bernstein, A. Dalgarno, H. S. W. Massey, and I. C. Percival, *Proc. Soc. A* **274**:427 (1963).
13. R. D. Levine, *Quantum Mechanics of Molecular Rate Processes* (Clarendon Press, Oxford, 1969).
14. R. D. Levine and R. B. Bernstein, *J. Chem. Phys.* **53**:686 (1970).
15. H. Feshbach, *Ann. Phys.* **19**:287 (1962).
16. H. S. Taylor, *Int. J. Quant. Chem.* **31**:747 (1987).
17. C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press, New York, 1965).
18. T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong, *Rev. Mod. Phys.* **53**:585 (1981).
19. J. Brickmann, Y. M. Engel, and R. D. Levine, *Chem. Phys. Lett.* **137**:441 (1987).
20. P. A. P. Moran, *An Introduction to Probability Theory* (Clarendon Press, Oxford, 1968).
21. E. T. Jaynes, *Phys. Rev.* **108**:171 (1957).
22. I. C. Percival, *Adv. Chem. Phys.* **36**:1 (1977).
23. R. D. Levine, in *Large Finite Systems*, J. Jortner, A. Pullman, and B. Pullman, eds. (Reidel, Dordrecht, 1987).
24. R. D. Levine, *J. Chem. Phys.* **84**:910 (1986).