International Journal of Theoretical Physics, Vol. 11, No. 6 (1974), pp. 357–378

Estimation of a Displacement Parameter of a Quantum System

CARL W. HELSTROM†
Laboratoire d'Etude des Phénomènes Aléatoires, Université de Paris-Sud
91405 *Orsay, France*

Received: 1 *May* 1974

Abstract

A displacement parameter such as the angle of rotation or the position of a quantum system, or the phase of a harmonic oscillation, is to be estimated by observing the system with an apparatus that applies to it an operator-valued measure (o.v.m.). The o.v.m. minimizing the average cost of errors in the estimate is determined by quantum estimation theory for a system in a pure state. The best estimate of the parameter is found to be the more accurate, the greater the uncertainty of the dynamical variable serving as the infinitesimal generator of the displacement group. The relation of this result to such uncertainty principles as those between angle and angular momentum, and between oscillator phase and photon number, is discussed. A lower bound to the variance of an unbiased estimate of the time of occurrence of an event in the evolution of a system is derived from the quantum-mechanical Cramér–Rao inequality. It is inversely proportional to the square of the uncertainty of the energy of the system.

1. *Quantum Estimation Theory*

The uncertainty principle for two quantum-mechanical variables, represented by Hermitian operators $\mathscr{A}$ and $\mathscr{B}$, is expressed by the formula

$$\Delta\mathscr{A}^2\,\Delta\mathscr{B}^2 \geqslant \tfrac{1}{4}\langle\mathscr{C}\rangle^2 \tag{1.1}$$

in which $i\mathscr{C} = [\mathscr{A},\mathscr{B}]$ is the commutator of $\mathscr{A}$ and $\mathscr{B}$, and $\langle\mathscr{C}\rangle$ denotes the expected value (Robertson, 1929). The uncertainties $\Delta\mathscr{A}$ and $\Delta\mathscr{B}$ are defined by

$$\Delta\mathscr{A}^2 = \langle(\mathscr{A}-\langle\mathscr{A}\rangle)\rangle^2, \qquad \Delta\mathscr{B}^2 = \langle(\mathscr{B}-\langle\mathscr{B}\rangle)^2\rangle \tag{1.2}$$

The most familar such relation connects the uncertainties of the position $\mathscr{Q}$ and the momentum $\mathscr{P}$ of a particle.

† On leave from the University of California, San Diego, La Jolla, California 92037, U.S.A. Publication assisted by NSF Grant GK-33811.

If one variable is the angular position or azimuth θ of a quantum system about an axis, and the other is the component J_z of angular momentum along that axis, one expects an uncertainty relation of the form

$$\Delta J_z \, \Delta\theta \geqslant ? \tag{1.3}$$

for like $\mathscr{P}$ and $\mathscr{Q}$ these are conjugate variables in classical mechanics; but as indicated by Judge (1964), an operator corresponding to the azimuth θ and possessing with J_z a commutator of the necessary form cannot be defined. The supposed uncertainty relation

$$\Delta n \, \Delta\theta \geqslant ? \tag{1.4}$$

between the number n of photons in an harmonic oscillator and the phase θ of its oscillation suffers a similar impediment. These difficulties have been circumvented by introducing operators C and S corresponding to $\cos\theta$ and $\sin\theta$ and by establishing for them a more complicated uncertainty relation that goes into (1.3) and (1.4) in the limit of large quantum numbers. This approach has been thoroughly reviewed by Carruthers & Nieto (1968). Here we shall describe how a more general formulation of quantum measurement, when utilized in quantum estimation theory, permits a simpler derivation and interpretation of uncertainty relations involving such variables as an azimuth or phase. These are viewed not as dynamical variables represented by operators, but as parameters of the state of the system.

Suppose that an apparatus A_p prepares a quantum system S in a state $|\psi\rangle$. If the combination of apparatus and system is rotated through an angle θ about the z-axis, the apparatus A_p will instead prepare the system S in the state

$$|\psi(\theta)\rangle = e^{iN\theta}|\psi\rangle \tag{1.5}$$

where N, the infinitesimal generator of the group of rotations about the z-axis, is the operator $J_z/\hbar$, $\hbar$ = Planck's constant $h/2\pi$. In a Stern-Gerlach experiment, for instance, a beam of spin $-\frac{1}{2}$ particles is divided into two beams, in one of which the spins point upward and in the other downward. The 'spin-up' beam passes through a hole in a screen, which intercepts the 'spin-down' beam. The whole device is rotated about an axis concident with the 'spin-up'beam, which then contains particles whose spin vector points in the direction θ with respect to a y-axis normal to the beam.

In this way the angle θ of rotation is a parameter of the state of the system S, specifying its orientation about the z-axis. If you did know through what angle θ the preparing apparatus had turned, you might ask how you could observe S in order to determine θ as accurately as possible. The results of your observations and your subsequent calculations with them would yield for the parameter θ an estimate $\hat{\theta}$ that in general would be somewhat in error, $\hat{\theta} \neq \theta$. We shall see that the azimuth θ can be the more accurately estimated, the more broadly the state $|\psi(\theta)\rangle$ is distributed among the eigenstates of the angular momentum J_z.

The phase θ_0 of an oscillation of frequency ν is related to the origin t_0 of the time scale,

$$\theta_0 = 2\pi\nu t_0 + \text{constant}$$

If the apparatus A_p preparing the state of the oscillator is turned on at a different time $t_0 + \tau$, the phase will be changed by $\theta = 2\pi\nu\tau$, and the state of the oscillator will be

$$|\psi(\theta)\rangle = e^{iN\theta}|\psi\rangle \tag{1.6}$$

instead of $|\psi\rangle$. Here N is the number operator a^+a, a and a^+ being the annihilation and creation operators for photons. The phase θ appears as a parameter of the state of the oscillator, and we shall study how it can best be estimated.

When a quantum system is observed, it interacts with an apparatus A_m that produces certain perceptible macroscopic phenomena such as the positions of meter needles or the readings of counting devices. The numbers so obtained may be subjected to certain calculations in order to produce other numbers, say $z_1, z_2, \ldots, z_m$, that are considered as the results of the observation. We can represent them as a point $\mathbf{z}$ in a certain 'outcome space' Z. They are random variables in the sense that repeating the observation on a number of systems S prepared in the same state yields in general a variety of sets $\mathbf{z}$ of outcomes, which must be described by a probability distribution on the space Z. The concept of quantum observation to be applied here associates with each apparatus A_m an operator-valued measure (o.v.m.) mapping arbitrary regions Δ of Z into non-negative definite Hermitian operators $\Pi(\Delta)$ acting on the Hilbert space $\mathscr{H}_S$ of the system. It is postulated that the probability that the results $\mathbf{z}$ fall into the region Δ is

$$\Pr\{\mathbf{z} \in \Delta\} = \mathrm{Tr}\,[\rho\Pi(\Delta)] \tag{1.7}$$

when ρ was the density operator of the system S before the observation; 'Tr' stands for the trace.

In order for the probability $\Pr\{\mathbf{z} \in \Delta\}$ to possess the properties of ordinary probability, the o.v.m. $\{\Pi\}$ must satisfy certain conditions.

(i) The empty set $\emptyset$ in Z maps into the zero operator of $\mathscr{H}_S$,

$$\emptyset \to \Pi(\emptyset) = \mathbf{0} \tag{1.8}$$

(ii) The entire space Z maps into the identity operator,

$$Z \to \Pi(Z) = \mathbf{1} \tag{1.9}$$

(iii) If the regions $\Delta_1, \Delta_2, \ldots$, are disjoint,

$$\Delta_1 + \Delta_2 + \cdots \to \Pi(\Delta_1 + \Delta_2 + \cdots) = \Pi(\Delta_1) + \Pi(\Delta_2) + \cdots$$
$$\Delta_1 \cap \Delta_2 \cap \ldots = \emptyset \tag{1.10}$$

even unto an infinite number of regions Δ_i. We say that the observing apparatus A_m 'applies' the o.v.m. $\{\Pi\}$ to the system S.

The common view of quantum measurement as yielding an eigenvalue of an Hermitian operator, or the eigenvalues of a number of commuting Hermitian operators, corresponds to the application of a projection-valued or orthogonal o.v.m., in which the operators $\Pi(\Delta)$ are projectors on to linear subspaces of $\mathscr{H}_S$. The need for the more general formulation has been brought out by Davies & Lewis (1970) and Benioff (1972a, b).

Let the density operator $\rho = \rho(\theta_1, \theta_2, \ldots, \theta_m)$ of the system S depend on certain parameters $\theta_1, \theta_2, \ldots, \theta_m$ that are to be estimated. We represent these estimanda as a point $\boldsymbol{\theta}$ in a parameter space Θ, writing $\rho = \rho(\boldsymbol{\theta})$. The estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ will be the results of observing the system S by an apparatus A_m of the kind just described. They correspond to a point $\hat{\boldsymbol{\theta}}$ in the space Θ, which is thus also the outcome space Z of the observation. The probability that the estimate $\hat{\boldsymbol{\theta}}$ lies in a certain region Δ of Θ when the true values of the parameters are specified by $\boldsymbol{\theta}$ is

$$\Pr\{\hat{\boldsymbol{\theta}} \in \Delta \,|\, \boldsymbol{\theta}\} = \mathrm{Tr}\,[\rho(\boldsymbol{\theta})\Pi(\Delta)] \tag{1.11}$$

where $\{\Pi\}$ is the o.v.m. characterising the observation. We seek the o.v.m. that produces the most accurate estimates.

Conceivably the purpose of an observation is to decide which of a set $\rho_1, \rho_2, \ldots, \rho_M$ of density operators best represents the state of the system S. The space Θ now contains only a finite set of points, which can be labeled with the integers $1, 2, \ldots, M$, and one can look for the o.v.m. that, for instance, minimises the average probability of error in choosing among them. This is a problem of quantum hypothesis testing, which has application to optical communications (Helstrom, Liu & Gordon, 1970). If the parameter space Θ is divided into M regions, and one asks only for the region in which the point $\boldsymbol{\theta}$ lies, estimation reduces to hypothesis-testing.

Here we shall suppose that the parameter space Θ is continuous and that the o.v.m. $\{\Pi\}$ can be generated by a set of infinitesimal, non-negative definite Hermitian operators $d\Pi(\hat{\boldsymbol{\theta}})$ associated with the points of Θ in such a way that

$$\Pi(\Delta) = \int_{\Delta} d\Pi(\hat{\boldsymbol{\theta}}) \tag{1.12}$$

and in particular

$$\Pi(\Theta) = \int_{\Theta} d\Pi(\hat{\boldsymbol{\theta}}) = \mathbf{1} \tag{1.13}$$

An o.v.m. so generated conforms to the rules (i)–(iii). The conditional probability density function (p.d.f.) of the estimates $\hat{\boldsymbol{\theta}}$ is then, in accordance with (1.11), given by

$$q(\hat{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta})\, d^m\hat{\boldsymbol{\theta}} = \mathrm{Tr}\,[\rho(\boldsymbol{\theta})\, d\Pi(\hat{\boldsymbol{\theta}})]\,, \tag{1.14}$$

$d^m\hat{\boldsymbol{\theta}} = d\hat{\theta}_1 \ldots d\hat{\theta}_m$ being the infinitesimal volume element of Θ.

Imitating conventional statistical estimation theory, one defines a function

$C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ that assesses the cost of errors in the estimates. A cost function often adopted for mathematical convenience is the quadratic form

$$C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{m} \sum_{j=1}^{m} G_{ij}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \tag{1.15}$$

where G_{ij} is an element of a positive-definite symmetric matrix. Another useful cost function is

$$C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = -\prod_{i=1}^{m} \delta(\hat{\theta}_i - \theta_i) \tag{1.16}$$

In ordinary statistics it leads to the maximum-likelihood estimators, the set of values $\boldsymbol{\theta}$ at which the posterior p.d.f. $p(\boldsymbol{\theta}|\mathbf{x})$ of the parameters $\boldsymbol{\theta}$, given the results x of observation, is maximum.

What is known in advance about the values of the parameters $\boldsymbol{\theta}$ to be anticipated is incorporated in their joint prior p.d.f. $z(\boldsymbol{\theta})$. If we suppose that the o.v.m. $\{\Pi\}$ is applied to many copies of the system S, with the parameters $\boldsymbol{\theta}$ of its density operator $\rho(\boldsymbol{\theta})$ distributed according to $z(\boldsymbol{\theta})$, the average cost incurred will be

$$\begin{aligned} \bar{C}[\Pi] &= \int_{\Theta} \int_{\Theta} z(\boldsymbol{\theta}) C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta})\, d^m\boldsymbol{\theta}\, d^m\hat{\boldsymbol{\theta}} \\ &= \mathrm{Tr} \int_{\Theta} \int_{\Theta} z(\boldsymbol{\theta}) C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \rho(\boldsymbol{\theta})\, d\Pi(\hat{\boldsymbol{\theta}})\, d^m\boldsymbol{\theta} \end{aligned} \tag{1.17}$$

and we want the o.v.m. $\{\Pi\}$ for which this average cost is minimum.

Equations determining the minimising or 'optimum' o.v.m. have been given by Holevo (1973). Defining the Hermitian operator $W(\hat{\boldsymbol{\theta}})$ by

$$W(\hat{\boldsymbol{\theta}}) = \int_{\Theta} z(\boldsymbol{\theta}) C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \rho(\boldsymbol{\theta})\, d^m\boldsymbol{\theta} \tag{1.18}$$

we write the average cost as

$$\bar{C}[\Pi] = \mathrm{Tr} \int_{\Theta} W(\hat{\boldsymbol{\theta}})\, d\Pi(\hat{\boldsymbol{\theta}}) \tag{1.19}$$

For the optimum o.v.m. the operator

$$\Upsilon = \int_{\Theta} W(\hat{\boldsymbol{\theta}})\, d\Pi(\hat{\boldsymbol{\theta}}) \tag{1.20}$$

must be Hermitian and satisfy

$$W(\hat{\boldsymbol{\theta}}) - \Upsilon \geqslant \mathbf{0}, \qquad \forall\, \hat{\boldsymbol{\theta}} \in \Theta \tag{1.21}$$

the inequality meaning that the operator on the left is non-negative-definite. The operator Υ corresponds to a Lagrange multiplier bringing in the constraint (1.13). Furthermore

$$[W(\hat{\boldsymbol{\theta}}) - \Upsilon]\, d\Pi(\hat{\boldsymbol{\theta}}) \equiv \mathbf{0} \tag{1.22}$$

and $d\Pi(\hat{\boldsymbol{\theta}})$ must be non-negative-definite and satisfy (1.13). The minimum attainable cost is then

$$\bar{C}_{\min} = \operatorname{Tr} \Upsilon \tag{1.23}$$

as follows from integrating (1.22) over Θ.

Suppose that $d\Pi'(\hat{\boldsymbol{\theta}})$ generates some other o.v.m. conforming to the rules (i)–(iii) and hence to (1.13). The difference between the cost it incurs and the minimum cost is, by (1.23),

$$\bar{C}[\Pi'] - \bar{C}_{\min} = \operatorname{Tr} \int_{\Theta} [W(\hat{\boldsymbol{\theta}}) - \Upsilon]\, d\Pi'(\hat{\boldsymbol{\theta}})$$

As the product of two non-negative-definite Hermitian operators has a non-negative trace, (1.21) shows that

$$\bar{C}[\Pi'] - \bar{C}_{\min} \geqslant 0$$

and if $d\Pi'(\hat{\boldsymbol{\theta}})$ equals the operator $d\Pi(\hat{\boldsymbol{\theta}})$ satisfying (1.22), this difference is equal to zero.

By using the delta-function cost function (1.16) in (1.18), we find that the quantum-mechanical equivalent of the maximum-likelihood estimator is the o.v.m. $\{\Pi_m\}$ generated by non-negative-definite Hermitian operators $d\Pi_m(\hat{\boldsymbol{\theta}})$ satisfying.

$$[z(\hat{\boldsymbol{\theta}})\rho(\hat{\boldsymbol{\theta}}) - \Upsilon_m]\, d\Pi_m(\hat{\boldsymbol{\theta}}) = \mathbf{0} \tag{1.24}$$

$$\Upsilon_m - z(\hat{\boldsymbol{\theta}})\rho(\hat{\boldsymbol{\theta}}) \geqslant \mathbf{0} \tag{1.25}$$

in which the Lagrange operator

$$\Upsilon_m = \int_{\Theta} z(\hat{\boldsymbol{\theta}})\rho(\hat{\boldsymbol{\theta}})\, d\Pi_m(\hat{\boldsymbol{\theta}}) \tag{1.26}$$

must be Hermitian.

As an example, consider a system whose Hilbert space is known to have finite dimension n. In terms of an orthogonal set of basic vectors $|\varphi_k\rangle$, the state vector of the system is expressed as

$$|\psi(\mathbf{c})\rangle = \sum_{k=1}^{n} c_k |\varphi_k\rangle, \qquad \mathbf{c} = (c_1, c_2, \ldots, c_n)$$

The n complex numbers $c_k = c_{kx} + ic_{ky}$ can be considered as parameters of the density operator

$$\rho(\mathbf{c}) = |\psi(\mathbf{c})\rangle\langle\psi(\mathbf{c})|$$

and we seek their maximum-likelihood estimates $\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n)$.

Because the vector $|\psi(\mathbf{c})\rangle$ must have unit length,

$$\sum_{k=1}^{n} |\hat{c}_k|^2 = \sum_{k=1}^{n} (\hat{c}_{kx}^2 + \hat{c}_{ky}^2) = 1$$

the point $\hat{\mathbf{c}} = (\hat{c}_{1x}, \hat{c}_{1y}, \ldots, \hat{c}_{nx}, \hat{c}_{ny})$ must lie on a $(2n)$-dimensional hypersphere S_{2n} of radius 1, which becomes the parameter space Θ. If nothing is known in advance about the state of the system, we can suppose that the point $\mathbf{c}$ may be anywhere on S_{2n}, and we assign it the prior p.d.f.

$$z(\mathbf{c}) = A_{2n}^{-1} = \tfrac{1}{2}\Gamma(n)\pi^{-n}$$

where A_{2n} is the area of the unit hypersphere.

The maximum-likelihood estimator is the o.v.m. generated by

$$d\Pi_m(\hat{\mathbf{c}}) = nA_{2n}^{-1}|\psi(\hat{\mathbf{c}})\rangle\langle\psi(\hat{\mathbf{c}})|\,dS$$

$$= nA_{2n}^{-1}\sum_{k=1}^{n}\sum_{m=1}^{n}\hat{c}_k\hat{c}_m^*|\varphi_k\rangle\langle\varphi_m|\,dS$$

where dS is an element of area on S_{2n}. When $d\Pi_m(\hat{\mathbf{c}})$ is integrated over S_{2n}, the terms with $k \neq m$ vanish, and for terms with $k = m$ one must calculate the average of $|\hat{c}_k|^2 = \hat{c}_{kx}^2 + \hat{c}_{ky}^2$ over the hypersphere; it comes out equal to n^{-1}. Thus $d\Pi_m(\hat{\mathbf{c}})$ integrates to the identity operator $\mathbf{1}$ as required. The reader can easily verify that the Lagrange operator is

$$\Upsilon_m = A_{2n}^{-1}\mathbf{1}$$

and that the optimisation equations (1.24) and (1.25) are satisfied. The conditional p.d.f. of the point $\hat{\mathbf{c}}$ on S_{2n} is given by

$$q(\hat{\mathbf{c}}\,|\,\mathbf{c})\,dS = \mathrm{Tr}\,[\rho(\mathbf{c})\,d\Pi(\hat{\mathbf{c}})] = nA_{2n}^{-1}|\langle\psi(\hat{\mathbf{c}})\,|\,\psi(\mathbf{c})\rangle|^2\,dS$$

This distribution is independent of the choice of basic vectors $|\varphi_k\rangle$. From it one can show that the quantity $r = |\langle\psi(\hat{\mathbf{c}})\,|\,\psi(\mathbf{c})\rangle|$, which is the absolute value of the complex cosine of the angle between the true state vector and the estimated state vector, has the conditional p.d.f.

$$q(r) = 2n(n-1)r^3(1-r^2)^{n-2}$$

in constrast to the p.d.f.

$$q_0(r) = 2(n-1)r(1-r^2)^{n-2}$$

that it would have if one picked the point $\hat{\mathbf{c}}$ at random on S_{2n}.

In the next section we shall apply the optimisation equations to the estimation of a displacement parameter of a quantum system.

2. *Estimation of Displacement*

The parameter θ of a density operator $\rho(\theta)$ is called a *displacement parameter* if the density operator has the form

$$\rho(\theta) = \mathrm{e}^{iN\theta}\rho_0\,\mathrm{e}^{-iN\theta}, \qquad \rho_0 = \rho(0) \tag{2.1}$$

where N is a constant operator serving as the infinitesimal generator of the group of displacements of the system. During most of the discussion we shall suppose that θ is an angle taking values in $\Theta = (-\pi, \pi)$. If θ is the azimuth of the system about the z-axis, $N = J_z/\hbar$; if θ is the phase of a harmonic oscillation, $N = a^+a$. By scaling θ, however, we can have it represent the position of a particle in a finite segment $-\frac{1}{2}L < x < \frac{1}{2}L$ of the x-axis, imposing periodic boundary conditions at each end; N is then proportional to the momentum operator $\mathscr{P}$. Similarly $T\theta/2\pi$ might be the time at which a signal in a transmission line or waveguide passes a certain point, T equaling the length of the line divided by the velocity of propagation. The operator N is then $(-H/\hbar)$, where H is the Hamiltonian operator for the electromagnetic field of the line.

Assuming that nothing at all is known about the true value of the displacement parameter θ before observation, we assign to it the uniform prior p.d.f.

$$z(\theta) = (2\pi)^{-1}, \qquad -\pi < \theta \leqslant \pi \tag{2.2}$$

We seek the o.v.m. $d\Pi_m(\hat{\theta})$ specifying the quantum maximum-likelihood estimator and satisfying the optimization equations (1.24)-(1.26). We shall find that the same o.v.m. minimises the average cost of error as assessed by the cost function

$$C(\hat{\theta}, \theta) = 4\sin^2\tfrac{1}{2}(\hat{\theta} - \theta) \tag{2.3}$$

This cost is approximately $(\hat{\theta} - \theta)^2$ for small errors, and it is easier to handle mathematically than the squared error, which for a periodic displacement parameter is most meaningfully defined by

$$C'(\hat{\theta}, \theta) = \min\{(\hat{\theta} - \theta)^2, (2\pi - \hat{\theta} + \theta)^2\}$$

for θ and $\hat{\theta}$ in $(-\pi, \pi)$. Our principal interest lies in errors much smaller than $\pi/2$.

Because two estimates $\hat{\theta}$ differing by an integral multiple of 2π are equivalent, the o.v.m. $\{\Pi_m\}$ must be generated by an operator

$$d\Pi_m(\hat{\theta}) = \xi(\hat{\theta})\,d\hat{\theta}, \qquad \xi(\hat{\theta}) \geqslant \mathbf{0} \tag{2.4}$$

for which the operator $\xi(\hat{\theta})$ is periodic in $\hat{\theta}$ with period 2π. Our problem is invariant to a change in the origin of the coordinate θ, the prior p.d.f. $z(\theta)$ being uniform, and the generating operator $\xi(\theta)$ must therefore have the form

$$\xi(\hat{\theta}) = \mathrm{e}^{iN\hat{\theta}}\xi_0\,\mathrm{e}^{-iN\hat{\theta}}, \qquad \xi_0 \geqslant \mathbf{0} \tag{2.5}$$

The conditional p.d.f. of the estimate $\hat{\theta}$ will then be

$$\begin{aligned} q(\hat{\theta}\,|\,\theta) &= \mathrm{Tr}\,[\rho(\theta)\xi(\hat{\theta})] \\ &= \mathrm{Tr}(e^{iN\theta}\rho_0 e^{-iN\theta}e^{iN\hat{\theta}}\xi_0 e^{-iN\hat{\theta}}) \\ &= \mathrm{Tr}\,[\rho_0 e^{iN(\hat{\theta}-\theta)}\xi_0 e^{-iN(\hat{\theta}-\theta)}] = \bar{q}(\hat{\theta}-\theta) \end{aligned}$$

and it will depend only on the error $\varphi = \hat{\theta} - \theta$, with

$$\bar{q}(\varphi) = \mathrm{Tr}(\rho_0 e^{iN\varphi}\xi_0 e^{-iN\varphi}) \tag{2.6}$$

We work with a representation whose basis is formed by the eigenstates $|n\rangle$ of the operator N,

$$N\,|n\rangle = n\,|n\rangle \tag{2.7}$$

The eigenvalues n are either integers or odd multiples of $\frac{1}{2}$, for displacing the system by $\theta = 2\pi$ brings it either back to the same state or into a state whose sign is reversed. Only differences of the eigenvalues enter our equations, and they must be integers. We assume that the eigenvalues n are simple; how to modify the analysis when they are degenerate is described in the Appendix.

The condition $\Pi_m(\Theta) = \mathbf{1}$ now yields, by (2.5),

$$\begin{aligned} \delta_{nm} &= \int_{-\pi}^{\pi} \langle n\,|\,\xi(\hat{\theta})\,|\,m\rangle\, d\hat{\theta} \\ &= \int_{-\pi}^{\pi} e^{i(n-m)\theta}\langle n\,|\,\xi_0\,|\,m\rangle\, d\theta = 2\pi\langle n\,|\,\xi_0\,|\,m\rangle\,\delta_{nm} \end{aligned} \tag{2.8}$$

in terms of the Kronecker delta; all the diagonal elements of the matrix $\langle n\,|\,\xi_0\,|\,m\rangle$ must be equal to $(2\pi)^{-1}$. The matrix for the Lagrange operator Υ_m is diagonal, for by (1.26), (2.1), and (2.5)

$$\begin{aligned} \langle n\,|\,\Upsilon_m\,|\,m\rangle &= \int_{-\pi}^{\pi} \langle n\,|\,\rho(\theta)\xi(\theta)\,|\,m\rangle\, d\theta/2\pi \\ &= \int_{-\pi}^{\pi} e^{i(n-m)\theta}\langle n\,|\,\rho_0\xi_0\,|\,m\rangle\, d\theta/2\pi = \langle n\,|\,\rho_0\xi_0\,|\,m\rangle\,\delta_{nm} \end{aligned} \tag{2.9}$$

Because Υ_m must be Hermitian, its diagonal elements

$$\langle n\,|\,\Upsilon_m\,|\,n\rangle = \langle n\,|\,\rho_0\xi_0\,|\,n\rangle = \langle n\,|\,\xi_0\rho_0\,|\,n\rangle$$

must be real; and because it is diagonal, Υ_m commutes with operator N. By (2.1) and (2.5) the optimization equations (1.24), (1.25) reduce to

$$[(2\pi)^{-1}\rho_0 - \Upsilon_m]\,\xi_0 = \mathbf{0} \tag{2.10}$$

$$\Upsilon_m - (2\pi)^{-1}\rho_0 \geqslant \mathbf{0} \tag{2.11}$$

At this juncture we postulate that the system S is in a pure state

$$|\psi(\theta)\rangle = e^{iN\theta}|\psi\rangle, \qquad \rho_0 = |\psi\rangle\langle\psi|$$

with θ to be estimated. The appendix to this paper shows that the solution of (2.9)–(2.11) is then

$$\xi_0 = (2\pi)^{-1}|\gamma\rangle\langle\gamma| \tag{2.12}$$

with the components $\langle n|\gamma\rangle$ of the vector $|\gamma\rangle$ given in the N-representation by

$$\begin{aligned} \gamma_n = \langle n|\gamma\rangle &= \exp(i \arg \psi_n), \qquad \psi_n \neq 0 \\ \psi_n = \langle n|\psi\rangle &= |\psi_n|\gamma_n \\ \gamma_n &= 1, \qquad \psi_n = 0 \end{aligned} \tag{2.13}$$

It is also demonstrated there that the o.v.m. $d\Pi_m(\hat{\theta})$ formed from ξ_0 through (2.4), (2.5) satisfies the optimization equations (1.20)–(1.22) for the cost function in (2.3).

The optimum o.v.m. in effect compensates for the phases of the components $\psi_n e^{in\theta}$ of the state vector $|\psi(\theta)\rangle$ in such a way that the posterior p.d.f. $\bar{q}(\hat{\theta} - \theta)$ of the estimates is as sharply peaked at $\hat{\theta} = \theta$ as possible. This p.d.f. is, by (2.6), given by

$$\bar{q}(\varphi) = (2\pi)^{-1}\Big|\sum_n |\psi_n| e^{-in\varphi}\Big|^2, \qquad \varphi = \hat{\theta} - \theta \tag{2.14}$$

The estimate $\hat{\theta}$ is unbiased because $\bar{q}(\varphi)$ is an even function of φ. A measure of the width in φ of this p.d.f. $\bar{q}(\varphi)$ is

$$\delta\varphi = \int_{-\pi}^{\pi} \bar{q}(\varphi)\, d\varphi / \bar{q}(0) = 2\pi \left[\sum_n |\psi_n|\right]^{-2} \tag{2.15}$$

It is the smaller, the more broadly the probabilities $P_n = |\psi_n|^2$ are distributed among the eigenstates $|n\rangle$ of the infinitesimal generator N. The minimum average value of the cost function $C(\hat{\theta}, \theta)$ from (2.3) is

$$\bar{C}_{\min} = \sum_n (|\psi_n| - |\psi_{n-1}|)^2 \tag{2.16}$$

and this too decreases as the probability distribution $\{P_n\}$ spreads out over more and more of the states $|n\rangle$.

If for instance

$$\begin{aligned} P_n &\equiv (2n_m + 1)^{-1}, \qquad |n| \leqslant n_m \\ P_n &\equiv 0, \qquad |n| > n_m \end{aligned}$$

we find

$$\bar{q}(\varphi) = \frac{\sin^2(n_m + \frac{1}{2})\varphi}{2\pi(2n_m + 1)\sin^2(\frac{1}{2}\varphi)}$$

$$\delta\varphi = 2\pi/(2n_m + 1), \qquad \bar{C}_{\min} = \delta\varphi/\pi$$

Thus if θ is the azimuth about the z-axis of a quantum system of total angular momentum $j_T\hbar$, the width $\delta\varphi$ of the p.d.f. of the best estimate of θ cannot be less than $2\pi/(2j_T + 1)$.

When the eigenvalues n are degenerate, it is only necessary to replace $|\psi_n|$ in (2.14)–(2.16) by $P_n^{1/2}$, where

$$P_n = \langle\psi|G_n|\psi\rangle \tag{2.17}$$

is the total probability attached to the eigenvalue n of N, G_n being the projector on to the linear subspace spanned by the eigenstates having that eigenvalue in common.

The optimum o.v.m. is orthogonal if the eigenvalue spectrum of the operator N contains all the integers from $-\infty$ to $+\infty$. Indeed,

$$\begin{aligned} d\Pi_m(\theta) &= (2\pi)^{-1} e^{iN\theta}|\gamma\rangle\langle\gamma| e^{-iN\theta}\, d\theta \\ &= |\gamma(\theta)\rangle\langle\gamma(\theta)|\, d\theta/2\pi \end{aligned} \tag{2.18}$$

and this will be an infinitesimal projector if the vectors $|\gamma(\theta)\rangle$ are orthogonal for all θ. The scalar product of two of them is

$$\langle\gamma(\theta)|\gamma(\theta')\rangle = \langle\gamma|e^{iN(\theta'-\theta)}|\gamma\rangle = \sum_n e^{in(\theta'-\theta)} \tag{2.19}$$

and this will equal $2\pi\delta(\theta - \theta')$ as required if all integers n in $(-\infty, \infty)$ appear in the sum. For estimation of the azimuth of a system with finite total angular momentum, the spectrum of N is finite, and for estimation of the phase of an oscillation, the spectrum contains only the non-negative integers. The optimum o.v.m.'s for these estimates are not orthogonal, and no Hermitian operator exists whose measurement on S in the conventional sense yields the best estimate $\hat{\theta}$.

The o.v.m. $\{\Pi_m\}$ can be approximated by a finite set of ν operators

$$\Pi_k = \int_{\theta_{k-1}}^{\theta_k} d\Pi_m(\hat{\theta})$$

$$\theta_k = 2\pi k/\nu, \qquad 1 \leqslant k \leqslant \nu \tag{2.20}$$

and as the estimate of θ one can take the mid-point of the interval (θ_{k-1}, θ_k) selected by an apparatus applying this finite o.v.m. $\{\Pi_k\}$ to the system. According to a theorem of Naĭmark's (1940), as interpreted by Holevo (1973), one can imbed the Hilbert space $\mathscr{H}_S$ in the product space $\mathscr{H}_S \otimes \mathscr{H}_A$ of the states of the system S and an ancillary system A that is in a pure state $|\psi_A\rangle$.

The state of the combination is

$$\rho(\theta) \otimes |\psi_A\rangle\langle\psi_A|$$

Vectors $|\Psi\rangle \in \mathscr{H}_S$ go into the vectors $|\Psi\rangle \otimes |\psi_A\rangle$, which span a subspace $\mathscr{H}'_S$ of $\mathscr{H}_S \otimes \mathscr{H}_A$. On $\mathscr{H}_S \otimes \mathscr{H}_A$ there operates an orthogonal o.v.m. composed of projectors E_k such that for all states $|\Psi\rangle$ in $\mathscr{H}_S$,

$$GE_k(|\Psi\rangle \otimes |\psi_A\rangle) = (\Pi_k|\Psi\rangle) \otimes |\psi_A\rangle$$

where

$$G = \mathbf{1} \otimes |\psi_A\rangle\langle\psi_A|$$

is the operator projecting arbitrary vectors in $\mathscr{H}_S \otimes \mathscr{H}_A$ on to $\mathscr{H}'_S$. Applying this projection-valued o.v.m. to the combination of S and A yields the same probabilities

$$\Pr\{\hat{\theta} \in (\theta_{k-1}, \theta_k) \,|\, \theta\} = \mathrm{Tr}\,[\rho(\theta)\Pi_k]$$

as applying the approximate o.v.m. $\{\Pi_k\}$. It is equivalent to measuring the Hermitian operator

$$\tilde{\theta} = \sum_{k=1}^{\nu} \tfrac{1}{2}(\theta_{k-1} + \theta_k)E_k$$

on S and A together. By taking the number ν of elements of this o.v.m. large enough, the optimum estimate $\hat{\theta}$ can be approximated as closely as desired.

The cosine and sine operators C and S defined by Carruthers & Nieto (1968) arise directly from the optimum o.v.m. for estimating the parameter θ when the components $\psi_n = \langle n|\psi\rangle$ of the state $|\psi\rangle$ in the N-representation are all real. (A phase factor $e^{i\eta}$ common to all components ψ_n is unobservable, and we suppose it has been eliminated.) We call such a state a *real* state, keeping in mind that its components may well not be real in any other representation. The phases of the components of the displaced states $|\psi(\theta)\rangle$ are then proportional to n. [*Note added in proof*: In this context 'real' means $\psi_n \equiv |\psi_n|$.]

We define the non-Hermitian operator E_- by

$$\begin{aligned} E_- &= \int_{-\pi}^{\pi} e^{i\theta}\, d\Pi_m(\theta) \\ &= \int_{-\pi}^{\pi} e^{i\theta} e^{iN\theta} |\gamma\rangle\langle\gamma| e^{-iN\theta}\, d\theta/2\pi \end{aligned} \tag{2.21}$$

where now $\gamma_n = \langle n|\gamma\rangle \equiv 1$ because all ψ_n are real. The matrix elements of E_- in the N-representation are, by (2.13),

$$\langle n|E_-|m\rangle = \int_{-\pi}^{\pi} e^{i(n-m+1)\theta}\, d\theta/2\pi = \delta_{n,m-1}$$

so that

$$E_-|m\rangle = |m-1\rangle \tag{2.22}$$

unless m is the least eigenvalue m' of the spectrum of N, in which case

$$E_-|m'\rangle = 0 \tag{2.23}$$

The conjugate operator

$$E_+ = \int_{-\pi}^{\pi} e^{-i\theta}\, d\Pi_m(\theta) \tag{2.24}$$

similarly yields

$$E_+|m\rangle = |m+1\rangle, \qquad m \neq m''$$
$$E_+|m''\rangle = 0 \tag{2.25}$$

where m'' is the largest eigenvalue of N. The operators C and S are then defined as usually by

$$C = \tfrac{1}{2}(E_+ + E_-), \qquad S = \frac{i}{2}(E_+ - E_-)$$

These operators have been further studied by Zak (1969), Lerner *et al.* (1970), Aharonov *et al.* (1973), and Volkin (1973). Perlman & Troup (1969) formulated the cosine and sine operators for angle variables in a manner similar to ours, but without interpreting them in terms of estimation. The optimum o.v.m. for estimating the parameter θ is equivalent to the generator of the C and S operators only when the basic state $|\psi\rangle$ is real. Otherwise the generator of C and S will yield estimates $\hat{\theta}$ whose p.d.f.

$$q'(\hat{\theta}\,|\,\theta) = (2\pi)^{-1}\Big|\sum_n \psi_n\, e^{-i(\hat{\theta}-\theta)}\Big|^2 \tag{2.26}$$

is broader than need be.

Suppose for example that the phase of a coherent oscillation is to be estimated, the basic state of the system being a coherent state $|\mu\rangle$ of the kind described by Glauber (1963). We can take the amplitude μ as real; $\bar{N} = \mu^2$ is the mean number of photons in the oscillation. In the number representation, with $N = a^+a$,

$$\psi_n = \langle n\,|\,\mu\rangle = (n!)^{-1/2}\mu^n \exp(-\tfrac{1}{2}\mu^2), \qquad n \geqslant 0$$

and the minimum average cost is, from (2.16),

$$\bar{C}_{\min} = 2\left[1 - \bar{N}^{1/2} e^{-\bar{N}} \sum_{n=0}^{\infty} \frac{\bar{N}^n}{n!}(n+1)^{-1/2}\right]$$

By using an asymptotic formula given by Carruthers & Nieto (1968), we find that for $\bar{N} \gg 1$

$$\bar{C}_{\min} = (4\bar{N})^{-1}\,[1 + O(\bar{N}^{-1})]$$

and as in this limit $\bar{C}_{\min}$ is asymptotically equal to the mean-square error,

$$\langle(\hat{\theta} - \theta)^2\rangle \sim (4\bar{N})^{-1}, \qquad \bar{N} \gg 1$$

Because the probabilities $P_n = |\psi_n|^2$ have the Poisson distribution, $\Delta N^2 = \bar{N}$, and

$$\Delta N^2 \langle(\hat{\theta} - \theta)^2\rangle \sim \tfrac{1}{4}, \qquad \bar{N} \gg 1 \tag{2.27}$$

Thus the mean-square error in an estimate $\hat{\theta}$ of the phase of the coherent oscillation can be as small as $(4\Delta N^2)^{-1}$ when the mean quantum number $\bar{N}$ is large.

3. *Estimation of Position*

A particle situated somewhere on a segment $-\frac{1}{2}L \leqslant x \leqslant \frac{1}{2}L$ of the x-axis is in a state

$$|\psi(x)\rangle = e^{i\mathscr{P}x/\hbar}|\psi\rangle \tag{3.1}$$

where $\mathscr{P}$ is the momentum operator and $|\psi\rangle$ is the state of the particle when its wavepacket is centered at $x = 0$. The parameter x of the density operator

$$\rho(x) = e^{i\mathscr{P}x/\hbar}|\psi\rangle\langle\psi|e^{-i\mathscr{P}x/\hbar}$$

is to be estimated. When we impose periodic boundary conditions at the ends of the segment, supposing it periodically repeated, the momentum operator $\mathscr{P}$ acquires eigenvalues

$$p_n = nh/L, \qquad h = 2\pi\hbar \tag{3.2}$$

for all integers n in $(-\infty, \infty)$. We identify the parameter θ of the previous section with $2\pi x/L$ and the generator N with $L\mathscr{P}/h$. The probability attached to the nth eigenvalue p_n is now

$$P_n = |\langle\langle p_n|\psi\rangle|^2 \tag{3.3}$$

where $|p_n\rangle\rangle$ is a normalised eigenvector of the operator $\mathscr{P}$. This operator induces a resolution of the identity

$$\sum_n |p_n\rangle\rangle\langle\langle p_n| = \mathbf{1} \tag{3.4}$$

of the Hilbert space $\mathscr{H}_S$.

Intending to pass to the limit $L \to \infty$, we define new eigenvectors $|p_n\rangle$ by

$$|p_n\rangle = (L/h)^{1/2}|p_n\rangle\rangle \tag{3.5}$$

Because the eigenvalues p_n are spaced by h/L, sums over n go in this limit into integrals as

$$\frac{h}{L}\sum_n \to \int dp$$

Thus the resolution of the identity in (3.4) becomes

$$\int_{-\infty}^{\infty} |p\rangle\langle p|\, dp = \mathbf{1}$$

From (2.14) the conditional p.d.f. of the maximum-likelihood estimate $\hat{x}$ of the position of the wavepacket is $q(\hat{x}\,|\,x) = \bar{q}(\hat{x} - x)$, where with $y = \hat{x} - x$

$$\bar{q}(y) = L^{-1}\Big|\sum_n |\langle\langle p_n|\psi\rangle| \exp(-ip_n y/\hbar)\Big|^2$$

$$\to h^{-1}\Big|\int_{-\infty}^{\infty} |\langle p|\psi\rangle| e^{-ipy/\hbar}\, dp\Big|^2 \tag{3.6}$$

in the limit $L \to \infty$. When the components $\langle p|\psi\rangle$ of the basic state $|\psi\rangle$ in the momentum representation are, within a common insignificant phase factor $e^{i\eta}$, all real and non-negative, the posterior p.d.f. of the estimate $\hat{x}$ becomes

$$\bar{q}(y) = |\langle y|\psi\rangle|^2, \qquad y = \hat{x} - x \tag{3.7}$$

where $|y\rangle$ is an eigenstate of the position operator $\mathcal{Q}$,

$$\mathcal{Q}|y\rangle = y|y\rangle, \qquad \langle p|y\rangle = h^{-1/2}\exp(-ipy/\hbar) \tag{3.8}$$

The best estimate of the position of the wavepacket is then made by an apparatus applying the orthogonal o.v.m. generated by

$$\Pi(d\hat{x}) = |\hat{x}\rangle\langle\hat{x}|\, d\hat{x}$$

$$\mathcal{Q}|\hat{x}\rangle = \hat{x}|\hat{x}\rangle$$

that is, in the conventional sense of the term it 'measures' the position operator $\mathcal{Q}$.

That the optimum estimator of the location of a system should, for a large class of states, involve the position operator $\mathcal{Q}$ is not surprising, but it is instructive to see how $\mathcal{Q}$ arises from estimation theory as the best estimator of a parameter of the state of the system, rather than as a dynamical variable. The outcome of 'measuring' $\mathcal{Q}$ gives not the exact position of a system such as a quantum particle, but an estimate of the location of its wavepacket.

If the momentum wave-function $\langle p|\psi\rangle$ is not real, the same conditional p.d.f. $q(\hat{x}\,|\,x)$ of the estimate of position as in (3.6) can be attained, but not

by measuring the operator $\mathcal{Q}$; the o.v.m. whose matrix elements in the momentum representation are

$$\langle p' \mid \Pi(d\hat{x}) \mid p'' \rangle = \frac{\langle p' \mid \psi \rangle \langle \psi \mid p'' \rangle}{h \mid \langle p' \mid \psi \rangle \langle \psi \mid p'' \rangle \mid} \exp\left[i(p' - p'')\hat{x}/\hbar\right] d\hat{x}$$

must be applied instead. It too is orthogonal and corresponds to measuring the operator $\mathcal{Q}'$ whose momentum representation is

$$\langle p' \mid \mathcal{Q}' \mid p'' \rangle = \frac{\hbar}{i} \frac{\langle p' \mid \psi \rangle \langle \psi \mid p'' \rangle}{\mid \langle p' \mid \psi \rangle \langle \psi \mid p'' \rangle \mid} \delta'(p' - p'')$$

where $\delta'(x)$ is the derivative of the delta-function. As in general, the optimum o.v.m. is adapted to the class of states anticipated.

4. *The Energy-Time Uncertainty Relation*

The epoch τ of a particular event in the temporal evolution of a system S can be regarded as a parameter of its density operator

$$\rho(\tau) = \mathrm{e}^{-iH\tau/\hbar} \rho_0 \, \mathrm{e}^{iH\tau/\hbar} \tag{4.1}$$

in which H is the Hamiltonian operator of the system, assumed independent of time. The event might be the time at which a particle crosses a given plane or leaves a certain region of space. The difficulties of constructing an operator corresponding to an arrival time τ and thence, by (1.1), finding an uncertainty relation between time and energy have been analysed by Allcock (1969). Nevertheless, Wigner (1972) has derived a time-energy uncertainty relation in terms of the continuous spectrum of the energy of the system, and Ekstein & Siegert (1971) have developed one for the time a particle remains within a given volume.

An observation to fix the epoch τ must start at some instant t_0 and end at a later time $t_0 + T$. One might reasonably assume a uniform prior p.d.f. $z(\tau) \equiv T^{-1}$ for τ, adopt a suitable cost $C(\hat{\tau}, \tau)$ of errors in estimating τ, and attempt to solve the equations of Section 1 to find the best o.v.m. The o.v.m. would be applied at the end of the interval $(t_0, t_0 + T)$ to determine at what time within it the event occurred. Only if the motion of the system S is periodic with period T, however, can the technique of Section 2 be utilised. A periodic system is exemplified by a lossless transmission line of length L containing a coherent signal propagated along it at velocity c, $L = cT$. Estimating the epoch τ is equivalent to estimating the phase θ of the carrier of the signal, and if the signal occupies only a narrow band of frequencies, the analysis in Section 2 applies. Periodic systems form a limited class, however, and for an aperiodic motion that method is inapplicable because the integrations over $0 < \theta < T$ do not reduce the optimisation equations to the simple forms in (2.9)–(2.11).

A lower bound to the variance $\mathrm{Var}\,\hat{\tau}$ of an unbiased estimate of a parameter such as the epoch τ can be established by means of the quantum-mechanical form of the Cramér–Rao inequality (Helstrom, 1973). If $d\Pi(\hat{\tau})$ generates the

o.v.m. applied to the system S by an apparatus for estimating τ, the conditional expected value of the estimate $\hat{\tau}$ is, with $t_0 = 0$,

$$\mathbf{E}(\hat{\tau}|\tau) = \mathrm{Tr} \int_0^T \hat{\tau}\rho(\tau)d\Pi(\hat{\tau})$$

and the estimate is unbiased if $\mathbf{E}(\hat{\tau}|\tau) = \tau$. Then

$$\begin{aligned} \mathrm{Var}\ \hat{\tau} &= \mathbf{E}[(\hat{\tau} - \tau)^2|\tau] \\ &\geqslant (\mathrm{Tr}\ \rho L^2)^{-1} = \{\mathrm{Tr}[L(\partial\rho/\partial\tau)]\}^{-1} \end{aligned} \tag{4.2}$$

where L is the symmetrised logarithmic derivative (s.l.d.) operator defined by

$$\frac{\partial\rho}{\partial\tau} = \tfrac{1}{2}(L\rho + \rho L) \tag{4.3}$$

and evaluated at the true value of the epoch τ.

If the system is in a pure state,

$$\begin{aligned} |\psi_\tau\rangle &= \mathrm{e}^{-iH\tau/\hbar}|\psi_0\rangle \\ \rho_0 &= |\psi_0\rangle\langle\psi_0| \end{aligned} \tag{4.4}$$

equation (4.3) can be solved by introducing a complete orthonormal set of states $|\psi_n\rangle$, of which $|\psi_\tau\rangle = |\psi_1\rangle$ is the first member,

$$\langle\psi_n|\psi_m\rangle = \delta_{nm}$$

By Schrödinger's equation, which is equivalent to (4.1) and (4.4)

$$\frac{\partial\rho}{\partial\tau} = \frac{i}{\hbar}(\rho H - H\rho) \tag{4.5}$$

and

$$\langle\psi_n|\frac{\partial\rho}{\partial\tau}|\psi_m\rangle = \frac{i}{\hbar}\langle\psi_1|H|\psi_m\rangle\delta_{n1} - \frac{i}{\hbar}\langle\psi_n|H|\psi_1\rangle\delta_{1m} \tag{4.6}$$

Forming the matrix elements of (4.3), we obtain similarly

$$\langle\psi_n|\frac{\partial\rho}{\partial\tau}|\psi_m\rangle = \tfrac{1}{2}\langle\psi_1|L|\psi_m\rangle\delta_{n1} + \tfrac{1}{2}\langle\psi_n|L|\psi_1\rangle\delta_{1m} \tag{4.7}$$

Hence

$$\begin{aligned} \langle\psi_1|L|\psi_1\rangle &= 0 \\ \langle\psi_1|L|\psi_m\rangle = \langle\psi_m|L|\psi_1\rangle^* &= \frac{2i}{\hbar}\langle\psi_1|H|\psi_m\rangle, \qquad m > 1 \\ \langle\psi_n|L|\psi_m\rangle &= 0, \qquad n > 1, m > 1 \end{aligned} \tag{4.8}$$

The quantity appearing on the right-hand side of (4.2) is now

$$\begin{aligned}\mathrm{Tr}[L(\partial\rho/\partial\tau)] &= \sum_m \sum_n \langle\psi_n|\frac{\partial\rho}{\partial\tau}|\psi_m\rangle\langle\psi_m|L|\psi_n\rangle \\ &= 4\hbar^{-2} \sum_{m>1} |\langle\psi_1|H|\psi_m\rangle|^2 \end{aligned} \tag{4.9}$$

By using the completeness relation

$$\sum_{m=1}^{\infty} |\psi_m\rangle\langle\psi_m| = \mathbf{1} \tag{4.10}$$

we can write this as

$$\begin{aligned}&\mathrm{Tr}[L(\partial\rho/\partial\tau)] \\ &= 4\hbar^{-2}\left\{\sum_{m=1}^{\infty}\langle\psi_1|H|\psi_m\rangle\langle\psi_m|H|\psi_1\rangle - |\langle\psi_1|H|\psi_1\rangle|^2\right\} \\ &= 4\hbar^{-2}[\langle\psi_1|H^2|\psi_1\rangle - |\langle\psi_1|H|\psi_1\rangle|^2] \\ &= 4\hbar^{-2}\Delta E^2 \end{aligned} \tag{4.11}$$

where ΔE is the uncertainty of the energy in the state $|\psi_1\rangle = |\psi_\tau\rangle$.

Hence from (4.2) the variance of an unbiased estimate of the epoch τ is bounded below by

$$\mathrm{Var}\,\hat{\tau} \geqslant \frac{\hbar^2}{4\Delta E^2} \tag{4.12}$$

The more precisely the energy of the system is fixed during its preparation, the less accurately will it be possible to estimate the epoch of some distinctive event in its subsequent evolution.

5. *Discussion*

The uncertainty relation between two dynamical variables $\mathscr{A}$ and $\mathscr{B}$ is usually interpreted in terms of measurements of a large ensemble of systems, all prepared in the same state. On half the systems $\mathscr{A}$ is measured, on the other half $\mathscr{B}$. The results of the two sets of measurements are random variables with variances $\Delta\mathscr{A}^2$ and $\Delta\mathscr{B}^2$ defined by (1.2). Whatever the common state of the systems, the product $\Delta\mathscr{A}\Delta\mathscr{B}$ cannot fall below a certain limit set by (1.1).

The uncertainty relations arising from quantum estimation theory bear a different interpretation. A great many systems are prepared in states specified by a density operator $\rho(\boldsymbol{\theta})$ depending on some parameter or set of parameters $\boldsymbol{\theta}$, which are themselves random variables with a joint prior p.d.f. $z(\boldsymbol{\theta})$. The values $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ are known in each case to the preparer of the system, but not to the observer. All the systems are observed by an apparatus A_m, possibly containing a computer, which produces a set of m numbers $\hat{\theta}_1, \ldots, \hat{\theta}_m$ as its estimates $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. As random variables they have a conditional probability density function $q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta})$. The differences $\hat{\theta}_i - \theta_i$ are errors whose seriousness is assessed by some cost function $C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$. The theory seeks a lower

bound to the average value of that cost over the entire ensemble of systems with its variety of true values of the parameters $\boldsymbol{\theta}$.

We have seen how in the estimation of a single displacement parameter θ of a pure state, the conditional p.d.f. $q(\hat{\theta}\,|\,\theta)$ of its maximum-likelihood estimate $\hat{\theta}$ is the broader, the narrower the distribution $P_n = \langle\psi\,|\,G_n\,|\,\psi\rangle$ of probability among the eigenstates of the infinitesimal generator N of the displacement group. Except in special cases the observing apparatus A_m does not measure an Hermitian operator corresponding in the conventional sense to the displacement θ; in general there exists no such operator in whose spectrum of eigenvalues the outcomes $\hat{\theta}$ are constrained to lie.

There are two reasons for observing a quantum system. One is to verify that an apparatus A_p prepares systems in the manner predicted. To this end one or more kinds of observational apparatus apply their defining operator-valued measures to an ensemble of systems prepared by A_p, and one checks whether the outcomes are described by the correct probability distributions.

The other reason for observing a system is to estimate certain parameters $\theta_1, \theta_2, \ldots, \theta_m$ of its state, or—as in quantum hypothesis testing—to choose the one of a set of possible density operators $\rho_1, \rho_2, \ldots, \rho_M$ that best represents the state of the system. The class of density operators $\rho(\boldsymbol{\theta})$ to which the observational apparatus must be adapted is fixed by what is known about how the quantum system has been prepared. By taking this viewpoint one avoids speciously identifying the outcome of a measurement as the value 'possessed' by a dynamical variable of the system before its interaction with the observing apparatus. The outcome is instead accepted as only an estimate of some parameter of the state of the system, and the way is opened to handling parameters such as angles, phases, and epochs that have no apparent representation as quantum-mechanical operators. A previous article (Helstrom, 1974) showed how the conceptual difficulties associated with simultaneous measurement of non-commuting observables can be circumvented by treating quantum measurement as parameter estimation.

Just as it is unknown how in general to construct an apparatus to measure a dynamical variable represented by an arbitrary Hermitian operator, so is it unknown how to build one to apply a given operator-valued measure to a quantum system. A future theory may delimit in some way the class of admissible o.v.m.'s. Quantum estimation theory can be regarded as furnishing lower bounds to the average costs of errors in parameter estimates, but whether those bounds can always be attained by real physical apparatus is an open question.

Appendix

In order to verify that the solution in (2.12)-(2.13) satisfies the reduced optimisation equations (2.10)-(2.11), we first calculate the elements of the matrix $\rho_0\xi_0$ in the N-representation,

$$\begin{aligned}\langle n\,|\,\rho_0\xi_0\,|\,m\rangle &= \langle n\,|\,\psi\rangle\langle\psi\,|\,\gamma\rangle\langle\gamma\,|\,m\rangle/2\pi \\ &= (2\pi)^{-1}\psi_n\gamma_m^* \sum_k \psi_k^*\gamma_k = K\psi_n\gamma_m^* \end{aligned} \tag{A.1}$$

where by (2.13)

$$K = (2\pi)^{-1} \sum_k |\psi_k| \tag{A.2}$$

Hence by (2.9) the Lagrange operator Υ_m is represented by the matrix

$$\langle n|\Upsilon_m|m\rangle = K|\psi_n|\delta_{nm} \tag{A.3}$$

and its diagonal elements are real, as required.

For (2.10) the (nm) matrix element is, by (A.1), (A.3), and (2.13),

$$\begin{aligned}(2\pi)^{-1}K\psi_n\gamma_m^* - (2\pi)^{-1}\langle n|\Upsilon_m|n\rangle\langle n|\gamma\rangle\langle\gamma|m\rangle \\ = (2\pi)^{-1}(K\psi_n\gamma_m^* - K|\psi_n|\gamma_n\gamma_m^*) = 0\end{aligned}$$

This verifies (2.10).

In order to show that the operator on the left-hand side of (2.11) is non-negative-definite, we form for an arbitrary state $|\zeta\rangle$, using (2.13),

$$\begin{aligned}\langle\zeta|\rho_0|\zeta\rangle &= |\langle\zeta|\psi\rangle|^2 = |\sum_k \zeta_k^*\psi_k|^2 \\ &= |\sum_k |\psi_k|^{1/2}\gamma_k\zeta_k^*|\psi_k|^{1/2}|^2 \leqslant \sum_k |\psi_k| \sum_m |\zeta_m|^2|\psi_m| \\ &= 2\pi\langle\zeta|\Upsilon_m|\zeta\rangle\end{aligned}$$

by the Schwarz inequality and (A.3). Hence

$$\langle\zeta|[\Upsilon_m - (2\pi)^{-1}\rho_0]|\zeta\rangle \geqslant 0$$

as required for (2.11).

Turning now to the minimisation of the average cost when assessed by the function in (2.3), we calculate the operator $W(\hat{\theta})$ from (1.18). Because of (2.1) and (2.2) and the fact that the cost function depends only on $\hat{\theta} - \theta$,

$$W(\hat{\theta}) = e^{iN\hat{\theta}}W_0e^{-iN\hat{\theta}} \tag{A.4}$$

where in the N-representation

$$\langle n|W_0|m\rangle = 2|\psi_n|^2\delta_{nm} - \psi_n\psi_{n-1}^*\delta_{n,m+1} - \psi_n\psi_{n+1}^*\delta_{n,m-1} \tag{A.5}$$

as is easily obtained by writing for (2.3)

$$C(\hat{\theta}, \theta) = 2 - e^{i(\hat{\theta}-\theta)} - e^{-i(\hat{\theta}-\theta)} \tag{A.6}$$

As for the maximum-likelihood estimator, the Lagrange operator Υ has a diagonal matrix

$$\langle n|\Upsilon|m\rangle = 2\pi\langle n|W_0\xi_0|m\rangle\delta_{nm}$$

by (1.20), (2.4), (2.5), (2.12), and (2.13), where

$$\begin{aligned}2\pi\langle n|W_0\xi_0|m\rangle &= \sum_k \langle n|W_0|k\rangle\langle k|\gamma\rangle\langle\gamma|m\rangle \\ &= (2|\psi_n|^2\gamma_n - \psi_n\psi_{n-1}^*\gamma_{n-1} - \psi_n\psi_{n+1}^*\gamma_{n+1})\gamma_m^* \\ &= (2|\psi_n|^2\gamma_n - \psi_n|\psi_{n-1}| - \psi_n|\psi_{n+1}|)\gamma_m^*\end{aligned} \tag{A.7}$$

Hence

$$\langle n|\Upsilon|m\rangle = (2|\psi_n|^2 - |\psi_n\psi_{n-1}| - |\psi_n\psi_{n+1}|)\delta_{nm} \tag{A.8}$$

and Υ is Hermitian as required.

The optimisation equations (1.21)-(1.22) now take the reduced form

$$W_0 - \Upsilon \geqslant \mathbf{0} \tag{A.9}$$

$$(W_0 - \Upsilon)\xi_0 = \mathbf{0} \tag{A.10}$$

because Υ commutes with $e^{iN\theta}$. From (A.7), (A.8), and (2.13) we obtain

$$\begin{aligned} 2\pi\langle n|\Upsilon\xi_0|m\rangle &= (2|\psi_n|^2 - |\psi_n\psi_{n+1}| - |\psi_n\psi_{n-1}|)\gamma_n\gamma_m^* \\ &= (2|\psi_n|^2\gamma_n - \psi_n|\psi_{n-1}| - \psi_n|\psi_{n+1}|)\gamma_m^* \\ &= 2\pi\langle n|W_0\xi_0|m\rangle \end{aligned}$$

verifying (A.10).

The matrix representation of $W_0 - \Upsilon$ has, by (A.5) and (A.8), real elements on the diagonal, and the elements on each side of the diagonal are

$$\begin{aligned} \langle n|(W_0 - \Upsilon)|n-1\rangle &= -\psi_n\psi_{n-1}^* \\ \langle n-1|(W_0 - \Upsilon)|n\rangle &= -\psi_{n-1}\psi_n^* \end{aligned} \tag{A.11}$$

All the rest of the elements of this matrix are zero. In order to calculate its eigenvalues and thus check its non-negative-definiteness, we could work out the determinant,

$$\det|\lambda\delta_{nm} - \langle n|(W_0 - \Upsilon)|m\rangle|$$

Upon expanding it, we should find the off-diagonal elements in (A.11) occurring only together as a product, which equals $|\psi_n\psi_{n-1}|^2$. Their phases disappear and have no bearing on whether $(W_0 - \Upsilon)$ is non-negative-definite. The question of definiteness can therefore be decided for the new matrix $\langle n|(W_0' - \Upsilon)|m\rangle$ obtained by replacing all ψ_n's in W_0 by their absolute values $|\psi_n|$. We now find for an arbitrary state $|\zeta\rangle$, after some calculation,

$$\langle\zeta|(W_0' - \Upsilon)|\zeta\rangle = \sum_n |\psi_n\psi_{n-1}|\,|\zeta_{n-1} - \zeta_n|^2 \geqslant 0$$

as required for (A.9). From (A.8), equation (2.16) for the minimum cost $\bar{C}_{\min}$ follows easily.

When the eigenvalues of the infinitesimal generator N are degenerate, the eigenstates common to a single eigenvalue n span a linear subspace $\mathscr{G}_n$ of $\mathscr{H}_S$. Let G_n be the projector on to $\mathscr{G}_n$, and let the dimension of $\mathscr{G}_n$ be ν_n. An orthonormal basis is set up in $\mathscr{G}_n$ by starting with the vector

$$\begin{aligned} |\eta_{n1}\rangle &= P_n^{-1/2}G_n|\psi\rangle \\ P_n &= \langle\psi|G_n|\psi\rangle \end{aligned} \tag{A.12}$$

The remaining $\nu_n - 1$ basis vectors $|\eta_{nk}\rangle$ are chosen by a Gram-Schmidt procedure from the eigenstates spanning $\mathscr{G}_n$. The array of vectors $|\eta_{nk}\rangle$ for all n

and k forms an orthonormal basis for representing arbitrary vectors and matrices in $\mathscr{H}_S$.

In this augmented N-representation the matrix of the density operator $\rho_0 = |\psi\rangle\langle\psi|$ takes on a block form; the element in the upper left-hand corner of each block $G_n\rho_0 G_m$ is equal to

$$\langle\eta_{n1}|\rho_0|\eta_{m1}\rangle = P_n^{1/2}P_m^{1/2} \tag{A.13}$$

The rest of the elements are zero. The matrix of the operator ξ_0 in (2.5) has $(2\pi)^{-1}$ on the diagonal. In the remaining blocks $G_n\xi_0 G_m$, $m \neq n$, the upper left-hand element is $(2\pi)^{-1}$ and the others are 0; thus

$$\xi_0 = (2\pi)^{-1}\left[\mathbf{1} + \sum_n \sum_{m\neq n} P_n^{-1/2}P_m^{-1/2}G_n|\psi\rangle\langle\psi|G_m\right]$$

In verifying the reduced optimisation equations (2.10), (2.11), and (A.9)-(A.10), only the upper left-hand elements of each block in the several matrices come into play, and the calculations go through just as before.

References

Aharonov, Y., Lerner, E. C., Huang, H. W. H. and Knight, J. M. (1973). *Journal of Mathematical Physics*, **14**, 746.

Allcock, G. R. (1969). *Annals of Physics*, **53**, 253.

Benioff, P. (1972a). *Journal of Mathematical Physics*, **13**, 231.

Benioff, P. (1972b). *Journal of Mathematical Physics*, **13**, 908.

Carruthers, P. and Nieto, M. M. (1968). *Reviews of Modern Physics*, **40**, 411.

Davies, E. B. and Lewis, J. T. (1970). *Communications in Mathematical Physics*, **17**, 239.

Ekstein, H. and Siegert, A. J. F. (1971). *Annals of Physics*, **68**, 509.

Glauber, R. J. (1963). *Physical Review*, **131**, 2766.

Helstrom, C. W., Liu, J. W. S. and Gordon, J. P. (1970). *Proceedings of the IEEE*, **58**, 1578.

Helstrom, C. W. (1973). *International Journal of Theoretical Physics*, Vol. 8, No. 5, p. 361.

Helstrom, C. W. (1974). *Foundations of Physics* (in press).

Holevo, A. S. (1973). *Journal of Multivariate Analysis*, **3**, 337.

Judge, D (1964). *Nuovo Cimento*, **31**, 332.

Lerner, E. C., Huang, H. W. and Walters, G. E. (1970). *Journal of Mathematical Physics*, **11**, 1679.

Naĭmark, M. A. (1940). *Izvestia Akademiĭ Nauk, USSR, math. series*, **4**, 277.

Perlman, H. S. and Troup, G. J. (1969). *American Journal of Physics*, **37**, 1060.

Robertson, H. P. (1929). *Physical Review*, **34**, 163.

Volkin, H. C. (1973). *Journal of Mathematical Physics*, **14**, 1965.

Wigner, E. P. (1972). *Aspects of Quantum Theory*, Chap. 14, p. 237 (Eds. A. S. Salam and E. P. Wigner). Cambridge.

Zak, J. (1969). *Physical Review*, **187**, 1803.