COMMUNICATION SCIENCES
AND
ENGINEERING

VII. PROCESSING AND TRANSMISSION OF INFORMATION

Academic Research Staff

Prof. P. Elias	Prof. R. S. Kennedy
Prof. R. G. Gallager	Prof. C. E. Shannon

Graduate Students

Donna Jean Brown	J. G. Himes	L. H. Ozarow
V. Chan	L. S. Metzger	L. Rapisarda
S. J. Dolinar, Jr.	A. Z. Muszynsky	S. R. Robinson
R. A. Flower	N-S. Myung	M. A. Sirbu, Jr.
D. A. Garbin	F. Nourani	V. Vilnrotter
L. M. Hawthorne		K. Yamada

A. CONSTRAINTS IMPOSED BY THE LAW OF CONSERVATION OF ENERGY ON POSSIBLE FORMS OF INTERACTION HAMILTONIANS

National Aeronautics and Space Administration (Grant NGL 22-009-013)

V. Chan

1. Introduction

A popular theory of measurement of quantum systems says that if we describe states of a quantum system S as elements in a Hilbert space $\mathcal{H}_S$, then any measurement can be characterized by a Hermitian operator in Hilbert space $\mathcal{H}_S$. This notion of quantum measurement is too restricted, and we shall consider measurements characterized by Hermitian operators in extended Hilbert spaces that include the original space $\mathcal{H}_S$. One possible way of implementing such measurement is to let an apparatus A interact with the system S and then perform subsequent measurement on the combined system S+A.[1-3] Therefore we want to know what types of interaction are feasible and, in particular, to find restrictions on possible types of interaction when we invoke some form of conservation law. We know that we can characterize interactions between two quantum systems by specifying[1, 2] the interaction Hamiltonian H_I. In this report we discuss the restriction of the law of conservation of energy on the allowable form of H_I.

2. States of S+A Described by Pure States

Let us assume that before a certain contact time t_c the two systems S and A are noninteracting and evolve independently according to their individual free Hamiltonians, $\mathcal{H}_s$ and $\mathcal{H}_A$, respectively. The Hamiltonian for the combined system S+A before t_c is then $H = H_s \otimes I_A + I_s \otimes H_A$, where I_A and I_s are the identity operators in $\mathcal{H}_A$ and $\mathcal{H}_s$, respectively. Let $\mathcal{M}_{S+A}$ be a complete linear subspace of $\mathcal{H}_s \otimes \mathcal{H}_A$ such that for $t > t_c$ the states of S+A described by vectors in $\mathcal{M}_{S+A}$ are noninteracting. Let $|a^o+s^o\rangle\rangle \in \mathcal{M}_{S+A}$ be the Schrödinger state of S+A at t_c. The energy of the combined system at this point is

$$E^{o}_{S+A} = \langle\langle s^{o}+a^{o}|H|a^{o}+s^{o}\rangle\rangle.$$

For $t > t_c$, $|a^{t}+s^{t}\rangle\rangle = U_t|a^{o}+s^{o}\rangle\rangle$, where $U_t \equiv \exp\{-\frac{1}{\hbar} H'(t-t_c)\}$. Then for $t > t_c$,

$$E^{t}_{S+A} = \langle\langle s^{t}+a^{t}|H'|a^{t}+s^{t}\rangle\rangle$$

$$= \langle\langle s^{o}+a^{o}|U_t^{\dagger}H'U_t|a^{o}+s^{o}\rangle\rangle.$$

Since H' is the generator of the unity group U_t, it commutes with U_t; that is, the commutator $[H', U_t] = 0$. Therefore, for $t > t_c$,

$$E^{t}_{S+A} = \langle\langle s^{o}+a^{o}|H'|a^{o}+s^{o}\rangle\rangle$$

$$= \langle\langle s^{o}+a^{o}|H|a^{o}+s^{o}\rangle\rangle + \langle\langle s^{o}+a^{o}|H_I|a^{o}+s^{o}\rangle\rangle$$

$$= E^{o}_{S+A} + \langle\langle s^{o}+a^{o}|H_I|a^{o}+s^{o}\rangle\rangle.$$

The law of conservation of energy requires

$$E^{t}_{S+A} = E^{o}_{S+A}; \quad \text{for all } t.$$

Hence this implies (*) $\langle\langle s^{o}+a^{o}|H_I|a^{o}+s^{o}\rangle\rangle = 0$.

A sufficient (but not necessary) condition for H_I to satisfy the constraint (*) is to have $|a^{o}+s^{o}\rangle\rangle$ in the null space ($\mathcal{N}_{H_I}$) of H_I (condition 1a). Or equivalently if we have already specified $\mathcal{M}_{S+A}$, we require H_I to be representable completely by vectors in $\mathcal{M}^{\perp}_{S+A}$, where $\mathcal{M}^{\perp}_{S+A} \equiv \mathcal{H}_s \otimes \mathcal{H}_A - \mathcal{M}_{S+A}$ is the orthogonal space of $\mathcal{M}_{S+A}$ (condition 1b).

A necessary and sufficient condition can be found. Let $H_I\mathcal{M}_{S+A} \equiv \{\text{all } x \in \mathcal{H}_s \otimes \mathcal{H}_A: x = H_I y, \text{ some } y \in \mathcal{M}_{S+A}\}$. Let $\mathcal{M}^{H_I}_{S+A}$ be the completion of this space. In other words, $\mathcal{M}^{H_I}_{S+A}$ is the range space of H_I with domain restricted to $\mathcal{M}_{S+A}$. Then a necessary and sufficient condition for (*) to hold is $\mathcal{M}^{H_I}_{S+A} \subset \mathcal{M}^{\perp}_{S+A}$ (condition 2).

3. States of S+A Described by Density Operators

Let the state of S+A at t_{c_-} be described by the density operator ρ^{o}_{S+A}. Requiring no interaction between S and A before t_c means that ρ^{o}_{S+A} can be represented completely by vectors in $\mathcal{M}_{S+A}$. For $t > t_c$ the density of the combined system is

$$\rho^{t}_{S+A} = U_t \rho^{o}_{S+A} U^{\dagger}_t .$$

The mean energy for $t > t_c$ is

$$\begin{aligned} E^{t}_{S+A} &= \operatorname{Tr}\left\{\rho^{t}_{S+A} H'\right\} \\ &= \operatorname{Tr}\left\{U_t \rho^{o}_{S+A} U^{\dagger}_t H'\right\}, \end{aligned}$$

where Tr denotes the trace.

Since $[H', U_t] = 0$ and $U^{\dagger}_t$ commute with H',

$$E^{t}_{S+A} = \operatorname{Tr}\left\{U_t \rho^{o}_{S+A} H' U^{\dagger}_t\right\}.$$

Since unitary transformations do not change the trace of operators,

$$\begin{aligned} E^{t}_{S+A} &= \operatorname{Tr}\left\{\rho^{o}_{S+A} H'\right\} \\ &= \operatorname{Tr}\left\{\rho^{o}_{S+A} H\right\} + \operatorname{Tr}\left\{\rho^{o}_{S+A} H_I\right\} \\ &= E^{o}_{S+A} + \operatorname{Tr}\left\{\rho^{o}_{S+A} H_I\right\}. \end{aligned}$$

Conservation of energy requires that, for all t, $E^{t}_{S+A} = E^{o}_{S+A}$, which implies

$$(**) \quad \operatorname{Tr}\left\{\rho^{o}_{S+A} H_I\right\} = 0.$$

Let $\mathcal{H}_o$ be the Hilbert space of bounded Hermitian operators defined on $\mathcal{H}_S \otimes \mathcal{H}_A$ and the inner product in $\mathcal{H}_o$ be defined as $(A, B) = \operatorname{Tr}\{A, B\}$ for all $A, B \in \mathcal{H}_o$. Then a necessary and sufficient condition to satisfy (**) is

$$\rho^{o}_{S+A} \in \mathcal{H}^{\perp}_{H_I} = \mathcal{H}_o - \{H_I\} \qquad \text{(condition 3a)},$$

where $\{H_I\}$ is the subspace generated by H_I, and $\mathcal{H}^{\perp}_{H_I}$ is the subspace orthogonal to it.

Furthermore, if we have already specified the possible choices of ρ^o_{S+A} by requiring that they be representable by vectors in $\mathcal{M}_{S+A}$ and denoted the subspace of $\mathcal{H}_o$ generated by these sets of possible density operators by $\mathcal{H}_{\rho_{S+A};\mathcal{M}_{S+A}}$, then an equivalent condition of (3a) is

$$H_I \in \mathcal{H}^{\perp}_{\rho_{S+A};\mathcal{M}_{S+A}} \equiv \mathcal{H}_o - \mathcal{H}_{\rho_{S+A};\mathcal{M}_{S+A}} \qquad \text{(condition 3b).}$$

A more illuminating sufficient (but not necessary) condition is easily found by noting that if we require H_I to be representable by vectors in $\mathcal{M}^{\perp}_{S+A}$ the condition (**) is always satisfied (condition 4).

References

1. V. Chan, "Interaction Formulation of Quantum Communication Theory," Quarterly Progress Report No. 106, Research Laboratory of Electronics, M.I.T., July 15, 1972, pp. 128-132.
2. V. Chan, "Sequential Detection of Signals Transmitted by a Quantum System (Equiprobable Binary Pure State)," Quarterly Progress Report No. 107, Research Laboratory of Electronics, M.I.T., October 15, 1972, pp. 89-93.
3. V. Chan, "Statistically Dependent Quantum Systems: The Paradox of Einstein, Podolsky, and Rosen," Quarterly Progress Report No. 108, Research Laboratory of Electronics, M.I.T., January 15, 1973, pp. 225-229.

B. REALIZATION OF AN OPTIMUM QUANTUM MEASUREMENT BY EXTENSION OF HILBERT-SPACE TECHNIQUE

National Aeronautics and Space Administration (Grant NGL 22-009-013)

V. Chan

A system S is in one of M equiprobable pure states $\{|f_i\rangle\}_{i=1}^{M}$ and these states are linearly dependent with certain symmetry such that $\sum_{i=1}^{M} b|f_i\rangle\langle f_i| = I_s$, $b > 0$, where I_s is the identity operator in Hilbert space $\mathcal{H}_s$ that describes the system S. To maximize the probability of correct detection, we want to observe S and determine in which of M states it is. It is known that there is a solution to this problem.[1]

The measurement operators are Hermitian and positive definite: $Q_i = b|f_i\rangle\langle f_i|$, $i = 1, \ldots, M$. These Q_i are not orthogonal in general, however, and do not correspond to any measurement on S alone that can be described by a Hermitian operator in $\mathcal{H}_s$. Helstrom and Kennedy[2] have proposed to synthesize these types of measurement by bringing into consideration another quantum system A, called the apparatus, described by Hilbert space $\mathcal{H}_A$ so that there exists a set of orthogonal positive definite Hermitian measurement operators $\{\pi_i\}_{i=1}^{M}$ in the tensor product Hilbert space $\mathcal{H}_s \otimes \mathcal{H}_A$ and a density operator ρ_A for A with $Q_i = \mathrm{Tr}_A\{\pi_i\rho_A\}$, where Tr_A indicates taking a partial trace over $\mathcal{H}_A$. Now the π_i correspond to a Hermitian measurement on $\mathcal{H}_s \otimes \mathcal{H}_A$. We have solved this problem for a particular case with M = 3.

If the possible states of S are the three shown in Fig. VII-1, it can be shown[1] that $\sum_{i=1}^{3} \frac{2}{3} |s_i\rangle\langle s_i| = I_s$, so that $Q_i = \frac{2}{3} |s_i\rangle\langle s_i|$, i = 1, ..., 3, and the probability of correct detection is $P[C] = \frac{1}{3} \sum_i \mathrm{Tr}\,[\rho_i Q_i] = \frac{2}{3}$.

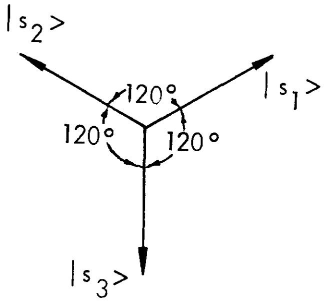


Fig. VII-1. Possible states of S.

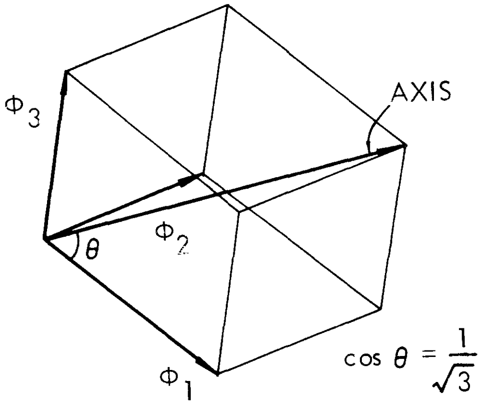


Fig. VII-2. Configurations of $\pi_i' = |\phi_i\rangle\langle\phi_i|$.

Pick any apparatus A described by Hilbert space $\mathcal{H}_A$ of dimension $N \geq 2$ (hence the dimension of $\mathcal{H}_s \otimes \mathcal{H}_A \geq 4$). Let $\rho_A = |a\rangle\langle a|$, where $|a\rangle$ is any pure state. Therefore the three possible joint states of S+A are $\{|s_i\rangle|a\rangle\}_{i=1}^{3}$, and again they span a two-dimensional subspace in $\mathcal{H}_s \otimes \mathcal{H}_A$, namely $\mathcal{H}_s \otimes \mathcal{M}_{|a\rangle}$, where $\mathcal{M}_{|a\rangle}$ is the subspace generated by $|a\rangle$. Choose any other one-dimensional subspace $\mathcal{M}_{S+A}$ in $\mathcal{H}_s \otimes \mathcal{M}^{\perp}_{|a\rangle}$, where $\mathcal{M}^{\perp}_{|a\rangle}$ is the orthogonal subspace of $\mathcal{M}_{|a\rangle}$. Then three orthogonal measurement operators $\{\pi_i'\}_{i=1}^{3}$ can be found in $\mathcal{H}_s \otimes \mathcal{M}_{|a\rangle} \cup \mathcal{M}_{S+A}$ to satisfy our requirements.

We shall show that $\pi_i' = |\phi_i\rangle\langle\phi_i|$, where the $|\phi_i\rangle$ are orthonormal vectors. By symmetry considerations, it is clear that we want the axis of the coordinate system made up of $|\phi_1\rangle$, $|\phi_2\rangle$, $|\phi_3\rangle$ to be perpendicular to the plane spanned by the $|s_i\rangle$, and the projections of the $|\phi_i\rangle$ on the plane of the $|s_i\rangle$ along the axis should coincide with these respective $|s_i\rangle$, so that $|\langle\phi_i|s_i\rangle|$ = a constant for all i is maximized (see

Fig. VII-2). By straightforward geometric calculations $|\langle \phi_i | s_i \rangle|^2 = \frac{2}{3}$ so that $P[C] = \frac{1}{3} \sum_{i=1}^{3} \mathrm{Tr}\{\rho_i \pi_i'\} = \frac{2}{3}$. Hence, the $|\phi_i\rangle$ are indeed optimum and if we wish to require in addition that the sum of the measurement operators equals the identity operator in $\mathcal{H}_s \otimes \mathcal{H}_A$, we need only define $\pi_i = \pi_i' \otimes I_d$, where I_d is the identity operator of $\mathcal{H}_S \otimes \mathcal{H}_A - \{\mathcal{H}_s \otimes \mathcal{M}_{|a\rangle} \cup \mathcal{M}_{S+A}\}$, then $\sum_{i=1}^{3} \pi_i = I_{S+A}$.

References

1. A. S. Holevo, "Statistical Problems in Quantum Physics," unpublished manuscript, Steklov Mathematical Institute, Moscow, U. S. S. R., to appear in Proc. Soviet-Japanese Symposium on Probability and Statistics, 1972.
2. C. W. Helstrom and R. S. Kennedy, "Noncommuting Observables in Quantum Detection and Estimation Theory" (submitted to IEEE Trans. on Information Theory).

C. ON THE OPTIMUM QUANTUM RECEIVER FOR THE M-ARY LINEARLY INDEPENDENT PURE STATE PROBLEM

National Aeronautics and Space Administration (Grant NGL 22-009-013)

R. S. Kennedy

It has been conjectured that the optimum quantum receiver for digital communication may not always be characterized by a set of (commuting) observables on the system space, i. e., a set of commuting Hermitian operators with complete sets of eigenvectors.[1-6] Recently, it has been demonstrated that this conjecture is, in fact, true[7-9] and there has been renewed interest in delineating conditions for which the optimum receiver can be characterized by observables.

It is known that the optimum receiver for binary signaling is characterized by an observable.[10, 3, 5] It has also been shown that, among those receivers characterized by an overcomplete set of measurement states, the optimum receiver for the M-ary linearly independent pure-state problem can be characterized by an observable.[11, 2] Not all quantum measurements can be characterized by such overcomplete sets,[8,9] however, and the characterization of the optimum receiver for the M-ary problem has remained open. In this report we show that the optimum quantum receiver for the M-ary pure state problem can, in fact, be characterized by an observable when the M pure states are linearly independent.

The problem of interest can be stated as follows. One of a set of M messages is transmitted, the i^{th} message occurring with (nonzero) a priori probability p_i. The transmission of message i causes the field at the receiver to be in the quantum state $|u_i\rangle$. That is, there is no randomness in the channel nor is there any additive noise.

It has been shown that any quantum receiver for M-ary digital communication can be characterized by a set of M nonnegative definite Hermitian operators, π_i, that sum to the identity operator.[8, 9] That is,

$$\pi_i \geq 0 \qquad i = 1, \dots, M \tag{1a}$$

$$\sum_{i=1}^{M} \pi_i = I. \tag{1b}$$

It has further been shown that the necessary and sufficient conditions for a set of π_i satisfying (1) to minimize the error probability are, in addition to (1),[3, 5, 8]

$$\sum_i p_i(\rho_i\pi_i - \pi_i\rho_i) = 0 \tag{2a}$$

$$\Pi_j(\sum_i p_i\pi_i\rho_i - p_j\rho_j) = 0 \tag{2b}$$

$$\sum_i p_i\pi_i\rho_i - p_j\rho_j \geq 0, \quad \text{all } j. \tag{2c}$$

Here ρ_i is the density operator of the received field when message i is transmitted.

For the pure-state problem, which is of interest here, the density operators ρ_i are given by

$$\rho_i = |u_i\rangle\langle u_i| \qquad i = 1, \dots, M. \tag{3}$$

Since we have assumed that the $|u_i\rangle$ are linearly independent, their weighted sum will vanish only if all of the weighting coefficients vanish. That is,

$$\Sigma\ a_i|u_i\rangle = 0 \tag{4}$$

implies that all of the a_i are zero. Although (3) and (4) yield some simplification of (2), they do not permit an explicit solution for the π_i.[11, 3, 5] As we shall show, they do imply that π_i satisfying (2) can be found which describe a receiver that is characterized by observables. Precisely stated, if the Hilbert space of the system is taken to be that spanned by the $|u_i\rangle$, the operators π_i that satisfy (2) subject to (1) are computing projection operators. That is,

$$\pi_i\pi_j = \delta_{ij}\pi_i \qquad i, j = 1, \dots, M, \tag{5}$$

where the δ_{ij} is the Kronecker delta; $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$.

Of course, there is no a priori reason for limiting the space of the system to be that spanned by the $|u_i\rangle$. Moreover, if larger spaces are considered, (2) will have solutions that do not satisfy (5). It can be shown, however, that the use of a larger space does not lead to performance improvement. That is, insofar as the performance is concerned, no generality is lost by assuming that the system space is spanned by the $|u_i\rangle$. The use of a larger space may, however, lead to simpler, and more easily interpreted, measurements. We assume that the Hilbert space associated with the system is spanned by the M linearly independent vectors $|u_i\rangle$ and prove that the π_i also satisfy (5).

The proof has three parts. First, we demonstrate that a set of vectors $|f_i\rangle$, i = 1, . . . , M can be found for which

$$\langle f_i u_j\rangle = 0 \qquad i \neq j, \qquad i, j = 1, \dots, M \tag{6a}$$

$$\langle f_i u_i\rangle \neq 0 \qquad i = 1, \dots, M. \tag{6b}$$

Using these vectors, we then show that the vectors $|v_i\rangle$ defined by

$$|v_i\rangle \equiv \pi_i |u_i\rangle \tag{7}$$

satisfy the expression

$$\pi_j |v_i\rangle = \delta_{ij} |v_i\rangle \qquad i, j = 1, \dots, M, \tag{8}$$

and are linearly independent when the π_j satisfy (2). The validity of (5) follows easily from (8) and the linear independence of the $|v_i\rangle$.

To demonstrate the existence of M vectors $|f_i\rangle$ satisfying (6), we invoke the assumptions that the $|u_i\rangle$ are linearly independent and span the space of the system. These assumptions imply that, for every i, there exists a vector $|f_i\rangle$ that is orthogonal to $u_i\rangle$ for all $j \neq i$ and is not orthogonal to $|u_i\rangle$. That is, $|f_i\rangle$ is orthogonal to the M-1 dimensional space spanned by $|u_j\rangle$, $j \neq i$. Thus (6) is proved.

To prove (8) we observe that, for any set π_i satisfying (2b),

$$\pi_j(\sum_i p_i\pi_i\rho_i - p_j\rho_j)\, |f_k\rangle = 0 \qquad j, k = 1, \dots, M. \tag{9a}$$

But

$$\rho_i |f_k\rangle = u_i\rangle \langle u_i f_k\rangle = \delta_{ik} C_i |u_i\rangle, \tag{9b}$$

with $C_k \neq 0$, the rightmost equality being a consequence of (6). Use of (9b) to eliminate the density operators from (9a) yields

$$\pi_j(\sum_i p_i\pi_i\delta_{ik}C_i\,|u_i\rangle) - p_j\pi_j\delta_{jk}C_j\,|u_j\rangle = 0 \qquad j, k = 1, \dots, M, \tag{10a}$$

or

$$p_kC_k\pi_j\pi_k\,|u_k\rangle = \delta_{jk}p_jC_j\pi_j\,|u_j\rangle \qquad j, k = 1, \dots, M. \tag{10b}$$

Introducing the vectors $|v_i\rangle$, defined by (7), in (10b) and noting from (6b) and (9b) that the C_j are nonzero, we obtain (8).

To prove that the $|v_i\rangle$ are linearly independent, we suppose, to the contrary, that they are linearly dependent. Then the space contains a vector, say $|g\rangle$, other than the null vector such that

$$\langle v_ig\rangle = 0 \qquad i = 1, \dots, M. \tag{11}$$

For this vector it follows from (2c) that

$$\sum_i \langle gp_i\pi_i\rho_ig\rangle - \langle gp_j\rho_jg\rangle \geq 0 \qquad j = 1, \dots, M. \tag{12}$$

But

$$\pi_i\rho_i\,|g\rangle = \pi_i\,|u_i\rangle\langle u_ig\rangle = |v_i\rangle\langle u_ig\rangle \qquad i = 1, \dots, M,$$

the rightmost equality being a consequence of (7). Thus, by virtue of the assumed condition (11),

$$\langle g\pi_i\rho_ig\rangle = \langle gv_i\rangle\langle u_i\,g\rangle = 0 \qquad i = 1, \dots, M. \tag{13}$$

Therefore, for (12) and hence (2c) to be satisfied, it is necessary that $\langle g\rho_jg\rangle$ vanish for all j. That is,

$$\langle gu_j\rangle = 0 \qquad j = 1, \dots, M. \tag{14}$$

But since the $|u_j\rangle$ span the space, (14) implies that $|g\rangle$ is the null vector. Consequently the $|v_i\rangle$ cannot be linearly dependent.

The proof is completed by noting that, since the M vectors $|v_i\rangle$ are linearly independent, any vector $|w\rangle$ in the system space may be expressed as

$$|w\rangle = \sum_i b_i\,|v_i\rangle. \tag{15}$$

Thus

$$\pi_j|w\rangle = \sum_i b_i\pi_j|v_i\rangle = \sum_i b_i\delta_{ij}|v_i\rangle = b_j|v_j\rangle, \tag{16}$$

the middle equality being a consequence of (8). Also

$$\pi_k\pi_j|w\rangle = \pi_k\{b_j|v_j\rangle\} = b_j\pi_k|v_j\rangle = b_j\delta_{jk}|v_j\rangle = \delta_{jk}\pi_j|w\rangle, \tag{17}$$

where the middle equality follows from (8) and the rightmost equality follows from (16). Since (17) must be true for all vectors $|w\rangle$, we conclude that

$$\pi_k\pi_j = \delta_j k\pi, \quad \text{all } j, k. \tag{18}$$

Summarizing, for the M-ary linearly independent pure state problem phrased on the space spanned by the M states, the π_i associated with the optimum receiver are commuting projection operators.

The author wishes to thank Professor C. W. Helstrom, Dr. H. P. H. Yuen, and Mr. V. Chan for their comments.

References

1. R. S. Kennedy and S. Karp (Eds.), "Optical Space Communication," NASA SP-217, Office of Technology Utilization National Aeronautics and Space Administration, Washington, D. C., 1969, Sec. 3.4, pp. 18-20.

2. J. P. Gordon and W. H. Louisell, "Simultaneous Measurement of Noncommuting Observables," in P. L. Kelley, B. Lax, and P. E. Tannenwald (Eds.), Physics of Quantum Electronics (McGraw-Hill Book Company, New York, 1966), pp. 833-840.

3. H. P. H. Yuen, "Communication Theory of Quantum Systems," Ph. D. Thesis, Department of Electrical Engineering, M.I.T., June 1970; also Technical Report 482, Research Laboratory of Electronics, M. I. T., August 30, 1971, p. 124.

4. S. D. Personick, "Efficient Analog Communication over Quantum Channels," Ph.D. Thesis, Department of Electrical Engineering, M. I. T., December 1969; also Technical Report 477, Research Laboratory of Electronics, M. I. T., May 15, 1970.

5. H. P. H. Yuen, R. S. Kennedy, and M. Lax, "On Optimal Quantum Receivers for Digital Signal Detection," Proc. IEEE 58, 1770-1773 (1970).

6. C. W. Helstrom, Jane W. S. Liu, and J. P. Gordon, "Quantum-Mechanical Communication Theory," Proc. IEEE 58, 1578-1598 (1970).

7. H. P. H. Yuen and M. Lax, "Multiple Parameter Quantum Estimation and Measurement of Non Self-adjoint Operators," a paper presented at IEEE International Symposium on Information Theory, Asilomar, California, January 31, 1972 (to appear in IEEE Trans. on Information Theory).

8. A. S. Holevo, "Statistical Problems in Quantum Physics," unpublished manuscript, Steklov Mathematical Institute, Moscow, U. S. S. R. (to appear in Proc. Soviet-Japanese Symposium on Probability and Statistics, 1972).

9. C. W. Helstrom and R. S. Kennedy, "Noncommuting Observables in Quantum Detection and Estimation Theory" (submitted for publication to IEEE Trans. (IT)).

10. C. W. Helstrom, "Detection Theory and Quantum Mechanics," Inform. Contr. 10, 254-291 (1967).

11. C. W. Helstrom, "Some Quantum-Mechanical Aspects of Optical Communications," in R. S. Kennedy and S. Karp (Eds.), "Optical Space Communication," NASA SP-217, op. cit., pp. 111-120.

D. COMPUTER UPDATING OF A DATA STRUCTURE

Joint Services Electronics Program (Contract DAAB07-71-C-0300)

R. A. Flower

1. Introduction

Let a data structure $D = \{d_1, d_2, \ldots, d_{|D|}\}$ be a set of data bases to be stored, accessed, and updated on a computer. An example of a data structure D is the set of all possible homework scores for four students with scores ranging between 0 and 7. A particular $d \in D$ is just one of the $|D| = 8^4$ possible homework scores for the students.

A problem of interest is how to store these scores efficiently on a computer so that they will be convenient to use. The user would like (i) to use a small amount of memory for storage, (ii) to be able easily to answer questions such as "what is the fourth student's homework score?," and (iii) to make updates easily; for example, adding late scores or changing scores of incorrectly graded homework. This report focuses primarily on how easily updates or changes can be made.

2. Computer Memory Model

As a model of the computer memory, let $\mathscr{M}$ be an updatable memory of L addressable cells each of which can be set to a binary value $a \in \{0, 1\} = A$. The state of memory is defined as a sequence of L binary values

$$\mathscr{M} = \mathscr{M}(1), \mathscr{M}(2), \ldots, \mathscr{M}(L) \qquad \mathscr{M}(i) \in \{0, 1\} \qquad i = 1, 2, \ldots, L,$$

where $\mathscr{M}(i)$ is the binary value of the i^{th} cell in memory.

To model the process of accessing memory, let $\mathscr{A}$ be a finite-state deterministic automaton connected to the memory $\mathscr{M}$ by four lines. An integer address $n \in N_L = \{1, 2, \ldots, L\}$ can be sent from $\mathscr{A}$ to $\mathscr{M}$ on the "read address" line and the binary value $\mathscr{M}(n) \in \{0, 1\}$ is returned from $\mathscr{M}$ to $\mathscr{A}$ on the "read value" line. An integer address $n \in N_L$ can be sent from $\mathscr{A}$ to $\mathscr{M}$ on the "write address" line and a binary value $a \in \{0, 1\}$ can be sent from $\mathscr{A}$ to $\mathscr{M}$ on the "write value" line, causing the cell $\mathscr{M}(n)$ to take on the value a, while leaving all other cells in their previous state.

3. Update Performance

Next we need some correspondence or map from the data structure D to the memory configurations that are used to represent each $d \in D$. Let A_L be the set of all possible sequences of L binary values. That is, A_L is the set of all 2^L possible configurations for memory. A particular configuration $m_L \in A_L$ is said to represent $d \in D$ if, whenever $\mathscr{M} = M_L$, the automaton $\mathscr{A}$ can access bits of $\mathscr{M}$ and answer certain questions about d. A subset of functions $M \subseteq A_L$ is said to represent D if each $m \in M$ represents some $d \in D$ and each $d \in D$ is represented by some $m \in M$. Denote by $\mu(d)$ the set of all $m \in M$ such that m represents d.

For example, with 16 bits of memory ($L = 16$) the first 4 bits of memory can be set to the binary number for the first student's score, the next 4 bits encode the second student's score, and so forth.

$$\mu((7,6,7,5)) = 0111011001110101 \tag{1}$$

Examples for μ and $\mathscr{M}$ will give the binary string of values stored in consecutive addresses in $\mathscr{M}$, as shown in (1).

Given $\mathscr{A}$ and $\mathscr{M}$, it is of interest to consider the difficulty in making an update from an unknown initial configuration $\mathscr{M} \in M$ to some representation of a specific data base $d_o \in D$. For example, one might want to enter this week's homework scores, now that last week's scores are no longer of interest. Note that this represents a total rewrite of the data base in memory.

It is possible to make an update without reading any bits of memory simply by writing a representation m of d_o $m \in \mu(d_o)$ into memory. It is also possible to make an update by writing only two bits of memory and using the unary encoding described below. Let $\mathscr{M}$ be a memory of $|D|$ bits and let each $d_i \in D$ correspond to the i^{th} bit of memory. Then a representation set M and mapping μ can be constructed by letting the representation of d_i set the i^{th} bit of memory to 1 and set the rest of memory to 0.

$$\mu(d_i) = 000\ \ldots\ 0\underset{\Big\uparrow\ i^{th}\ \text{bit}}{1}0\ \ldots\ 00$$

An update to d_i from an unknown initial configuration $\mathscr{M} \in M$ would only require setting the single initial cell with value 1 to value 0 and then setting the i^{th} cell to value 1.

In summary, it is possible to make updates without reading any bits of memory, and it is possible to make updates by writing only two bits of memory. Neither bits read nor bits written alone, however, is a good measure of the effort required to make an update. When no bits are read, all of memory must be overwritten. When only two bits are written, many cells must be read in order to find the single initial cell with value 1. A more reasonable measure of the effort involved in an update is the sum of these two measures – the number of cells read <u>and</u> the number of cells overwritten.

We are concerned with finding optimum update strategies. Update algorithms U', which read the value at a given address more than once, write a new value into a given address more than once, or read the value at an address already overwritten, may all be improved to algorithms U which have some memory to save the appropriate values. The algorithms U will require fewer accesses, and hence algorithms U' will not be considered.

For any update algorithm $U(m_o)$ which reads and writes in memory until $\mathcal{M} = m_o$, a corresponding binary tree may be constructed as follows. For an $m \in M$ construct a root node, attach to this node a node label containing the address in memory accessed i, and a type label, containing the action of the automaton (read, write 0, or write 1). Also construct an output branch from the node and label it with the binary value initially in memory at the address accessed m(i). If the algorithm halts here with $\mathcal{M} = m_o$, construct a leaf node; otherwise repeat this process, adding new nodes and new output branches until the update algorithm halts with $\mathcal{M} = m_o$. Repeat this process for each $m \in M$, adding new nodes and/or new output branches as necessary. Note that since the update algorithm is deterministic, there is a single root node and the resulting tree is a binary tree. As an example, consider the representation for the three-element data structure $D = \{d_1, d_2, d_3\}$:

$$M = \{000, 101, 010, 111\} \qquad \mu(d_1) = 000$$

$$\mu(d_2) = 101$$

$$\mu(d_3) = 010, 111$$

An update algorithm U(010) leaving an unknown initial memory configuration $m \in M$ in the final state $\mathcal{M} = 010$ is the following.

```
U(010):  Read cell 1:  If the value is 0 go to B
         Write value 0 in cell 1
         Write value 0 in cell 3
     B:  Write value 1 in cell 2   END.
```

The corresponding tree is shown in Fig. VII-3a. Figure VII-3b is a more general update, which we shall discuss.

Each leaf node is reached by updating only one initial $m \in M$ (shown in parenthesis below the leaf nodes for the example). If a leaf node were reached by m and $m' \in M$, $m \neq m'$, then m and m' must differ in some position not accessed by the update algorithm. If this were the case, however, when the update algorithm halted there might be one of two different final representations in memory corresponding to the different values in the unaccessed position. But $U(m_o)$ was an algorithm that guaranteed the final state of memory to be $\mathcal{M} = m_o$, so each leaf node must be reached by only a single $m \in M$.

Let $r_U(m, m_o)$ be the number of bits of memory read by algorithm $U(m_o)$ in updating the initial configuration $\mathscr{M} = m$ to the final configuration $\mathscr{M} = m_o$, and let $c_U(m, m_o)$ be the number of bits overwritten by algorithm $U(m_o)$. The path length from the root node to the leaf node corresponding to initial configuration m is $r_U(m, m_o) + c_U(m, m_o) \triangleq \ell(m, m_o)$. It is known that the path lengths of a binary tree must satisfy an inequality known as the Kraft inequality.[2] For the binary tree corresponding to $U(m_o)$ this inequality takes the form

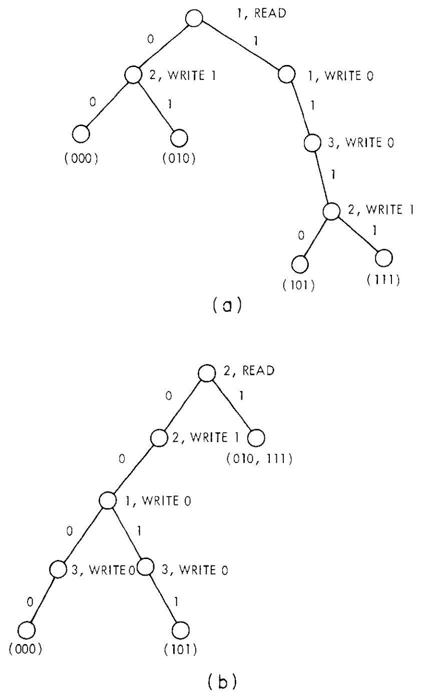


Fig. VII-3. Update algorithms for D.
$D = \{d_1, d_2, d_3\}$ $M = \{000, 101, 010, 111\}$
$\mu(d_1) = 000$
$\mu(d_2) = 101$
$\mu(d_3) = 010, 111$

(a) Binary tree corresponding to update algorithm U(010), an algorithm that updates an unknown initial configuration to final representation 010 of d_3.

(b) Binary tree corresponding to $U(d_3)$, an algorithm that updates an unknown initial configuration to some representation of d_3.

$$\sum_{m \in M} 2^{-\ell(m, m_o)} \leq 1. \tag{2}$$

If there is a unique representation for each $d \in D \, |\mu(d)| = 1$, then for each d there is some initial configuration m that requires reading and writing at least $\log |M| = \log |D|$ bits in updating to $\mathscr{M} = \mu(d)$, since if all $\ell < \log |M|$, inequality (2) would not hold. (All logs are taken to base 2.) On the other hand, all $d \in D$ may be represented by distinct binary sequences of length $\lceil \log |D| \rceil < 1 + \log |D|$. Thus the lower bound for the longest update is nearly achieved by an algorithm which simply writes m_o in memory.

For some cases an average update might be of more interest than the worst-case update. Let there be a probability distribution Q on the initial representations $q(d) = \Pr\{\mathscr{M}_{\text{initial}} = \mu(d)\}$, where each d has a unique representation. For example, all of the students may be bright (or the grader may be soft) and homework scores of 6 or 7 would be much more likely than lower scores. Then it is known[1] that inequality (2) implies that $\bar{\ell} = \overline{r+c} \geq H(D)$, where H(D) is given by $H(D) = -\sum_{d \in D} q(d) \log q(d)$. This reduces to the earlier inequality when all data bases are equiprobable, so that $q = 1/|D|$ and $H(D) = \log |D|$.

To apply these results to the homework example, some update will require reading and/or changing $\log |D| = \log |8^3| = 12$ bits of memory. For the discussed representation M, an update from $\mu(0,0,0,0) = 0000000000000000$ to $\mu(7,7,7,7) = 0111011101110111$ will require changing 12 bits.

Suppose the homework scores have the following probabilities:

$$\Pr\{7777\} = \frac{1}{2}$$

$$\Pr\{7677\} = \Pr\{7767\} = \Pr\{7776\} = \frac{1}{8}$$

$$\Pr\{\text{any other scores}\} = \frac{1}{8} \cdot \frac{1}{2^{12}-4}$$

For this distribution $H(D) = \frac{1}{2} \log 2 + 3 \cdot \frac{1}{8} \cdot \log 8 + (2^{12}-4) \cdot \frac{1}{8(2^{12}-4)} \log (8(2^{12}-4))$, which is slightly less than 7/2.

If each update for the representation is made by writing in the new scores, then each update will require 12 writes, which is far above the lower bound to the average. For the homework scores, one might try to improve the average update with the following encoding: If all scores are 7, set $m(1) = 1$. If the second, third, or fourth score is 6 and the rest of the scores are 7, then set $m(1) = 0$ and set $m(2)$ and $m(3)$ to the binary number for the student who scored 6. For any other scores, set $m(1) = m(2) = m(3) = 0$ and use the remaining memory to encode the scores in four 3-bit fields.

If updates are made by writing an encoding of the scores into memory, then an update to (7777) requires writing only a single bit of memory, an update to (7677) (7767) or (7776) requires writing 3 bits of memory, and the remaining updates require writing 15 bits of memory. The average update requires

$$\bar{\ell} = \frac{1}{2} \cdot 1 + \frac{3}{8} \cdot 3 + (2^{12}-4) \frac{1}{8(2^{12}-4)} \cdot 15 = \frac{7}{2}$$

accesses, which is only slightly greater than H(D) and is less than H(D) + 1. Note that such an encoding no longer has a single unique $m = \mu(d)$ for each d. For example, any memory m with $m(1) = 1$ is a representation of (7777) and there are 2^{L-1} of them, $|\mu(7777)| = 2^{L-1}$.

Such an encoding can be constructed for any D and any probability distribution on D by using a result from information theory. In information theory there is a well-known procedure[1] for constructing a set c of code words of variable length corresponding to a set of messages such that the length of the average code word is less than H(c) + 1, where H(c) is the entropy of the message set.[1]

Such an encoding may be used as a representation M of D by writing the code word of length j for d into consecutive memory locations m(1), m(2), . . . , m(j) and allowing the remaining memory locations m(j+1), m(j+2), . . . , m(L) to retain their initial values. Such a Huffman code obeys the prefix condition which means that only the bits set by the encoding for d need to be read in determining which d is in memory. Hence, the remaining bits of memory may be used for some other purpose if desired. If updates are made simply by writing the code words for d into memory, then the average update will be less than H(D) + 1.

Further analysis is required, however, to show that H(D) is still a lower bound for the average update when multiple representations are allowed.

4. Analysis of Multiple Representations

If there is more than one representation of a data base d, then it is unnecessarily restrictive to require the update algorithm U(d) to finish with a specific predetermined representation $m \in \mu(d)$ for d. Any $m \in \mu(d)$ will do equally well (see Fig. VII-3b).

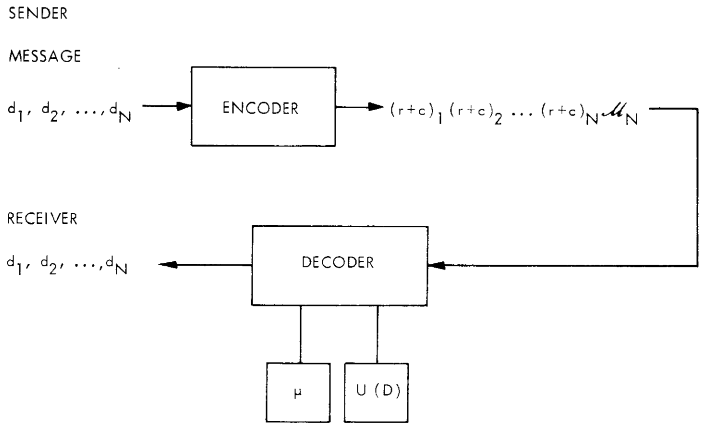
Let U(d) be an update algorithm that only requires that the final memory configuration be some representation for d. It can be shown that each leaf of the tree corresponding to U(d) can be reached by no more than $|\mu(d)|$ of the possible initial memory configurations. Then $\ell(m, d)$, the number of bits read or written by U(d) in updating $\mathscr{M}_i = m$ to some representation for d, may be shown to satisfy the following extension of the Kraft inequality

$$\sum_{m \in M} 2^{-\ell m, d)} \leq |\mu(d)|. \tag{3}$$

It follows from (3) that some initial configuration requires reading and writing at least $\frac{|M|}{|\mu(d)|}$ bits to update to a representation of d for each d. Since not all data bases can have a larger-than-average number of representations, for some d $|\mu(d)| \leq \frac{|M|}{|D|}$. Thus for some d, some initial configuration requires reading and writing at least $\log |D|$ bits to update to $m \in \mu(d)$. Allowing multiple representations has not improved the bound on the longest update. Nor can allowing multiple representations improve the average update for a family of update algorithms U(D) capable of an indefinite number of updates. Let the probability distribution on updates be $q(d) = \Pr\{\text{updating to } d\}$. A series of N updates may be considered to be a message to be transmitted from a sender to a receiver. For each update i, the sender transmits the bits read or overwritten in reverse chronological order (i.e., the last bit read or overwritten in the i^{th} update is the first bit sent in the block of bits corresponding to the i^{th} update). Following the last update, the final memory representation $\mathscr{M}_N$ is sent (see Fig. VII-4).

The receiver has a copy of the update trees for U(D) and a copy of the data structure-memory representation map μ. He decodes the N-update message as follows. From the final memory configuration $\mathscr{M}_N$ and the data structure map he can determine for which d, $\mathscr{M}_N \in \mu(d)$, and hence the N^{th} update instruction d_N. From the update tree for d_N and the sequence $(r+c)_N$ of bits read or overwritten in the N^{th} update, the receiver can reconstruct the memory configuration $\mathscr{M}_{N-1}$ just after the $(N-1)^{\text{th}}$ update and just before the N^{th} update. From μ he can then determine for which d, $\mathscr{M}_{N-1} \in \mu(d)$, and hence d_{N-1}. This procedure can be iterated to recover the entire message $d_1 d_2 \ldots d_N$.

The encoding is a uniquely decodable code, so by the converse to the coding theorem of information theory,[1] if successive updates are chosen with statistical independence,



d_i THE i^{th} UPDATE INSTRUCTION
(i.e., THE DATA BASE IN MEMORY AFTER THE i^{th} UPDATE)

$(r+c)_i$ THE VALUES IN THE CELLS OF MEMORY WHEN READ OR OVERWRITTEN IN MAKING THE i^{th} UPDATE IN REVERSE CHRONOLOGICAL ORDER

$\mathscr{M}_N$ FINAL STATE OF MEMORY $\mathscr{M}_N(1), \mathscr{M}_N(2), \ldots, \mathscr{M}_N(L)$

Fig. VII-4. Model of updates as a communication problem.

$\bar{n} \geq$ H(message) = N H(D), but $\bar{n} = N(\overline{r+c}) + L$ (where L is the number of bits in memory). This means that $\overline{r+c} \geq H(D) + \frac{L}{N}$ and any family of update algorithms U(d) that are capable of making an indefinite number of updates on a computer memory must satisfy $\bar{\ell} = \overline{r+c} \geq H(D)$.

A map μ which Huffman encodes D and a family of update algorithms U(D) which essentially writes that encoding into memory can attain $\overline{r+c} < H(D) + 1$.

5. Conclusions

We have considered the effort involved in updating a computer representation for a data structure. Multiple representations for each $d \in D$ cannot reduce the longest update, nor can it, in general, reduce the average update. Any family of algorithms U(D) capable of an indefinite number of updates must satisfy $\overline{r+c} \geq H(D)$. There is a set of representations and a family of update algorithms for which $\overline{r+c} < H(D) + 1$.

References

1. R. G. Gallager, Information Theory and Reliable Communication (John Wiley and Sons, Inc., New York, 1968), see especially Chap. 3.

2. L. G. Kraft, "A Device for Quantizing, Grouping and Coding Amplitude Modulated Pulses," S.M. Thesis, Department of Electrical Engineering, M.I.T., June 1949.