Quantum-Mechanical Linear Filtering of Random Signal Sequences

JOHN S. BARAS, MEMBER, IEEE, ROBERT O. HARGER, MEMBER, IEEE, AND YOUNG H. PARK

Abstract—**The problem of estimating a member of a scalar random signal sequence with quantum-mechanical measurements is considered. The minimum variance linear estimator based on an optimal present quantum measurement and optimal linear processing of past measurements is found. When the average optimal measurement without post-processing, for a fixed signal, is linear in the random signal and the signal sequence is pairwise Gaussian, the optimal processing separates: the optimal measurement is the same as the optimal measurement without regard to past data, and the past and present data are processed classically. The results are illustrated by considering the estimator of the real amplitude of a laser signal received in a single-mode cavity along with thermal noise; when the random signal sequence satisfies a linear recursion, the estimate can be computed recursively. For a one-step memory signal sequence it is shown that the optimal observable generally differs from the optimal observable disregarding the past; the optimal measurement can be computed recursively.**

Manuscript received February 7, 1975; revised August 1, 1975. This work was supported in part by the National Science Foundation under Grant GK-14920.

J. S. Baras and R. O. Harger are with the Electrical Engineering Department, University of Maryland, College Park, Md. 20742.

Y. H. Park is with Bendix Corporation, Columbia, Md.

I. Introduction

Detection and estimation problems have recently been studied [1]–[3] employing measurement models correctly incorporating quantum mechanics. Such work applies directly, e.g., to establishing fundamental limitations in optical communication systems [4]. More recently, the analog of filtering a random signal sequence has been considered [5], [6], [13]. Here the problem of estimating x_k, a member of a "signal" sequence $\{x_0,x_1,\cdots,x_k,\cdots\}$ of scalar random variables, is considered; the parameter k is conveniently regarded as discrete time. To be chosen are the optimal measurements at time k and the optimal linear combination of present and past measurements at times $j = 0, 1,\cdots, k - 1$. The random sequence so obtained is defined precisely below, but it is simply described in the optical communication setting as follows.

At time k a laser signal modulated in some fashion by x_k is received in a cavity containing otherwise only an electromagnetic field due to thermal noise; the total field is in a state described by a density operator $\rho(x_k)$ that depends on x_k (but not otherwise on k). If x_k is a scalar, the measurement at time k, whose outcome is denoted by v_k, will correspond to a self-adjoint operator V_k [7, p. 192]. If x_k is a vector the essentially quantum problem of simultaneous measurement arises, and a more general concept of measurement [1], [2], [8], [9] must be resorted to [14].

By optimal is meant *minimum mean-square error*; the implied average is over the classical distributions of $\{x_k\}$ and the distributions due to quantum-mechanical measurement.

An ultimate objective would include efficient computation; e.g., suppose that x_k is a scalar "dynamical state" generated by the recursive equation

$$x_{k+1} = \phi_k x_k + w_k \tag{1}$$

where $\{\phi_k\}$ is a sequence of scalars and $\{w_k\}$ is a sequence of independent Gaussian random variables with zero mean and variance Q_k. Solutions in a form that recursively compute the optimal estimate and measurement at time k would be highly desirable. In specific situations below, this is achieved.

II. Filtering Problem

The customary formulation of quantum mechanics [10, sec. 8.5] associates a self-adjoint operator V on a Hilbert space $\mathscr{H}$ with each measurement and incorporates *a priori* statistical information with a density operator ρ on $\mathscr{H}$ (ρ a self-adjoint, positive semidefinite operator with unit trace). The measurement represented by V produces a real number v (the outcome) whose expectation is

$$E_v\{v\} = \mathrm{tr}\,\{\rho V\}$$

where $\mathrm{tr}\,\{\cdot\}$ denotes trace.[1] In case the density operator ρ depends on a random variable x with distribution function F_x, E_v should be replaced with the conditional expectation $E_{v|x}$. The unconditional expectation is then

$$E\{v\} = \int \mathrm{tr}\,\{\rho(x)V\}F_x(dx). \tag{2}$$

Here the following sequence of measurements is of interest. At each time j, $j = 0,1,\cdots$, a measurement represented by the self-adjoint operator V_j is made, with outcome v_j. The state of the system prior to the measurement is described by $\rho(x_j)$. The outcomes v_j are classical random variables which, conditioned upon a fixed signal sequence $\{x_j\}$, are independent.[2] This conditional independence of the measurement outcomes implies[3] that for any multinomial of $\{v_0,\cdots,v_k\}$

$$E_{v|x}\{v_0{}^m\cdots v_k{}^n\} = \mathrm{tr}\,\{\rho(x_0)V_0{}^m\}\cdots \mathrm{tr}\,\{\rho(x_k)V_k{}^n\}$$

for any integers $m,\cdots,n$. The unconditional expectation is then

$$E\{v_0{}^m\cdots v_k{}^n\} = \int \mathrm{tr}\,\{\rho(x_0)V_0{}^m\} \cdots \mathrm{tr}\,\{\rho(x_k)V_k{}^n\}F_{x_0,\cdots,x_k}(dx_0,\cdots,dx_k). \tag{3}$$

The linear filtering problem is the following. At time k, $k = 0,1,\cdots$, the previous outcomes $\{v_j,\ j = 0,1,\cdots, k-1\}$ and the present measurement outcome v_k are used to form a linear estimate

$$\hat{x}_k = \sum_{j=0}^{k} c_j(k)v_j \tag{4}$$

of x_k. Then the $\{c_j(k), j = 0,\cdots,k\}$ and the present measurement represented by V_k are to be chosen to minimize the mean-square error $E\{\mathscr{E}_k{}^2\}$, where $\mathscr{E}_k \equiv x_k - \hat{x}_k$ and the expectation is as in (3). Clearly one may set $c_k(k) = 1$.

Explicitly writing out that part of $E_{v|x}$ for the kth stage yields

$$E\{\mathscr{E}_k{}^2\} = E_x E_{v|x}\,\mathrm{tr}\left\{\rho(x_k)\left[x_k I - V_k - I\sum_{j=0}^{k-1} c_j(k)v_j\right]^2\right\} \tag{5}$$

where I is the identity operator on $\mathscr{H}$. It is also convenient to note that

$$E\{\mathscr{E}_k{}^2\} = E\{x_k{}^2\} - 2\sum_{j=0}^{k} c_j(k)\,\mathrm{tr}\,(\delta_{kj}V_j) + \sum_{i,j=0}^{k} c_i(k)c_j(k)\,\mathrm{tr}\,\{\zeta_{ji}V_j\} \tag{6}$$

where

$$\delta_{kj} \equiv E_x\{x_k\rho(x_j)\} \tag{7}$$

$$\eta_k \equiv E_x\{\rho(x_k)\} \tag{8}$$

and

$$\zeta_{ij} \equiv \begin{cases} E_x\{(\mathrm{tr}\,\rho(x_j)V_j)\rho(x_i)\}, & i \neq j;\\ \eta_i V_i, & i = j. \end{cases} \tag{9}$$

In the sequel, $\hat{V}_k$ will denote the optimal observable and $\{\hat{c}_j(k),\ j = 0,\cdots,k-1\}$ the optimal processing coefficients at the kth stage. Applying the calculus of the variations argument of [12] to $[V_k + I\sum_{j=0}^{k} c_j(k)v_j]$ in (5) formally gives a necessary condition for $\hat{V}_k$ to minimize separately $E\{\mathscr{E}_k{}^2\}$

$$\eta_k\hat{V}_k + \hat{V}_k\eta_k = 2\delta_{kk} - 2\sum_{j=0}^{k-1}\hat{c}_j(k)\zeta_{kj}. \tag{10}$$

Simple differentiation on (6) shows that a necessary and sufficient condition that the $\{\hat{c}_j(k)\}_{j=0}^{k-1}$ minimize separately

[1] It is worthwhile to note the distribution function F_v of the classical random variable v. The spectral theorem [7, p. 249] associates with each self-adjoint operator V on $\mathscr{H}$ a unique spectral measure M_V, a mapping of Borel sets of the real line into projection operators on $\mathscr{H}$. Then the distribution function is $F_v(v) = \mathrm{tr}\,\{\rho M_V(-\infty,v]\}$. The spectral theorem also yields the moments of the random variable v via $E_v(v^m) = \mathrm{tr}\,\{\rho V^m\}$, $m = 1,2,\cdots$.

[2] In the optical communication example cited above, this conditional independence corresponds to "clearing" the receiver cavity prior to each reception.

[3] The conditional independence assumed here is best described by stating that the joint distribution function $F_{v|x}$ is the product $F_{v_0|x_0}\cdots F_{v_k|x_k}$, where each $F_{v_j|x_j}$ has already been described.

$E\{\mathscr{E}_k{}^2\}$ is

$$\sum_{j=0}^{k-1} \hat{c}_j(k) \operatorname{tr}\{\zeta_{ji}V_j\} = \operatorname{tr}\{\delta_{ki}V_i\} - \operatorname{tr}\{\zeta_{ki}\hat{V}_k\}, \qquad i = 0, 1, \cdots, k-1. \quad (11)$$

It is important for the subsequent results of this paper to establish necessary and sufficient conditions for $\hat{V}_k$ and $\{\hat{c}_j(k), j = 0, \cdots, k-1\}$ to minimize jointly $E\{\mathscr{E}_k{}^2\}$. This is done in the following theorem which employs the projection theorem [11, p. 49]. It also settles the question of the existence of optimal V_k and $\{c_j(k)\}$.

Theorem 1: There exists an optimum observable $\hat{V}_k$ and optimal processing coefficients $\hat{c}_j(k)$, $j = 0, 1, \cdots, k-1$, if and only if there exists a solution to (10) and (11).

Proof: Let $\mathscr{L}$ be the set of operator-valued functions of the form

$$f(x) \equiv \beta x I + I \sum_{j=0}^{k-1} \alpha_j v_j + V_k$$

where x is a random variable, β and $\{\alpha_i\}$ are real scalars, and V_k is a self-adjoint operator on $\mathscr{H}$. With the ordinary addition of scalars and operators and the multiplication by scalars, $\mathscr{L}$ is seen to be a linear space. For $f,g \in \mathscr{L}$, define the form

$$(f,g) \equiv E_x E_{v|x} \operatorname{tr}\{\rho(x_k) \cdot [f(x_k)g(x_k) + g(x_k)f(x_k)]\}.$$

Let $\mathscr{L}' \subset \mathscr{L}$ be the subspace of elements f such that (f,f) is finite. Then $(\cdot,\cdot)$ is a degenerate inner product on $\mathscr{L}'$ in the sense that $\|\cdot\| \equiv (\cdot,\cdot)^{1/2}$ is a seminorm [11, p. 45]. It is not a norm since $\|f\| = 0$ does not imply $f = 0$.

Let $\mathscr{L}'' \subset \mathscr{L}'$ be the subspace of operator-valued functions of the form

$$h \equiv I \sum_{j=0}^{k-1} \alpha_j v_j + D_k.$$

Then (see (5)) the problem of minimizing the mean-square error is a minimum norm problem, and the projection theorem [11, p. 44] provides necessary and sufficient conditions for a solution. $\{\hat{c}_j(k),\ j = 0, 1, \cdots, k-1\}$ and $\hat{V}_k$ are the solution if and only if, for any real scalars α_j, $j = 0, \cdots, k-1$, and self-adjoint operator D_k on $\mathscr{H}$,

$$\begin{aligned} 0 = E_x E_{v|x} \operatorname{tr}\Bigg[\rho(x_k) \cdot \Bigg\{\Bigg[x_k I - I \sum_{j=0}^{k-1} \hat{c}_j(k) v_j - \hat{V}_k\Bigg] &\cdot \Bigg[I \sum_{j=0}^{k-1} \alpha_j v_j + D_k\Bigg] + \Bigg[I \sum_{j=0}^{k-1} \alpha_j v_j + D_k\Bigg] \\ &\cdot \Bigg[x_k I - I \sum_{j=0}^{k-1} \hat{c}_j(k) v_j - \hat{V}_k\Bigg]\Bigg\}\Bigg]. \end{aligned} \quad (12)$$

Two necessary conditions, which together are sufficient, may be obtained from (12), the first by setting the $\{\alpha_j\} \equiv 0$ and the second by setting $D_k \equiv 0$.

Setting the $\{\alpha_j\} \equiv 0$ and interchanging the trace over $\mathscr{H}$ with expectation $E_x E_{v|x}$, one obtains

$$0 = \operatorname{tr}\Bigg\{D_k \Bigg[2\delta_{kk} - 2 \sum_{j=0}^{k-1} \hat{c}_j(k)\zeta_{kj} - \eta_k \hat{V}_k - \hat{V}_k \eta_k\Bigg]\Bigg\}$$

for any self-adjoint operator D_k. The arbitrariness of D_k implies this last equality holds if and only if

$$\eta_k \hat{V}_k + \hat{V}_k \eta_k = 2\delta_{kk} - 2 \sum_{j=0}^{k-1} \hat{c}_j(k)\zeta_{kj}.$$

Setting $D_k \equiv 0$, one is similarly led to the condition

$$\sum_{j=0}^{k-1} \hat{c}_j(k) \operatorname{tr}\{\zeta_{ji}V_j\} = \operatorname{tr}\{\delta_{ki}V_i\} - \operatorname{tr}\{\zeta_{ki}\hat{V}_k\}, \qquad i = 0, 1, \cdots, k-1.$$

Q.E.D.

Equations (11) are the normal equations [11, p. 56] for the $\{\hat{c}_j(k)\}$. A redundant equation may be obtained from (10) by multiplying through by $\hat{V}_k$ and tracing; adding it to the above set, one has the complete set of normal equations. It is special in that necessarily $\hat{c}_k(k) \equiv 1$

$$\sum_{j=0}^{k} \hat{c}_j(k) E\{v_i v_j\} = E\{v_i x_k\}, \qquad i = 0,1,\cdots,k \quad (11a)$$

where the expectation is as in (3) and $v_k = \hat{v}_k$, the outcome of the optimal measurement. The equations (10) and (11) can be cast in a more convenient form.

Corollary 1: The optimal observable $\hat{V}_k$ and processing coefficients $\{\hat{c}_j(k)\}$ satisfy the equations

$$\hat{V}_k = T_k - \sum_{j=0}^{k-1} \hat{c}_j(k)\sigma_{kj} \quad (13)$$

and

$$\sum_{j=0}^{k-1} \hat{c}_j(k) \operatorname{tr}\{\zeta_{ji}V_j - \zeta_{ki}\sigma_{kj}\} = \operatorname{tr}\{\delta_{ki}V_i - \zeta_{ki}T_k\}, \qquad i = 0, 1, \cdots, k-1 \quad (14)$$

where T_k and σ_{kj} are such that

$$\eta_k T_k + T_k \eta_k = 2\delta_{kk} \quad (15)$$

and

$$\eta_k \sigma_{kj} + \sigma_{kj}\eta_k = 2\zeta_{kj}. \quad (16)$$

Proof: Substituting (15) and (16) into (10) immediately yields (13). Multiplying (13) on the right by ζ_{ki}, $i = 0, 1, \cdots, k-1$, tracing over $\mathscr{H}$, and substituting for $\operatorname{tr}\{\zeta_{ki}\hat{V}_k\}$ in (11) yields (14). Q.E.D.

Equations (13) and (14) are "decoupled" in the sense that, after solving (15) and (16), the $\{\hat{c}_j(k)\}$ are found via (14); then $\hat{V}_k$ is found via (13). Note also that conditions for existence of solutions in (10) imply existence of solutions for (15) and (16), and conversely.

It is remarked that (13) and (14) apply for any set of $k+1$ jointly distributed random variables $\{x_0,x_1,\cdots,x_k\}$ and for any set of k prior measurements represented by $\{V_0,V_1,\cdots,V_{k-1}\}$. If, additionally, the $\{x_j\}$ satisfy a recursion such as (1) there is the hope that a recursive determination of $\hat{V}_k$ and, at least implicitly, of the $\{\hat{c}_j(k)\}$ could be obtained, especially if the $\{V_j\}$ are chosen optimally at each time j. This would avoid a calculation of growing complexity at each time k. It is also of interest to know when

$\hat{V}_k$ depends in a significant structural way, on k—for then a new measuring device is required at each time k! We now turn to examples that partially answer such questions.

III. Filter Separation

Assume the $\{x_j, j = 0,1,\cdots,k\}$ are pairwise Gaussian random variables and that the observables $\{V_j, j = 0,1 \cdots,k\}$ have each been chosen optimally according to (13) and (14); suppose further

$$\operatorname{tr}\{\rho(x_j)T_j\} = \Gamma_j x_j, \qquad j = 0,1,\cdots,k \tag{17}$$

where Γ_j is a scalar, that is, the average optimal measurement without postprocessing (see [12]), for a fixed signal, is proportional to said signal.

Theorem 2: If $\{x_j, j = 0,1,\cdots,k\}$ are pairwise Gaussian random variables, if measurements $\{\hat{V}_j, j = 0,1,\cdots,k\}$ are optimally chosen (according to (13) and (14)), if $\{T_j, j = 0,1,\cdots,k\}$ are given by (15), and if (17) holds, then

$$\hat{V}_k = B_k T_k \tag{18}$$

where

$$B_k = 1 - \sum_{j=0}^{k-1} \hat{c}_j(k) B_j \Gamma_j A_{jk}, \qquad B(0) = 1 \tag{19}$$

and A_{jk} is such that $E(x_j \mid x_k) = A_{jk}x_k$.

Proof: Trivially, (18) holds at $k = 0$. At time $k = 1$, $\hat{V}_1 = T_1 - \hat{c}_0(1)\sigma_{10}$; to find σ_{10} by (16) note that (using (17))

$$\begin{aligned}\zeta_{10} &= E_x\{\operatorname{tr}\{\rho(x_0)T_0\} \cdot \rho(x_1)\}\\ &= E_x\{\Gamma_0 x_0 \rho(x_1)\}\\ &= \Gamma_0 E_x\{\rho(x_1)E(x_0 \mid x_1)\}.\end{aligned}$$

Since x_0 and x_1 are jointly Gaussian random variables there exists a constant A_{01} such that $E(x_0 \mid x_1) = A_{01}x_1$, therefore,

$$\begin{aligned}\zeta_{10} &= \Gamma_0 A_{01} E_x\{\rho(x_1)x_1\}\\ &= \Gamma_0 A_{01}\delta_{11}.\end{aligned}$$

Using this result in (16) and comparing to (15), one sees that $\sigma_{10} = \Gamma_0 A_{01} T_1$. So (13) yields

$$\hat{V}_1 = B_1 T_1$$

where

$$B_1 \equiv 1 - \hat{c}_0(1) B_0 \Gamma_0 A_{01}.$$

Assuming (18) and (19) hold at time $k - 1$, again one finds

$$\begin{aligned}\zeta_{kj} &\equiv E_x\{\operatorname{tr}\{\rho(x_j)V_j\} \cdot \rho(x_k)\}\\ &= E_x\{\operatorname{tr}\{\rho(x_j)B_jT_j\} \cdot \rho(x_k)\}\end{aligned}$$

using the induction hypothesis,

$$\zeta_{kj} = B_j \Gamma_j A_{jk} \delta_{kk}$$

where $E(x_j \mid x_k) = A_{jk}x_k$. Using this result in (16) and comparing to (15), one sees $\sigma_{kj} = B_j \Gamma_j T_k$, thus (13) gives

$$\hat{V}_k = B_k T_k$$

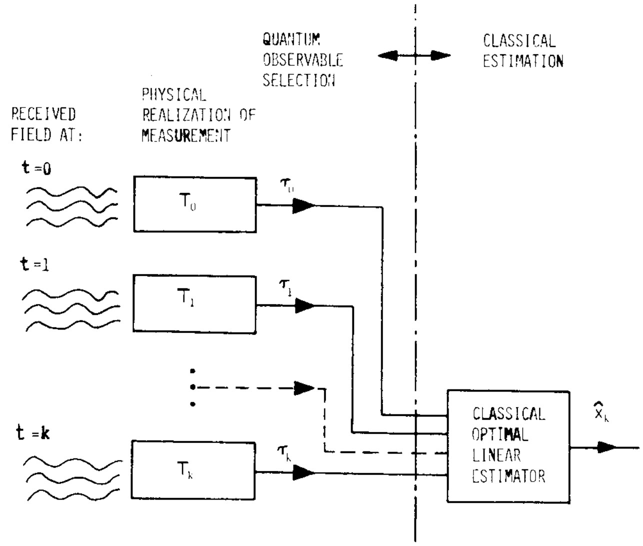

Fig. 1. Separation of optimal filter when signal sequence is pairwise Gaussian and average optimal measurement without postprocessing, for fixed signal, is linear in signal.

where

$$B_k = 1 - \sum_{j=0}^{k-1} \hat{c}_j(k) B_j \Gamma_j A_{jk}.$$

Q.E.D.

Note that the observable $\hat{V}_k$ of (18) is proportional to T_k, *the optimal measurement if the past measurements are disregarded* (*proof: set* $k = 0$). This $\hat{V}_k$ is greatly simpler than that of (13) and yields the following "separation." *The optimal quantum observables are chosen separately from the optimal classical postprocessing of the measurement outcomes.* This is illustrated in Fig. 1.

Note that the left side of (17) is $E(\tau_j \mid x_j)$, where τ_j is the outcome of the measurement represented by T_j, which is, therefore, linear in x_j—as is true if τ_j and x_j are jointly Gaussian when necessarily $\Gamma_j = E(\tau_j x_j)/E(x_j^2)$.

Lemma 1: If (17) holds then i) $\Gamma_j = E(\tau_j x_j)/E(x_j^2)$ and ii) $0 \le \Gamma_j \le 1$.

Proof: Multiplying (17) through by x_j and taking $E_x\{\cdot\}$, one finds $E(x_j\tau_j) = \Gamma_j E(x_j^2)$ establishing i). However, $E(x_j\tau_j) = \operatorname{tr}\{\delta_{jj}T_j\}$, which by (15) is $\operatorname{tr}\{\eta_j T_j^2\} = E\{\tau_j^2\}$, thus

$$\Gamma_j = \frac{E(x_j\tau_j)}{E(x_j^2)} = \frac{E(\tau_j^2)}{E(x_j^2)} \ge 0.$$

However, $[E(x_j\tau_j)]^2 \le E(x_j^2)E(\tau_j^2)$ so that

$$\Gamma_j = \frac{E(\tau_j^2)}{E(x_j^2)} \le 1.$$

Q.E.D.

In view of (18) the optimal estimate is

$$\begin{aligned}\hat{x}_k &\equiv \hat{v}_k + \sum_{j=0}^{k-1} \hat{c}_j(k)\hat{v}_j\\ &= B_k\tau_k + \sum_{j=0}^{k-1} B_j \hat{c}_j(k)\tau_j.\end{aligned} \tag{20}$$

The normal equations (11a) become

$$B_i \sum_{j=0}^{k} B_j \hat{c}_j(k) E\{\tau_i \tau_j\} = B_i E\{\tau_i x_k\}, \qquad i = 0,1,\cdots,k.$$

Without loss of generality one can assume each $B_j \neq 0$, for if $B_j = 0$, for any j, $\hat{c}_j(k)$ is indeterminate but does not affect $\hat{x}(k)$; thus the jth equation can be deleted along with the jth column of the matrix of elements $B_i B_j E(\tau_i \tau_j)$, and this reduced matrix equation can be solved instead. Dividing B_j out of the jth equation, one has

$$\sum_{j=0}^{k} B_j \hat{c}_j(k) E\{\tau_i \tau_j\} = E\{\tau_i x_k\}, \qquad i = 0,1,\cdots,k. \quad (21)$$

Comparing equations (20) and (21), one has the following.

Theorem 3: If $\{x_j, j = 0,1,\cdots,k\}$ are jointly Gaussian random variables, if measurements $\{\hat{V}_j, j = 0,1,\cdots,k\}$ are optimally chosen (according to (10) and (11)), and if $\mathrm{tr}\{\rho(x_j)T_j\} = \Gamma_j x_j$, $j = 0,1,\cdots,k$, then the $\{\tau_j, j = 0,1,\cdots,k\}$ are a sufficient statistic for $\hat{x}_k$.

Theorem 3 makes it clear that the estimator including the past measurements will perform at least as well as an estimator using only the present measurement. Also, if the measurement outcomes $\{\tau_j, j = 0,1,\cdots,k\}$ allow a (classical) recursive estimate of $\hat{x}_k$, the quantum filtering problem will have a recursive solution, such an example follows.

Example: Suppose that x_k is a Gaussian random variable and is transmitted as the real amplitude of a laser signal (assumed monochromatic) and received, along with thermal noise, in a single-mode cavity upon which an optimal measurement is to be made. The density operator in the coherent state or P-representation is then [4]

$$\rho(x_k) = \frac{1}{\pi N} \int \exp\left(-\frac{|\alpha - x_k|^2}{N}\right) |\alpha\rangle\langle\alpha| d^2\alpha$$

and the solution to (15) is known [12] to be

$$T_k = D_k \cdot \frac{a + a^+}{2} \qquad D_k \equiv \frac{2\lambda_k}{N + 2\lambda_k + \frac{1}{2}}$$

here N defines the thermal noise level and $\lambda_k \equiv E(x_k^2)$. A measurement of $(a + a^+)/2$, assuming fixed x_k, results in a Gaussian random variable with mean x_k and variance $(N/2 + 1/4)$ and is realized by homodyning [12].

Thus x_k and τ_k are jointly Gaussian random variables and $E\{\tau_k \mid x_k\} = \mathrm{tr}\{\rho(x_k)T_k\} = D_k x_k$. Theorem 3 applies here with $\Gamma_j = D_j$. Moreover, in this case T_k is proportional to an observable $(a_k + a_k^+)/2$ that is structurally independent of k, only one type of device is required—a simplification of great practical importance. Clearly the $(k + 1)$ measurements of $Y_j \equiv (a_j + a_j^+)/2$ at the times $j = 0,1,\cdots,k$ gives a sufficient statistic for the optimal estimate $\hat{x}_k$. Now

$$\hat{x}_k = B_k \Gamma_k y_k + \sum_{j=0}^{k-1} B_j \Gamma_j \hat{c}_j(k) y_j \quad (20a)$$

where the $\{y_j\}$ are the $(k + 1)$ outcomes of the measurements represented by $\{Y_k\}$, and the normal equations (21)

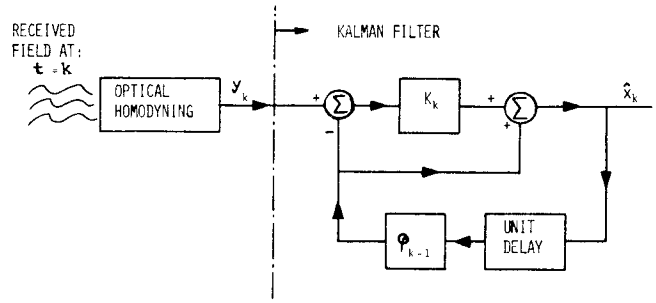

Fig. 2. Optimal filter for signal sequence of (1) when received as amplitude of a laser signal in single-mode cavity along with thermal noise.

become

$$\sum_{j=0}^{k} \Gamma_j B_j \hat{c}_j(k) E\{y_i y_j\} = E\{y_i x_k\}, \qquad i = 0,1,\cdots,k. \quad (21a)$$

Equations (20a) and (21a) describe an equivalent, although fictitious, classical estimation problem. Estimate x_k given a sequence of observations

$$y_k = x_k + u_k$$

where $\{u_j, j = 0,1,\cdots,k\}$ is a sequence of independent zero-mean identically distributed Gaussian random variables with variance $(N/2 + 1/4)$.

Furthermore, if the sequence $\{x_j, j = 0,1,\cdots,k,\cdots\}$ satisfies the recursion (1), then $\hat{x}_k$ can be recursively calculated by the Kalman–Bucy filtering equations [11, p. 96]

$$\hat{x}_k = \phi_{k-1}\hat{x}_{k-1} + K_k[y_k - \phi_{k-1}\hat{x}_{k-1}]$$

where the so-called Kalman gain is

$$K_k = P_k \left[P_k + \left(\frac{N}{2} + \frac{1}{4}\right)\right]^{-1}$$

and

$$P_k = \phi_{k-1}^2\{P_{k-1}[1 - K_{k-1}]\} + Q_{k-1}$$

is the error variance based on the past k observations. See Fig. 2.

IV. Finite-Memory Signal Process

As an example in a different direction, suppose $\{x_j, j = 0,1,\cdots,k,\cdots\}$, a sequence of zero-mean random variables, is such that x_j and x_i are independent if $|j - i| > 1$. Such a random sequence is said to have a "one-step memory."

Theorem 4: If $\{x_j, j = 0,\cdots,k,\cdots\}$ has a one-step memory and each observable $\hat{V}_j$, $j = 0,1,\cdots,k$, is chosen optimally according to (13) and (14), then

$$\hat{V}_k = T_k - \hat{c}_{k-1}(k)\sigma_{k,k-1}, \qquad k \geq 1, \quad \hat{V}_0 = T_0. \quad (22)$$

Proof: For $k = 1$, trivially, the relation is true. For time $k + 1 > 2$, by (13)

$$\hat{V}_{k+1} = T_{k+1} - \sum_{j=0}^{k} \hat{c}_j(k+1)\sigma_{k+1,j} \quad (23)$$

where $\sigma_{k+1,j}$ is determined by (16); in turn, by (9),

$$\zeta_{k+1,j} \equiv E_{\boldsymbol{x}}\{\operatorname{tr}[\rho(x_j)\hat{V}_j]\rho(x_{k+1})\}. \tag{24}$$

For $j < k$, using the assumption of pairwise independence, $\zeta_{k+1,j} = [\operatorname{tr}\eta_j\hat{V}_j]\eta_{k+1}$ implying that $\sigma_{k+1,j} = [\operatorname{tr}\eta_j\hat{V}_j]I$. However, using the induction hypothesis,

$$\begin{aligned}\operatorname{tr}\eta_j\hat{V}_j &= \operatorname{tr}\eta_j\{T_j - \hat{c}_{j-1}(j)\sigma_{j,j-1}\}\\ &= \operatorname{tr}\delta_{jj} - \hat{c}_{j-1}(j)\operatorname{tr}\zeta_{j,j-1}\end{aligned}$$

(using (15) and (16)); as $\operatorname{tr}\delta_{jj} = E(x_j) = 0$, using (24) gives

$$\operatorname{tr}\eta_j\hat{V}_j = -\hat{c}_{j-1}(j)\operatorname{tr}\eta_{j-1}\hat{V}_{j-1}.$$

Iterating this procedure, eventually a product with the factor $\operatorname{tr}\eta_0\hat{V}_0 = \operatorname{tr}\delta_{00} = E(x_0) = 0$ appears. Thus $\zeta_{k+1,j}$ and $\sigma_{k+1,j} \equiv 0, j < k$. Q.E.D.

Note (22) may be written (using (9), (15), and (16))

$$\eta_k\hat{V}_k + \hat{V}_k\eta_k = 2\delta_{kk} - 2\hat{c}_{k-1}(k)E_{\boldsymbol{x}}\{\operatorname{tr}[\rho(x_{k-1})\hat{V}_{k-1}]\rho(x_k)\}$$

which, knowing $\hat{c}_{k-1}(k)$, gives $\hat{V}_k$ recursively in terms of $\hat{V}_{k-1}$. Recursive calculations of $\hat{c}_{k-1}(k)$, the mean-square error at time k, and $\hat{x}_k$ can also be given, and these results extend to the "n-step memory" case for $n > 1$ [15].

Acknowledgment

The authors wish to thank an anonymous reviewer for suggestions simplifying the original version of this paper.

References

[1] C. W. Helstrom, "Detection theory and quantum mechanics," *Inform. Contr.*, vol. 10, pp. 254–291, Mar. 1967.
[2] A. S. Holevo, "Statistical problems in quantum physics," in *Proc. Soviet-Japanese Symp. on Probability and Statistics*, vol. 1, pp. 22–40, 1972.
[3] ——, "Statistical decision theory for quantum systems," *J. Multivariate Analysis*, vol. 3, pp. 337–394, 1973.
[4] C. W. Helstrom, J. W. Liu, and J. P. Gordon, "Quantum-mechanical communication theory," *Proc. IEEE*, vol. 58, pp. 1578–1598, Oct. 1970.
[5] Y. H. Park, "Quantum linear recursive minimum mean-square error estimation," Ph.D. dissertation, Dep. Elec. Eng., Univ. Md., College Park, Aug. 1974.
[6] J. S. Baras, R. O. Harger, and A. Ephremides, "Recursive filtering of operator-valued processes in quantum estimation," in *Proc. 1974 Int. Symp. on Information Theory*, Notre Dame, Ind., Oct. 1974, p.5.
[7] E. Prugovečki, *Quantum Mechanics in Hilbert Space*. New York: Academic, 1971.
[8] H. P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of nonselfadjoint observables," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 740–750, Nov. 1973.
[9] C. W. Helstrom and R. S. Kennedy, "Noncommuting observables in quantum detection and estimation theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 16–24, Jan. 1974.
[10] J. M. Jauch, *Foundations of Quantum Mechanics*. Reading, Mass.: Addison-Wesley, 1968.
[11] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[12] S. D. Personick, "Application of quantum estimation theory to analog communication over quantum channels," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 240–246, May 1971.
[13] J. S. Baras and Y. H. Park, "Estimation of random signals based on quantum mechanical measurements," in *Proc. 8th Asilomar Conf. on Circuits, Systems, and Computers*, 1974, pp. 533–539.
[14] J. S. Baras and R. O. Harger, "Multiparameter filtering with quantum measurements," in *Proc. Johns Hopkins Conf. on Information Sciences and Systems*, 1975, pp. 178–184.
[15] R. O. Harger and L. Wilhelm, "Quantum mechanical filtering of finite memory signal sequences," in *Proc. Johns Hopkins Conf. on Information Sciences and Systems*, 1975, (Abstract), p. 185.