Quantum Mechanical Linear Filtering of Vector Signal Processes

JOHN S. BARAS, MEMBER, IEEE, AND ROBERT O. HARGER, MEMBER, IEEE

***Abstract*—The problem of estimating a member of a discrete-time vector process from past and present quantum mechanical measurements is considered; specifically, the minimum-variance linear estimator based on optimal present measurements is studied. Necessary and sufficient conditions that characterize the optimal processing matrix coefficients and the optimal measurements are discussed and interpreted. The optimal linear filter is compared to the optimal quantum estimator without postprocessing of past data. When the signal sequence is pairwise Gaussian and the optimal quantum measurement without postprocessing has the properties that it is linear in a specific sense and that its outcome and the corresponding element of the signal sequence are jointly Gaussian, then the optimal linear filter separates. That is, the optimal measurement can be taken to be 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 filtering of the in-phase and quadrature amplitudes of a laser field received in a single-mode cavity along with thermal noise. In this case, when the random signal sequence satisfies a linear recursion, the estimate can be computed recursively in a very efficient manner.**

I. INTRODUCTION

DETECTION and estimation problems with quantum statistics, such as arise in the theory of optical communication systems, have been heavily studied [1]–[4]. In previous work [5], we analyzed the problem of linearly filtering a scalar signal sequence utilizing quantum mechanical measurements. In this paper, the problem of estimating $x(k)$, a member of the signal sequence $\{x(0), x(1), \cdots, x(k), \cdots\}$ of vector random variables is considered; the parameter k is conveniently regarded as discrete time. As is well-known [2], [6], [7] the multiparameter (or vector) case is much more complex than the corresponding scalar problem, and its treatment necessitates the use of *generalized quantum measurements* in the sense of Holevo [2]. This is solely due to the essential quantum mechanical limitation on simultaneous measurements [8, p. 260], [9, p. 101].

Briefly, the problem is to choose the optimal measurement at time k and the optimal linear processing of present and past measurements at times $i = 0,1, \cdots, k-1$. Optimal is understood here as *minimum mean-square error,* and the implied average is over the classical distributions of $\{x(k)\}$ and the distributions due to quantum mechanical measurements.

A convenient example for physical motivation is provided by the following optical communication setting [3]–[5]: at time k, a laser field 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. The filtering problem consists of estimating $x(k)$ from quantum mechanical measurements via the procedure described above.

An ultimate objective of this work is to find interesting practical cases that result in simplification of the filter structure or that make feasible the computations indicated by the necessary and sufficient conditions for optimality. In particular, suppose $x(k)$ is generated by a linear recursive equation

$$x(k+1) = \Phi(k)x(k) + w(k), \qquad k = 0,1, \cdots, \qquad (1)$$

where $\{\Phi(k)\}$ is a sequence of matrices, and $\{w(k)\}$ is a sequence of independent Gaussian random vectors with zero mean and covariance matrix $Q(k)$. Recursive computation of the optimal estimate and measurement at time k would be highly desirable. This is achieved in a specific situation stemming from an optical communication problem. Some of the results of this paper have been announced in [10] and [11]; the mathematical proofs of some of the results will appear elsewhere [12].

II. THE FILTERING PROBLEM

The customary formulation of quantum mechanics [8, p. 258] associates a self-adjoint operator V on a Hilbert

Manuscript received June 10, 1976; revised March 7, 1977. This work was supported in part by the National Science Foundation under Grant ENG 75-20900.

The authors are with the Department of Electrical Engineering, University of Maryland, College Park, MD 20742.

space H with a measurement, incorporates *a priori* statistical information with density operator ρ on H (ρ is a self-adjoint positive definite operator with unit trace, and represents *the state* of the quantum system [9, p. 94, p. 132]). The measurement represented by $\boldsymbol{V}$ produces a real number v (the outcome) whose expectation is $E\{v\} = \text{Tr}\,[\rho \boldsymbol{V}]$ (where Tr denotes the operation of taking the trace of an operator on H [8, p. 374]). This formulation is adequate for restricted estimation problems only, in particular for the estimation of a scalar. When a vector is to be estimated, the essentially quantum mechanical problem of simultaneous measurements arises, and a more general concept of measurement must be resorted to [2, p. 341].

To assist in motivating the concept of a generalized quantum measurement, we first elaborate slightly on the customary formulation. The spectral theorem [8, p. 249] associates with each self-adjoint operator $\boldsymbol{V}$ on H a unique spectral measure $\boldsymbol{M}_{\boldsymbol{V}}(\cdot)$, a mapping of the Borel sets of the real line into projection operators on H. The distribution function of the outcome v is then $F_v(\xi) = \text{Tr}\,[\rho \boldsymbol{M}_{\boldsymbol{V}}(-\infty,\xi]]$. The spectral theorem yields the moments $E\{v^m\} = \text{Tr}\,[\rho V^m]$, $m = 1,2,\cdots$. The spectral measure $\boldsymbol{M}_{\boldsymbol{V}}(\cdot)$ is fundamental, and is termed a *simple measurement* [2]. Following Holevo [2, p. 341], a *generalized measurement* is a map $\boldsymbol{M}$ from the σ-algebra of Borel sets $\mathcal{B}^n$ of the n-dimensional space $\mathbb{R}^n$ to the algebra $\mathfrak{B}(H)$ of all bounded linear operators on H such that i) $\boldsymbol{M}(B) \geq 0$, for every $B \in \mathcal{B}^n$; ii) if $\{B_i\} \subseteq \mathcal{B}^n$ is a partition of $\mathbb{R}^n$, then $\Sigma_i \boldsymbol{M}(B_i) = \boldsymbol{I}$, where the series converges weakly in $\mathfrak{B}(H)$ [13, p. 53] and where $\boldsymbol{I}$ is the identity operator on H. That is, a measurement is a *positive operator-valued measure* (POM) [14, p. 6], or a *generalized resolution of the identity* [15, p. 121]. It is worth noting that if $\boldsymbol{M}$ is an *orthogonal* resolution of the identity, i.e., if in addition $B \cap C = \phi$, for B, $C \in \mathcal{B}^n$, implies that $\boldsymbol{M}(B)\boldsymbol{M}(C) = \boldsymbol{0}$, then $\boldsymbol{M}$ is necessarily a spectral measure [14, p. 12], and thus we have a simple measurement. A POM $\boldsymbol{M}$ induces a probability measure $\mu_{\boldsymbol{M}}$ on $\mathcal{B}^n$ via $\mu_{\boldsymbol{M}}(B) = \text{Tr}\,[\rho \boldsymbol{M}(B)]$, for $B \in \mathcal{B}^n$, as is readily verified; thus $\boldsymbol{M}$ is also sometimes termed a *probability operator measure*. The interpretation of this mathematical construct is that a generalized measurement $\boldsymbol{M}$ represents a physical measurement process with outcomes $u \in \mathbb{R}^n$, with distribution function

$$F_u(\xi) = F_u(\xi_1, \cdots, \xi_n) = \text{Tr}\,[\rho \boldsymbol{M}(-\infty,\xi]], \tag{2}$$

where $(-\infty,\xi] \equiv (-\infty,\xi_1] \times (-\infty,\xi_2] \times \cdots \times (-\infty,\xi_n]$.

Consider now the moment $E\{u_i\}$, the expectation of the ith component of the outcome u of the measurement represented by the POM M

$$E\{u_i\} = \int_{\mathbb{R}^n} u_i F_u(du_1, \cdots, du_n)$$

$$= \int_{\mathbb{R}^n} u_i \, \text{Tr}\,[\rho \boldsymbol{M}(du)], \qquad i = 1, \cdots, n.$$

Assuming the interchange is permitted, this point is discussed carefully in Holevo [2, sec. 6], we have

$$E\{u_i\} = \text{Tr}\left[\rho \int_{\mathbb{R}^n} u_i \boldsymbol{M}(du)\right].$$

The integral is a well-defined self-adjoint operator on H [2], [14] that we denote by $\boldsymbol{U}_i$. Then $E\{u_i\} = \text{Tr}\,[\rho \boldsymbol{U}_i]$, $i = 1, \cdots, n$.

Consider next the second-order moment

$$E\{u_i u_j\} = \int_{\mathbb{R}^n} u_i u_j F_u(du_1, du_2, \cdots, du_n)$$

$$= \text{Tr}\left[\rho \int_{\mathbb{R}^n} u_i u_j \boldsymbol{M}(du)\right],$$

where again the operator integral is a well-defined self-adjoint operator on H [2], [14] which we denote by $\boldsymbol{U}_{ij}$. Clearly $\boldsymbol{U}_{ij} = \boldsymbol{U}_{ji}$, but, unlike the special case when $\boldsymbol{M}$ is a spectral measure, $\boldsymbol{U}_{ij} \neq \boldsymbol{U}_i \boldsymbol{U}_j$. Holevo termed the operators

$$\boldsymbol{U}_{i_1,\cdots,i_k} = \int_{\mathbb{R}^n} u_{i_1} \cdots u_{i_k} \boldsymbol{M}(du)$$

the *operator moments* of the POM $\boldsymbol{M}$. In particular, the $\boldsymbol{U}_i$ are the *first operator moments* and the $\boldsymbol{U}_{ij}$ are the *second operator moments*. Observe that in the case of a simple measurement, the POM is uniquely defined by its first operator moment. A direct consequence here of $\boldsymbol{U}_{ij} \neq \boldsymbol{U}_i \boldsymbol{U}_j$ is that the mean-square error (mse) will not be expressible directly in terms of self-adjoint operators in a quadratic form (as in the scalar filtering problem [5]), but rather will remain expressed in terms of POM's. Since the set of self-adjoint operators on H is a linear space while the set of POM's is only a convex set, the nature of the optimization problem will be different.

As pointed out by Holevo [2, p. 343], this generalization of the concept of a quantum measurement is well-justified in view of Naimark's theorem [15, p. 124] which asserts that, given a generalized measurement $\boldsymbol{M}$ in H, there exists an auxiliary Hilbert space H_e, a (pure) density operator ρ_e on $\mathfrak{B}(H_e)$, and a simple measurement $\boldsymbol{E}_{\boldsymbol{M}}$ in $H \otimes H_e$ (the tensor product of Hilbert spaces H and H_e [8, p. 144] such that

$$\text{Tr}\,[\rho \boldsymbol{M}(B)] = \text{Tr}\,[(\rho \otimes \rho_e)\boldsymbol{E}_{\boldsymbol{M}}(B)], \tag{3}$$

for every $B \in \mathcal{B}^n$ and every density operator ρ on H. That is, the distribution functions of the measurement outcomes induced by the generalized measurement $\boldsymbol{M}$ and the simple measurement $\boldsymbol{E}_{\boldsymbol{M}}$ are the same. The physical interpretation is [2], [7] that a generalized quantum measurement is realized by the measurement of compatible observables (i.e., a simple measurement) on a composite quantum system produced by adjoining to the original system characterized by (ρ,H) an auxiliary system characterized by (ρ_e,H_e). Thus, justifiably, the triple $(H_e,\rho_e,\boldsymbol{E}_{\boldsymbol{M}})$ is called a *realization* of the measurement represented by the POM $\boldsymbol{M}$. For simplicity, we shall refer hereafter to generalized quantum measurements as quantum measurements unless explicitly stated otherwise.

Let x be a vector random variable with distribution function F_x on which the density operator ρ depends; i.e.,

$\rho = \rho(x)$. Then the distribution function $F_u(\xi)(2)$ of the vector outcome u of the generalized quantum measurement $\boldsymbol{M}$ becomes a conditional distribution function $F_{u|x}(\xi,\zeta)$. The first moments of u are

$$E\{u_i\} = E_x\{E\{u_i|x\}\} = \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} u_i \operatorname{Tr}[\rho(x)\boldsymbol{M}(du)]\right] F_x(dx).$$

Interchanging the order of $\boldsymbol{M}(du)$, integration and trace yields

$$E\{u_i\} = \int_{\mathbb{R}^n} \operatorname{Tr}[\rho(\xi)U_i]F_x(d\xi).$$

Similarly,

$$E\{u_i u_j\} = \int_{\mathbb{R}^n} \operatorname{Tr}[\rho(\xi)U_{ij}]F_x(d\xi).$$

In this paper, the following sequence of measurements is of interest. At each time $i = 0,1,\cdots$, a measurement represented by the POM $\boldsymbol{M}_i$ is made with outcome $\upsilon(i) \in \mathbb{R}^n$. The state of the system prior to measurement is characterized by the density operator $\rho(x(i))$, where $x(i)$ is a member of the sequence of random vectors $x(j)$, $j = 0,1,\cdots$. The outcomes $\{\upsilon(i)\}$ (classical vector random variables) are assumed to be independent when conditioned upon the sequence $\{x(i)\}$. In the optical communication example cited in the introduction, this conditional independence corresponds to "clearing" the receiver cavity prior to each reception [5]. Therefore, the joint distribution function of the measurement outcomes $\upsilon(0),\cdots,\upsilon(k)$ is given by

$$F_{\upsilon(0),\cdots,\upsilon(k)}(\nu(0),\cdots,\nu(k)) = E_x\{\operatorname{Tr}[\rho(x(0))\boldsymbol{M}_0(-\infty,\nu(0)]] \cdots \operatorname{Tr}[\rho(x(k))\boldsymbol{M}_k(-\infty,\nu(k)]]\}, \quad (4)$$

where $E_x\{\cdot\}$ denotes expectation with respect to the joint distribution of $x(0),\cdots,x(k)$.

The linear filtering problem is the following. Let $\upsilon(i)$, $i = 0,\cdots,k-1$ be the outcomes of measurements, represented by the POM's $\boldsymbol{M}_i$, which were made at times $t = 0,\cdots,k-1$. A new measurement, represented by $\boldsymbol{M}_k$, is performed at $t = k$, and the present and past outcomes are combined linearly to give the estimator

$$\hat{x}(k) = \sum_{i=0}^{k} C_i(k)\upsilon(i), \quad (5)$$

where $C_i(k)$, $i = 0,\cdots,k$, are $n \times n$ matrices. The problem then is to find a POM $\boldsymbol{M}_k$ and matrices $C_i(k)$, $i = 0,\cdots,k$ to minimize the mean-square error

$$\text{mse} = E\{(x(k) - \hat{x}(k))^t(x(k) - \hat{x}(k))\}, \quad (6)$$

where the expectation is with respect to the distributions of $\{x(i)\}$ and the measurement distributions, as in (4). Here, the superscript t denotes transpose.

One might think that a technique similar to the one used in the corresponding scalar problem (cf. [5]) employing the projection theorem in an appropriate space of operator valued functions would work for the vector case as well. That is, it might seem that one should employ Naimark's theorem to construct auxiliary Hilbert spaces H_{ei}, $i = 0,\cdots,k$, and simple measurements $\boldsymbol{E}_i$, $i = 0,\cdots,k$, on $H \otimes H_{ei}$ (cf. [3]) statistically equivalent to $\boldsymbol{M}_i$, $i = 0,\cdots,k$, and then proceed as in the scalar case. However, a careful review of Naimark's theorem [15, p. 124] reveals that the inner product of the space $H \otimes H_{ek}$ *depends on* $\boldsymbol{M}_k$ *explicitly*, and therefore we cannot apply projection theorem techniques since the inner product of the underlying Hilbert space depends on one of the variables over which we are optimizing, namely $\boldsymbol{M}_k$.

First, we observe that we can set $C_k(k) = I_n$ (the identity matrix on $\mathbb{R}^n$) without loss of generality. Indeed, consider any pair of POM $\boldsymbol{X}$ and $n \times n$ matrix C, and let $\upsilon \in \mathbb{R}^n$ be the outcomes of the measurement represented by $\boldsymbol{X}$. Let $c(x) = Cx$ be a linear map from $\mathbb{R}^n$ into $\mathbb{R}^n$ and define, for every $A \in \mathcal{B}^n$, $\boldsymbol{X}'(A) = \boldsymbol{X}(c^{-1}(A))$. It is easy to verify that $\boldsymbol{X}'$ is a POM which represents the measurement with outcomes $C\upsilon$. Thus, we shall hereafter take $C_k(k) = I_n$.

The calculation of the mse (6), using (4) and (5), is straightforward, given that certain interchanges in the order of trace and $\boldsymbol{M}_k(du)$-integration are permitted; the calculation is rigorously justified in [2] and [12], generally requiring that each $x(i)$ and the past measurement outcomes $\upsilon(i)$ have finite second moments. We find

$$\text{mse} = \operatorname{Tr}\int_{\mathbb{R}^n} \boldsymbol{G}(u, \boldsymbol{C}(k))\boldsymbol{M}_k(du) + E\left\{\left(\sum_{i=0}^{k-1} C_i(k)\upsilon(i)\right)^t \left(\sum_{j=0}^{k-1} C_j(k)\upsilon(j)\right)\right\},$$

where $\boldsymbol{C}(k) = [C_0(k), C_1(k),\cdots,C_{k-1}(k), I_n]$ and where

$$\boldsymbol{G}(u,\boldsymbol{C}(k)) = E_x\left\{\left[(x(k)-u)^t(x(k)-u) - 2(x(k)-u)^t \cdot \left(\sum_{i=0}^{k-1} C_i(k)E\{\upsilon(i)|x(i)\}\right)\right]\cdot \rho(x(k))\right\}$$

is an operator-valued function for each $\boldsymbol{C}(k)$. To obtain a more compact expression for $\boldsymbol{G}(u,\boldsymbol{C}(k))$, we introduce several operators, namely,

$$\eta(i) = E\{\rho(x(i))\} = \int_{\mathbb{R}^n} \rho(\xi)F_{x(i)}(d\xi), \qquad i = 0,1,\cdots, \quad (7a)$$

which, under our assumptions, is a nonnegative self-adjoint trace-class operator [8, p. 374] on H;

$$\lambda(i) = E\{x(i)^t x(i)\rho(x(i))\} = \int_{\mathbb{R}^n} \xi^t\xi\rho(\xi)F_{x(i)}(d\xi), \qquad i = 0,1,\cdots, \quad (7b)$$

which is a nonnegative self-adjoint trace-class operator on H;

$$\delta(i) = E\{x(i)\rho(x(i))\} = \int_{\mathbb{R}^n} \xi\rho(\xi)F_{x(i)}(d\xi), \qquad i = 0,1,\cdots, \quad (7c)$$

which is an n-vector of self-adjoint trace-class operators on H with component operators $\delta(i)_l = E\{x(i)_l\rho(x(i))\}$, $l = 1,\cdots,n$;

$$\gamma(k,i) = E\{v(i)\rho(x(k))\}$$
$$= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} E\{v(i)|x(i)=\xi\}\rho(\zeta)F_{x(i),x(k)}(d\xi,d\zeta),$$
$$i = 0,1,\cdots, \tag{7d}$$

which is an n-vector of self-adjoint trace-class operators on H with component operators $\gamma_l(k,i) = E\{v_l(i)\rho(x(k))\}$, $l = 1,\cdots,n$; and finally

$$\pi(k,i) = E\{x(k)v(i)^t\rho(x(k))\}$$
$$= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \zeta E\{v(i)^t|x(i)=\xi\}\rho(\zeta)F_{x(i),x(k)}(d\xi,d\zeta),$$
$$i = 0,1,\cdots, \tag{7e}$$

which is an $n\times n$ matrix of self-adjoint trace-class operators on H with element operators $\pi(k,i)_{lj} = E\{x(k)_l v(i)_j\rho(x(k))\}$. We introduce also the following notation: a) for $a\in\mathbb{R}^n$ and β an n-vector of operators, $a^t\beta = \beta^t a = \sum_{i=1}^n a_i\beta_i$, b) for A an $n\times n$ matrix and σ an $n\times n$ matrix of operators, $\operatorname{tr} A\sigma = \operatorname{tr}\sigma A = \sum_{i=1}^n\sum_{j=1}^n A_{ij}\sigma_{ji}$. Using these definitions, we may write

$$G(u,C(k)) = \lambda(k) - 2u^t\delta(k) + u^tu\eta(k) + 2\sum_{i=0}^{k-1} u^tC_i(k)\gamma(k,i) - 2\sum_{i=0}^{k-1}\operatorname{tr} C_i(k)^t\pi(k,i). \tag{8}$$

Furthermore, the operator

$$\zeta(C(k)) = E\left\{\left(\sum_{i=0}^{k-1} C_i(k)v(i)\right)^t\left(\sum_{j=0}^{k-1} C_j(k)v(j)\right)\cdot\rho(x(k))\right\}$$

is a nonnegative self-adjoint trace-class operator on H. If we define the new operator-valued function

$$\mathcal{F}(u,C(k)) = G(u,C(k)) + \zeta(C(k)), \tag{9}$$

we may write

$$\text{mse} = \operatorname{Tr}\int_{\mathbb{R}^n}\mathcal{F}(u,C(k))M_k(du). \tag{10a}$$

Therefore, following the terminology and notation of Holevo [2], we have expressed the mse as the *trace-integral* of the operator valued function $\mathcal{F}(\cdot,C(k))$ with respect to the POM M_k over $\mathbb{R}^n$. We denote the trace-integral by

$$\langle\mathcal{F}(\cdot,C(k)),M_k\rangle_{\mathbb{R}^n} = \operatorname{Tr}\int_{\mathbb{R}^n}\mathcal{F}(u,C(k))M_k(du). \tag{10b}$$

If we now let $\mathcal{M}$ be the convex set of POM's on H, we see that the linear filtering problem becomes: find a POM $\hat{M}_k$ and $n\times n$ matrices $\hat{C}_i(k)$, $i = 0,\cdots,k-1$, which minimize (10) over the set $\mathcal{M}\times(\mathbb{R}^{n\times n})^k$. In [12], the following theorem, which establishes the existence of an optimal measurement and linear filter, is proved.

Theorem 1: Suppose that the signal sequence $\{x(i)\}$ and the measurement outcomes at time $0,1,\cdots,k-1$ have finite second moments. Then there exist POM $\hat{M}_k$ and $n\times n$ matrices $\hat{C}_i(k)$, $i = 0,\cdots,k-1$, which minimize mse. Moreover, the optimal measurement outcome also has finite second moments.

III. Necessary and Sufficient Conditions for Optimality

In this section, we derive necessary and sufficient conditions on the optimal measurement $\hat{M}_k$ and optimal processing coefficient matrices $\hat{C}_i(k)$, $i = 0,1,\cdots,k-1$. Our first result is given by the following theorem.

Theorem 2: Necessary and sufficient conditions for $\hat{C}_0(k),\hat{C}_1(k),\cdots,\hat{C}_{k-1}(k)$ and $\hat{M}_k$ to be optimal processing coefficients and optimal measurement at time k are

i) $\langle\mathcal{F}(\cdot,\hat{C}(k)),X\rangle_{\mathbb{R}^n}\ge\langle\mathcal{F}(\cdot,\hat{C}(k)),\hat{M}_k\rangle_{\mathbb{R}^n}$, for every $X\in\mathcal{M}$

ii) $\sum_{j=0}^k E\{v(i)v(j)^t\}\hat{C}_j^t(k) = E\{v(i)x(k)^t\}$, $i = 0,1,\cdots,k$,

where $\hat{C}_k(k) = I_n$ and $v(k) = \hat{v}(k)$.

Proof: We give only a sketch of the argument; the mathematical details are in [12]. The necessity is clear. Note that ii) are the *normal equations* for the minimum-variance linear estimate of $x(k)$ based on the random variables $v(0),\cdots,v(k-1)$ and $\hat{v}(k)$, with the constraint $\hat{C}_k(k) = I_n$. The sufficiency is more complicated. It is based on the fact that mse is a quadratic function of $C(k)$ and a linear function of X. Given any fixed POM X, we define matrices $D_0(k),\cdots,D_{k-1}(k)$ which satisfy ii) (the normal equations) when the densities are those induced by X. Then, for any set of matrices $C_0(k),\cdots,C_{k-1}(k)$, we clearly have that

$$\langle\mathcal{F}(\cdot,C(k)),X\rangle_{\mathbb{R}^n}\ge\langle\mathcal{F}(\cdot,D(k)),X\rangle_{\mathbb{R}^n}. \tag{11}$$

Now it is proved in [12] that i) and ii) above imply that, for any $X\in\mathcal{M}$,

$$\langle\mathcal{F}(\cdot,D(k),X\rangle_{\mathbb{R}^n}\ge\langle\mathcal{F}(\cdot,\hat{C}(k)),\hat{M}_k\rangle_{\mathbb{R}^n}. \tag{12}$$

But then (11) and (12) prove the optimality of $\hat{C}_0(k),\cdots,\hat{C}_{k-1}(k)$ and $\hat{M}_k$.

We now concentrate on condition i) of Theorem 2 in our effort to improve these necessary and sufficient conditions. This condition represents an optimization problem with respect to the POM X (while $\hat{C}(k)$ is held fixed) similar to those studied extensively by Holevo in [2, pp. 368–372]. Note that, in our case, the operator-valued function $\mathcal{F}(u,\hat{C}(k))$ is quadratic in u.

The following lemma, an application of the Lagrange duality theorem [17, p. 94], has been announced by Holevo in a more general setting [18]; a proof is also given in [12].

Lemma 1: Let $\mathcal{F}$ be a continuous operator-valued function on $\mathbb{R}^n$ such that, for every $u\in\mathbb{R}^n$, $\mathcal{F}(u)$ is a nonnegative self-adjoint operator with finite trace on a

Hilbert space H. Consider the set of operators on H, $S_{\mathcal{F}} = \{\tau | \tau$ is self-adjoint with finite trace, $\tau \geq 0$, and $\tau \leq \mathcal{F}(u)$ for all $u \in \mathbb{R}^n\}$. Then

$$\inf_{X \in \mathcal{M}} \langle \mathcal{F}, X \rangle_{\mathbb{R}^n} = \max_{\tau \in S_{\mathcal{F}}} \mathrm{Tr}\,[\tau].$$

As a consequence, we have the following lemma.

Lemma 2: Necessary and sufficient conditions for the POM $\hat{M}_k$ to solve the optimization problem described in i) of Theorem 2 are

i) $\mathcal{F}(\cdot, \hat{C}(k))$ is integrable with respect to $\hat{M}_k$,
ii) $\mathcal{F}(u, \hat{C}(k)) \geq \hat{\tau}$, for all $u \in \mathbb{R}^n$,

where $\hat{\tau} \triangleq \int_{\mathbb{R}^n} \mathcal{F}(u, \hat{C}(k)) \hat{M}_k\,(du)$ is well-defined in view of i).

Proof: (Necessity.) From the lemma above we have that there exists $\tau_0 \in S_{\mathcal{F}(\cdot, \hat{C}(k))}$ such that

$$\inf_{X \in \mathcal{M}} \langle \mathcal{F}(\cdot, \hat{C}(k)), X \rangle_{\mathbb{R}^n} = \langle \mathcal{F}(\cdot, \hat{C}(k)), \hat{M}_k \rangle_{\mathbb{R}^n}$$

$$= \max_{\tau \in S_{\mathcal{F}(\cdot, \hat{C}(k))}} \mathrm{Tr}\,[\tau] = \mathrm{Tr}\,[\tau_0]. \quad (13)$$

It remains only to show that $\tau_0 = \hat{\tau}$. Since $\mathcal{F}(\cdot, \hat{C}(k))$ is a quadratic polynomial in $(u_1, \cdots, u_n)$, and hence locally integrable with respect to $\hat{M}\,(du)$, we have that

$$\int_A (\mathcal{F}(u, \hat{C}(k)) - \tau_0) \hat{M}_k\,(du) \geq 0,$$

$$\text{for every bounded } A \in \mathcal{B}^n. \quad (14)$$

Now if there exists bounded $A_0 \in \mathcal{B}^n$ such that

$$\int_{A_0} (\mathcal{F}(u, \hat{C}(k)) - \tau_0) \hat{M}_k\,(du) > 0,$$

we must have $\langle \mathcal{F}(\cdot, \hat{C}(k)), \hat{M}_k \rangle_{\mathbb{R}^n} > \mathrm{Tr}\,[\tau_0]$, which is a contradiction to (13). So (14) is in fact an equality for any bounded $A \in \mathcal{B}^n$. Choosing now an increasing sequence of bounded sets $A_i \in \mathcal{B}^n$, $A_i \rightarrow \mathbb{R}^n$, we have from (14) that $\mathcal{F}(\cdot, \hat{C}(k))$ is integrable with respect to $\hat{M}_k$ over $\mathbb{R}^n$ (see [2] or [12] for definitions). Therefore, from (14), $\tau_0 = \int_{\mathbb{R}^n} \mathcal{F}(u, C(k)) \hat{M}_k\,(du) \triangleq \hat{\tau}$. This completes the proof of necessity.

For the sufficiency, we observe that i) and ii) imply that for any POM X, $\langle \mathcal{F}(\cdot, \hat{C}(k)), X \rangle_{\mathbb{R}^n} \geq \mathrm{Tr}\,[\hat{\tau}] = \langle \mathcal{F}(\cdot, \hat{C}(k)), \hat{M}_k \rangle_{\mathbb{R}^n}$. Then clearly Lemma 1 implies that $\langle \mathcal{F}(\cdot, \hat{C}(k)), \hat{M}_k \rangle_{\mathbb{R}^n} = \inf_{X \in \mathcal{M}} \langle \mathcal{F}(\cdot, \hat{C}(k)), X \rangle_{\mathbb{R}^n}$, and the proof is complete.

As a consequence, we have the following necessary and sufficient conditions for the optimization problem of this section.

Theorem 3: Necessary and sufficient conditions for $\hat{C}_0(k), \cdots, \hat{C}_{k-1}(k)$ and $\hat{M}_k$ to be optimal processing coefficient matrices and an optimal measurement at time k are

i) $\sum_{j=0}^{k} E\{v(i)v(j)^t\}\hat{C}_j^t(k) = E\{v(i)x(k)^t\}$, $i = 0, 1, \cdots, k$, where $\hat{C}_k(k) = I_n$ and $v(k) = \hat{v}(k)$,
ii) $\mathcal{F}(\cdot, \hat{C}(k))$ is integrable with respect to $\hat{M}_k$,
iii) $\mathcal{F}(u, \hat{C}(k)) \geq \hat{\tau}$, for all $u \in \mathbb{R}^n$, where $\hat{\tau} = \int_{\mathbb{R}^n} \mathcal{F}(u, \hat{C}(k)) \hat{M}_k(du)$.

Some remarks are now in order. We observe that the solution to the optimal linear filtering problem in the multiparameter (or vector) case is not as explicit as the solution in the scalar case (cf. [5, eq. (13)–16)]). This was expected since the optimization problem here cannot be formulated as a quadratic problem (cf. our previous remarks on operator moments of measurements). Observe that in the scalar case the measurement (which in that case is simple and represented by a projection-valued measure) is uniquely defined by its first operator moments. That is why the conditions of Theorem 3 can be transformed into the convenient form of [5, corollary 1]. In the vector case, however, the best that can be done generally is to derive explicit necessary and sufficient conditions which characterize the first and second operator moments of the optimal measurements. Since these moments do not determine uniquely the optimal measurement (see also [19, p. 536]), there exists freedom in further restricting the measurement to belong to certain convenient classes. Such a route has been followed by Holevo who used canonical measurements for estimation problems concerning Gaussian states [19].

We note that, although the results of this section do not generally provide an explicit closed form solution for the optimal measurement and optimal processing coefficients, they can be used to establish optimality for suggested processing and measurement schemes. This approach has been successfully employed in similar problems by Holevo [2], [23] and by Belavkin [24], [25] (including problems with non-Gaussian states). The role played by the conditions of Theorem 3 in linear quantum filtering theory is central. It is therefore a natural task to analyze these conditions in detail and to discover cases that permit explicit solution. This will be done for Gaussian statistics in the next section. Other cases will be treated elsewhere.

We also note that the question of implementation is a hard and mostly unanswered one, even for the scalar case. There are very few cases where we know how to implement (with devices) the optimal measurements which result from the solution of the problem. In the vector case, we have *in addition* the problem of finding the auxiliary system and simple measurement necessary to implement a POM. The only example of an explicit construction appears in [20] and with further generalizations in [19].

IV. Gaussian Statistics: Filter Separation

A case which is of particular interest for applications and which admits explicit computation of optimal measurements and processing coefficients is that where the process and measurement statistics are Gaussian. We assume in this section that the vector signal process $\{x(j)\}$ is pairwise Gaussian, and that the outcome $\hat{v}(j)$ of the optimal measurement $\hat{M}_j$, and $x(j)$ are jointly Gaussian for each j. For some of the lemmas, we will need only the linearity in $x(j)$ of the conditional expectation of $\hat{v}(j)$ given $x(j)$, which is a slightly weaker assumption. As in [5], we want to compare the optimal measurements appearing in the optimal filter with the optimal measurements that represent the mimimum-variance estimator of $x(k)$ without postprocessing (i.e., where $C_i(k) = 0$, $i = 0, \cdots, k-1$). This is a natural question to investigate since we would like to find in what way the correlation between the $\{x(i)\}$ affects the optimal measurement; when $\{x(i)\}$ is an uncorrelated sequence, the optimal measurement of the filter coincides with the optimal measurement selected independently for each $x(i)$. The following lemma establishes that this effect is to weigh the estimanda (i.e., $x(j)$) with certain matrices that carry information about the cross correlation of the $\{x(i)\}$.

Lemma 3: Suppose the vector random sequence $\{x(i)\}$ is pairwise Gaussian and that the POM's $\hat{M}_j$, $j = 0,1, \cdots, k$, and the $n \times n$ matrices $\hat{C}_i(j)$, $i = 0,1, \cdots, j-1$, $j = 0,1, \cdots, k$, are optimal measurements and optimal processing matrix coefficients for the optimal linear filter for time $j = 0,1, \cdots, k$. Suppose in addition that $\hat{v}(j)$, the outcomes of $\hat{M}_j$, satisfy

$$E\{\hat{v}(j)|x(j)\} = \Delta(j)x(j) \tag{15}$$

for some $n \times n$ matrices $\Delta(j)$, $j = 0,1, \cdots, k$. Then $\hat{M}_j$ is the optimal measurement for the minimum-variance estimation of $B(j)x(j)$, $j = 0,1, \cdots, k$, without regard to past data. The $n \times n$ matrices $B(j)$ are defined by

$$B(j) = I_n - \sum_{i=0}^{j-1} \hat{C}_i(j)\Delta(i)A(i,j),$$
$$B(0) = I_n, \quad j = 0,1, \cdots, k, \tag{16}$$

where $A(i,j)$ are $n \times n$ matrices such that $E\{x(i)|x(j)\} = A(i,j)x(j)$ and exist since $\{x(i)\}$ is pairwise Gaussian.

Proof: Clearly the result is correct for $j = 0$. Since $\hat{M}_j, \hat{C}_0(j), \hat{C}_1(j), \cdots, \hat{C}_{j-1}(j)$ are optimal for the linear filtering problem at time $j > 0$, they satisfy the conditions of Theorem 3. Therefore, $\mathcal{F}(\cdot, \hat{C}(j))$ in (9) is integrable with respect to $\hat{M}_j$, and

$$\mathcal{F}(u, \hat{C}(j)) - \int_{\mathbb{R}^n} \mathcal{F}(u, \hat{C}(j))\hat{M}_j\,(du) \geq 0, \qquad \text{for } u \in \mathbb{R}^n. \tag{17}$$

Utilizing (8) and (9) in (17) results in

$$\eta(j)\left[u^t u - \int_{\mathbb{R}^n} u^t u \hat{M}_j\,(du)\right] - 2\left[\delta(j) - \sum_{i=0}^{j-1} \hat{C}_i(j)\gamma(j,i)\right]^t \cdot \left[u - \int_{\mathbb{R}^n} u\hat{M}_j\,(du)\right] \geq 0, \tag{18}$$

for $u \in \mathbb{R}^n$. From (7) and the assumptions of the lemma, we have

$$\gamma(j,i) = E\{E\{\hat{v}(i)|x(i)\}\rho(x(j))\} = \Delta(i)A(i,j)\delta(j), \qquad i = 0,1, \cdots, j-1. \tag{19}$$

Because of (19) and the definition of $B(j)$, (18) becomes

$$\eta(j)\left[u^t u - \int_{\mathbb{R}^n} u^t u \hat{M}_j\,(du)\right] - 2[B(j)\delta(j)]^t \cdot \left[u - \int_{\mathbb{R}^n} u\hat{M}_j\,(du)\right] \geq 0, \qquad u \in \mathbb{R}^n. \tag{20}$$

Define the new random variable $y(j) = B(j)x(j)$. Consider the problem of determining the optimal measurement for the minimum-variance estimation of $y(j)$ without regard to past data. We have implicitly studied this problem: it suffices to set $C_i(j) = 0 (i \neq j)$ in the results of the previous sections. So, from Theorem 3, necessary and sufficient conditions for the POM M_j^* to be optimal for this problem are that the operator-valued function (obtained from (8) and (9) by setting $C_i(j) = 0$)

$$\mathcal{F}^*(u;j) = \lambda^*(j) - 2u^t\delta^*(j) + \eta^*(j)u^t u \tag{21a}$$

be integrable with respect to M_j^*, and that

$$\mathcal{F}^*(u;j) - \int_{\mathbb{R}^n} \mathcal{F}^*(u;j)M_j^*\,(du) \geq 0, \qquad \text{for } u \in \mathbb{R}^n. \tag{21b}$$

We can assume, without loss of generality, that $B(j)$ is invertible (cf. comments after (28) below). Therefore, in (21),

$$\eta^*(j) = E\{\rho^*(y(j))\} = E\{\rho(B(j)^{-1}y(j))\} = E\{\rho(x(j))\} = \eta(j), \tag{22}$$

and similarly

$$\delta^*(j) = B(j)\delta(j). \tag{23}$$

Utilizing (19) and the definition (9) of $\mathcal{F}(u, \hat{C}(j))$ and $B(j)$, we see that

$$\mathcal{F}(u, \hat{C}(j)) = \lambda(j) - 2(B(j)\delta(j))^t u + \eta(j)u^t u + \phi, \tag{24}$$

where ϕ is a constant (independent of u) operator. Since $\mathcal{F}(u, \hat{C}(j))$ is integrable with respect to $\hat{M}_j$ by assumption, (21a), (22), (23), and (24) imply that $\mathcal{F}^*(u;j)$ is integrable with respect to $\hat{M}_j$ also. Finally, (20) implies that (21b) is satisfied by $\hat{M}_j$. This completes the proof of the lemma.

We note again that condition (15) in the statement of Lemma 3 is clearly satisfied if $\hat{v}(j)$ and $x(j)$ are jointly Gaussian, $j = 0,1, \cdots, k$.

Lemma 3 provides a simplification in the structure of the filter. Namely, it establishes the fact that the optimal measurement in the filter at time j can be obtained as the solution of the minimum-variance estimation problem for $B(j)x(j)$. This result cannot be used in the construction of the filter, however, for the following reasons: a) to check for the essentially joint Gaussian assumption on $\hat{v}(j)$, $x(j)$ (cf. (15)), we must already know the optimal measurement, b) to construct the matrices $B(j)$, we must already know the optimal matrices $\hat{C}_i(j)$. What is needed is a converse

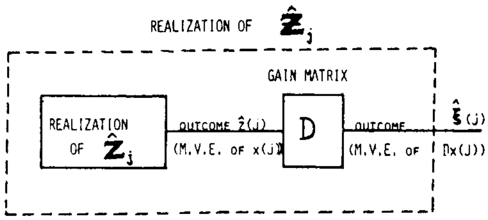

Fig. 1. Illustrating relation between POM's $\hat{\mathcal{Z}}_j$ and $\hat{Z}_j$.

to Lemma 3, namely, a procedure whereby one solves a simpler estimation problem for each $x(i)$ independently, and then combines the results to construct the filter. We proceed now to establish specific conditions sufficient for such a converse.

We consider first the problem of finding the optimal measurement for the minimum-variance estimator of $x(j)$ without regard to past data. Let $\hat{Z}_j$ be an optimal POM for this problem. Let us consider also the same problem for the new vector random variable $Dx(j)$, where D is an $n \times n$ invertible matrix. In general, there may be no relation between the optimal POM's for these two problems. The following special case is of importance for the results of this section: the minimum-variance quantum estimator (MVQE) of $x(j)$ without regard to past data will be called *linear with respect to the $n \times n$ invertible matrix D* if the POM defined by $\hat{\mathcal{Z}}_j(A) = \hat{Z}_j(d^{-1}(A))$, for $A \in \mathcal{B}^n$, where d is the linear map $d(x) = Dx (d: \mathbb{R}^n \to \mathbb{R}^n)$, provides (via its outcomes) a minimum-variance estimator for $Dx(j)$. This notion is illustrated in Fig. 1. It is important to realize that a MVQE can be linear with respect to one matrix but not linear with respect to another.

It is immediately seen, from (22), (23) and the operator equation satisfied by the optimum observable in the scalar case (cf. for example [5, eq. (10) with $k = 0$]), that the minimum-variance quantum estimator of a scalar random variable is linear in the above sense.

Let us now assume (in addition to the hypotheses of Lemma 3) that the minimum-variance quantum estimator of $x(j)$ without regard to past data is linear with respect to $B(j)$. As a result of Lemma 3 then, the optimum measurements that appear in the linear filter will be of the form

$$\hat{M}_j(A) = \hat{Z}_j(\beta_j^{-1}(A)), \qquad j = 0,1,\cdots,k, \tag{25}$$

where β_j: $\mathbb{R}^n \to \mathbb{R}^n$, $\beta_j(x) = B(j)x$, $B(j)$ as given by (16), and the outcome of $\hat{Z}_j$ is a minimum-variance estimator of $x(j)$ without regard to past data. Then (15) implies

$$E\{\hat{z}(j)|x(j)\} = B^{-1}(j)\Delta(j)x(j) \triangleq \Gamma(j)x(j). \tag{26}$$

Note that, if $\hat{v}(j)$ and $x(j)$ are jointly Gaussian, then $\hat{z}(j)$ and $x(j)$ are also, which allows us to pass the joint Gaussian assumption on the estimation problem for $x(j)$ directly. Under these assumptions, the optimal linear (filter) estimator takes the form

$$\hat{x}(k) = B(k)\hat{z}(k) + \sum_{j=0}^{k-1} \hat{C}_j(k)B(j)\hat{z}(j). \tag{27}$$

The normal equations for the $\hat{C}_j(k)$ (Theorem 2) then become

$$\sum_{j=0}^{k} B(i)E\{\hat{z}(i)\hat{z}(j)^t\}B(j)^t\hat{C}_j(k)^t = B(i)E\{\hat{z}(i)x(k)^t\},$$
$$i = 0,1,\cdots,k. \tag{28}$$

We observe that, without loss of generality, one can assume that the $B(j)$, $j = 0,1,\cdots,k$ are nonsingular. For, if any $B(j)$ is singular, it can be restricted to the complement of its null space without affecting either $\hat{x}(k)$ or the normal equations. Hence, the latter may be written

$$\sum_{j=0}^{k} E\{\hat{z}(i)\hat{z}(j)^t\}[\hat{C}_j(k)B(j)]^t = E\{\hat{z}(i)x(k)^t\},$$
$$i = 0,1,\cdots,k. \tag{29}$$

We summarize these results in the following.

Lemma 4: Under the hypotheses of Lemma 3 and the additional hypothesis that the minimum variance quantum estimator of $x(j)$ without regard to past data is linear with respect to $B(j)$, the optimal linear filtered estimate of $x(k)$ takes necessarily the form (27) and the normal equations (29) hold. Moreover, (26) holds.

We now turn to a converse of Lemma 4. This is of particular importance for the synthesis of the filter.

Lemma 5: Let the vector signal process $\{x(i)\}$ be pairwise Gaussian, $\hat{Z}_i$ be the optimal POM (with outcome $\{\hat{z}(i)\}$ such that $\hat{z}(i)$ and $x(i)$ are jointly Gaussian) without regard to past measurements, and $\{D_j(i), j = 0,1,\cdots,i\}$ solve the normal equations (29) based on the $\{\hat{z}(i)\}$. Assume that $\hat{z}(k)$ is linear with respect to $D_k(k)$ and set the POM $\hat{M}_i(A) \equiv \hat{Z}_i[d_i^{-1}(A)]$, for $A \in \mathcal{B}^n$, $i = 0,\cdots,k$, where $d_0(x) \equiv x$, $d_i(x) \equiv D_i(i)x$, $i = 1,\cdots,k$, with outcomes $\{\hat{v}(i), i = 0,1,\cdots,k\}$, and let $\{\hat{C}_j(i), j = 0,1,\cdots,i\}$ solve the normal equations (29) based on the $\{\hat{v}(i)\}$. Then, for each $i = 0,1,\cdots,k$, the POM $\hat{M}_i$ and the matrices $\{\hat{C}_j(i)\}$ are optimal.

Proof: We need to show that conditions i), ii), and iii) of Theorem 3 are satisfied by $\hat{M}_i$ and $\hat{C}(i)$ for each $i = 0,1,\cdots,k$. Now i) is satisfied by our construction of the $\{\hat{C}_j(i)\}$.

By the optimality of $\hat{Z}_i$ and Theorem 3, the operator-valued function

$$\mathcal{F}'(u;i) = \lambda(i) - 2u^t\delta(i) + \eta(i)u^tu \tag{30}$$

is integrable with respect to $\hat{Z}_i$ and

$$\eta(i)\left[u^tu - \int_{\mathbb{R}^n} u^tu\hat{Z}_i(du)\right] - 2\delta(i)^t\left[u - \int_{\mathbb{R}^n} u\hat{Z}_i(du)\right] \ge 0,$$
$$\text{for } u \in \mathbb{R}^n. \tag{31}$$

On the other hand, the hypotheses imply that there exist $n \times n$ matrices $A(j,i)$ and $\Gamma(i)$ such that

$$E\{x(j)|x(i)\} = A(j,i)x(i) \tag{32}$$

$$E\{\hat{z}(i)|x(i)\} = \Gamma(i)x(i), \tag{33}$$

and therefore, similarly to (19), we obtain $\gamma(i,j) = D_j(j)\Gamma(j)A(j,i)\delta(i)$. From (8) and (9),

$$\mathcal{F}(u, \hat{C}(i)) = \lambda(i) + \eta(i)u^t u - 2u^t \left(I_n - \sum_{j=1}^{i-1} \hat{C}_j(i)D_j(j)\Gamma(j)A(j,i)\right) \delta(i) + \phi, \quad (34)$$

where ϕ is an operator independent of u.

From the last block row of the normal equations for the $\hat{C}_j(i)$, we have

$$\sum_{j=1}^{i-1} E\{\hat{v}(i)\hat{v}(j)^t\}\hat{C}_j^t(i) + E\{\hat{v}(i)\hat{v}(i)^t\} = E\{\hat{v}(i)x(i)^t\} \quad (35)$$

or

$$\left[\sum_{j=1}^{i-1} \hat{C}_j(i)D_j(j)E\{\hat{z}(j)\hat{z}(i)^t\} + D_i(i)E\{\hat{z}(i)\hat{z}(i)^t\}\right] \cdot D_i^t(i) = E\{x(i)\hat{z}(i)^t\}D_i^t(i).$$

Now for similar reasons as those for the matrices $B(j)$ (see comments below (28) above), we can assume without loss of generality that $D_i(i)$ is nonsingular. Then using (32) we have

$$\sum_{j=1}^{i-1} \hat{C}_j(i)D_j(j)\Gamma(j)A(j,i)E\{x(i)\hat{z}(i)^t\} + D_i(i)E\{\hat{z}(i)\hat{z}(i)^t\} = E\{x(i)\hat{z}(i)^t\}$$

or

$$D_i(i)E\{\hat{z}(i)\hat{z}(i)^t\} = \left(I_n - \sum_{j=1}^{i-1} \hat{C}_j(i)D_j(j)\Gamma(j)A(j,i)\right) E\{x(i)\hat{z}(i)^t\}. \quad (36)$$

It is easy to see that the assumptions imply that $E\{x(i)\hat{z}(i)^t\} = E\{\hat{z}(i)\hat{z}(i)^t\}$, $i = 0,1,\cdots,k$. So, trivial cases apart, (36) implies

$$D_i(i) = I_n - \sum_{j=1}^{i-1} \hat{C}_j(i)D_j\Gamma(j)A(j,i). \quad (37)$$

Using (37) in (34), we obtain

$$\mathcal{F}(u, \hat{C}(i)) = \lambda(i) + \eta(i)u^t u - 2(D_i(i)\delta(i))^t u + \phi. \quad (38)$$

From (30), (38), and the definition of $\hat{\boldsymbol{Z}}_i$, $\hat{\boldsymbol{M}}_i$, it follows that $\mathcal{F}(\cdot, \hat{C}(i))$ is integrable with respect to $\hat{\boldsymbol{M}}_i$. So condition ii) of Theorem 3 is also satisfied.

Finally, since the minimum-variance quantum estimator of $x(i)$ without regard to past data is linear with respect to $D_i(i)$, we have, using (22), (23) and Theorem 3, that

$$\eta(i)\left[u^t u - \int_{\mathbb{R}^n} u^t u \hat{\boldsymbol{M}}_i(du)\right] - 2(D_i(i)\delta(i))^t \left[u - \int_{\mathbb{R}^n} \hat{\boldsymbol{M}}_i(du)\right] \geq 0, \quad \text{for } u \in \mathbb{R}^n. \quad (39)$$

But then (38) and (39) imply

$$\mathcal{F}(u, \hat{C}(i)) - \int_{\mathbb{R}^n} \mathcal{F}(u, \hat{C}(i))\hat{\boldsymbol{M}}_i(du) \geq 0, \quad \text{for } u \in \mathbb{R}^n,$$

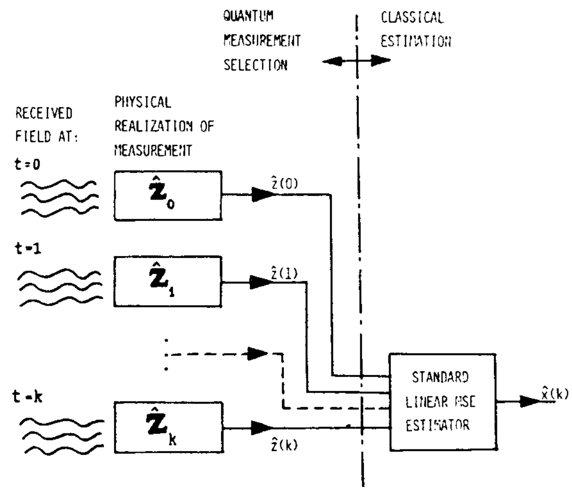

Fig. 2. Illustrating separation of optimal linear filter.

which is the condition iii) of Theorem 3. This completes the proof of the lemma.

We note that Lemma 5 is required because of the nonuniqueness of the optimal linear filter and of the optimal measurement without postprocessing.

From the recursions (16), (26), and (37), the matrices $D_i(i)$, $i = 0,1,\cdots,k$, of Lemma 5 are identical to the matrices $B(i)$, $i = 0,1,\cdots,k$, in Lemma 4. So the filtered estimate of $x(k)$ utilizing the POM's and processing matrices constructed in Lemma 5 is again given by (27), and the normal equations (29) hold. So we have established the following "separation" theorem.

Theorem 4: Under the hypothesis of Lemma 5, the outcomes $\hat{z}(j)$, $j = 0,1,\cdots,k$, of the measurements $\hat{\boldsymbol{Z}}_j$, $j = 0,1,\cdots,k$, are a sufficient statistic for the linear minimum-variance estimate (LMVE) $\hat{x}(k)$ of $x(k)$.

This theorem establishes, under the hypotheses stated, the following important "separation property" of the optimal linear quantum filter: *the optimal quantum measurements are chosen separately from the optimal (classical) linear postprocessing of the measurement outcomes.* This is illustrated in Fig. 2.

Of obvious interest are conditions under which the minimum-variance quantum estimator of $x(j)$ without regard to past data is linear with respect to the matrices $B(j)$. The following theorem provides a sufficient condition.

Lemma 6: If $B(j)$ as constructed above is of the form $B(j) = b(j)U(j)$, where $b(j)$ is a scalar $\neq 0$ and $U(j)$ an $n \times n$ orthogonal matrix, then the minimum-variance quantum estimator of $x(j)$ without regard to past data is linear with respect to $B(j)$.

Proof: Let $\hat{\boldsymbol{Z}}_j$ be the optimal POM for the estimation of $x(j)$ without postprocessing. Then (30) and (31) hold. Let $\boldsymbol{M}_j^*$ be defined via (25). For $\boldsymbol{M}_j^*$ to be the optimal POM for the estimation of $B(j)x(j)$ (without postprocessing), we must show that the operator-valued function (21) is integrable with respect to M_j^* and that

$$\Lambda \triangleq \eta^*(j)\left[u^t u - \int_{\mathbb{R}^n} u^t u \, M_j^*(du)\right] - 2\delta^*(j)^t\left[u - \int_{\mathbb{R}^n} u\, M_j^*(du)\right] \geq 0, \text{ for } u \in \mathbb{R}^n, \tag{40}$$

where $\eta^*(j) = \eta(j)$ and $\delta^*(j) = B(j)\delta(j)$ from (22), (23). The integrability of (21) with the respect to M_j^* is immediate. Utilizing the definition of M_j^* in (40), we have that

$$\Lambda = \eta(j)\left[u^t u - \int_{\mathbb{R}^n} u^t u\, \hat{Z}_j(B^{-1}(j)\,du)\right] - 2\delta^t(j)B^t(j)\left[u - \int_{\mathbb{R}^n} u\, \hat{Z}_j(B^{-1}(j)\,du)\right]$$
$$= \eta(j)\left[u^t u - b^2(j)\int_{\mathbb{R}^n} z^t z\, \hat{Z}_j(dz)\right] - 2\delta^t(j)\left[B(j)^t u - b(j)^2 \int_{\mathbb{R}^n} z\, \hat{Z}_j(dz)\right]$$

since $U(j)^t U(j) = I_n$. Making the substitution $u = B(j)z$, we have

$$\Lambda = b^2(j)\eta(j)\left[z^t z - \int_{\mathbb{R}^n} z^t z\, \hat{Z}_j(dz)\right] - 2b^2(j)\delta^t(j)\left[z - \int_{\mathbb{R}^n} z\, \hat{Z}_j(dz)\right] \geq 0, \quad \text{for } z \in \mathbb{R}^n$$

because of (31). Since $B(j)$ is clearly invertible, this completes the proof.

The following theorem gives sufficient conditions for $B(j)$ to be of the form described in Lemma 6.

Theorem 5: Let $\hat{Z}_i$ be the optimal POM estimating $x(i)$ without regard to past data and suppose that

i) $E[\hat{z}(i)|x(i)] = \gamma(i)x(i)$, $\gamma(i)$ a scalar,
ii) $E[\hat{z}(i)\hat{z}(i)^t] = \theta(i)I_n$, $\theta(i)$ a scalar, and
iii) the component Gaussian random processes $\{x_j(i)\}$, $j = 1, \cdots, n$, are uncorrelated and identically distributed.

Then $B(i) = b(i)I_n$.

Proof: Clearly $B(i) = b(i)I_n$, for $i = 0$. The defining equations for $B(i)$, $i > 0$, are the normal equations (29) or (30). We note that under the assumptions of the theorem, for $i > 0$, $E\{x(j)|x(i)\} = \alpha(j,i)x(i)$; $j = 0,1,\cdots,i-1$, where $\alpha(j,i)$ is a scalar, and that $E\{x(i)x(i)^t\} = \lambda(i)I_n$, $\lambda(i)$ a scalar. Therefore $E\{\hat{z}(j)x(i)^t\} = \gamma(j)\alpha(i,j)\lambda(j)I_n$, $j = 0,1,\cdots,i-1$, and $E\{\hat{z}(j)\hat{z}(i)^t\} = \gamma(i)\gamma(j)\alpha(i,j)\lambda(j)I_n$, $j = 0,1,\cdots,i-1$. Therefore, (29) yields

$$\begin{bmatrix} \theta(0)I_n & \gamma(0)\gamma(1)\alpha(1,0)\lambda(0)I_n & \cdots \gamma(0)\gamma(i)\alpha(i,0)\lambda(0)I_n \\ \gamma(0)\gamma(1)\alpha(1,0)\lambda(0)I_n & \theta(1)I_n & \cdots \gamma(1)\gamma(i)\alpha(i,1)\lambda(1)I_n \\ \vdots & \vdots & \vdots \\ \gamma(i)\gamma(0)\alpha(i,0)\lambda(0)I_n & \cdots & \theta(i)I_n \end{bmatrix} \cdot \begin{bmatrix} [\hat{C}_0(i)B(0)]^t \\ \vdots \\ [\hat{C}_{i-1}(i)B(i-1)]^t \\ B(i)^t \end{bmatrix} = \begin{bmatrix} \gamma(0)\lambda(0)\alpha(i,0)I_n \\ \gamma(1)\lambda(1)\alpha(i,1)I_n \\ \vdots \\ \gamma(i-1)\lambda(i-1)\alpha(i,i-1)I_n \\ \gamma(i)\lambda(i)I_n \end{bmatrix}.$$

Therefore, the vectors of off-diagonal elements of the matrices $[\hat{C}_0(i)B(0)]^t, \cdots, [C_{i-1}(i)B(i-1)]^t, B(i)^t$ satisfy the equation

$$\begin{bmatrix} \theta(0) & \gamma(0)\gamma(1)\alpha(1,0)\lambda(0) & \cdots \gamma(0)\gamma(i)\lambda(0)\alpha(i,0) \\ \gamma(0)\gamma(1)\alpha(1,0)\lambda(0) & \theta(1) & \cdots \gamma(1)\gamma(i)\lambda(1)\alpha(i,1) \\ \vdots & \vdots & \vdots \\ \gamma(i)\gamma(0)\alpha(i,0)\lambda(0) & \cdots & \theta(i) \end{bmatrix} \cdot \begin{bmatrix} [\hat{C}_0(i)B(0)]_{lm}^t \\ \vdots \\ [\hat{C}_{i-1}(i)B(i-1)]_{lm}^t \\ B(i)_{lm}^t \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \text{for } l \neq m. \tag{41}$$

Since, trivial cases apart, the matrix in (41) is nonsingular, the only solution is the zero-vector solution. Therefore, the matrices $\hat{C}_0(i)B(0), \cdots, \hat{C}_{i-1}(i)B(i)$, $B(i)$ are diagonal. Now, for $l = m$, the diagonal elements of these matrices satisfy, for every l, the same equation as (41) with the right side replaced by the column matrix $[\gamma(0)\lambda(0)\alpha(i,0), \cdots, \gamma(i-1)\lambda(i-1)\alpha(i,i-1), \gamma(i)\lambda(i)]^t$. Therefore, $B(i)$ is of the form $b(i)I_n$ as is, in fact, every matrix in sight. This completes the proof of the theorem.

The most obvious consequences of the separation theorem are a) when it holds, we need only implement the optimal measurements $\hat{Z}_j$ that should be thought of as intrinsic to the quantum field and can be found *a priori*, we have thus a considerable reduction on the number of measuring devices needed; b) when the physical implementation of the POM $\hat{Z}_j$ does not depend explicitly on j (i.e., on time) and the classical estimation problem (i.e., the normal equations (29)) admits a recursive solution, then the filter simplifies even further. Such an example follows.

Example: Suppose that $x(k)$ is a 2-vector signal process $(x_1(k),x_2(k))$, where $x_1(k)$ and $x_2(k)$ are independent zero-mean identically distributed Gaussian random sequences with variances $\lambda(k)$, and that $x(k)$ is transmitted as the in-phase $(x_1(k))$ and quadrature $(x_2(k))$ amplitudes of a monochromatic laser that is received, along with thermal noise, in a single-mode cavity upon which measurements can be made, $k = 0,1,\cdots$. The density operator in the coherent state representation (or P-representation) [21] is

$$\rho(x(k)) = \frac{1}{\pi n_0}\int e^{-|\alpha-x_1(k)-ix_2(k)|^2/n_0}|\alpha\rangle\langle\alpha|\, d^2\alpha,$$

where the coherent states $|\alpha\rangle$ [32] are eigenstates of the photon annihilation operator of the mode $\boldsymbol{a}$, n_0 is the mean number of thermal photons in the mode, and the integration is taken over the entire complex α-plane ($\alpha = \mathrm{Re}\,\alpha + i\,\mathrm{Im}\,\alpha$, $d^2\alpha = d(\mathrm{Re}\,\alpha)d(\mathrm{Im}\,\alpha)$).

It is known [6], [7], [22] that the POM defined by

$$\hat{Z}_k(A) = \int_A \left|\frac{\alpha}{d(k)}\right\rangle\left\langle\frac{\alpha}{d(k)}\right|\frac{d^2\alpha}{d^2(k)\pi}, \qquad \text{for } A \in \mathcal{B}^n \tag{42}$$

$d(k) \equiv 2\lambda(k)[n_0 + 2\lambda(k) + 1]^{-1}$, represents an optimal measurement for the minimum variance estimation of $x(k)$ without regard to past data. For completeness, we show that the above POM satisfies the necessary and sufficient conditions for optimality (30) and (31) from conditions ii) and iii) of Theorem 3. Using the properties of the coherent states, it is straightforward to show that $\eta(k) = [1 - \exp(-\sigma(k))]\exp(-\sigma(k)\boldsymbol{a}^\dagger\boldsymbol{a})$, where $\sigma(k) \equiv \ln e(k)$, $e(k) \equiv (n_0 + 2\lambda(k) + 1)\cdot(n_0 + 2\lambda(k))^{-1}$ and $\boldsymbol{a}^\dagger\boldsymbol{a}$ is the "number operator" [21]. Also $\delta_1(k) = \lambda(k)(n_0 + 2\lambda(k))^{-1}[\boldsymbol{a}\eta(k) + \eta(k)\boldsymbol{a}^\dagger]$ and $\delta_2(k) = -i\lambda(k)(\eta_0 + 2\lambda(k))^{-1}[\boldsymbol{a}\eta(k) - \eta(k)\boldsymbol{a}^\dagger]$. (Compare with [6, eqs. (6.16)–(6.18)]. To evaluate (33), we need the first and second moments of the POM defined in (42): we find, via the coherent state representation, $\boldsymbol{U}_1(k) = d(k)(\boldsymbol{a} + \boldsymbol{a}^\dagger)/2$, $\boldsymbol{U}_2(k) = d(k)(\boldsymbol{a} - \boldsymbol{a}^\dagger)/2i$, $\boldsymbol{U}_{11}(k) = \boldsymbol{U}_1^2 + d(k)^2\boldsymbol{I}/4$, $\boldsymbol{U}_{22}(k) = \boldsymbol{U}_2^2 + d^2(k)\boldsymbol{I}/4$, and $\boldsymbol{U}_{12}(k) = \boldsymbol{U}_{21}(k) = \boldsymbol{U}_1\boldsymbol{U}_2 - d(k)^2\boldsymbol{I}/4i$. It follows immediately that the operator-valued function in (32) is integrable with respect to $\hat{Z}_k$ in (42). Finally, to show (33) we must establish, for all complex scalars α, that

$$\eta(k)\alpha\bar{\alpha} - d^2(k)\eta(k)\boldsymbol{a}\boldsymbol{a}^\dagger + \frac{2\lambda(k)d(k)}{n_0 + 2\lambda(k)}\eta(k)\boldsymbol{a}^\dagger\boldsymbol{a} + \frac{2\lambda(k)d(k)}{n_0 + 2\lambda(k)}\boldsymbol{a}\eta(k)\boldsymbol{a}^\dagger - \frac{2\lambda(k)}{n_0 + 2\lambda(k)}\boldsymbol{a}\eta(k)\bar{\alpha} - \frac{2\lambda(k)}{n_0 + 2\lambda(k)}\eta(k)\boldsymbol{a}^\dagger\alpha \geq 0. \tag{43}$$

Working with the "number representation" [7] for $\boldsymbol{a}$, $\boldsymbol{a}^\dagger$ and $\eta(k)$, it is easy to establish that $\eta(k)\boldsymbol{a} = e(k)\boldsymbol{a}\eta(k)$ and $\boldsymbol{a}^\dagger\eta(k) = e(k)\eta(k)\boldsymbol{a}^\dagger$. Thus (43) demands that $(\bar{\alpha} - d(k)\boldsymbol{a}^\dagger)\eta(k)(\alpha - d(k)\boldsymbol{a})$ be nonnegative, which is true since $\eta(k)$ is nonnegative definite.

We proceed now with the solution to the filtering problem. From (42), we note that $\hat{Z}_k(A) = \Psi_k[d_k^{-1}(A)]$, $A \in \mathcal{B}^n$, where

$$\Psi_k(A) \equiv \int_A |\alpha\rangle\langle\alpha|\frac{d^2\alpha}{\pi} \tag{44}$$

and where the mapping $d_k\colon R^2 \to R^2$ is defined by $d_k(x) = d(k)x$, $x \in \mathbb{R}^2$. So if we let $y(k) \in \mathbb{R}^2$ denote the outcome of a measurement represented by the POM Ψ_k, we see that $\hat{z}(k) = d(k)y(k)$. Moreover, the first and second moments of Ψ_k are obtained from the moments calculated above by replacing $d(k)$ by unity. Now the probability density function of the outcome $y(k)$, conditioned on $x(k)$, is given by $\mathrm{Tr}\,[\rho(x(k))|\alpha\rangle\langle\alpha|/\pi]$: the calculation shows that $y_1(k)$ and $y_2(k)$, conditioned on $x(k)$, are uncorrelated Gaussian random variables with means $x_1(k)$ and $x_2(k)$, respectively, and identical variances $(n_0 + 1)/2$. The physical realization of the POM Ψ_k is optical heterodyning [8], [9], [34], [36] of the received field. It is known [22] that the measurement represented by the POM Ψ_k is "realized" (see Section II) by the simultaneous measurement of the commuting operators $[(\boldsymbol{a} + \boldsymbol{a}^\dagger)/2 - (\boldsymbol{a}_e + \boldsymbol{a}_e^\dagger)/2]$ and $[(\boldsymbol{a} - \boldsymbol{a}^\dagger)/2i + (\boldsymbol{a}_e - \boldsymbol{a}_e{}^\dagger)/2i]$ on the Hilbert space $H \otimes H_e$ representing the receiver cavity adjoined by an harmonic oscillator in the ground state $\rho_e = |0_e\rangle\langle 0_e|$.

A simple calculation shows that $E\{\hat{z}(k)\hat{z}(k)^t\} = (n_0 + 2\lambda(k) + 1)I_2/2$. Therefore, the hypotheses of Theorem 5 are satisfied and the separation theorem (Theorem 4) holds. It is clear from (49) and from the above discussion that the optimal measurement without postprocessing $\hat{Z}_k$ does not depend explicitly on k, a great practical advantage; the k-dependence is entirely accountable in the classical postprocessing.

Thus the optimal filtering estimator becomes

$$\hat{x}(k) = b(k)d(k)y(k) + \sum_{i=0}^{k-1}\hat{C}_i(k)b(i)d(i)y(i),$$

where the coefficient matrices $\{\hat{C}_0(k)b(0)d(0),\cdots,b(k)d(k)I_2\}$ satisfy the normal equations for the LMVE of $x(k)$ based on $y(0),y(1),\cdots,y(k)$.

The outcomes $y(i)$ are statistically equivalent to the following fictitious observation process

$$y(i) = x(i) + \epsilon(i), \qquad i = 0,1,\cdots,k, \tag{45}$$

where $\epsilon(i)$ is a white zero-mean Gaussian random vector sequence with covariance matrix $(n_0 + 1/2)I_2$.

Finally, if the $\{x(k)\}$ sequence satisfies the recursive relation (1), then the optimal estimator is given by the well-known Kalman–Bucy filtering equations [16, sect. 4.7] for the classical problem (1) and (45), namely,

$$\hat{x}(k) = \Phi(k-1)\hat{x}(k-1) + K(k)[y(k) - \Phi(k-1)\hat{x}(k-1)],$$

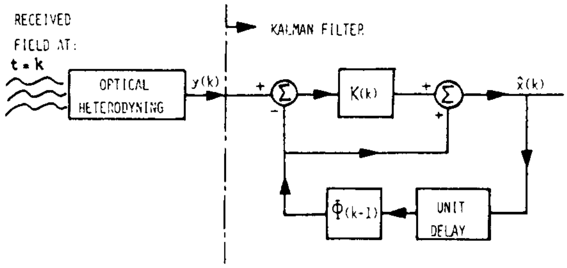

Fig. 3. Illustrating form of filter in this example.

where

$$K(k) = P(k)\left[P(k) + \frac{n_0 + 1}{2} I_2\right]^{-1}$$

and

$$P(k) = \Phi(k-1)[P(k-1) - K(k-1)P(k-1)]\Phi(k-1)^t + Q(k-1).$$

The realization of this computation is shown in Fig. 3.

REFERENCES

[1] A. S. Holevo, "Statistical problems in quantum physics," *Proc. of 1972 Soviet-Japanese Symposium on Probability and Statistics*, vol. 1, pp. 22–40, 1972.

[2] ——, "Statistical decision theory for quantum systems," *J. Multivariate Analysis*, vol. 3, pp. 337–394, 1973.

[3] C. W. Helstrom, "Detection theory and quantum mechanics," *Inform. Contr.*, vol. 10, pp. 254–291, Mar. 1967.

[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] J. S. Baras, R. O. Harger, and Y. H. Park, "Quantum mechanical linear filtering of random signal sequences," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 59–64, Jan. 1976.

[6] H. P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of non-selfadjoint observables," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 740–749, Nov. 1973.

[7] C. W. Helstrom and R. S. Kennedy, "Non-commuting observables in quantum detection and estimation theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 16–24, Jan. 1974.

[8] E. Prugovecki, *Quantum Mechanics in Hilbert Space*. New York: Academic, 1971.

[9] J. M. Jauch, *Foundations of Quantum Mechanics*. Reading, MA: Addison-Wesley, 1968.

[10] J. S. Baras and R. O. Harger, "Multiparameter filtering with quantum measurements, *Proc. of 1975 Conference on Information Sciences and Systems*, Johns Hopkins University, April 1975, pp. 178–184.

[11] ——, "Nonclassical filtering problems in quantum electronics," *Proc. of 1975 IEEE Conference on Decision and Control*, Houston, TX, Dec. 1975, pp. 804–806.

[12] J. S. Baras, "An optimization problem from linear filtering with quantum measurements," to appear.

[13] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*. A.M.S. Colloquium Publication, 1957.

[14] S. K. Berberian, *Notes on Spectral Theory*. New York: D. Van Nostrand, 1966.

[15] N. I. Akhiezer and I. M. Glazman, *Theory of Operators in Hilbert Space*, vol. II. New York: Frederick Ungar, 1963.

[16] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[17] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, CA: Stanford University Press, 1958.

[18] A. S. Holevo, "The theory of statistical decisions on an operator algebra," *Dokl. Akad. Nauk SSSR*, Tom 218, no. 1, pp. 1276–1281, 1974.

[19] ——, "Some statistical problems for quantum Gaussian states," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 533–543, Sept. 1975.

[20] C. W. Helstrom, "Cramér-Rao inequalities for operator-valued measures in quantum mechanics," *Int. J. Theor. Physics*, vol. 8, no. 5, pp. 361–376, 1973.

[21] R. J. Glauber, "Coherent and incoherent states of the radiation field," *Phys. Rev.*, vol. 131, pp. 2766–2788, Sept. 15, 1963.

[22] S. D. Personick, "An image-band interpretation of optical heterodyne noise," *Bell Syst. Tech. J.*, vol. 50, pp. 213–216, Jan. 1971.

[23] A. S. Holevo, *Proc. 4th International Symposium on Information Theory*, Leningrad 1976, part I, pp. 150–152.

[24] V. P. Belavkin, "Optimal linear randomized filtration of quantum boson signals," *Problems of Control and Information Theory*, vol. 3, no. 1, pp. 47–62, 1974.

[25] ——, *Proc. 4th International Symposium on Information Theory*, Leningrad 1976, part I, pp. 17–19.