INFORMATION THEORY AND REDUCED-ORDER FILTERING

by

John C. Doyle

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1977

Signature of Author ....
Department of Electrical Engineering and Computer Science
19 May 1977

Certified by ....
Sanjoy K. Mitter, Thesis Supervisor
Professor of Electrical Engineering

Accepted by ....
Chairman, Departmental Committee on Graduate Students

INFORMATION THEORY

AND

REDUCED-ORDER FILTERING

by

John C. Doyle

Submitted to the Department of Electrical Engineering and Computer Science on May 23, 1977 in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

The relationship between Shannon's Information Theory and filtering and estimation is examined and the intrinsic limitations to applying information and entropy analysis are discussed. Through the use of the functions of entropy and information, a novel reduced-order filter design procedure is presented.

Acknowledgements

I would like to thank all those whose suggestions aided the completion of this thesis, especially my advisor Professor Sanjoy K. Mitter.

Table of Contents

I. Introduction

The objective of this thesis is to investigate the relationship and practical relevance of Shannon's Information Theory to linear system theory and linear filtering. This relationship has long been questioned but until now there has been no fully satisfactory answer. In fact, the literature has often served to confuse rather than resolve the issues concerned.

The focus for the investigation is the Reduced-Order Filter Problem. This arises when some subset of the state of a linear dynamic system is to be estimated from measurements corrupted by additive white Gaussian noise. When there are no complexity constraints on the estimator, the optimal solution is the well-known Kalman filter ([1],[2]). In practice, however, the Kalman filter may present unacceptable on-line computational requirements or exhibit poor sensitivity to modeling errors and a suboptimal reduced-order filter may be sought.

While the reduced-order filter design procedure may utilize one of the approaches appearing in the literature (i.e. [3],[4],[5]), it usually requires considerable trial and error or excessive computation. Although no useful procedure is entirely automatic, a systematic design scheme is developed in this paper using entropy and information analysis.

While concepts from dynamic systems theory have been used in modeling elements of communication systems ([6]), little has been done in applying information theory to filter design. The bulk of the work has dealt with information properties of optimal least square (MSE) estimates ([7],[8],[9]), providing little insight into design or evaluation of practical estimators.

Zakai and Ziv [10] derive upper and lower bounds for the estimation error for certain diffusion processes using linear filtering considerations and information theoretic arguments. More recently, Bobrovsky and Zakai [11] present new lower bounds based on Van Trees' [12] version for the Cramer-Rao inequality. Weidemann and Stear [13] and Galdos-Gustafson [14] deal more directly with the filtering problem and their papers will be discussed in some detail in Section II.

The remainder of this thesis will be organized as follows. In Section II the intrinsic limitations involved in applying a noncausal information theory to dynamic problems are discussed. The pitfalls associated with a failure to appreciate these limitations will be exposed. With these considerations in mind, a novel procedure for reduced-order filter design is developed using information theoretic concepts in Section III. The range of applicability of the design procedure is discussed.

Because the concepts discussed here are basically quite simple, care has been taken to avoid unnecessary complication. The appendices contain the important relations used.

II. Information Theory and the Reduced-Order Filter Problem

The purpose of this section is to examine the limitations in applying information theory to the Reduced-Order Filtering Problem. The basic difficulty is that Shannon's Information Theory is noncausal and must be severely strained in order to have relevance to dynamic or causal problems.

The foundations of information theory were laid by C.E. Shannon in his celebrated journal article [15], "A Mathematical Theory of Communication". The concepts of entropy and mutual information have been extremely useful and effective in analysis of communication systems, but relatively few conclusive results have been obtained in applying these concepts to estimation or control problems. Of course, this is no surprise to those familiar with these areas.

Two papers which have attempted to apply information theoretic concepts to estimation are by Weidemann and Stear [13] (also Weidmann [16] and Galdos and Gustafson [14]). Weidemann and Stear, based on the work of Goblick [17], use entropy of the estimation error as a criterion function for analyzing estimating systems. Unfortunately, the estimating system analyzed was not causal and their results were existence theorems analogous to the coding theorem. Although of little practical relevance, the paper is interesting in that it proposes the use of entropy analysis, so successful for communication systems, in estimation. Although they point out certain advantages for entropy over a mean square error criterion, in the linear Gaussian case the advantages are not obvious. In fact, entropy reduces to MSE in the scalar case. Entropy analysis will form a basis for the design procedure developed in Section III.

An analysis of the recent paper by Galdos and Gustafson reveals some of the pitfalls in applying Information theory to filtering. They attempt to show "how information theory concepts can be modified to realistically imbed the Reduced-Order Filter Problem into an information framework which allows <u>systematic</u> filter <u>design</u>, <u>evaluation</u>, and <u>comparison</u>". (Emphasis in original.)

Galdos and Gustafson develop some machinery related to the rate distortion function[1], prove several theorems, and finally propose a two step design procedure. The two step procedure involves first maximizing the information in the estimate (regardless of error) until the information content is sufficient to guarantee a desired performance. The second step is simply adjusting certain parameters (which do not effect information) to reduce error by using all the information in the estimate. This decoupling is claimed to reduce the computational burden. The question is: compared to what?

The problem that Galdos and Gustafson are solving is exactly a static estimation problem at each time instant. That is, given estimates at time k and measurements at time k+1, what is the best estimate at time k+1? This procedure is only meaningful if there are either no dynamics or no information loss. It is easy to construct examples where using static estimation at each step results in a poor filter design. If there is no information loss, the filter will be the full-dimension Kalman filter.

Furthermore, the answer may be found by simply computing $E(x_{k+1}|\hat{x}_k, y_{k+1})$ using the appropriate convariance matrices, which are needed for Galdos-Gustafson's approach anyway. A two-step optimization represents an unnecessary

---

1. See Berger [18].

complication and greatly increases the computational burden.

Even though Galdos-Gustafson's procedure is useless in a practical sense, it appears initially to be interesting in that the problem can be decomposed into information maximization followed by error minimization. However, this property is a trivial consequence of the definitions of information and entropy. The relevant equations are developed in Appendix B and used in the design procedure is Section III.

The limitations in the applicability of information theory to filtering are far more fundamental than Weidemann-Stear or Galdos-Gustafson let on. To see this, consider a source modeled as a discrete-time system driven by noise. Suppose that a subset of the state of the system is to be estimated from noisy measurements. Each measurement provides information[1] about the state of the system and the optimal filter will keep all the information. For some systems (i.e. linear with Gaussian statistics) the optimal filter may be realized with finite memory, but in general it will require infinite memory. The problem is to selectively discard or save the information to reduce the memory requirements.

Some information will be kept in order to achieve an accurate estimate now, while some information will be saved that may be used for future estimates. However, in order to make the decision as to what information will be kept, a full knowledge is needed of what information will be available in the future. This means that in order to apply information theoretic tools, the joint probabilities of all future system states and measurements and current filter states must be computed.[2] Except in certain special cases

---

1. The term is being used very loosely here.
2. An example of this is in Berger [18], Chapter 4.

this is impossible. Finding the optimal finite memory filter is intrinsically a two-point boundary value problem [19].

In order to apply the concepts of information and entropy to filtering, the dynamic nature of the problem must be removed. Galdos and Gustafson did this by solving a static estimation problem at each time step but thereby made the solution not meaningful.

An important subclass of filtering problems occurs when the system has reached a statistical steady-state. In the following section this problem is treated in detail and a formulation developed which allows the application of information and entropy.

It should be noted that the introduction of more advanced concepts such as rate distortion theory will not change the noncausal nature of the theory and generally only serve to complicate and cloud the issues.

III. Steady-State Reduced-Order Filter Design

Consider the n^{th} order Gauss-Markov process modeled by the dynamic system

$$x(k+1) = \Phi\, x(k) + w(k+1) \qquad k = ,0,1,\ldots$$
$$y(k+1) = H\, x(k+1) + v(k+1) \tag{1}$$

where:

$$\dim\{x\} = \dim\{w\} = n$$
$$\dim\{y\} = \dim\{v\} = n_0$$

and w and v are zero-mean uncorrelated Gaussian white sequences with covariances Q and R respectively.[1]

The problem is to design a linear time-invariant causal filter of the form

$$z(k+1) = Az(k) + By(k+1)$$
$$r(k+1) = Cz(k+1) \tag{2}$$
$$\varepsilon(k+1) = Tx(k+1) - r(k+1)$$

where:

$$\dim\{z\} = m_1$$
$$\dim\{r\} = m_2 < m_1$$
$$C = m_2 \times m_1 \text{ matrix}$$
$$T = m_2 \times n \text{ matrix}$$

The quantities in the filter equations are the filter states z, the estimate r, and the error ε.

---

1. When no time-reference for a variable is given, it implies an arbitrary time instant or steady-state condition. While this involves some notational sloppiness, no confusion should arise.

Any filter of this type belongs to an equivalence class of filters which have the same error statistics but differ by a nonsingular state transformation. The performance measure for the filter is $h(\varepsilon)$, the error entropy in the steady state.[1] There may be several design objectives, such as minimizing $h(\varepsilon)$ for a given filter size, or minimizing filter order (m_1) given a maximum tolerable $h(\varepsilon)$. Of course, it may be that the filter designer has no prespecified criteria but seeks to find a reasonable tradeoff between filter order and error entropy.

Without loss of generality, suppose

$$T = [I_{m_2 \times m_2} \vdots 0_{m_2 \times (n-m_2)}]$$

and

$$x = \begin{bmatrix} \bar{x} \\ \tilde{x} \end{bmatrix}, \ \dim(\bar{x}) = m_2, \ \dim(\tilde{x}) = n-m_2.$$

That is, some subset $\bar{x}$ of the full state x is to be estimated.

Suppose some system (1) is in the statistical steady-state with some filter(2)[2]. Recognizing

$$h(\varepsilon,z) = h(\bar{x}-r,z) = h(\bar{x},z) \qquad (3)$$

gives

$$\begin{aligned} h(\varepsilon) &= -h(z) + h(\varepsilon,z) + I(\varepsilon;z) \\ &= h(\bar{x}) - h(\bar{x}) - h(z) + h(\bar{x},z) + I(\varepsilon;z) \qquad (4) \\ &= h(\bar{x}) - I(\bar{x};z) + I(\varepsilon;z) \end{aligned}$$

If $r = E(\bar{x}|z)$ then $I(\varepsilon;z) = 0$ and $h(\varepsilon) = h(\bar{x}) - I(\bar{x};z)$.[3] (5)

---

1. See Appendix A for definition of entropy and information.
2. See Appendix B for details of following derivations.
3. The same result may be found using rate distoriton theory (see Berger [18]), but the added complication is unnecessary. It should be noted that (5) implies that the Shannon Lower Bound may be achieved for the sensor plus filter.

Since $h(\bar{x})$ is fixed by (1), the error entropy depends only on $I(\bar{x};z)$. So in order to minimize error entropy we:

(1) Maximize $I(\bar{x};z)$ by choice of A and B

(2) Pick C such that $r = C\,z = E[\bar{x}|z]$ so that $I(\varepsilon;z) = 0$. This assures that all the information about $\bar{x}$ in z is used.

Step (2) is a straightforward static estimation problem, so attention may be focused on step (1). From Appendix A

$$I(\bar{x};z) = \frac{1}{2}\log \frac{|\Lambda_z|\;|\Lambda_{\bar{x}}|}{\begin{vmatrix}\Lambda_{\bar{x}} & \Lambda_{\bar{x}z}\\ \Lambda_{\bar{x}z}^T & \Lambda_z\end{vmatrix}}$$

$$= \frac{1}{2}\log \frac{|\Lambda_z|}{|\Lambda_z - \Lambda_{\bar{x}z}^T\Lambda_{\bar{x}}\Lambda_{\bar{x}z}|}$$

$$= \frac{1}{2}\log \frac{|\Lambda_{\bar{x}}|}{|\Lambda_{\bar{x}} - \Lambda_{\bar{x}z}\,\Lambda_z^{-1}\,\Lambda_{\bar{x}z}^t|}\;. \tag{6}$$

Here $\Lambda_z = E(zz^T)$ and $\Lambda_{\bar{x}z} = E(\bar{x}z^T)$, and $\Lambda_{\bar{x}}$ and Λ_z are positive definite.

If the filter order is specified, minimizing $h(\varepsilon)$ is equivalent to selecting A and B in (4) to minimize

$$|\Lambda_{\bar{x}} - \Lambda_{\bar{x}z}\,\Lambda_z^{-1}\Lambda_{\bar{x}z}|$$

where $\Lambda_{\bar{x}}$ is known from

$$\Lambda_x - \Phi\,\Lambda_x\,\Phi' - Q = 0 \tag{7}$$

and

$$\Lambda_{xz} - \Phi\,\Lambda_{xz}\,A' = \phi(I+\Lambda_x)\,H'B' \tag{8}$$

$$\Lambda_z = A\,\Lambda_z\,A' = BH\,\Lambda_x H'B' + (BH\Lambda_{xz}A' + A\Lambda_{xz}'\,H'B') \tag{9}$$

The matrices $\Lambda_{\bar{x}}$ and $\Lambda_{\bar{x}z}$ may be found by taking the appropriate submatrices from Λ_x and Λ_{xz}.

An optimization of this type may be very difficult computationally. Since this is not in general a convex problem, it is important to begin with an initial guess which is stable and near the optimum. Furthermore, there may not be a prespecified desired filter order.

In view of these considerations, it would be desirable to have a design procedure for reducing the order of the filter without involving such a computational burden. Suppose the steady-state Kalman filter has been computed:

$$\hat{x}_{k+1} = \hat{A}\,\hat{x}_k + \hat{B}\,y_{k+1} \tag{10}$$

States may not simply be removed since generally there is coupling between the states. This coupling prevents isolation of the information contribution from each component. In order to apply information theoretic concepts the problem must be made noncausal. This may be achieved by nonsingular transformation of the filter states so that they are mutually decoupled. For simplicity, suppose the eigenvalues of $\hat{A}$ are distinct.[1] Then let

$$z = P\,\hat{x}$$

so

$$z_{k+1} = P\,\hat{A}\,P^{-1}\,z_k + P\,\hat{B}\,y_{k+1}$$

$$= A\,z_k + B\,y_{k+1}$$

where $A = P\,\hat{A}\,P^{-1}$ is diagonalized.

---

1. If there are multiple roots, the Jordan form must be used throughout. While this involves minor additional complexity, it does not change the basic approach.

Since each mode, a component of z (or pair of components for complex eigenvalues), is decoupled from the other components, the information contained in each mode is restricted dynamically to that mode. Modal reduction of the filter may be performed with a clear measure of information loss.

From (6), note that once the matrices

$$\Lambda_z$$

and $\Lambda_{z|\bar{x}} = \Lambda_z - \Lambda_{\bar{x}z}' \Lambda_{\bar{x}}^{-1} \Lambda_{\bar{x}z}$

have been computed, it is a simple matter to compute the mutual information $I(\bar{x};z)$ for any subvector $\bar{z}$ of z, since

$$I(\bar{x};z) = \frac{1}{2} \log \frac{|\Lambda_{\bar{z}}|}{|\Lambda_{\bar{z}}|\bar{x}|} .$$

The matrices $\Lambda_{\bar{z}}$ and $\Lambda_{\bar{z}|\bar{x}}$ are found from Λ_z and $\Lambda_{z|\bar{x}}$ by deleting the rows and columns corresponding to the deleted modes. The m^{th} order subvector $\bar{z}$ of z which maximizes $I(\bar{x};\bar{z})$ (and consequently minimizes $h(\varepsilon)$) is that which maximizes the ratios of the corresponding principle minors of Λ_z and $\Lambda_{z|\bar{x}}$.

While this method should be most helpful for large systems, because of the dimension of the covariance matrices involved, it may be impossible to compute all minors of a given order. To avoid this, filter states may be removed individually or multiply until the desired order versus information tradeoff is achieved. Of course, removal of m states one at a time where each one deleted minimizes the one--state information loss will not in general result in the removal of the m least important states. While no a priori bounds are available, degradation due to state-by-state reduction should be small provided the information loss for each state removed is small.

To reiterate, given a diagonalized filter as in (11), the suggested procedure involves three techniques:

(i) Modal reduction by deleting filter states while minimizing information loss.

(ii) Adjustment of filter parameters by performing optimization using result from (1) as initial guess.

Steps (i) and (ii) may be repeated until a satisfactory design is achieved. Step (ii) may not be necessary if step (i) is successful.

(iii) Pick C in (2) so that $r = Cz = E(\bar{x}|z)$ where z is the vector of filter states and r is the estimate of $\bar{x}$. A nonsingular transformation of filter states may be performed if desired such that the first m_2 filter states[1] are exactly the estimates for $\bar{x}$.

Certain remarks may be made at this point:

(1) This design procedure has a pleasing interpretation in terms of familiar frequency domain concepts. The modal reduction involves deleting poles which have, in some sense, a small effect on the filter shape. Step (iii) involves picking the zeroes such that the resulting filter is as close a fit as possible to the original filter. The optimization in step (ii) adjusts the reduced order filter poles to further improve the fit. The suggested procedure may be thought of as a generalized pole-zero cancellation. In this light, it is interesting to note the complete decoupling of the procedure into steps (i) and (ii), followed by step (iii).

---

1. Recall $\dim(\bar{x}) = m_2$.

(2) The measurement vector y may be used nondynamically to improve the error entropy without changing the design procedure significantly. Then the output of the filter is given by

$$r = C_1 z + C_2 y = E(\bar{x}|y,z).$$

Details are given in Appendix C.

(3) The concepts of entropy and information were useful in developing the suggested procedure because the dynamic and causal nature of the estimation problem was suppressed.

(4) Equation (5) implies the notion of relative equality of information. That is, it does not matter how the information about $\bar{x}$ is gotten, from past, current or even future measurements. It may initially seem surprising that (5) holds when the sensor is unavailable for design. However, eq. 5 is merely a relation between covariance matrices for what amounts to static variables. It is equivalent to

$$\log |\Lambda_{\bar{x}|z}| = \log |\Lambda_{\bar{x}}| - \log \frac{|\Lambda_{\bar{x}}|}{|\Lambda_{\bar{x}|z}|}$$

which is, of course, true for any nonsingular matrices $\Lambda_{\bar{x}}$ and $\Lambda_{\bar{x}|z}$. Because the problem is nondynamic, the issue of fixed sensors is circumvented.

(5) Since the problem treated is in the steady-state, the entire procedure applies directly to continuous-time problems. Only equations (7), (8), and (9) need be changed (along with the system and filter equations). In fact, the procedure is applicable to any Gaussian steady-state or static estimation problem where a reduced filter is sought.

(6) There are serious questions concerning the appropriateness of entropy as a performance measure. In the case considered here, error entropy $h(\varepsilon)$

reduces to the determinant of the error covariance matrix Λ_ε, whereas MSE involves minimizing a trace function of Λ_ε. There previously have been no design procedures developed using a determinant measure so it is difficult a priori to determine the impact of this new approach. It is interesting to note, however, that Bucy [7] has shown that a necessary condition for optimality of a MSE estimator is that it maximize the mutual information between the system and filter states. This is equivalent to minimizing error entropy.

(7) It must be emphasized that the development in this section had very little to do with information theory as it is generally applied to communication systems. Only the entropy and information functions were borrowed.

(8) The apparent success found from using the information and entropy functions in estimation might suggest their applicability to control problems. Weidemann [6], using entropy of the state vector as a performance criteria, derived an equation for the control problem analogous to (4). However, as he pointed out, the equations for the control problem required a measure of mutual information between open and closed loop quantities, which is impossible.

Furthermore, the technique of modal reduction is inapplicable because due to feedback, the modes of the filter cannot be decoupled. The concepts of entropy and information seem applicable only to the situation where some linear function of the state is to be estimated for a system in the steady-state. This linear function could be a control signal and modal reduction of the compensator could be performed open-loop. This would give no guarantee of closed-loop stability, but the procedure might be useful in some instances.

Appendix A[1]

The most often used information quantity is the entropy function h(X), which, for the n-dimensional random vector X with continuous density function $p_x(X)$ is defined as

$$h(X) \triangleq -\int_{-\infty}^{\infty} dX\, p_x(X) \log_e p_x(X)$$

The entropy function gives a measure of the "randomness" or spread of the random vector X and is analogous to variance. In fact, for the vector X having the Gaussian probability density

$$p(X) = [(2\pi)^n \quad |R|]^{-1/2} \quad \exp(-1/2\ X^T R^{-1} X)$$

where R is the covariance matrix, the total entropy is directly related to the variance, i.e.,

$$h(X) = \frac{1}{2} \log\ [(2\pi e)^n \ |R|].$$

Of course, entropy is a relative measure depending on the coordinate system, but this does not restrict its usefulness since results always depend on the difference between two entropies. The terms related to coordinate system thus will cancel.

Other related entropy and information functions associated with a pair of vectors X and Y which posses continuous joint and marginal densities are

(1) the joint entropy

$$h(X,Y) \triangleq -\int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY\ p(X,Y) \log p(X,Y)$$

(2) the conditional entropy

$$h(X|Y) \triangleq -\int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY\ p\ (X,Y) \log p(X|Y)$$

---

1. This discussion is drawn from [16], [20], and [21].

(3) the mutual information

$$I(X;Y) = \int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dY \; p(X,Y) \; \log \frac{p(X,Y)}{p_x(X)p_y(Y)}$$

Some useful properties of these functions are:

(1) $I(X;Y) = h(X) + h(Y) - h(X,Y)$

$$= h(X) - h(X|Y) = h(Y) - h(Y|X)$$

$$\geq 0$$

(2) $I(X;Y) \geq I(X;AY)$

Where A is a linear transformation. Equality holds when A is non-singular.

(3) If X and Y are jointly normal with covariance

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

where $A = E[XX^T]$, $B = E[XY^T]$ and $C = E[YY^T]$, then

$$I(X;Y) = \frac{1}{2} \log \frac{|A| \; |B|}{\begin{vmatrix} A & B \\ B^T & C \end{vmatrix}}$$

$$= \frac{1}{2} \log \frac{|A|}{|C - B^T A^{-1} B|}$$

$$= \frac{1}{2} \log \frac{|C|}{|A - BC^{-1} B^T|}$$

(4) The information I(X;Y) is mutual. It is equally a measure of the information about X in Y and about Y in X.

$$I(X;Y) = I(Y;X).$$

Appendix B

There are certain relationships between entropy and information which play a fundamental role in the application of these concepts to estimation. The most important of these will be developed and discussed below.

Consider two continuous random vectors X and Z with joint density $p_{xz}(X,Z)$. Suppose an estimate $\hat{x} = F(z)$ of x given z is desired, where the performance measure to be minimized is the entropy of the error vector $\varepsilon = \hat{x} - x$.

Since

$$p_{xz}(X,Z) = p_{\varepsilon z}(X - F(Z),Z)$$
$$= p_{\varepsilon z}(X - \hat{X},Z)$$

and therefore, from the definition of joint entropy

$$h(\varepsilon,z) = h(x,z).$$

From the identity between entropy and information

$$I(\varepsilon;z) = h(\varepsilon) + h(z) - h(\varepsilon,z)$$
$$= h(\varepsilon) - h(x) + h(x) + h(z) - h(x,z)$$
$$= h(\varepsilon) - h(x) + I(x;z)$$

Rearranging,

$$h(\varepsilon) = h(x) - I(x;z) + I(\varepsilon;z)$$

Now, suppose that the joint density

$$p_{xz}(X,Z) = p_{z|x}(Z|X)\ p_y(X)$$

is not entirely specified; that p_x is fixed but that $p_{z|x}$ is variable. This is analogous to building an estimator where the conditional density $p_{z|x}$ of the random vector from which the estimate $\hat{x} = F(z)$ is to be generated is available for design. (Within constraints imposed by the particular problem.)

Equation (1) holds for any choice of $p_{z|x}$ and F and can be used to provide insight into their selection. The first term on the right hand side is just the a priori entropy of x.

The other terms indicate that the error entropy is proportionally reduced by the amount of information contained in z about x and increased by the amount contained in z about ε. This implies that $p_{z|x}$ and F may be designed separately. That is, $p_{z|x}$ is chosen so as to maximize $I(\hat{x};z)$. Then F is chosen so that all the information in z about x is used in the estimate. When ε and z are independent

$$I(\varepsilon;z) = 0 .$$

For jointly Gaussian variables this simply means letting $\hat{x} = E(x|z)$.

This decoupling of the "estimation" problem is a natural consequence of the definitions of entropy and information.[1] It will not generally be too helpful since the problem can usually be solved directly much more easily.[2] However, as shown in Section III, in some cases this decoupling leads to a useful design procedure.

Note that this develoment has dealt only with random vectors and no references were made to sources or dynamical systems.

---

1. Weidemann [16] and Weidemann-Stear [13] derive a similar equation.
2. As in the case of Galdos-Gustafson [14].

Appendix C

The measurement vector y may be used nondynamically to reduce the error entropy without changing significantly the design procedure developed in Section III. The relevant equation is

$$h(\varepsilon) = h(\bar{x}) - I((y,z);\varepsilon) + I((y,z);\varepsilon)$$

The last term may be made zero by letting

$$r = c_1 z + c_2 y = E(\bar{x}|y,z).$$

The second term is found from

$$I((y,z);\bar{x})$$

$$= \frac{1}{2}\log \frac{|\Lambda_{\bar{x}}| \begin{vmatrix} \Lambda_y & \Lambda_{yz} \\ \Lambda_{yz}^T & \Lambda_z \end{vmatrix}}{\begin{vmatrix} \Lambda_{\bar{x}} & \Lambda_{\bar{x}y} & \Lambda_{\bar{x}z} \\ \Lambda_{\bar{x}y}' & \Lambda_y & \Lambda_{yz} \\ \Lambda_{\bar{x}z}' & \Lambda_{yz}' & \Lambda_z \end{vmatrix}}$$

$$= \frac{1}{2}\log \frac{|\Lambda_y|\ |\Lambda_z - \Lambda_{yz}'\Lambda_y^{-1}\Lambda_{yz}|}{|\theta_y|\ |\theta_z - \theta_{yz}'\theta_y^{-1}\theta_{yz}|}$$

where

$$\theta_y = \Lambda_y - \Lambda_{\bar{x}y}'\Lambda_{\bar{x}}^{-1}\Lambda_{\bar{x}y}$$

$$\theta_{yz} = \Lambda_{yz} - \Lambda_{\bar{x}y}'\Lambda_{\bar{x}}^{-1}\Lambda_{\bar{x}z}$$

$$\theta_z = \Lambda_z - \Lambda_{\bar{x}z}'\Lambda_{\bar{x}}^{-1}\Lambda_{\bar{x}z}.$$

So the procedure developed in Section III will carry through exactly with only the numerical matrices involved being changed.

References

1. R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, J. Basic Eng., Vol. 82, pp. 34-45, March 1960.
2. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Eng., Vol. 83, pp. 95-108, March 1971.
3. A.R. Stubberud and D.A. Wismer, "Suboptimum Kalman Filtering Techniques", in C. Leondes, ed., Theory and Applications of Kalman Filtering, Agardograph, February 1970.
4. C.G. Sims, "An Algorithm for Estimating a Portion of the State Vector", IEEE Trans. Aut. Con., Vol. AC-19, pp. 341-393, August 1974.
5. A. Gelb, ed., Applied Optimal Estimation, The MIT Press, 1974.
6. W. Toms and T. Berger, "Information Rates of Stochastically Driven Dynamic Systems", IEEE Trans. Info. Th., Vol. IT-17, pp. 113-114, January 1971.
7. R.S. Bucy and P.D. Joseph, Filtering for Stochastic Processes with Application to Guidance, Wiley, 1968.
8. T. Duncan, "On the Calculation of Mutual Information", SIAM J. Appl. Math., Vol. 19, pp. 215-220, July 1970.
9. T. Kailath, "The Innovations Approach to Detection and Estimation Theory", Proc. IEEE, Vol. 58, No. 5, pp. 680-695, May 1970.
10. M. Zakai and J. Ziv, "Lower and Upper Bounds on Optimal Filtering Error on Certain Diffusion Processes", IEEE Trans. Info. Th., Vol. IT-18, pp. 325-331, May 1972.
11. B.Z. Bobrousky and M. Zakai, "A Lower Bound on the Estimation Error for Markov Processes", IEEE Trans. Aut. Cont., Vol. AC-19, pp. 785-788, December 1975.

12. H.L. Van Trees, Detection, Estimation and Modulation Theory, Pt. 1, New York: Wiley, 1968.

13. H.L. Weidemann and E.B. Stear, "Entropy Analysis of Estimating Systems", IEEE Trans. Info. Th., Vol. IT-16, pp. 264-270, May 1970.

14. J.I. Galdos and D.E. Gustafson, "Information and Reduced-Order Filter Design", IEEE Trans. Inf. Th., Vol. IT-23, pp. 183-194, March 1977.

15. C.E. Shannon, "A Mathematical Theory of Communication", Bell Syst. Tech. O., Vol. 27, pp. 379-423, 623-656, July and October 1948.

16. H.L. Weidemann, "Entropy Analysis of Feedback Control Systems", in C. Leondes, ed., Advances in Control, Vol. 17, pp. 225-255, 1969.

17. T.J. Goblick, "Theoretical Limitations on the Transmission of Data From Analog Sources", IEEE Trans. Inf. Th., Vol. IT-11, pp. 558-567, October 1965.

18. T. Berger, Rate Distortion Theory, Prentice-Hall, 1971.

19. D.E. Johansen, "Optimal Control of Linear Stochastic Systems with Complexity Constraints", in C. Leondes, ed., Advances in Control, Academic Press, Vol. 4, pp. 182-278, 1966.

20. R.G. Gallager, Information Theory and Reliable Communication, Wiley, 1968.

21. I.M. Gelfand and A.M. Yaglom, "Calculation of the Amount of Information About a Random Function Contained in Another Such Function", Am. Math. Soc. Translations, Providence, R.I., Series 2, Vol. 12, 1959.