Some Engineering Applications of Matrix Spectral Factorization

Prepared by
T.H. Kerr
Date: 16 March 1972

Table of Contents

Impact of Matrix Spectral Factorization

1. Matched filtering problem (vector case).
2. Putting a problem having colored system or measurement noise into the proposed form for applying Kalman-Bucy filtering.
3. A general synthesis technique for linear multiport networks.
4. Inability to perform matrix spectral factorization is precisely what limited the Wiener filtering theory to the scalar case.
5. Modeling a black box having multiple inputs and multiple outputs (also includes a hypothesis test to determine if the proposed mathematical model is adequate).

Item 2 may be performed for the discrete systems also by making use of a simple trick involving a bilinear transformation.

The factorization will be of the form $S(p) = W^T(-p)W(p)$, where

(a) the matrix $W(p)$ has elements that are ratios of polynomials in p,

(b) each element of $W(p)$ is analytic in the right half p-plane, $Re(p) > -\mu$, $\mu > 0$,

(c) each element of $W^{-1}(p)$ is analytic in the right half p-plane, $Re(p) > -\mu$, $\mu > 0$,

(d) the matrix $W(p)$ is real, (i.e., $W(p) = W(\overline{p})$) when $S(p)$ is real.

1 Applications of Matrix Spectral Factorization in Communications

(Specifically, applying matrix spectral factorization to the n-dimensional matched filtering problem having correlated colored noise)

Prepared by

T. H. Kerr

Date: 29 Dec 71

Approach:

(1) review the solution of the one-dimensional matched filter problem,*

(2) significance of using white noise,

(3) how procedure is altered (pre-whitening) if the noise is colored,

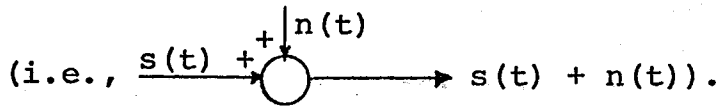
(4) how procedure is altered for the n-dimensional problem (pre-whitening) when the noise is colored and cross correlated. (This is where matrix spectral factorization is used.)

---

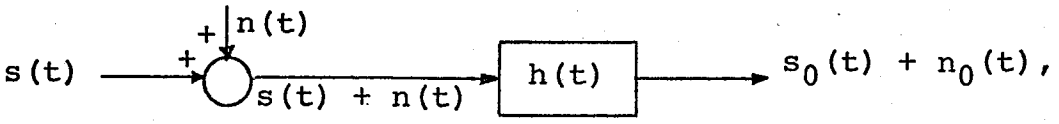
*An obvious application of matched filtering is in radar.

One Dimensional Matched Filtering

Given a deterministic signal s(t) corresponding to a particular message, and noise n(t), where the effect of transmitting the signal through a communications channel is as though the noise is added to it



If at the receiving end of the channel we have a linear filter having impulse response h(t) as shown in the figure below



and we define the output signal-to-noise ratio at time t as

$Z(t) = \dfrac{s_0^2(t)}{n_0^2(t)}$, then we can pose the following problem:

Find the causal linear filter h(t) such that Z(t) is a minimum at time $\hat{t}$. This problem is <u>most easily</u> solved if the input noise is white. (Even if the input noise is not white the problem may still be solved; the trick of pre-whitening may be used to an advantage.)

Theoretical solution to the one dimensional matched filtering problem.

n(t) is white; $S_{nn}(p) = G_0 \Leftrightarrow R_{nn}(\tau) = G_0 \delta(t)$

$$R_{n_0 n_0}(\tau) = \int_{-\infty}^{\infty} h(\tau_1) \int_{-\infty}^{\infty} h(\tau_2) R_{nn}(\tau-\tau_2+\tau_1)\, d\tau_2 d\tau_1$$

$$E[n_0^2(t)] = R_{n_0 n_0}(0) = \int_{-\infty}^{\infty} h(\tau_1) \int_{-\infty}^{\infty} h(\tau_2) G_0 \delta(0-\tau_2+\tau_1)\, d\tau_2 d\tau_1$$

$$= \int_{-\infty}^{\infty} h(\tau_1) h(\tau_1) G_0 d\tau_1 = G_0 \int_{-\infty}^{\infty} h^2(\tau_1) d\tau_1.$$

The output signal at a given time $\hat{t}$ is, by convolution,

$$s_0(\hat{t}) = \int_0^{\infty} s(\hat{t}-u) h(u)\, du.$$

"The problem is to maximize the function

$$Z_0(\hat{t}) = \frac{\left| \int_0^{\infty} s(\hat{t}-u) h(u)\, du \right|^2}{G_0 \int_{-\infty}^{\infty} h^2(\tau_1) d\tau_1}$$ by picking the proper h(t). For

a given $s_0(\hat{t})$, the maximum value of $Z_0(\hat{t})$ occurs when the output noise $\overline{n^2(t)}$ is a minimum.

Using variational methods, we then consider a function of the form

$$Q = \overline{n_0^2(t)} + \mu s_0(\hat{t})$$

where μ is the Lagrangian multiplier. The problem is now one of minimizing Q since $s_0(\hat{t})$ is considered constant. The function whose minimum is sought becomes

$$Q = \int_0^{\infty} [G_0 h^2(\lambda_1) + \mu s_i(\hat{t}-\lambda_1) h(\lambda_1)]\, d\lambda_1.$$

This function may be minimized by first rewriting it as follows:

$$Q = G_0 \int_0^{\infty} h(\lambda_1) [h(\lambda_1) + \frac{\mu}{G_0} s_i(\hat{t}-\lambda_1)]\, d\lambda_1.$$

This function is clearly a minimum when

$$h(\lambda_1) = -\frac{\mu}{G_0} s_i(\hat{t}-\lambda_1).$$

It now becomes evident that μ is simply a gain constant for the network. Since the output signal-to-noise ratio will not depend upon μ, we are at liberty to choose μ for our convenience. Thus, let $\mu = -G_0$. The solution now becomes

$$h(\lambda_1) = s_i(\hat{t}-\lambda_1).$$

If an impulse is applied to a realizable network at $t = 0$, the impulse response only exists for $t \geq 0$. The impulse response for the optimum realizable filter is then

$$h(t) = s_i(\hat{t}-t), \qquad t \geq 0."$$

Significance of Using White Noise

As you are probably aware, white noise is a mathematical idealization which does not really exist. The white noise assumption is made when the noise has a power spectral density function which is essentially flat well beyond (or above) the response frequency of the components. For a white noise process, the process at t_1 is completely uncorrelated with the process at t_2 for all t_1, t_2 $(t_1 \neq t_2)$. Analyzing a process corrupted by white noise is similar to a "worse case" analysis since if the process had correlation between two different times there would be "something known" about the trend of the process; however, with white noise there is nothing known.

The white noise assumption entails such physically impossible implications as the noise having infinite power (as

made famous in the "ultraviolet catastrophe"). Even though the white noise assumption has all these drawbacks, still it is useful since it simplifies the analytical calculations, and even after making the white noise assumption, the results of analysis are close to what is actually observed physically.

How Matched-Filtering Procedure Is Altered (Pre-whitening) if the Noise Is Colored

There are situations, however, where the white noise assumption would be bad and it would be more reasonable to acknowledge that we have colored noise. The matched-filtering problem is most easily solved when we have white noise. What we can do is pre-whiten the noise (i.e., find the filter that will give a white noise output for a colored noise input). After this pre-whitening has been applied, the remainder of the problem is exactly the same as what was done in the white noise case (John C. Hancock, 1961, pp. 142-43).

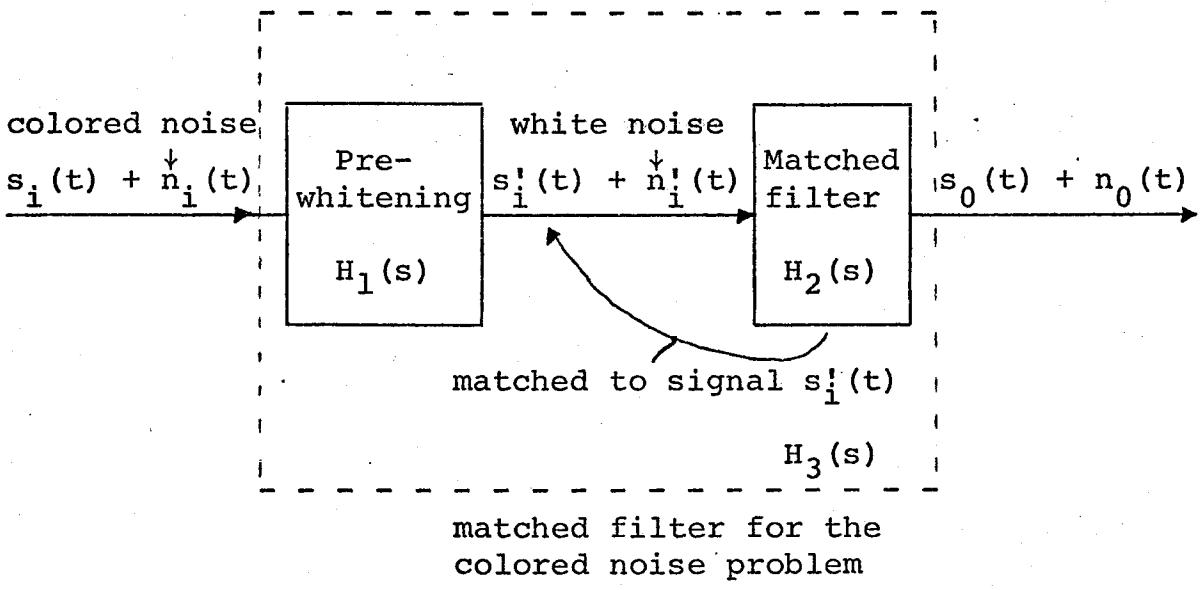
The filter $H_1(s)$ is chosen so that the noise $n_i'(t)$ is white with unit variance. ($1 = S_{n_i'n_i'}(s) = H_1(s)H_1(-s)S_{n_in_i}(s)$.) The signal at the output of the pre-whitening filter is found from $s_i'(s) = H_1(s)s_i(s)$ or

$$s_i'(t) = \int_{-\infty}^{\infty} h_1(t-\tau)s_i(\tau)d\tau.$$

The filter $H_2(s)$ is chosen so that

$$h_2(t) = s_i'(\hat{t}-t), \qquad t \geq 0$$

so that the signal-to-noise ratio at $t = \hat{t}$ is a maximum. The two filters $H_1(s)$ and $H_2(s)$ can be considered together as $H_3(s)$ or $h_3(t) = \int_{-\infty}^{\infty} h_2(t-\lambda)h_1(\lambda)d\lambda$, the matched filter for the colored noise $n_i(t)$.



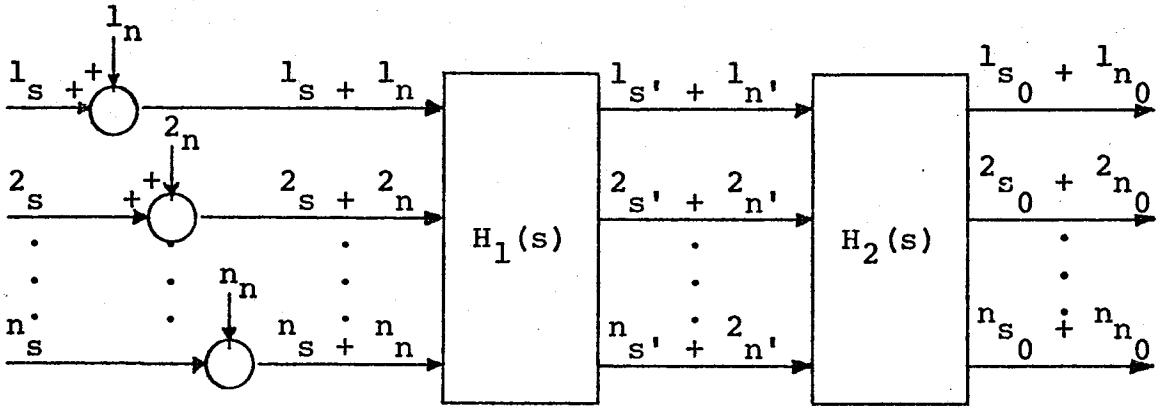
How Procedure Is Altered for the n-dimensional Problem when the Noise Is Colored (Multiple Matched Filter Detection)

Assume $N = [{}^1n, {}^2n, \ldots, {}^nn]^T$ consists of nonwhite, perhaps cross correlated noise. The signals ${}^is$ on each of the n channels corresponding to a particular message may be different or the same.

The relationship between the input to $H_1(s)$ and output power spectral density matrices is

$$S_{n'n'}(s) = H_1(-s)S_{nn}(s)H_1^T(s).$$

If n' is uncorrelated, unit variance, white noise, then



$S_{n'n'}(s) = I$, the identity matrix. If we apply a matrix spectral factorization to $S_{nn}(s)$ to yield $S_{nn}(s) = W^T(-s)W(s)$, then all we need to pre-whiten the noise is a linear filter having the transfer function matrix $H_1(s) = [W^T(s)]^{-1}$. This can be seen since

$S_{n'n'}(s) = H_1(-s)S_{nn}(s)H_1^T(s) = H_1(-s)W^T(-s)W(s)H_1^T(s)$ or

$S_{n'n'}(s) = [W^T(-s)]^{-1}W^T(-s)W(s)[W(s)]^{-1} = I.$

There is no worry about $[W(s)]^{-1}$ existing and being realizable since the matrix factorization procedure yields matrix $W(s)$ which has an inverse $W^{-1}(s)$ both of which have all their poles in $Re(s) < 0$ (i.e., the L.H.P.).

From here, the matched filtering problem is the standard problem. It may be formulated and solved in vector notation in a manner completely analogous to what was done in the review of the scalar problem.

One problem that exists for the n-dimensional matched filter that was nonexistent in the scalar case is that the signal-to-noise ratio may be defined in several different

ways.

$$\left(\frac{S}{N}\right)_1 = \frac{\text{squared magnitude of signal at time } \hat{t}}{\text{average squared magnitude of noise}}$$

$$\left(\frac{S}{N}\right)_2 = \frac{\text{square of the sum of the signal components at time } \hat{t}}{\text{average squared magnitude of noise}}$$

$$\left(\frac{S}{N}\right)_3 = \text{sum of signal-to-noise ratios of each component.}$$

A. V. Balakrishnan (1961, p. 52) showed that all of the above signal-to-noise ratio definitions for the vector process problem yield the same matched filter.

References

Balakrishnan, A. V., "Matched Filters for Multiple Processes," I.R.E. Trans. on Information Theory, Vol. IT-7, January, 1961, p. 52.

Hancock, J. C., The Principles of Communication Theory, McGraw-Hill Book Co., Inc., New York, 1961.

2 Applications of Matrix Spectral Factorization in Communications

(Specifically, applying matrix spectral factorization to put a system into the proper form for applying Kalman-Bucy Filtering Theory)

Prepared by

T. H. Kerr

Date: 23 Dec 71

In recent years there have been several exciting applications of the Kalman-Bucy theory for optimal linear estimation. There have been many applications of the theory in the aerospace area and a few applications in the process control area.

Before the Kalman-Bucy theory can be applied, the problem must be put in the proper form for applying the theory. The proper form is that the system be modeled by

(1) $\dot{x}(t) = F(t)x(t) + D(t)u(t) + G(t)w(t)$, $x(t_0) = x_0$, $t \geq t_0$

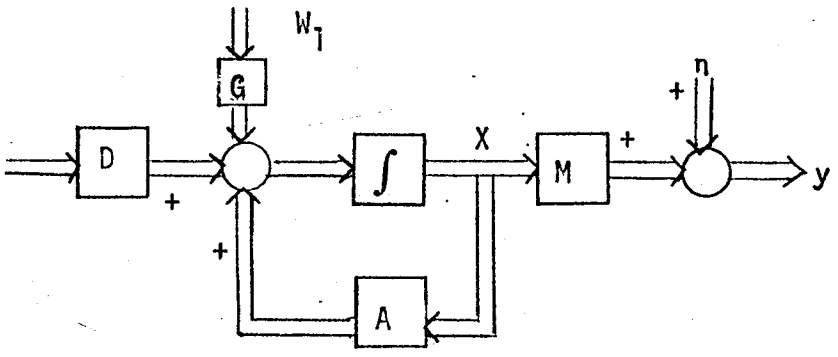
and that the measurements be modeled by

(2) $y(t) = M(t)x(t) + N(t)V(t)$,

where $w(t)$ and $V(t)$ are zero mean, Gaussian, white noises, $u(t)$ is a deterministic input, x is the state, and the y-vector constitutes the measurements.

When $w(t)$ and $v(t)$ are not white, but colored or correlated Gaussian noise, the system is not in proper form for Kalman-Bucy filtering to be applied. Applying matrix spectral factorization allows one to transform (by using state-space augmentation) a problem that is not in the proper form to one that is of the proper form.

Given:



$$\dot{x} = Ax + Du + Gw_1$$

$$y = Mx + \eta$$

w_1 is white, zero-mean, Gaussian noise $W_1W_1(\tau) = Q\delta(\tau)$, $Q > 0$

η is colored Gaussian noise, $S_{\eta\eta}(p)$ is given.

The above system is not of the proper form for applying the Kalman-Bucy filtering theory.

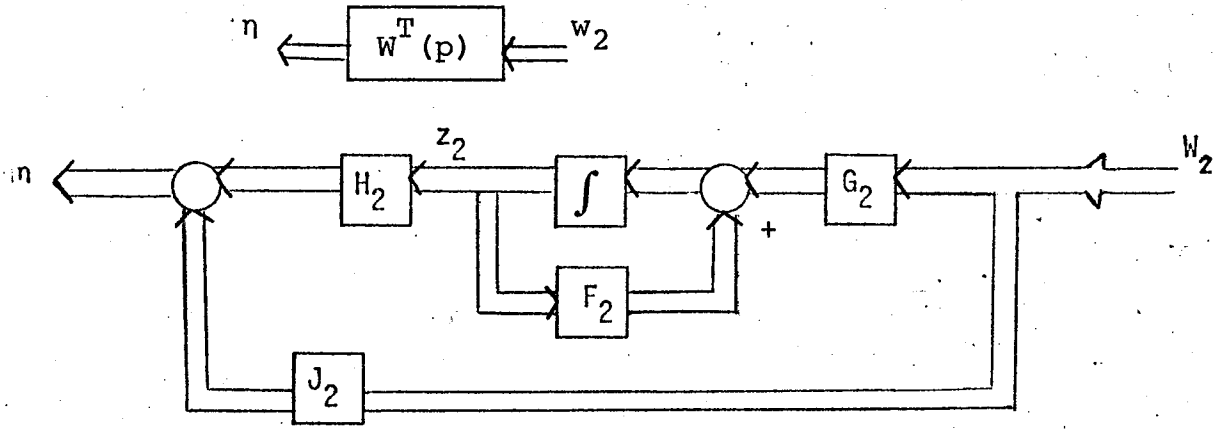
The above problem can be put in the proper form for applying Kalman-Bucy filtering theory by using matrix spectral factorization and state augmentation.

spectral factorization

$$S_{\eta\eta}(p) = W^T(-p)W(p)$$

$$W^T(p) = H_2(pI - F_2)^{-1}G_2 + J_2$$

Kalman method for obtaining a realization from the transfer function.



Let $R = \begin{bmatrix} x \\ \hline z_2 \end{bmatrix}$ ← state augmentation

$$\dot{z}_2 = F_2z_2 + G_2W_2$$

$$\eta = H_2z_2 + J_2W_2$$

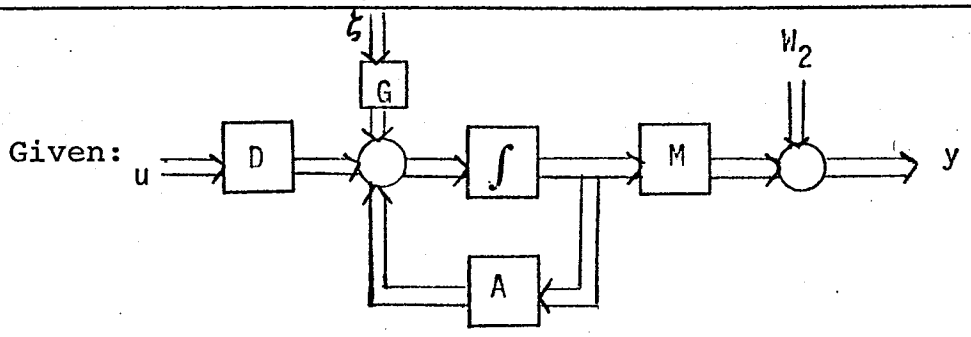
Correct Kalman-Bucy Form

$$\mathbb{R}_{W_1W_1}(\tau) = Q\delta(\tau);\ \mathbb{R}_{W_2W_2}(\tau) = I\delta(\tau)$$

$$\dot{R} = \begin{bmatrix} \dot{x} \\ \dot{z}_2 \end{bmatrix} = \left[\begin{array}{c|c} A & 0 \\ \hline 0 & F_2 \end{array}\right] R + \begin{bmatrix} D \\ 0 \end{bmatrix} u + \left[\begin{array}{c|c} G & 0 \\ \hline 0 & G_2 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

State and measurement noise are correlated.

$$y = [M \mid H_2]R + J_2W_2$$



$\dot{x} = Ax + Du + G\zeta$

$y = Mx + w_2$

W_2 is white, zero mean Gaussian noise, $\mathbb{R}_{W_2W_2}(\tau) = R\delta(\tau)$

ζ is colored Gaussian noise $S_{\zeta\zeta}(p)$ is given.

The above system is not of the proper form for applying the Kalman-Bucy filtering theory.

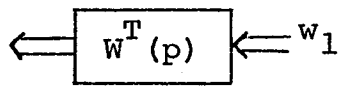
The above problem can now be put in the proper form

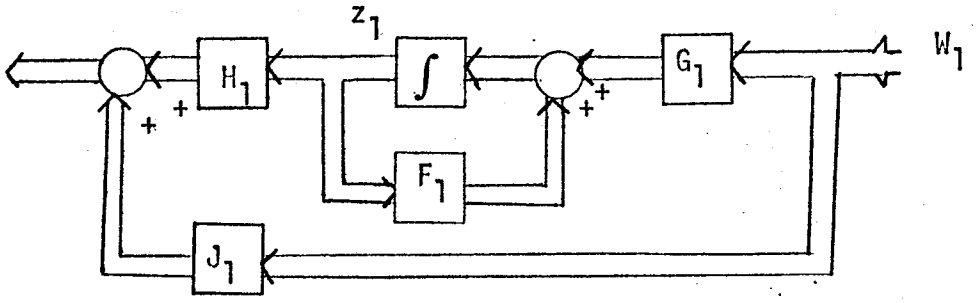
spectral factorization

$$S_{\zeta\zeta}(p) = W^T(-p)W(p)$$

$$W^T(p) = H_1(pI - F_1)^{-1}G_1 + J_1$$

Kalman method for obtaining a realization





Let $R = \begin{bmatrix} x \\ z_1 \end{bmatrix}$ ← state augmentation

$$\dot{z}_1 = F_1 z_1 + G_1 W_1$$

$$\zeta = H_1 z_1 + J_1 W_1$$

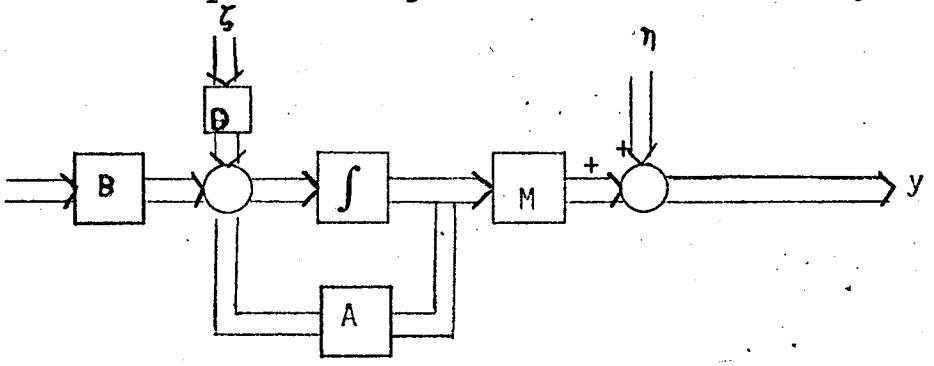
Correct Kalman Form

$$\mathbb{R}_{W_2W_2}(\tau) = R\delta(\tau); \quad \mathbb{R}_{W_1W_1}(\tau) = I\delta(\tau)$$

$$\dot{R} = \begin{bmatrix} \dot{x} \\ \dot{z}_1 \end{bmatrix} = \left[\begin{array}{c|c} A & 0 \\ \hline 0 & F_1 \end{array}\right] R + \begin{bmatrix} D \\ - \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ -- \\ G_1 \end{bmatrix} W_1$$

$$y = [C\ 0]R + W_2$$

As a final example of how spectral factorization can help in Kalman-Bucy filtering consider the following:



Measurement and system noise are both colored and possibly cross correlated.

$$\dot{x} = Ax + Bu + D\zeta$$

$$y = Mx + \eta$$

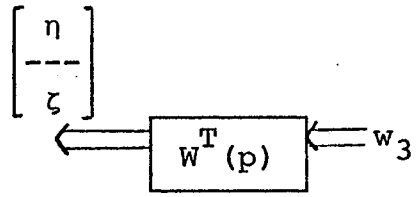
$S_{\eta\eta}(p)$; $S_{\zeta\zeta}(p)$; $S_{\zeta\eta}(p)$ are given.

Define $$\mathcal{S}(p) = \left[\begin{array}{c|c} S_{\eta\eta}(p) & S_{\zeta\eta}(p) \\ \hline S_{\eta\zeta}(p) & S_{\zeta\zeta}(p) \end{array}\right] = W^T(-p)W(p)$$

spectral factorization

$$W^T(p) = H_3(pI - F_3)^{-1}G_3 + J_3$$

Kalman realization



$$R = \begin{bmatrix} x \\ z_3 \end{bmatrix}; \quad \dot{z}_3 = F_3 z_3 + G_3 W_3$$

$$\begin{bmatrix} \eta \\ \hline \zeta \end{bmatrix} = H_3 z_3 + J_3 W_3$$

Correct Kalman Form

$$\dot{R} = \left[\begin{array}{c:c} A & D\begin{bmatrix} 0 & 0 \\ 0 & I_{q\times q} \end{bmatrix} H_3 \\ \hdashline 0 & F_3 \end{array}\right] R + \begin{bmatrix} B \\ \hline 0 \end{bmatrix} u + \left[\begin{array}{c} D\begin{bmatrix} 0 & 0 \\ 0 & I_{q\times q} \end{bmatrix} J_3 \\ \hline G_3 \end{array}\right] W_3$$

$$y = \left[M \vdots \begin{bmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{bmatrix} H_3\right] R + J_3 W_3$$

$\mathbb{R}_{W_3W_3}(\tau) = I\delta(\tau)$. Here the state and measurement white noise are the same.

In all truthfulness, the first example demonstrated on page 1 may be solved another way as shown in a paper by Edwin Stear and Allen R. Stubberud, "Optimal Filtering for Gauss-Markov Noise," International Journal of Control,

Vol. 8, No. 2, 1968, pp. 123-30. This approach does not involve state augmentation, however it is restrictive in the sense that it assumes N^{-1} exists in eq. (1) ($y = Mx + Nv(t)$). In reality, N need not be square let alone be nonsingular.

In the recent text by Sage and Melsa, Estimation Theory with Applications to Communication and Control, McGraw-Hill, 1971, it is stated that if the Gauss-Markov process $\dot{\gamma}(t)$ can be generated by $u(t)$, a white noise process in such a way that $\dot{\gamma}(t) = \Xi(t)\gamma(t) + \Sigma(t)u(t)$, then the state equation may be augmented so that the whole model is driven by white noise. Here they assumed that $\Xi(t)$ and $\Sigma(t)$ were already known. In a realistic situation, you would only have the power spectral density matrix obtained from the measurements. From this you would obtain the Ξ and Σ by spectral factorization and obtaining a realization.

3 Applications of Matrix Spectral Factorization in Network Synthesis

(Specifically, a general synthesis technique for linear multiport networks using the scattering matrix)

Includes:

Sufficient Conditions for Matrix Factorization and How Satisfied by the Intended Applications

I All Applications of Matrix Factorization in Communications Theory Encountered Thus Far Involve a Power Spectral Density Matrix

II A Network Synthesis Application of Matrix Factorization

Sufficient Conditions for Matrix Factorization and How Satisfied by the Intended Applications

All that is needed to apply the mathematical technique of the matrix factorization procedure of Youla, Davis, or Anderson to a matrix $S(p)$ is that the following four conditions be satisfied (sufficiency):

(i) the elements of $S(p)$ are ratios of polynomials in the complex variable p,

(ii) $S(p) = S(\overline{p})$, where the vinculum denotes taking the complex conjugate,

(iii) $S^T(-p) = S(p)$ (i.e., para-Hermetian),

(iv) $S(p)\big|_{p=j\omega} = S(j\omega)$ is positive semidefinite (i.e., $S(j\omega) \geq 0$ for all real ω).

I All Applications of Matrix Factorization in Communications Theory Encountered Thus Far Involve a Power Spectral Density Matrix

It will now be shown that an arbitrary power spectral density matrix satisfies conditions (ii), (iii), and (iv) and may be approximated arbitrarily closely by a matrix that, in addition to satisfying (ii), (iii), and (iv), also satisfied (i).

For any two real valued random processes $\{y_i(t)\}_{t\in T}$ and $\{y_j(t)\}_{t\in T}$ that are second order processes (i.e., $E[|y_i(t)|^2] < \infty$ and $E[|y_j(t)|^2] < \infty$ for all elements t of an ordered index set T, which by Holders inequality assures

that $E[|y_i(t)y_i(s)|] < \infty$ for every $t, s \in T$), the correlation function is $R_{y_iy_j}(t,s) = E[y_i(t)y_j(s)] = E[y_j(s)y_i(t)] \triangleq R_{y_jy_i}(s,t)$. If, in addition, $y_i(t)$ and $y_j(t)$ are jointly wide sense stationary, then $R_{y_iy_j}(t-s) = R_{y_iy_j}(t-s,0) = R_{y_iy_j}(t,s) = R_{y_jy_i}(s,t) = R_{y_jy_i}(s-t,0) = R_{y_jy_i}(s-t)$. Upon substituting $\tau = t - s$ in the above expression, the result is $R_{y_iy_j}(\tau) = R_{y_jy_i}(-\tau)$. Since the power spectral density function is here defined to be the bilateral Laplace transform of the correlation function, we have that

$$S_{y_iy_j}(p) \triangleq \int_{-\infty}^{\infty} R_{y_iy_j}(\tau)e^{-p\tau}\,d\tau = \int_{\infty}^{\infty} R_{y_jy_i}(-\tau)e^{-p\tau}\,d\tau$$

$$= \int_{-\infty}^{\infty} R_{y_jy_i}(\tau')e^{-(-p)\tau'}\,d\tau' \triangleq S_{y_jy_i}(-p)$$

by a change of variable ($\tau = -\tau'$). Therefore, for power spectral density matrices, it is always true that $S^T_{\tilde{y}\tilde{y}}(-p) = S_{\tilde{y}\tilde{y}}(p)$, where $y^T(t) = [y_1(t), y_2(t), \ldots, y_n(t)]$; so power spectral density matrices always satisfy condition (iii).

For $\tau \in \mathbb{R}$ and the fact that the conjugate of a product is the product of the conjugates, we have that $\overline{e^{-p\tau}} = \overline{1 - p\tau/1! + (p\tau)^2/2! - \ldots} = 1 - \bar{p}\tau/1! + (\bar{p}\tau)^2/2! - \ldots = e^{-\bar{p}\tau}$. The preceding is a useful lemma in showing that $\overline{S_{\tilde{y}\tilde{y}}(p)} = \overline{\int_{-\infty}^{\infty} R_{\tilde{y}\tilde{y}}(\tau)e^{-p\tau}\,d\tau} = \int_{\infty}^{\infty} \overline{R_{\tilde{y}\tilde{y}}(\tau)e^{-p\tau}}\,d\tau = \int_{-\infty}^{\infty} R_{\tilde{y}\tilde{y}}(\tau)\overline{e^{-p\tau}}\,d\tau = \int_{-\infty}^{\infty} R_{\tilde{y}\tilde{y}}(\tau)e^{-\bar{p}\tau}\,d\tau = S_{\tilde{y}\tilde{y}}(\bar{p})$. Therefore, for power spectral density matrices it is always true that

$S_{\tilde{y}\tilde{y}}(p) = S_{\tilde{y}\tilde{y}}(\overline{p})$; so, power spectral density matrices always satisfy condition (ii).

Since the power spectral density matrix satisfies both (ii) and (iii), it follows that $S^{T}_{\tilde{y}\tilde{y}}(p) = S^{T}_{\tilde{y}\tilde{y}}(\overline{p}) = S_{\tilde{y}\tilde{y}}(-\overline{p})$, which is the definition of a paraconjugate Hermitian matrix (Youla, 1961, p. 173); thus, power spectral density matrices are paraconjugate Hermetian.

As pointed out in Kraus and Pötzl, 1956, p. 283, equation 6, a necessary and sufficient condition for a matrix $S_{yy}(p)$ to be the power spectral density matrix of a wide sense stationary vector random process is that $\overline{b}^{T} S_{\tilde{y}\tilde{y}}(j\omega)\, b \geq 0$, for every complex n-vector $b \neq \underset{n\times 1}{0}$, for all $\omega \in \mathbb{R}$. Therefore, power spectral density matrices always satisfy condition (iv).

Since power spectral density matrices satisfy conditions (ii) and (iii), it is true that $S^{T}_{\tilde{y}\tilde{y}}(p) = S_{\tilde{y}\tilde{y}}(-p)$. The Hermitian part, $S_{H}(p)$, of a matrix $S(p)$ is defined as follows $S_{H}(p) = (1/2)[S^{T}(p) + S(p)]$. For power spectral density matrices, the Hermitian part, for $p = j\omega$; is $S_{\tilde{y}\tilde{y}H}(j\omega) = (1/2)[S^{T}_{\tilde{y}\tilde{y}}(j\omega) + S_{\tilde{y}\tilde{y}}(j\omega)] = (1/2)[S^{T}_{\tilde{y}\tilde{y}}(-j\omega) + S_{\tilde{y}\tilde{y}}(j\omega)] = (1/2)[S_{\tilde{y}\tilde{y}}(j\omega) + S_{\tilde{y}\tilde{y}}(j\omega)] = S_{\tilde{y}\tilde{y}}(j\omega)$. Therefore, for power spectral density matrices, the condition that $\overline{b}^{T} S_{\tilde{y}\tilde{y}}(j\omega)\, b \geq 0$, for every complex n-vector $b \neq \underset{n\times 1}{0}$, for all $\omega \in \mathbb{R}$ is equivalent to the condition that $\overline{b}^{T} S_{\tilde{y}\tilde{y}H}(j\omega)\, b \geq 0$,

for every complex n-vector $b \neq 0_{n\times 1}$, for all $\omega \in \mathbb{R}$. This last test is one of several conditions that must be satisfied in order that a matrix be positive real as defined in Newcomb, 1966, p. 117 or Anderson, 1967, p. 171. However, the fact that a matrix satisfies this last condition is not equivalent to the matrix being positive real since the matrix must also satisfy several other conditions, one being that the matrix be analytic in Re(p) > 0 which, in general, is not satisfied by power spectral density matrices.

Laning and Battin (1956, p. 381) show that an arbitrary correlation function of a real wide sense stationary process may be approximated to within any desired accuracy in the integral square error sense by a class of exponential functions. These functions have bilateral Laplace transforms that are ratios of polynomials having real coefficients; thus, the power spectral density matrix of any arbitrary wide sense stationary process can be approximated in this manner by a matrix consisting of the ratios of polynomials and still satisfying conditions (ii), (iii), and (iv).

Approximation of an arbitrary power spectral density matrix by a power spectral density matrix having elements that are ratios of polynomials with real coefficients can be done directly in the frequency domain by any of the methods of approximating the spectral density functions by

meromorphic functions (i.e., functions only having singularities consisting of a finite number of poles) mentioned in Solodovnikov, 1960, p. 169. These methods for approximating the power spectral density function in the frequency domain are:

(a) interpolation,

(b) approximation based on a Fourier series expansion,

(c) approximation based on a Laguerre series expansion,

(d) approximation of the logarithmic curve of the power spectral density by intersecting straight line segments.

By any of the above mentioned methods, an approximate power spectral density matrix can be obtained that has elements consisting of ratios of polynomials with real coefficients. This approximation to the power spectral density matrix satisfies (i), (ii), (iii), and (iv), so that the various factorization procedures may be applied.

II A Network Synthesis Application of Matrix Factorization

A. The general n-port synthesis method that Newcomb describes (p. 309, p. 145, Newcomb) which is originally due to Belevitch uses a Gauss diagonalization procedure to effect the required factorization of R and R_{II} into $R = \tilde{\Sigma}_{21*}\Sigma_{21}$ and $R_{II} = \Sigma_{12}\tilde{\Sigma}_{12*}$, respectively. An important

step in the procedure involves using Σ_{12}^{-1}, which Newcomb admits only exists "formally." The reason for this is that the Gauss diagonalization procedure does not guarantee anything about the existence or analyticity in $\mathrm{Re}(p) \geq 0$ of the inverses of the factors; some factors may have inverses while other factors do not; however, the Youla, Davis, and Anderson factorizations do guarantee the existence of inverses that are analytic in $\mathrm{Re}(p) \geq 0$. It is the Youla, Davis, Anderson method that we desire to develop a computer program for.

B. The matrices R and R_{II} satisfy conditions (i) - (iv) for applying the factorization procedures.

1. Conditions (i) and (ii) are trivially satisfied since all elements of the matrices R and R_{II} are ratios of polynomials

2. Since $R(p) \triangleq I - \tilde{S}_*(p)S(p)$

$$R_{II} \triangleq I - S(p)\tilde{S}_*(p), \text{ then}$$

$\tilde{R}(p) = [I - \tilde{S}_*(p)S(p)] = I - \tilde{S}(p)S_*(p)$

$\tilde{R}_{II}(p) = [I - S(p)\tilde{S}_*(p)] = I - S_*(p)\tilde{S}(p)$

and $\tilde{R}_*(p) = \tilde{R}(-p) = I - \tilde{S}_*(p)S(p) = R(p)$

and $\tilde{R}_{II*}(p) = \tilde{R}_{II}(-p) = I - S(p)\tilde{S}_*(p) = R_{II}(p)$; hence condition (iii) is satisfied.

3. We can verify condition (iv) by assembling the proper string of proved theorems from Newcomb's textbook.

a) Table on page 105 of Newcomb's text states that a property of a passive (linear, solvable, time invariant) network is that the impedance matrix Z is positive real.

b) Theorem 5-2 ([p. 121] An Alternate PR Test): An $n \times n$ matrix Z is positive real iff $S \triangleq (Z + I)^{-1}(Z - I)$ exists (see p. 52, p. 130) and satisfies the following

1. S is real rational
2. S has no poles in $\sigma \geq 0$
3. $I_n - \tilde{S}^*(j\omega)S(j\omega) \geq 0$ for all real ω.

c) Theorem 5-13 (p. 130): If Z is PR, then $S = (Z + I)^{-1}(Z + I)$ exists and is bounded real (BR). By a), c), and b), S exists and $I_n - \tilde{S}^*(j\omega)S(j\omega) \geq 0$ for all real ω. Since $R(p) \triangleq I_n - \tilde{S}_*(p)S(p)$, we have that $R(j\omega) = I_n - \tilde{S}^*(j\omega)S(j\omega) \geq 0$ for all real ω; hence condition (iv) is satisfied for R.

d) Theorem 5-21 (p. 147): If S(p) is BR, then $R(p) = I - \tilde{S}_*S$ and $R_{II}(p) = I - S\tilde{S}_*$ have the same rank and $R_{II}(j\omega) \geq 0$. Therefore $R_{II}(p)$ satisfies condition (iv).

References

Anderson, B. D. O., "A System Theory Criterion for Positive Real Matrices," J.S.I.A.M. Control, Vol. 5, May, 1967, pp. 171-82.

Kraus, G., and Pötzl, H., "Limiting Conditions on the Correlation Properties of Random Signals," I.R.E. Trans. on Circuit Theory, Vol. CT-3, December, 1956, pp. 282-85.

Laning, J. H., and Battin, R. H., Random Processes in Automatic Control, McGraw-Hill Book Co., Inc., New York, 1956.

Newcomb, R. W., Linear Multiport Synthesis, McGraw-Hill Book Co., Inc., New York, 1966.

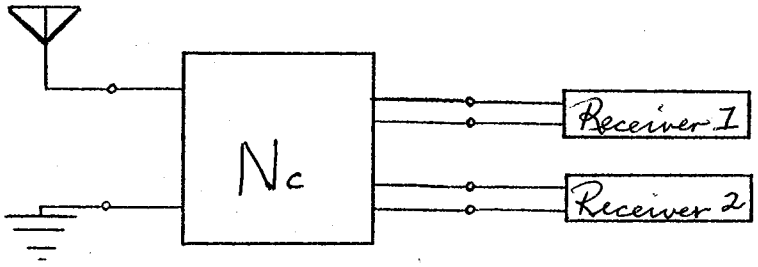
Solodovnikov, V. V., Introduction to the Statistical Dynamics of Automatic Control Systems, Translation Edited by J. B. Thomas and L. A. Zadeh, Dover Publications, Inc., New York, 1960.

Youla, D. C., "On the Factorization of Rational Matrices," I.R.E. Trans. on Information Theory, Vol. IT-7, July, 1961, pp. 172-89.

Two Typical Examples of Problems That Linear n-Port Network Synthesis Theory Solves

Problem 1: Antenna Feed Structure (three ports)

(R. W. Newcomb, Linear Multiport Synthesis, McGraw-Hill Book Co., Inc., New York, 1966, p. 235)



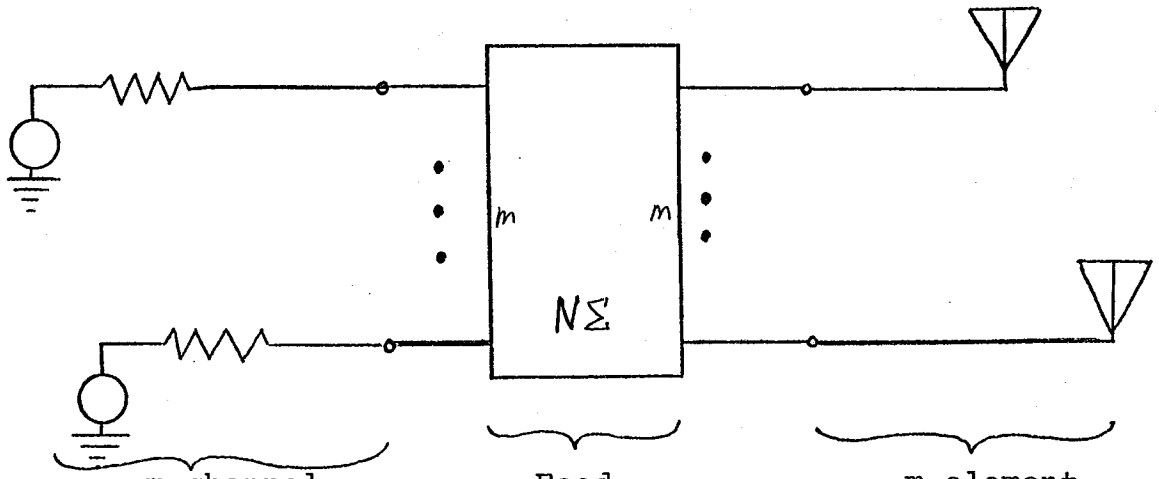
"It is desired to have an antenna feed two separate receivers, as shown in the above figure. Upon normalizing to 1-ohm terminations (ibid., p. 145), all ports are to be matched (i.e., for a scattering matrix $S(p)$, $s_{ii}(p) = 0$). It is also desired to have no coupling between receivers, $s_{23} = s_{32} = 0$, or transmission between receivers and antennas in the backward direction, $s_{13} = s_{12} = 0$. Equal transmission from the antenna to the two receivers of a second order Butterworth characteristic with a zero at the origin is also desired; thus $s_{21} = s_{31} = p/(p^2 + \sqrt{2}\, p + 1)$. The coupling network, N_c, is then described by

$$S(p) = \begin{bmatrix} 0 & 0 & 0 \\ \dfrac{p}{p^2 + \sqrt{2}\, p + 1} & 0 & 0 \\ \dfrac{p}{p^2 + \sqrt{2}\, p + 1} & 0 & 0 \end{bmatrix}."$$

This synthesis problem is solved in Newcomb's textbook.

Problem 2: Antenna Feed Structure (2m-ports)

(ibid., p. 104)



If the antennas are to be used in an identical manner for transmission and reception, we desire that the feed network be reciprocal, and certainly it is reasonably required to be lossless. It is also desirable to match and isolate the m transmitter channels. Thus we desire the scattering matrix Σ of the 2m-port feed network N as

$$\Sigma = \begin{bmatrix} O_m & \tilde{\Sigma}_{21} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Since we want N_Σ to be lossless, Σ is para-unitary. By the para-unitary property of Σ, Σ_{21} is para-unitary and $\Sigma_{22} = O_m$. Thus, at a fixed real frequency $p_0 = j\omega_0$, $\Sigma_{21}(j\omega_0) =$

$[\alpha_{ij}]$ with $1 = \sum_{i=1}^{m} |\alpha_{ij}|^2 = \sum_{i=1}^{m} |\alpha_{ji}|^2$ for all $j = 1, \ldots, m$.

Now, α_{ij} represents the illumination of the $i^{\underline{th}}$ antenna due to an excitation at the $j^{\underline{th}}$ transmitter, and thus, if we require similar excitations to give similar responses, we can specify $|\alpha_{j1}| = |\alpha_{ji}|$, $i = 1, \ldots, m$ for each j; that is, the magnitude of amplitudes in a fixed row may be chosen equal. Using this in the above shows that $|\alpha_{ij}| = 1/\sqrt{m}$. If, further, a fixed wave sent into ports, 1, 2, ..., m must be progressed in phase to add at the first antenna, then we can extract the phase progressions as factors $\exp[j\theta_i]$, $i = 1, \ldots, m$ from the $i^{\underline{th}}$ column of Σ_{21}. The remaining columns must still have unit magnitude, and if we distribute the beams in a maximally flat manner when any one feed is excited, each column can be considered as consisting of the m roots of unity. By, perhaps, a proper ordering of the output ports we can then guarantee that

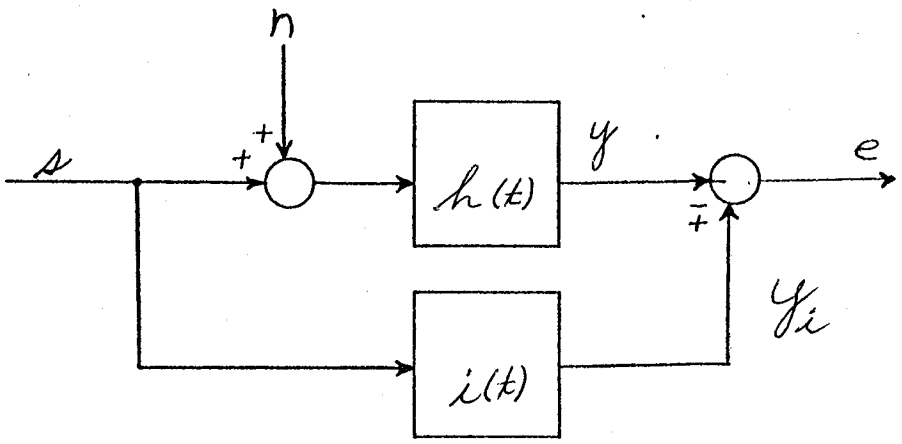
$$\Sigma_{21}(j\omega_0) = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \alpha & \alpha^2 & \ldots & \alpha^{m-1} \\ \alpha^2 & \alpha^3 & \ldots & \alpha \\ \cdot & & & \\ \cdot & & & \\ \alpha^{m-1} & \alpha & \ldots & \alpha^{m-2} \end{bmatrix} \begin{bmatrix} e^{j\theta_1} & 0 & \ldots & 0 \\ 0 & e^{j\theta_1} & & 0 \\ 0 & 0 & e^{j\theta} & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \ldots & e^{j\theta_1} \end{bmatrix},$$

where α is an $m^{\underline{th}}$ root of unity. This then gives a possible specification for the feed network for single-frequency operation.

4 Applications of Matrix Spectral Factorization in Communications

(Specifically applied to Wiener filtering)

Wiener filtering formulation:



Find the h(t) (realizable) to minimize $\sigma_e^2 = E[e(t)^2]$, the mean-square error. (Assumptions: stationary s(t), stationary noise n(t), corresponding power spectral densities of both are known, noise is additive, impulse responses h(t) and i(t) correspond to time invariant linear systems so problem may be solved in the frequency domain.)

In order to obtain a causal filter, must perform a scalar spectral factorization (Sage, Optimum System Control, 1968, p. 190, eq. 8.3-12). Manipulations are used which assure that filter has poles in LHP only.

The analogous Wiener filtering problem may be posed and solved for the multidimensional case (ibid., pp. 209-11) but matrix spectral factorization is required (ibid., p. 211, eq. 8.5-36). Matrix spectral factorization is a very formidable problem unless a computer method is available. This is why Wiener filters are usually only for the scalar case.

5 Application of Matrix Spectral Factorization in Modeling a Particular Class of Multiple Input Multiple Output Black Boxes

A Synopsis of the Paper as Applied to the Scalar Case (More readable)

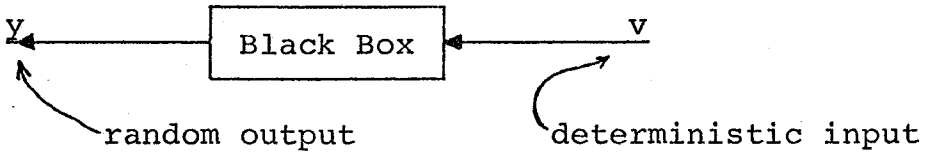
B Paper Which Generalizes the Idea to the Multiple-Input Multiple-Output Case

A The Scalar Case

Synopsis of the paper as applied to a single-input single-output black box rather than to a general multi-input multi-output black box. (The paper solves the general problem but the notation becomes cumbersome.)

Modeling a scalar problem having inputs and outputs

Consider the following black box:



Want to use a model of the form

(I)
$$\frac{dx}{dt} = Fx + Gw(t) + Mv$$

($w(t)$: white noise, Gaussian)

$$x(0) = C, \text{ a random variable}$$

$$y(t) = Hx(t) + m,$$

where $E[C] = 0$, $E[CC] = P$, $E[w(t)] = 0$ for all t,

$$E[w(t)w(s)] = Q\delta(t-s),\ Q > 0.$$

Unknown parameters: F, G, H, m, P, Q, M (7 unknown).

Approach: test to see if the above mathematical model is adequate for the black box. (Hypothesis testing, where we determine the number of trials we must make to have α confidence in the outcome of the test.)

By an argument involving equivalent Fokker-Plank equations, mathematical model (I) is equivalent to mathematical model (II).

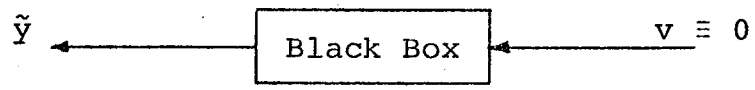
(II)
$$\begin{cases} \dfrac{dx}{dt} = Fx + G_0 W(t) + Mv, \\ \qquad\qquad\qquad \text{white noise} \\ x(0) = C, \text{ a random variable,} \\ y(t) = Hx(t) + m, \\ \text{where } E[C] = 0,\ E[CC] = P,\ E[w(t)] = 0 \text{ for all } t, \\ E[w(t)w(s)] = I\delta(t-s),\ Q > 0. \end{cases}$$

Unknown parameters: F, G_0, H, m, P, M (6 unknown).

(Modified Bucy-Joseph approach)

Procedure:

Clamp $v \equiv 0$



Make measurements for $v \equiv 0$:

$$R_{\tilde{y}\tilde{y}}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \tilde{y}(u+\tau)\tilde{y}(u)du \quad \text{(from data)}$$

$$s_{\tilde{y}\tilde{y}}(p) = \mathcal{L}_{II}[R_{\tilde{y}\tilde{y}}(\tau)]$$

bilateral Laplace transform

$$m = \lim_{T\to\infty} \frac{1}{T} \int_0^T \tilde{y}(u)du \quad \text{(from data)}$$

Let $\tilde{z} = \tilde{y} - m$

$$R_{\tilde{z}\tilde{z}}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T [\tilde{y}(u+\tau) - m][\tilde{y}(u) - m]du \quad \text{(from data)}$$

$$s_{\tilde{z}\tilde{z}}(\tau) = \mathcal{L}_{II}[R_{\tilde{z}\tilde{z}}(\tau)]$$

Can derive theory for model (II) which says that for z = y - m,

$$S_{zz}(p) = H(pI - F)^{-1}G_0G_0(-pI - F)^{-1}H$$

E[CC] = P, where P satisfies $FP + PF + G_0G_0 = 0$,

$$(P = \int_0^\infty e^{Ft}\, G_0 G_0\, e^{Ft}\, dt).$$

Factor $S_{\tilde{z}\tilde{z}}(p) = W(-p)W(p)$,

$W(-p)W(p) = S_{\tilde{z}\tilde{z}}(p) = S_{zz}(p) = H(pI - F)^{-1} G_0 G_0 (-pI - F)^{-1} H$

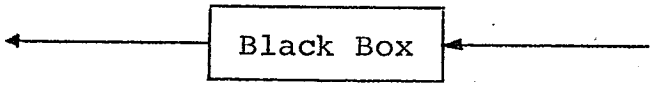
factored — measurements — theoretical model

Let $W(p) = H(pI - F)^{-1} G_0$

Can find H, F, G_0 easily (Kalman) (Gibson). Once we know (H, F, G_0), can find P: either by $P = \int_0^\infty e^{Ft}\, GG_0\, e^{Ft}\, dt$, or, since this is the scalar case, by $(2F)P = -G^2$ or $P = -\frac{G^2}{2F}$. Recall that $m = \lim_{T\to\infty} \frac{1}{T} \int_0^\infty \tilde{y}(u)\, du$.

We now know 5 of the 6 parameters. Must now test to see if the linear mathematical model (II) is adequate and if it is, determine the sixth parameter, M.

Now for $v \neq 0$



Theoretically:

If the mathematical model (II) is correct, then x(t) is a Gaussian process, with mean

$$E[x(t)] = \int_0^t e^{F(t-\tau)}\, Mv(\tau)\, d\tau$$

(M: unknown)

and variance

$$\mathrm{Var}[x(t)] = \int_0^t e^{Ft}\, G_0 G_0\, e^{Ft}\, dt + e^{Ft}\, P e^{Ft}$$

(at a particular time, x(t) is a Gaussian random variable with unknown mean but known variance).

Now $y(t) = Hx(t) + m$.

1. Hx is a linear transformation on x and so is also Gaussian.

2. Adding the constant m maintains Gaussian.

At a particular time, x(t) is a Gaussian random variable with unknown mean but known variance

$$E[y(t)] = \int_0^T H\, e^{F(t-\tau)} M v(\tau)\, d\tau$$

$$\mathrm{Var}[y(t)] = \int_0^t H\, e^{Ft}\, G_0 G_{0_1}\, e^{Ft}\, H\, dt + H\, e^{Ft}\, P\, e^{Ft}\, H$$

$$= \int_0^t H^2\, G_0{}^2\, e^{2Ft}\, dt + H^2\, e^{2Ft}\, P$$

H_0: The black box is adequately described by the mathematical model (II). We can now determine the number of trials we need to make to have α-confidence ($0 \leq \alpha \leq 1$) in the outcome of the test.

The test is essentially a test to see if the actual black box behaves linearly and in a time invariant manner. If it does, then superposition should apply.

Apply an input $v^A(t)$, where $v^A(t)$ is an arbitrary scalar function of time. Apply this input several times, say N times (it is this N, the number of trials that will be determined a priori as will be shown below), and record the outputs, $y^A(t)$, corresponding to the N trials (for each trial, "enough" time should elapse between trials to allow the system to decay back to its original status [how much time is "enough" should be evident from the previously

determined F matrix: 10 times the reciprocal of the smallest eigenvalue]); the record should extend from t = 0 to t = T*, where T* is chosen for convenience. Divide the time interval [0,T*] into equispaced points, so that $\Delta = T^*/m$. For $t = k\Delta$, for every fixed k, a member of [1, 2, ..., m], average the N outputs $[y^A(k\Delta)]$ to yield

$$\bar{y}^A(k\Delta) = \frac{1}{N} \sum_{n=1}^{N} [y^A(k\Delta)]_n.$$

Now apply an input $v^B(t)$, where $v^B(t)$ is an arbitrary scalar function of time different from $u^A(t)$. Similarly obtain

$$\bar{y}^B(k\Delta) = \frac{1}{N} \sum_{n=1}^{N} [y^A(k\Delta)]_n.$$

Finally, apply an input $v^{(A+B)}(t) = v^A(t) + v^B(t)$. Similarly record the response and obtain

$$\bar{y}^{A+B}(k\Delta) = \frac{1}{N} \sum_{n=1}^{N} [y^{A+B}(k\Delta)]_n.$$

If the black box were linear and the inputs entered in a time-invariant manner, then $[y^A(k\Delta)_n]_{n=1}^N$, for a fixed k, is a sample of size N from a Gaussian population where the population has a known variance of

$$\mathrm{Cov}[y^A(k\Delta)] = \int_0^{k\Delta} H\, e^{F(k\Delta-u)}\, G_0 G_0\, e^{F(k\Delta-u)}\, H\, du + H\, e^{Fk\Delta}\, P\, e^{Fk\Delta}\, H$$

and unknown mean

$E[y^A(k\Delta)] = [\int_0^{k\Delta} H\, e^{F(k\Delta-\tau)}\, u^A(\tau)d\tau]M$, where M is an unknown vector.

For $t = k\Delta$, for any fixed k, the probability that the sample mean $\bar{y}^A(k\Delta)$ of a sample of size N is within ε of the population mean $\mu^A(k\Delta)$ is given by

$$P\left[\frac{1}{N}\frac{(\bar{y}^A(k\Delta) - \mu^A(k\Delta))^2}{\mathrm{Cov}[y^A(k\Delta)]} \leq \varepsilon^2\right] = P\left[\frac{1}{N}|\bar{z}|^2 \leq \varepsilon^2\right]$$

$$= P[|\bar{z}|^2 \leq N\,\varepsilon^2] = \alpha,$$

where use has been made of the transformation

$$z = \frac{y^A(k\Delta) - \mu^A(k\Delta)}{\mathrm{Cov}[y^A(k\Delta)]},$$

where z is a Gaussian random variable having $p_z(B) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2} z^2]$. Note that for $t = k\Delta$, for any fixed k, the problem transforms into the same problem in z with the same sample size N. From a table of areas under the Gaussian curve, it is possible to calculate, a priori, the sample size N required so that we have α confidence that the sample mean (a maximum likelihood, sufficient, unbiased, "efficient," and "consistent" statistic for the population mean [Hogg and Craig, 1970, p. 255]) is within ε of the true population mean, where α and ε are set in advance.

Returning to test the linearity hypothesis, if the actual black box were linear, then

$$||\bar{y}^A(k\Delta) - \mu^A(k\Delta)||^2_{R-1} = ||\bar{z}||^2_I \leq \varepsilon^2$$

and we would have:

$$||\bar{y}^A(k\Delta) + \bar{y}^B(k\Delta) - y^{A+B}(k\Delta)||_{R-1} \leq ||y^A(k\Delta) - \mu^A(k\Delta)||_{R-1}$$
$$+ ||\bar{y}^B(k\Delta) - \mu^B(k\Delta)||_{R-1} + ||y^{A+B}(k\Delta) - \mu^{A+B}(k\Delta)||_{R-1}$$
$$+ ||\mu^A(k\Delta) + \mu^B(k\Delta) - \mu^{A+B}(k\Delta)||_{R-1},$$

where

$$||\bar{y}^A(k\Delta) - \mu^A(k\Delta)||^2_{R-1} \leq \varepsilon^2 N, \text{ etc.},$$

$$||\mu^A(k\Delta) + \mu^B(k\Delta) - \mu^{A+B}(k\Delta)||^2_{R-1} = 0;$$

therefore

$$||\bar{y}^A(k\Delta) + \bar{y}^B(k\Delta) = \bar{y}^{A+B}(k\Delta)||^2_{R-1} \leq 9\ \varepsilon^2\ N.$$

(R was used in the above to represent the appropriate covariance.)

Define

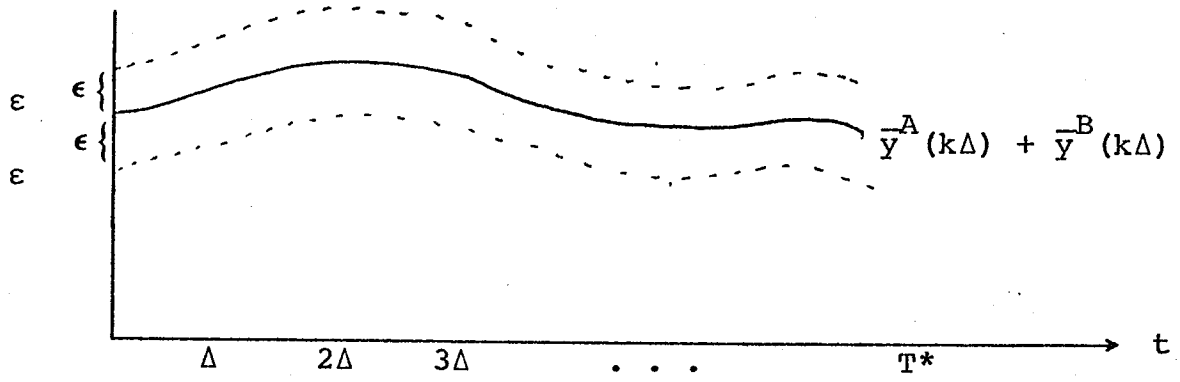
$$\gamma = \sum_{k=1}^{m} ||\bar{y}^A(k\Delta) + y^B(k\Delta) - y^{A+B}(k\Delta)||^2_{R-1}.$$

Therefore, to accept the hypothesis that the actual black box is linear and time invariant, it must be that

$$\gamma \leq 9\ m\ \varepsilon^2\ N.$$

If $\gamma > 9\ m\ \varepsilon^2\ N$, the conclusion is that the black box under consideration cannot be modeled by the mathematical model (II).

The criterian can be interpreted pictorially in the following figure



If $\bar{y}^{A+B}(k\Delta)$ is within the ε-sheath to a degree that, at the m time points at which it is checked, the sum of the excursions outside the ε-sheath are compensated for by its proximity to $\bar{y}^A(k\Delta) + \bar{y}^B(k\Delta)$ at other times such that $\gamma \leq 9\ m\ \varepsilon^2\ N$, the hypothesis that the actual black box is linear is accepted.

Now returning to the problem of identifying M, we have that

$$\bar{y}^A(T^*) = \frac{1}{N} \sum_{n=1}^{N} [y^A(T^*)]_n \doteq E[y^A(T^*)] = \left[\int_0^{T^*} H\ e^{F(T^*-\tau)} u^A(\tau)\, d\tau\right] M;$$

$$\therefore M = \frac{y^A(T^*)}{\left[\int^{T^*} H\ e^{F(T^*-\tau)}\ u^A(\tau)\, d\tau\right]} .$$

MODELING A PARTICULAR CLASS OF MULTIPLE INPUT MULTIPLE OUTPUT BLACK BOXES WITH STOCHASTIC INTEGRAL EQUATIONS**

Earl D. Eyman and Thomas Kerr*
Department of Electrical Engineering
University of Iowa, Iowa City

ABSTRACT

In this paper, a method is given for obtaining a mathematical model of a class of black boxes having multiple inputs and multiple outputs in terms of Ito stochastic integral equations. This method is applicable to the class of black boxes having ergodic correlation functions when there is zero applied input. The point of view adopted in this paper is phenomenological in that it is desired that calculations made using the mathematical model should be "close" to what is actually observed at the output of the black box.

STATEMENT OF THE PROBLEM

Given a black box, as shown in Figure 1:

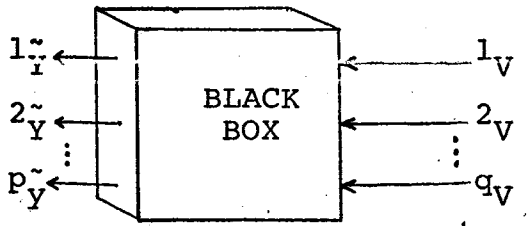


Figure 1

The black box has outputs $1_{\tilde{Y}}$ to $p_{\tilde{Y}}$ and deterministic inputs 1_V to q_V. If when the input $V = 0$, the output Y has an ergodic correlation function matrix; then it is desired to use a mathematical model of the form

$$X_t = C + \int_o^t F\, X_u\, du + (I) \int_o^t G\, d\beta_u + \int_o^t MV(u)\, du\,, \qquad (1)$$

$$Y_t = H\, X_t + m\,, \qquad (2)$$

** This work was partially supported by a NASA Grant NGR-001-090.

* Presently at General Electric, Corporate Research & Development Laboratories, Schenectady, New York.

$$E[C] = 0 \ , \ E[CC^T] \triangleq P \ ,$$

$$E[\beta_u] = 0 \ \forall \ u, \ E[\beta_s \beta_t^T] = Q \min \ (t,s), \ Q > 0,$$

as the model for the black box, where Y_t represents the p-dimensional random vector output of the black box, V(t) is the q-dimensional deterministic input of the black box, C is a Gaussian random vector initial condition, $\{\beta_u\}_{u \in \mathbb{R}}$ is a vector Wiener process independent of C, and (I) $\int_o^t G \, d\beta_u$ is an Ito stochastic integral. Equation 1 is a linear Ito stochastic integral equation. The Gaussian random vector initial condition C is completely characterized by its mean, $E[C] = 0$, and its variance, $E[CC^T] = P$. The Wiener process $\{\beta_u\}_{u \in \mathbb{R}}$ is completely characterized by its mean vector, $E[\beta_u] = 0 \ \forall \ u$, and its correlation function matrix, $E[\beta_s \beta_t^T] = Q \min (t,s)$. Essentially, the problem is to find the matrix parameter constants F, G, M, H, m, P, Q [in Appendix 4, it is shown how to reduce the number of unknown matrices by one] so that the solution of equations 1 and 2 satisfy the conditions that:

(i) the mathematical model has the same output mean vector as the black box,

(ii) the mathematical model has the same output correlation function matrix as the black box,

(iii) the output of the mathematical model has the same type of sample functions (either continuous or piecewise continuous) as the output of the black box. [When the black box has piecewise continuous sample functions, the Wiener process $\{\beta_u\}_{u \in \mathbb{R}}$ should be replaced by a centered Poisson process and the integral will still have meaning as a stochastic integral with all its desired properties (Anderson, 1966)].

The above conditions (i), (ii), (iii) are the criteria of "closeness" that were alluded to in the Abstract.

Overview of What is Done

A procedure is given for testing the actual black box under consideration to determine whether the assumed form of the mathematical model, equations 1 and 2, is acceptable. This test is based on manipulating the problem into a hypothesis testing situation, where the hypothesis is: "does the black box behave in a manner corresponding to the mathematical model of the particular form of equations 1 and 2?". A test

procedure is formulated and a method is derived for determining, a priori, the number of trials, N, required for a certain confidence, α, in the conclusion of whether to accept or reject the hypothesis. If the hypothesis is accepted, that is, if it is found that equations 1 and 2 do adequately describe the behavior of the black box, then the methods of the pseudo-inverse are employed to determine M. Prior to the determination of M, the other matrix parameters are determined by the method of Bucy and Joseph (see Appendices 4 and 5) with the input $V(t) = 0$.

PROCEDURE AND DERIVATIONS

The solutions of a system of equations that are equivalent to (1) and (2), which bear the same relationship that (1") has to (1') in Appendix 4, are

$$X_t = \Phi(t,0)\ C + (I) \int_0^t \Phi(t,\tau) G_o d\beta_\tau + \int_0^t \Phi(t,\tau) M V(\tau) d\tau ,$$

$$Y_t = H\Phi(t,0)\ C + (I) \int_0^t H\Phi(t,\tau) G_o d\beta_\tau + \int_0^t H\Phi(t,\tau) M V(\tau) d\tau ,$$

where $\Phi(t,\tau) = e^{F(t-\tau)}$ (see Appendix 1 for a pertinent discussion). For $t = T^*$, we have that

$$Y(T^*) = H\Phi(T^*,0)C + (I) \int_0^{T^*} H\Phi(T^*,\tau) G_o d\beta_\tau + \int_0^{T^*} H\Phi(T^*,\tau) M V(\tau) d\tau .$$

Since $\{\Phi(t,0)C + (I) \int_0^t \Phi(t,\tau) G_o d\beta_\tau\}_{t\varepsilon}$ is a Gauss-Markov process (Jazwinski, 1970, p. 111) because it is the solution of the stochastic integral equation

$$X_t = C + \int_0^t F\, X_\tau\, d\tau + (I) \int_0^t G_o d\beta_\tau ,$$

we have that for fixed $t = T^*$, $\Phi(T^*,0)C + (I) \int_0^{T^*} \Phi(T^*,\tau) G_o d\beta_\tau$ is a Gaussian random n-vector (recall that a Gaussian process is completely characterized by the first and second order distributions which are Gaussian and jointly Gaussian, respectively). Since premultiplying the above by a constant H to obtain $H\Phi(T^*,0)C + (I) \int_0^{T^*} H(T^*,\tau) G_o d\beta_\tau$ just represents a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^p$, the result is still a Gaussian random p-vector. Adding the constant term $\int_0^{T^*} H\Phi(T^*,\tau) V(\tau) d\tau$ to the above, for a specific $V(\cdot)$, to yield

$$Y(T^*) = H\Phi(T^*,0)\ C + (I)\int_0^{T^*} H\Phi(T^*,\tau)G_o d\beta_\tau + \int_0^{T^*} H\Phi(T^*,\tau)MV(\tau)d\tau,$$

which, is, again, a Gaussian random p-vector.

Since M is an (n x q) constant matrix, let $M=[m_1 | m_2 | \ldots | m_q]$, where each m_j is a column vector. For an input

$V^{\delta j}(t) \Delta [\delta_{1j},\delta_{2j},\ldots,\delta_{qj}]^T\ \delta(t-T^*/2)$, where δ_{ik} is the Kronecker delta, let $[Y^\delta(T^*)]^j$ represent the corresponding p-vector output. An expression for $[Y^\delta(T^*)]^j$ using the sifting property of the Dirac delta function is

$$[Y^\delta(T^*)]^j = H\Phi(T^*,0)\ C + (I)\int_0^{T^*} H\Phi(T^*,\tau)G_o d\beta_\tau + \int_0^{T^*} H\Phi(T^*,\tau)MV^{\delta j}(\tau)d\tau$$

$$= H\Phi(T^*,0)C + (I)\int_0^{T^*} H\Phi(T^*,\tau)G_o d\beta_\tau + \{\int_0^{T^*} H\Phi(T^*,\tau)\delta(\tau-\frac{T^*}{2})d\tau\}m_j$$

$$= H\Phi(T^*,0)C + (I)\int_0^{T^*} H\Phi(T^*,\tau)G_o d\beta_\tau + H\Phi(T^*,\frac{T^*}{2})m_j.$$

(The use of the "impulse function" in this analysis is just for convenience, later it will be replaced by any arbitrary, easily generated function, without affecting the conclusions of this section).

The mean of the impulse-excited output is

$$E[Y^\delta(T^*)]^j = H\Phi(T^*,0)E[C] + E\{(I)\int_0^{T^*} H\Phi(T^*,\tau)G_o d\beta_\tau\} + H\Phi(T^*,\frac{T^*}{2})m_j.$$

Since E[C] = 0 (see Appendices 1 and 2) and the expectation of the Ito integral is 0 (Varadhan, 1968, p. 129), we have that

$$E[Y^\delta(T)^*]^j = H\Phi(T^*,\frac{T^*}{2})m_j. \qquad (4)$$

The covariance matrix is

$$Cov[Y^\delta(T^*)]^j = \int_0^{T^*} H\Phi(T^*,u)G_o G_o^T\Phi^T(T^*,u)H^T du + H\Phi(T^*,0)P\Phi^T(T^*,0)H^T;$$

this result is obtained by using the fact that C and β_t are independent for $t \geq 0$ and other properties of the expectation

of the Ito integral and of the Ito integral squared (Varadhan, 1968, p. 129). Note that the covariance, $\mathrm{Cov}[Y(T^*)]^j$, is independent of j and is the same for any input.

With the aid of the above established results and the material of the Appendices, the procedure for modeling a black box having outputs and inputs will now be given in detail.

First, clamp the inputs $V(t) = 0$. If the output of the actual black box behaves in such a manner that the correlation function is ergodic, evaluate the unknown matrices H, F, G_Q, P, m as described in Appendix 5. Now everything is known in the mathematical model, equations 1 and 2, except M. We must first determine if the added term $\int_0^t MV(\tau)d\tau$ validly represents the manner in which the input affects the actual black box under discussion, (i.e., is linearity valid?). A test procedure will now be given and a criterion set to determine if this added input term is valid for the specific black box. If the criterion is satisfied, a method is given for determining M from the same data used in the test.

The test is essentially a test to see if the actual black box behaves linearly and in a time invariant manner. If it does, then superposition should apply. For fixed j, apply an input $V^{Aj}(t) = [\delta_{1j},\delta_{2j},\ldots,\delta_{qj}]^T u^A(t)$, where $u^A(t)$ is an arbitrary scalar function of time. Apply this input several times, say N times (it is this N, the number of trials that will be determined a priori as will be shown below), and record the corresponding N outputs ${}^i[Y^A(t)]^j$, $i = 1,\ldots,p$; the record should extend from $t = 0$ to $t = T^*$ where T^* is chosen for convenience. Divide the time interval $[0,T^*]$ into equispaced points, so that $\Delta = T^*/m$. For $t = k\Delta$, for every fixed $k\in[1,\ldots,m]$, average the N outputs ${}^i[Y^A(k\Delta)]^j$ to yield

$$ {}^i[\bar{Y}^A(k\Delta)]^j = \frac{1}{N}\sum_{n=1}^{N} {}^i[Y^A(k\Delta)]_n^j \qquad (i = 1,2,\ldots p). $$

Still for fixed j, apply an input

$V^{Bj}(t) = [\delta_{1j},\delta_{2j},\ldots,\delta_{qj}]^T u^B(t)$, where $u^B(t)$ is an arbitrary scalar function of time different from $u^A(t)$. Similarly obtain

$$ {}^i[\bar{Y}^B(k\Delta)]^j = \frac{1}{N}\sum_{n=1}^{N} {}^i[Y^A(k\Delta)]_n^j \qquad (i = 1,2,\ldots,p). $$

Finally, for the same fixed j, apply an input $V^{(A+B)j}(t) = [\delta_{1j},\delta_{2j},\ldots,\delta_{qj}]^T [u^A(t) + u^B(t)]$. Then, again for every fixed k, average the outputs to obtain

$$ {}^{i}[\bar{Y}^{A+B}(k\Delta)]^{j} = \frac{1}{N}\sum_{n=1}^{N} {}^{i}[Y^{A+B}(k\Delta)]_{n}^{j} \qquad (i = 1,\ldots,p). $$

This same test procedure is followed for each of the q components of the input.

If the black box were linear and the inputs entered in a time-invariant manner, then $\{[Y^{A}(k\Delta)]_{n}^{j}\}_{n=1}^{N}$, for a fixed k, is a sample of size N from a Gaussian population where the population has a known variance (eq. 5) of

$$ \mathrm{Cov}[Y^{A}(k\Delta)]^{j} = \int_{o}^{k\Delta} H\Phi(k\Delta,u)G_{o}G_{o}^{T}\Phi^{T}(k\Delta,u)H^{T}du + H\Phi(k\Delta,0)P\Phi^{T}(k\Delta,0)H^{T}; $$

and unknown mean

$$ E[Y^{A}(k\Delta)]^{j} = [\int_{o}^{k\Delta} H\Phi(k\Delta,\tau)u^{A}(\tau)d\tau]m_{j}, $$

where m_j is an unknown vector.

For $t = k\Delta$, for any fixed k, the probability that the sample mean $[\bar{Y}^{A}(k\Delta)]^{j}$ of a sample of size N is within ε of the population mean μ^{Aj} is given by

$$ P[\frac{1}{N}|| [\bar{Y}^{A}(k\Delta)]^{j} - \mu^{Aj} ||^{2}_{(\mathrm{Cov}[\bar{Y}^{A}(k\Delta)]^{j})^{-1}} \leq \varepsilon^{2}] $$

$$ = P[\frac{1}{N}||\bar{Z}||_{I}^{2} \leq \varepsilon^{2}] = P[||\bar{Z}||_{I}^{2} \leq N\varepsilon^{2}] = \alpha $$

where use has been made of the transformation $Z = \{[Y^{A}(k\Delta)]^{j} - \mu^{Aj}\}\{\mathrm{Cov}[Y^{A}(k\Delta)]\}^{-1/2}$ where Z is a p-dimensional Gaussian vector having $p_{Z}(\beta) = (1/(2\pi)^{p/2}\exp\{-(1/2)||\beta||_{I}^{2}\}$ as a probability density function. Note that for $t = k\Delta$, for any fixed k, the problem transforms into the same problem in Z with the same sample size N. Since $[\bar{Y}^{A}(k\Delta)]^{j}$ was Gaussian, $\bar{Z}$ is Gaussian since only the above linear transformation was used.

Since the ${}^{i}Z$'s are independent and have a Gaussian distribution, their squares have a χ^2 distribution and are also independent. The sum of p independent random variables having a χ^2 distribution is also χ^2 distributed with p degrees of freedom, χ_{p}^{2}. This distribution is well tabulated. From

the χ^2_p table, it is possible to calculate, a priori, the sample size N required so that we have α confidence that the sample mean [a maximum likelihood, sufficient, unbiased, "efficient", and "consistent" statistic for the population mean (Hogg and Craig, 1970, p. 255)] is within ε of the true population mean, where α and ε are set in advance. The number of degrees of freedom p is the number of outputs of the black box.

Returning to test the linearity hypothesis, if the actual black box were linear, then

$$||[\bar{Y}^A(k\Delta)]^j - \mu^{Aj}||^2_R - 1 = ||\bar{Z} - \mu_Z||^2_I \leq \varepsilon^2,$$

and we would have:

$$||[\bar{Y}^A(k\Delta)]^j + [\bar{Y}^B(k\Delta)]^j - [\bar{Y}^{A+B}(k\Delta)]^j||_R - 1$$

$$\leq ||[\bar{Y}^A(k\Delta)]^j - \mu^{Aj}||_R - 1 + ||[Y^B(k\Delta)]^j - \mu^{Bj}||_R - 1$$

$$+ ||[\bar{Y}^{A+B}(k\Delta)]^j - \mu^{(A+B)j}||_R - 1 + ||\mu^{(A+B)j} - \mu^{Aj} - \mu^{Bj}||_R - 1;$$

$$||[\bar{Y}^A(k\Delta)]^j - \mu^{Aj}||^2_R - 1 \leq \varepsilon^2 N; \text{ etc., } ||\mu^{Aj} + \mu^{Bj} - \mu^{(A+B)j}||^2_R - 1 = 0$$

$$||[\bar{Y}^A(k\Delta)]^j + [\bar{Y}^B(k\Delta)]^j - [\bar{Y}^{A+B}(k\Delta)]^j||^2_R - 1 \leq 9\varepsilon^2 N.$$

(R was used in the above to represent the appropriate covariance). Define

$$\gamma = \sum_{k=1}^{m} ||[\bar{Y}^A(k\Delta)]^j + [\bar{Y}^B(k\Delta)]^j - [\bar{Y}^{A+B}(k\Delta)]^j||^2_R - 1.$$

Therefore, to accept the hypothesis that the actual black box is linear and time invariant it must be that

$$\gamma \leq 9m\varepsilon^2 N \quad \text{for} \quad j = 1,2,\ldots,q.$$

If $\gamma > 9m\varepsilon^2 N$, the conclusion is that the black box under consideration cannot be modeled by the methods presented.

The criterion can be interpreted pictorially in (p + 1) - dimensional Euclidean space in Figure 2. Let the solid line represent $[\bar{Y}^A(k\Delta)]^j + [\bar{Y}^B(k\Delta)]^j$ in p-dimensional Euclidean space. When ε, α, and consequently N have been specified, an ε sheath is defined around the solid line in p-dimensional Euclidean space as represented by the dashed lines in Figure 2.

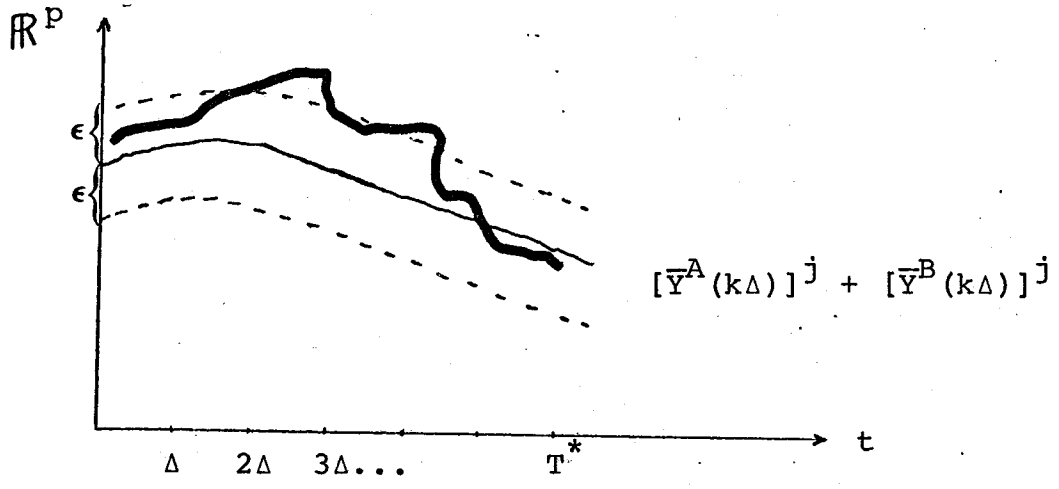


Figure 2

If $[\bar{Y}^{A+B}(k\Delta)]^j$ is within the ε-sheath to a degree that, at the m time points at which it is checked, the sum of the excursions outside the ε-sheath are compensated for by its proximity to $[\bar{Y}^A(k\Delta)]^j + [\bar{Y}^B(k\Delta)]^j$ at other times such that $\gamma \leq 9m\varepsilon^2 N$, the hypothesis that the actual black box is linear is accepted.

Now returning to the problem of identifying M, we have that

$$[\bar{Y}^A(T^*)]^j = \frac{1}{N}\sum_{n=1}^{N}[Y^A(T^*)]_n^j \doteq E[Y^A(T^*)]^j = [\int_0^{T^*} H\Phi(T^*,\tau)u^A(\tau)d\tau]m_j.$$

Now, $[\bar{Y}^A(T^*)]^j$ is a known p-vector, H is a known (p x n) matrix, and $\Phi(T^*,T^*/2) = e^{FT^*/2}$ is a known (n x n) matrix, and $u^A(t)$ is a known deterministic scalar control; therefore,

$$[\bar{Y}^A(T^*)]^j = [\int_0^{T^*} H\Phi(T^*,\tau)u^A(\tau)d\tau]m_j$$

is of the form of the algebraic equation $Y = Ax$, where A is a known (p x n) matrix, x is an unknown q-vector, and Y is a known p-vector. We wish to solve the above equation of m_j, which corresponds to solving $Y = Ax$ for x. If A were square and nonsingular, the solution would be $x = A^{-1}Y$. Even when A^{-1} does not exist, it is desirable to solve $Y = Ax$ in some approximate sense; the theory of the pseudo-inverse and how to find it is given in Appendix II of Aoki (1967, pp. 318-324). Let A^+ represent the pseudo-inverse of the (p x n) matrix A; then the solution x is $x = A^+Y$. The above analysis

can be applied for each $j(j = 1,\ldots,q)$ so that each m_j $(j = 1,\ldots,q)$ is determined. The $(n \times q)$ matrix $M = [m_1\ m_2\ \ldots\ m_q]$, has been identified. The problem of modeling a black box by a linear, constant coefficient, stochastic integral equation has now been solved.

REFERENCES

1. Aoki, M., Optimization of Stochastic Systems, Academic Press, New York, 1967.
2. Anderson, B.D.O., "A System Theory Criterion for Positive Real Matrices", J.S.I.A.M. Control, Vol. 5, pp. 171-182, May 1967.
3. Anderson, M.W., "Stochastic Integral Equations", Ph.D. Thesis, University of Tennessee, 1966.
4. Bucy, R.S., and Joseph, P.D., Filtering for Stochastic Processes with Applications to Guidance, Interscience Publishers, a Division of John Wiley & Sons, Inc., New York, 1968.
5. Davis, M.C., "Factoring the Spectral Matrix", I.E.E.E. Trans. on Automatic Control, Vol. A.C.-12, pp. 296-305, October 1963.
6. Jazwinski, A.N., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
7. Kalman, R.E., "Mathematical Description of Linear Systems", J.S.I.A.M. Control, Ser. a., Vol. 1, Nov. 2, 1963.
8. Kerr, T.H., "Applying Stochastic Integral Equations to Solve a Particular Stochastic Modeling Problem", Ph.D. Thesis, University of Iowa, Iowa City, Iowa, 1971.
9. Laning, J.H., and Battin, R.H., Random Processes in Automatic Control, McGraw-Hill Book Co., Inc., New York, 1956.
10. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Co., Inc., New York, 1965.
11. Solodovnikov, V.V., Introduction to the Statistical Dynamics of Automatic Control Systems, Translation Edited by J.B. Thomas and L.A. Zadeh, Dover Publications, Inc., New York, 1960.
12. Wong, E., and Zakai, M., [1], "On the Convergence of Ordinary Integrals", Ann. of Math. Stat., Vol. 36, pp. 1560-1564, 1965.
13. Wong, E., and Zakai, M., [2], "On the Relation Between Ordinary and Stochastic Differential Equations", Int. J. Engng. Sci., Vol. 3, pp. 213-229, 1965.

APPENDIX 1: The argument for the zero mean restriction follows. The equations of the mathematical model

$$X_t = C + \int_o^t F X_u \, du + (I) \int_o^t G \, d\beta_u$$

$$Y_t = H X_t$$

have the solutions

$$X_t = \Phi(t,0)\ C + (I) \int_o^t \Phi(t,u)\ G\ d\beta_u$$

$$Y_t = H\Phi(t,0)\ C + (I) \int_o^t H\Phi(t,u)\ G\ d\beta_u,$$

where $\Phi(t,u)=\exp F(t-u)$, as can be verified by applying Ito's lemma to each scalar component of the vector solution for X to obtain the original stochastic integral equations (Bucy and Joseph,1969,p.24). To satisfy the condition that the output of the black box have the same mean as the output of the model requires that

$$m = E[\tilde{Y}(t)] = E[Y(t)] = H\Phi(t,0)\ E[C], \forall t.$$

Since m is a constant and $\Phi(t,0)$ is varying with time, in general, this equation is satisfied if and only if $m = 0 = E[C]$.

APPENDIX 2: The zero mean restriction can be removed by assuming a mathematical model of the same form except that $Y_t = H X_t + d$. Now $m = E[\tilde{Y}(t)] = E[Y(t)] = H\Phi(t,0)\ E[C] + d, \forall t$ is satisfied if and only if $m = d$ and $E[C] = 0$. This causes no added difficulty since the covariance can be transformed, factored, and manipulated in the same way that the correlation function is transformed, factored, and manipulated in Bucy and Jospeh (1968, pp.25-26, 29-42). Using this approach with $Y_t = H X_t + d$, we have that the mathematical model and the black box have the same mean vector and covariance function matrix; therefore, they have the same correlation function.

APPENDIX 3: From the solution of the stochastic integral equation mentioned in Appendix 1, from the unique properties of the Ito integral (Varadhan, 1967-68, p.129) or Jazwinski, 1971,p.99) and the fact that the black box is wide-sense stationary, the following equations are derived which hold true for the problem of modeling the black box with no inputs (Bucy and Joseph, 1968, pp.39-42).

$$S_{\tilde{y}\tilde{y}}(p) = H(pI-F)^{-1}\ G_o G_o^T\ (-pI-F^T)^{-1}\ H^T,$$

$$E[CC^T] = E[X(t)X^T(t)] \triangleq P, \text{ where } FP + PF^T + G_o G_o^T = 0,$$

$$W^T(p) = H(pI-F)^{-1}\ G_o, \text{ where } S^T_{\tilde{y}\tilde{y}}(p) \text{ factors}$$

into $S^T_{\tilde{y}\tilde{y}}(p) = W^T(-p)\ W(p)$ (Please see Appendix 5d).

Appendix 4: If $GQG^T = G_o G_o^T$, then the system of equations used as a mathematical model in this paper,

$$(1') \quad X_t = C \int_o^t F X_u \, du + (I) \int_o^t G \, d\beta_u$$

$$(2') \quad Y_t = H X_t$$

$$E[C] = 0; \; E[CC^T] = P; \; E[\beta_t] = 0, \forall t$$

$$E[\beta_t \beta_s^T] = Q \min(t,s), \quad Q \geq 0,$$

where F, G, H, P, Q are the five unknown matrices, can be replaced by the equivalent system of equations

$$(1'') \quad X_t = C + \int_o^t F X_u \, du + (I) \int_o^t G_o \, d\beta_u$$

$$(2'') \quad Y_t = H X_t$$

$$E[C] = 0; \quad E[CC^T] = P; \quad E[\beta_t] = 0, \forall t$$

$$E[\beta_t \beta_s^T] = I \min(t,s),$$

where F, G_o, H, P are only four unknown matrices. This replacement can be done since the solution of the two X_t stochastic integral equations in both (1') and (1") are Gauss-Markov processes (Jazwinski,1970,p.79) and as Markov processes do not require specification of the entire family of all finite dimensional distributions for a complete characterization; knowledge of the transition probability densities of the form $p(X,t \mid y,s)$ suffices. The solutions of the X_t stochastic integral equations in (1') and (1") have transition probability density functions which satisfy the forward Kolmogorov or Fokker-Planck equations

$$\frac{\partial p(X,t|X_o,t_o)}{\partial t} = -(\frac{\partial}{\partial X})^T [FX p(X,t|X_o,t_o)] + \frac{1}{2} tr\{(\frac{\partial}{\partial X})(\frac{\partial}{\partial X})^T [GQG^T \; p(X,t|X_o,t_o)]\}$$

and

$$\frac{\partial p(X,t|X_o,t_o)}{\partial t} = -(\frac{\partial}{\partial X})^T [FX_p(X,t|X_o,t_o)] + \frac{1}{2} tr\{(\frac{\partial}{\partial X})(\frac{\partial}{\partial X})^T [G_o G_o^T \; p(X,t|X_o,t_o)]\}$$

respectively (Jazwinski,1970,p.130). If $G_o G_o^T = GQG$, these two Kolmogorov equations are the same, so their solutions are the same; therefore, the X_t processes in (1') and (1") are the same process.

APPENDIX 5: The procedure for identifying the unknown matrices of (1"), (2") from measurements made at the output of a black box follows below.

a. Obtain an extensive time record of $\tilde{y}(t)$, the actual output of the black box.

b. Process this data by time averaging to obtain the correlation function matrix, $R_{\tilde{y}\tilde{y}}(\tau)$, and the mean, $E[y(t)]$. The only assumption on the whole procedure is that the correlation function matrix is ergodic. (However, this one assumption implies ergodicity of the mean and wide-sense stationarity [Papoulis, 1965,p.329].)

c-i. Approximate this correlation function matrix in the τ-domain by an exponential series (Lang and Battin, 1956,p.381). Then take the bilateral Laplace transform of the approximating correlation function to obtain the power spectral density matrix, $S_{\tilde{y}\tilde{y}}(p)$. Since the approximating

correlation function consists of exponential terms, the power spectral density matrix has elements that are rational functions (i.e., ratios of polynomials).

c-ii. An alternate procedure to c-i. Instead of approximating in the τ-domain, first obtain the power spectral density matrix by taking the bilateral Laplace transform of the correlation function matrix, then obtain an approximation for the elements of the matrix in terms of rational functions by any one of the four methods mentioned in Solodovnikov, Chapter V; Methods of Approximation of Spectral Density Functions by Meromorphic Functions, 1960.

d. Since every power spectral density matrix which has elements that are ratios of polynomials satisfies the sufficient conditions for applying the matrix factorization procedure (Kerr,1971,p.330-333), factor $S^T_{\tilde{y}\tilde{y}}(p)$ into $S^T_{\tilde{y}\tilde{y}}(p) = W^T(-p)\, W(p)$, where $W(p)$ is analytic in p in $\mathrm{Re}(p) \geq 0$. This factorization can be accomplished by either of the two methods presented by Youla (1961, method I is theorem 2, method II is theorem 3) or by the method of Davis (1963,pp. 296-305). Since $S^T_{\tilde{y}\tilde{y}}(p)$ is known, $W^T(p)$ is known. Let $W^T(p) = H(pI-F)^{-1}G_o$, where the triple (H,F,G_o) is to be determined.

e. A triple (H,F,G_o) [not necessarily unique] can be found which satisfies $W^T(p) = H(pI-F)^{-1}G_o$, and such that (H,F) is observable, (F,G) is controllable, and F is stable, either by the methods of obtaining a realization from a "transfer function" as mentioned in Kalman(1963,p.152) or by an original method in Kerr (1971, p.255).

f. Once (H,F,G_o) is known, the solution of $0 = FP + PF^T + G_oG_o^T$, where (F,G_o) is completely controllable, is known: $P = \int_o^\infty e^{Ft}\, G_oG_o^T\, e^{F^Tt}\, dt$ (Anderson, 1967,p.173).

g. From Appendix 2, we have that $E[\tilde{y}(t)] = m$; therefore the matrices H, F, G_o, P, m in the mathematical model of the black box have all been determined. The modeling problem for the black box without inputs is solved.

APPENDIX 6: From the method of Appendix 5 and from the main method of this paper, mathematical models in terms of Ito stochastic integrals were obtained. Eventually, these mathematical models will be used to make computations which represent what actually occurs at the outputs of the black box. Digital computers do not normally perform Ito integrations, but this can be resolved by using the methods of Wong and Zakai which relate Ito integrals to ordinary integrals (1965 [1], p. 1560 and 1965 [2], p. 213).