TRANSPORT THEORY AND STATISTICAL PHYSICS, 8(3), 163-168 (1979)

A VARIATIONAL PRINCIPLE FOR THE LINEAR FILTER MATRIX AND AN INTERPRETATION FOR THE MAXIMUM VALUE OF THE FUNCTIONAL

M. Kanal* and H.E. Moses**

Center for Atmospheric Research
College for Pure and Applied Science
University of Lowell
Lowell, MA 01854

S.K. Mitter

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

A functional of a trial matrix and suitable correlation matrix is constructed which is an absolute maximum when the trial matrix satisfies the Wiener-Hopf equation for the filter matrix. The maximum value of the functional is essentially the sum of the squares of the minimum errors of the observations and is thus of interest in its own right.

---

*Present address: Physics Department, Clark University, Worcester, MA, 01610.

**Research sponsored by the Air Force Office of Scientific Research. Air Force Systems Command, USAF, under Grant No. AFOSR-77-3169. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

The filter equation (i.e., Wiener-Hopf equation) bears a close resemblance to the Gelfand-Levitan equation and, in fact, this resemblance led to the variational principle of the present paper from that for the Gelfand-Levitan equation.

1. Inner Products for Matrices. Positive-Definite Operators. Abstract Form of the Variational Principle.

Let us consider matrices $H(s,t,) = \{H_{ij}(s,t)\}$ such that $i,j = 1,2,..p$ and the continuous variables s and t have the range $a<s,\ t<b$. The hermitian adjoint of $H(s,t)$ will be denoted by a prime

$$H'(s,t) = \{H^*_{ji}(s,t)\}. \tag{1}$$

We use the usual notation for product of two matrices:

$$A(s,t)B(s',t') = \{\sum_k A_{ik}(s,t)B_{kj}(s',t')\} \tag{2}$$

We use column vectors $y(t) = \{y_i(t)\}$ where the components $y_i(t)$ are complex. The corresponding complex conjugate row vector is denoted by $y'(t)$. The inner product of the vector $z(t)$ and $y(t)$ is given by

$$(z,y) = \int_a^b z'(t)y(t)dt, \tag{3}$$

where the integrand uses the form Eq. (2) for the matrix product of a row vector by a column vector. A matrix $A(s,t)$ is said to be positive-definite in the vector sense, if $A(s,t)$ is hermitian, i.e. $A_{ij}(s,t) = A^*_{ji}(t,s)$ and if

$$(y,Ay) \geq 0, \tag{4}$$

for all $y(t)$, in which the equality holds only for $y(t)\equiv 0$. In Eq. (4) the vector

$$Ay(t) = \{\sum_j \int_a^b A_{ij}(t,s)y_j(s)ds\}.$$

We now want to introduce the notion of inner products of matrices of one variable $B(t) = \{B_{ij}(t)\}$ where the range of i,j and t are as before and where the prime continues to mean hermitian adjoint as in Eq. (1), where, however, only one continuous variable appears. The inner product of $B(t)$ and $C(t)$ is defined by

$$(B,C) = \mathrm{tr} \int_a^b B'(t)C(t)dt, \tag{5}$$

where the usual matrix product is meant in the integrand. It is seen that this matrix inner product has the usual inner product properties. We can define the matrix $(AB)t$, using the matrix $A(t,s)$ as the kernel of an integral operator, by

$$AB(t) = \{\sum_k \int_a^b A_{ik}(t,s,)B_{kj}(s)ds\} \equiv \int_a^b A(t,s,)B(s)ds. \qquad (6)$$

The operator A with kernel A(t,s) is said to hermitian in the matrix sense if for all matrices B(t), C(t)

$$(B,AC) = (C,AB)^*. \qquad (7)$$

The following theorem is easily shown to hold:

An operator A is hermitian in the matrix sense if and only if it is hermitian in the earlier (vector) sense.

An operator A is said to be positive-definite in the matrix sense if it is hermitian and if

$$(B,AB) \geq 0 \qquad (8)$$

for all matrices B(t), where the equality holds only if B(t) = 0. The following theorem holds:

An operator A is positive-definite in the matrix sense if and only if it is positive-definite in the usual vector sense.

We can now state the fundamental variational principle which is proved abstractly in Ref. 1 and given again in Ref. 2. Let B(t) be the solution of the integral equation

$$AB(t) = C(t), \qquad (9)$$

where the matrix B(t) is the unknown matrix, the matrix C(t) is given and A represents a positive-definite integral operator with the matrix kernel A(t,s).

Then the following theorem holds. Let N(t) be any p x p matrix function of t. The functional F(N) defined by

$$F(N) = (N,C) + (C,N) - (N,AN) \qquad (10)$$

is an absolute maximum if and only if N(t) is the solution B(t) of Eq. (9).

2. Application to the Filter Equation (Wiener-Hopf Equation).

For the sake of brevity, we shall refer to equations in Kailath's excellent monograph (Ref. 3) as we need them. The equations from Ref. 3, when used in the present paper will be prefixed by the letter K.

The equation for the filtering matrix is

$$h(t,s) + \int_{t_0}^{t} h(t,\tau)K(\tau,s)d\tau = K(t,s) \qquad (11)$$

{Eq. (11) is the same as Eq. (K10).} The kernel K(t,s) is given and is defined in Eq. (K5) and (K6). Of particular interest is the fact the integral operator with the kernel $R_y(t,s) = I_p\delta(t-s) + K(t,s)$ (Eq. (K6)) is positive-definite in both the vector and matrix senses.

We note that Eq. (11) is a matrix generalization of the Gelfand-Levitan equation for which the operator corresponding to $R_y(t,s)$ is also positive-definite (see Ref. 2). It was our ability to obtain a variational principle for the Gelfand-Levitan equation which led us to the variational principle of the present paper.

Eq. (11) may be written as

$$h'(t,s) + \int_{t_o}^{t} K(s,\tau)h'(t,\tau)d\tau = K'(t,s). \tag{12}$$

In obtaining Eq. (12) from Eq. (11) we have used the fact that the operator with the kernel $R_y(t,s)$ is hermitian. Eq. (12) is precisely of the form Eq. (9) where t in Eq. (9) is replaced by s and A is the operator whose kernel $A(s,\tau)\equiv R_y(s,\tau)$. In Eq. (12) the variable t is simply a parameter in the Freholm equation for the filter matrix.

Now, according to Section 1, the functional

$$F(N,t) = \mathrm{tr}\ \{\int_{t_o}^{t} N(t,s)K'(t,s)ds + \int_{t_o}^{t} K(t,s)N'(t,s)ds - \int_{t_o}^{t} N(t,s)N'(t,s)ds - \int_{t_o}^{t} ds \int_{t_o}^{t} d\tau\ N(t,s)K(s,\tau)N'(t,\tau)\}, \tag{13}$$

where N(t,s) is a trial matrix for h(t,s). According to the theorem of the previous section, the functional F(N,t) reaches its maximum value when N(t,s) = h(t,s).

It is of great interest to find the maximum value of the functional, i.e. to evaluate F(,t). It is readily seen from Eq. (12) that

$$F(h,t) = \mathrm{tr}\{K(t,t) - h(t,t)\}. \tag{14}$$

Hence

$$\mathrm{tr}\ h(t,t) = \mathrm{tr}\ K(t,t) - F(h,t). \tag{15}$$

In the next section we shall show that tr h(t,t) gives the minimum least square error for the observations. It is interesting to note that tr h(t,t) plays the role of K(x,x) of the Gelfand-Levitan equation. In the Gelfand-Levitan equation K(x,x) is the integral of the scattering potential which is the quantity to be found, whereas tr h(t,t) gives a quantity of great interest in filter theory. {One should not confuse K(s,t) of the present paper with the Gelfand-Levitan kernel K(x,y) of Ref. 2.}

3. Significance of tr h(t,t).

To find the significance of tr h(t,t) we turn to the Kalman filter, as described in Ref. 3. From Eq. (K22)-(K24)

$$\frac{d\hat{x}(t)}{dt} = \{F(t) - K(t)H(t)\}\hat{x}(t) + K(t)y(t). \tag{16}$$

Hence

$$\hat{x}(t) = \int_{t_0}^{t} W(t,s)K(s)y(s)ds, \tag{17}$$

where W(t,s) is the fundamental matrix solution of Eq. (16) when $y(t)\equiv 0$.

As is well known

$$W(t,t) = I, \tag{18}$$

where I is the identity matrix with dimension corresponding to that of x(t).

From Eq. (K22)

$$\hat{z}(t) = \int_{t_0}^{t} H(t)W(t,s)K(s)y(s)ds. \tag{19}$$

However, by definition of the filter matrix

$$\hat{z}(t) = \int_{t_0}^{t} h(t,s)y(s)ds. \tag{20}$$

Hence, from Eq. (18), (K26), (K19a), together with Eq. (K25),

$$\mathrm{tr}\ h(t,t) = \mathrm{tr}\ E\{z(t) - \hat{z}(t)\}\{z(t) - \hat{z}(t)\}' + \mathrm{tr}\ H(t)G(t)C(t). \tag{21}$$

Since G(t), H(t)C(t) are known quantities, we have the result that tr h(t,t) gives the square of the error of observation. However, the filter equation was obtained by <u>minimizing</u> this error with the use of Eq. (20). Hence tr h(t,t) gives the <u>minimum error</u> for the Kalman filter. Hence the use of a good trial function N(t,s) leads to a good upper bound for tr h(t,t) and hence the minimum error for the filter.

Before leaving the topic, it should be noted that the variational principle can be extended to more general kernels than h(t,s), for example to the kernel H(t,s) of Eq. (K9).

In Ref. 4 another variational principlie is derived for the filter matrix using a different approach. The functional used in Ref. 4 is essentially the integral over t of our functional.

REFERENCES

1. Mikhlin, S.G., <u>Variational Methods in Mathematical Physics</u>, Macmillan, New York (1964). (Trans. by T. Boddington, edited by L.I.G. Chambers.)

2 Kanal, M. and Moses, H.E., "A Variational Principle for the Gelfand-Levitan Equation and the Korteweg-deVries Equation," Jour. Math. Phys. 18, 2445-2447 (1977).

3. Kailath, T., "A View of Three Decades of Linear Filtering Theory," IEEE Transactions on Information Theory, IT-20, 146-181 (1974).

4. Dyson, F.J., "Old and New Approaches to the Inverse Scattering Problem," Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann. Edited by E.H. Lieb, B. Simon, and A.S. Wightman. Princeton Series in Physics, Princeton, New Jersey. (1976).