ESD-TR-66-22
ESTI FILE COPY

ESD-TDR-66-22

ESD RECORD COPY
RETURN TO
SCIENTIFIC & TECHNICAL INFORMATION DIVISION
(ESTI), BUILDING 1211

ESD ACCESSION LIST
ESTI Call No. AL 49854
Copy No. 1 of 1 cys.

Technical Note 1966-7

J. F. Claerbout

A Summary, by Illustrations, of Least Squares Filters with Constraints

31 January 1966

Prepared for the Advanced Research Projects Agency
under Electronic Systems Division Contract AF 19(628)-5167 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

AD0629961

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This research is a part of Project Vela Uniform, which is sponsored by the U.S. Advanced Research Projects Agency of the Department of Defense; it is supported by ARPA under Air Force Contract AF 19(628)-5167 (ARPA Order 512).

This report may be reproduced to satisfy needs of U.S. Government agencies.

Distribution of this document is unlimited.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

A SUMMARY, BY ILLUSTRATIONS, OF LEAST SQUARES FILTERS WITH CONSTRAINTS

J. F. CLAERBOUT

Group 64

TECHNICAL NOTE 1966-7

31 JANUARY 1966

LEXINGTON MASSACHUSETTS

ABSTRACT

Several methods of combining a number of time series into a single series are discussed. They are all individual filtering followed by summation and are somewhat like Wiener filtering in that a least squares criterion is used to define the filter coefficients. They differ from Wiener filtering in that signal information is given in the form of various linear constraints on the filter coefficients rather than being given as a signal correlation function. The formulas are worked out explicitly for the case of two time series and three filter points and presented in such a way as to make generalization clear.

Accepted for the Air Force
Franklin C. Hudson
Chief, Lincoln Laboratory Office

DEFINITIONS

Let $f_1(\tau)$ be an ordinary sampled-data filter on an $x_1(t)$ input time series and let $f_2(\tau)$ be likewise on $x_2(t)$. The final output $y(t)$ is the sum of the separate outputs. Mathematically we have

$$y(t) = \sum_{k=1}^{2} \sum_{\tau=0}^{\ell-1} x_k(t-\tau)\, f_k(\tau) \tag{1}$$

For the sake of illustration, we will use only $n = 2$ time filters and 2 time series and we will further choose three point filters, i.e., $\ell = 3$. Results will be presented so that generalization should be clear to the reader. Formula (1) may be written explicitly as

$$\begin{aligned} y(t) &= [\ x_1(t) \text{ filtered} \qquad\qquad\qquad\qquad] + [\ x_2(t) \text{ filtered} \qquad\qquad\qquad\qquad] \\ &= [\ x_1(t)\, f_1(0) + x_1(t-1)\, f_1(1) + x_1(t-2)\, f_1(2)] + [\ x_2(t)\, f_2(0) + x_2(t-1)\, f_2(1) + x_3(t-2)\, f_2(2)] \end{aligned} \tag{2}$$

Illustration #1 Minimum variance (power)

It is desired to find the filter coefficients which will minimize the power in the filter's output.

$$P = \text{power} = \sum_t y(t)^2 \tag{3}$$

If the filter coefficients are zero, the output power will be zero so the problem is trivial and not fully formed. Later we will introduce constraints which bring about more interesting solutions. However, it is worthwhile going through with this problem as it contains the basic mathematics in simplest form. The situation can be compared with Laplace's equation where the solution is a constant until boundary values are considered.

The power is a quadratic function of the filter coefficients. To find the minimum we treat each coefficient as a variable and set partial derivatives of power with respect to the filter coefficients equal zero. Because of this the method is called least squares. This will give a set of $n \times \ell = 6$ equations for the six unknowns.

For example, to get one equation we insert (2) into (3) and differentiate with respect to $f_1(0)$.

$$0 = \frac{\partial}{\partial f_1(0)} \sum_t y(t)^2 = \sum_t 2\, x_1(t)\, y(t) \tag{4}$$

$$0 = r_{11}(0)\, f_1(0) + r_{12}(0)\, f_2(0) + r_{11}(-1)\, f_1(1) + r_{12}(-1)\, f_2(1) + r_{11}(-2)\, f_1(2) + r_{12}(-2)\, f_2(2) \tag{5}$$

where we have introduced the definition of correlation

$$r_{ij}(\tau) = \sum_t x_i(t)\, x_j(t+\tau) \tag{6}$$

If the data x(t) is imagined to have infinite duration in time, it is necessary to introduce a normalizing factor into (6). In practice it is unimportant.

Likewise differentiating with respect to each of the other six coefficients of f gives six equations which may be arranged as:

$$\begin{bmatrix} r_{11}(0) & r_{12}(0) & r_{11}(-1) & r_{12}(-1) & r_{11}(-2) & r_{12}(-2) \\ r_{21}(0) & r_{22}(0) & r_{21}(-1) & r_{22}(-1) & r_{21}(-2) & r_{22}(-2) \\ r_{11}(1) & r_{12}(1) & r_{11}(0) & r_{12}(0) & r_{11}(-1) & r_{12}(-1) \\ r_{21}(1) & r_{22}(1) & r_{21}(0) & r_{22}(0) & r_{21}(-1) & r_{22}(-1) \\ r_{11}(2) & r_{12}(2) & r_{11}(1) & r_{12}(1) & r_{11}(0) & r_{12}(0) \\ r_{21}(2) & r_{22}(2) & r_{21}(1) & r_{22}(1) & r_{21}(0) & r_{22}(0) \end{bmatrix} \begin{bmatrix} f_1(0) \\ f_2(0) \\ f_1(1) \\ f_2(1) \\ f_1(2) \\ f_2(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

We abbreviate this set as

$$RF = 0$$

If we take the 6×6 matrix of coefficients and partition it into a 3×3 matrix of 2×2 submatrices, we have

$$R = \begin{bmatrix} R(0) & R(-1) & R(-2) \\ R(1) & R(0) & R(-1) \\ R(2) & R(1) & R(0) \end{bmatrix} \tag{8}$$

where $R(\tau)$ is the 2×2 matrix

$$R(\tau) = \begin{bmatrix} r_{11}(\tau) & r_{12}(\tau) \\ r_{21}(\tau) & r_{22}(\tau) \end{bmatrix} = R^T(-\tau) \tag{9}$$

It may be noticed that all the submatrices on the main diagonal are identical to each other. The same may be said of other diagonals. Furthermore, $R(\tau) = R^T(-\tau)$ By inspection the matrix is symmetric. The matrix is positive semi-definite because the power output of the filter is a quadratic function of the f coefficients $F^T RF$, which must be positive for any F. All this is the Toeplitz property of R.

The solution to this problem is that all the filter coefficients equal zero because the right side of Equation (7) equals zero, as we anticipated. The importance of this example comes thru the introduction of the Toeplitz matrix which plays a central role in time series analysis and the examples to follow.

Illustration #2 The constraint $f_1(0) = 1$

Suppose we wish to minimize $\sum_t y(t)^2$ as before but to make the additional constraint $f_1(0) = 1$. One method would be to proceed as before replacing $f_1(0)$ by 1 wherever it occurs and dropping the equation obtained by $\frac{\partial \text{ power}}{\partial f_1(0)} = 0$.

The trouble with that approach is that we will no longer have a Toeplitz matrix like that in formula (7) but rather a matrix with much less symmetry. For this reason we will introduce the method of Lagrange multipliers. It will still give us a Toeplitz matrix like (7) for a large class of problems with linear and quadratic constraints. Several problems of this class will be discussed in later examples.

The Lagrange method, as applied to illustration #2, is to minimize

$$E = [\sum y(t)^2] - \lambda [f_1(0) - 1] \tag{10}$$

by setting it's derivatives with respect to each of the six coefficients of f and with respect to λ equal to zero. Then solve the seven equations for the f's and λ.

To see that the Lagrange method gives the desired result, note that $\frac{\partial E}{\partial \lambda}$ gives the constraint equation. Then at the minimum of E, λ will multiply $[f_1(0) - 1] = 0$. Thus minimizing E is equivalent to minimizing $\sum_t y(t)^2$.

Partial derivative of E with respect to the f coefficients gives

$$\begin{bmatrix} & & \\ & R & \\ & & \end{bmatrix} \begin{bmatrix} \\ F \\ \\ \end{bmatrix} = \begin{bmatrix} \lambda \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{11}$$

If we divide F by λ on the left side and λ by λ on the right, we will have a set of simultaneous equations for F/λ. Then we can pick $f_1(0) = 1$ and this determines λ and all the f coefficients.

Illustration #3 Levin's filter

Again we minimize $\sum y(t)^2$ but with the constraints

$$\begin{aligned} 1 &= f_1(0) + f_2(0) \\ 0 &= f_1(1) + f_2(1) \\ 0 &= f_1(2) + f_2(2) \end{aligned} \tag{12}$$

Levin[1] arrived at these constraints by wanting the filter to pass that signal which is common to $x_1(t)$ and $x_2(t)$ without distortion. To show there is no distortion of the common part $s(t)$ of $x_1(t)$ and $x_2(t)$ write

$$\begin{aligned} x_1(t) &= s(t) + z_1(t) \\ x_2(t) &= s(t) + z_2(t) \end{aligned} \tag{13}$$

Then from (2)

$$y(t) = s(t) + (\text{terms in } z_1 \text{ and } z_2 \text{ but not } s)$$

Using Lagrange's method, we differentiate

$$E = \{ \sum_t y(t)^2 \} - \lambda_1[f_1(0) + f_2(0) - 1] - \lambda_2[f_1(1) + f_2(1)] - \lambda_3[f_1(2) + f_2(2)] \tag{14}$$

getting for $\dfrac{\partial E}{\partial f_k(\tau)}$

$$\begin{bmatrix} & & \\ & R & \\ & & \end{bmatrix} \begin{bmatrix} f_1(0) \\ f_2(0) \\ f_1(1) \\ f_2(1) \\ f_1(2) \\ f_2(2) \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1 \\ \lambda_2 \\ \lambda_2 \\ \lambda_3 \\ \lambda_3 \end{bmatrix} \tag{15}$$

which we may write as

$$\begin{bmatrix} \\ \\ \\ R \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ \\ \\ F \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \\ \\ \end{bmatrix} \tag{16}$$

which we abbreviate as

$$R\,F = G\Lambda \tag{17}$$

We can write the constraint equation down directly or go through the formalism of taking $\frac{\partial E}{\partial \lambda} = 0$ for each λ.

$$\begin{bmatrix} \\ \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ \\ \\ \end{bmatrix} \begin{bmatrix} f_1(0) \\ f_2(0) \\ f_1(1) \\ f_2(1) \\ f_1(2) \\ f_2(2) \end{bmatrix} = \begin{bmatrix} \\ \\ 1 \\ 0 \\ 0 \\ \\ \\ \end{bmatrix} \tag{18}$$

which we may abbreviate as

$$G^T F = D \tag{19}$$

Now we want to solve the set (17) and (19)

$$G^T F = G^T R^{-1} G \Lambda \qquad \text{multiply (17) by } G^T R^{-1}$$

$$D = G^T R^{-1} G \Lambda \qquad \text{insert (19)}$$

$$\Lambda = (G^T R^{-1} G)^{-1} D \qquad \text{solve for } \Lambda \qquad (20)$$

$$RF = G(G^T R^{-1} G)^{-1} D \qquad \text{insert into (17)}$$

$$F = R^{-1} G(G^T R^{-1} G)^{-1} D \qquad \text{solve for F} \qquad (21)$$

which is the desired solution. From (21) it may be seen that a scale factor in R does not affect F. This justifies the earlier assertion that a scale factor in R is unimportant.

The actual power minimum obtained is

$$F^T RF = D^T (G^T R^{-1} G)^{-1} D \qquad (22)$$

The quantity $R^{-1} G$ is computed efficiently by the Levinson recursion[2].

Repositioning the one in the vector D will produce Levin filters with delayed signal outputs. Some positions will reduce the output noise power $F^T RF$ more than others.

Illustration #4 A modification of Levin's constraint

In one application of Levin's filter it was observed that the filter coefficients weighted some channels very weakly or even with negative weights. This seemed unsatisfactory because in the experimental data the channel gains were not well controlled and the signals s(t) were not truly identical in the channels. A modification of Levin's constraints which retains the idea of passing the common part of $x_1(t)$ and $x_2(t)$ without distortion but leads to an even weighting of signals is

$$\begin{aligned} 1/2 &= f_1(0) \\ 1/2 &= f_2(0) \\ 0 &= f_1(1) + f_2(1) \\ 0 &= f_1(2) + f_2(2) \end{aligned} \tag{23}$$

These constraints lead to the same formal equations as before

$$RF = G\Lambda \tag{17}$$

$$G^T F = D \tag{19}$$

where in this case

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{bmatrix} . \tag{24, 25}$$

Although signals are forced to be weighted equally some channels may still be weighted more heavily than others, due to other filter lags than the zeroth lag. Although this might be good it led to trouble in one application when experimental data channel timing was not well controlled.

Illustration #5 Prediction Error

The constraints

$$\begin{aligned} f_1(0) &= 1 \\ f_2(0) &= 0 \end{aligned} \tag{26}$$

give rise to the prediction error filter[3] at unit span for the signal $x_1(t)$. This filter is so named because (2) can be written

$$y(t) = x_1(t) - (\text{terms dependent on time before } t) \tag{27}$$

Since $\sum_t y(t)^2$ is minimum, the "terms dependent on time before t" try to cancel $x_1(t)$. Those "terms" are the prediction filter of $x_1(t)$ and $y(t)$ is the prediction error signal.

The prediction error filter for span = 2 is given by the constraints

$$\begin{aligned} f_1(0) &= 1 \\ f_2(0) &= 0 \\ f_1(1) &= 0 \\ f_2(1) &= 0 \end{aligned} \tag{28}$$

Span = 2 means the prediction is done with terms depending on time at and before t-2.

Illustration #6 The general linear constraint

A fully generated example of two linear constraints is:

$$d_1 = v_{10} f_1(0) + v_{11} f_1(1) + v_{12} f_1(2) + v_{20} f_2(0) + v_{21} f_2(1) + v_{22} f_2(2)$$

$$d_2 = w_{10} f_1(0) + w_{11} f_1(1) + w_{12} f_1(2) + w_{20} f_2(0) + w_{21} f_2(1) + w_{22} f_2(2)$$

This leads to

$$G = \begin{bmatrix} v_{10} & w_{10} \\ v_{20} & w_{20} \\ v_{11} & w_{11} \\ v_{21} & w_{21} \\ v_{12} & w_{12} \\ v_{22} & w_{22} \end{bmatrix} \qquad D = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Illustration #7 Quadratic constraints

Quadratic constraints have not become well known but they are tractable using similar methods.

This will be illustrated by means of the annihilation filter.* The constraint is that the power coming out of the $f_1(\tau)$ filter plus the power coming out of the $f_2(\tau)$ filter

---

* It utilizes coherencies among the channels to annihilate the output y(t).

should have a constant (= 1) expected value but there should still be minimum power in the sum (3) as described in Illustration #1.

Explicitly, this constraint is

$$1 = \sum_t [f_1(0)\, x_1(t) + f_1(1)\, x_1(t-1) + f_1(2)\, x_1(t-2)]^2 + [f_2(0)\, x_2(t) + f_2(1)\, x_2(t-1) + f_2(2)\, x_2(t-2)]^2$$

Taking derivatives as in example #1 leads to

$$\left\{ R - \lambda \left[\begin{array}{cc|cc|cc} r_{11}(0) & 0 & r_{11}(-1) & 0 & r_{11}(-2) & 0 \\ 0 & r_{22}(0) & 0 & r_{22}(-1) & 0 & r_{22}(-2) \\ \hline r_{11}(1) & 0 & r_{11}(0) & 0 & r_{11}(-1) & 0 \\ 0 & r_{22}(1) & 0 & r_{22}(0) & 0 & r_{22}(-2) \\ \hline r_{11}(2) & 0 & r_{11}(1) & 0 & r_{11}(0) & 0 \\ 0 & r_{22}(2) & 0 & r_{22}(1) & 0 & r_{22}(0) \end{array}\right] \right\} F = 0$$

This is a generalized eigenvector-eigenvalue problem whose solution may be obtained by standard methods. A recursion scheme to speed computations like the Levinson method has not yet been developed for problems with quadratic constraints.

JC/lmm

ACKNOWLEDGMENT

The author is indebted to Ed Kelly who introduced him to Levin's method and Roy Greenfield who suggested the modification to the Levin filter.

REFERENCES

1. M. J. Levin, Maximum-Likelihood Array Processing, Section IV Array Processing, Seismic Discrimination Semiannual Technical Summary Report, M. I. T., Lincoln Laboratory, 31 December 1964, DDC455743.

2. R. A. Wiggins and E. A. Robinson, Recursive Solution to the Multichannel Filtering Problem, Journal of Geophysical Research, 70, 1965 (1885-92).

3. J. Claerbout, Detection of P-waves from Weak Sources at Great Distances, Geophysics, April 1964.

DOCUMENT CONTROL DATA – R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Lincoln Laboratory, M.I.T.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP None
3. REPORT TITLE A Summary, by Illustrations, of Least Squares Filters with Constraints	
4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* Technical Note	
5. AUTHOR(S) *(Last name, first name, initial)* Claerbout, Jon F.	
6. REPORT DATE 31 January 1966	7a. TOTAL NO. OF PAGES 18 7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. AF 19(628)-5167 b. PROJECT NO. ARPA Order 512 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) TN 1966-7 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* ESD-TDR-66-22
10. AVAILABILITY/LIMITATION NOTICES Distribution of this document is unlimited.	
11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Advanced Research Projects Agency, Department of Defense

13. ABSTRACT

Several methods of combining a number of time series into a single series are discussed. They are all individual filtering followed by summation and are somewhat like Wiener filtering in that a least squares criterion is used to define the filter coefficients. They differ from Wiener filtering in that signal information is given in the form of various linear constraints on the filter coefficients rather than being given as a signal correlation function. The formulas are worked out explicitly for the case of two time series and three filter points and presented in such a way as to make generalization clear.

14. KEY WORDS

time series
linear constraints
filters