

Correspondence

A Maximum Likelihood Estimator of Channel Impulse Response

J. A. BUCKLEW

Abstract—It is assumed that a receiver's input is a digital signal corrupted by white Gaussian noise. The information symbols in the digital data stream are in general allowed to be correlated. The small signal maximum likelihood estimate of the channel impulse response is derived. We show that the estimator problem leads to a generalized eigenvalue problem. An example is presented.

I. INTRODUCTION

Consider the following signal:

$$x(t) = \sum_{k=-\infty}^{\infty} u_k h(t - kT) \quad -\infty < t < \infty. \quad (1)$$

This functional form can arise in several different situations. For example, a digital signal would be modeled by the u_k 's being information-bearing random variables and the $h(\cdot)$ would represent a channel impulse response. A band-limited random process passed through a linear time-invariant system could also be modeled by the above equation where the u_k 's would represent time samples of the random process and $h(\cdot)$ would be a sinc (\cdot) type function convolved with the linear system impulse response.

In many cases of practical interest the function $h(\cdot)$ is unknown and needs to be measured. This could occur, for example, in a communication system where the channel impulse response may be time varying, perhaps due to multipath or some other randomness in the channel structure itself. If the channel impulse response in a digital communications system extends over a time interval longer than a bit period, then one also has to contend with intersymbol interference (ISI). It is known [1] that the maximum likelihood receiver for ISI channels can be implemented using a Viterbi algorithm. Unfortunately, this optimum receiver structure requires that one has complete knowledge of the channel impulse response.

For the case when (1) represents a "filtered" band-limited random process, we may desire to "unfilter" or deconvolve the signal. We are then faced, in effect, with a system identification problem where, preparatory to deconvolution, we must first identify the system impulse response of the distorting filter.

The idea of estimating a channel impulse response from a

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received March 17, 1982; revised July 28, 1982.

The author is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706.

noisy data bearing signal is not new. Perhaps the most well-known example of such estimators would be decision feedback equalizers [2]–[4]. Basically, the idea behind these systems is to make use of previously decoded information (and sometimes by incorporating delay the tail of the current pulse) to decode the incoming bit, then correct its sign and pass the resulting "corrected" signal into a circulating memory or averager where an estimate of the impulse response function will appear. Decision feedback equalizers are just one example of a class of communication receivers known as adaptive equalizers [5], [6]. These systems usually bypass the direct problem of channel impulse response estimation and instead try (usually in a minimum mean squared error sense) to estimate the ISI in a bit period and then subtract it out. Most of these systems use some sort of transversal filter structure and one could in principle obtain an estimate of the impulse response from the delay line multiplier coefficients. A good overview and more references of these types of systems are included in [9].

The methodology we will employ is to derive the small signal maximum likelihood estimate of $h(\cdot)$. Our philosophy in doing this follows that of locally optimum decision theory, where it is felt that if a receiver performs optimally for very small signals, then it should also perform well for larger signals. This methodology seems to lead to simple intuitive results. Section II of this paper is devoted to deriving the estimator; in Section III we give some simulation results and Section IV presents a discussion of the results and some conclusions to be drawn from them.

II. DEVELOPMENT

We will make the following assumptions about our estimation problem.

1) The receiver has access only to a finite time record of a noisy version of $x(t)$, i.e., the receiver knows only $y(t)$ where

$$y(t) = \sum_{k=-N}^{N-1} u_k h(t - kT) + n(t) \quad -NT \leq t \leq NT \quad (2)$$

and $n(t)$ is zero mean white Gaussian noise with spectral density $N_0/2$.

Remark: Note that time or bit synchronization is not assumed, i.e., if we have $y(t - \epsilon)$, $-NT \leq t \leq NT$, known at the receiver, then we will obtain an estimate for $h(t - \epsilon)$. Therefore, our estimate can also be used to synchronize the receiver.

2) N will be assumed to be large.

3) The $\{u_k\}_{k=-\infty}^{\infty}$ will be assumed to be samples from a stationary ergodic sequence with $E\{u_k^2\} \ll N_0/2$ and $E\{u_k\} = 0$. We also assume $E\{u_k u_{k'}\} = R(k - k') \approx 0$ for $|k - k'| > N' \ll N$.

4) $h(t)$ will be assumed to be nonzero only over some interval $(-mT, mT)$ where $m \ll N$.

It is well known [7] that the likelihood function for this problem is

$$\int_{R^{2N}} \exp \left[\left(2 \int_{-NT}^{NT} y(t)x(t) dt - \int_{-NT}^{NT} x(t)^2 dt \right) / N_0 \right] p(\mathbf{u}) d\mathbf{u} \quad (3)$$

where $\mathbf{u} = (u_{-N}, u_{-N+1}, \dots, u_0, u_1, \dots, u_{N-1})$ and $p(\mathbf{u})$ is its probability density. Our problem then becomes to maximize $I(h)$ over all $h(t)$, where

$$I(h) = \int_{R^{2N}} \exp \left[\left(2 \int_{-NT}^{NT} y(t) \sum_k u_k h(t-kT) dt - \int_{-NT}^{NT} \left(\sum_k h(t-kT)u_k \right)^2 dt \right) / N_0 \right] p(\mathbf{u}) d\mathbf{u}. \quad (4)$$

We will employ a standard calculus of variations approach now. Let $h(t) = h_0(t) + \epsilon r(t)$ where $h_0(t)$ is the optimum estimate and $r(t)$ is some square integrable time function such that $r(t)$ is orthogonal (in an L^2 sense) to $h_0(t)$. Then we must have

$$\left. \frac{\partial I(h)}{\partial \epsilon} \right|_{\epsilon=0} = 0. \quad (5)$$

If N is very large and $h(t)$ is nonzero only over a few bits, then we have

$$\begin{aligned} & \sum_k \int_{-NT}^{NT} y(t)h(t-kT) dt \\ & \approx \sum_k \int_{-NT}^{NT} h(t)y(t-kT) dt. \end{aligned} \quad (6)$$

Performing the operations implied by (5) and making use of (6) it is simple (but lengthy) to show that the optimal estimate $h_0(t)$ is given by

$$h_0(t) = \sum_{k=-N}^{N-1} a_k y(t-kT) \quad (7)$$

where the a_k are constants that need to be determined. If N is very large, we have

$$\frac{1}{N} \int_{-NT}^{NT} x(t)^2 dt \approx \text{constant} = E \left\{ \frac{1}{N} \int_{-NT}^{NT} x(t)^2 dt \right\}$$

due to the assumed ergodicity of the $\{u_k\}$ sequence and where $E\{\cdot\}$ denotes the expectation operator. Hence, (4) can be recast into a Lagrange multiplier problem with the problem of maximizing $I(h)$ the same as maximizing over the $\mathbf{a} = (a_{-N}, \dots,$

$a_{N-1})$ vector in the following expression,

$$E_{\mathbf{u}} \left\{ \exp \left[\left(2 \int_{-NT}^{NT} y(t)x'(t) dt / N_0 \right) \right] \right\} + \lambda E_{\mathbf{u}} \left\{ \int_{-NT}^{NT} \left(\sum_k u_k h_0(t-kT) \right)^2 dt \right\} \quad (8)$$

where $x'(t) \triangleq \sum_k u_k h_0(t-kT)$ and $h_0(t)$ is given by (7).

Consider the term multiplying the Lagrange multiplier λ . This equals

$$\begin{aligned} & E_{\mathbf{u}} \left\{ \sum_k \sum_{k'} u_k u_{k'} \int_{-NT}^{NT} h_0(t-kT)h_0(t-k'T) dt \right\} \\ & \cong \sum_{k=-N}^{N-1} \sum_{k'=-N}^{N-1} R(k-k') \int_{-\infty}^{\infty} h_0(t-kT)h_0(t-k'T) dt \end{aligned}$$

where

$$\begin{aligned} E\{u_k u_{k'}\} & \triangleq R(k-k') \\ & \cong \sum_{k=-2N}^{2N} R(k)(2N-|k|) \\ & \quad \cdot \int_{-\infty}^{\infty} h_0(t)h_0(t-kT) dt \\ & = \sum_{k=-2N}^{2N} R(k)(2N-|k|) \sum_{k'=-N}^N \sum_{k''=-N}^N a_{k'} a_{k''} \\ & \quad \cdot \int_{-mT}^{mT} y(t-k'T)y(t-(k''+k)T) dt. \end{aligned} \quad (9)$$

Let us define $\int_{-mT}^{mT} y(t-k'T)y(t-(k'')T) dt \triangleq y_{k',k''}$, where we append zeros if necessary to the $y(t)$ record. Note that $y_{k',k''} = y_{k'',k'}$. Let us now consider taking the partial with respect to a_j of the above expression [which we need to do in the maximization of (8)]. We first note that

$$\begin{aligned} & \frac{\partial}{\partial a_j} \left(\sum_{k'} \sum_{k''} a_{k'} a_{k''} y_{k',k''+k} \right) \\ & = \sum_{k''} a_{k''} y_{j,k''+k} + \sum_{k'} a_{k'} y_{k',j+k}. \end{aligned}$$

Let $Y = (y_{jk})$ be a $2N \times 2N$ matrix, and let $Y^{-1} = (M_{n,l})$ be its inverse. The partial of (9) with respect to a_j leads to the following set of $2N$ equations:

$$\begin{aligned} A(j) & = \sum_{k=-2N}^{2N} R(k)(2N-|k|) \\ & \quad \cdot \left[\sum_{k''} a_{k''} y_{j,k''+k} + \sum_{k'} a_{k'} y_{k',j+k} \right] \end{aligned}$$

for $j = -N, \dots, N-1$. Let $A^T = [A(-N), A(-N+1), \dots, A(N)]$. Now right multiply A^T by Y^{-1} . The l th component of the resulting vector is

$$\sum_{k=-2N}^{2N} R(k)(2N-|k|) \left[\sum_{k''} a_{k''} \sum_j y_{j,k''+k} M_{j,l} + \sum_{k'} a_{k'} \sum_j y_{k',j+k} M_{j,l} \right].$$

But noting that $y_{j,k''+k} = y_{k''+k,j}$ and

$$\sum_j y_{k''+k,j} M_{j,l} = 1 \quad \text{if } k''+k=l \text{ or } k=l-k'' \\ = 0 \quad \text{otherwise}$$

we have the above equal to

$$\sum_{k=-2N}^{2N} R(k)(2N-|k|) \left[\sum_{k''} a_{k''} \delta_{k''+k,l} + \sum_{k'} a_{k'} \sum_j y_{j,j+k} M_{j,l} \right] \\ = \sum_{k''} R(l-k'')(2N-|l-k''|) a_{k''} \\ + \sum_k R(k)(2N-|k|) \cdot \left(\sum_{k'} a_{k'} \sum_j y_{k',j+k} M_{j,l} \right).$$

If $R(k) \approx 0$ for $k > N' \ll N$ the second term can be approximated as

$$\sum_{k=-\infty}^{\infty} R(k)(2N-|k|) \sum_{k'} a_{k'} \sum_j y_{k',j+k} M_{j,l} \\ = \sum_{k=-\infty}^{\infty} R(k-j)(2N-|k-j|) \sum_{k'} a_{k'} \sum_j y_{k',k} M_{j,l} \\ = \sum_j \sum_{k=-\infty}^{\infty} R(k-j)(2N-|k-j|) (a^T Y)_k M_{j,l} \\ = \sum_{j=-N}^N \sum_{k=-N}^N R(k-j)(2N-|k-j|) M_{j,l} (a^T Y)_k$$

where $(a^T Y)_k$ indicates the k th component of the vector $a^T Y$. Let $R = (R(i-j))$ and $R' = (R(i-j)(i-j))$ which are both $2N \times 2N$ matrices. The above then is the l th component of the second term in (10). It is simple to see that the whole second term vector can be written as $a^T Y R' Y^{-1}$. Similarly the first term vector is written as $a^T R'$. Therefore, we have

$$\left(\frac{\partial}{\partial a_j} E_u \left\{ \int_{-NT}^{NT} x'(t)^2 dt \right\} \right)^T Y^{-1} \\ = a^T (R' + Y R' Y^{-1}).$$

We now go back to (8) and consider the first term.

$$E_u \left\{ \exp \left[2 \int_{-NT}^{NT} y(t) x'(t) dt / N_0 \right] \right\} \\ = E_u \left\{ \exp \left[2 \int_{-NT}^{NT} y(t) \sum_k u_k h_0(t-kT) dt / N_0 \right] \right\} \\ \cong E_u \left\{ \exp \left[2 \int_{-NT}^{NT} \sum_k u_k y(t-kT) h_0(t) dt / N_0 \right] \right\} \\ = E_u \left\{ \exp \left[\frac{2}{N_0} \sum_k \sum_{k'} u_k a_{k'} \cdot \int_{-mT}^{mT} y(t-kT) y(t-k'T) dt \right] \right\}.$$

Let us take the partial of the above expression with respect to a_j , i.e., we have

$$E_u \left\{ \frac{2}{N_0} \sum_k u_k y_{k,j} \exp \left[\frac{2}{N_0} \sum_{k'} \sum_{k''} u_{k'} a_{k''} y_{k',k''} \right] \right\}.$$

Now since the $E\{u_k^2\} \ll N_0$, we can approximate $\exp[x] \cong 1+x$. Making use of the fact that $E\{u_k\} = 0$, we have the above equal to

$$E_u \left\{ \frac{4}{N_0^2} \sum_{k''} \sum_k \sum_{k'} u_k u_{k'} a_{k''} y_{k',k''} y_{k,j} \right\} \\ = \frac{4}{N_0^2} \sum_{k''} \sum_{k'} \sum_k R(k-k') a_{k''} y_{k',k''} y_{k,j}$$

is the j th component of the vector

$$\left(\frac{\partial}{\partial a_j} E_u \left\{ \exp \left[2 \int_{-NT}^{NT} x'(t) y(t) dt / N_0 \right] \right\} \right)^T$$

Hence, the whole vector can be written as $(4/N_0^2) a^T Y R Y$. Multiply by Y^{-1} and we obtain $(4/N_0^2) a^T Y R$. Hence, the maximization of (8) is equivalent to solving the matrix equation

$$a^T Y R Y = \lambda a^T (R' Y + Y R') \quad (11)$$

for a . Since $Y R Y$ and $R' Y + Y R'$ are symmetric matrices, this is called a generalized eigenvalue problem [8]. It is known then that we want to find the eigenvector a for (11) corresponding to the largest eigenvalue. We note that if $R = I$,

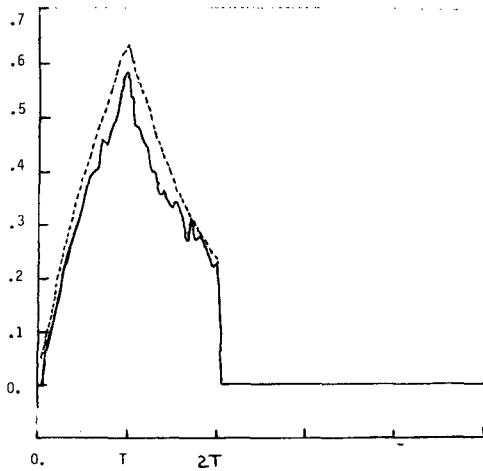


Fig. 1. A plot of the channel impulse response (dotted line) and the estimated channel impulse response (solid line) for $E_b/N_0 = 4.4$.

which corresponds to uncorrelated sample points, then $R' = I$ and (11) becomes $a^T Y = \lambda a^T$, the simple eigenvalue problem.

III. SIMULATION RESULTS

We prepared a computer simulation example to demonstrate the technique embodied in (11). We let $2N = 60$, $T = 1$, and the $\{u_k\}$ sequence was assumed to be independent identically distributed Bernoulli $(1/2) \pm 1$ random variables. The noise was Gaussian and the ratio of energy per bit (E_b) to N_0 was 4.4. We took

$$h(t) = 1 - e^{-t} \quad 0 \leq t \leq 1$$

$$= (1 - e^{-1})e^{-(t-1)} \quad 1 \leq t \leq 2.$$

In Fig. 1 we see the results of this experiment where the "noisy" curve is our estimate plotted on the same graph as $h(t)$. The eigenvector problem embodied in (12) was solved by a standard vector iteration technique [8].

IV. DISCUSSION AND CONCLUSIONS

Consider the form of (7). Ideally, a_k should have the same sign as u_k , in effect making our estimate the average of many noisy copies of $u_k a_k h(t)$. The magnitude of a_k should be inversely proportional to the amount of noise that comes in with the k th information bit. The ISI part of $h(t)$ is recovered since if t is not in the range $(0, T)$, then the contribution of all the "copies" of $h(t)$ outside the range $(0, T)$ are adding together with the wrong signs and, for N large enough, tend to cancel out. According to Lucky [9], this tail cancellation is a very old idea and almost all decision feedback equalizers employ it. It is interesting to see this same structure appear in a maximum likelihood format.

We also note that this estimator requires no prior information (other than the correlation structure of the data) to "start it up." In fact, this structure can even be used to achieve bit synchronization. However, there is a sign ambiguity in the

estimate of $h(t)$, i.e., both a and $-a$ will generate the low signal maximum likelihood estimate of $h(t)$. We could incorporate prior information about the $\{u_i\}$ sequence by choosing $+a$ or $-a$ to maximally correlate in some sense with u . The best solution, of course, would be to go back to the original likelihood equation (3) and incorporate prior information into the probability density of u .

The estimator's main drawback is that it requires finding the eigenvector of a $2N \times 2N$ matrix where $2N$ is the number of bits to be used in making the estimate. In the case where the data are correlated the computation is even more complex. If N is kept small, however, this is not too major a problem as eigenvector calculations can be done much faster than matrix inversions. We can always keep N small by dividing our data up into small subblocks and averaging the resulting estimates from each subblock. This seems to work well as long as the subblocks are still large in comparison to the ISI. We applied this technique to the example in the previous section; we averaged our estimate over 30 blocks (1800 total bits) giving (visually at least) a perfect reproduction of $h(t)$.

In conclusion, we have presented a technique for estimating the channel impulse response from noisy data. Its advantages are that no prior information about the bit sequence is needed, the data sequence need not be uncorrelated, it performs optimally in small signal environments, and synchronization is not needed. The possible disadvantages include a sign ambiguity and some increased system complexity and computation.

ACKNOWLEDGMENT

I would like to thank W. Birkemeier of the University of Wisconsin, Madison for first posing this problem. I would especially like to thank J. Weitzen of the University of Wisconsin, Madison for his assistance in preparing the example.

REFERENCES

- [1] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363-378, May 1972.
- [2] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 56-65, Jan. 1971.
- [3] D. A. George, R. R. Brown, and J. R. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 281-293, June 1971.
- [4] J. Salz, "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, pp. 1341-1373, Oct. 1973.
- [5] J. G. Proakis and J. H. Miller, "An adaptive technique for digital signaling through channels with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 484-497, July 1969.
- [6] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, vol. 48, pp. 55-70, Jan. 1969.
- [7] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-15, May 1969.
- [8] H. R. Schwarz, H. Rutishauser, and E. Stiefel, *Numerical Analysis of Symmetric Matrices*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [9] R. W. Lucky, "A survey of the communication theory literature: 1968-1973," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 725-739, Nov. 1973.