THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 62 | July–August 1983 | Number 6, Part 1

On the Start-Up Problem in Digital Echo Cancelers

By J. SALZ*

(Manuscript received October 5, 1982)

Digital echo cancellation techniques make it possible to realize efficient full-duplex data transmission over a single loop. The purpose of this paper is to elucidate the solution to the start-up problem in these devices and to present a new, fast, and simple tap-adjustment procedure. The theory indicates that a modified stochastic gradient tap-adjustment algorithm, using pseudorandom input data sequences for the initial training period, converges in N steps, where N is the total number of canceler taps, and that this is the fastest possible convergence time.

I. INTRODUCTION

Two-way voice communication over a single loop is made possible by the use of a hybrid bridge. However, the suppression of echoes by fixed hybrids is insufficient to support full-duplex data transmission, and therefore makes adaptive data echo cancelers necessary.

* Bell Laboratories.

In two-way data communications, transmitted data "echoes" back to the near-end receiver after being reflected and dispersed through an unknown return path. So, if one assumes that the echo path is linear, the estimation of its overall impulse response is sufficient to allow the synthesis of one's own echo signal. This synthesized version is subtracted from the received signal, which then makes it possible for the receiver to extract the data intended for it.

The impulse response of the echo channel (local transmitter output to local receiver input) can be measured in several ways. An obvious way to do this is to transmit a single impulse and measure the echo. However, among other defects of this procedure, the average power would be very low and the resulting signal-to-noise ratio (s/n) would be inadequate. If a pseudorandom sequence (+1 or −1) is transmitted instead, the average power would be much greater and would be essentially constant on the line, more nearly representing a true data signal. The latter is the preferable approach.

Digital data echo cancelers operate in two modes. In the acquisition mode, or start-up, the impulse response of the echo path is measured. This is best accomplished, as we shall see, with the use of fixed data sequences. Since, during this period, no information is conveyed to the far-end, the time allotted for this purpose should be as short as possible. Although it is conceptually possible to start an echo canceler either blind or with random data, the convergence of the taps, or the reliable measurements of the impulse response, are known to require a long time. In the subsequent mode, or during actual data transmission, a tracking algorithm is initiated whose function is to update the measurements when slight changes occur in the impulse response. We focus on the more critical start-up algorithm.

From an operational point of view, it is desirable to implement these algorithms in a recursive fashion, or in a closed-loop manner. This means that the canceler tap coefficients, which represent the sampled impulse response, are updated in response to a measured error between the actual impulse response and the one estimated at any particular instant. Conventional gradient adjustment algorithms, even with fixed data sequences, are known to converge very slowly.

This research was motivated by the need for a theory capable of explaining the behavior of tap-adjustment algorithms. During the course of this investigation a modified stochastic gradient algorithm that is simple to implement and converges in the theoretically smallest number of steps was discovered.

II. PROBLEM FORMULATION

Figure 1 shows a full-duplex data modem employing a digital echo canceler. We are concerned with the sampled signal values

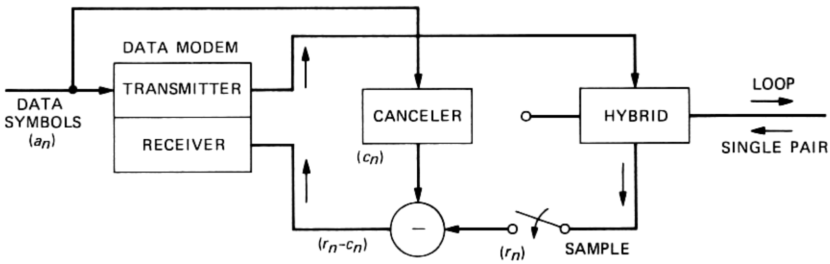

Fig. 1—Digital data canceler.

$$r_n = \sum_{k=-\infty}^{\infty} h_k a_{n-k} + \nu_n, \tag{1}$$

where the a_n's are the transmitted data symbols that echo back to the receiver, the h_n's are the overall impulse response values of the echo path, and the ν_n's are the desired received signal values plus noise. The object of the canceler is to synthesize signal values, c_n, which are estimates of $\sum_k h_k a_{n-k}$ and subtract them from the r_n's. The receiver then proceeds to process the difference signal values, $r_n - c_n$, to extract the data intended for it.

The fundamental problem is to devise procedures for estimating the h_n's from observations of the r_n's, while treating the ν_n's as undesirable noise. Practically, it must be assumed that only a finite number, N, of the h_n's can be estimated, and so we express (1) compactly as

$$r_n = H^t A_n + \nu_n, \tag{2}$$

where H and A_n are finite dimensional column vectors,*

$$H = \begin{pmatrix} h_N \\ \vdots \\ h_1 \end{pmatrix}, \quad \text{and} \quad A_n = \begin{pmatrix} a_{n-N} \\ \vdots \\ a_{n-1} \end{pmatrix},$$

and where $(\ )^t$ indicates the transpose of a matrix.

In the absence of precise statistical knowledge of the ν_n's and H, a natural procedure for choosing an estimator of H is to minimize the sum of squared errors from time 1 to time l

$$\epsilon_l = \sum_{n=1}^{l} (r_n - \hat{H}^t A_n)^2.$$

This is a standard problem and the solution is immediate. It involves the solution of a set of linear equations

$$Z_l \hat{H}_l = U_l \tag{3}$$

* We deal with a baseband model for notational convenience. By using complex numbers throughout, the treatment generalizes to passband models.

for the best estimator at time l, $\hat{H}_l$.

In (3)

$$U_l = \sum_{n=1}^{l} A_n r_n$$

and

$$Z_l = \sum_{n=1}^{l} A_n A_n^t.$$

In applications it is usually desirable to solve these equations recursively with a minimum computational effort and to assure rapid convergence to the actual H. Our attention in the next sections is directed toward these aims. However, before dealing with our main subject we wish first to examine some asymptotic behaviors of the standard solution.

2.1 Random input data

In some applications, echo cancelers must start blind, i.e., from random data. This is the case in echo cancelers used to suppress speech. However, in full-duplex data communications, a preamble word, or words, can be sent first to assure rapid convergence. To emphasize these differences we examine the behavior of the estimator with random data first, where one has no choice in the selection of the starting sequences. So, consider random data such that the a_n's assume ± 1 independently with equal probability and examine the limit as $l \rightarrow \infty$. It is found that

$$U_l \rightarrow lE\{A_n r_n\}$$
$$= lE\{A_n A_n^t\}H,$$
$$Z_l \rightarrow lE\{A_n A_n^t\} = lI,$$

and, consequently,

$$Z_l^{-1} U_l \rightarrow H. \tag{4}$$

In the above we made use of the fact that the ν_n's and A_n's are naturally independent. We assumed that these sequences are ergodic and so replaced time averages with mathematical expectations, $E\{\cdot\}$. This then demonstrates that if one is willing to wait forever, it is conceptually possible to determine H exactly—not a terribly startling result.

While the asymptotic behavior is easy to deduce, the statistical behavior of the estimator for finite l is difficult to glean. One immediately encounters an unsolved mathematical problem that involves the conditions on the random sequences that would guarantee the existence of the inverse matrix Z_l^{-1}. Clearly, l has to be greater than N for the inverse to even have a chance to exist, and for those

Likewise, the vector

$$U_l = \sum_{n=1}^{l} A_n r_n = U_{l-1} + A_l r_l. \tag{13}$$

Now, applying the matrix inversion lemma[2] to (12), and assuming that the inverses exist, we find that

$$Z_l^{-1} = Z_{l-1}^{-1} - \frac{Z_{l-1}^{-1} A_l A_l^t Z_{l-1}^{-1}}{1 + A_l^t Z_{l-1}^{-1} A_l}. \tag{14}$$

This is the key equation and, in conjunction with (13), makes it possible to claim (11). We note that the algorithm expressed in (13) is computationally complicated since it requires the calculation of a matrix recursion, (14), and multiplication of matrices by vectors at each iteration. For large N this becomes infeasible and simpler procedures are therefore sought. However, before proposing a simpler algorithm we need to review some properties of pseudorandom sequences.

The pseudorandom data sequences that are the inputs to the canceler during the start-up period derive from the binary sequences

$$X = x_1 x_2 x_3 x_4 \cdots$$

($x_i = 0$, or 1). The nth digit is computed from certain of the earlier digits by means of the recurrence relation

$$x_n = x_{n-c} + x_{n-b} \bmod 2,$$

where c and b are integers, $0 < c < b$. The actual data sequences $\{a_n\}$ that are applied to the canceler are the x_n's with "0" replaced by "-1".

Returning now to the sequence X, we remark that in spite of the fact that x_n is completely determined by the digits that precede it, the sequence X resembles in some respects a completely random sequence. The calculation of the sequence X is carried out in a shift register working in a closed loop and a mod 2 adder. It turned out that for special choices of c and b the sequence X is periodic with period $2^b - 1$.[2]

In our application, we will make use of the following known properties of the sequences

$$1)\ A_n^t A_m = \sum_{i=1}^{N} a_{n-i} a_{m-i} = \begin{cases} N, & n = m \\ -1, & n \neq m \end{cases}$$

$$2)\ Q^t A_n = 1, \quad \text{for} \quad n = 1, \cdots, N$$

$$Q = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}_1^N,$$

the all "1" vector

$$3)\ A_{n+N} = A_n, \text{ periodicity.}$$

Data sequences possessing properties 1 through 3 are also referred to as pseudorandom sequences. As a consequence of these properties, it is easy to verify the following:
Property (a):

$$Z_N = \sum_{n=1}^{N} A_n A_n^t$$
$$= (N+1)I - QQ^t,\ N \text{ x } N \text{ matrix,}$$

where I is the identity matrix. This decomposition is possible since the ijth element of the matrix Z_N, $(Z_N)_{ij} = N$ for $i = j$ and -1 on the off-diagonals. This can readily be seen from property 1.
Property (b): The inverse matrix

$$Z_N^{-1} = \frac{1}{N+1}(I + QQ^t).$$

This can be derived from the matrix inversion lemma or verified by actually computing $Z_N Z_N^{-1} = I$.
Property (c): The set of vectors

$$B_n = Z_N^{-1} A_n,\ n = 1 \cdots N$$

are orthonormal to the vectors A_n, $n = 1, \cdots N$. This is a crucial property to what follows, so we prove that

$$A_m^t Z_N^{-1} A_n = \frac{A_m^t}{N+1}(I + QQ^t)A_n$$
$$= \frac{A_m^t A_n + (A_m^t Q)(Q^t A_n)}{N+1} \begin{matrix} = 0,\ n \neq m \\ = 1,\ n = m\,. \end{matrix}$$

These properties suggest an approach to a simple and rapidly converging tap-adjustment algorithm.

The key to the simplification of the algorithm, (11), is the recognition that Z_l^{-1} for $l < N$ does not exist, and so it is not possible to start the algorithm at $l = 1$. The basic idea is to replace Z_l by Z_N and thus obtain the simpler algorithm

$$\hat{H}_{n+1} = \hat{H}_n - Z_N^{-1} A_n e_n, \tag{15}$$

where the error at time n is again

$$e_n = A_n^t \hat{H}_n - r_n$$
$$= A_n^t(\hat{H}_n - H) - \nu_n.$$

Note that this is a measurable quantity at each iteration and $A_n e_n$ is just the gradient of the instantaneous squared error.

The algorithm expressed in (15) is remarkably simple since

$$
\begin{aligned}
Z_N^{-1} A_n &= \frac{1}{N+1} (I + QQ^t) A_n \\
&= \frac{1}{N+1} (A_n + Q),
\end{aligned}
$$

which follows from the definition of Z_N and property (b). Inserting this into (15) we obtain

$$
\hat{H}_{n+1} = \hat{H}_n - \frac{1}{N+1} (A_n + Q) e_n. \tag{16}
$$

This form is immediately recognized as a slightly modified stochastic gradient algorithm with step size equal to $1/N + 1$ and the gradient vector, A_n, replaced by $A_n + Q$. It is nothing more than the original vector, A_n, in which a_n's equal to -1 are replaced by zero and $a_n = 1$ is replaced by $a_n = 2$. We wish to acknowledge that during the course of the development of this theory C. W. Farrow anticipated the form of this algorithm.

It now remains to demonstrate that (16) indeed converges "fast", by which we mean that it converges in N steps. Toward this end, define the error vector

$$
\epsilon_n = \hat{H}_n - H,
$$

and rewrite (16) in the form,

$$
\begin{aligned}
\epsilon_{n+1} &= \epsilon_n - Z_N^{-1} A_n (A_n^t \epsilon_n - \nu_n) \\
&= (I - B_n A_n^t) \epsilon_n + B_n \nu_n.
\end{aligned} \tag{17}
$$

Iterating (17) yields explicitly

$$
\begin{aligned}
\epsilon_{n+1} = {} & \prod_{k=1}^{n} (I - B_k A_k^t) \epsilon_1 \\
& + \sum_{k=1}^{n} \prod_{j=k}^{n-1} (I - B_{j+1} A_{j+1}^t) B_k \nu_k.
\end{aligned} \tag{18}
$$

This is the general solution but because of property (c), which states that $A_n^t B_{n-1} = 0$, we get a much simpler solution, which is the chief reason for the rapid convergence, namely,

$$\epsilon_{n+1} = \left(I - \sum_{k=1}^{n} B_k A_k^t\right) \epsilon_1 + \sum_{k=1}^{n} B_k \nu_k = \left[I - Z_N^{-1} \left(\sum_{k=1}^{n} A_k A_k^t\right)\right] \epsilon_1 + Z_N^{-1} \left(\sum_{k=1}^{n} A_k \nu_k\right). \tag{19}$$

This simple form results because the product of the matrices in (18) reduces to

$$\prod_{k=1}^{n} (I - B_k A_k^t)\epsilon_1 = (I - B_1 A_1^t)(I - B_2 A_2^t) \cdots (I - B_n A_n^t)\epsilon_1 = [(I - B_1 A_1^t)(I - B_2 A_2^t) \cdots -B_n A_n^t]\epsilon_1 = \left[I - \sum_{k=1}^{n} B_k A_k^t\right] \epsilon_1.$$

The evolution of the error vector ϵ_n is guided by two components, the transient

$$\tau_n = \left[I - Z_N^{-1} \left(\sum_{k=1}^{n} A_k A_k^t\right)\right] \epsilon_1, \tag{20}$$

and the steady-state component

$$S_n = Z_N^{-1} \left(\sum_{k=1}^{n} A_k \nu_k\right). \tag{21}$$

A most crucial property of the transient solution is that at $n = N$, τ_n vanishes, and this is the reason for claiming "fast" convergence. Clearly, the transient solution cannot vanish before this time since the inverse doesn't even exist, and therefore we claim the algorithm convergence in the least possible number of steps. This most important property of the algorithm can be seen from (20), since

$$\tau_N = (I - Z_N^{-1} Z_N)\epsilon_1 = 0.$$

Consequently, the error vector at time $N + 1$ consist only of measurement noise, or the steady-state component

$$\epsilon_{N+1} = Z_N^{-1}\left(\sum_{k=1}^{N} A_k\nu_k\right). \tag{22}$$

We have thus demonstrated that the algorithm converges to the true solution in N steps since the transient vanishes at the end of the Nth iteration and, from that time on, the taps fluctuate around the true value due to measurement noise, ν_n, alone. The variance of the tap fluctuations can be calculated from the variance matrix

$$\begin{aligned}\rho_N &= E\{\epsilon_{N+1}\epsilon_{N+1}^t\}\\ &= Z_N^{-1}\left(\sum_{k,k'=1}^{N} A_kA_{k'}^tE\{\nu_k\nu_{k'}\}\right)Z_N^{-1}\\ &= \sigma^2 Z_N^{-1} = \frac{\sigma^2}{N+1}(I+QQ^t),\end{aligned} \tag{23}$$

where again we assumed that the ν_k's are identically and independently distributed. The error variance, σ_H^2, is therefore,

$$\begin{aligned}\sigma_H^2 &= \sigma^2 \text{ Trace } Z_N^{-1}\\ &= \frac{\sigma^2}{N+1}\text{Trace}[I+QQ^t]\\ &= \sigma^2\frac{2N}{N+1}.\end{aligned} \tag{24}$$

This is precisely the value we obtained from solving the set of linear equations, (3), with pseudorandom input sequences. Thus, iterating N times provides the solution in a simple fashion.

It may turn out in applications that the value of variance obtained in N iterations is not sufficiently small. As seems reasonable, the noise variance can be reduced to any desired value by repeating the pseudorandom sequence of length N. To see this, consider a slightly modified tap-adjustment algorithm

$$H_{n+1} = H_n - \alpha Z_N^{-1}A_ne_n, \tag{25}$$

where now the scaler, α, is a fixed step size yet to be determined. Proceeding as before, the recursion for the tap error $\epsilon_n = H_n - H$ now becomes

$$\epsilon_{n+1} = (I - \alpha B_nA_n)\epsilon_n + \alpha B_n\nu_n, \tag{26}$$

with the concomitant solution

$$\epsilon_{n+1} = \left[I - \alpha Z_N^{-1} \left(\sum_{k=1}^{n} A_k A_k^t \right) \right] \epsilon_1 + \alpha Z_N^{-1} \left(\sum_{k=1}^{n} A_k \nu_k \right). \tag{27}$$

Again, examine the solution at $n = pN + 1$, where p is a positive integer, to obtain

$$\epsilon_{pN+1} = \left[I - \alpha Z_N^{-1} \left(\sum_{k=1}^{pN} A_k A_k^t \right) \right] \epsilon_1 + \alpha Z_N^{-1} \left(\sum_{k=1}^{pN} A_k \nu_k \right). \tag{28}$$

Since A_n is periodic with period N (property 3), we conclude that

$$\epsilon_{pN+1} = (1 - \alpha p)\epsilon_1 + \alpha Z_N^{-1} \left(\sum_{k=1}^{pN} A_k \nu_k \right), \tag{29}$$

and so we see that the transient component can be made to vanish when $\alpha = 1/p$. A straightforward calculation indicates that with this choice of α the variance is

$$\sigma_H^2 = \frac{2N}{p(N+1)} \sigma^2, \tag{30}$$

indicating a reduction by a factor of p—the number of times the pseudorandom sequence is repeated.

REFERENCES

1. M. S. Mueller, "Least-Squares Algorithms for Adaptive Equalizers," B.S.T.J., *60*, No. 8 (October 1981), pp. 1905–25.
2. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Information and System Science Series, Thomas Kailath, ed., Englewood Cliffs, NJ: Prentice-Hall, 1979.
3. J. J. Stiffler, *Theory of Synchronous Communications*, Englewood Cliffs, NJ: Prentice-Hall, 1971, pp. 178–84.

AUTHOR

Jack Salz, B.S.E.E., 1955, M.S.E., 1956, and Ph.D., 1961, University of Florida; Bell Laboratories, 1961—. Mr. Salz first worked on the electronic switching system. Since 1968 he has supervised a group engaged in theoretical studies in data communications and is currently a member of the Communications Methods Research Department. During the academic year 1967–68, he was on leave as Professor of Electrical Engineering at the University of Florida. He was a visiting lecturer at Stanford University in Spring 1981 and a visiting MacKay Lecturer at the University of California, at Berkeley, in Spring 1983.