THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 10, December 1981
Printed in U.S.A.

On the Rapid Initial Convergence of Least-Squares Equalizer Adjustment Algorithms

By M. S. MUELLER

(Manuscript received May 11, 1981)

Adjustment algorithms for transversal equalizers derived from least-squares cost functions are known to converge extremely fast. While various simulation results confirming this fact abound in the literature, a theory explaining the fast convergence has been lacking. This paper reports on steps toward such a theory. For some commonly used start-up data sequences it was found that algebraic properties of the sampled signal vectors play a critical role in the transient behavior of these algorithms; namely, successive signal vectors are linearly independent for a large class of transmission channels in the absence of noise. After N iterations (N being the number of taps), the resulting coefficient vector is found to be related to well-known equalizer coefficient vectors. If a single pulse is used as a training signal, the zero forcing equalizer is obtained; if a pseudo random noise sequence, with a period in symbols equal to the number of coefficients is used, the steady-state solution of the cyclic equalization is obtained. Thus, after only N iterations, the least-squares algorithms yield a coefficient vector which is only asymptotically obtainable by gradient techniques.

I. INTRODUCTION

Adaptive equalizers are important building blocks in modems for digital data transmission over linear dispersive channels. They adaptively mitigate the adverse effects of intersymbol interference. A critical parameter in the start-up performance of modems is the speed of convergence of the equalizer adjustment algorithm. The overall data throughput depends on it and, consequently, a high convergence speed is desirable.

Various different equalizer structures are known at this time. In the following, we concentrate on the frequently used transversal filter structure.[1] Many equalizer update algorithms are based on the steepest descent, or gradient technique, which minimizes the mean-squared error (mse) between the equalizer output and the transmitted data symbols.[2] In particular, the stochastic approximation of the gradient algorithm with an mse criterion is used frequently. The convergence speed of this algorithm was analyzed in Refs. 3 and 4. It was found to be dependent on the number of coefficients used and, to a lesser degree, on the eigenvalue spread of the channel autocorrelation matrix.

Several methods to improve the convergence speed of the gradient algorithm were published in Refs. 5 to 8. In Ref. 5, prior knowledge of the transmission channel is assumed and a transformation of the received signal is proposed which reduces the effect of a large eigenvalue spread on the convergence speed, whereas in Ref. 6 a transformation of the correction vectors, yielding the same performance, is proposed. Used in conjunction with a stochastic gradient algorithm, these methods reduce the convergence time to the minimal time which is obtainable with an ideal channel having an eigenvalue spread of one. In Refs. 7 and 8 cyclic equalization was proposed as a means to speed up the convergence time. The number of iterations required for convergence of the algorithm is about the same as for the stochastic gradient algorithm, but theoretically, the convergence time can be reduced to the time required to fill the register of the equalizer. Practically, however, the convergence speed is limited by the available computational power of the implemented algorithm.

In Ref. 9, Godard cast the equalizer adjustment problem as an estimation of a stationary state vector in Gaussian noise—a classical Kalman filtering problem. This resulted in a new, powerful, and rapidly converging equalizer adjustment algorithm. While this algorithm was familiar in the area of stochastic approximation theory,[10] it was never applied to equalizers prior to Godard's work. It was shown by computer simulations,[9] that this coefficient adjustment algorithm converges considerably faster than the stochastic gradient algorithm and virtually independently of the channel used. After about N iterations, where N is the number of coefficients of the equalizer, the mse of the equalized signal is generally close to the minimal obtainable. This is an improvement by a factor of three to ten[9,11–16] compared with the performance of the stochastic gradient algorithm. The exact improvement factor depends on the channel involved and on the modulation scheme used. Godard showed further that under certain modeling assumptions, the excess mse converges asymptotically, as the inverse of the number of iterations.

More recently, various methods were published[12,13,14] that reduce the computational complexity in the implementation of the Godard algorithm. These methods exploit the fact that only one new element, which may be a vector, is introduced in the vector of the received signal at each iteration. They avoid the processing of large matrices which are required in the Godard algorithm. Accordingly, the number of arithmetic operations can be reduced. For the Godard algorithm those grow quadratically, while for the algorithms described in Refs. 12 to 14 they grow proportionally to the number of equalizer coefficients—a considerable reduction for long equalizers. In Refs. 15 and 16 all these so-called least-squares algorithms are extended to include fractionally spaced, complex equalizer structures. The least-squares lattice algorithm was also extended to a decision feedback equalizer in Ref. 17. Investigations regarding the implementation are reported in Ref. 18.

The objective of this paper is to provide insight, on a fundamental level, into the rapid initial convergence of the least-squares algorithms relative to the stochastic gradient algorithm. Godard's approach is probabilistic and aimed at minimizing the ensemble mse as rapidly as possible. He assumed that all the involved random variables have joint Gaussian distributions. For this case, the Kalman filtering technique gives the fastest possible convergence. Although these assumptions are not generally satisfied in data transmission applications, a very powerful algorithm emerged.

In Ref. 19, a general equivalence between Kalman filtering and least-squares estimation techniques was exhibited. For the problem at hand, it implies that the algorithm obtained, while not optimal in a probabilistic sense, is the solution of a deterministic least-squares problem. This fact was stated for equalizers in Ref. 12. The Godard algorithm, as well as the ones proposed in Refs. 12 to 17, minimize the sum of the squared equalizer output errors under the condition that the coefficient vector remains constant from the start of the session to the current time. Note that this particular cost function is not the mse. Therefore, it does not explain directly why the actual mse converges so rapidly.

One obvious reason for the fast convergence of the least-squares algorithms is that all the information available from the start of the equalizer adaptation is stored and exploited in the update procedure, while the stochastic gradient algorithm relies mostly on current information. In Ref. 11, the Godard algorithm was interpreted as a stochastic gradient algorithm, where the coefficient corrections are transformed by an estimate of the inverse of the channel autocorrelation matrix. The fast initial convergence was attributed there to "self-orthogonalization" of the equalizer adjustments. However, this can only account for part of the high speed. It cannot explain the fact that the least-squares algorithms converge considerably faster than the stochastic gradient algorithms under the best conditions, i.e., when an ideal channel is involved or, equivalently, when the channel autocorrelation matrix is known exactly.

Here, we investigate the initial convergence of the deterministic least-squares algorithms from an algebraic point of view and offer alternative interpretations for the fast initial convergence. In Section II, the problem is stated and certain algebraic properties of the commonly utilized pseudo random start-up sequences are derived. In Section III, we relate the coefficient vector that results after N iterations (N being the number of coefficients) to the zero forcing equalizer and to the coefficients resulting with cyclic equalization.[7] The influence of the added noise is taken into account in Section IV.

II. THE OPTIMAL EQUALIZER COEFFICIENTS

Let $[a(k)]$ denote the complex data sequence which is transmitted at a signaling rate of $1/T$ over a channel with sampled impulse response $h(nT) = h_n$. The samples $\xi(nT)$ of the received signal, can be expressed as

$$\xi(nT) = \sum_{k=-\infty}^{\infty} a(k)h|(n+M-k)T| + \nu(nT), \tag{1}$$

where $\nu(nT)$ denotes the channel noise, and the equalizer length $N = 2M + 1$ is an odd number.

Let $x(n)$ denote the complex vector of the past N received signal samples

$$x(n)^T = [\xi|(nT)|, \xi|(n-1)T|, \cdots \xi|(n-N+1)T|], \tag{2}$$

and let $c(k)$ denote the complex coefficient vector of the adaptive equalizer. Then the equalizer output $y(n)$ at time nT, using the coefficient vector $c(k)$, can be written as the scalar product

$$y(n) = c(k)^* x(n). \tag{3}$$

The * denotes conjugate transposed vectors (matrices) or conjugate complex scalars.

2.1 The mean-square criterion

For the mean-square criterion, the equalizer coefficients are to be chosen such that the equalizer output $y(n)$ approximates the transmitted data value $a(n)$ as closely as possible. The optimum coefficient minimizes the mse[1,2]

$$\epsilon(k) = E[|c(k)^* x(n) - a(n)|^2], \tag{4}$$

where $E[\cdot]$ denotes ensemble averaging over the sequence $[a(n)]$. The vector c_{opt} which minimizes eq. (4) is given by

$$c_{opt} = A^{-1}v, \tag{5}$$

where

$$A = E[x(n)x(n)^*] \tag{6}$$

is the autocorrelation matrix of the channel, and

$$v = E[x(n)a(n)^*] \tag{7}$$

is the cross-correlation vector between the signal vector and the transmitted data value. In eq. (5), c_{opt} is the coefficient vector which, on the average, will perform better than any other. Ref. 2 indicates that the solution of eq. (5) exists if the absolute square of the transfer function and the spectral density of the noise have no zeros.

2.2 The least-squares criterion

The least-squares algorithms[9,11–16] minimize the following cost function

$$z(n) = \sum_{k=1}^{n} |c(n)^*x(k) - a(k)|^2, \tag{8}$$

i.e., the equalizer coefficient vector $c(n)$ minimizes the sum of error squares which resulted if $c(n)$ was used from the beginning of the transmission to the present instant nT. Usually, this is performed iteratively, i.e., $c(n)$ is calculated recursively for $n = 1 \cdots \infty$. Note that for $n \to \infty$, time invariant channels, ergodic data sequences $[a(n)]$ and noise, the cost function approaches asymptotically the ensemble mse, $\epsilon(n)$. Also, the solution $c(n)$ converges then to c_{opt}.

Differentiating eq. (8) with respect to $c(n)$ and setting the derivative to zero yields the following set of linear equations for the coefficient vector $c(n)$:

$$\sum_{k=1}^{n} x(k)x(k)^*c(n) = \sum_{k=1}^{n} x(k)a(k)^*. \tag{9}$$

Although the least-squares algorithms do not attempt to minimize the mse, it was observed[9,11–12] that the mse

$$\epsilon(n-1) = E[|c(n-1)^*x(n) - a(n)|^2], \tag{10}$$

which results if the equalizer's coefficient vectors stemming from the previous iteration at $(n - 1)T$ is used, converges roughly when n reaches N.

To find reasons for this interesting property, we examine closely in the following, the solution of eq. (9) after exactly N iterations and derive relations between the coefficient vector $c(N)$ and some well-known equalizer coefficient vectors, namely, the zero forcing and the cyclic equalizer.

In the noiseless case, we show that the vectors $x(1), \cdots x(N)$ are linearly independent of each other for the data sequences that are usually used for equalizer start-up. Under this hypothesis of linearly independent signal vectors $x(n)$, it can only be that the linear combination

$$\sum_{k=1}^{N} x(k)b(k) = 0, \tag{11}$$

if $b(k) = 0$ for all $k = 1 \cdots N$. Then, it follows from eqs. (9) and (11) that

$$x(k)^*c(N) = a(k)^* \qquad k = 1, \cdots N. \tag{12}$$

This means that $c(N)$ equalizes the first N received signal vectors ideally. All errors are zero and accordingly $z(N) = 0$. In the special case where the channel transfer function is of the all-pole type and of order $N - 1$, $z(n)$ will remain zero for $n > N$ and $c(n) = c(N)$. Therefore, the optimal coefficient vector is found exactly when $n = N$. Although this is not the case generally, $c(N)$ still has interesting properties.

III. ALGEBRAIC PROPERTIES OF SIGNAL VECTORS AND PARTICULAR EQUALIZER VECTORS WITHOUT NOISE

Using a notation which is similar to the one used in Ref. 4, it is possible to express the N-dimensional vector $x(k)$ in the noiseless case as follows:

$$x(k) = Bd(k), \tag{13}$$

$$\text{where} \quad B = \begin{vmatrix} \cdots h_0 & \cdots h_M & \cdots h_{N-1}\cdots \\ \vdots & \vdots & \vdots \\ \cdots h_{-M} & \cdots h_0 & \cdots h_M\cdots \\ \vdots & \vdots & \vdots \\ \cdots h_{1-N} & \cdots h_{-M} & \cdots h_0\cdots \end{vmatrix} \tag{14}$$

and

$$d(k)^T = [\cdots, a(k+M), \cdots, a(k), \cdots, a(k-M), \cdots]. \tag{15}$$

In eq. (14), B is a stationary $N \times L$ matrix, where L is at least the sum of the channel memory plus $(N - 1)$. The center part of length $N = 2M + 1$ is shown in eq. (14). In eq. (15), $d(k)$ is a stationary L-dimensional vector.

The vectors $x(1) \cdots x(N)$ are linearly independent, if they span the N dimensional space. This is equivalent to the $N \times N$ matrix

$$[x(1)\,|\,x(2)\,|\,\cdots\,|\,x(N)] = B[d(1)\,|\,\cdots\,|\,d(N)] \tag{16}$$

having rank N. This can only be true if both matrices on the right-hand side have rank N. This is a necessary but not a sufficient condition. Although eq. (16) has a very particular form, namely Toeplitz, it is not easy to find sufficient conditions for linear independence. We, therefore, investigate two interesting special cases involving particular start-up data sequences; namely, a sequence consisting of just a single pulse, and the other a periodic pseudo random noise sequence. Before pursuing this line of attack, we need additional facts about the matrix B.

For B to have rank N, the row vectors of B must be linearly independent. A necessary and sufficient condition is Gram's criterion: N vectors $b_1, \cdots b_N$ are linearly independent if and only if

$$\det \begin{vmatrix} b_1{}^*b_1 & b_1{}^*b_2 \cdots & b_1{}^*b_N \\ b_2{}^*b_1 & b_2{}^*b_2 \cdots & b_2{}^*b_N \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ b_N{}^*b_1 & \cdots\cdots & b_N{}^*b_N \end{vmatrix} \neq 0. \tag{17}$$

We identify $b_1{}^*$ as the first row of B and in general $b_n{}^*$ as the nth row of B. Consequently, the matrix in eq. (17) is the autocorrelation matrix of the channel. Gram's criterion then requires that the autocorrelation matrix be nonsingular. This is the same condition as the one required for the existence of a solution of eq. (5) in the noiseless case. Thus, whenever an optimal coefficient vector (in the mean-square sense) exists, then the matrix B has full rank N.

3.1 Single training pulse

Consider now the transmission of a single pulse at $k = 1 + M$. Then, inserting eqs. (13) to (15) into eq. (12) yields the following set of equations for the equalizer coefficient vector after N iterations:

$$\begin{vmatrix} h_0 \cdots & h_M \cdots & h_{N-1} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ h_{-M} \cdots & h_0 \cdots & h_M \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ h_{1-N} & h_{-M} \cdots & h_0 \end{vmatrix}^* c(N) = \begin{vmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}. \tag{18}$$

This is precisely the equation which defines the zero forcing equalizer.[1] In case the peak distortion of the channel impulse response is smaller than one, i.e.,

$$D = \frac{1}{|h_0|} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} |h_k| \leq 1, \tag{19}$$

the resulting equalizer minimizes the peak distortion of the overall channel,[1] and Gershgorin's criterion guarantees a unique solution of eq. (18). This is a sufficient but not a necessary condition.

This example shows that the least-squares equalizer adjustment technique yields the zero forcing equalizer in N iterations [if a unique solution of eq. (18) exists], whereas the gradient technique only attains this asymptotically in the steady state and practically requires a multiple of N iterations to obtain a good approximation.

As the time instant approaches infinity, the n equations

$$x(k)^{*}c(n) = a(k)^{*} \quad \text{and} \quad k = 1, \cdots n \tag{20}$$

cannot be satisfied simultaneously any more. In this case, $c(n)$ will be determined from eq. (9). Inserting eqs. (13) to (15) into eq. (9) yields

$$\lim_{n\geq\infty} BB^{*}c(n) = B \begin{vmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}. \tag{21}$$

Since BB^{*} is proportional to A as defined in eq. (6) and the right-hand side of eq. (21) is proportional to v as defined in eq. (7), with the same proportionality constant, it follows that

$$\lim_{n\geq\infty} c(n) = c_{\text{opt}}, \tag{22}$$

i.e., the least-squares algorithms converge in the noiseless case to the optimal coefficients in the mse sense.

3.2 Periodic pseudo random training sequence

While the technique of sounding the channel through isolated test pulses is technically possible, a different method has found wider use. In Refs. 7, 8, and 20, periodic pseudo random noise sequences (PRNSs) were proposed and analyzed for equalizer training purposes. When these sequences are used, the resulting equalizer coefficient vector in the steady state is found to be different from the optimal one when random data was used. Nevertheless, the former vector approaches the optimal solution very closely even for short periods. If a PRNS of period length $P = N$ is used, then the vectors of the sampled signal are periodic also and may be written as:

$$\tilde{x}(k) = \tilde{B}\tilde{d}(k), \tag{23}$$

$$\text{where} \quad \tilde{B} = \begin{vmatrix} \tilde{h}_0 \cdot\cdot & \tilde{h}_{N-1} \cdot\cdot \\ \cdots\cdot & \cdots\cdot \\ \cdots\cdot & \cdots\cdot \\ \tilde{h}_{1-N} & \tilde{h}_0 \end{vmatrix}, \tag{24}$$

$$\tilde{h}_n = \sum_{k=-\infty}^{\infty} h_{n+kN}, \tag{25}$$

and

$$\tilde{d}(k)^T = |\underset{\sim}{a}_{k+M}, \cdots \underset{\sim}{a}_k, \cdots \underset{\sim}{a}_{k-M}|. \tag{26}$$

The rows of $\tilde{B}$ and the vector $\tilde{d}\,(k)$ have dimension N.

Analyzing eqs. (24) and (25) reveals that $\tilde{B}$ is a $N \times N$ circulant matrix. It is well known from Ref. 7 that the eigenvalues of circulant matrices are determined by the discrete Fourier transform of the first row. Since the first row of the mentioned $N \times N$ circulant is formed by samples of the periodically repeated channel impulse response, it is concluded that $\tilde{B}$ has full rank N, provided that the discrete Fourier transform of the periodically repeated channel impulse response has no zero values.

This condition is very similar to the one stated for the existence of an optimal coefficient vector in the mse sense,[2] which, in the noiseless case, requires that the absolute value of the channel transfer function have no zeros.

From eq. (26) and the fact that the data sequence is periodic, i.e., $a(k + N) = a(k)$, it follows that N successive data vectors $d(k) \cdots \cdot d(k + N - 1)$ form an $N \times N$ circulant matrix. Arguing as above, it follows in general that these vectors are linearly independent if the dft of the data sequence has no zero values.

For PRNSs of period N in particular it is known that

$$\begin{aligned} d(k)^* d(k) &= N \\ d(k)^* d(j) &= -1 \qquad k \neq j. \end{aligned} \tag{27}$$

The matrix in Gram's criterion (17) is then a circulant with eigenvalues

$$\begin{aligned} \lambda_1 &= 1 \\ \lambda_j &= N + 1 \qquad j = 2 \cdots N. \end{aligned} \tag{28}$$

The determinant is $(N + 1)^{N-1} \neq 0$; therefore, N successive data vectors of a PRNS with period N are always linearly independent. This, together with the fact that $\tilde{B}$ has rank N, implies that any N different $\tilde{x}$ vectors are linearly independent. Therefore, the coefficient vector after N iterations is given as the solution of

$$\tilde{x}(k)^* c(N) = \tilde{a}(k)^* \quad \text{for} \quad k = 1, \cdots N \tag{29}$$

and is guaranteed to exist. In Ref. 7 it was shown that for the particular case of $N = P$, i.e., when the equalizer length equals the period P of the PRNS, the solution of eq. (29) has an interesting interpretation in the frequency domain: it equalizes the channel transfer function of the periodically repeated impulse response at N equidistant points. Again, this solution is obtained by the least-squares algorithms after N iterations, whereas the gradient technique obtains this only asymptotically as the number of iterations becomes large.

If the equalizer length is smaller than the period of the PRNS, i.e., $N < P$ no specific information on the nature of $c(N)$ can be obtained. We, therefore, consider the solution $c(P)$ after P iterations. Inserting eq. (23) in eq. (9) and using eq. (27) yields

$$\tilde{B}(I - D)\tilde{B}^* c(P) = \tilde{B} \begin{vmatrix} -1 \\ \vdots \\ -1 \\ P \\ -1 \\ \vdots \\ -1 \end{vmatrix} \frac{1}{P+1}, \tag{30}$$

where D is a matrix containing identical elements

$$D_{i,j} = \frac{1}{P+1}. \tag{31}$$

Using the fact that all rows of $\tilde{B}$ have the same sum

$$\sum_{K=1}^{N} \tilde{h}_K = \sum_{K=-\infty}^{\infty} h_K,$$

it follows that

$$\tilde{B}\tilde{B}^* c(P) = \tilde{v} + q, \tag{32}$$

where q is a vector with identical elements

$$q_i = \frac{1}{P+1} \left| \; \left| \sum_{k=-\infty}^{\infty} h_k \right|^2 \sum_{k=1}^{N} c_k(P) - \sum_{k=-\infty}^{\infty} h_k \right| \tag{33}$$

and

$$\tilde{v}_i = \tilde{h}_{-i}. \tag{34}$$

Observe that

$$\sum_{k=-\infty}^{\infty} h_k$$

and

$$\sum_{k=1}^{N} c_k(P)^\dagger$$

equal the *dc* value of the transfer function of the sampled channel and

---

† $c_k(P)$ denotes the kth element of $c(P)$.

of the equalizer, respectively. In the absence of noise and for $P \geq \infty$, their product will be one. Then, according to eq. (33), $q_i = 0$. Therefore, q_i will be small even for finite P, and $c(P)$ may be approximated as follows

$$c(P) \cong (\tilde{B}\tilde{B}^*)^{-1}\tilde{v}. \tag{35}$$

This result may be interpreted as the optimal solution in the mean-square sense for a channel with an impulse response of finite duration P which is identical to the periodically repeated impulse response in the base interval $|-PT/2, PT/2|$. If P is large enough to span the channel impulse response, then it follows that $c(P)$ is very close to the optimal coefficient vector after only P iterations.

IV. THE INFLUENCE OF NOISE

Generally, the channel noise is not negligible as assumed in the previous section. In the presence of additive noise the vector $w(k)$ of the sampled signal can be written as

$$w(k) = x(k) + r(k). \tag{36}$$

If only one single pulse is transmitted, $x(k)$ is defined in eq. (13) and $r(k)$ is the noise vector. In this case, a coefficient vector $c_a(n)$ is obtained after N iterations, which is the solution of the following equation

$$|H + R|\, c_a(N) = \begin{vmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}, \tag{37}$$

where

$$H = \begin{vmatrix} h_0 & \cdot\cdot & h_M & \cdot\cdot & h_{N-1} \\ & \cdot\cdot & & & \cdot \\ & \cdot & \cdot & & \cdot \\ & \cdot & & \cdot & \cdot \\ & \cdot & & \cdot & \cdot \\ & \cdot & & \cdot & \\ h_{1-N} & \cdot\cdot & h_{-M} & \cdot & \cdot\, h_0 \end{vmatrix}^* \tag{38}$$

and

$$R = \begin{vmatrix} \nu_1 & \nu_2 \cdots\cdots & \nu_N \\ \nu_0 & \nu_1 \cdots\cdots & \\ \vdots & \vdots & \vdots \\ \nu_{2-N} & \cdots\cdots & \nu_1 \end{vmatrix}^*. \tag{39}$$

The difference between this solution and the one given in eq. (18) equals

$$c_a(N) - c(N) = H^{-1}Rc_a(N). \tag{40}$$

If $c_a(N)$ is used instead of $c(N)$ to compute the value of the cost function defined in eq. (8), we obtain

$$z_a(N) = |c_a(N) - c(N)|^* H^* H |c_a(N) - c(N)|. \tag{41}$$

Substituting eq. (40) into eq. (41) and evaluating the expected value of the cost function yields

$$E|z_a(N)| = E|c_a(N)^* R^* R c_a(N)| = c_a(N)^* E|R^* R| c_a(N), \tag{42}$$

where $E|R^*R|$ is N times the correlation matrix of the random noise. Using Parseval's theorem, eq. (42) can be expressed in terms of the transfer function $C_a(\omega)$ of the equalizer and of the power density spectrum $S_\nu(\omega)$ of the noise $\nu(n)$

$$E|z_a(N)| \cong TN/2\pi \int_{-\pi/T}^{\pi/T} |C_a(\omega)|^2 S_\nu(\omega) d\omega. \tag{43}$$

Thus, increase of the cost function is N times the average noise power after the equalizer. This means that the average squared error (12) per equalized symbol is equal to the noise variance after the equalizer.

We now examine the solution of eq. (9) for $n > N$. In this case, the equation for the equalizer coefficient vector becomes

$$\left| B(n)B(n)^* + \sum_{k=1}^{n} r(k)r(k)^* \right| c(n) = B(n) \begin{vmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{vmatrix}, \tag{44}$$

where $B(n)$ is similar to eq. (14) with row length equal to n. This indicates that, as n becomes large, the influence of the noise grows proportionally. Since $B(n)B(n)^*$ converges to the channel correlation matrix $A = BB^*$, which is stationary, it can be concluded that for very large n the influence of the noise becomes dominant. Therefore, it is not advisable to use the sounding technique for more than N to $2N$ iterations.

If a PRNS with period P symbols is used during start-up and noise is present, the coefficient vector after P iterations is not determined anymore by eq. (35). Inserting eqs. (36) and (23) into eq. (9) yields instead

$$\left| \tilde{B}\tilde{B}^* + \frac{1}{P+1} \sum_{k=1}^{P} r(k)r(k)^* \right| c(P) \cong \tilde{v}, \tag{45}$$

where it was assumed that cross products of $r(k)$ and $x(k)$ may be neglected.

The matrix of eq. (45) will, in the mean, become the correlation matrix of the channel plus the noise. Therefore, the solution will, in the mean, be the optimal solution as given by eq. (5). Note that now there is no danger in letting the algorithm run for an indefinite time, since both terms of the matrix, as well as the right-hand side of the equation, grow proportionally.

V. CONCLUSION

The initial convergence of least-squares equalizer adjustment algorithms was analyzed to determine why the least-squares algorithms converge so much faster than the widely used stochastic gradient algorithms. The algebraic properties of the sampled signal vectors were found to be of crucial importance for the convergence behavior. In particular, it was found that, for a wide class of transmission channels and for commonly used data sequences, successive sampled signal vectors are linearly independent. This ensures a unique equalizer coefficient vector after exactly N iterations, where N is the dimension of the equalizer. In the noiseless case, this coefficient vector was found to correspond to particular equalizer coefficients which were reported and studied earlier. If a single pulse is transmitted, the zero forcing equalizer is obtained. If a pseudo random noise sequence with a period in symbols equal to the number of equalizer coefficients is used, the steady state solution of the cyclic equalization technique results. This explains why the least-squares adjustment algorithms converge much faster than the gradient techniques: the above-mentioned particular equalizer coefficients are obtained after only N iterations, whereas with the stochastic gradient techniques they are only approximated as the number of iterations becomes very large. The influence of the inevitable channel and measurement noises was evaluated. Approximations show that similar performance, as in the noiseless case, is obtainable.

REFERENCES

1. R. W. Lucky, J. Salz, and E. J. Weldon, Jr, *Principles of Data Communication*, New York: McGraw-Hill, 1968.
2. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels," B.S.T.J., *48*, No. 1 (January 1969), pp. 55–70.
3. G. Ungerboeck, "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communication," IBM J. Res. and Dev., *16*, No. 6 (November 1972), pp. 546–55.

4. J. E. Mazo, "On the Independence Theory of Equalizer Convergence," B.S.T.J., *58*, No. 5 (May–June 1979), pp. 963–93.
5. R. W. Chang, "A New Equalizer Structure for Fast Start-Up Digital Communication," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1969–2014.
6. K. H. Mueller, "A New, Fast-Converging Mean-Square Algorithm for Adaptive Equalizers with Partial-Response Signaling," B.S.T.J., *54*, No. 1 (January 1975), pp. 143–53.
7. K. H. Mueller and D. A. Spaulding, "Cyclic Equalization—A New Rapidly Converging Equalization Technique for Synchronous Data Communication," B.S.T.J., *54*, No. 7 (February 1975), pp. 369–406.
8. S. U. H. Qureshi, "Fast Start-Up Equalization with Periodic Training Sequences," IEEE Trans. Inform. Theory, *IT-23*, No. 5 (Sept. 1977), pp. 553–63.
9. D. Godard, "Channel Equalization Using a Kalman Filter for Fast Data Transmission," IBM J. Res. Develop., *18*, No. (May 1974).
10. A. E. Albert and L. A. Gardner, Jr., *Stochastic Approximation and Nonlinear Regression*, Cambridge: M.I.T. Press, 1967, Chapter 7.
11. R. D. Gitlin and F. R. Magee, "Self-Orthogonalizing Adaptive Equalization Algorithms," IEEE Trans. Commun., *COM-25*, No. 7 (July 1977), pp. 666–72.
12. D. D. Falconer and L. Ljung, "Application of Fast Kalman Estimation to Adaptive Equalization," IEEE Trans. Commun., *COM-26*, No. 10 (October 1978), pp. 1439–46.
13. E. H. Satorius and T. D. Pack, unpublished work.
14. E. Shichor, unpublished work.
15. T. L. Lim and M. S. Mueller, "Rapid Equalizer Start-Up Using Least-Squares Algorithms," Conf. Rec., Int. Conf. Commun., 1980, Seattle, Wa., Paper No. 57.7.
16. M. S. Mueller, "Least-Squares Algorithms for Adaptive Equalizers," B.S.T.J., *60*, No. 8 (October 1981), pp. 193–213.
17. M. T. Shensa, "A Least-Squares Lattice Decision Feedback Equalizer," Conf. Rec., Int. Conf. Commun., 1980, Seattle, Wa., Paper No. 57.6.
18. D. D. Falconer, V. B. Lawrence, and S. K. Tewksbury, "Processor-Hardware Considerations for Adaptive Digital Filter Algorithms," Conf. Rec., Int. Conf. Commun., 1980, Seattle, Wa., Paper No. 57.5.
19. H. W. Sorenson, "Comparison of Kalman, Bayesian and Maximum Likelihood Estimation Techniques," *Theory and Applications of Kalman Filtering*, NATO, Agardograph, No. 139, 1970, Chapter 6.
20. R. W. Chang and E. Y. Ho, "On Fast Start-Up Data Communication Systems Using Pseudo-Random Training Sequences," B.S.T.J. *51*, No. 9 (November 1972), pp. 2013–27.