D. Godard

Channel Equalization Using a Kalman Filter for Fast Data Transmission

Abstract: This paper shows how a Kalman filter may be applied to the problem of setting the tap gains of transversal equalizers to minimize mean-square distortion. In the presence of noise and without prior knowledge about the channel, the filter algorithm leads to faster convergence than other methods, its speed of convergence depending only on the number of taps. Theoretical results are given and computer simulation is used to corroborate the theory and to compare the algorithm with the classical steepest descent method.

Introduction

Data transmission systems generally use voiceband communication channels. These are characterized by a relatively narrow bandwidth (300 to 3000 Hz), a high signal-to-noise ratio (about 20 to 30 dB), and amplitude and phase distortion slowly varying in time. High speed data transmission then requires equalization. Many presently used modem receivers are equipped with a matched filter to maximize the signal-to-noise ratio and an equalizer to minimize the inter-symbol interference due to distortion. Equalizers usually are of the transversal filter type (tapped delay line filter) with tap gains adjusted to minimize some error criterion. An automatic equalization process requires an initial training period during which the equalizer reduces the error. In "preset equalization," isolated pulses are transmitted prior to data transmission, and the derived tap-gain settings are kept constant during the data transmission itself. Periodically, a short training period may be entered to update the tap gains.

A second kind of equalization process is known as "adaptive equalization." Here the equalizer settings are derived from the received signal. During the training period, the equalizer continuously seeks to minimize the deviation of its sampled output signal from an ideal reference signal generated internally in proper synchronism in the receiver. When the residual distortion is small enough, actual data may be transmitted. The equalizer is then switched into the "decision-directed mode," using as reference a reconstructed signal obtained by thresholding the output signal of the equalizer. Adaptive equalization has many advantages over preset equalization, among them being the ability to adapt to changes in channel characteristics during the transmission.

Clearly, there is a delay in data transmission proportional to the length of the training period, and a decrease in this delay is desirable. Many adjustment algorithms [1-11] have been described in the literature, often emphasizing the speed of convergence. For the well-known mean-square algorithm Gersho [4] showed that the speed of convergence is largely determined by the maximum and minimum values of the power spectrum of the unequalized signal. Similar results for more sophisticated algorithms have been reported by Chang [5] and Kobayashi [6]. To achieve fast equalization, a new equalizer structure has been developed by Sha and Tang [11], although their theory can fail for large distortion. Devieux and Pickholtz [7] studied adaptive equalization with a second-order gradient algorithm, but that algorithm requires computation of the covariance matrix of the sampled received signal. More recently, Ungerboeck [8] showed that, in the speed of convergence of the mean-square algorithm, the influence of the number of taps usually dominates. He gave a new criterion for convergence and an optimal step size parameter in the adjustment loops.

In this paper, a new algorithm, based on Kalman filtering theory [12, 13] is proposed for obtaining fast convergence of the tap gains of transversal equalizers to their optimal settings. A Kalman filter had previously been applied to channel equalization by Lawrence and Kaufman [15] but in a quite different way, since in their study the equalizer is replaced by the filter.

It is shown here that the convergence of mean-square distortion is obtained, under noisy conditions, within a number of iterations determined only by the number of taps, without prior information about the channel. First, fundamentals of Kalman's theory are reviewed. Then we show how to apply the Kalman filter to the equalizer and derive an expression for the speed of convergence. Finally, computer simulations are used to check the validity of the theory and to compare the proposed algorithm with the steepest descent method. Comparison is also made with other sophisticated algorithms.

Kalman-Bucy filtering theory

Over a decade ago, with publication of now classical papers, Kalman [12] and Kalman and Bucy [13] defined a recursive method and deeply transformed filtering and predicting theories.

Application of the Kalman filter supposes the studied system to be described by a set of linear difference equations, in the case of discrete systems that are of interest in this study. Let $\mathbf{x}_k$ be, at the kth sampling instant, the N-dimensional vector of the N state variables [14] of a system modeled as follows:

$$\mathbf{x}_k = \mathbf{\Phi}(k, k-1)\mathbf{x}_{k-1} + \mathbf{W}_k, \quad (1)$$

$$\mathbf{Z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{V}_k, \quad (2)$$

where $\mathbf{\Phi}(k, k-1)$ is the $N \times N$ state transition matrix; $\mathbf{Z}_k$, the M-dimensional measurement vector; $\mathbf{H}_k$, the $M \times N$ measurement matrix; and $\mathbf{W}_k$ and $\mathbf{V}_k$, respectively, N- and M-dimensional vectors of white noise processes of zero mean. Later we use the covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ of $\mathbf{W}_k$ and $\mathbf{V}_k$. It is assumed that the noise processes $\mathbf{W}$ and $\mathbf{V}$ are statistically independent; i.e., if E denotes expectation, then

$$\mathrm{E}[\mathbf{W}_k\mathbf{V}_j'] = 0, \quad k, j = 1, 2, \cdots.$$

Throughout the paper, a prime (′) denotes matrix transposition.

Assume we know at the $(k-1)$th sampling instant an estimate $\hat{\mathbf{x}}_{k-1}$ of the actual state vector $\mathbf{x}_{k-1}$ and the error covariance matrix

$$\mathbf{P}_{k-1} = \mathrm{E}[(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})']. \quad (3)$$

It is possible to derive from (1) a predicted value of the state at the kth sampling instant,

$$\hat{\mathbf{x}}_{k,k-1} = \mathbf{\Phi}(k, k-1)\hat{\mathbf{x}}_{k-1}, \quad (4)$$

and a predicted error covariance matrix defined by

$$\mathbf{P}_{k,k-1} = \mathrm{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k,k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k,k-1})']. \quad (5)$$

It can be shown that

$$\mathbf{P}_{k,k-1} = \mathbf{\Phi}(k, k-1)\mathbf{P}_{k-1}\mathbf{\Phi}'(k, k-1) + \mathbf{Q}_k. \quad (6)$$

A predicted measurement $\hat{\mathbf{Z}}_k$ is derived from (2):

$$\hat{\mathbf{Z}}_k = \mathbf{H}_k\hat{\mathbf{x}}_{k,k-1}. \quad (7)$$

The deviation of the predicted measurement from the actual measurement $\mathbf{Z}_k$ is used to define a new estimate of the state vector as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k,k-1} + \mathbf{K}_k(\mathbf{Z}_k - \hat{\mathbf{Z}}_k). \quad (8)$$

The $N \times M$ correction matrix $\mathbf{K}_k$ is computed in order to minimize the trace of the error covariance matrix $\mathbf{P}_k$. Then it can be shown that $\mathbf{K}_k$ and $\mathbf{P}_k$ are given by

$$\mathbf{K}_k = \mathbf{P}_{k,k-1}\mathbf{H}_k'(\mathbf{H}_k\mathbf{P}_{k-1}\mathbf{H}_k' + \mathbf{R}_k)^{-1} \quad (9)$$

and

$$\mathbf{P}_k = \mathbf{P}_{k,k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k,k-1}. \quad (10)$$

Application of the Kalman filter requires an estimate $\hat{\mathbf{x}}(0)$ of the initial state vector and computation of the corresponding error covariance matrix $\mathbf{P}_0$. Usually $\hat{\mathbf{x}}(0)$ is chosen as the mean value of $\mathbf{x}(0)$. Then it can be shown that the subsequent estimates are unbiased.

The Kalman filter defined by recursive formulas (4) through (10) leads to minimization of the trace of the error covariance matrix. More generally, if $\mathbf{A}$ is a symmetric positive definite matrix, minimization of the product $(\mathbf{x}_k - \hat{\mathbf{x}}_k)'\mathbf{A}(\mathbf{x}_k - \hat{\mathbf{x}}_k)$ is obtained.

Application of a Kalman filter to equalizers

For simplicity we limit ourselves to a pulse-amplitude-modulated (PAM) baseband system. The equalizer is an N-tap delay line filter. The input signal of the equalizer is the PAM baseband signal

$$U(t) = \sum_{k=-\infty}^{\infty} a_k h(t - kT) + v(t), \quad (11)$$

where $\{a_k\}$ is the sequence of quantized pulses to be transmitted, $h(t)$ is the channel response, T is the intersymbol separation (or baud interval), and $v(t)$ is the additive noise. Both sequence $\{a_k\}$ and noise $v(t)$ are stationary. Let $\mathbf{u}_k$ be the vector of tap output signals and $\mathbf{c}_k$ be the vector of tap gains of the equalizer, both at the kth sampling instant. These two vectors are N-dimensional. The output signal of the equalizer is

$$s_k = \mathbf{u}_k'\mathbf{c}_k, \quad (12)$$

and we define the error signal as

$$e_k = a_k - s_k. \quad (13)$$

We choose as the error criterion the minimization of the expected mean-square distortion,

$$\mathscr{E}^2 = \mathrm{E}[e_k^2]. \quad (14)$$

With sequence $\{a_k\}$ and noise $v(t)$ being assumed stationary, $\mathscr{E}^2$ does not depend on k for a given tap-gain setting.

It can be shown [4] that the optimal tap gains $\mathbf{c}_{opt}$ are given by

$$\mathbf{c}_{opt} = \mathbf{A}^{-1}\mathbf{b}, \tag{15}$$

where $\mathbf{A}$ is the $N \times N$ symmetric positive definite matrix

$$\mathbf{A} = \mathrm{E}[\mathbf{u}_k\mathbf{u}_k'], \tag{16}$$

and $\mathbf{b}$ is the N-dimensional vector

$$\mathbf{b} = \mathrm{E}[a_k\mathbf{u}_k]. \tag{17}$$

When $\mathbf{c}_k$ is chosen equal to $\mathbf{c}_{opt}$, the mean-square distortion assumes its minimum value $\mathscr{E}^2_{opt}$. Even if no noise is present, $\mathscr{E}^2_{opt}$ is not zero because of the finiteness of the number of taps. Let $e_{k\ opt}$ be the error signal with the optimal adjustment of the tap gains. Then we have

$$a_k = \mathbf{u}_k'\mathbf{c}_{opt} + e_{k\ opt}, \tag{18}$$

with

$$\mathscr{E}^2_{opt} = \mathrm{E}[e^2_{k\ opt}]. \tag{19}$$

During the equalization process, the expected mean-square distortion at the kth sampling instant is, using (18),

$$\mathscr{E}_k{}^2 = \mathrm{E}[\{\mathbf{u}_k'(\mathbf{c}_{opt} - \mathbf{c}_k) + e_{k\ opt}\}^2]. \tag{20}$$

It is well known [3] that

$$\mathrm{E}[e_{k\ opt}\mathbf{u}_k] = \mathbf{0}, \tag{21}$$

so the variables $\mathbf{u}_k'(\mathbf{c}_{opt} - \mathbf{c}_k)$ and $e_{k\ opt}$ are uncorrelated and (20) may be rewritten as

$$\mathscr{E}_k{}^2 = \mathrm{E}[(\mathbf{c}_{opt} - \mathbf{c}_k)'\mathbf{u}_k\mathbf{u}_k'(\mathbf{c}_{opt} - \mathbf{c}_k)] + \mathscr{E}^2_{opt}. \tag{22}$$

As in [4] and [8] we assume that the dependence between $\mathbf{u}_k$ and $\mathbf{c}_k$ may be neglected. Denoting by $\mathbf{P}_k$ the covariance matrix,

$$\mathbf{P}_k = \mathrm{E}[(\mathbf{c}_{opt} - \mathbf{c}_k)(\mathbf{c}_{opt} - \mathbf{c}_k)'], \tag{23}$$

and using (16), the expected mean-square distortion at the kth sampling instant is given by

$$\mathscr{E}_k{}^2 = \mathrm{trace}\ \mathbf{P}_k\mathbf{A} + \mathscr{E}^2_{opt}. \tag{24}$$

Since $\mathbf{A}$ is a symmetric positive definite matrix, minimization of (24) is done by the algorithm by choosing the tap gains as state variables. To simplify the theory we assume the channel characteristics to be stationary over the training period, so that the optimum tap-gain values are constant during the settling time of the equalizer. This is not a severe assumption since, for normal transmission speeds, the equalizer settles in a few milliseconds.

The problem is now stated: One wants to identify the optimal equalizer characterized by the constant state variables $\mathbf{c}_{opt}$, knowing that the output signal $\mathbf{u}_k'\mathbf{c}_{opt}$ satisfies

$$a_k = \mathbf{u}_k'\mathbf{c}_{opt} + e_{k\ opt}. \tag{18}$$

Referring to the general state equations (1) and (2), one sees that the state transition matrix is here the identity matrix and that (18) is the measurement equation, $e_{k\ opt}$ appearing as the measurement noise. Clearly, sequence $\{a_k\}$ and noise $v(t)$ being of zero mean, $e_{k\ opt}$ is a random variable of zero mean. To apply the Kalman-filter approach, we assume that $e_{k\ opt}$ may be considered as a white noise process of zero mean and variance $\mathscr{E}^2_{opt}$. This seems to be a reasonable approximation because the optimal mean-square distortion is usually small, including intersymbol interference still present at the output of the optimal equalizers, so that the correlation between successive samples of the noise may be neglected. However, this assumption is not necessary for the convergence of the algorithm.

Let us assume we know at the $(k-1)$th sampling instant an estimate $\mathbf{c}_{k-1}$ of the state vector and the error covariance matrix $\mathbf{P}_{k-1}$. Obviously, since the state transition matrix is the identity matrix, the predicted state vector $\mathbf{c}_{k,k-1}$ is equal to $\mathbf{c}_{k-1}$, and the predicted matrix $\mathbf{P}_{k,k-1}$ is equal to $\mathbf{P}_{k-1}$. The predicted measurement is

$$\hat{s}_k = \mathbf{u}_k'\mathbf{c}_{k-1},$$

and the new estimate $\mathbf{c}_k$ is given by

$$\mathbf{c}_k = \mathbf{c}_{k-1} + \mathbf{K}_k(a_k - \hat{s}_k), \tag{25}$$

with

$$\mathbf{K}_k = \mathbf{P}_{k-1}\mathbf{u}_k(\mathbf{u}_k'\mathbf{P}_{k-1}\mathbf{u}_k + \mathscr{E}^2_{opt})^{-1} \tag{26}$$

and

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{K}_k\mathbf{u}_k'\mathbf{P}_{k-1}. \tag{27}$$

Computation of the Kalman gain $\mathbf{K}_k$, here reduced to an N-dimensional vector, involves only inversion of a scalar quantity, but requires prior knowledge of the optimal mean-square distortion. Clearly, $\mathscr{E}^2_{opt}$ cannot be known a priori. Usually, after equalization, the signal-to-noise ratio at the output of the equalizer is between 20 and 30 dB, so that to compute the Kalman gain one can use an estimated value $\hat{\mathscr{E}}^2_{opt}$ of the optimal mean-square distortion included between 0.01 and 0.001. Later we show that the estimated value of $\mathscr{E}^2_{opt}$ has no influence on the successive estimates of the tap gains.

As initial estimate $\mathbf{c}_0$ of the tap gains, we choose $\mathbf{c}_0 = \mathbf{0}$. The optimal tap gains being assumed to be uniformly distributed between plus and minus 1.5, the initial covariance matrix $\mathbf{P}_0$ is the matrix with elements

$$(\mathbf{P}_0)_{ij} = 0.75\ \delta_{ij}, \quad i, j = 1, \cdots, N,$$

δ_{ij} being the Kronecker function. The off-diagonal terms of the matrix are zero, to reflect the statistical independence of the initial estimates.

• *Speed of convergence*

Equation (26) may be rewritten as

$$(\mathbf{I} - \mathbf{K}_k\mathbf{u}_k')\mathbf{P}_{k-1}\mathbf{u}_k = \mathbf{K}_k\mathscr{E}^2_{\text{opt}}, \tag{26'}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. Multiplying (27) on the right by $\mathbf{u}_k$ and using (26'), we obtain

$$\mathbf{P}_k\mathbf{u}_k = \mathbf{K}_k\mathscr{E}^2_{\text{opt}},$$

or

$$\mathbf{K}_k = \mathbf{P}_k\mathbf{u}_k/\mathscr{E}^2_{\text{opt}}. \tag{28}$$

Eliminating $\mathbf{K}_k$ from (27), we obtain

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{P}_k\mathbf{u}_k\mathbf{u}_k'\mathbf{P}_{k-1}/\mathscr{E}^2_{\text{opt}}. \tag{27'}$$

The $\mathbf{P}_k$ matrix, being a covariance matrix, is positive definite and has an inverse. Thus multiplication of (27') on the left by $\mathbf{P}_k^{-1}$ and on the right by $\mathbf{P}_{k-1}^{-1}$ leads to

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k-1}^{-1} + \mathbf{u}_k\mathbf{u}_k'/\mathscr{E}^2_{\text{opt}}. \tag{29}$$

From (29) we derive

$$\mathbf{P}_k = \mathscr{E}^2_{\text{opt}}\left(\mathscr{E}^2_{\text{opt}}\mathbf{P}_0^{-1} + \sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}. \tag{30}$$

After a few iterations, $\mathscr{E}^2_{\text{opt}}$ usually being smaller than 0.01, the diagonal matrix $\mathscr{E}^2_{\text{opt}}\mathbf{P}_0^{-1}$ may be neglected in (30) so that $\mathbf{P}_k$ becomes

$$\mathbf{P}_k = \mathscr{E}^2_{\text{opt}}\left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}, \tag{31}$$

and (28) becomes

$$\mathbf{K}_k = \left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}\mathbf{u}_k, \tag{32}$$

showing that the dependence between the Kalman gains and $\mathscr{E}^2_{\text{opt}}$ may be neglected for usual values of $\mathscr{E}^2_{\text{opt}}$.

Clearly, the matrix $k^{-1}(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i')$ converges to the $\mathbf{A}$ matrix when k goes to infinity. Nevertheless, one can estimate that after 30 or 40 iterations it is possible to write

$$\text{trace } \mathbf{P}_k\mathbf{A} \approx \mathscr{E}^2_{\text{opt}}Nk^{-1}. \tag{33}$$

Then, from (24), if the optimal mean-square distortion is known a priori, we derive

$$\mathscr{E}_k^{\,2} \approx \mathscr{E}^2_{\text{opt}}(1 + Nk^{-1}), \tag{34}$$

which means that convergence can theoretically be obtained within less than $2N$ steps.

• *Further analysis of the algorithm*

It has been seen that the algorithm can be applied to the identification of the optimal equalizer if the optimal error $e_{k\text{ opt}}$ may be considered as white noise. We now show that this assumption is not necessary.

Denote by $\mathbf{S}_k$ the $N \times N$ matrix defined by

$$\mathbf{S}_k = \mathbf{P}_k/\mathscr{E}^2_{\text{opt}}, \quad k = 0, 1, 2, \cdots. \tag{35}$$

Then (26) and (27) become

$$\mathbf{K}_k = \mathbf{S}_{k-1}\mathbf{u}_k/(1 + \mathbf{u}_k'\mathbf{S}_{k-1}\mathbf{u}_k) \text{ and} \tag{36}$$

$$\mathbf{S}_k = \mathbf{S}_{k-1} - \mathbf{K}_k\mathbf{u}_k'\mathbf{S}_{k-1}. \tag{37}$$

It is easy to verify that equations (28) and (29) may be rewritten as

$$\mathbf{K}_k = \mathbf{S}_k\mathbf{u}_k \text{ and} \tag{38}$$

$$\mathbf{S}_k^{-1} = \mathbf{S}_{k-1}^{-1} + \mathbf{u}_k\mathbf{u}_k', \text{ or} \tag{39}$$

$$\mathbf{S}_k = \left(\mathbf{S}_0^{-1} + \sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}. \tag{40}$$

Substituting (38) into (25) we obtain

$$\mathbf{c}_k = \mathbf{c}_{k-1} + \mathbf{S}_k\mathbf{u}_k a_k - \mathbf{S}_k\mathbf{u}_k\mathbf{u}_k'\mathbf{c}_{k-1}. \tag{41}$$

Substituting $\mathbf{S}_k^{-1} - \mathbf{S}_{k-1}^{-1}$ from (39) for $\mathbf{u}_k\mathbf{u}_k'$ and multiplying both sides of (41) by $\mathbf{S}_k^{-1}$, we have

$$\mathbf{S}_k^{-1}\mathbf{c}_k = \mathbf{S}_{k-1}^{-1}\mathbf{c}_{k-1} + a_k\mathbf{u}_k. \tag{42}$$

With $\mathbf{c}_0 = \mathbf{0}$, (42) may be rewritten as

$$\mathbf{c}_k = \mathbf{S}_k\sum_{i=1}^{k}a_i\mathbf{u}_i$$

or, using (40),

$$\mathbf{c} = \left(\mathbf{S}_0^{-1} + \sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}\sum_{i=1}^{k}a_i\mathbf{u}_i. \tag{43}$$

For sufficiently large k, such as $k \geqslant N$, the diagonal matrix $\mathbf{S}_0^{-1}$ may be neglected in (43), so that the kth estimate of the tap gains is given by

$$\mathbf{c}_k = \left(\sum_{i=1}^{k}\mathbf{u}_i\mathbf{u}_i'\right)^{-1}\sum_{i=1}^{k}a_i\mathbf{u}_i \tag{44}$$

and is independent of the previous estimates if the initial estimate of $\mathbf{c}_0$ is $\mathbf{0}$.

From (44) it is obvious that the kth estimate $\mathbf{c}_k$ is the optimal one for the received sequence up to the kth sampling instant. For example, if the pseudorandom sequence used during the training mode is periodic, sending only one period of the sequence is sufficient. The algorithm must give good results even if the signal-to-noise ratio is small, since the algorithm builds up the inverse of the correlation matrix of the sampled received signal, which is corrupted by noise. The assumption that $e_{k\text{ opt}}$ is a white noise process is necessary only to express the $\mathbf{P}_k$ matrix as the error covariance matrix of the tap gains.

• *Adaptation to slowly varying channels*

The optimal tap-gain values are time varying as a consequence of the amplitude and phase characteristics of real channels being not stationary. From (31) and (32) the $\mathbf{P}_k$ matrix elements and the Kalman gain $\mathbf{K}_k$ converge to zero when k goes to infinity, so the equalizer cannot adapt itself to changes in channel characteristics during the transmission. Nevertheless, one can easily derive adaptive techniques from previous theory:

1. One can assume that the optimal tap-gain values are randomly varying about a mean value. This leads to the state equation

   $$\mathbf{c}_{k\,\mathrm{opt}} = \mathbf{c}_{(k-1)\,\mathrm{opt}} + \Delta\mathbf{c}_k, \tag{45}$$

   where $\Delta\mathbf{c}_k$ is considered as a white noise process. Then one has to calculate the correlation matrix

   $$\mathbf{Q} = \mathrm{E}[\Delta\mathbf{c}_k \Delta\mathbf{c}_k']$$

   and, at each step from (6), the predicted error covariance matrix

   $$\mathbf{P}_{k,k-1} = \mathbf{P}_{k-1} + \mathbf{Q}.$$

   Although Eq. (45) does not describe the true situation, it could give good results in the case of rapidly varying channels.
2. One can freeze the $\mathbf{P}_k$ matrix after, say, $5N$ sampling intervals. The Kalman gain stays sufficiently large to ensure adaptation. (This procedure may be compared with the one used by Chang [5] when his prefixed weighting matrix is not perfectly suited to the $\mathbf{A}$ matrix.)
3. When the equalizer is switched into the decision-directed mode, the $\mathbf{P}_k$ matrix is restated and fixed to a diagonal matrix with elements $(\mathbf{P})_{ij} = \alpha \mathscr{E}^2_{\mathrm{opt}} \delta_{ij}$, where α is the step-size parameter usually used in the stochastic gradient method [8]. It is easy to verify, referring to (28) and (25), that in this case the equalization process becomes the same as in the steepest descent method, which gives good enough results when the equalizer has only to track slow changes in channel characteristics during data transmission.

This procedure would be attractive in a signal processor in which a large part of the computation power could be used during a brief portion of the start-up phase to achieve a fast reduction of mean-square distortion.

Computer simulation

During the equalization process, the expected mean-square distortion at the kth sampling instant is approximately given by

$$\mathscr{E}_k^{\,2} = \mathscr{E}^2_{\mathrm{opt}}(1 + Nk^{-1}), \tag{34}$$

showing that convergence must be obtained within less than $2N$ steps, and that the speed of convergence does not depend on the characteristics of the channel. Computer simulation has been used to check the validity of these assertions and to compare the speed of convergence of the proposed algorithm with that in the steepest descent method, where estimates of the tap gains are iteratively given by

$$\mathbf{c}_{k+1} = \mathbf{c}_k + \mu e_k \mathbf{u}_k.$$

In addition, the influence of the estimated value $\hat{\mathscr{E}}^2_{\mathrm{opt}}$ of the optimal mean-square distortion was investigated. As step-size parameter we chose the optimal one defined by Ungerboeck [8]:

$$\mu = 1/N\langle \mathbf{u}^2 \rangle,$$

where $\langle \mathbf{u}^2 \rangle$ denotes the energy of the unequalized signal.

The algorithms have been tested with three channels:

Channel 1	Moderate amplitude and phase distortion
Channel 2	Heavy amplitude distortion, no phase distortion
Channel 3	Heavy amplitude and phase distortion.

For Channels 2 and 3 a large spread of the eigenvalues of the $\mathbf{A}$ matrix occurs, leading to slow convergence with the steepest descent method.

Two programs have been written. The first one, for a given voiceband communications channel, determines $\mathbf{A}$, $\mathbf{b}$, $\mathbf{c}_{\mathrm{opt}}$, and $\mathscr{E}^2_{\mathrm{opt}}$. The second program generates a random sequence of bipolar signals $(a_k = \pm 1)$, simulates the channel, adds white Gaussian noise, and simulates the equalizer. A signal-to-noise ratio of 30 dB is assumed at the input of the equalizer. At each sampling instant the mean-square distortion $\langle e_k^{\,2} \rangle$ is computed from

$$\langle e_k^{\,2} \rangle = (\mathbf{c}_k - \mathbf{c}_{\mathrm{opt}}) \mathbf{A} (\mathbf{c}_k - \mathbf{c}_{\mathrm{opt}}) + \mathscr{E}^2_{\mathrm{opt}}$$

for various estimates of the optimal mean-square distortion.

The results of the simulation are presented in Figs. 1 and 2. From them we draw the following conclusions:

1. The speed of convergence of $\langle e_k^{\,2} \rangle$ does not depend on the choice of $\hat{\mathscr{E}}^2_{\mathrm{opt}}$, provided that the value chosen is reasonably small but not zero. A zero value would mean that no noise is present and that the equalizer is of infinite length. For a given channel, three computer runs with $\hat{\mathscr{E}}^2_{\mathrm{opt}} = 0.1$, 0.001, and 0.0001 gave identical results for the same sequence $\{a_k\}$ and the same sequence of noise samples $v(kT)$.
2. The speed of convergence is independent of the characteristics of the channel. A large spread of the eigenvalues of the $\mathbf{A}$ matrix, as is the case for Channels 2 and 3, leads to a slow convergence with the steepest descent method but has no effect on the speed of convergence obtained through our algorithm.
3. Good agreement of the expected mean-square distortion theoretically predicted by (34) and $\langle e_k^{\,2} \rangle$ ob-

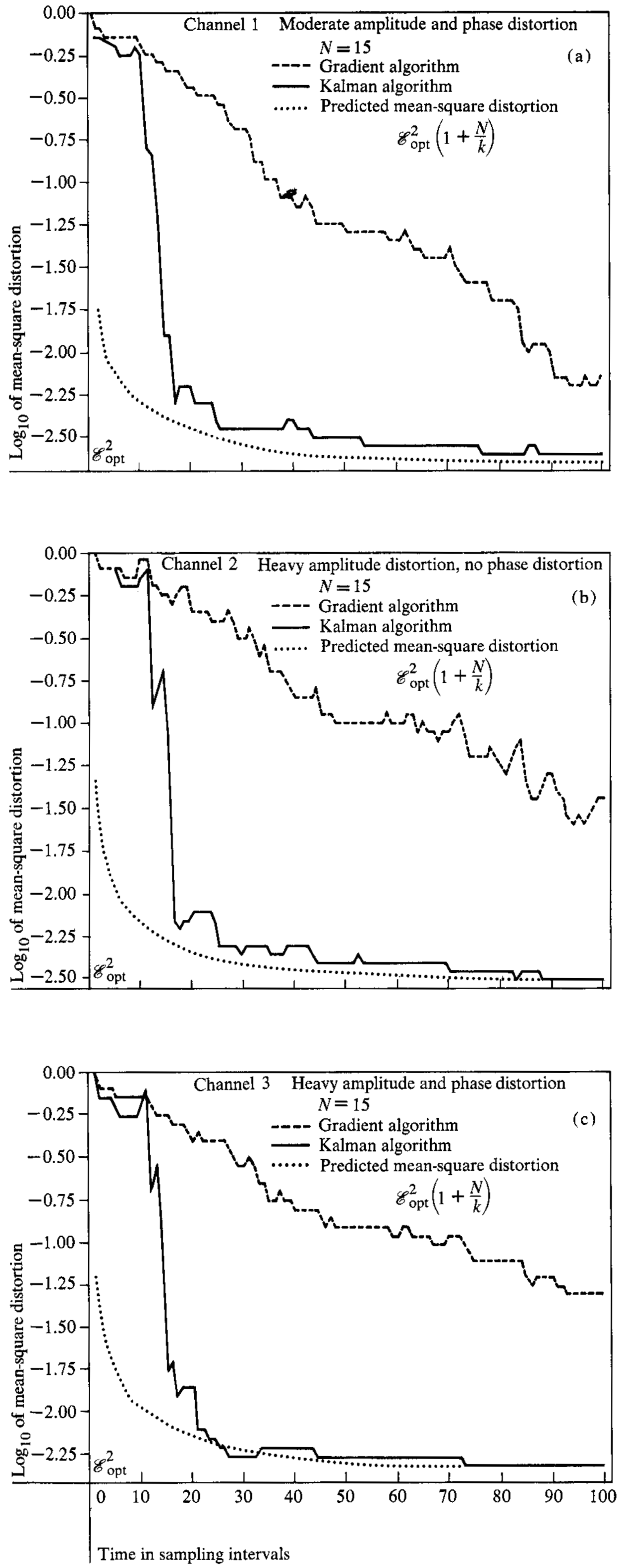

Figure 1 Results of the computer simulation of the Kalman-filter algorithm with three test channels: (a) Channel 1 – moderate amplitude and phase distortion; (b) Channel 2 – heavy amplitude distortion, no phase distortion; and (c) Channel 3 – heavy amplitude and phase distortion.

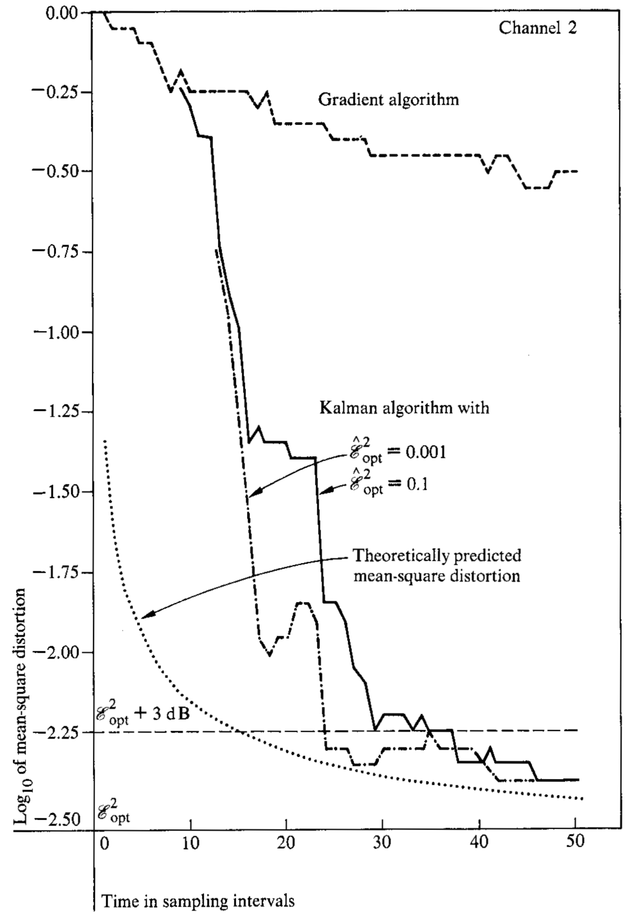

Figure 2 Results of simulation with Channel 2 to test the assumption of statistical independence of $\mathbf{c}_k$ and $\mathbf{u}_k$.

tained by simulation may be observed in spite of the various approximations that were made, among them being the statistical independence of $\mathbf{c}_k$ and $\mathbf{u}_k$.

To investigate the influence of this assumption, a computer run was made with five baud intervals introduced between sampling instants when tap-gain corrections are made. As a consequence the successive tap output signals are forced to be quasi-statistically independent of one another and the noise $e_{k\ opt}$ at the output of the optimal equalizer is white noise. It can be seen in Fig. 2 that, without counting the additional baud intervals, the speed of convergence is unchanged.

All simulations showed that convergence towards $\mathscr{E}^2_{opt} + 3$ dB was obtained within less than $2N$ steps. The equalizer may be switched into the decision-directed mode after about $2N$ sampling intervals. With an N of 15 and a transmission speed of 2400 bauds, the settling time of the equalizer is about 12 ms.

It would be interesting to compare these results with those of Chang [5] and Sha and Tang [11]. Their structures, however, use equally spaced, isolated test pulses during the training period. For a given distortion smaller than one, the Sha and Tang equalizer is optimally settled when it has received, on the average, four or five isolated pulses. Thus the settling time is at least $4N$ or $5N$ sampling intervals. When the **A** matrix is perfectly known and under noise-free conditions, the Chang equalizer requires only one training pulse. But when the **A** matrix is not precisely known and with a high signal-to-noise ratio, the settling time is again about $4N$ or $5N$ sampling intervals.

Assume now that the equalizer using the Kalman filter algorithm receives one isolated test pulse and that no noise is present. If the diagonal elements of the $\mathbf{P}_0$ matrix, or of the $\mathbf{S}_0$ matrix, are chosen large enough, Eq. (44) holds and the equalizer is optimally settled when one test pulse has been received, without prior knowledge about the channel characteristics. The algorithm then leads to the optimal speed of convergence, but it requires a larger amount of computation than other methods.

We also made a computer run with the $\mathbf{P}_k$ matrix reduced to its nine main diagonals, its other elements being zero. This resulted for Channel 3 in only a small degradation of the speed of convergence, since convergence below $\mathscr{E}^2_{\mathrm{opt}} + 3$ dB was obtained within about 60 sampling intervals, while halving the required amount of computation.

References

1. R. W. Lucky, "Automatic Equalization for Digital Communication," *Bell Syst. Tech. J.* **44,** 547 (1965).
2. R. W. Lucky and H. R. Rudin, "An Automatic Equalizer for General Purpose Communication Systems," *Bell Syst. Tech. J.* **45,** 225 (1966).
3. R. W. Lucky, J. Salz and E. J. Weldon, *Principles of Data Communication*, McGraw-Hill Book Co., Inc., New York, 1968.
4. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission," *Bell Syst. Tech. J.* **48,** 55 (1969).
5. R. W. Chang, "A New Equalizer Structure for Fast Start-Up Digital Communications," *Bell Syst. Tech. J.* **50,** 1969 (1971).
6. H. Kobayashi, "Application of Hestenes-Stiefel Algorithm To Channel Equalization," *Conference Record*, IEEE International Conference on Communications, Montreal, 1971.
7. C. Devieux and R. L. Pickholtz, "Adaptive Equalization With a Second Order Gradient Algorithm," *Proceedings of the Symposium on Computer Processing in Communications*, Polytechnic Press, Polytechnic Inst. of Brooklyn, New York, 1970.
8. G. Ungerboeck, "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communication," *IBM J. Res. Develop.* **16,** 546 (1972).
9. T. J. Schonfeld and M. Schwartz, "A Rapidly Converging First Order Training Algorithm for an Adaptive Equalizer," *IEEE Trans. Inf. Theory* **IT-17,** 431 (1971).
10. D. A. George, D. C. Coll, A. R. Kaye, and R. R. Bowen, "Channel Equalization for Data Transmission," *Proceedings of the 83rd Annual Meeting of the Engineering Institute of Canada*, Vancouver, B.C., September 1969, pp. 20-31.
11. R. T. Sha and D. T. Tang, "A New Class of Automatic Equalizers," *IBM J. Res. Develop.* **16,** 556 (1972).
12. R. E. Kalman, "A New Approach to Linear and Prediction Problems," *J. Basic Eng., Trans. ASME* **82,** 35 (1960).
13. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *J. Basic Eng., Trans. ASME* **95,** 107 (1961).
14. M. DeRusso, R. Ray and C. Close, *State Variables for Engineers*, John Wiley & Sons, Inc., New York, 1965.
15. R. E. Lawrence and H. Kaufman, "The Kalman Filter for the Equalization of a Digital Communications Channel," *IEEE Trans. Commun. Tech.* **COM-19,** 1137 (1971).

Received September 17, 1973

The author is located at the Centre d'Etudes et de Recherches, Compagnie IBM France, 06 La Gaude, Alpes Maritimes, France.