BLIND EQUALIZATION BY DIRECT EXAMINATION OF THE INPUT SEQUENCES

Fredrik Gustafsson

Dept. of Electrical Engineering

Linköping University, Sweden

Bo Wahlberg

S^3 – Automatic Control

Royal Institute of Technology, Sweden

Abstract

This paper presents a novel approach to blind equalization (deconvolution), which is based on direct examination of possible input sequences. In contrast to many other approaches, it does not rely on a model of the approximative inverse of the channel dynamics. To start with, the blind equalization identifiability problem for a noise-free finite impulse response channel model is investigated. A necessary condition for the input, which is algorithm independent, for blind deconvolution is derived. This condition is expressed in an information measure of the input sequence. A sufficient condition for identifiability is also inferred, which imposes a constraint on the true channel dynamics. The analysis motivates a recursive algorithm where all permissible input sequences are examined. The exact solution is guaranteed to be found as soon as it is possible. An upper bound on the compuational complexity of the algorithm is given. This algorithm is then generalized to cope with time-varying infinite impulse response channel models with additive noise. The estimated sequence is an arbitrary good approximation of the *maximum a posteriori* estimate. The proposed method is evaluated on a Rayleigh fading communication channel. The simulation results indicate fast convergence properties and good tracking abilities.

1

1 INTRODUCTION

1.1 Preliminaries

The problem of channel equalization is of considerable interest in data communication and related fields. Given a received output sequence, we want to determine (recover) the transmitted input sequence. In the case the channel is modeled as a known tapped-delay line (finite impulse response filter) and the input has a finite number of possible values, the Viterbi algorithm provides the optimal estimate of the input signal, see [12] and [21]. If the channel is unknown we have the problem of blind deconvolution, or equalization. Methods for blind deconvolution are discussed in the surveys [2, 3, 5, 6, 13]. The most common approach is to filter the output by an estimate of the inverse channel followed by some decision device. Limitations of this approach is discussed in Section 2. If the input contains a known training sequence, it is straightforward to estimate a finite impulse response (FIR) model of the channel. The input signal can then be recovered by applying the Viterbi algorithm to the estimated model. However, in many applications this approach cannot be used, for example when the channel is time varying. Another example is when the length of the transmitted data sequence is limited, so it is desirable to have as short training sequence as possible. In both cases, the FIR model needs to be updated continuously even after the training sequence.

The key idea of the current paper is as follows. Assume that the input signal belongs to a finite alphabet. Thus there are only a finite number of possible input sequences. By considering each of these as a training sequence a finite bank of FIR models is estimated, each associated with one input sequence. By associating a cost function to the estimates, namely the *a posteriori* probabilities of the input sequences, we can determine which one is the most likely and thus use this as an estimate. However, the number of possible input sequences increases exponentially with time. To limit the computationally complexity we propose an approximate algorithm, where only the most likely estimates are kept at each time instant. The properties of the proposed blind equalization scheme is evaluated by applying the method to a Rayleigh fading channel. The results are encouraging.

Consider for a moment the problem of system identification, for instance estimating a channel model from a known training sequence. There are two fundamental questions:

- Is it possible to identify the model from the actual observations?

- Will a particular estimator ever find the true model?

2

In the context of system identification, these properties are called *identifiability*, which relates to the model and the data, and *convergence*, which depends on the applied estimation method. It is well-known, see for instance [16], that for linear regression models, as will be used in this paper, a necessary and sufficient condition for identifiability is that the input is persistently exciting of order n (as will be defined in Section 3).

Reported analysis of blind equalizers deals with the convergence properties of specific methods, and almost nothing seems to be known about identifiability. That is, under what circumstances is it possible to recover the input sequence? Obviously, this is a general property which is independent of the actual blind equalizer which is going to be applied. One result in this direction is reported in [22], although their approach assumes a specific equalizer. We will provide an answer to this question by showing that a necessary condition is a persistently exciting input sequence of order $2n - 1$. By also require a certain condition of the channel impulse response, a sufficient condition for identifiability is obtained.

1.2 Problem Formulation

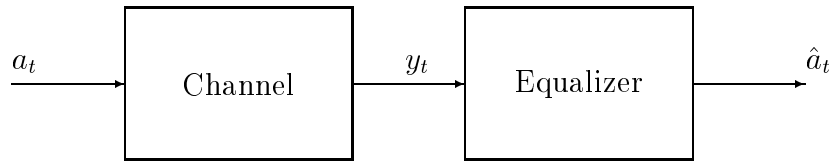Consider the simplified but yet realistic digital communication system illustrated in Figure 1.

Figure 1: A digital communication system

The transmitter generates a sequence $\{a_t\}$ of encoded information belonging to a finite alphabet, which is sent over a channel before it reaches the receiver as a sequence $\{y_t\}$. The channel can be accurately modeled as a linear system, which on physical grounds often can be approximated by a a non-minimum phase FIR filter. Due to the non-ideal channel, the problem of so called intersymbol inference implies that the sent symbol a_t cannot be reconstructed from y_t alone. Thus, there is a need for equalizing the channel distortion. This is done in the second block in Figure 1.

The mathematical relations and notations are as follows. The output of the channel is given by

$$y_t = b_1 a_t + b_2 a_{t-1} \cdots b_n a_{t-n+1}$$

$$= B(q)a_t \tag{1}$$

where $B(q) = b_1 + b_2 q^{-1} + ... + b_n q^{-(n-1)}$. Here q^{-1} denotes the backward shift operator, $q^{-1}a_t = a_{t-1}$. We will use the regression form of (1) as well,

$$y_t = \varphi_t^T b, \tag{2}$$

where the $n \times 1$ vector b contains the unknown parameters and the regression vector is $\varphi_t = (a_t, a_{t-1}, ..., a_{t-n+1})^T$. The outputs are collected into a $(t - n + 1) \times 1$ vector $Y_t = (y_n, ..., y_t)^T$ and the t inputs into the $t \times 1$ vector $A_t = (a_1, ..., a_t)^T$. We will sometimes refer to A_t and Y_t as the sequences $\{a_k\}$ and $\{y_k\}$, respectively. One can now rewrite the $t - n + 1$ equations in (2) in matrix form,

$$Y_t = \Phi_{t,n} b. \tag{3}$$

Here $\Phi_{t,n}$ is the Toeplitz matrix with n columns containing the input sequence A_t,

$$\Phi_{t,n} = \begin{pmatrix} a_n & a_{n-1} & \cdots & a_1 \\ a_{n+1} & \cdots & & a_2 \\ \vdots & & & \vdots \\ a_t & \vdots & & a_{t-n+1} \end{pmatrix}. \tag{4}$$

The blind equalization problem can now be stated as solving the, possibly perturbed, constrained non-linear equation system $Y_t = \Phi_{t,n} b$ with respect to b and $\Phi_{t,n}$. This seems like an underdetermined problem, even in the noise-free case, with more unknown parameters than equations. For instance, if $\overline{A}_t = cA_t$ is a permissible input sequence, where c is a constant, then $Y_t = c\Phi_{t,n} \cdot 1/c\, b$ is another solution. Nevertheless, it will be shown that equation (3) can be solved under quite general conditions due to the finite alphabet property of a_t. Furthermore, it is shown that all solutions can be written cA_t. This observation motivates the following definition of identifiability for the blind equalization problem.

Definition 1 *The input sequence and the channel model are said to be identifiable from observations of the output sequence if all solutions to equation (3) can be written cA_t and $1/c \cdot b$ for some c, where A_t is the true input sequence and b the true channel model.*

This symmetry property of the problem does not cause any problems in practice, since the information is encoded in differential form so it is a_t/a_{t-1} that contains the information rather than a_t itself. Also notice that only constants c such that ca_t belongs to the alphabet are possible.

1.3   Outline Of The Paper

In Section 2, we will give a short review of methods for blind equalization and related convergence properties. The identifiability issue of the blind equalization problem without assuming any specific structure of the equalizer is examined in in Section 3. The obtained result is then used to derive a novel blind equalization approach, which is presented in Section 4. A simulation evaluation of the method is undertaken in Section 5. Section 6 concludes the paper. Parts of the results in the current paper have been presented in the conference papers [9, 10, 11].

2   BLIND EQUALIZATION BY INVERSE MODEL FILTERING

A standard approach to equalizing is to try to find an explicit model of the inverse channel, and then recover the input using a simple static decision device as shown in Figure 2.
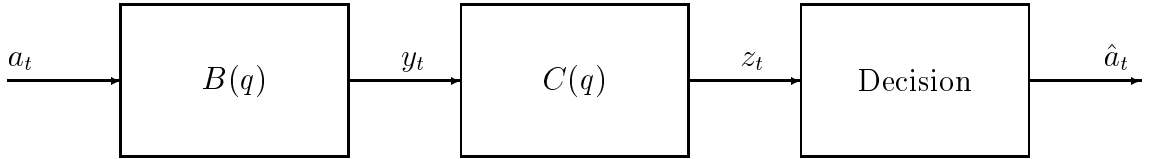
Figure 2: Equalizing by using an inverse filter

Assume that $B(q)$ is a non-minimum phase filter, and that the inverse channel model filter is specified as a FIR filter

$$z_t = C(q;c)y_t = c_1 y_t + c_2 y_{t-1} + ... + c_m y_{t-m+1}. \tag{5}$$

Define $H(q)$ as the combined channel–equalizer, $H(q) = B(q)C(q;c)$, which ideally should be equal to 1. However, for nonminimum phase systems one has to, at best, accept a time delay $H(q) = q^{-k}$, for some unknown k.

The classical way of constructing equalizers $C(q;c)$ is to use a known training sequence to estimate the parameters c and then apply a simple decision device on its output in the transmitting mode, see Figure 2. If the channel is time varying or the training sequence is too short to obtain a good estimate of the channel inverse one can try to continue adjusting the parameters c in the inverse filter $C(q;c)$ even after the training sequence. This is the

problem of blind equalization. The key question here is if the blind equalizer will converge to a value corresponding to an open-eye equalizer, that is $\hat{a}_t = a_t$ in Figure 2. This is usually called the admissibility problem. Admissibility is a weaker condition than identifiability, as defined in Section 1, since the actual parameter estimates are not considered in this context (because there are no "true" values). In blind equalization, the filter $C(q;c)$ is adjusted to resemble the inverse channel $B(q)$ by minimizing some loss function in z_t. This can for example be done using stochastic gradient algorithms resembling the least mean squares (LMS) method, see e.g. [17, 20, 7, 2, 4, 19]. It is clear from examples that the aforementioned algorithms sometimes fail to converge to an open-eye condition as shown in [13], or they may even diverge. Nevertheless, some convergence results are known. They all apply under the assumption that the equalizer is infinite dimensional. Then under certain conditions the overall impulse response will converge to $\pm q^{-k}$ for some time delay k.

The so called *decision directed* algorithm is shown to converge in [20] if the initial parameter setting is such that the overall impulse response satisfies $\sum_{k=1}^{\infty} |h_k| < |h_0|$. If the input is restricted to ± 1, then $\hat{a}_t = \text{sign}(z_t) = \text{sign}(a_t) = a_t$, so this assumption corresponds to open-eye initialization. That open-eye initialization is generally sufficient for convergence is proved in [18].

The *modulus restoral* algorithm is shown to converge for an appropriate initial setting in [7]. In [6], the convergence to the desired over-all impulse response is proven if the equalizer is infinite dimensional (that is, $m = \infty$ in (5)).

The conclusions of this discussion are as follows. The advantage with the inverse filtering approach is that the algorithms are simple to implement and computationally very fast. On the other hand, there are a number of basic disadvantages:

- The choice of loss function is rather *ad hoc*.

- All suggested loss functions have undesired local minima.

- The inverse channel, which is often is an infinite impulse response model, must be approximated by a FIR filter.

- The over-all impulse response contains an unknown delay.

- Not even asymptotic convergence of the parameter vector b can be expected in the noise-free case of (14), since a constant step size is used in proposed gradient schemes.

However, these drawbacks are not inherent in the problem formulation (14) but depend on the inverse filtering approach. In the next sections, it is shown that these problems all can be overcome.

6

3   IDENTIFIABILITY

We will here investigate the important question of identifiability of the parameters for a noise-free FIR channel. Recall the problem formulation (3) of the noise-free FIR channel model, i.e. given the measurements Y_t solve the bilinear equation system,

$$Y_t = \Phi_{t,n} b. \tag{6}$$

with respect to b and $\Phi_{t,n}$. Here b is the n unknown FIR parameters in the channel model and $\Phi_{t,n}$ is a Toeplitz matrix with n columns constructed by the input sequence of length t. The following simple example clearly illustrates the problem at hand.

Example 1 *The bilinear equation system*

$$\begin{pmatrix} -0.65 \\ -1.35 \\ 0.65 \\ -0.65 \end{pmatrix} = \begin{pmatrix} a_2 & a_1 \\ a_3 & a_2 \\ a_4 & a_3 \\ a_5 & a_4 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

has the unique solution $b = (1,\ 0.35)^T$, $A_t = (1, -1, -1, 1, -1)^T$, while

$$\begin{pmatrix} -0.65 \\ -1.35 \\ 0.65 \\ 1.35 \end{pmatrix} = \begin{pmatrix} a_2 & a_1 \\ a_3 & a_2 \\ a_4 & a_3 \\ a_5 & a_4 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

has the two solutions $b = (1,\ 0.35)^T$, $A_t = (1, -1, -1, 1, 1)^T$ and $b = (0.35,\ -1)^T$, $A_t = (1, 1, -1, -1, 1)^T$.

The explanation turns out to be in terms of an information measure of the input sequence.

    If the input sequence was known, then the FIR parameters can be computed uniquely if and only if $\Phi_{t,n}$ has full column rank. Then, we have

$$b = (\Phi_{t,n}^T \Phi_{t,n})^{-1} \Phi_{t,n}^T Y_t. \tag{7}$$

This is a classical result in system identification *c.f.* [16] which has motivated the following definition of input excitation.

Definition 2 *The sequence $\{a_t\}$ is persistently exciting (P.E.) of order k at time t if $\Phi_{t,k}$ has full column rank, where the Toeplitz matrix $\Phi_{t,k}$ is defined in (4).*

Note that the number of columns in $\Phi_{t,k}$ is a free parameter here. In the current application where the input sequence belongs to a finite set, this basically means that the input sequence may not be periodical with a period shorter than k. This is certainly true in practice, where the input contains information and is encoded to resemble white noise.

One can argue that it is more logical to study the identifiability of the input sequence directly since the channel model is not of interest in itself. However, it is clear from (7) that if the input sequence is P.E. of order n and known, then the channel model can be calculated. Conversely, if the channel model is known then the input can always be calculated. Thus, the two problems of identifiability of the input and the channel model can be considered as equivalent if the input is P.E. of order n (which is in accordance with Definition 1).

Let $\bar{A}_t$ and $\bar{b}$ denote the true values. Assume now that there exists another solution A_t and b such that

$$Y_t = \overline{\Phi}_{t,n}\,\bar{b} = \Phi_{t,n}b. \tag{8}$$

Lemma A.2 proves that also $\Phi_{t,n}$ must have full column rank. Hence, $b = \Phi_{t,n}^{\dagger}\overline{\Phi}_{t,n}\bar{b}$ where $\Phi_{t,n}^{\dagger} = (\Phi_{t,n}^{T}\Phi_{t,n})^{-1}\Phi_{t,n}^{T}$ denotes the pseudo-inverse. Equation (8) then implies

$$\left(\Phi_{t,n}\Phi_{t,n}^{\dagger}\overline{\Phi}_{t,n} - \overline{\Phi}_{t,n}\right)\bar{b} \triangleq \Gamma\bar{b} = 0. \tag{9}$$

The question of identifiability is now equivalent to prove that $\Gamma\bar{b} = 0$ implies $\Phi = c\overline{\Phi}$, which is a trivial solution to $\Gamma = 0$.

We will proceed in two steps. First it is shown that $\Gamma \neq 0$ if and only if $\bar{A}_t$ is P.E. of order $2n-1$, which means that P.E. of order $2n-1$ is a necessary condition for identifiability. Then a condition on $\bar{b}$ is derived guaranteeing that $\Gamma\bar{b} \neq 0$ for all possible $A \neq 0$ that the finite alphabet can generate.

Theorem 3.1 *Consider the two sequences $\bar{A}_t$ and A_t, not necessarily belonging to a finite alphabet. The equation $\Gamma = \Phi_{t,n}\Phi_{t,n}^{\dagger}\overline{\Phi}_{t,n} - \overline{\Phi}_{t,n} = 0$ with $\Phi_{t,n}$ defined in (4) has only the trivial solution $\Phi = c\overline{\Phi}$ if and only if $\bar{A}_t$ is P.E. of order $2n-1$. Thus, a necessary condition for identifiability is P.E. of order $2n-1$.*

Proof: Define the $n \times n$ matrix $S \triangleq \Phi_{t,n}^{\dagger}\overline{\Phi}_{t,n} = (s_1, ..., s_n)$, where s_i is the ith column of S. $\Gamma = 0$ can now be expressed as

$$\Phi S = \overline{\Phi},$$

8

or

$$\begin{pmatrix} a_n & \cdots & a_1 \\ \vdots & & \vdots \\ a_t & \cdots & a_{t-n+1} \end{pmatrix} (s_1, \cdots, s_n) = \begin{pmatrix} \overline{a}_n & \cdots & \overline{a}_1 \\ \vdots & & \vdots \\ \overline{a}_t & \cdots & \overline{a}_{t-n+1} \end{pmatrix}. \qquad (10)$$

Eliminating $\overline{a}_k$ for $k = n, n+1, .., t-n+1$, using the Toeplitz structure, and solving (10) for s_1 and s_n gives the following system of equations:

$$\begin{pmatrix} a_{2n-1} & \cdots & a_n & a_n & \cdots & a_1 \\ a_{2n} & \cdots & a_{n+1} & a_{n+1} & \cdots & a_2 \\ \vdots & & & & & \vdots \\ a_t & \cdots & a_{t-n+1} & a_{t-n+1} & \cdots & -a_{t-2n+2} \end{pmatrix} \begin{pmatrix} -s_n \\ s_1 \end{pmatrix} \triangleq F_t x = 0. \quad (11)$$

Note that F_t is identical to $\Phi_{t,2n-1}$ except for the middle column which is repeated twice in F_t.

Firstly, assume P.E. of order $2n-1$. Then rank F_t = rank $\Phi_{t,2n-1} = 2n-1$. But F_t contains $2n$ columns so there exists exactly one non-zero linearly independent solution x, which is trivially seen to be

$$x = c(0, \cdots 0, -1, 1, 0, \cdots, 0)^T, \qquad (12)$$

where c is a (possibly complex) constant. Hence $s_1 = ce_1$ and $s_n = ce_n$ where e_i is the ith column in the identity matrix I. Continuing solving (10) for s_1 and s_i immediately gives $s_i = ce_i$ and we conclude that

$$S = cI$$

and thus $\Phi_{t,n} = 1/c\, \overline{\Phi}_{t,n}$ so $b = c\overline{b}$ and $A_t = 1/c\, \overline{A}_t$.

Secondly, assume rank $\Phi_{t,2n-1} < 2n - 1$. Then there exists at least one more solution of $F_t x$, which is linearly independent of (12). Again, we can continue solving (10) for s_i, and we get a solution $\Phi_{t,n} S = \overline{\Phi}_{t,n}$, where $S \neq cI$. This proves non-identifiability in the case of P.E. less than $2n - 1$. $\qquad \square$

Theorem 3.1 shows that P.E. of order $2n - 1$ is a necessary condition in all blind deconvolution problems, even if the input is not in a finite alphabet.

Example 2 *Theorem 3.1 now explains the result in Example 1. The first input sequence in Example 1 is P.E. of order 3 = $2n - 1$, while the two solutions in the second case are only P.E. of order 2. This explains why there are two solutions in the second case.*

Next we state a sufficient condition for blind deconvolution.

9

Theorem 3.2 *A sufficient condition for identifiability if the input belongs to the finite alphabet $a_t \in \{\pm 1, \pm 3, .., \pm(M-1)\}$ is that the input sequence is P.E. of order $2n-1$ and that the FIR coefficients $\{b_i\}$ are linearly independent with respect to coefficients in the set*

$$Z_n = \left\{ 0, \pm 1, \pm 2, ..., \pm 2\,(M-1)^{2n+1}\,n^{n/2}(t_0 - n + 1)^n \right\}. \qquad (13)$$

Here t_0 is defined as the first time instant $\bar{A}_t$ is P.E. of order n.

Proof: Theorem 3.1 shows that $\Gamma \neq 0$ if A_t P.E. of order $2n - 1$. Now, Lemma A.1 proves that for a certain integer K the elements of $K\Gamma$ belong to Z_n. Since the coefficients of b are supposed to be linearly independent with respect to elements in Z_n it follows that $\Gamma b \neq 0$ which proves identifiability. □

The condition on b means that there must not exist relationships like $3b_1 + 7b_3 = 0$. This restriction is not too severe. Losely speaking it is satisfied with probability one. Even if it is not satisfied, simulations have shown that A_t is still identifiable which can be explained as follows. Assume that b is linearly dependent over Z_n. Then one of the rows of Γ may be orthogonal to b although it is not likely. Equation (10) still holds except for some rows that must be deleted. The key point is that an equation like (11) still holds if there are enough rows in A that are not orthogonal to b and the conclusion $S = cI$ remains. The problem is that the complicated interaction of $\overline{\Phi}_{t,n}$ and $\bar{b}$ makes it difficult to give any necessary conditions on A_t in the theorem. Thus, this sufficient condition is rather conservative.

Example 3 *The first input sequence in Example 1 is P.E. of order $2 = n$ at time $t_0 = 3$. The integer set is thus with $M = 2$, $Z_n = \{0, \pm 1, .., \pm 2 \cdot 1 \cdot 2^1 \cdot (3 - 2 + 1)^2\} = \{0, \pm 1, .., \pm 16\}$. The smallest integer solution to $m \cdot 1 + n \cdot 0.35 = 0$ is $7 \cdot 1 - 20 \cdot 0.35 = 0$. Since 20 does not belong to Z_n, we could have concluded directly without any computations that there is only one solution to the first problem. On the other hand, if the true channel parameters were $b = (1,\, 0.3)^T$ the sufficient condition is not satisfied since $3 \cdot 1 - 10 \cdot 0.3 = 0$. However, the solution is still unique which shows that the sufficient condition is conservative.*

Another approach to identifiability is examined in [22]. They note that also the output belongs to a finite alphabet (though quite large). This fact is used to come up with an algebraic solution to the non-linear equation system (14). The idea is to define equivalent output measurement sets. Identifiability conditions, which are based on the specific algorithm used, are also given. However, these conditions are not easy to relate to standard identifiability assumptions.

4 A DIRECT EQUALIZER

4.1 The Noise-Free Case

We will first present an algorithm to solve the non-linear system of equations

$$Y_t = \Phi_{t,n}b, \tag{14}$$

consistent with the identifiability result in Section 3, which gives the correct values of A_t and b as fast as possible. It will be extended in the next subsections to cover more realistic channel models. The algorithm is graphically illustrated in Figure 3. Notice first that if $\Phi_{t,n}$ is one solution to (14), then it holds that $Y_t = \Phi_{t,n}\Phi_{t,n}^\dagger Y_t$.
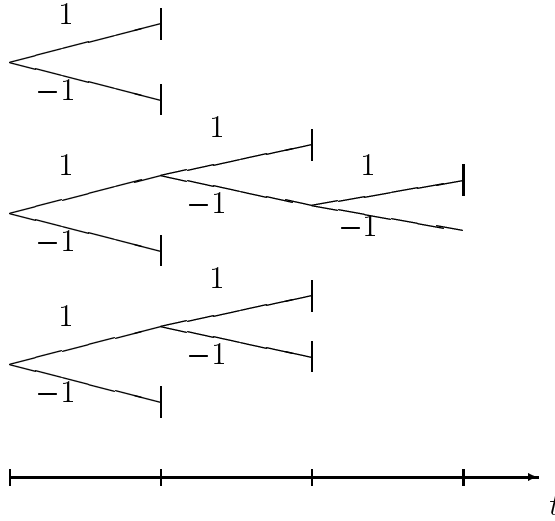

Figure 3: Recursive solution of $Y_t = \Phi_{t,n}b$ for $M = 2$

Algorithm 1 *A recursive solution to $Y_t = \Phi_{t,n}b$ is given by the following scheme:*

1. *At time t, there are L_t permissible sequences A_t^i, $i = 1, 2, ..L_t$ that satisfy $Y_t = \Phi_{t,n}^i(\Phi_{t,n}^i)^\dagger Y_t$. The corresponding channel models are given by $b^i = (\Phi_{t,n}^i)^\dagger Y_t$.*

2. *Let each of the permissible sequences split into M sequences where M is the size of the alphabet of a_t. These are the* a priori *permissible sequences at time $t + 1$.*

11

3. *Let the* a posteriori *permissible sequences at time $t + 1$ be those which satisfy*

$$Y_{t+1} = \Phi^i_{t+1,n}(\Phi^i_{t+1,n})^\dagger Y_{t+1}. \tag{15}$$

If A^i_t is P.E. of order n, this condition can be replaced by

$$y_{t+1} = (\varphi^i_{t+1})^T b^i. \tag{16}$$

4. *Repeat the above steps.*

We remark that if a sequence is not P.E. of order n, then the pseudo-inverse $(\Phi^i_{t,n})^\dagger$ cannot be computed as $(\Phi^T_{t,n}\Phi_{t,n})^{-1}\Phi^T_{t,n}$, but it can always be computed by the singular value decomposition, see [8].

Consider now the sequences that are P.E. of order n at time t, so that b^i is uniquely determined. It is then clear that each survivor at time t will have at most one survivor at time $t + 1$, since the relation $a^i_{t+1} = 1/b^i_1(y_{t+1} - b^i_2 a_t - ...b^i_n a_{t-n+2})$ defines a^i_{t+1} uniquely and gives at most one permissible value in the finite alphabet of a_t. This means that $\{L_t\}$ will be a non-increasing sequence if only sequences which are P.E. of order n are considered. Lemma A.2 in the appendix strengthens this result, since it claims that if the true sequence is P.E. of order n then all other sequences that satisfies (14) must be P.E. of order n as well. The conclusion is that if the true input sequence is P.E. of order n at time t_0, then there exists an upper bound M^{t_0} (M is the size of the alphabet) on the number of sequences that have to be examined in the algorithm. Thus, in some sense, there is no exponential complexity in the problem as could be expected as a consequence of the exponential increase of input sequences.

Theorem 4.1 *Consider the channel description (14) and assume that the input sequence $\{a_t\}$ is P.E. of order n at time t_0. If t_0 is known a priori then the number of sequences that have to be considered in Algorithm 1 is bounded by M^{t_0}, where M is the size of the alphabet of a_t. Furthermore, if t_0 is unknown but the parameter vector b is linearly independent over Z_n, as defined in Theorem 3.2, then the number of sequences is still bounded by M^{t_0},*

Proof: The first part follows immediately from the discussion above. If b is linearly independent over Z_n, then Lemma A.2 gives that all $\Phi_{t_0,n}$ that satisfy (14) for some b must have rank n. According to the definition of P.E., all permissible input sequences are P.E. of order n and the discussion above still holds. □

In practice it is not realistic to assume that t_0 is known since the input is stochastic. The first statement in the theorem is still useful for designing

recursive algorithms which will work with a high probability, since the number of sequences that can be examined must always be limited. This can be achieved by assuming a large enough t_0.

We have from Theorems 3.1 and 3.2 that identifiability is determined by the first time instant when the input sequence is P.E. of order $2n - 1$. The second statement of Theorem 4.1 shows that the complexity of the problem is determined by the time instant when the input sequence is P.E. of order n. Thus, the goal of the encoder should be to generate an input sequence that becomes P.E. as quickly as possible.

4.2   A Heuristical Motivation

The algorithm above is easy to explain but the channel model is not very realistic. We will now motivate how it should be extended to cover a possibly time-varying channel disturbed by noise. First, consider the model (14) with additive noise e_t collected in the vector E_t,

$$Y_t = \Phi_{t,n} b + E_t.$$

The algorithm above can still be used if the condition (15) is replaced by

$$||Y_{t+1} - \Phi^i_{t+1,n} b^i|| < \epsilon \qquad (17)$$

for some norm and threshold ϵ. This is intuitively appealing and the problem is how to choose the norm and threshold in an optimal way and to minimize (17) in an efficient and recursive way. The choices of norm and threshold are of course dependent on the noise but also on the uncertainty in b^i caused by the noise.

If the channel is time-varying so $y_t = \varphi_t^T b_t + e_t$, the estimate of b_t must be updated recursively in some way and used in (17) instead of b^i.

We will now derive this heuristically motivated algorithm in a mathematical way by recursively computing the exact *a posteriori* distribution of A_t and then using a search scheme to obtain an implementable algorithm like Algorithm 1.

4.3   Optimal Estimates

Consider the channel model in Figure 4.

The output of the channel is given by

$$y_t = -d_1(t)y_{t-1} - \cdots - d_{n_d}(t)y_{t-n_d} + b_1(t)a_t + \cdots + b_{n_b}(t)a_{t-n_b+1} + e_t.$$
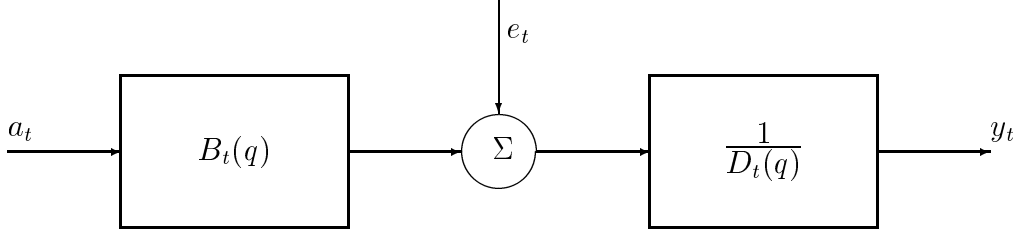
13

Figure 4: Time-varying channel model with noise

This model is equivalent to (14) if the parameters are constant, $n_d = 0$ and $e_t \equiv 0$. The auto-regressive part is introduced to be as general as possible without increasing the complexity. We will assume that the parameter variation can be modeled as a random walk so a total state space linear regression model is

$$
\begin{aligned}
\theta_{t+1} &= \theta_t + v_t \\
y_t &= \varphi_t^T \theta_t + e_t.
\end{aligned}
\tag{18}
$$

Here,

$$
\begin{aligned}
\varphi_t &= (-y_{t-1}, \cdots, -y_{t-n_d}, a_t, \cdots, a_{t-n_b+1})^T &\tag{19} \\
\theta_t &= (d_1(t), \cdots, d_{n_d}(t), b_1(t), \cdots, b_{n_b}(t))^T. &\tag{20}
\end{aligned}
$$

We assume that v_t and e_t are uncorrelated white Gaussian noises with covariance matrices Q_t and Λ_t, respectively. Assume we have an arbitrary enumeration of the M^t sequences A_t, say A_t^i, $i = 1, 2..., M^t$, and denote the corresponding regression vector φ_t^i. The *a posteriori* probabilities of a sequence A_t^i, given the observations Y_t, is given by the following theorem.

Theorem 4.2 *Consider the model (18). The* a posteriori *probability of A_t^i is given by*

$$
\begin{aligned}
p(A_t^i|Y_t) &= C_t p(a_t^i|A_{t-1}^i)\gamma(y_t - (\varphi_t^i)^T\hat{\theta}_t^i, (\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t) \cdot p(A_{t-1}^i|Y_{t-1}) &\tag{21} \\
&= C p(A_t^i) \prod_{k=1}^{t} \gamma(y_k - (\varphi_k^i)^T\hat{\theta}_k^i, (\varphi_k^i)^T P_k^i \varphi_k^i + \Lambda_t) &\tag{22}
\end{aligned}
$$

where C and C_t are constants given by the condition $\sum_{i=1}^{M^t} p(A_t^i|Y_t) = 1$. Here $\gamma(x-\mu, P)$ denotes the value of the Gaussian probability density function with

14

mean μ and covariance P evaluated in x. The parameter estimate $\hat{\theta}_t^i$ and its covariance matrix P_t^i are computed recursively by

$$\hat{\theta}_{t+1}^i \;=\; \theta_t^i + P_t^i \varphi_t^i \left((\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t \right)^{-1} \left(y_t - (\varphi_t^i)^T \hat{\theta}_t^i \right) \qquad (23)$$

$$P_{t+1}^i \;=\; P_t^i + P_t^i \varphi_t^i \left((\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t \right)^{-1} (\varphi_t^i)^T P_t^i + Q_t \qquad (24)$$

with the initial values θ_0 and P_0.

Proof: Repeated use of Bayes' rule gives

$$p(A_t^i|Y_t) \;=\; \frac{p(Y_t, A_t^i)}{p(Y_t)} = \frac{p(A_t^i)}{p(Y_t)} p(Y_t|A_t^i). \qquad (25)$$

Now $1/p(Y_t)$ can be regarded as a constant $\overline{C}_t$. Bayes' rule gives,

$$\begin{aligned}
p(A_t^i|Y_t) \;&=\; \overline{C}_t p(A_t^i) p(y_t, Y_{t-1}|A_t^i) \\
&=\; \overline{C}_t p(A_t^i) p(Y_{t-1}|A_t^i) p(y_t|A_t^i, Y_{t-1}) \\
&=\; \overline{C}_t p(a_t^i|A_{t-1}^i) p(A_{t-1}^i) p(Y_{t-1}|A_{t-1}^i) p(y_t|\varphi_1^i, ..., \varphi_t^i, Y_{t-1}) \\
&=\; C_t p(A_{t-1}^i|Y_{t-1}) p(a_t^i|A_{t-1}^i) \gamma(y_t - (\varphi_t^i)^T \hat{\theta}_t^i, (\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t)
\end{aligned}$$

which is (21) and (22) follows by expanding the recursion. The last constant C_t is equal to $\overline{C}_t/\overline{C}_{t-1}$. The last equality is a consequence of a well-known result from linear filtering theory that the prediction error of (18) is Gaussian if φ_t is known. See for instance [1]. The equations (23) and (24) are just the Kalman filter equations for the state space model (18). $\qquad\square$

A logical estimate of the input sequence is the *maximum a posteriori* (MAP) estimate,

$$\hat{A}_t^{MAP} = \arg\max_i p(A_t^i|Y_t). \qquad (26)$$

The *maximum likelihood* (ML) estimate,

$$\hat{A}_t^{ML} = \arg\max_i p(Y_t|A_t^i). \qquad (27)$$

is closely related to the MAP estimate as seen from (25). It can be computed from the MAP estimate by letting the prior being non-informative, that is $p(A_t) = 1/M^t$.

The prior information $p(a_t^i|A_{t-1}^i)$ in (21) can be used to decode the information by rejecting "impossible" sequences, thus eliminating the need of a separate decoder. It can also be used to incorporate a training sequence in a very natural way, by letting $p(A_t^{train}) = 1$. However, most often the inputs

15

are considered as independent variables so the ML estimate is equivalent to the MAP estimate.

Theoretically, Theorem 4.2 holds in the noiseless case as well, especially for the channel model (14). What happens is that the *a posteriori* probabilities become either zero, if $y_t - (\varphi_t^i)^T \hat{\theta}_t^i \neq 0$ for some t, or Dirac impulses, if $y_t - (\varphi_t^i)^T \hat{\theta}_t^i = 0$ for all t. A consequence is the following:

Corollary 4.3 *Consider the noise-free FIR channel model (14) and assume that the same conditions as in Theorems 3.1 and 3.2 hold. Then the MAP estimate (26) yields the true sequence (except for a possible scaling factor c) if the true sequence is P.E. of order $2n-1$. Furthermore, if the true sequence is P.E. of order n at time t_0, then the number of filters needed to compute the MAP estimate is bounded by M^{t_0}.*

It is reasonable to believe that the result still holds if the noise level is small enough.

4.4   Basic Limitations

The MAP estimate (26) completely eliminates the disadvantages of equalizing by an inverse filter which are mentioned in Section 2. However, it introduces some new problems:

- The computational complexity is exponential increasing, since it requires Mt Kalman filters at time t.

- It is not guaranteed that $\hat{A}_t^{MAP}$ resembles $\hat{A}_{t-1}^{MAP}$, although it is very likely. Thus, a new measurement can alter the entire estimated sequence.

In the next section, we will present an approximative MAP estimate that contains a fixed number of filters. It turns out that the second disadvantage disappears as a consequence of the approximation.

4.5   A Local Search Algorithm

We will now give a recursively implementable approximation of the MAP estimate. It contains a fixed number, K, of filters. In words, only sequences which have turned out to be likely are considered. The others are rejected.

Algorithm 2 *Assume there are K sequences A_t^i given at time t and that their relative* a posteriori *probabilities $p(A_t^i|Y_t)$ have been computed.*

16

1. *Compute $p(A_{t+1}^i|Y_{t+1})$ by using Theorem 4.2 for the KM sequences obtained by considering all expansions of the sequences at time t.*

2. *Reject all but the K most probable sequences — that is, those which have the largest $p(A_{t+1}^i|Y_{t+1})$.*

3. *Repeat from step 1.*

We conclude from Corollary 4.3 that this algorithm is asymptotically optimal in the measurement noise for a time-invariant FIR model if $K \geq M^{t_0}$, that is, if the number of parallel filters are chosen to be large enough.

This remarkably simple algorithm works very well in simulations as will be demonstrated in Section 5. The second step resembles the Viterbi algorithm, see [21], because unlikely sequences are rejected. In the Viterbi algorithm the channel model b is assumed to be known and the most probable sequence is saved for every possible combination of the last n inputs, so there are totally M^n sequences under consideration. All other sequences are rejected, because the MAP estimate of the input sequence is guaranteed to be among these M^n sequences. Here, where b is unknown, the uncertainty in the estimate $\hat{b}^i$ is taken consideration of automatically and the sequences that are not rejected are not restricted to be different in the last n inputs (because this is no longer optimal).

5   SIMULATION RESULTS

In this section we will examine how Algorithm 2 performs in the case of a Rayleigh fading communication channel. Rayleigh fading is an important problem in mobile communication. The motion of the receiver causes a time-varying channel characteristics. The Rayleigh fading channel is simulated using the following premises: The frequency of the carrier wave is 900 MHz and the baseband sampling frequency is 25 kHz. The receiver is moving with the velocity 83 km/h so the maximum doppler frequency can be shown to be approximately 70 Hz. A channel with two complex time-varying taps, corresponding to this maximum doppler frequency, will be used.[1] An example of a tap is shown in Figure 5. For more details and a thoroughly treatment of fading in mobile communication, see [15].

The input is assumed to belong to the finite alphabet $\{-1, +1, -i, +i\}$, with equal probability for each symbol. An input sequence of length 100

---

[1]The taps are simulated by filtering white Gaussian noise with unit variance by a second order resonance filter, with the resonance frequency equal to 70/25000 Hz, followed by a 7'th order Butterworth low-pass filter with cut-off frequency $\pi/2 \cdot 70/25000$.
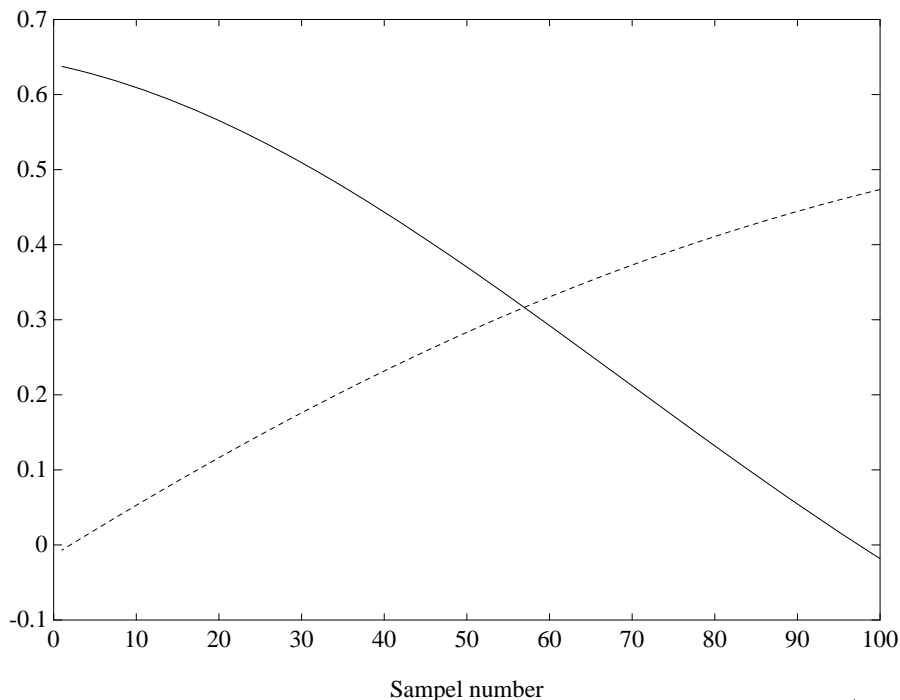
Figure 5: Example of a complex tap in a Rayleigh fading channel (real and imaginary parts)

is filtered through a simulated Rayleigh fading channel and Gaussian noise with variance λ is added. 100 different realizations of the input sequence, the noise sequence and the channel is used throughout all simulations. The magnitude of the noise was changed so the noise variance λ is 0.001, 0.002, 0.005, 0.01, 0.02, 0.05 and 0.1, respectively.

To obtain a feeling for the problems involved in estimating the input sequence we begin with computing an upper bound of the performance of any algorithm. This is here done by assuming that the time-varying channel really is known to the receiver and using the Viterbi algorithm, which is in this case optimal in the maximum likelihood sense. The estimated bit error probability is shown in Figure 6, where the first value, corresponding to the smallest λ, is zero and is not shown.

The input sequence is then estimated by Algorithm 2 with a number of parallel filters. The true measurement noise variance was used and the variance (Q) of the random walk was chosen to 0.01. Before showing the result, let us comment on the number of filters. From Theorem 4.1 we have an upper bound on this number as M^{t_0}. This was derived for the time-invariant noise-free case, but it should provide a reasonable guideline here as well. Assuming that the input is P.E. of order n already at time $t_0 = n$ this upper bound implies $M^{2n-1} = 4^3 = 64$ parallel filters. We will compare this
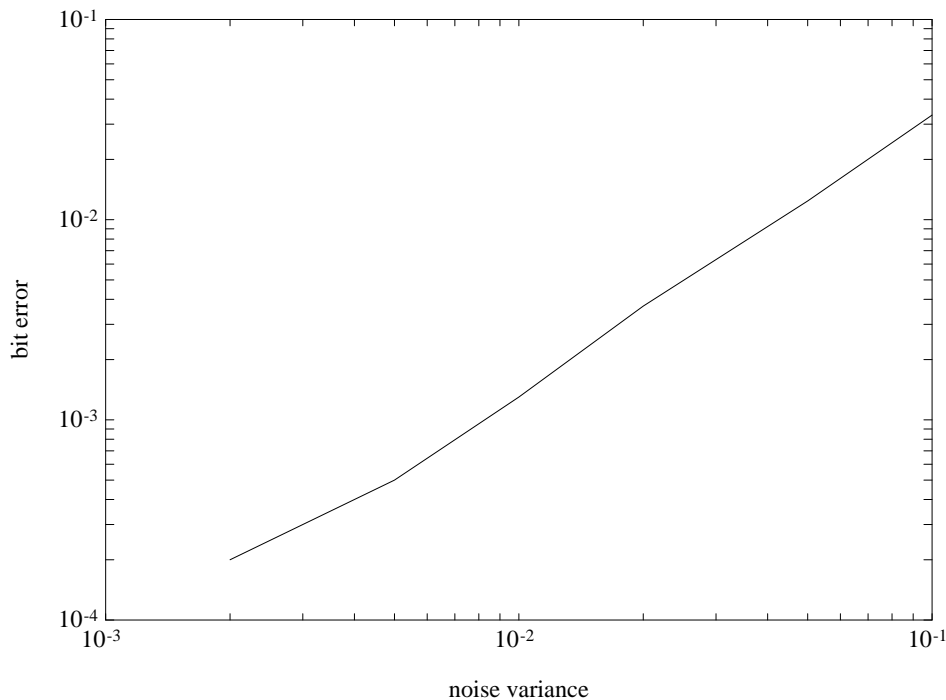
18

Figure 6: Bit-error as a function of the measurement noise for the Viterbi algorithm when the channel model is assumed to be known.

choice with a simpler algorithm with 16 filters.

The total bit error probability is estimated by comparing the estimated and the true input sequence and is shown by the dashed line in Figure 7 in the case of 16 parallel filters. Since no training sequence is used most of the erroneously estimated inputs are caused by transients. A better estimate of the bit error in the long run is computed by only comparing the last 20 inputs in each sequence as shown by the solid line.

It is almost inevitable to avoid so called zero crossings where all taps in the channel model are approximately zero at the same time. This phenomenon results in a very low signal-to-noise ratio for a while and it was observed that the algorithm was not capable to recover after a zero-crossing in some cases. Non-convergence of the algorithm is also possible when the input sequence is not persistently exciting for a long time in the beginning. To get insight into the influence of totally erroneously estimated sequences, the dotted line in Figure 7 shows the total bit error probability when sequences with bit error less than 10% are considered.

Figure 8 shows the same estimates for the same realizations when the number of parallel filters is increased to 64 in the algorithm. In this case, the difference between the solid and the dashed line is less significant, so the effect of the transients is almost negligible. Furthermore, the bit error rate
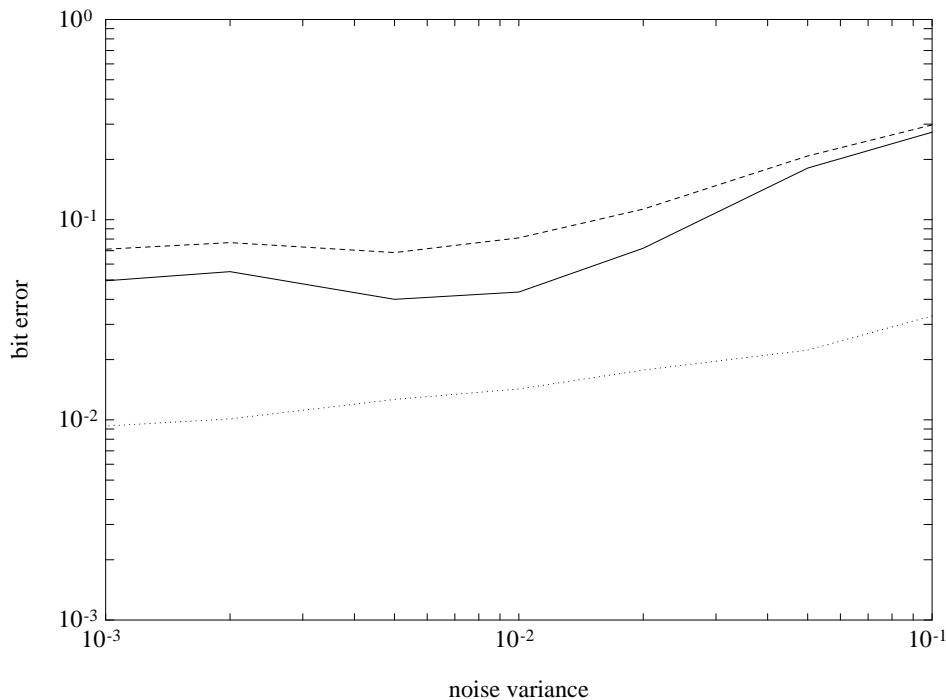
19

Figure 7: Bit-error as a function of the measurement noise with 16 parallel filters. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

is much lower. Even compared to the lower bound in Figure 6 the bit error is quite small, it only differs a factor 10 approximately. It is noteworthy, that for the largest noise variance, $\lambda = 0.1$, the bit error rate for simulations with a total bit error less than 10 % is the same as for the Viterbi algorithm. The conclusion is that the effect of transients and zero crossings are less the better approximation of the MAP estimate is used.

The problem of non-convergence of the algorithm for some simulations is a bit discouraging. However, it is important to note that this is an observable phenomenon, since it can be concluded from a perpetually switching between completely different estimated input sequences. Thus, it is easy to incorporate this test in the algorithm and in that case restart the algorithm, for instance by temporarily increasing Q. This idea is not persuaded here.

In Figure 9 a typical parameter convergence is shown. The true FIR real and imaginary parameter values are here compared to the least squares estimates conditioned on the estimated input sequence at time t. The convergence to the true parameter settings is quite fast and the tracking ability very good.

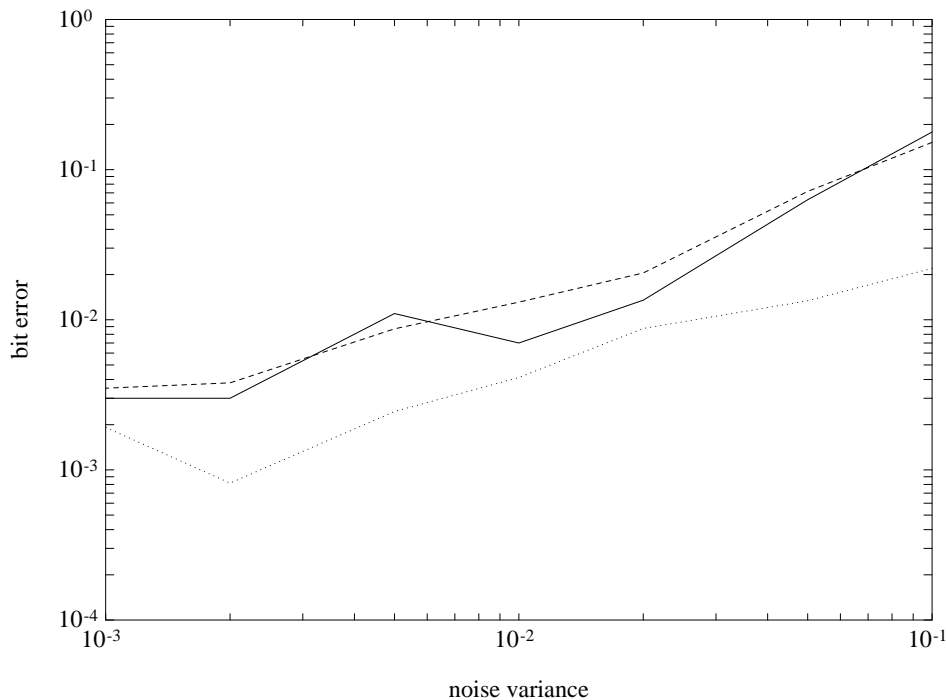As previously mentioned it is easy to incorporate a known training se-

Figure 8: Bit-error as a function of the measurement noise with 64 parallel filters. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

quence with Algorithm 2. One may believe that this would increase the performance drastically. Figure 10 shows the same estimates as in Figure 7 for the same realizations but where the first 10 samples of the input sequence are used as a training sequence. As expected there is no difference between the stationary (dashed line) bit error and the bit error rate when the transients are included (solid line). Compared with Figure 8 we see that the bit error rates are comparable. The conclusion, from this example, is that Algorithm 2 performs equally well with training sequence and without training sequence, but with a greater complexity of the algorithm.

6   CONCLUSIONS

We have herein studied the problem of blind deconvolution by direct examination of the input sequences. First, the identifiability problem of a noise-free FIR channel model was investigated. We know from the theory of system identification that the channel model is identifiable if the input is known and persistently exciting of order n. Here, when the input is unknown but
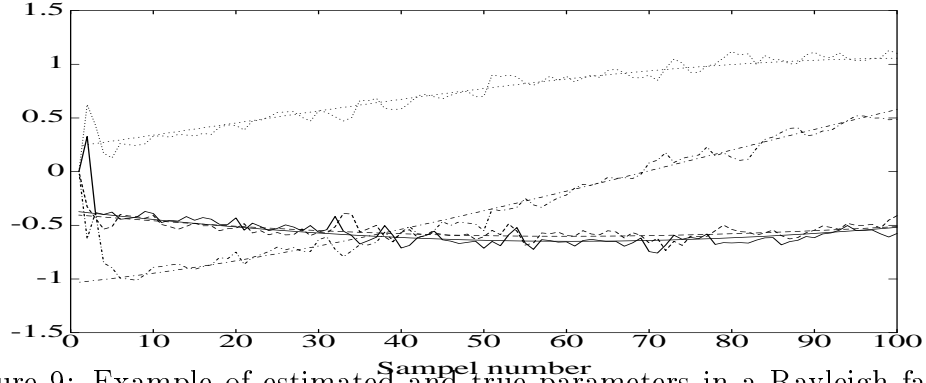
Figure 9: Example of estimated and true parameters in a Rayleigh fading channel.

belongs to a finite alphabet we have shown the following result:

- The channel model and the input sequence are simultaneously identifiable only if the input sequence is persistently exciting of order $2n - 1$.

- The complexity of the problem is determined by the first time instant when the input sequence is persistently exciting of order n.

Algorithm 1 gives a recursive scheme to solve the aforementioned deconvolution problem which works as follows. If the input sequence is known, then it is straight-forward to compute the channel model exactly (since there is no noise) and the next measurement can be computed exactly as well. Since the input belongs to a finite alphabet, there is only a finite number of input sequences. By computing the corresponding prediction to each sequence and rejecting all sequences that gives non-zero prediction error, the correct sequence is sooner or later found. The first point above gives a sufficient condition for this and the second point concerns an upper bound of the number of sequences that have to be examined.

A noise-free FIR channel model is not realistic in practice so next a time-varying IIR channel model with additive noise was studied. The *maximum a posteriori* estimate was derived. It can be computed recursively for each input sequence but the problem is that the number of input sequences increases
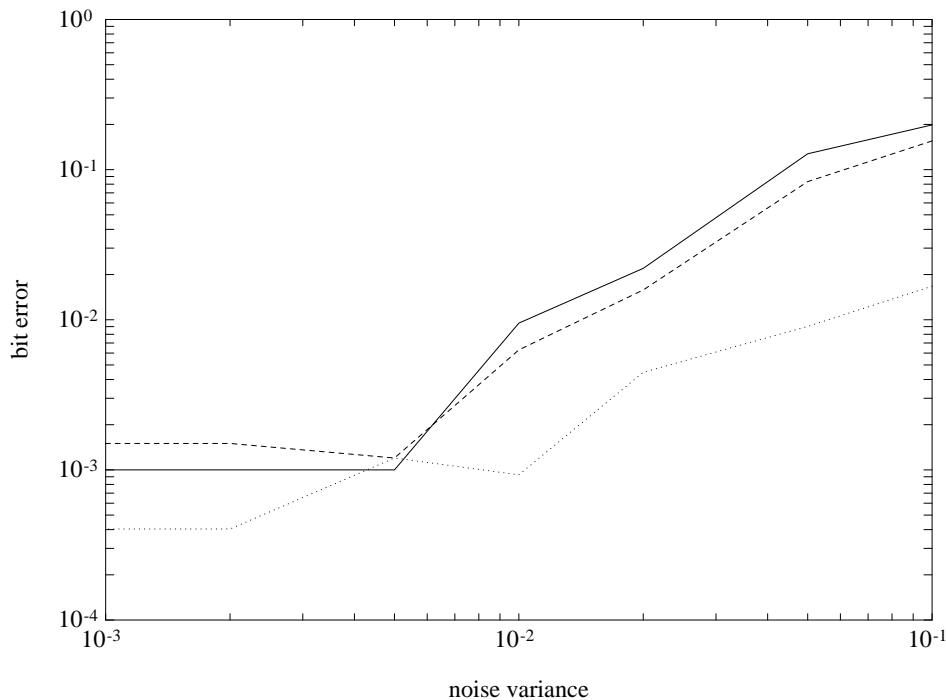
Figure 10: Bit-error as a function of the measurement noise with 16 parallel filters and a training sequence. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

exponentially. Now the theoretically results above were used to motivate a truly recursive approximation of the statistically optimal estimate, which only uses a fixed number of filters. It is given in Algorithm 2. We pointed out that the algorithm can be designed to be asymptotically optimal in the measurement noise for a time-invariant FIR filter.

The performance of Algorithm 2 was evaluated on a Rayleigh fading channel model. The bit error rate was compared for different complexities of the algorithm and also with the Viterbi algorithm when the true channel model was used. The algorithm turned out to show a very fast convergence to the true parameter settings, low bit error rate and it is fairly robust to a high noise level and zero crossings of the parameters.

Acknowledgement

A   Two lemmas

Here we use a somewhat different definition of Γ than in (9). If $\bar{A}_t$ is P.E. of order n at time m, we can solve (8) for b as $b = (\Phi_{m,n}^T \Phi_{m,n})^{-1} \Phi_{m,n}^T Y_m$. This gives the following alternate definition of Γ in the equation $\Gamma b = 0$:

$$\Gamma = \Phi_{t,n}(\Phi_{m,n}^T \Phi_{m,n})^{-1} \Phi_{m,n}^T \overline{\Phi}_{m,n} - \overline{\Phi}_{t,n}.$$

This reduces the set Z_n somewhat.

Lemma A.1 *Consider the $t - n + 1 \times n$ Toeplitz matrices $\Phi_{t,n}$ and $\overline{\Phi}_{t,n}$ containing elements in the set $\{\pm 1, \pm 3, ..., \pm(M-1)\}$. The elements of*

$$\det(\Phi_{m,n}^T \Phi_{m,n}) \left(\Phi_{t,n}(\Phi_{m,n}^T \Phi_{m,n})^{-1} \Phi_{m,n}^T \overline{\Phi}_{m,n} - \overline{\Phi}_{t,n} \right) \stackrel{\triangle}{=} \tilde{\Gamma} \qquad (28)$$

belong to the finite integer set

$$Z_n = \left\{ 0, \pm 1, \pm 2, ..., \pm 2 \, (M-1)^{2n+1} \, n^{n/2} (m - n + 1)^n \right\}.$$

Proof: First we have from $\mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n}) = \det(\Phi_{m,n}^T \Phi_{m,n})(\Phi_{m,n}^T \Phi_{m,n})^{-1}$ that the elements of

$$\tilde{\Gamma} = \Phi_{t,n}\mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n})\Phi_{m,n}^T \overline{\Phi}_{m,n} - \det(\Phi_{m,n}^T \Phi_{m,n})\overline{\Phi}_{t,n} \qquad (29)$$

must be integers, since they can be computed by only additions and multiplications. Next we find an upper bound on them.

Consider first the second term in (29). The elements of $\Phi_{m,n}^T \Phi_{m,n}$ are trivially bounded by $(m - n + 1)(M - 1)^2$. From *Hadamard's inequality*

$$| \det A|^2 \leq \prod_{j=1}^{n} \sum_{i=1}^{n} |a_{ij}|^2,$$

see [14] page 65, we have

$$\det(\Phi_{m,n}^T \Phi_{m,n}) \leq (M - 1)^{2n}(m - n + 1)^n n^{n/2}.$$

Thus, the elements of the second term in (29) are bounded by $(M-1)^{2n+1}(m - n + 1)^{2n} n^{n/2}$.

In the first term in (29), the elements of $\mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n})$ are computed by determinants of $(n-1) \times (n-1)$ submatrices and are thus bounded by $(M - 1)^{2(n-1)}(m-n+1)^{n-1}(n-1)^{(n-1)/2}$ so the elements of $\Phi_{t,n} \, \mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n})\Phi_{m,n}^T \overline{\Phi}_{m,n}$ are bounded by $(M-1)n(M-1)^{2(n-1)}(m-n+1)^{n-1}(n-1)^{(n-1)/2}(M-1)^2(m-n+1) = (M - 1)^{2n+1}(n - 1)^{(n-1)/2}n(m - n + 1)^n$ and the result follows. $\square$

The important implication of this lemma is that we now can conclude that $\tilde{\Gamma} = 0$ if $\tilde{\Gamma} b = 0$.

Lemma A.2 *Assume*

$$\Phi b = \overline{\Phi}\,\overline{b}, \tag{30}$$

where Φ *is a Toeplitz matrix as defined in (4) (for simplicity, we drop the indices here), and that b is linearly independent over Z_n (as defined in Lemma A.1). Then $\overline{\Phi}$ has full column rank if Φ has full column rank.*

Proof: Assume $\overline{\Phi}$ has rank k. Let $\tilde{\Phi}$ consist of k linearly independent columns of $\overline{\Phi}$. Then there must exist a $\tilde{b} = (\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T Y$ such that $Y = \tilde{\Phi}\tilde{b}$. Thus, we have

$$(I - \tilde{\Phi}(\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T)\Phi b = (I - P)\Phi b = 0, \tag{31}$$

where $P = \tilde{\Phi}(\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T$. Since the elements in $(I - P)\Phi$ belong to Z_n we have by assumption that $(I - P)\Phi = 0$, or equivalently $P\Phi = \Phi$. Now P is a projection matrix so range $\tilde{\Phi} = \text{range}P \supseteq \text{range}\Phi$ and thus $\tilde{\Phi}$ is of full column rank if Φ is of full column rank. $\qquad\square$

The conclusion of this lemma is that sequences that are not P.E. of order n are out of question if the true sequence is P.E. and b is linearly independent over Z_n.

References

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ., 1979.

[2] S. Bellini. Blind equalization. *Alta Frequenza*, LVII:445–450, 1988.

[3] A. Benveniste and M. Goursat. Blind equalizers. *IEEE Transactions on Communications*, 32:871–883, 1984.

[4] A. Benveniste, M. Goursat, and G. Ruget. Robust identification of a non-minimum phase system: Blind adjustment of a linear equalizer in data communication. *IEEE Transactions on Automatic Control*, 25:385–399, 1980.

[5] Z. Ding. *Application Aspects of Blind Adaptive Equalizers in QAM Data Communications*. PhD thesis, Cornell University, 1990.

[6] G.J. Foschini. Equalizing without altering or detecting data. *AT&T Technical Journal*, 64, 1985.

[7] D.N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transactions on Communications*, 28:1867–1875, 1980.

[8] G.H. Golub and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press, 1989.

[9] F. Gustafsson and B. Wahlberg. Blind equalization by direct examination of the input sequences. In *Proceedings on ICASSP '92*, pages IV,701–704, San Francisco, USA, March 23-26 1992.

[10] F. Gustafsson and B. Wahlberg. On simultaneous system and input sequence estimation. In *4th IFAC Symposium on Adaptive Systems in Control and Signal Processing*, pages 93–98, Grenoble, France, 1-3 July 1992.

[11] F. Gustafsson and B. Wahlberg. Identifiability in blind equalization. In *Proc. on the IFAC World Congress, 1993*, volume VII, pages 125–128, Sydney, Australia, 1993.

[12] S. Haykin. *Digital Communication*. John Wiley & Sons, Inc., 1988.

[13] C.R. Johnson. Admissibility in blind adaptive channel equalization. *IEEE Control Systems Magazine*, 11:3–15, 1991.

[14] P. Lancaster and M. Tismenetsky. *The theory of matrices*. Academic Press, Inc, London, 1985.

[15] W.C.Y Lee. *Mobile Communications Engineering*. McGraw-Hill, 1982.

[16] L. Ljung and T Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge MA, 1983.

[17] R.W. Lucky. Techniques for adaptive equalization of digital communiqation systems. *Bell System Technical Journal*, 45:255–286, 1966.

[18] O. Macchi and E. Eweda. Convergence analysis of self-adaptive equalizers. *IEEE Transactions on Information Theory*, 30:161–176, 1984.

[19] G. Picchi and G. Prati. Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm. *IEEE Transactions on Communications*, 35:877–887, 1987.

[20] Y. Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *IEEE Transactions on Communications*, 23:679–682, 1975.

[21] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inform. Theory*, 13:260–269, 1967.

[22] D. Yellin and B. Porat. Blind identification of FIR systems excited by discrete-alphabet inputs. *IEEE Transactions on Signal Processing*, pages 1331–1339, 1993.