BLIND EQUALIZATION BY DIRECT EXAMINATION OF THE INPUT SEQUENCES

Fredrik Gustafsson
Bo Wahlberg
Dept. of Electrical Engineering
Linköping University, Sweden

Submitted to IEEE Trans. on Communication

Abstract

The problem of blind deconvolution is addressed. The new approach presented herein examines the possible input sequences directly by using a bank of filters and, in contrast to common approaches, does not try to find an approximative inverse of the channel dynamics. First, the identifiability question of a noise-free FIR channel model is investigated. Sufficient condition for the input sequence and the channel are given which guarantee that both the channel model and the input sequence can be determined exactly in finite time. The condition on the input sequence is that it is sufficiently exciting (informative enough). This result motivates a recursive algorithm where all permissible input sequences are examinedand it is guaranteed to find the exact solution as fast as possible. An upper bound on the compputional complexity of the algorithm is given.

This algorithm is then generalized to cope with a time-varying IIR channel model 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. It shows fast convergence properties and good tracking ability.

1 INTRODUCTION

1.1 Preliminaries

The problem of channel equalization is of considerable interest in data communication and related fields. Given a recieved output sequence, we want to determine (recover) the transmitted input sequence. In the case the channel is modelled 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 [10] and [18]. 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, 7, 11]. The most common approach is to filter the output by an estimate of the inverse channel followed by some decision device. This approach is discussed in Section 2. If the input contains a known training sequence, it is straightforward to estimate a finite impulse respons (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. We assume that the input signal belongs to a finite alphabet, thus there are only a finite number of possible input sequences. Hence, by considering each of these as a training sequence we can estimate a finite bank of FIR models, 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. A crucial question is whether the input signal and channel model are identifiable from the measured output sequence. We will give sufficient conditions for this to hold.

The properties of the proposed blind equalization scheme is evaluated by applying the method to a Rayleigh fading channel. The results are very encouraging.

1.2 Problem Formulation

Consider the simplified but yet realistic digital communication system illustrated in Figure 1. The transmitter generates a sequence $\{u_t\}$ 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 modelled 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 u_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.

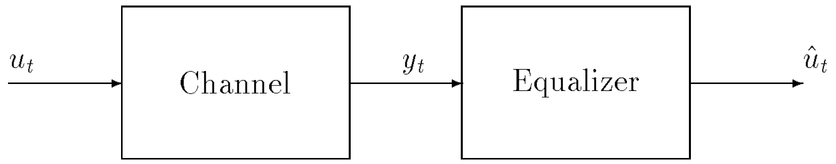

Figure 1: A digital communication system

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

$$
\begin{aligned}
y_t &= b_1 u_t + b_2 u_{t-1} \cdots b_n u_{t-n+1} \\
&= B(q) u_t
\end{aligned}
\tag{1}
$$

where $B(q) = b_1 + b_2 q^{-1} + ... + b_n q^{-(n-1)}$. Here q^{-1} denotes the backward shift operator, $q^{-1}u_t = u_{t-1}$. 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 $\varphi_t = (u_t, u_{t-1}, ..., u_{t-n+1})^T$. 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 $U_t = (u_1, ..., u_t)^T$. We will sometimes refer to U_t and Y_t as the sequences $\{u_k\}$ 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 U_t,

$$\Phi_{t,n} = \begin{pmatrix} u_n & u_{n-1} & \cdots & u_1 \\ u_{n+1} & \cdots & & u_2 \\ \vdots & & & \vdots \\ u_t & \vdots & & u_{t-n+1} \end{pmatrix}. \tag{4}$$

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 $\overline{U}_t = cU_t$ 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 of u_t. Furthermore, it is shown that all solutions can be written cU_t. This observation motivates the following definition of identifiability for the blind equalization problem.

Definition 1 *The input sequence and the channel model will be said to be identifiable from the measurements if all solutions to equation (3) can be written cU_t and $1/c \cdot b$, where U_t is the true input sequence and b the true channel model.*

This symmetric property of the problem does not cause any problems in practice, since the information is encoded in differential form so it is u_t/u_{t-1} that contains the information rather than u_t itself.

1.3 Outline Of The Paper

In Section 2, we will give a short review of proposed methods to blind equalization by using an inverse channel model filter, and related identifiability properties. In Section 3, we examine the identifiability issue of the blind equalization problem without assuming any particular structure of the equalizer. The obtained result is used in the equalizer presented in Section 4, which does not use an inverse model. In Section 5 simulation results are presented.

2 BLIND EQUALIZATION BY INVERSE MODEL FILTERING

2.1 Introduction

The standard approach to equalizing is to try to find a model of the inverse channel explicitly and then use a simple static decision device as shown in Figure 2. We will in this presentation follow the praxis that $B(q)$ is a non-minimum phase finite impulse response (FIR) filter, and that the inverse channel model filter is also chosen to have a FIR structure,

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

We define $H(q)$ as the combined channel–equalizer,

$$H(q) = B(q)C(q;c) = \sum_{k=0}^{\infty} h_k q^{-k}, \tag{6}$$

which ideally should equal 1. However, $H(q) = 1$ cannot be achieved which can be explained as follows. Since $B(q)$ is non-minimum phase its inverse must be split into one causal stable and one anti-causal stable part. Both these filters have infinite impulse responses, namely $\sum_{k=0}^{\infty} g_k^c q^{-k}$ and $\sum_{k=1}^{\infty} g_k^a q^k$, respectively ($c$ stands for causal and a for anti-causal). Truncating them to nc and na terms, where $nc + na = m$, and introducing a delay of na units give the approximative channel inverse impulse response

$$\tilde{C}(q) = g_{na}^a + g_{na-1}^a q^{-1} + \cdots + g_1^a q^{-na} + g_0^c q^{-na-1} + \cdots + g_{nc-1}^c q^{-nc-na+1}. \tag{7}$$

Comparing (7) with (5) gives that the non-minimum phase property of $B(q)$ introduces an unknown delay (na) in the equalizer which can neither be removed nor estimated. Thus, the best result which can be obtained is

$$H(q) = q^{-na} \tag{8}$$

for some unknown na.

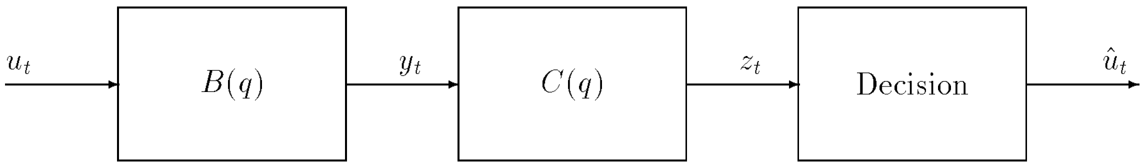

Figure 2: Equalizing by using an inverse filter

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 $\hat{u}_t = u_t$ 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. We use a linear regression form for the inverse filter (5),

$$z_t = \overline{\varphi}_t^T c, \tag{9}$$

where

$$\begin{aligned} \overline{\varphi}_t &= (y_t, \dots, y_{t-m+1})^T \\ c &= (c_1, c_2, \dots, c_m)^T. \end{aligned}$$

Note that the regression vectors in (2) and (9) are basically the same since both consist of old inputs, but that they are refering to different filters and thus different inputs.

2.2 Stochastic Gradient Based Algorithms

A vast number of methods for blind equalization has been suggested in the literature. We will here describe the most important ones and for simplicity it is assumed that $u_t = \pm 1$ although this is not necessary for their applicability. A common feature is that they attempt to minimize a loss function of the type

$$\mathrm{E}[\psi(z_t)] = \mathrm{E}[\psi(C(q;c)y_t)],$$

where E is the expectation operator, as pointed out in [11]. This leads to stochastic gradient algorithms resembling the least mean squares (LMS) method. The parameter estimate is here updated according to

$$\hat{c}_{t+1} = \hat{c}_t - \mu \frac{d}{dc}\psi(z_t)\mid_{c=\hat{c}_t}, \tag{10}$$

where μ is a stepsize. The loss function in LMS is $\mathrm{E}[e_t^2/2]$, where $e_t = u_t - z_t$, but the true output u_t is not available here, so the usual error signal $e_t = u_t - z_t$ cannot be used.

The perhaps most logical loss function, suggested in [14], is to approximate the error by $\hat{e}_t = \hat{u}_t - z_t = \mathrm{sign}(z_t) - z_t$. This is an example of decision feedback, were the decided symbol $\hat{u}_t$ is fed back as being the true one. The corresponding loss function is

$$\psi(z_t) = \frac{1}{2}(\mathrm{sign}(z_t) - z_t)^2 \tag{11}$$

which results in the stochastic gradient algorithm

$$\hat{c}_{t+1} = \hat{c}_t + \mu\overline{\varphi}_t(\mathrm{sign}(z_t) - z_t) \tag{12}$$

For a good initial estimate of c, $\hat{e}_t$ will equal e_t most of the time and the convergence properties can be expected to be the same as for the LMS method.

Another approach is based on the observation that $u_t^2 = 1$. The loss function

$$\psi(z_t) = \frac{1}{4}(1 - z_t^2)^2. \tag{13}$$

aims to adjust the parameters so that $z_t^2 = 1$. The corresponding algorithm is

$$\hat{c}_{t+1} = \hat{c}_t + \mu\overline{\varphi}_t z_t(1 - z_t^2). \tag{14}$$

This is usually called a modulus restoral algorithm and was first presented in [8]. The two loss functions (11) and (13) are the most considered ones, and several other suggested ones, see *e.g.* [2, 4, 16], can be interpreted as modifications of these.

2.3 Least Squares Based Algorithms

If the input sequence is known, for instance during a training period, the recursive least squares (RLS) algorithm can be used to estimate c. The loss function in RLS is

$$\psi(z^t) = \frac{1}{t}\sum_{k=1}^{t} \lambda^{t-k}|u_k - z_k|^2, \tag{15}$$

where λ is a forgetting factor. The minimization of (15) is easily done recursively, see for instance [13], and the convergence is known to be much faster than for LMS.

In case the input is unknown, the prediction error $e_t = u_t - z_t$ can be replaced by $\mathrm{sign}(z_t) - z_t$ by using decision feedback in the recursive updates, just as in (11). In [6] it is suggested to minimize the off-line expression (15) with respect to the parameters c and the entire sequence $\{u_k\}$ which hopefully will give better robustness properties. The drawback is that the minimization cannot be performed recursively and is computationally very expensive.

2.4 Convergence Properties

It is clear from examples that the aforementioned algorithms sometimes fail to converge to an open-eye condition as shown in [11], 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. We will discuss admissibility results of the loss functions (11), (13) and (15).

The decision directed algorithm is shown to converge in [17] 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 $\hat{u}_t = \mathrm{sign}(z_t) = \mathrm{sign}(u_t) = u_t$, so this assumption corresponds to open-eye initialization. That open-eye initialization is generally sufficient for convergence is proved in [15].

The modulus restoral algorithm is shown to converge for an appropiate initial setting in [8], but is shown to have undesirable local—but not global—minima, see [6]. In [7], the convergence to the desired over-all impulse response is proven if the equalizer is infinite dimensional (that is, $m = \infty$ in (5)).

Consider now the off-line minimization of (15) as proposed in [6]. If the true channel is described by (9) then they show that (15) has just one minima corresponding to the true values. The conditions are quite restrictive: infinitely many special input sequences of infinite length are assumed to have appeared.

2.5 Basic Limitations

To conclude this section we summarize the disadvantages of the inverse model filtering approach.

- The choice of loss function $\psi(z_t)$ is rather *ad hoc*.
- All suggested loss functions have undesired local minima.
- The inverse channel, which is IIR, must be approximated by a FIR filter.
- The over-all impulse response contains an unknown delay.
- Not even asymptotic convergence in the parameter vector b can be expected in the noise-free case of (22), since a constant stepsize is used in the proposed gradient schemes.

In the next section, it is shown that the above mentioned problems are not inherent in the model (22) but depend on the inverse filtering approach.

3 IDENTIFIABILITY

We will here investigate the important question of identifiability of the parameters for a noise-free FIR channel. It turns out that the parameters actually are identifiable, if the input is informative enough, as will be defined exactly later on. First, the problem is formulated once more. We use the matrix formulation (3) of the noise-free FIR channel model, reproduced here for convenience,

$$Y_t = \Phi_{t,n} b. \tag{16}$$

Here Y_t contains the measurements, 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 information in the input is measured in the usual way, *c.f.* [13].

Definition 2 *The sequence $\{u_t\}$ 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).*

If the input sequence belongs to a finite set this basically means that the input sequence may not be periodical with a period shorter than n. 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 well-known 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).

The aforementioned classical result in system identification, where the input is known, is that if the system is noise-free then the parameter vector b is identifiable and equal to $(\Phi_{t,n}^T \Phi_{t,n})^{-1} \Phi_{t,n}^T Y_t$ if u_t is P.E. of order n, that is $\Phi_{t,n}$ has full rank. The following theorem claims that $\Phi_{t,n}$ and b actually are simultaneously identifiable if the input is persistently exciting of order $2n-1$. We assume that the coefficients b_i are linearly independent over a certain finite field of integers, which is defined in the theorem. This means that there must not exist relationships like $3b_1 + 7b_3 = 0$. This restriction is, however, not so severe in practice as motivated after the proof.

Theorem 3.1 *Let the true input sequence and parameter vector be* $\overline{U}_t$ *and* $\overline{b}$ *so that* $Y_t = \overline{\Phi}_{t,n}\, \overline{b}$, *where* $\overline{u}_t \in \{\pm 1, \pm 2, ..., \pm M/2\}$. *Assume that* $\overline{U}_t$ *is P.E. of order* $2n-1$ *at time* t *and P.E. of order* n *at time* t_0. *Suppose that* $\{b_i\}$ *is linearly independent over the finite integer field*

$$Z_n = \left\{ -2\left(\frac{M}{2}\right)^{n+1} n!(t_0 - n + 1)^n, ..., -1, 0, 1, ..., 2\left(\frac{M}{2}\right)^{n+1} n!(t_0 - n + 1)^n \right\}. \tag{17}$$

Then $\overline{b}$ *and* $\overline{U}_t$ *are identifiable from* Y_t.

Proof: Suppose that there exist U_t and b such that

$$Y_t = \overline{\Phi}_{t,n}\, \overline{b} = \Phi_{t,n} b. \tag{18}$$

According to the assumptions, $\overline{\Phi}_{t_0,n}$ has full rank and Lemma A.2 gives that the search over $\Phi_{t_0,n}$ can be restricted to full rank matrices. We can now solve (18) for b,

$$b = \Phi_{t_0,n}^{\dagger} \overline{\Phi}_{t_0,n}\, \overline{b} = (\Phi_{t_0,n}^T \Phi_{t_0,n})^{-1} \Phi_{t_0,n}^T \overline{\Phi}_{t_0,n}\, \overline{b}, \tag{19}$$

where $\dagger$ denotes pseudo-inverse, which gives

$$\left(\Phi_{t,n} \Phi_{t_0,n}^{\dagger} \overline{\Phi}_{t_0,n} - \overline{\Phi}_{t,n} \right) \overline{b} \triangleq A \overline{b} = 0.$$

According to Lemma A.1 the elements of $\det(\Phi_{t_0,n}^T \Phi_{t_0,n}) A$ belongs to Z_n so we must have $A = 0$ by the assumption of linear independence of $\{b_i\}$.

Now, define $S = (s_1, ..., s_n) = \Phi_{t_0,n}^{\dagger} \overline{\Phi}_{t_0,n}$, where s_i is the i^{th} column of S, which is an $n \times n$ matrix. $A = 0$ can now be expressed as

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

or

$$\begin{pmatrix} u_n & \cdots & u_1 \\ \vdots & & \vdots \\ u_t & \cdots & u_{t-n+1} \end{pmatrix} (s_1, \cdots, s_n) = \begin{pmatrix} \overline{u}_n & \cdots & \overline{u}_1 \\ \vdots & & \vdots \\ \overline{u}_t & \cdots & \overline{u}_{t-n+1} \end{pmatrix}. \tag{20}$$

Eliminating $\overline{u}_k$, using the Toeplitz structure, and solving (20) for s_1 and s_n gives

$$\begin{pmatrix} u_{2n-1} & \cdots & u_n & u_n & \cdots & u_1 \\ u_{2n} & \cdots & u_{n+1} & u_{n+1} & \cdots & u_2 \\ \vdots & & & & & \vdots \\ u_t & \cdots & u_{t-n+1} & u_{t-n+1} & \cdots & -u_{t-2n+2} \end{pmatrix} \begin{pmatrix} -s_n \\ s_1 \end{pmatrix} \triangleq F_t x = 0. \qquad (21)$$

Note that F_t is identical to $\Phi_{t,2n-1}$ except for that the middle column appears twice in F_t. According to the assumption of P.E., 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

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

where c is a (possibly complex) constant. Hence $s_1 = ce_1$ and $s_n = ce_n$ where e_i is the i^{th} column in the identity matrix I. Continuing solving (20) 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 $U_t = 1/c\,\overline{U}_t$. □

We remark that the important assumption that u_t belongs to a finite set is necessary in the first half of the proof to conclude that $A = 0$ from $Ab = 0$, and the Toeplitz structure of $\Phi_{t,n}$ is necessary in the second part for S to equal the identity matrix.

The restriction on b is not too severe, losely speaking it is satisfied with probability one. Even if it is not satisfied, examples have shown that U_t is still identifiable which can be explained as follows. Suppose that b is linearly dependent over Z_n. Then one of the rows of A may be orthogonal to b although it is not likely. Equation (20) still holds except for some rows that must be deleted. The key point is that an equation like (21) 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 $\overline{b}$ makes it difficult to give any necessary conditions on U_t in the theorem.

The symmetry property inherent in the problem formulation is clarified in the last sentence of the proof where it is concluded that all sequences which can be expressed as a constant times the true sequence solves the problem. For instance, if $u_t = \pm 1$, then we must have $\overline{U}_t = \pm U_t$.

4 A DIRECT EQUALIZER

4.1 The Noise-Free Case

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

$$Y_t = \Phi_{t,n} b, \qquad (22)$$

consistent with the identifiability result in Section 3, which gives the correct values of U_t and b as fast as possible. It will be extended in the next subsections to cover more realistic channel models. In words it works as follows, see also Figure 3.

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 U_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 u_t. These are the* a priori *permissible sequences at time $t+1$.*
3. *Let the* a posteriori *permissible sequences at time $t+1$ be those which satisfy*

$$Y_{t+1} = \Phi_{t+1,n}^i (\Phi_{t+1,n}^i)^\dagger Y_{t+1}. \tag{23}$$

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

$$y_{t+1} = (\varphi_{t+1}^i)^T b^i. \tag{24}$$

4. *Repeat the above steps.*

We remark that if a sequence is not P.E. of order n, then the pseudo-invers $(\Phi_{t,n}^i)^\dagger$ cannot be computed as $(\Phi_{t,n}^T \Phi_{t,n})^{-1} \Phi_{t,n}^T$, but it can always be computed by the singular value decomposition, see [9]. Consider now the sequences that are P.E.

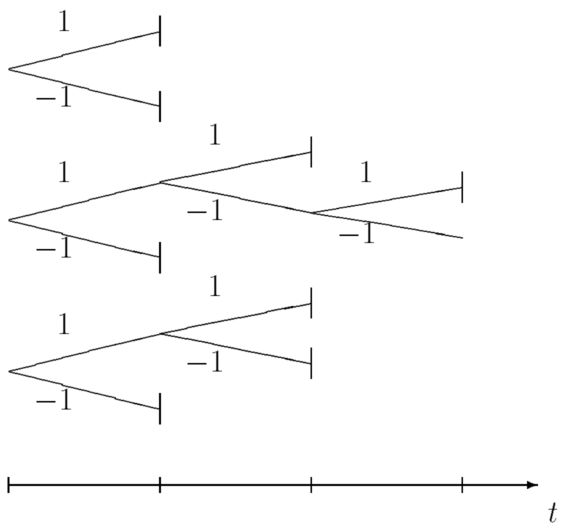

Figure 3: Recursive solution of $Y_t = \Phi_{t,n} b$ for $M = 2$

of order n at time t, so that b^i is unambigous. It is then clear that each survivor at time t will have at most one survivor at time $t+1$, since the relation $u_{t+1}^i = 1/b_1^i(y_{t+1} - b_2^i u_t - ... b_n^i u_{t-n+2})$ defines u_{t+1}^i uniqely and gives at most one permissible value in the finite alphabet of u_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 (22) 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 (22) and suppose that the input sequence $\{u_t\}$ 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}. Furthermore, if t_0 is unknown but the parameter vector b is linearly independent over Z_n, as defined in Theorem 3.1, 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 u_t.*

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 (22) for some b must have rank n. According to the definition of P.E., all permissible input sequences is 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 Theorem 3.1 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 gives 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. very quickly.

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 possible time-varying channel disturbed by noise. First, consider the model (22) 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 (23) is replaced by

$$||Y_{t+1} - \Phi^i_{t+1,n} b^i|| < \epsilon \tag{25}$$

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 (25) 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 (25) 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 U_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

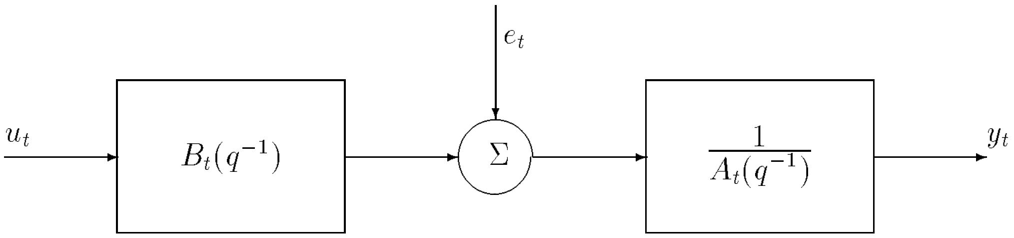

Figure 4: Time-varying channel model with noise

$$y_t = -a_1(t)y_{t-1} - \cdots - a_{n_a}(t)y_{t-n_a} + b_1(t)u_t + \cdots + b_{n_b}(t)u_{t-n_b+1} + e_t.$$

This model is equivalent to (22) if the parameters are constant, $n_a = 0$ 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 modelled as a random walk so a total state space linear regression model is

$$\begin{aligned} \theta_{t+1} &= \theta_t + v_t \\ y_t &= \varphi_t^T \theta_t + e_t. \end{aligned} \tag{26}$$

Here, φ_t is the regression vector $(-y_{t-1}, \cdots, -y_{t-n_a}, u_t, \cdots, u_{t-n_b+1})^T$ and θ_t is the unknown parameter vector $(a_1(t), \cdots, a_{n_a}(t), b_1(t), \cdots, b_{n_b}(t))^T$. We assume that v_t and e_t are uncorrelated white Gaussian noises with covariance matrices Q_t and Λ_t, respectively. Suppose we have an arbitrary enumeration of the M^t sequences U_t, say U_t^i, $i = 1, 2..., M^t$, and denote the corresponding regression vector φ_t^i. The *a posteriori* probabilities of a sequence U_t^i, given the observations Y_t, is given by the following theorem.

Theorem 4.2 *Consider the model (26). The* a posteriori *probability of U_t^i is given by*

$$\begin{aligned} p(U_t^i|Y_y) &= C_t p(u_t^i|U_{t-1}^i)\gamma(y_t - (\varphi_t^i)^T\hat{\theta}_t^i, (\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t) \cdot p(U_{t-1}^i|Y_{t-1}) \quad (27) \\ &= C p(U_t^i) \prod_{k=1}^{t} \gamma(y_k - (\varphi_k^i)^T \hat{\theta}_k^i, (\varphi_k^i)^T P_k^i \varphi_k^i + \Lambda_t) \quad (28) \end{aligned}$$

where C and C_t are constants given by the condition $\sum_{i=1}^{M^t} p(U_t^i|Y_t) = 1$. Here $\gamma(x-\mu, P)$ denotes the value of the Gaussian probability density function with mean μ and covariance P evaluated in x. The parameter estimate $\hat{\theta}_t^i$ and its covariance matrix P_t^i is computed recursively by

$$\hat{\theta}_{t+1}^i = \theta_t^i + P_t^i\varphi_t^i\left((\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t\right)^{-1}\left(y_t - (\varphi_t^i)^T\hat{\theta}_t^i\right) \tag{29}$$

$$P_{t+1}^i = P_t^i + P_t^i\varphi_t^i\left((\varphi_t^i)^T P_t^i \varphi_t^i + \Lambda_t\right)^{-1}(\varphi_t^i)^T P_t^i + Q_t \tag{30}$$

with the initial values θ_0 and P_0.

Proof: Repeated use of Bayes' rule gives

$$\begin{aligned} p(U_t^i|Y_t) &= \frac{p(Y_t, U_t^i)}{p(Y_t)} \\ &= \frac{p(U_t^i)}{p(Y_t)} p(Y_t|U_t^i). \end{aligned} \tag{31}$$

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

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

which is (27) and (28) 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 (26) is Gaussian if φ_t is known, see for instance [1]. The equations (29) and (30) are just the Kalman filter equations for the state space model (26). □

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

$$\hat{U}_t^{MAP} = \arg\max_i p(U_t^i|Y_t). \tag{32}$$

The *maximum likelihood* (ML) estimate,

$$\hat{U}_t^{ML} = \arg\max_i p(Y_t|U_t^i). \tag{33}$$

is closely related to the MAP estimate as seen from (31). It can be computed from the MAP estimate by letting the prior being non-informative, that is $p(U_t) = 1/M^t$.

The prior information $p(u_t^i|U_{t-1}^i)$ in (27) can be used to decode the information by rejecting "impossible" sequences, thus eliminating the use of a separate decoder. It can also be used to incorporate a training sequence in a very natural way, by letting $p(U_t^{train} = 1)$. However, most often the inputs 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 (22). 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 (22) and suppose that the same conditions as in Theorem 3.1 hold. Then the MAP estimate (32) 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 is small enough.

4.4 Basic Limitations

The MAP estimate (32) completely eliminates the disadvantages of equalizing by an inverse filter which are mentioned in Section 2.5. However, it introduces some new problems:

- The computational complexity is exponential increasing, since it requires M^t Kalman filters at time t.
- It is not guaranteed that $\hat{U}_t^{MAP}$ resembles $\hat{U}_{t-1}^{MAP}$, 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 that have turned out to be likely are considered, the others are rejected.

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

1. *Compute $p(U_{t+1}^i|Y_{t+1})$ 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 who have the largest $p(U_{t+1}^i|Y_{t+1})$.*
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 [18], 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).

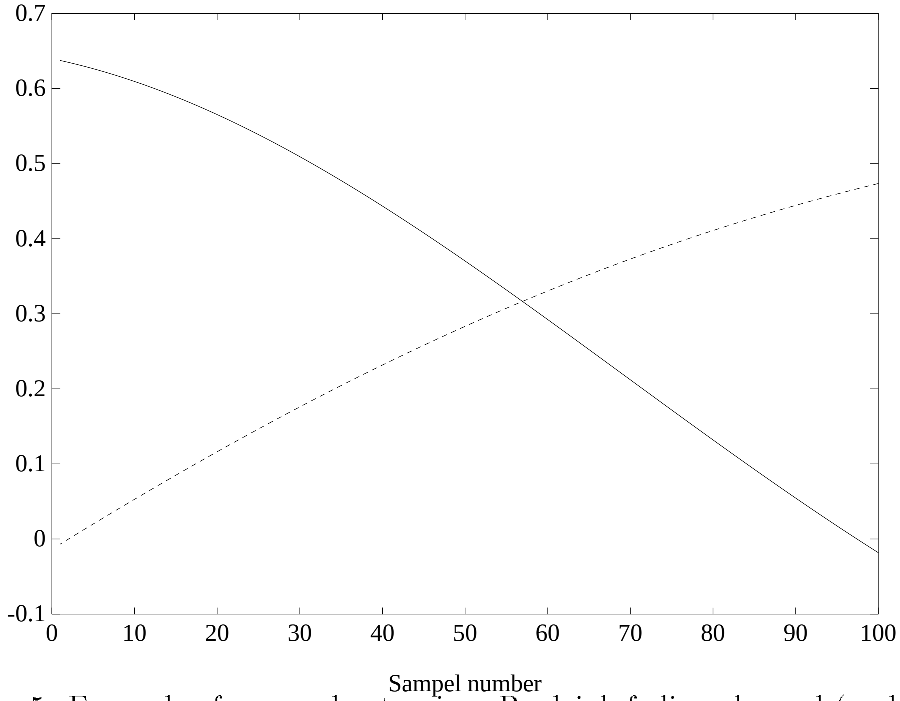

Figure 5: Example of a complex tap in a Rayleigh fading channel (real and imaginary parts)

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 where the receiver is moving, which 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 [12]. 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 was filtered through a simulated Rayleigh fading channel and Gaussian noise with variance λ was added. 100 different realizations of the input sequence, the noise sequence and the channel was 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 get 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

[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 $\pi/2 \cdot 70/25000$.

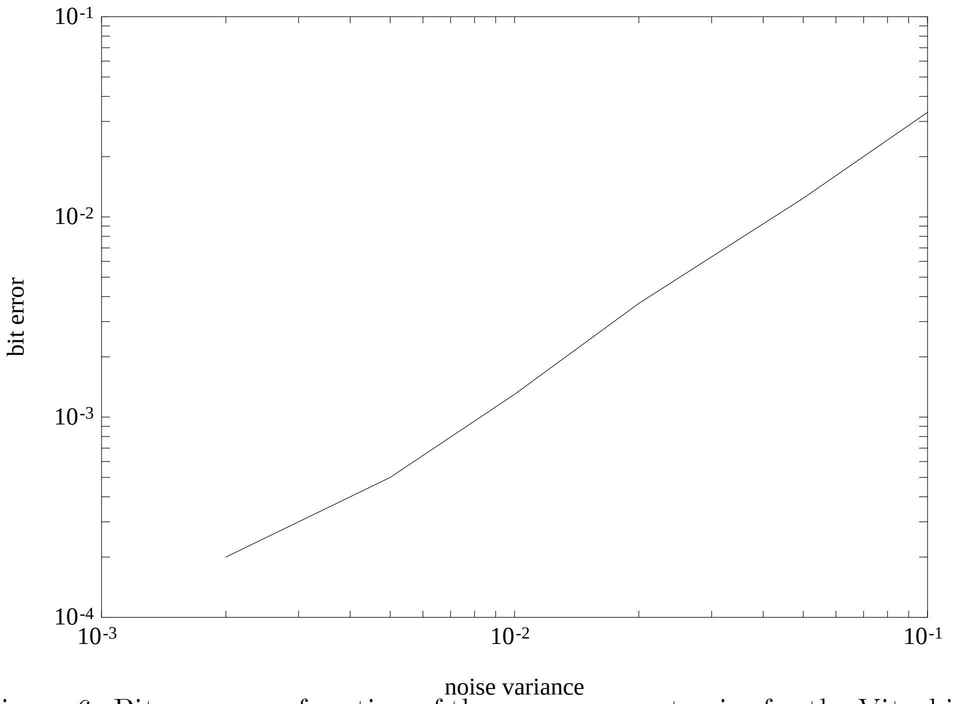

Figure 6: Bit-error as a function of the measurement noise for the Viterbi algorithm when the channel model is known.

Figure 6, where the first value, corresponding to the smallest λ, is zero and is not shown. The input sequence was then estimated by Algorithm 2 with 16 parallel filters. The true measurement noise variance was used and the variance (Q) of the random walk was chosen to 0.01. 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. Since no training sequence was 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 a feeling for the influence of totally erroneously estimated sequences, the dotted line in Figure 7 shows the total bit error probability when sequences with worse bit error than 10% are removed.

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

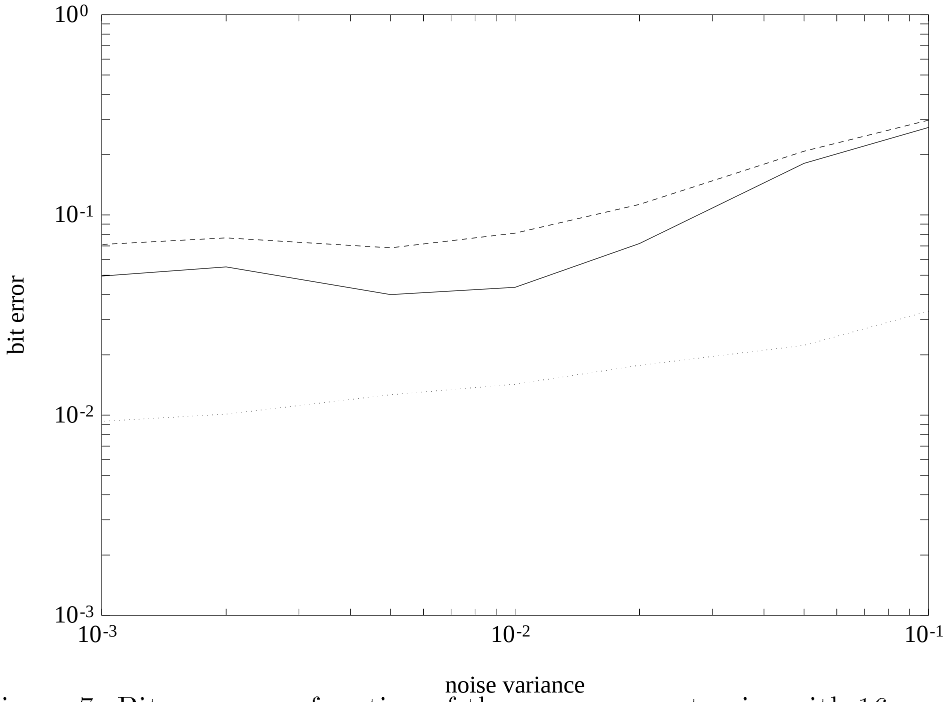

Figure 7: Bit-error as a function of the measurement noise with 16 parallel filters. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

little discouraging. However, it is observable phenomenon for the user, since it can be concluded from a perpetually switching between completely different estimated input sequences. Thus, the user in principle knows when the estimate is worthless and can try to restart the algorithm, for instance by temporarily increasing Q. This idea is not examined here.

In Figure 9 a typical parameter convergence is shown. The true FIR real and imaginary parameter vaules 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 sequence 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 here 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

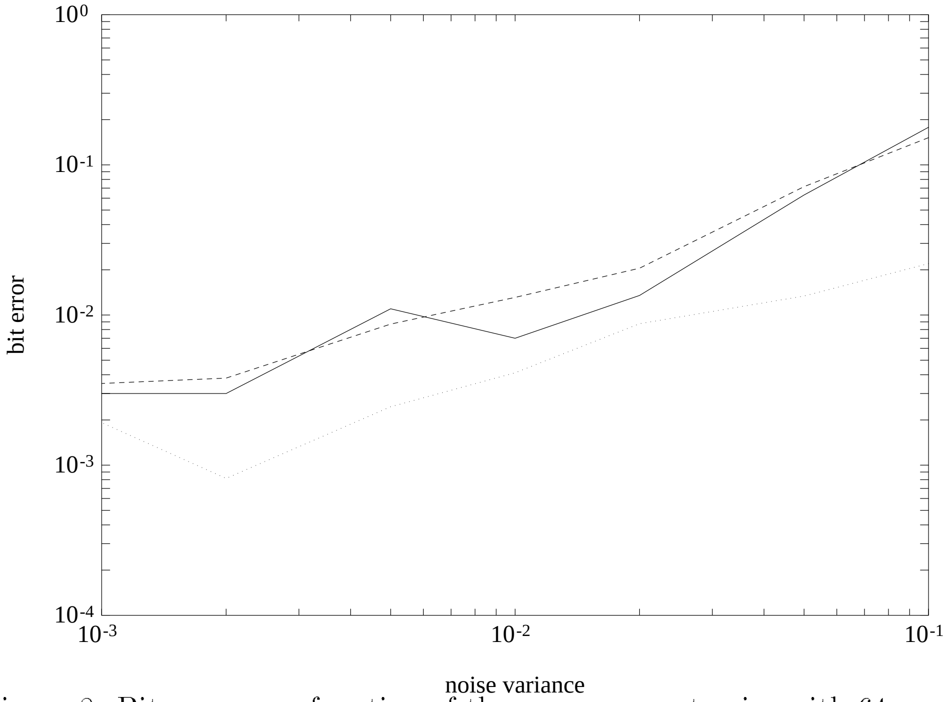

Figure 8: Bit-error as a function of the measurement noise with 64 parallel filters. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

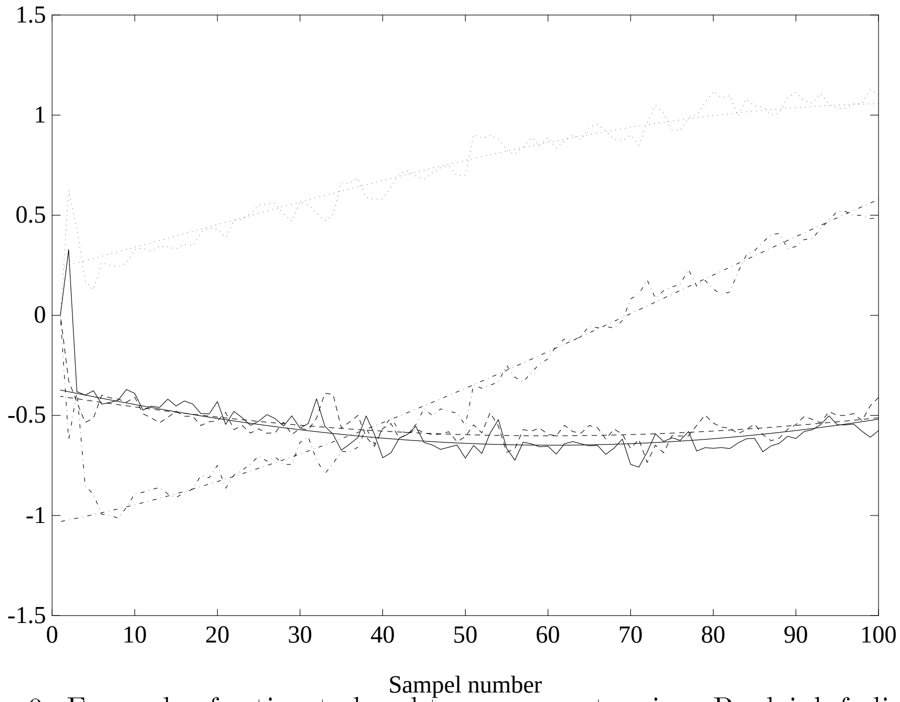

Figure 9: Example of estimated and true parameters in a Rayleigh fading channel.

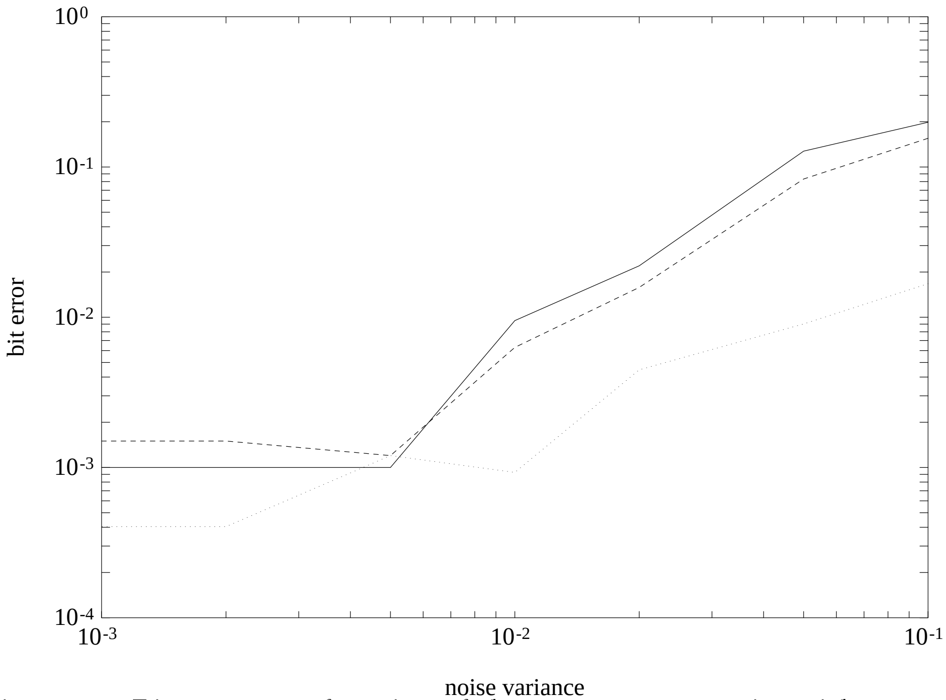

Figure 10: Bit-error as a function of the measurement noise with 16 parallel filters and a training sequence. Averages over all inputs and all simulations (dashed), only the last 20 samples of the inputs (solid) and only for simulations with less than 10 % bit error (dotted).

of order n. Here, when the input is unknown but belongs to a finite alphabet we have the following:

- The channel model and the input sequence are simultaneously identifiable if the input sequence is persistently exciting of order $2n-1$.
- The complexity of the problem is determined by the 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 straightforward 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 very 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 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

The authors would like to thank Dr. Michael Sternad and Mr. Lars Lindbom for providing the Rayleigh fading channel model used in Section 5.

A TWO LEMMAS

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 $\{\pm 1, \pm 2, ..., \pm M/2\}$. The elements of*

$$\det(\Phi_{m,n}^T \Phi_{m,n}) \left(\Phi_{t,n} (\Phi_{m,n}^T \Phi_{m,n})^{-1} \Phi_{m,n}^T \overline{\Phi}_{m,n} - \overline{\Phi}_{t,n} \right) \triangleq \tilde{A} \tag{34}$$

belong to the finite integer set

$$Z_n = \left\{ -2 \left(\frac{M}{2} \right)^{n+1} n!(t_0 - n + 1)^n, ..., -1, 0, 1, ..., 2 \left(\frac{M}{2} \right)^{n+1} n!(t_0 - n + 1)^n \right\}.$$

Proof: First we have that the elements of $\tilde{A} = \Phi_{t,n}(\Phi_{m,n}^T \Phi_{m,n})^{-1} \Phi_{m,n}^T \overline{\Phi}_{m,n} - \det(\Phi_{m,n}^T \Phi_{m,n}) \overline{\Phi}_{t,n}$ must be integers, since they can be computed by only additions and multiplications. Next we find an upper bound on them.

Since the elements of $\Phi_{m,n}^T \Phi_{m,n}$ are bounded by $M(m-n+1)$ we first have from the definition of a determinant that the elements of $\det(\Phi_{m,n}^T \Phi_{m,n})$ are trivially bounded by $M^n(m-n+1)^n n!$. Next, the elements of $\mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n})$ are computed by determinants of $(n-1) \times (n-1)$ matrices and are thus bounded by $M^{n-1}(m-n+1)^{n-1}(n-1)!$ so the elements of $\Phi_{m,n}$ $\mathrm{adj}(\Phi_{m,n}^T \Phi_{m,n}) \Phi_{m,n}^T \overline{\Phi}_t$ are bounded by $MnM^{n-1}(m-n+1)^{n-1}(n-1)!M(m-n+1) = M^{n+1}n!(m-n+1)^n$ and the result follows. □

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

Lemma A.2 *Suppose*

$$\Phi b = \overline{\Phi}\, \overline{b}, \tag{35}$$

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: Suppose $\overline{\Phi}$ 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

$$(I - \tilde{\Phi}(\tilde{\Phi}^T \tilde{\Phi})^{-1} \tilde{\Phi}^T) \Phi b = (I - P) \Phi b = 0, \tag{36}$$

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 $\tilde{\Phi} = \mathrm{range} P \supseteq \mathrm{range} \Phi$ and thus $\tilde{\Phi}$ is of full column rank if Φ is of full column rank. □

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 nonminimum 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] Z. Ding, C.R. Johnson, and T.L. Fine. Decision-directed with memory: a rls adaptive equalization algorithm.

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

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

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

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

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

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

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

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

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

[16] 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.

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

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