ON THE IMPACT OF USING UNRELIABLE DATA ON THE BOOTSTRAP CHANNEL ESTIMATION PERFORMANCE

Samson Lasaulce¹ and Noura Sellami²

1France Télécom R&D, 38-40 rue du Général Leclerc, 92794 Issy-les-Moulineaux, France, samson.lasaulce@francetelecom.com
2ENSEA Laboratoire ETIS, 6 avenue du Ponceau, 95014 Cergy-Pontoise, France, sellami@ensea.fr

ABSTRACT

In this paper, we study the performance of the bootstrap channel estimation scheme, which consists in this paper in using the hard-decided outputs of the channel decoder in order to extend the training sequence. Based on a simple large system analysis, we derive the expression of the channel estimation MSE in order to evaluate the impact of using wrong decisions on channel estimation performance. In particular, it is shown that it depends on the first and second moment of the number of errors per block of symbols. Interestingly, it is also proven that the bootstrap estimator does not always improve channel estimation accuracy. The performance of the Viterbi equalizer based on bootstrap channel estimation is assessed through a SNR analysis.

1. INTRODUCTION

The most conventional way of estimating the propagation conditions in mobile radio systems is to model the channel by a Finite Impulse Response filter and to transmit a data sequence known to the receiver. If channel noise is additive, white and Gaussian, the Maximum Likelihood (ML) estimator based on the observations generated by the training sequence (TS) consists in minimizing the mean square error between the signal received during the emission of the TS and its noiseless counterpart (the filtered version of the TS), which is a Least-Squares (LS) estimator.

The performance (in terms of error variance) of the Least-Squares channel estimator is roughly inversely proportional to the training sequence length (see [1] for example). Therefore, a simple idea to increase the number of training symbols is to use the hard decisions associated with the outputs of the symbol detector or channel decoder (e.g. [2]). Of course, it is also possible to exploit the soft information associated with the latter if it is available. In this paper, we will only consider the case where the channel estimator is fed with hard decisions. As the decoder output is generally more reliable than the symbol detector output, it is more efficient to use the decoded bits. The latter are then re-encoded and used to extend the initial training sequence. Generally, several iterations are performed in order to improve the overall receiver performance. The corresponding estimation strategy is often referred to as bootstrap estimation or iterative estimation based on hard decisions.

In contrast with the conventional estimation procedure, in which the training symbols are perfectly known, the bootstrap estimator does not coincide with the ML estimator because of the presence of errors in the extended training sequence. Such an introduction of "noise" into the training data affects the performance of the bootstrap channel estimator. The corresponding performance degradation has been evaluated in [2] and [3] based on the Cramer-Rao Bound (CRB). Performance analyses respectively rely on an asymptotic analysis (large extended TS length) in [2] and a deterministic/Gaussian symbol model in [3]. One of the disadvantages of considering lower performance bounds is that the behavior of the real estimation performance is partially hidden. In particular, the fact that the CRB decreases with the "extended TS" length does not imply that the bootstrap estimator variance is also a decreasing function of this parameter. A relevant analysis of iterative channel estimation performance in turbo equalization has been done in [4] but only the soft information case is considered.

It turns out that under the assumptions of [3], it is possible to obtain the estimation variance of the bootstrap estimator itself. This is precisely the main purpose of this paper. Furthermore, papers dealing with iterative channel estimation generally do not consider the receiver performance through a Block Error Rate (BER) analysis but a Bit Error Rate (BER) analysis whereas the "irregular behavior" of the number of errors per block plays an important role in the design of the bootstrap estimator.

This paper is organized as follows. The signal model and description of the bootstrap estimation are provided in section 2. Section 3 briefly reviews the main assumptions that are used to derive the expression of the bootstrap channel estimation mean square error (MSE). The main part of this paper is in section 4 where it is showed how to obtain the estimation error variance expression. The corresponding expression is discussed in section 5. This section also provides simulation results. Section 6 reviews the main contributions of this paper.

General notations. In this paper, notations \( s, \overrightarrow{v} \) and \( \mathbf{M} \) stand for scalar, vector and matrix respectively. Notations \( (\cdot)^t \) and \( (\cdot)^H \) stand for transpose and Hermitian operators.

2. DISCRETE-TIME SIGNAL MODEL AND PROBLEM STATEMENT

2.1. Framework

In this paper we consider a simple scenario where there is one transmitter (figure 1) and one receiver (figure 2). A single antenna is assumed at both sides of the transmission. The transmitter sends coded and interleaved BPSK symbols \( \{x(t) \in \{-1, +1\}, t \in \mathbb{Z}\} \) to the receiver over a multipath channel (figure 3) that includes the emission filter, the reception filter
and the propagation channel. Notation \( \Pi \) stands for matrix interleaving. Additionally, a block transmission is assumed. Each block comprises two parts: a training sequence of size \( T_0 \) (for instance it can be located at the beginning of the block) and an information data part. In total, there are \( T_i \) symbols. The channel is modeled by its equivalent symbol-rate discrete-time impulse response and a white and Gaussian noise \( \nu = N(0, \sigma^2) \) is added to the filtered symbols as described in figure 3. We will see throughout the paper to what extent the results provided in this paper can be applied to more complex situations.

2.2. Expression of the received signal

Assuming the channel impulse response to be constant over the transmitted block under consideration, the observation equation corresponding to figure 3 writes:

\[
\forall t \in [0, T_i - 1], y(t) = \sum_{\ell=0}^{\ell-1} h(\ell)x(t - \ell) + v(t) \quad (1)
\]

As we want to estimate the channel impulse response \( \hat{h} = [h(0)\ldots h(T)]^T \), it is more convenient to rewrite (equation) in the following vector-form:

\[
\mathbf{Y} = \mathbf{Xh} + \mathbf{V} \quad (2)
\]

where \( \ell + 1 \) is the assumed channel length, \( T = T_0 + \delta T \leq T_i \) is the number of symbols used to estimate the channel, \( \mathbf{X} \) is the \( (T + \ell) \times (\ell + 1) \) filtering matrix, \( \mathbf{V} = [v(0)\ldots v(T + \ell)]^T \) is the noise vector and \( \mathbf{Y} = [y(0)\ldots y(T + \ell)]^T \) the observation vector. Note that the filtering matrix includes both training and information symbols since we assume that all the symbols of the transmitted blocks are in the same alphabet.

2.3. Review of the bootstrap channel estimation procedure and problem statement

The idea of bootstrap estimation is to make use of the entire block of symbols to estimate the channel. In practice, the corresponding symbols \( \{x(t), t \in [0, T_i - 1]\} \) are perfectly known for \( t \in [0, T_i - 1] \) and estimated for \( t \in [T_0, T_i - 1] \). In order to estimate the channel, the observation vector is therefore approximated as follows:

\[
\mathbf{Y} \approx \hat{\mathbf{X}} + \mathbf{V} \quad (3)
\]

where \( \hat{\mathbf{X}} \) is the estimated version of the matrix \( \mathbf{X} \), which is obtained after decoding and re-encoding. Of course, the iteration process can be repeated several times and here the matrix \( \hat{\mathbf{X}} \) corresponds to the estimated symbols from the last iteration. The Least-Squares channel estimate of the channel is given by:

\[
\hat{h} = \arg\min_{\hat{h}} \| \mathbf{Y} - \hat{\mathbf{X}} \hat{h} \|_2 \quad (4)
\]

It is well known that LS channel estimation accuracy increases with the number of training symbols. This still assures perfectly reliable training symbols, which is not the case for bootstrap estimation because the extended training sequence used to perform channel estimation comprises some decision errors. For instance, if the BER target is 0.1% and the BLER target is around 1%, then 10% of the symbols of a "false block" correspond to wrong decisions. This would mean that in certain situations, there would be a trade-off between the extended TS size and the "average symbol reliability" to be found. In particular, in order to gain a more complete understanding of this kind of phenomena, we propose to theoretically evaluate the performance of the bootstrap estimator. To this end, we first detail the assumptions needed to this calculation.

3. ASSUMPTIONS

Assumptions (A1) to (A4) correspond to the main hypotheses we have made so far. On the other hand, assumptions (A5) to (A8) are additional assumptions that are needed to derive the expressions of the channel estimation error variance. Assumption (A9) is used to get a simplified expression of the estimation error (subsection 4.4) and assumption (A10) is just used to better interpret the provided results (section 5).

\[
\begin{align*}
\text{(A1)} & \quad \text{Transmitted symbols are independent and identically distributed.} \\
\text{(A2)} & \quad \text{BPSK modulation is assumed.} \\
\text{(A3)} & \quad \text{The vector channel is considered as a random process but is assumed to be constant over a block duration.} \\
\text{(A4)} & \quad \text{The coefficients of the channel impulse response are decorrelated.} \\
\text{(A5)} & \quad \text{The vector error is assumed to be independent of the channel.} \\
\text{(A6)} & \quad \text{The vector error is assumed to be independent of the noise vector.} \\
\text{(A7)} & \quad \text{In average, the error is uniformly distributed over the data block.} \\
\text{(A8)} & \quad \text{The block size is assumed to be very large: } T >> \ell, \\
\text{(A9)} & \quad \text{which means that } X^\dagger X \approx I_{T_i} \text{ and } \hat{X}^\dagger \hat{X} \approx I_{T_i}. \\
\text{(A10)} & \quad \text{with } \delta T \equiv T - T_0 \\
\end{align*}
\]
4. ESTIMATION ERROR VARIANCE OF THE BOOTSTRAP ESTIMATOR

4.1. Channel estimation error

First of all, it is useful to notice that under assumption (A2) the estimated filtering matrix \( \tilde{X} \) writes simply as
\[
\tilde{X} = X + \sigma \delta X
\]
where the entries of the error matrix are in the set \(-2, 0, +2\). Using (2), (4) and assumptions (A1-A2) gives the following expression of the channel estimate:
\[
\tilde{h} = (X^T \tilde{X})^{-1} \tilde{X}^T y = (T_{\tilde{XX}})^{-1}(X^T + \sigma \delta X)^T (Xh + V).
\]
(5)
Finally the channel estimation error is:
\[
\delta h = \tilde{h} - h = T^{-1}X^T V + T^{-1}X^T \delta X h + T^{-1} \delta X^T V,
\]

4.2. Mean of the channel estimation error

As we know that the channel noise is independent of the transmitted symbols and symbols errors (A6), we can write:
\[
E(\tilde{h}) = -2T^{-1}E(n)E(h)
\]
where \("n\" is the number of errors in the considered detected data block. We see that the estimation bias does not vanish if and only if
- the bit error rate is not zero and
- the channel expectation is not zero.
As the channel is generally a central random process, the estimation bias is always zero independently of the bit error rate value.

4.3. General expression of the channel estimation error variance

Based on assumption (A6), it can be shown that
\[
E(\delta h) = E(\tilde{h}^{T} \delta h) = E(\tilde{h})^{T} E(\delta h) + E(h) E(\tilde{h})^{T} E(\delta h) = E(\epsilon) + \epsilon
\]
with \(\forall (i, j) \in [1, 3]^2, \epsilon_{ij} = E(E_{ij}^{T} E_{ij})\).

Under assumption (A2) the error terms \(\epsilon_{11}, \epsilon_{13}, \epsilon_{14}, \text{ and } \epsilon_{33}\) are easy to evaluate and can be expressed as follows:
- \(\epsilon_{11} = \sigma^2 T^{-1} E(h_{11}^2) = \sigma^2 \frac{T + 1}{T}\)
- \(\epsilon_{13} = \sigma^2 T^{-1} E(h_{13}^2) = -2 \sigma^2 \frac{T}{T} (\ell + 1) E(n)\)
- \(\epsilon_{14} = \sigma^2 T^{-1} E(h_{14}^2) = -2 \sigma^2 \frac{T}{T} (\ell + 1) E(n)\)
- \(\epsilon_{33} = \sigma^2 T^{-1} E(h_{33}^2) = 4 \sigma^2 \frac{T}{T} (\ell + 1) E(n)\)
The last error term \(\epsilon_{22}\) is more difficult to evaluate. We will only provide the main steps of the proof. First of all, it is very convenient to notice that this error term writes as:
\[
\epsilon_{22} = E(\tilde{h}^2) = T^{-2} E(\tilde{h}^2) = T^{-2} E(T_{\tilde{XX}}^2 X^T \delta X^T \delta X + \delta X^T \delta X^T)
\]
\[
= T^{-2} E \left[ T^{-2} E \left[ \sum_{i=1}^{3} E_{ij}^2 \right] \right] = T^{-2} E \left[ \sum_{i=1}^{3} E_{ij}^2 \right] = T^{-2} \sum_{i=1}^{3} E_{ij}^2
\]
where the vectors \(u_j\) and the scalar constants \(E_j\) are respectively defined by \(\delta X^T X = [u_1, \ldots, u_3]\) and \(E_j = E[h_{ij}]\). It can be checked that entries \(h_{ij}, i, j \in [0, \ell]\) of each vector \(u_j\) correspond to an inner product between two vectors (say \(v_k\) and \(v_{k+1}\)) which depends only on the difference \(k - j\). Vectors \(v_k\) and \(v_{k+1}\) are defined such that
\[
\forall k \in [0, T - 1], v_k(e(k) \in [0, 2, +2], E[e(k)] = 4 \beta \quad \text{and} \quad \forall k \in [d, T - 1 + d], E[e(k)] = 0,
\]
which leads then to:
\[
E(\tilde{h}^2) = 4 E(n^2) + 4 \beta^2 \sum_{i=1}^{3} E_{ij}^2
\]
where \(n\) is the number of errors in a given block of symbols and \(\beta\) is the bit error rate. Finally, if we plug this relation into equation (3), it can be seen that the error term \(\epsilon_{22}\) also depends on the first and second central moment of the channel power-delay profile as:
\[
\epsilon_{22} = \frac{4 E(n^2)}{T} \left[ T^{-1} E(n) + \epsilon \delta h \right] - \frac{1}{2} \left[ T^{-1} E(n) + \epsilon \delta h \right] + \frac{1}{2} \left( T^{-1} E(n) + \epsilon \delta h \right)^2
\]
with \(j = \frac{\sum_{i=1}^{3} E_{ij}}{\sum_{j=1}^{3} E_{ij}}\) and \(\tilde{j} = \frac{\sum_{i=1}^{3} j E_{ij}}{\sum_{j=1}^{3} E_{ij}}\).

From this equation it is seen that the channel estimation error variance depends both on the channel and number of errors statistics. The general expression of the estimation error variance is then obtained by summing the different error terms of equation (3). In order to better interpret these results, we will assume that \(\delta T^2 \gg \ell^2\). Under this condition, the expression of the estimation error can be considerably simplified as indicated below. This assumption is in fact not very restrictive. In practice, it is seen that the influence of the error terms associated with the moments of the channel is not significant.

4.4. Simplified expression of the channel estimation error variance

By assuming that \(\delta T^2 \gg \ell^2\) it is easily found that the MSE expression boils down to:
\[
E(\tilde{h}^2) = \sigma^2 (\ell + 1) + \frac{4 E(n)}{T} \left[ E(n^2) + \epsilon \delta h \right]
\]
(9)
By introducing the notations \(\beta = E(n / \delta T)\), \(\beta^2 = E(n^2 / \delta T^2)\), \(\rho = E[n^2 / (\ell + 1)^2]\) and \(\tau = \delta T / T\), that is to say that \(\delta T^2 = \ell^2 \tau^2\), \(E(n) = \beta T \tau\) and \(E(n^2) = \beta^2 T^2 \tau^2\) we get a closed-form expression of the estimation error
\[
\epsilon(\tau) = \sigma^2 \frac{T + 1}{T} \times \frac{1 + a(\tau) \tau}{1 + \tau}
\]
\[
a(\tau) = 4 \rho \frac{\tau + \beta}{1 + \tau}
\]
(10)
We can see that if the "extended TS" is large enough the bootstrap estimator performance does not depend on the
moments of the channel power-delay profile anymore. Similarly to the conventional training-based estimator, it is related to the channel only through its length $\ell + 1$. From equation (10), we also notice that the MSE converges to a finite value that is

$$\lim_{r \to \infty} \varepsilon(r) = 4\beta^2$$

(11)

where we assumed a normalized channel power that is $E[h^2] = 1$.

5. DISCUSSION AND SIMULATION RESULTS

Equation (10) can be used even if $\tau = 0$ small provided condition (A8) is met, which implies that condition (A10) has to be met. From now on, let assume that the latter is met. Then the estimation error taken in $\tau = 0$ namely $\varepsilon(0) = [\sigma^2(\ell + 1)]/T_0$ coincides with the conventional training-based estimator performance. This means that the bootstrap estimator performs better than the conventional training-based estimator if only and only if $d(\tau) < 1$. It turns out that this is not always the case. This depends on the value of the channel length, the training sequence size, the signal-to-noise ratio, the mean of the number of errors (bit error rate) and its variance. Figure 4 depicts the normalized MSE ($\varepsilon(\tau)/\varepsilon(0)$ in [dB]) as a function of the normalized number of symbols that are used to extend the training sequence ($\delta T/T_0 = \tau$) for different scenarios described in the table below:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\ell + 1$</th>
<th>$T_0$</th>
<th>SNR,</th>
<th>$\beta$</th>
<th>$\beta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>5</td>
<td>50</td>
<td>0</td>
<td>10.5%</td>
<td>1.1%</td>
</tr>
<tr>
<td>S2</td>
<td>4</td>
<td>50</td>
<td>0</td>
<td>15.9%</td>
<td>2.6%</td>
</tr>
<tr>
<td>S3</td>
<td>5</td>
<td>50</td>
<td>0</td>
<td>22%</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

Table 1: scenarios definition

The three solid curves correspond to simulation results and the dashdot curves represents the asymptotic MSE given by (10).

The values of $\beta$ and $\beta^2$ have been generated by making hard decisions on the log-likelihood ratios $\log[P(y(t)/x(t) = 1)/P(y(t)/x(t) = -1)]$, which can be modeled [5] by Gaussian random variables $N(0.5\sigma_x^2, \sigma_y^2)$. The value of $\sigma_y^2$ depends on the system under consideration. In the table, for scenarios (S1-S3), $\sigma_y^2$ respectively equals 2.5, 2 and 1.5. The multipath channels used for simulations comprise $\ell + 1$ Rayleigh fading taps with equal energy. From figure 4, we see that asymptotic analysis of section 4 leads to a very good approximation of the bootstrap estimation MSE. Regarding this aspect, it can be checked that if $\delta T > T_0 = 50$ then the main diagonal of matrices $X^T X$ and $\tilde{X}^T \tilde{X}$ contains more than 95% of its total energy, which means that the main assumptions (A8) is well verified.

From figure 4, it is also seen that the bootstrap estimator does not always perform better than the conventional training-based estimator. Scenario 2 even shows an optimum number of training symbols to be added to the initial TS. The essential question that arises is that to know to what extent these types of scenarios can be encountered in the real life. The main issue is about the values of the mean and variance of the number of errors namely $\beta$ and $\beta^2$. First of all, attention of the reader is drawn upon the fact that $\beta$ can be either the bit error rate (BER) or the ratio of the bit error rate to the block error rate (BER/BLER). The latter situation occurs when only the false blocks are considered. Indeed, as the BLER is generally the parameter of interest for higher layers, a possible reception strategy can consist in optimizing the BER for a given BLER target. In that case, the channel estimation algorithm is reapplied to the false blocks only, which generally contain a large amount of errors (see also end of section 2). This is one of the reasons why it makes sense to consider values of $\beta$ as high as 10% or 20% (depending on the system), which are often obtained in practice. Regarding the variance of the number of errors it is even more dependent on the system under consideration (propagation environment, channel coding, interleaving size, etc). Similarly to the BER $\beta^2$ can be calculated in some special cases. For example, for an uncoded transmission over a flat-fading channel ($\ell = 0$), it is ready to show that $\beta^2 = 1/4 - 1/(\tan^{-1}(\sqrt{s}))$ with $s = (\rho + 1)/\rho$. In this case, the maximum value of $\beta^2$ is 25%. In any case, $\beta^2$ is greater than $\beta$ (by definition).

At last, we will discuss the impact of the channel estimation error on the receiver performance. For this purpose we consider a Viterbi equalizer. In [6], Gorokhov studied the signal-to-noise (SNR) degradation due to channel estimation errors. Under certain (realistic) conditions, which are assumed to be met here, the SNR loss can be expressed through the following relation:

$$SNR_{\text{req}} = SNR \times \left[1 + \left(1 + 1\varepsilon(\tau)\right)^{-1}\right].$$

Using this relation and scenarios definitions from table 1 we can compute the equivalent SNR as a function of the number of additional training symbols. The corresponding curve is represented in figure 5. In scenario 1 around 1 dB is lost by using all the symbols of the block (50+450). In scenario 2, 0.3 dB can be gained by adding the optimum number of training symbols (around 25 more symbols) and in scenario 3, the conventional bootstrap estimator (using the entire block of symbols) allows us to gain 0.8 dB on the SNR even if it is applied on false blocks.

6. CONCLUSIONS

In this paper, we have evaluated the bootstrap estimator performance under asymptotic assumptions for a BPSK modulation. The corresponding assumptions proved to be not very limiting: theoretical and experimental MSE curves are very close (figure 4). We showed that the best strategy is not always to add the maximum number of decoded symbols to the initial training sequence. The best scheme to select depends in particular on the BER and the BLER. If the BLER is fixed, it is better in certain scenarios not to use all the decoded symbols or not to use them at all. This allows the receiver to avoid a signal-to-noise ratio degradation due to the large number of errors present in false blocks. Therefore, a good strategy can be to apply bootstrap estimation on all the received blocks to achieve the desired BLER and then to apply the conventional TS-based estimation on the false blocks. For a Viterbi equalizer, it was shown that this strategy can provide a 1 dB gain on the SNR.
could be useful to extend the provided results to higher order modulation and propagation channels with correlated taps.

could be useful to extend the provided results to higher order modulation and propagation channels with correlated taps.

Figure 4: possible behaviors of the bootstrap estimator

Figure 5: SNR loss due to bootstrap channel estimation

REFERENCES


