Training-based Channel Estimation and De-noising for the UMTS TDD Mode

S. Lasaulce\textsuperscript{1}, P. Loubaton, E. Moulines and S. Buljore
\textsuperscript{1}MOTOROLA Research Center – Paris
91193 Gif-sur-Yvette
FRANCE
lasaulce@crm.mot.com

Abstract- In this paper we provide a theoretical framework of de-noising for UMTS TDD-like mobile radio communication systems. Based on the Bayesian approach, we show how to de-noise channel estimates provided by the Training-based estimation procedure. The proposed schemes allow not only for eliminating major drawbacks of hard thresholding but also for a low complexity implementation.

I. INTRODUCTION

In mobile radio communication systems, there is a need to estimate the effect caused by the radio communication channel on the data being transmitted. The most conventional strategy in estimating radio channel propagation conditions is to model the radio channel by a Finite Impulse Response and then to perform the channel estimation by transmitting a known data sequence (the Training Sequence). For practical reasons due to unknown propagation environments, an assumed length of the channel has to be used.

In the UMTS TDD (Time Division Duplex) mode, for the aforementioned reason and a complexity constraint \cite{1}, the assumed channel length of the overall channel (emission filter, propagation channel and reception filter) is taken to be 57 chips in the uplink and 64 chips in the downlink.

For a typical vehicular propagation environment (say the ETRVA channel specified by the ETSI), the propagation channel length is approximately 10 chips. When including the emission and reception filters the channel length extends to approximately 19 chips. In the uplink, this equates to 67\% of the estimated coefficients being pure noise. Such an introduction of noise into the data recovery process is known to significantly affect the performance of the symbol detector. For a MMSE Block Linear Joint Detector \cite{2} operating at typical Bit Error Rates (less than 5\% in the uncoded case and voice service), there is at least a 3 dB loss on the Signal to Noise Ratio (SNR) because of the flawed channel knowledge \cite{3}.

To bridge the huge gap between performance provided by the trained case and those provided by the case where the channel is known from the receiver, one possible approach is to de-noise the trained channel estimates before performing the data detection step. The purpose of this paper is thus threefold:

- To provide a theoretical and efficient tools to de-noise the trained channel estimates
- To provide algorithms for which complexity is kept reasonable
- To show how the new algorithms perform in comparison with other classical estimation schemes in a realistic context (TDD mode specifications)

II. THE STRONGEST-PATH METHOD

A. Training-based estimation in the TDD mode

In the European option of the TDD mode, referred as the wide band TDD mode, each frame is divided into 15 time slots (667 \(\mu\)s). In each time slot, the communication resource is shared by allocating each communication a spreading code (Code Division Multiple Access). For each user, one part of the time-slot, which is called midamble, is dedicated to the channel estimation purpose.

In the more general case of the uplink, the m-dimensional signal received by one antenna (one sensor) at the base station can be expressed as:

\[
Y = S g + \mathcal{V}
\]

where \(S\) is a m by K(l+1) matrix constructed from the midambles of the K active users per time slot. As for \(g\), it is obtained by stacking the (l+1)-chip discrete channel impulse responses of the various users:

\[
g = [g_1^T,...,g_{k+1}^T]^T
\]

\(V\) is a complex additive white and Gaussian noise of variance \(\sigma^2\).

Therefore, the Maximum Likelihood estimate of the channels based on the observation \(Y\) coincides with the Least-Squares solution:

\[
\hat{g} = (S^H S)^{-1} S^H Y
\]

In the sequel, (3) will be referred as the trained estimate of the channels. At last, estimation performance of the trained procedure can be characterized by:

\[
E[\|\hat{g} - g\|^2] = \sigma^2 Tr(R_s^{-1})
\]

with \(R_s = S^H S\).

B. Hard thresholding

According to (4), the estimation error associated with a given channel coefficient \(g(i)\) is:

\[
E[\|\hat{g}(i) - g(i)\|^2] = \sigma^2 u_i^T R_s^{-1} u_i
\]

with \(u_i = (0...010...0)^T\).

If instead of estimating \(g(i)\) by \(\hat{g}(i)\), it was estimated by 0, then the estimation error would be the energy of the coefficient under consideration. This means that if this energy is less than the
estimation noise variance, it is better to replace by 0 the trained estimate of the considered coefficient. This is known as the hard thresholding scheme [4].

The strongest-path method, which is equivalent to a hard thresholding strategy, consists of selecting only the coefficients of the channel estimate that are the most significant in terms of energy. Although this kind of approach is very used in the real life because it is easy to implement, it has at least four major drawbacks:

I. There is no theoretical ground to select the threshold value or the number of path.

II. It is generally not consistent: the channel coefficients whose energy are found to be beneath the threshold are systematically discarded from the channel impulse response, even under good propagation conditions (high signal to noise ratios in particular).

III. A further problem is apparent in that, after selection, the coefficients remain noisy as estimation noise is not removed.

IV. A yet further problem is that hard thresholding does not exploit fully the statistics of the channel.

III. NEW ESTIMATION SCHEMES BASED ON THE BAYESIAN APPROACH

A. Assumptions

In order to design more efficient de-noising strategies, we will exploit one or several of the following assumptions:

(a) The channel is considered as a Gaussian and complex random process (Rayleigh fading).

(b) There is a basis where the channel can be described by only a few components.

(c) Channel coefficients are not correlated.

(d) Training sequences have ideal auto-correlation and inter-correlation properties (\(R_y = mI\)).

Depending on the desired performance level, the complexity constraint imposed by the system and the channel nature we show, in the next section, what is the best scheme to select.

B. The corresponding schemes

As the proposed estimators are based on the Bayesian approach, assumption (a) will hold for all the sequel. Depending on the validity of (b), (c) and (d), different estimators can be designed. In order to find the desired trade-off between performance and complexity, we tackle the estimation problem in two steps. First, we focus on performance improvements (sections III.B.1 and III.B.2). Second, from the designed estimators we propose several ways to reduce complexity (sections III.B.3 and III.B.4). Assuming (a), eight denoising schemes are possible, which are indicated in Table I.

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>True = 1</td>
<td>S1</td>
</tr>
<tr>
<td>False = 0</td>
<td></td>
</tr>
<tr>
<td>1 0 0 0</td>
<td>S1</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>S2</td>
</tr>
<tr>
<td>1 0 1 0</td>
<td>S3</td>
</tr>
<tr>
<td>1 0 0 1</td>
<td>S4</td>
</tr>
<tr>
<td>1 1 1 0</td>
<td>S5</td>
</tr>
<tr>
<td>1 1 0 1</td>
<td>S6</td>
</tr>
<tr>
<td>1 0 1 1</td>
<td>S7</td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>S8</td>
</tr>
</tbody>
</table>

1) The conventional Bayesian approach (S1). It turns out that a good way to de-noise channel estimates by using statistics of the channel is to apply a Bayesian approach in a spirit similar to [5], which corresponds to assume that assumption (a) holds.

As the hard thresholding procedure can be viewed as a binary weighting procedure [4], it is likely that estimation performance can be improved by extending the weights set \([0,1]\) to the continuous interval \([0,1]\). By considering the channel as a random process, one can easily show that looking for the weighting matrix (W) minimizing the estimation error of the de-noising procedure

\[
W = \arg \min_{\mathbf{w}} E[\|\hat{\mathbf{g}}_w - \mathbf{g}\|^2] \quad (5)
\]

is equivalent to maximize the A Posteriori Probability (APP) \(P(\hat{\mathbf{g}}/\mathbf{Y})\) when the channel is Gaussian:

\[
\hat{\mathbf{g}} = W \hat{\mathbf{g}} = \arg \max_{\mathbf{z}} P(\mathbf{g}/\mathbf{Y}) \quad (6)
\]

where \(W\) is expressed by

\[
W = [R_y + \sigma^2 R_x^{-1}]^{-1} R_y \quad (7)
\]

and at last

\[
R_y = E[\mathbf{g} \mathbf{g}^H] \quad (8)
\]

In the sequel, we assume that assumption (a) holds that is the best de-noising scheme consists of maximization of the APP.

2) The low rank approximation (S2). Additionally, if (b) is verified, performance of the Bayesian approach-based estimator can be improved by reducing the number of parameters to estimate. In such a basis, \(\mathbf{g}\) is a linear combination (\(\gamma\)) of a low number (say \(p\)) of \(p\) column vectors from the matrix \(\mathbf{P}\) such that:

\[
\mathbf{g} = \mathbf{P} \gamma \quad (9)
\]

For each time slot, only \(\gamma\) is estimated whereas \(\mathbf{P}\) is assumed constant over the \(M\) time slots used for the estimation procedure.

By maximizing \(P(\mathbf{\gamma}/\mathbf{Y})\) with respect to \(\mathbf{\gamma}\), one shows that:
The estimator $\hat{Y}$ can be constructed from the most significant eigenvectors of $P$. Specifically, $P$ can be approximated by:

$$\hat{Y} = SP\gamma^* + \bar{V}$$

where

$$R_\gamma = E[\gamma\gamma^*]$$

As for the optimum choice of $P$ that is the best basis for representing a Gaussian random vector by a minimum number of vectors, [6] shows that it is the Karhunen-Loeve (KL) basis. Thus $P$ can be constructed from the most significant eigenvectors of $R_\gamma$.

If (9) is not perfectly true, which is the general case, one can approximate the received signal by:

$$Y = SP\gamma + \bar{V}$$

which is what we call low rank approximation.

3) How to exploit the correlation properties. In IIIB.1 and IIIB.2, we see that the proposed channel estimators exploit fully the statistics of the channel. However, a high additional computational cost has to be paid. Indeed, according to (7) and (10), in the proposed estimation procedures there is the need to perform large dimension matrices inversions and eigen decompositions. Typically, in the uplink of the TDD mode, for 8 active users the size of the channel is $8 \times 57 = 456$ chips, which is the dimension of $R_\gamma$.

But complexity of the two proposed channel estimation procedures can be significantly reduced in contexts where assumptions (c) and (d) are well verified. For estimation strategies corresponding to S3, S4 and S6, complexity reduction is still very small because inversion and eigen decomposition have to be performed.

The most interesting estimation schemes for reducing complexity while maintaining good performance are obtained in the scenarios S5, S7 and S8. For S5, $R_\gamma$ is diagonal, which means that the Karhunen Loeve basis is the canonical basis. Therefore, no eigen decomposition is needed. For S6, which is the most favorable case, $R_\gamma$ and $R_i$ are both diagonal. Consequently, the matrix weighting procedure associated with (7) boils down to a one-dimensional scalar weighting procedures, which means that each channel coefficient is multiplied by a reliability factor. For S7, it is easy to show that the corresponding estimation strategy is equivalent to a hard thresholding case where the selected coefficients are weighted by the “Bayesian reliabilities”. In the next section we show how to implement the ODB estimator (S6).

4) One-dimensional Bayesian estimation (ODB) (S6). From (7), it is easily found that each coefficient of the channel impulse response estimate is weighted by:

$$\tilde{\delta}^{(i)}(k) = \frac{E_i^{(k)}}{\sum_{i=1}^{K} E_i^{(k)}}$$

where

$$E_i^{(k)} = E\|g_i^{(k)}\|^2$$

and $\sigma_{\alpha}^2(i) = \sigma_{\alpha}^2 R_i^{-1} R_i^{-1}$. (14)

Of course, in real situations, both the energies of the channel coefficients and the noise variance have to be estimated.

But even in the case where the noise variance is assumed known and the number of samples is large enough (high number of time slots), (13) has a drawback: as estimation of the coefficients energies is performed from the estimated coefficients, they also include noise energy. Specifically, for high path delays, which are the paths for which there is almost no signal, $\delta^{(i)}(k)$ tends to 1/2. In order to avoid this pitfall, we propose to modify (13) as follows:

$$\tilde{\delta}^{(i)}(k) = \frac{\hat{E}_i^{(k)} - \sigma_{\alpha}^2}{\hat{\sigma}_i^{(k)}}$$

with

$$\hat{E}_i^{(k)} = (1/M) \sum_{m=1}^{M} \|g_i^{(k)}(i,m)\|^2$$

“m” being the time slot index.

One suitable way to estimate the noise variance is to make use of the less significant paths. It is especially true when channel length is overestimated, which is often the case in real life. For instance, in the uplink, using the queue of the channel estimates gives:

$$\hat{\sigma}_i^{(k)} = (1/KQ) \sum_{q=1}^{Q} \hat{E}_{(i,k,q)}^{(k)}$$

where $K$ is the number of active users, $Q$ the number of channel coefficients assumed as pure noise (for a given user) and $\{i(k,q)\}$ are the corresponding indices.

IV. SIMULATION EXAMPLES

A. Simulation setup

The simulator used follows the ETSI specifications [7] for the TDD physical layer. Propagation environment tests are: the ETRIB (Indoor B, max. delay = 700 ns), the ETRVA (Vehicular A, max. delay = 2510 ns) and the ETRPB (Pedestrian B, max. delay = 3700 ns). Simulations have been done in the uplink and 4 active users have been assumed. One single antenna is assumed at the base station. There is no channel coding. Speed of the various users is fixed at 3 km/h. At last, Bit Error Rates are evaluated over 1000 time slots, each of them comprising 122 QPSK symbols and a 512-chip midamble.

B. Simulation results

In this paper, we provide only simulation results for the ODB de-noiser developed in IIIIB.4. More results are provided in [8].
This gain is even more significant for higher signal to noise ratios, which is the "consistency effect".

It is also of interest to see what are the influence of the number of available time slots (M) and the propagation environment type. In Table 1, the ODB estimator is compared with the conventional trained estimator, we see that the SNR gain at for a 2% BER ranges from 1.6 dB to 2.2 dB for 11 and 61 time-slots.

As for the influence of the propagation environment, we see in Table 2 that longer the channel impulse response is, higher is the SNR gain. This could be explained by the fact that ETRPB-like channel also comprise medium-energy paths that are almost systematically neglected in the hard thresholding procedure. It can also be noticed that the Bayesian de-noiser is compared with the hard one.

In the framework of continuous time slotted transmissions, we showed that the best way to de-noise the trained channel estimates is to use the Bayesian approach.

We proposed de-noising schemes both efficient from the statistical point of view and easy to implement. Depending on the context (channel coefficients correlated or not, ideal Training Sequences or not, sparse channels or not), we saw how to choose between the different de-noising schemes in order to obtain the desired trade-off between performance and complexity. Note also that in the proposed estimation procedures, almost all the drawbacks (I, II, III and IV) of the hard thresholding procedure have been removed.

Simulations results performed in the UMTS TDD mode show that the one-dimensional Bayesian de-noiser allows for a 1 dB SNR gain over the hard thresholder (at BER=2%), which means that the one-dimensional Bayesian post-processing provides a 2 dB gain to the trained channel estimation.

**REFERENCES**


**TABLE 1 : INFLUENCE OF THE NUMBER OF TIME SLOTS ON THE SNR GAIN OF 1-D BAYESIAN ESTIMATOR OVER THE TRAINED ESTIMATOR**

<table>
<thead>
<tr>
<th>M</th>
<th>11</th>
<th>21</th>
<th>61</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain [dB] at BER=2%</td>
<td>1.6</td>
<td>2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**TABLE 2 : INFLUENCE OF THE CHANNEL TYPE ON THE SNR GAIN OF 1D BAYESIAN ESTIMATOR OVER THE HARD THRESHOLDER**

<table>
<thead>
<tr>
<th>Environment</th>
<th>ETRIB</th>
<th>ETRVA</th>
<th>ETRPB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain [dB] at BER=2%</td>
<td>0.7</td>
<td>1.2</td>
<td>1.6</td>
</tr>
</tbody>
</table>

**FIG. 1 : INFLUENCE OF THE ESTIMATION ACCURACY ON THE MMSE-BL JOINT DETECTOR**

**V. CONCLUSIONS**