Parameter Estimation of a Convolutional Encoder from Noisy Observations

Janis Dingel and Joachim Hagenauer
Institute for Communications Engineering (LNT), Munich University of Technology (TUM)
80290 München, Germany
Email: {janis.dingel, hagenauer}@tum.de

Abstract— **We consider the problem of estimating the parameters of a convolutional encoder from noisy data observations, i.e. when encoded bits are received with errors. Reverse engineering of a channel encoder has applications in cryptanalysis when attacking communication systems and also in DNA sequence analysis, when looking for possible error correcting codes in genomes. We present a new iterative, probabilistic algorithm based on the Expectation Maximization (EM) algorithm. We use the concept of log-likelihood ratio (LLR) algebra which will greatly simplify the derivation and interpretation of our final algorithm. We show results indicating the necessary data length and allowed channel error rate for reliable estimation.**

I. Introduction

We address a problem strongly related to cryptanalysis and data security. In a reverse engineering context, an observer wants to extract the transmitted information from a received data stream without knowing all the parameters of the transmission. The observed signal may have been corrupted by noise during transmission. Even without employing advanced protocols from cryptology, modern communication systems are hard to decipher if the parameters of the different elements of the transmission chain are not or only partially known. Research on the reverse engineering of a channel encoder, as a special subproblem, has been conducted for communication systems [1], [2], [3], [4] and also for DNA sequences [5], looking for possible error correcting codes in the genetic code. Most of these approaches concentrate on linear block codes. In this paper, we derive a new algorithm for the estimation of encoder parameters of a convolutional code from a noisy data stream. This problem has been considered before by Rice [3] and later by Filiol [2], where an algebraic estimation procedure has been proposed. A candidate is recovered from a subsequence of bits that is hopefully unaffected by noise and then tested for significance on the whole observed sequence. However, if no noisefree subsequence exists for which the parameters can be algebraically recovered, the method will fail. A method for the reconstruction of punctured convolutional codes has also been presented by Filiol [6]. Here, we introduce an iterative, probabilistic approach based on the Expectation Maximization (EM) algorithm. The EM algorithm is a strong tool that has proven useful in many communications and signal processing problems such as blind channel estimation [7] and system identification [8], which are related to the one considered here. In fact the problem is essentially the same as estimating an unknown hidden markov model, which is typically done with the Baum-Welch algorithm [9]. However, here we investigate systems where computations are carried out in a finite field which requires different methods from those developed for traditional blind estimation scenarios. The latter implies that an approach has to combine concepts from both, coding theory and blind signal processing. In a similar way, the EM algorithm has been applied to the synthesis of linear feedback shift-registers from noisy sequences [10]. However, here we use the concept of log-likelihood ratio (LLR) algebra, introduced in [11], which will greatly simplify the derivation and interpretation of our final algorithm. We achieve this by transforming the problem into the LLR domain, a natural step when looking at it from a coding perspective.

The paper is outlined as follows. In Section II, we will specify and formalize the problem. Section III introduces the application of the EM algorithm and Section IV shows the transformation to the LLR domain and the derivation of the estimation algorithm. Section V discusses simulation results before we conclude with Section VI.

II. Problem Statement

We assume that an information stream $\boldsymbol{u} = [u_0, u_1, \ldots, u_t, \ldots]$, $u_t \in \mathcal{A}$ is encoded with a convolutional encoder of memory M. Here we will consider only codes defined over $GF(2)$ with elements $\mathcal{A} = \{+1, -1\}$, where $+1$ is the "null" element under the $\oplus$ addition. We present our approach for codes of rate $\frac{1}{n}$ and we exclude encoders with feedback. However, the algorithm can be generalized to any rate $\frac{k}{n}$. In Figure 1, the situation for the ith output of the convolutional encoder is shown. The parameters $\boldsymbol{g}^{(i)} = [g_0^{(i)}, \ldots, g_M^{(i)}]$ determine how the

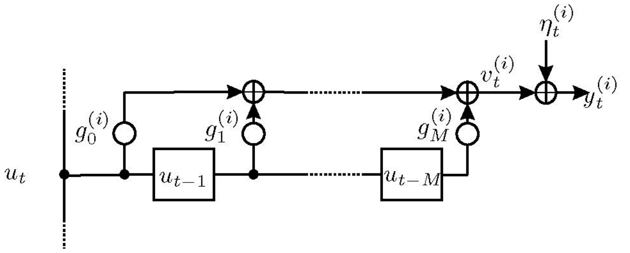


Fig. 1. The ith output of the convolutional encoder, $v_t^{(i)}$, is transmitted over an additive, memoryless channel resulting in the observation $y_t^{(i)}$.

1-4244-1429-6/07/$25.00 ©2007 IEEE

information symbol u_t and the content of the memory elements $[u_{t-1}, \ldots, u_{t-M}]$, at time t, are mapped to the encoded symbol $v_t^{(i)}$ of output i, $i = 1, \ldots, n$. Throughout this paper, we assume that the received symbol $y_t^{(i)}$ is observed through a BSC, i.e. $y_t^{(i)} = v_t^{(i)} \oplus \eta_t^{(i)}$. The state of the encoder $\boldsymbol{s}_t$ will be defined as the concatenation of the current input symbol with the content of the memory elements, i.e. $\boldsymbol{s}_t = [s_{t_0}, \ldots, s_{t_M}] = [u_t, u_{t-1}, \ldots, u_{t-M}]$. Then we can write

$$y_t^{(i)} = \sum_{k=0}^{M} \oplus s_{t_k} g_k^{(i)} \oplus \eta_t^{(i)}, \tag{1}$$

where multiplication and addition are in $GF(2)$. Note that the ith output only depends on the ith parameter subset $\boldsymbol{g}^{(i)}$ but that the outputs are not independent as each output at time t is determined by the same encoder state $\boldsymbol{s}_t$. Such an encoder is specified by the vector $\boldsymbol{g} = [\boldsymbol{g}^{(1)}, \ldots, \boldsymbol{g}^{(n)}] \in \mathcal{A}^{n(M+1)}$, which are the parameters we want to estimate from the observation $\boldsymbol{y} = [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(n)}]$, where $\boldsymbol{y}^{(i)} = [y_1^{(i)}, \ldots, y_T^{(i)}]$ and the length of $\boldsymbol{y}$ is nT. The likelihood for the observation $p(\boldsymbol{y}; \boldsymbol{g})$ depends on the deterministic parameter $\boldsymbol{g}$. [1]

III. Maximum Likelihood estimation via EM-Iteration

The Expectation Maximization (EM) Algorithm is a powerful concept for iteratively approaching the maximum likelihood solution in what are called incomplete-data problems [8]. Let $\boldsymbol{y}$ be the observed data that depends on the unobservable data $\boldsymbol{x}$ and a deterministic parameter $\boldsymbol{\theta}$ that we want to estimate. The so called missing data $\boldsymbol{x}$ is modeled as a random variable. Instead of obtaining the maximum likelihood for $\boldsymbol{\theta}$ only from the observed data, the complete data $(\boldsymbol{x}, \boldsymbol{y})$ is incorporated in an iterative procedure [8]. With a fixed current estimate on the parameter, $\boldsymbol{\theta}^{[k]}$, the expected value $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{[k]})$ of $\log(p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}))$ is evaluated with respect to $\boldsymbol{x}$ and conditional on $\boldsymbol{y}$ in the so called E-Step. In the M-step of the algorithm, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{[k]})$ is maximized with respect to $\boldsymbol{\theta}$ to yield a new estimate $\boldsymbol{\theta}^{[k+1]}$ in iteration $k+1$

$$\boldsymbol{\theta}^{[k+1]} = \arg\max_{\boldsymbol{\theta}} \underbrace{\left\{ \sum_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\theta}^{[k]}) p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}).) \right\}}_{Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{[k]})}. \tag{2}$$

The following theorem guarantees that in well behaved problems the sequence of EM iterates converges to a stationary point which is a global maximum, in which case this yields the unique ML estimate of $\boldsymbol{\theta}$.

Theorem 1: Let $l_{\boldsymbol{y}}(\boldsymbol{\theta}) = \log(p(\boldsymbol{y}; \boldsymbol{\theta}))$, then $l_{\boldsymbol{y}}(\boldsymbol{\theta}^{[k+1]}) - l_{\boldsymbol{y}}(\boldsymbol{\theta}^{[k]}) \geq Q(\boldsymbol{\theta}^{[k+1]}; \boldsymbol{\theta}^{[k]}) - Q(\boldsymbol{\theta}^{[k]}; \boldsymbol{\theta}^{[k]})$ with equality iff $\mathcal{D}\left(p(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\theta}^{[k+1]}) || p(\boldsymbol{x}|\boldsymbol{y}; \boldsymbol{\theta}^{[k]})\right) = 0$, where $\mathcal{D}(\cdot||\cdot)$ is the Kullback-Leibler distance.

The proof and further properties of the EM algorithm are discussed for example in [8], [9].

In our problem setting, we identify $\boldsymbol{y}$ as the observed and $\boldsymbol{s}$ as the missing data. The unobserved state at time t, $\boldsymbol{s}_t$, is mapped to the observation $\boldsymbol{y}_t = [y_t^{(1)}, \ldots, y_t^{(n)}]$. The joint distribution of $(\boldsymbol{s}, \boldsymbol{y})$ depends on the parameter $\boldsymbol{g}$ of the convolutional encoder (see [9], [10])

$$p(\boldsymbol{s}, \boldsymbol{y}; \boldsymbol{g}) = p(\boldsymbol{s}_0) \prod_t p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t) p(\boldsymbol{y}_t|\boldsymbol{s}_t; \boldsymbol{g}).$$

The transition to the next state does not depend on $\boldsymbol{g}$ and the same holds for the distribution of the initial state $p(\boldsymbol{s}_0)$. Hence, applied to our problem, the maximization in (2) can be written as

$$\boldsymbol{g}^{[k+1]} = \arg\max_{\boldsymbol{g}} \left\{ \sum_{t,\ \boldsymbol{s}_t} p(\boldsymbol{s}_t|\boldsymbol{y}; \boldsymbol{g}^{[k]}) \log(p(\boldsymbol{y}_t|\boldsymbol{s}_t; \boldsymbol{g})) \right\}. \tag{3}$$

Note, given the state $\boldsymbol{s}_t$, the n outputs $\boldsymbol{y}_t^{(i)}$ of the convolutional transducer are independent from each other and only influenced by the respective subset of parameters $\boldsymbol{g}^{(i)}$:

$$\log(p(\boldsymbol{y}_t|\boldsymbol{s}_t; \boldsymbol{g})) = \sum_{i=1}^{n} \log(p(y_t^{(i)}|\boldsymbol{s}_t; \boldsymbol{g}^{(i)})). \tag{4}$$

So far, we considered $\boldsymbol{g}$ as an element from a discrete parameter space. However, the EM based approach requires the parameter space to be continuous. In [10], this has been achieved by introducing a probabilistic model. Here, we will transform the parameters into the LLR space, which will greatly simplify the calculations applying log-likelihood algebra [11].

IV. Transformation into the Log-Likelihood Domain

A. Application of Log Likelihood Ratios

Instead of regarding the encoder parameters as discrete values, we will work with probabilities. In the context of iterative decoding it has proven useful to work with log-likelihood ratios and to apply log-likelihood algebra [11]. We will see that it is also a natural choice for our problem. In general, a log-likelihood ratio $L(x)$ for the binary random variable x with pdf $p(x) = [p(x = +1),\ p(x = -1)]$ is denoted as $L(x) = \log\left(\frac{p(x=+1)}{p(x=-1)}\right)$ and the soft bit $\bar{x}$, which is the expected (average) value of x, can be shown to be $\bar{x} = E\{x\} = \tanh\left(\frac{L(x)}{2}\right)$. Throughout this paper, we will abbreviate the log-likelihood ratio and the soft bit of the parameter $g_m^{(i)}$ as

$$L_m^{(i)} = \log\left(\frac{p(g_m^{(i)} = +1)}{p(g_m^{(i)} = -1)}\right) \text{ and } \bar{g}_m^{(i)} = \tanh\left(\frac{L_m^{(i)}}{2}\right). \tag{5}$$

Furthermore, we denote $\boldsymbol{L} = [L_0^{(1)}, L_1^{(1)}, .., L_M^{(n)}] = [\boldsymbol{L}^{(1)}, .., \boldsymbol{L}^{(n)}]$ as our parameter vector. In our EM approach

[1] We will denote the dependency of a pdf $p(x)$ on a deterministic parameter $\boldsymbol{\theta}$ by $p(x; \boldsymbol{\theta})$ whereas we will write $p(x|\boldsymbol{\theta})$ when we mean to express the conditional pdf of x on a random variable $\boldsymbol{\theta}$.

we now consider $Q(\boldsymbol{L}, \boldsymbol{L}^{[k]})$ and the objective function in (3) is transformed into

$$\sum_{t,\ \boldsymbol{s}_t} p(\boldsymbol{s}_t|\boldsymbol{y}; \boldsymbol{L}^{[k]}) \log(p(\boldsymbol{y}_t|\boldsymbol{s}_t; \boldsymbol{L})), \tag{6}$$

which is now maximized for $\boldsymbol{L} \in \mathbb{R}^{n(M+1)}$, i.e. in a continuous space. In order to find an analytical solution, we relax the strict maximization of (6) and require only $Q(\boldsymbol{L}^{[k+1]}, \boldsymbol{L}^{[k]}) > Q(\boldsymbol{L}^{[k]}, \boldsymbol{L}^{[k]})$, which, according to Theorem 1, suffices to increase the likelihood in iteration step $k+1$. This approach is commonly referred to as the *generalized* EM algorithm [9]. An increase of $Q(\cdot,\cdot)$ is achieved by a steepest ascent approach, i.e. the maximization is replaced by

$$L_m^{(i),[k+1]} = L_m^{(i),[k]} + \mu \frac{\partial Q(\boldsymbol{L}, \boldsymbol{L}^{[k]})}{\partial L_m^{(i)}}, \tag{7}$$

where μ is a suitable chosen step size. The gradient is evaluated at $L_m^{(i),[k]}$.

As we mentioned before, most of the terms in $Q(\boldsymbol{L}, \boldsymbol{L}^{[k]})$ do not depend on $\boldsymbol{L}$ and as a result, only equation (6) has to be considered for the derivative of the Q-function. With the independence property from (4), this yields

$$\frac{\partial Q(\boldsymbol{L}, \boldsymbol{L}^{[k]})}{\partial L_m^{(i)}} = \sum_{t,\ \boldsymbol{s}_t} p(\boldsymbol{s}_t|\boldsymbol{y}; \boldsymbol{L}^{[k]}) \times \frac{\partial}{\partial L_m^{(i)}} \log\left(p(y_t^{(i)}|\boldsymbol{s}_t; \boldsymbol{L}^{(i)})\right). \tag{8}$$

We identify $p(\boldsymbol{s}_t|\boldsymbol{y}; \boldsymbol{L}^{[k]})$ as the a posteriori probability distribution for the state of the encoder at time t, given all observations and the current guess of the LLRs. This distribution is of great interest in coding and signal processing and a number of computationally efficient forward-backward recursions have been developed to calculate this distribution, (an overview is provided e.g. in [9]). For the evaluation of (8), we still have to find an analytical expression for the derivative of the conditional log-likelihood of $y_t^{(i)}$. At this point, the choice of log-likelihood ratios for the parameters turns out to be an elegant solution. Remember that, assuming a BSC channel model, $y_t^{(i)}$ is a modulo 2 sum (Eq. (1)). We make use of the boxplus operator $\boxplus$, a result from LLR algebra [11]. In the LLR domain, (1) is transformed to

$$L(y_t{}^{(i)}|\boldsymbol{s}_t) = \sum_{\substack{k=0\\ s_{t_k}=-1}}^{M}\!\boxplus\; L_k^{(i)} \boxplus L(\eta)$$
$$= 2\tanh^{-1}\left(\prod_{\substack{k=0\\ s_{t_k}=-1}}^{M} \tanh(\frac{L_k^{(i)}}{2}) \tanh(\frac{L(\eta)}{2})\right). \tag{9}$$

For the BSC channel model with transition probability ε we set $L(\eta) = \log\left(\frac{1-\varepsilon}{\varepsilon}\right)$. Note that $L(y_t^{(i)}|\boldsymbol{s}_t)$ inherently depends on the parameter $\boldsymbol{L}^{(i)}$ which shall not explicitly appear in the notation for the sake of simplicity. Introducing LLRs allows for a convenient way to process soft channel values and different expressions of $L(y_t^{(i)}|\boldsymbol{s}_t)$ for the Gaussian or the multiplicative fading channel can be derived [11]. This is an obvious advantage of our probabilistic approach compared to the algebraic solution presented in [2] and is expected to lead to better results alike decoding methods that process soft-values outperform their algebraic counterparts [11]. In the following, we concentrate on the BSC channel model.

B. Analytical Derivation of the Gradient

In order to derive an analytic solution for (8), we have to evaluate an expression of the form

$$\frac{\partial}{\partial \theta_m} \log\left(p(y; \boldsymbol{\theta})\right), \tag{10}$$

a derivative of a log-likelihood that depends on a parameter vector $\boldsymbol{\theta}$ with respect to a single parameter θ_m.

Lemma 1:

$$\frac{\partial \log\left(p(y = \pm 1; \boldsymbol{\theta})\right)}{\partial \theta_m} = \pm p(y = \mp 1; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_m} L(y; \boldsymbol{\theta}). \tag{11}$$

Above statement can be easily verified substituting

$$p(y = \pm 1; \boldsymbol{\theta}) = \frac{e^{\pm L(y;\boldsymbol{\theta})/2}}{e^{L(y;\boldsymbol{\theta})/2} + e^{-L(y;\boldsymbol{\theta})/2}}, \tag{12}$$

in (10) and carrying out some simple calculations. In the log-domain, Eq. (10) now reads as

$$\frac{\partial}{\partial \theta_m} \log\left(p(y = \pm 1; \boldsymbol{\theta})\right) = \frac{\pm e^{\mp L(y;\boldsymbol{\theta})/2}}{e^{L(y;\boldsymbol{\theta})/2} + e^{-L(y;\boldsymbol{\theta})/2}} \frac{\partial L(y; \boldsymbol{\theta})}{\partial \theta_m}. \tag{13}$$

In our problem, we identify $L(y; \boldsymbol{\theta})$ with $L(y_t^{(i)}|\boldsymbol{s}_t)$ that is a boxplus-sum as shown in equation (9) the derivative of which can be given as

$$\frac{\partial L(y_t^{(i)}|\boldsymbol{s}_t)}{\partial L_m^{(i)}} = \frac{\partial}{\partial L_m^{(i)}} \sum_{k, s_{t_k}=-1}\!\boxplus\; L_k^{(i)} \boxplus L(\eta) \tag{14}$$

$$= \frac{\bar{\eta} \prod\limits_{k, s_{t_k}=-1} \bar{g}_k^{(i)}}{1 - (\bar{\eta} \prod\limits_{k, s_{t_k}=-1} \bar{g}_k^{(i)})^2} \frac{E\{(g_m^{(i)} - \bar{g}_m^{(i)})^2\}}{\bar{g}_m^{(i)}}. \tag{15}$$

Eqn. 15 is obtained by using the derivative of the hyperbolic tangent and its inverse. Equation (15) depends only on soft bits and the variance of $g_m^{(i)}$ and we used the relation $E\{(g_m^{(i)} - \bar{g}_m^{(i)})^2\} = 1 - (\bar{g}_m^{(i)})^2 = 2\frac{\partial \bar{g}_m^{(i)}}{\partial L_m^{(i)}}$. The numerator (divided by $\bar{g}_m^{(i)}$) measures the reliability of all the other parameters about the mth parameter. The denominator is a measure of uncertainty when the mth parameter is included. This ratio is weighted with the variance of $g_m^{(i)}$ which is between zero and 1. The higher the uncertainty about $g_m^{(i)}$ the higher its variance. For the estimation update in (7), the gradient is evaluated at $L_m^{(i),[k]}$. Our algorithm starts with an initialization of $\boldsymbol{L}^{[0]}$. Setting the LLRs of the parameters close to zero indicates maximum uncertainty. The choice of a good initial estimate is an important step in EM techniques as only local convergence is guaranteed. However, we will only consider the case where no a priori knowledge on the encoder structure is available. One EM iteration is performed by multiplying the terms obtained from the considerations in (13), (15) with the probabilities $p(\boldsymbol{s}_t|\boldsymbol{y};\boldsymbol{L})$, summing up over all possible states and times. The gradient (8) is then used to update the parameter estimates according to (7), where a suitable step size μ has to be chosen. Iterations are performed until a certain stop condition is met.

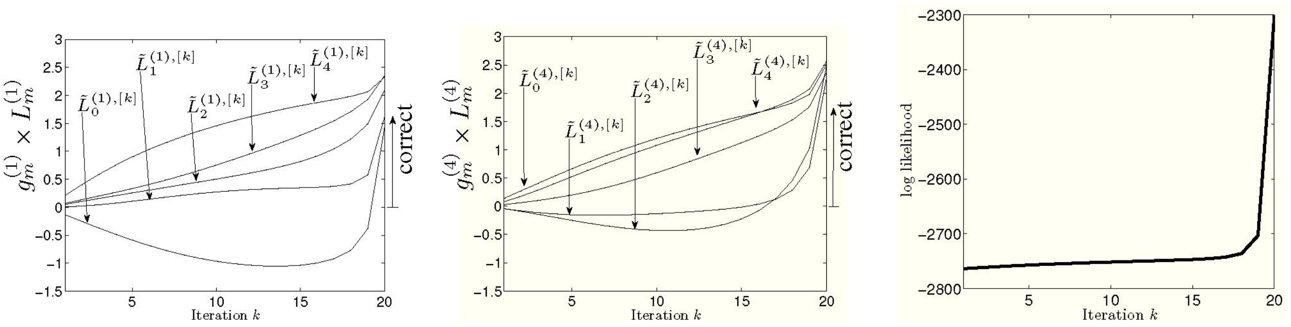
V. Simulation Results

The visualization of the iterative estimation of the parameters of a rate $\frac{1}{4}$ nonsystematic encoder with $M = 4$ ($[53, 51, 63, 73]$ in octal form) is shown in Figure 2. The encoded stream was sent over the memoryless BSC channel with transition probability $\varepsilon = 0.1$. It was assumed that the partitioning of the stream yielding $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(4)}$ is known at the receiver. 1000 noisy bits were observed per output, i.e. we expect 100 wrong bits per output. Full maximum likelihood estimation would mean to evaluate and compare 2^{20} likelihood values. In Figure 2(a) and 2(b) we see the progression of the LLRs of the parameters $\boldsymbol{g}^{(1)}$ and $\boldsymbol{g}^{(4)}$ over 20 EM iterations. On the y-axis, we show the LLRs multiplied with the true values, i.e. $\tilde{L}_m^{(i)} = L_m^{(i)} g_m^{(i)}$. Positive values indicate the correct estimation of a parameter, whereas at iterations where a negative value is shown, a hard decision would lead to a false estimate. We can observe that several parameters start to move in the wrong direction, reaching even high reliability values. However, after a certain number of iterations, they start to move back in the correct direction. Around iteration 19 the LLRs predict the correct encoder. As the absolute values of the LLRs reflect the reliability about an estimate, we stop iterating when there are no unreliable values left in $\boldsymbol{L}^{[k]}$. Figure 2(c) shows the log-likelihood $\log\big(p(\boldsymbol{y};\boldsymbol{L}^{[k+1]})\big)$. noise levels in case of *systematic* convolutional encoding. The streams were encoded with a rate $\frac{1}{4}$ memory $M = 4$ optimal distance profile code with generator polynomial $[\boldsymbol{g}^{(2)}, \ldots, \boldsymbol{g}^{(4)}] = [56, 62, 72]$ [12]. Correct convergence means that the $\boldsymbol{L}$ indicate high reliability (here defined as $|L_m^{(i)}| \geq 2\ \forall m, i$) *and* the true $\boldsymbol{g}$ is obtained after hard decision. Figure 3(b) shows the mean number of iterations until reliablility is achieved and the maximum number of iterations performed was set to 50. We observe high rates of convergence even at high noise levels for a relatively small number of observed coded bits. For example with 400 coded bits at a noise level $\varepsilon < 0.05$ we were always able to recover the correct encoder. However, in an attacking scenario we think that it can be assumed that an attacker has long data streams available and it is not expected that the encoding scheme changes every few bits. For the same reason, we ignore the computational complexity which is not expected to be of great relevance for the attacker. Considering nonsystematic encoders, we observed poor convergence rates with the setup presented. In this case, the success heavily depends on the start value and the gradient ascent approach often causes the algorithm to run in a local maximum immediately. Global maximization techniques such as simulated annealing or stochastic optimization [13] are expected to significantly improve the algorithm in this case and will be investigated in the future. Even though the restriction to systematic encoders is yet a disadvantage compared to the algebraic method presented in [2], we believe that the probabilistic approach has several advantages:

1) a priori information about the parameters is easily incorporated by adapting the initial ratios. In this way, knowledge on the design rules of convolutional encoders (see for example [12]) could help to choose good starting values. For instance every good convolutional encoder will have $L_0^{(i)} = -\infty$ and $L_M^{(i)} = -\infty$. In a similar fashion, constraints on the state sequence imposed by tail-biting codes could be incorporated in our algorithm.
2) The probabilistic approach can be extended to AWGN or fading channels and thus will be able to process the soft values coming from the channel.
3) With the results from [10], it will be possible to consider convolutional encoders with feedback, a situation which has neither been addressed in [2] nor, to our knowledge, in other publications.
4) The algebraic method requires information about the noise level for statistical hypothesis testing. Our EM approach allows for the co-estimation of the noise level as an additional parameter.

VI. Conclusion

We considered the problem of estimating the parameters of a convolutional encoder given only the noisy observations. Algebraic methods to tackle this problem have been considered before [3], [2]. Here, we presented an iterative, probabilistic approach based on the EM algorithm. Our approach is similar to the one presented by Moon for the synthesis of linear feedback shift registers in [10]. However, we used LLRs and applied log-likelihood ratio algebra, which better suits the underlying coding nature of this problem. We showed that this leads to simple derivations of the algorithm and allows for the processing of soft values. Furthermore, our LLR based approach provides a reliability measure on the obtained estimates. Alternatively, soft bits could be used to model the parameters. We ran our method on distorted data streams and showed that high rates of correct reconstruction can be achieved even at high noise levels assuming a systematic encoder. The algorithm shows poor convergence rates for non-systematic encoders as the gradient ascent approach runs the algorithm into local maxima. Global maximization techniques such as simulated annealing or stochastic optimization will be investigated in the future to overcome this problem. Even though our method is yet restricted to systematic codes, we believe that it could have several advantages compared to the algebraic approach as discussed at the end of Section V. We presented our approach for codes of rate $\frac{1}{n}$ and excluded the possibility of feedback from our considerations. However, in principle it is possible to apply our method to codes of any rate

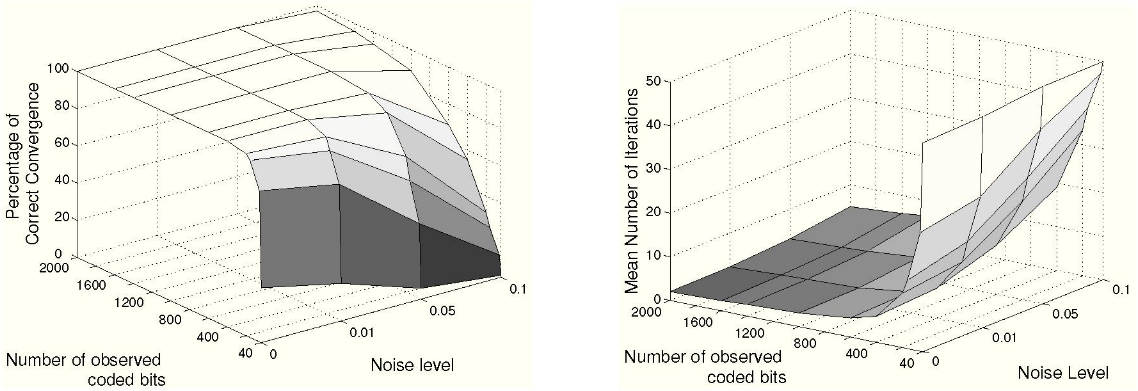


(a) LLR estimates on $\boldsymbol{g}^{(1)}$ vs. iterations.

(b) LLR estimates on $\boldsymbol{g}^{(4)}$ vs. iterations.

(c) Progression of $\log(p(\boldsymbol{y}; \boldsymbol{L}^{[k+1]}))$ vs. iterations.

Fig. 2. Sample of a successful estimation after 20 iterations of a rate $\frac{1}{4}$, $M = 4$, nonsystematic encoder from 4000 observed coded bits and $\varepsilon = 0.1$.



(a) Percentage of frames with correct convergence.

(b) Mean number of required iterations until convergence.

Fig. 3. Percentage of correct convergence and mean number of required iterations for different numbers of observed coded bits and noise levels in case of systematic convolutional encoding.

$\frac{k}{n}$ and the extension to the cases with feedback or puncturing should be possible with the results from [10], [6].

Acknowledgment

The authors would like to thank the Deutsche Forschungsgemeinschaft (DFG) and the Bund der Freunde der TUM for supporting their work.

References

[1] M. Cluzeau, "Block code reconstruction using iterative decoding techniques," in *Proc. IEEE International Symposium on Information Theory, ISIT06, Seattle*, July 2006.

[2] E. Filiol, "Reconstruction of convolutional encoders over $GF(q)$," *Lecture Notes In Computer Science*, vol. 1355, pp. 101–109, 1997.

[3] B. Rice, "Determining the parameters of a rate $\frac{1}{n}$ convolutional code over $GF(q)$," in *Proc. Third International Conference on Finite Fields and Applications, Glasgow*, 1995.

[4] A. Valembois, "Detection and recognition of a binary linear code," *Discrete Applied Mathematics*, vol. 111, pp. 199–218, 2001.

[5] G. L. Rosen, "Examining coding structure and redundancy in DNA," *IEEE Engineering in Medicine and Biology*, vol. 25, no. 1, pp. 62–68, January 2006.

[6] E. Filiol, "Reconstruction of punctured convolutional encoders," in *Proc. International Symposium on Information Theory and Applications (ISITA'00), Hawai*, November 2000.

[7] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406–2413, July 1994.

[8] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, pp. 47–60, November 1996.

[9] Y. Ephraim and N. Merhav, "Hidden markov processes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1518–1569, June 2002.

[10] T. K. Moon, "Maximum-likelihood binary shift-register synthesis from noisy observations," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 2096–2104, July 2002.

[11] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, March 1996.

[12] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, ser. Series on Digital & Mobile Communication, J. B. Anderson, Ed. IEEE Press, 1999.

[13] S. Schäffler and J. Hagenauer, "Soft decision MAP decoding of binary linear block codes via global optimization," in *Proc IEEE International Symposium on Information Theory, Ulm, Germany*, June 1997.