HMM Modeling of Burst Error Channels by Particle Swarm Optimization of the Likelihood Function

M. Vinicius dos Santos,
Ernesto L. Pinto
Electrical Engineering Department
Military Institute of Engineering, IME
Rio de Janeiro, Brazil
e-mail: marcusvinicius@ime.eb.br,
ernesto@ime.eb.br

Marco Grivet
Telecommunication Studies Center
Pontifical Catholic University of Rio de Janeiro, PUC/Rio
Rio de Janeiro, Brazil
e-mail: mgrivet@cetuc.puc-rio.br

***Abstract*— This paper proposes a new strategy for fitting Hidden Markov Models to error processes of channels with memory. Our approach consists of obtaining the analytical expression of the likelihood function of the model parameters and applying particle swarm optimization (PSO) to obtain their maximum likelihood (ML) estimates. In particular, this approach is here applied to the well known single error-state (simplified) Fritchman models, which have been recognized as a very useful tool for modeling error process of several communications systems over the last decades. The paper also addresses the mathematical analysis of several statistics of burst errors produced by these models. Some numerical examples are given in order to illustrate the effectiveness of the approach here proposed.**

Keywords*— *Hidden Markov models; Fritchman models; ML estimation; PSO.

I. INTRODUCTION

Most communications systems and networks of current interest for the scientific community are characterized by the occurrence of statistically dependent (burst) errors at the bit and packet levels.

A great research effort has been made during the last decades in order to provide tools for investigating the impact of errors on the performance of networks, by means of analysis and simulation [1].

The establishment of simple models with the ability to capture the most significant statistical properties of the error processes is an issue of utmost importance and the use of Markov Models (MM) and Hidden Markov Models (HMM) has been frequently claimed in this context.

Hidden Markov Models have been successfully applied to satellite and wireless communications [1]–[5], [13]. Besides, simple HMM such as the Gilbert-Elliot model have been shown to be useful for performance analysis, not only to evaluate the impact of errors generated in the physical layer, but also in regard of the errors produced in higher layers [1].

A class of Markov models of great interest is that of Fritchman models, in which a finite number of states is partitioned in error states, that generate errors with probability 1, and error-free states [6]. In particular, the single-error state model usually named simplified Fritchman model (SFM for short) has been applied to several wireless channels and frequently used as a benchmark to evaluate other strategies for burst error modeling [7].

The parametrization of HMM models is usually obtained on the grounds of maximum likelihood (ML) estimation, using the well-known Baum-Welch algorithm. This algorithm is guaranteed to converge to parameter estimates that locally maximize the likelihood metrics. As the likelihood functions of HMM models often have multiple modes, the quality of the models so obtained is heavily dependent on the initialization of the algorithm.

In the case of Fritchman models, the mathematical analysis of some statistics of the burst error process and the use of curve fitting techniques have been frequently adopted for adjusting the model parameters [4], [7].

In the present work we propose another approach for fitting those models, which is rooted on the derivation of a convenient analytical expression for the likelihood of the parameters. With this aim, the observed error sequence is regarded as a sequence formed by pairs of lengths of clusters (groups of consecutive errors) and gaps (groups of consecutive correct decisions), instead of being expressed as a sequence of binary observations. Besides, we propose the use of the particle swarm optimization (PSO) technique to maximize the likelihood function, in order to better deal with its multi-modality.

To the best of our knowledge, no similar approach has been proposed so far. Our guess is that it may be applied as an effective tool for deriving HMM models for errors in channels with memory. However, we have just started to investigate the use of this approach and have no rigorous characterization of its applicability at this moment.

We focus on the application of the proposed approach to fitting simplified Fritchman models. Besides deriving the likelihood function, we also present analytical expressions of several statistics of the error process generated by an SFM. Numerical examples of the application of these expressions to check the consistency of the proposed approach for model parametrization are also presented.

The paper is organized in 5 sections. A brief presentation of simplified Fritchman models is given in Section II, as well as the analysis of their likelihood function. The proposed method for model parametrization applied to SFM is presented in section III. Section IV is devoted to the analysis of some statistics usually applied in the investigation of burst-error processes. Numerical results are given is section IV. Finally, our conclusions and proposals for future works are summarized in section V.

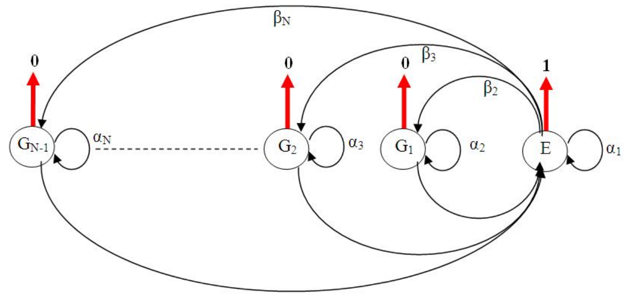

Fig. 1. Simplified Fritchman model of N states

II. Likelihood Function of Simplified Fritchman Model

The general form of the simplified Fritchman model can be seen in figure 1, where states are numbered from 1 to N, from right to left. This model is characterized by $(2.N-1)$ parameters, namely, $\underline{\Theta} = (\underline{\alpha}, \underline{\beta}) = ((\alpha_1, ..., \alpha_N), (\beta_2, ..., \beta_N))$, where all of them are in the range $[0,1]$ and $\alpha_1 + \sum_{n=2}^{N} \beta_n = 1$.

Without any loss of generality, any error sequence can be compactly written in the form $((0^{n_1}1^{m_1}), (0^{n_2}1^{m_2}), ...(0^{n_K}1^{m_K}))$, where it is assumed that all $m's$ and $n's$ are positive and an error has occurred just prior to this sequence and a no-error has occurred just after this sequence. We use the terms "gap" and "cluster" to respectively designate these groups of $0's$ and $1's$. "Bursts" are defined here as a sequence beginning and ending with an error, preceded and succeeded by gaps with length greater or equal to a certain value L, such that no interior gap has length greater than L. From an historical point of view, this is the so-called "CCITT" definition of burst.

Straightforward calculations reveal that the probability of this sequence, under the conditions assumed, can be written as:

$$P[\bigcap_{k=1}^{K} (0^{n_k}1^{m_k})|1] = \prod_{k=1}^{K} P[0^{n_k}1^{m_k}|1] \tag{1}$$

where:

$$P[0^n 1^m|1] = \alpha_1^{m-1}. \sum_{i=2}^{N} \beta_i.(1-\alpha_i).\alpha_i^{n-1} \tag{2}$$

Hence, if the observed error sequence is characterized by $((m_1,n_1),(m_2,n_2),...,(m_K,n_K))$, then the ML estimation of parameter $\underline{\Theta}$ can be pursued by solving the following optimization problem:

$$max \quad z = \prod_{k=1}^{K} [\alpha_1^{m_k-1}. \sum_{i=2}^{N} \beta_i.(1-\alpha_i).\alpha_i^{n_k-1}] \tag{3}$$

subject to:

$$\begin{gathered} 0 \leq \alpha_i \leq 1 \quad i = 1,...,N \\ 0 \leq \beta_i \leq 1 \quad i = 2,...,N \\ \alpha_1 + \textstyle\sum_{n=2}^{N} \beta_n = 1 \end{gathered} \tag{4}$$

An important observation that should be made at this point is the trivial fact that $M_K = \sum_{k=1}^{K} m_k$ is a sufficient statistics for the sequence $m_1, m_2, ..., m_K$, since the objective function of the above optimization problem can be rewritten as:

$$max \quad z = \alpha_1^{M_K-K}. \prod_{k=1}^{K} [\sum_{i=2}^{N} \beta_i.(1-\alpha_i).\alpha_i^{n_k-1}] \tag{5}$$

Another point worth noticing is the fact that the objective function is a product of probabilities and if their number K is high, it is likely that figures below the computer's number range are reached (typically 10^{-300}). Note that the quality of the parameter's estimation is enhanced as K increases. The simplest strategy to circumvent this problem is the use of logarithms, but these probabilities are made by sums whose individual terms can also reach values below computer's resolution. Hence we developed a very simple approach that can deal with this limitation [14].

III. Proposed Method for Model Fitting

Although we are facing an optimization problem with linear constraints, this by no means assure us to find global optimal solutions. On the contrary, the literature is full of similar cases and situations where the multi-modal behavior of the objective function (*OF* for short) prevent us to find the global solution. The case here discussed is no exception.

The consequence of this fact is that the quest of global solution with any gradient-optimization method is extremely dependent on the proposed initial solution. Several strategies were envisaged to produce methods that can circumvent this fact, but up to these days, there is no such method capable to cope with a general multi-modal objective function. Most of them are based in some convenient metaphor, which is useful to produce a procedure that probes the *OF* beyond the "well of attraction" of a local optimal solution. We can mention the *Genetic Algorithms* (GA) [9], *Simulated Annealing* (SA) [10], *Evolutionary Algorithms* (EA) [11], *Particle Swarm Optimization* (PSO) [12] among others. All these methods have their pros and cons.

For a long time, the authors were involved with the PSO technique and our belief is that this method suits our needs as far as this application is concerned. In short, the PSO technique consists of initially spreading "particles" uniformly over the feasible region and evaluate a fitness function (in this case the objective function) at each of these particle locations. After that, each particle moves with a specific velocity which is an weighted vectorial composition of two components: one is given by the direction towards the best position visited by the particle (local information) and the other is given by the direction towards the best position visited by all particles (global information). This process is repeated until some stopping criterion is met. No question that this method has a strong brute-force feature but we must acknowledge that all metaphoric-based optimization procedures have this characteristic. For those interested in knowing more about a comparison between PSO and some evolutionary procedures, we recommend [8]

IV. Some Useful Statistics of the Burst-Error Process of a SFM

The limiting state probabilities $x_i, i = 1,..,N$ of the SFM can be easily calculated by solving the classical equation $\underline{x}^T = \underline{x}^T.P$ where P is the state transition probability matrix of this model. This solution is given by:

$$x_1 = [1 + \sum_{n=2}^{N} \frac{\beta_n}{1-\alpha_n}]^{-1} \tag{6}$$

$$x_i = \frac{\beta_i}{1-\alpha_i}.x_1 \quad i = 2,..,N \tag{7}$$

If G, EC, EFB respectively represents the random variables "gap length", "error cluster length" and "error-free burst length" (see these and more definitions in [7]), then their distributions are as follows:

$$P[G=m] = \frac{\sum_{i=2}^{N} \beta_i.(1-\alpha_i).\alpha_i^{m-1}}{\sum_{i=2}^{N} \alpha_i.\beta_i} \tag{8}$$

$$P[EC=m] = \alpha_1.(1-\alpha_1)^{m-1} \tag{9}$$

$$P[EFB=m] = \frac{\sum_{i=2}^{N} \beta_i.(1-\alpha_i).\alpha_i^{m-1}}{\sum_{i=2}^{N} \alpha_i^{L-1}.\beta_i} \tag{10}$$

There are two other distributions that are of particular interest to those who work in the coding community. They prefer to approach this problem by means of the concept of "burstiness" instead of a definition of "burst". This concept is related to the length of sequences of the form $(0^m|1)$ and $(1^m|0)$. If errors are random and uniformly distributed over the time, these length distributions do present a power law, indicated by a linear graph in a log scale. "Burstiness" is seen as a depart from this linear situation.

For the SFM here discussed these distributions are given by the following expressions:

$$P[1^m|0] = (\frac{x_1}{1-x_1}).(1-\alpha_1).\alpha_1^{m-1} \tag{11}$$

$$P[0^m|1] = \sum_{n=2}^{N} \beta_n.\alpha_n^{m-1} \tag{12}$$

It is conceivable that someone interested in having a Fritchman's model (or any other model, as a matter of fact) that adjusts his data can approach this problem by means of fitting a particular distribution to this data. This approach is in the realm of the so called "curve fitting" techniques and by doing so there is no guarantee that other important statistics do also fit the same data.

What we would like to make crystal clear at this point is the fact that we are not here proposing a curve-fitting solution but instead an approach of finding model's parameters by means of the maximum likelihood standpoint. It is a fact that we also have no guarantee that the model hereby found do fit all the above mentioned distributions. Nevertheless what we have observed in practical cases, and is illustrated in next section, is that models estimated by this manner tend to fit all the mentioned distributions.

V. Some Numerical Results

In order to test and validate this procedure as a mean of estimating SFM from data, we adopted the following methodology: (1) we randomly select an order and the parameters of a SFM; (2) we generate a large error sequence according tho the previous item; (3) we find the ML estimates of the SFM's parameters by means of the PSO technique; (4) since we have analytical expressions for some statistics, we evaluate them for the true and estimated parameter values of the SFM.

For the purpose of this article, we select a SFM of 4 states, corresponding to 7 parameters. The real and estimated parameters values are shown in table I and the selected statistics are graphically presented in figure 2. We would like to draw attention to the following facts:

- Although the estimated parameters are not very closed to the real ones, the matchings of the statistics are excellent. This

TABLE I
True and Estimated SMF Parameters

Parameter	Real Value	Estimated Value
α_1	5.054^{-1}	5.182^{-1}
α_2	9.841^{-1}	9.930^{-1}
α_3	4.233^{-1}	4.051^{-1}
α_4	9.937^{-1}	6.811^{-1}
β_2	1.940^{-2}	7.572^{-2}
β_3	4.160^{-1}	4.060^{-1}
β_4	5.926^{-2}	0.000

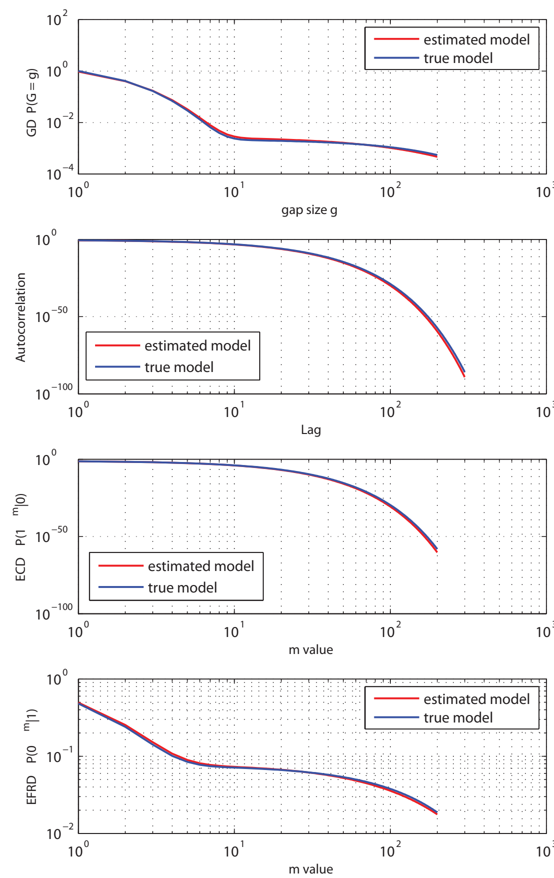

Fig. 2. Real and Estimated Statistics

fact suggests some robustness of the implemented procedure and it was observed in all tests which we made with models of order from 4 to 10.

- The ML estimation of the model parameters was able to produce a fairly acceptable fitting of four different statistics, although none of them was involved in the original procedure.
- The number of PSO particles used for the example presented was 10000, which represents $10000^{1/7} = 3.727$ particles per dimension, an extremely low value. This pattern was also observed in higher dimensions, suggesting that even with sparse filling feasible regions, the PSO algorithm fortunately behaves well.
- The number of $(0^m 1^n)$ data pairs used for this example was 1000, which roughly means a data error sequence 2.000.000 bits long. As stated before, the traditional ML estimation by Baum-Welch procedure operates on the data error sequence while the PSO operates on the gap-cluster pair sizes, thus giving an advantage to the latter as far as computer processing time is concerned.

VI. Conclusions

A new approach for fitting Hidden Markov Models to burst error channels has been proposed. This method is rooted on the derivation of a convenient expression for the observed error sequences, in terms of gap and cluster lengths, as well as on the use of particle swarm optimization for performing maximum likelihood estimation of the model parameters. An application of this approach to simplified Fritchman models has been addressed. Besides, the analysis of a number of statistics usually adopted in the investigation of burst errors has also been presented. The expressions so obtained were used in numerical examples of application of the approach here presented for fitting SFM models. In the continuation of this work we intend to apply this method to the parametrization of other Hidden Markov models. Besides, an in-depth investigation of the class of models that can be adjusted by the proposed method will also be pursued.

Acknowledgment

The authors would like to express their thanks to the Brazilian Council for Scientific and Technological Development (CNPq) ant to the Foundation for Research Support of the State of Rio de Janeiro (FAPERJ), for the partial financial support to this work.

References

[1] M. Zorzi and R. R. Rao. Perspectives an the impact of error statistics on protocols for wireless networks. *IEEE Personal Communications*, 6(5):32–40, October 1999.

[2] J. Paffett V.Y.Y. Chu, P. Sweeney and M. N. Sweeting. Characterising error sequences of the low earth orbit satellite channel and optimisation with hybrid-arq schemes. In *Proc. IEEE Global Telecommunications Conference, GLOBECOM'1998*, volume 5, pages 2930–2935, 1998.

[3] A. Beverly and K. S. Shanmugan. Hidden markov models for burst errors in gsm and dect channels. In *Proc. IEEE Global Telecommunications Conference, GLOBECOM'1998*, volume 6, pages 3692–3698, 1998.

[4] J.S. Swarts and H.C. Ferreira. On the evaluation and application of markov channel models in wireless communications. In *Proc. IEEE Vehicular Technology Conference, VTC'1999 Fall*, volume 1, pages 117–121, 1999.

[5] C. Thirouard J. Antoine A. Wautier N.Paul, J. Brouet and L. Husson. Efficient evaluation of voice quality in geran (gsm edge radio access network). In *Proc. IEEE Vehicular Technology Conference, VTC'2001 Fall*, volume 3, pages 1402–1406, 2001.

[6] B. D. Fritchman. A binary channel characterization using partitioned markov chains. *IEEE Transactions on Information Theory*, 13(2):221–227, apr 1967.

[7] C. X. Wang and W. Xu. A new class of generative models for burst-error characterization in digital wireless channels. *IEEE Transactions on Communications*, 55(3):453–462, 2007.

[8] Peter J. Angeline Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences. *Evolutionary Programming VII - Lecture Notes in Computer Sciences - Springer Berlin/Heidelberg*, Volume 1447, 1998.

[9] Affenzeller M., Winkler S., Wagner S., Beham A. Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications *Chapman & Hall - 1st. edition* , 2009.

[10] P.M.J. van Laarhoven Simulated Annealing: Theory and Applications *Springer Netherlands* , 2009.

[11] Ashlock D. Evolutionary Computation for Modeling and Optimization *Springer New York* , 2009.

[12] Parsopoulos K.E, Vrahatis M.N. Particle Swarm Optimization and Intelligence: Advances and Applications *Information Science Publishing* , 2010.

[13] A. Umbert, P. Diaz, On the Importance of Error Memory in UMTS Radio Channel Emulation using Hidden Markov Models (HMM) *15th. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (IEEE PIMRC 2004), vol.4, pp.2998-3002, Sept. 2004.

[14] Fernandes, M.V.S Modeling Digital Communication Channels Subject to Error Bursts *Ph.D. Thesis - Electrical Department - Pontifical Catholic University of Rio de Janeiro*, May 2010.