



**TEL AVIV UNIVERSITY**  
**The Iby and Aladar Fleischman Faculty of Engineering**  
**The Zandman-Slaner School of Graduate Studies**

High-Order Hidden Markov Models with  
Applications to Rainfall Estimation from  
Cellular Network Measurements

A thesis submitted toward the degree of Master of Science  
in Electrical and Electronic Engineering

by

Uri Hadar

under the supervision of Prof. Hagit Messer-Yaron

March 2009





**TEL AVIV UNIVERSITY**  
**The Iby and Aladar Fleischman Faculty of Engineering**  
**The Zandman-Slaner School of Graduate Studies**

# High-Order Hidden Markov Models with Applications to Rainfall Estimation from Cellular Network Measurements

A thesis submitted toward the degree of Master of Science  
in Electrical and Electronic Engineering

by

Uri Hadar

This research was carried out  
in the Department of Electrical Engineering - Systems  
under the supervision of Prof. Hagit Messer-Yaron

March 2009



## Acknowledgments

I would like to thank Prof. Hagit Messer-Yaron for the support, and for the helpful counsel and guidance, sharing me with her vast knowledge, experience and uncompromising professionalism.

I would like to thank Prof. Pinhas Alpert, and my colleagues of the ‘Cellular Rain’ research group: Artem Zinevich for introducing me to the world of HMMs, Oren Goldshtein, Asaf Rayitsfeld and Noam David for the fruitful cooperation.

I would like to thank the following individuals for fertile discussions and helpful advices: Hanan Shteingart, Yair Noam and Noam Presman.

I Would like to thank ‘Pelephone’ and ‘Cellcom’ for providing us with data, and for the mostly appreciable cooperation.

Most and foremost, I Would like to thank my parents for endlessly supporting me and my decisions.



## Abstract

While the Hidden Markov Model (HMM) has been used for a wide range of applications, an efficient procedure for estimating the model parameters and finding the optimal state sequence has not been generally formulated for orders higher than first, i.e., for models that assume higher order of either the state sequence memory, or the dependency of the observations on the states. We propose a simple method that transforms any high order HMM (including models in which the state sequence and observation dependency are of different orders) into an equivalent first order one, and thus makes the first order HMM formulation applicable to models of any order.

The transformation forms an equivalent first-order model by stacking the state variables into vectors of appropriate length. The Baum-Welch reestimation procedure is modified to fit the transformed model, in cases where the order of the state sequence is different from that of the observations. A method for implementation is suggested, along with several efficient ways to carry out the transformation. Also discussed is order selection for high-order hidden Markov modeling, where to be determined are the order of the state sequence memory, and that of the observation dependency.

The high-order HMM is applied to the problem of rainfall estimation from power measurements of cellular networks. The measurements were recorded from a network of fixed terrestrial line-of-sight backhaul microwave links, that are employed for transmission purposes by cellular providers. Hidden Markov modeling is used, where the power measurements are modeled as the observations, and the hidden state sequence is the rain indicator process. Applying model selection methods to the data shows that the specifications of the first-order HMM are not sufficient to capture the dynamics of rain-induced attenuation, and thus, higher order modeling is required. A method is introduced for estimating rain rate from the power measurements, that is based on using the HMM, and the power law- a well known empirical relation between rain rate and microwave attenuation. The results presented are promising in terms of comparison with rain measured by rain gauges.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Rain estimation from microwave links . . . . .	1
1.2	Hidden Markov modeling . . . . .	2
1.3	The power law . . . . .	3
1.4	The data . . . . .	3
1.5	Outline of the method . . . . .	4
<b>2</b>	<b>Hidden Markov models</b>	<b>7</b>
2.1	The first-order hidden Markov model . . . . .	7
2.1.1	Markov chains . . . . .	7
2.1.2	Extending to HMM . . . . .	8
2.2	Generalizing to higher orders . . . . .	9
2.2.1	The high-order model . . . . .	10
2.2.2	Transforming into first-order . . . . .	11
2.2.3	The HMM parameters . . . . .	13
2.2.4	Incorporating additional information . . . . .	25
2.3	Order selection in high-order HMMs . . . . .	26
2.3.1	Determining $r$ . . . . .	27
2.3.2	Determining $m$ . . . . .	28
2.3.3	Summary . . . . .	34
<b>3</b>	<b>Using the HMM to detect wet and dry periods</b>	<b>35</b>
3.1	Using the first order HMM . . . . .	35
3.1.1	Initializing the HMM . . . . .	36
3.1.2	Results . . . . .	36
3.2	Does the first order assumption hold? . . . . .	37
3.3	Applying the higher order HMM . . . . .	38
3.3.1	Results . . . . .	39

<b>4</b>	<b>Estimating rain</b>	<b>41</b>
4.1	Smoothing the Data . . . . .	41
4.1.1	Smoothing by constrained nonlinear optimization . . .	41
4.1.2	Smoothing by linear filtering . . . . .	42
4.2	Converting to rain rate . . . . .	43
4.3	Results . . . . .	45
<b>5</b>	<b>Summary</b>	<b>47</b>
<b>A</b>	<b>Rainfall estimation using BSS</b>	<b>49</b>
A.1	A brief overview of BSS . . . . .	49
A.2	Applying BSS . . . . .	50
A.3	Experimental results . . . . .	50

# List of Figures

2.1	High-order HMM examples . . . . .	14
3.1	Results of applying 1st-order HMM . . . . .	38
3.2	Applying AIC to estimate HMM orders . . . . .	39
3.3	Results of high-order HMM . . . . .	40
4.1	Smoothing by constrained nonlinear optimization . . . . .	42
4.2	Frequency response of dequantization filter $H(e^{j\omega})$ . . . . .	44
4.3	Smoothing by linear filtering . . . . .	44
4.4	Rain estimated from link against measurements from rain gauge	45
A.1	Map of few of the links in the network . . . . .	51
A.2	Two input links. SOBI algorithm . . . . .	52
A.3	Four input links. JADE-op algorithm . . . . .	52
A.4	Four input links. SOBI algorithm . . . . .	53
A.5	Five input links. SOBI algorithm . . . . .	53



# List of Tables

2.1	High-order HMM summarising table . . . . .	26
3.1	Results of high-order HMM . . . . .	39



# Chapter 1

## Introduction

The problem of estimating spatial-temporal rainfall has been intensively investigated worldwide and has important implications in meteorology, hydrology, agriculture, environmental policies and weather forecasting. Several techniques have been developed to monitor rainfall. The prevailing methods use a network of rain gauges or a weather radar. Rain gauges are considered to provide accurate ground level fixed point measurements (though prone to errors due to wind, human etc.). However, it is generally impractical to install and maintain a sufficient number of closely spaced rain gauges required to monitor highly variable distribution of rainfall. Weather radars, on the other hand, estimate rainfall field over a large contiguous areas with high temporal and spatial resolution, but are affected by high degree of uncertainty due to several reasons. For example, differences between measuring aloft and at ground level, and reflections from other forms of precipitation such as hail or melting particles [2]. Some researchers have suggested a combined approach that uses radar reflectivity and rain gauge data, in order to produce more accurate precipitation estimation (see [14] for example).

The increasing popularity of cellular communications and the global spread of wireless communication networks bring about a great opportunity for their use in environmental studies in general, and precipitation monitoring in particular. Wireless communication systems are wide-spread spatially, and operating in real time, 24 hours a day, and therefore, can provide high temporal and spatial resolution environmental monitoring without additional cost [26].

### 1.1 Rain estimation from microwave links

Several methods were proposed to estimate rainfall from microwave links, e.g. by Dino Giuli in [19] and by the MANTISSA Project in [31]. The methods

discussed were based on a pre-determined setup of links, characterized by the spatial geometry or the wave frequencies. In a published pioneer work, Messer et al [26] have demonstrated that RSL measurements from fixed terrestrial line-of-sight microwave links, deployed by cellular providers, can be used to estimate space-time rainfall intensities. In this work we continue focusing on using existing wireless systems, where the geometry and other parameters of the monitored system, designed by the cellular operators to maximize communication performance, are given. This enforced us to use an entirely different method for estimating the rainfall intensity. The monitored system is composed of existing, backhaul microwave links. Timely attenuation data are recorded from every link in the monitored area and transmitted to a central processing facility. Goldshtein [20] developed a method for generating a rain field of regular grid, based on the cellular measurements. In that method, the microwave links, which are arbitrarily located, are represented by points, and then spatially interpolated onto a regular grid to create the rain field.

## 1.2 Hidden Markov modeling

The problem of identifying wet and dry spells in microwave link data is addressed. Rahimi et al. [31] suggested a method which is based on computing the correlation between measurements from opposite links, as the attenuation tends to change similarly in the presence of rain. The main fault of this method is that it can be used only with links in which both ends transmit and receive. For identifying wet and dry spells we used hidden Markov modeling. A Hidden Markov model (HMM) is a model in which an underlying Markovian state sequence generates an observation sequence. An efficient algorithm for solving the HMM (i.e., estimating the model parameters and finding the hidden state sequence) given a set of observations exists for the case of first-order HMM [29]. Using model selection methods, we examined the compatibility of the statistical properties of the problem to the first-order HMM specifications. The aforementioned examination showed that the first-order HMM does not model the problem accurately enough to capture its dynamics, and thus a high-order generalization to the first-order HMM algorithm is needed. We propose a method that transforms high-order models into first-order ones [21] and thus provides the aforementioned generalization.

The results presented in this work show that reliable estimates of rainfall can be obtained using samples from existing infrastructure of fixed terrestrial microwave links in terms of comparability with rain gauge estimates. Based on the analysis made in this work, it is envisaged that the current infrastruc-



ture of fixed line of sight microwave links can provide an alternative mean to reconstruct rainfall fields at ground level, or can be considered integrative approach with the current weather radar and gauge methods in order to improve resolution and over all performance of rainfall estimation.

### 1.3 The power law

The propagation of electromagnetic microwaves traveling through air is influenced by atmospheric conditions. In particular, electromagnetic signals traveling at frequencies of above 10GHz, where the radio waves and raindrops are of comparable sizes, are subject to large variations in the attenuation due to precipitation. The basic idea is to collect measurements of the received signal level (RSL) from a network of microwave links, and process it for precipitation monitoring purposes. The microwave propagation impairments, caused by scattering and absorption of water droplets, have been extensively studied for years by telecomm engineers to provide reliable microwave communication system design [20]. The most common relation between rain rate and attenuation in a given microwave link, is known as ‘The Power Law’, and is given by

$$A = aR^bL \quad (1.1)$$

where  $A[\text{dB}]$  is the rain induced attenuation,  $R[\text{mm/hr}]$  is the rain rate,  $L[\text{km}]$  is the length of the link, and  $a, b$  are functions of the frequency, temperature and drop size distribution. The power law assumes constant rain rate along the link. The validity of the power law is now well established, and though the relation is regarded as empirical, a strong theoretical justification exists for this choice. For example, See [28]. See also propagation tests and validation in specific frequencies in [22].

In this work we regard the attenuation as a source for atmospheric information. Given a set of attenuation measurements from a link, the estimation of the rain induced attenuation from the total attenuation provides information on the rain rate by inversion of the power law:

$$R = \left( \frac{A}{aL} \right)^{\frac{1}{b}} \quad (1.2)$$

### 1.4 The data

The work presented here has been implemented on data records from an actual network of fixed terrestrial line-of-sight microwave links, employed for transmission purposes by ‘Pelephone’, an israeli cellular provider. Each

link has its own carrier frequency. All frequencies range between 17 and 23 GHz. The received signal level (RSL) has been recorded from each link in the monitored system, every minute, by a central server. The RSL is displayed in dBm units, with 1 dBm resolution. The links include a fade mitigation capability called Adaptive Transmitted Power Control (ATPC), a mechanism that is used to improve the spectrum efficiency of the links by limiting the transmitted power to that required to maintain a constant bit error rate (BER) regardless of the propagation conditions. The ATPC automatically controls the transmitted power through indication of the received power. Due to limitations of the system, records of the transmitted power level were not available. Our interest is in the attenuation, which is the difference between the received power and the transmitted power (in dB). Therefore, we only used measurements that were taken while the ATPC was turned off, that is, while the transmitted power was constant.

We denote the RSL of a given link at time  $n$  by  $P_n$ [dBm]. The total attenuation is the difference between the transmitted power and the received power (the RSL). We do not have the transmitted power level but we know it is constant. We arbitrarily take this constant to be  $\max\{P_n\}$ . The attenuation at time  $n$ ,  $A_n^{total}$ [dB] is then given by

$$A_n^{total} = \max_l \{P_l\} - P_n. \quad (1.3)$$

We denote the rain induced attenuation at time  $n$  by  $A_n$ [dB], and the attenuation that is not rain induced, i.e., the zero rain attenuation by  $Z_n$ [dB]. The total attenuation is the sum

$$A_n^{total} = A_n + Z_n. \quad (1.4)$$

## 1.5 Outline of the method

Since the attenuation is affected by a variety of physical phenomena, we first try to separate, or to estimate the rain induced attenuation from the total attenuation in order to estimate rainfall intensity. Given a time series of RSL measurements from a link, our goal is, then, to estimate the rain induced attenuation from these measurements. At times of no rain, the rain induced attenuation is zero ( $A_n = 0$ ) and thus (1.4) becomes  $Z_n = A_n^{total}$ . At times of rain the values of  $Z_n$  were interpolated from its values at dry times. In order to identify the wet and dry periods, the Hidden Markov Model (HMM) was employed. The power measurements were modeled as the observations and the hidden state sequence is the rain indicator process - zero when there is no rain and one otherwise.

In the next chapter an HMM overview is given, and a generalization of the first-order HMM to higher orders is suggested. In chapter 3 we apply the HMM to identify wet and dry periods, and examine, using Akaike's information criterion, whether the first-order assumptions hold. In chapter 4 we use the results obtained by the HMM to estimate rain rate and compare to rain measured by rain gauges.



# Chapter 2

## Hidden Markov models

A Hidden Markov Model (HMM) is a model in which the observed signal is a probabilistic function of an underlying Markovian state sequence, where the model parameters are generally unknown. The basic theory of HMMs was published in a series of classic papers by Baum and his colleagues (e.g.[5]) in the late 1960s and early 1970s, and the statistical methods of HMMs have become increasingly popular over the last few decades. Since the models are very rich in mathematical structure, they can form the theoretical basis for use in a wide range of applications. Some examples are speech recognition [3], biomedical [4], image restoration [7] and economics [32].

In this chapter we propose a generalization of the first-order HMM to higher orders. This generalization makes the well known HMM formulation applicable to models of any order.

### 2.1 The first-order hidden Markov model

#### 2.1.1 Markov chains

A first-order Markov process, denoted here by  $\{q_n\}$ , is a stochastic process that satisfies

$$P(q_n|\{q_l\}_{l<n}) = P(q_n|q_{n-1}). \quad (2.1)$$

A Markov process  $\{q_n\}$  is said to be homogeneous if the transition probability  $P(q_n|q_{n-1})$  is independent of  $n$ . A Markov process taking a finite number of values is called a Markov Chain. These values are called states. We denote the number of states by  $M$ , and the states by  $\sigma_1, \sigma_2, \dots, \sigma_M$ , i.e.,  $q_n \in \{\sigma_1, \sigma_2, \dots, \sigma_M\} \forall n$ .

The probability of an homogeneous Markov chain of length  $N$  is given by

$$P(q_1, q_2, \dots, q_N) = P(q_1)P(q_2|q_1) \cdots P(q_N|q_{N-1}, q_{N-2}, \dots, q_1) =$$

$$= P(q_1) \prod_{n=1}^N P(q_n | q_{n-1}), \quad (2.2)$$

and thus, a full probabilistic description of such a process consists of the following parameters

- Transition probabilities

$$a_{ij} = P(q_n = \sigma_j | q_{n-1} = \sigma_i), \quad 1 \leq i, j \leq M. \quad (2.3)$$

- Initial state probabilities

$$\pi_i = P(q_1 = \sigma_i), \quad 1 \leq i \leq M. \quad (2.4)$$

### 2.1.2 Extending to HMM

A Hidden Markov Model is a model in which the state sequence (which is a Markov chain) is not directly observable. It generates an observation sequence according to a set of  $M$  probability distributions, associated with the  $M$  different states of the chain. At each time step  $n$ , an observation  $O_n$  is drawn from the (discrete or continuous) probability distribution associated with the current state  $q_n$ . The HMM formulation is based on the assumption that each observation is conditionally independent on the previous ones and on the states given the current state, i.e.,

$$P(O_n | \{O_l\}_{l < n}, \{q_l\}_{l \leq n}) = P(O_n | q_n). \quad (2.5)$$

We denote the probability distribution of state  $\sigma_i$  by  $b_i(O_n)$ , i.e.,

$$b_i(O_n) = P(O_n | q_n = \sigma_i), \quad 1 \leq i \leq M. \quad (2.6)$$

The set of parameters  $\lambda = \{\{a_{ij}\}, \{\pi_i\}, \{b_i(\cdot)\}\}$  gives a full probabilistic description of the aforementioned model.

Considering the model described above, solving the following problems, known as the three basic problems of the HMM [29], is of great interest:

- P1** Given the observation sequence  $\{O_n\}$  and the model parameters  $\lambda$ , how do we efficiently compute  $P(\{O_n\}; \lambda)$ ?
- P2** Given the observation sequence  $\{O_n\}$  and the model parameters  $\lambda$ , how do we choose a corresponding state sequence  $\{q_n\}$  which is optimal in some meaningful sense?
- P3** How do we adjust the model parameters  $\lambda$  to maximize  $P(\{O_n\}; \lambda)$ ?

An efficient procedure for solving problems P1-P3 in the first-order HMM has been formulated and is detailed in [29] (We refer to that procedure as E1HMM<sup>1</sup>)

## 2.2 Generalizing to higher orders

Despite the popularity of the E1HMM, it is limited to cases where the following assumptions hold:

1. The hidden state sequence is a first order homogeneous Markov process.
2. The observation at each time step is conditionally independent of the observation history and state history, given the current state.

While in some applications these assumptions are valid, other find them restrictive and demand modeling of higher order. In the analysis of DNA sequences, for example, where high order Markov models are often used, the specifications of the first order HMM are not sufficient [12]. In the field of speech recognition, it has been shown in [24] that a 2nd order generalization of the E1HMM outperforms the E1HMM. In communications, an example of a high order hidden Markov model is the transmission of a high-order Markov source over a multipath channel with additive noise, i.e.,

$$O_n = \sum_{l=0}^{m-1} h_l \tilde{q}_{n-l} + v_n$$

where  $\tilde{q}_n$  is an  $r$ th order Markov process over a finite alphabet,  $h_n$  is the impulse response of the channel,  $v_n$  is an i.i.d. noise and  $O_n$  is the received signal. Note that the order of the hidden state sequence  $r$  and the order of the observation dependency on the states  $m$  are not necessarily equal. A high-order generalization of the E1HMM can provide a very powerful tool for the above scheme compared to other common methods (e.g. Wiener filtering or MAP) because (1) The E1HMM does not require stationarity of the signal  $\tilde{q}_n$  (It only requires  $\tilde{q}_n$  to be an homogeneous Markov process, which is not necessarily stationary), and (2) The E1HMM does not require prior knowledge of the parameters (and thus, for example, can provide an estimate of the source without knowing the impulse response of the channel).

---

<sup>1</sup>The term HMM is somewhat ambiguous in the literature because it refers to both the model itself, as well as to the procedure for solving the model. Therefore, in this work, the efficient procedure for solving the model in the first order case, as described in [29], will be referred to as E1HMM. The model itself will be simply referred to as HMM (first order HMM or high order HMM).

Few generalizations had been proposed to overcome the limitations of the E1HMM. Mari et al. proposed a 2nd order generalization to the E1HMM [24]. However, in their model only the state sequence is of 2nd order and furthermore, higher orders are not discussed. JA du Preez presented an algorithm that transforms any high order HMM into an equivalent first order HMM [15]. In that work, high order is considered only for the state sequence, and his algorithm is complicated compared to our proposed method. Wai Ki Ching et al. developed a method for solving the high order HMM [12], yet did not address the very important problem of adjusting the model parameters. Another such method was proposed by L. M. Lee and J. C. LEE [23], yet their method is applicable only to the left-to-right structure common in speech processing applications.

In the next sections, we present a simple and elegant method that transforms any high order HMM into an equivalent first order HMM and thus makes the E1HMM applicable to HMMs of any order. First, we describe the high order model on which we intend to apply the E1HMM. Next, we introduce the transformation and an efficient way to perform it (and its inverse), and describe how to modify the E1HMM to fit the transformed model, where special handling is required for the procedure of the parameter reestimation (iterative update and improvement [29]) in the cases where the orders of the state sequence and observation dependency are different.

### 2.2.1 The high-order model

Assume the following model: The hidden state sequence  $\{\tilde{q}_n\}_{n=1}^N$  is an homogeneous Markov process of order  $r$  with  $S$  states, i.e., a stochastic process that satisfies

$$P(\tilde{q}_n|\{\tilde{q}_l\}_{l<n}) = P(\tilde{q}_n|\{\tilde{q}_l\}_{l=n-r}^{n-1}) \quad (2.7)$$

where the transition probability  $P(\tilde{q}_n|\{\tilde{q}_l\}_{l=n-r}^{n-1})$  is independent of  $n$ , and  $\tilde{q}_n$  takes  $S$  possible values. Without loss of generality, let these values be  $\{0, 1, \dots, S-1\}$ .

The process  $\{\tilde{q}_n\}_{n=1}^N$  is called the *hidden* state sequence because it is not directly observable. The sequence of (possibly vector valued) observations  $\{O_n\}_{n=1}^N$  is generated by the state sequence according to a set of probability distributions that satisfy

$$P(O_n|\{O_l\}_{l<n}, \{\tilde{q}_l\}_{l\leq n}) = P(O_n|\{\tilde{q}_l\}_{l=n-(m-1)}^n), \quad (2.8)$$

i.e., at each time step  $n$ , an observation  $O_n$  is generated by the state sequence according to a probability distribution associated with the last  $m$



states  $\{\tilde{q}_l\}_{l=n-(m-1)}^n$ . Each observation is conditionally independent on the previous ones and on the state sequence history, given the current and the preceding  $m - 1$  states.

To model the above process, the following parameters are needed.

- $S^{r+1}$  transition probabilities,

$$\tilde{a}_{i_r \dots i_0} = P(\tilde{q}_n = i_0 | \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-r} = i_r). \quad (2.9)$$

- $S^m$  observation (discrete or continuous) probability distributions,

$$\tilde{b}_{i_0 \dots i_{m-1}}(O_n) = P(O_n | \tilde{q}_n = i_0, \dots, \tilde{q}_{n-(m-1)} = i_{m-1}). \quad (2.10)$$

- $S^\nu$  initial state probabilities,

$$\tilde{\pi}_{i_1 \dots i_\nu} = P(\tilde{q}_1 = i_1, \tilde{q}_0 = i_2, \dots, \tilde{q}_{2-\nu} = i_\nu) \quad (2.11)$$

where  $\nu = \max\{r, m\}$ .

We denote the set of all model parameters by  $\tilde{\lambda}$ , i.e.,

$$\tilde{\lambda} = \{\{\tilde{\pi}_{i_1 \dots i_\nu}\}, \{\tilde{b}_{i_0 \dots i_{m-1}}(\cdot)\}, \{\tilde{a}_{i_r \dots i_0}\}\}.$$

Having defined the model, we wish to have an efficient procedure for solving the problems P1-P3 for the high-order case.

Note that the special case of  $r = m = 1$  is the First Order HMM, in which the state sequence is denoted by  $\{q_n\}$ , and (2.7),(2.8) take the forms of (2.1) and (2.5).

## 2.2.2 Transforming into first-order

Consider the high order model of section 2.2.1, i.e., an  $r$ th order Markov state sequence  $\{\tilde{q}_n\}$  over  $\{0, 1, \dots, S - 1\}$  that generates an observation sequence  $\{O_n\}$  with state dependency of order  $m$ . Our transformation is based on the following, well known property of high-order Markov chains.

**Proposition 1** *Let  $\overline{Q}_n^\nu = [\tilde{q}_n, \tilde{q}_{n-1}, \dots, \tilde{q}_{n-(\nu-1)}]^T$ . The (vector valued) process  $\{\overline{Q}_n^\nu\}$  is a first order Markov process for any  $\nu \geq r$ .*

Proof:  $P(\overline{Q}_n^\nu | \{\overline{Q}_l^\nu\}_{l < n}) = P(\{\tilde{q}_l\}_{l=n-(\nu-1)}^n | \{\tilde{q}_l\}_{l < n}) = P(\tilde{q}_n | \{\tilde{q}_l\}_{l < n}) = P(\tilde{q}_n | \{\tilde{q}_l\}_{l=n-\nu}^{n-1}) = P(\{\tilde{q}_l\}_{l=n-(\nu-1)}^n | \{\tilde{q}_l\}_{l=n-\nu}^{n-1}) = P(\overline{Q}_n^\nu | \overline{Q}_{n-1}^\nu)$ .

**Definition 1** *Let*

$$f : \{0, 1, \dots, S - 1\}^\nu \longrightarrow \{0, 1, \dots, S^\nu - 1\}$$

*be the mapping of any base  $S$  number to its decimal value, i.e., if  $i_0, i_1, \dots, i_{\nu-1} \in \{0, 1, \dots, S - 1\}$ , then*

$$\begin{aligned} f([i_0, i_1, \dots, i_{\nu-1}]) &= [S^{\nu-1}, \dots, S, 1] \cdot [i_0, i_1, \dots, i_{\nu-1}]^T \\ &= \sum_{l=0}^{\nu-1} i_l S^{\nu-1-l}. \end{aligned} \quad (2.12)$$

*Note that  $f$  is bijective and thus invertible. We denote its inverse by  $f^{-1}$ .*

**Proposition 2** *Let  $\nu = \max\{r, m\}$  and let*

$$q_n = f(\overline{Q}_n^\nu) = \sum_{l=0}^{\nu-1} \tilde{q}_{n-l} S^{\nu-1-l}. \quad (2.13)$$

*The (scalar valued) process  $\{q_n\}$  and the observation sequence  $\{O_n\}$  satisfy (2.1) and (2.5) and thus constitute a first order HMM.*

*Proof:*  $q_n$  is obtained by a memoryless operation on  $\overline{Q}_n^\nu$  and thus,  $\{q_n\}$  is a first order homogeneous Markov process as in (2.1).

Now,  $P(O_n | \{O_l\}_{l < n}, \{q_l\}_{l \leq n}) = P(O_n | \{O_l\}_{l < n}, \{\overline{Q}_l^\nu\}_{l \leq n}) = P(O_n | \{O_l\}_{l < n}, \{\tilde{q}_l\}_{l \leq n}) = P(O_n | \{\tilde{q}_l\}_{l=n-(\nu-1)}^n) = P(O_n | \overline{Q}_n^\nu) = P(O_n | q_n)$  which is (2.5).

$q_n$  is the decimal value of the base  $S$  number  $[\tilde{q}_n, \tilde{q}_{n-1}, \dots, \tilde{q}_{n-(\nu-1)}]$  (where  $\tilde{q}_n$  is the most significant digit and  $\tilde{q}_{n-(\nu-1)}$  is the least significant one), and thus the following holds:

- $q_n$  has  $S^\nu$  possible states  $\{0, 1, \dots, S^\nu - 1\}$ .
- The transformation (2.13) can be implemented with the recursive equation

$$q_n = \tilde{q}_n S^{\nu-1} + \left\lfloor \frac{q_{n-1}}{S} \right\rfloor \quad (2.14)$$

where  $\lfloor \cdot \rfloor$  denotes the floor operation.

- $\{\tilde{q}_n\}$  can be perfectly reconstructed from  $\{q_n\}$ . The inverse transformation can be implemented as follows:

$$\tilde{q}_n = \left\lfloor \frac{q_n}{S^{\nu-1}} \right\rfloor. \quad (2.15)$$

The transformation and its inverse can also be carried out using linear, time invariant (LTI) filters. Applying the  $Z$  transform on both sides of (2.13) gives the transformation filter

$$T(z) = \frac{q(z)}{\tilde{q}(z)} = S^{\nu-1} \sum_{l=0}^{\nu-1} (zS)^{-l} = \frac{S^\nu - z^{-\nu}}{S - z^{-1}} \quad (2.16)$$

and its inverse, the reconstruction filter of  $\{\tilde{q}_n\}$  from  $\{q_n\}$

$$T^{-1}(z) = \frac{\tilde{q}(z)}{q(z)} = \frac{1}{T(z)} = \frac{S - z^{-1}}{S^\nu - z^{-\nu}}. \quad (2.17)$$

One must bear in mind, though, that the above filters might output non-integer values due to finite machine accuracy and thus, should be implemented carefully.

Depicted in Fig. 2.1 are two examples for high-order models and their transformed, first-order equivalents. In A.1 is an example of a high order model in which  $r = 2$ ,  $m = 1$  and  $S = 2$ . The high order hidden state sequence  $\{\tilde{q}_n\}$  has two states  $\{0, 1\}$ . The state sequence is a 2nd order Markov process. Therefore, every transition depends not only on the current state, but also on the preceding one. The observation sequence has a first order dependency on the states, i.e., each observation is dependent only on the current state and thus there are only two probability distributions, one for each state. Fig. A.2 depicts the equivalent transformed model, which is of first order. The state sequence of the transformed model  $\{q_n\}$  is a first order Markov process with four states  $\{0, 1, 2, 3\}$ . Note the following two properties of the transformed model: (1) Not all transitions between the four states are possible. This matter of "illegal transitions" is discussed in subsection 2.2.3. (2) There are four states, yet only two different probability distributions. This follows from the fact that  $m < r$  and explained in subsection 2.2.3.

### 2.2.3 The HMM parameters

Once our model is transformed, it meets the first order HMM conditions (2.1) and (2.5) and we now wish to solve the three HMM problems P1-P3 using the E1HMM. The first two problems need no special attention and their solution is as detailed in [29]. In this section we address the parameter estimation problem P3, which requires careful handling for the cases in which  $m \neq r$ . We first address the case  $m = r$  and then expand the discussion to the cases  $m < r$  and  $m > r$ .

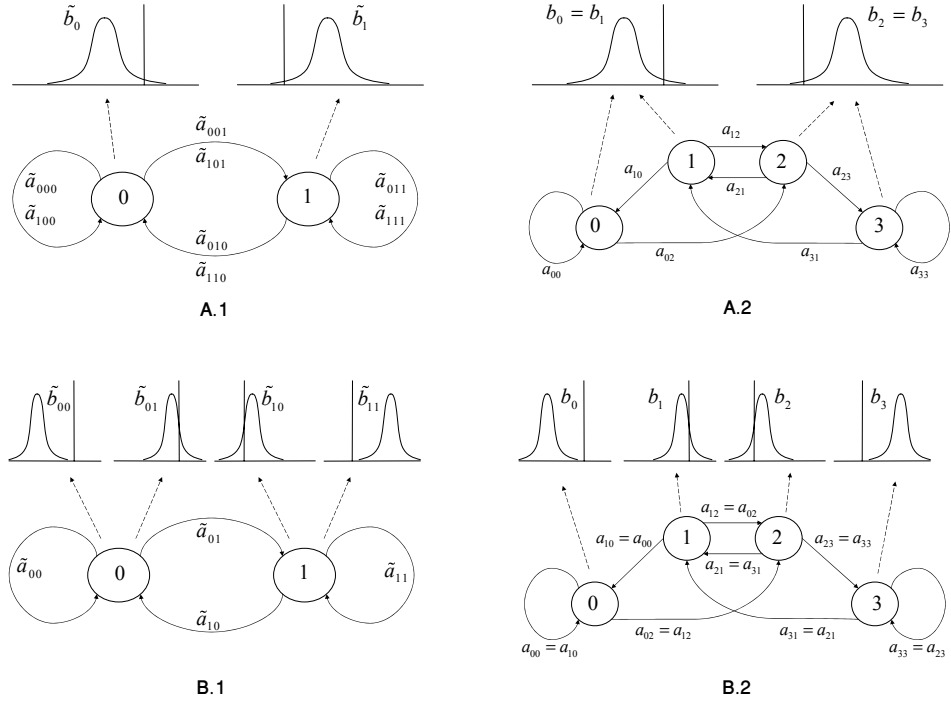


Figure 2.1: Two examples for high-order models with  $S = 2$ , and their transformed, first-order equivalents. A.1) Model A: A high-order model with  $r = 2$  and  $m = 1$ , i.e., a 2nd order Markov chain generating an observation sequence with 1st order state dependency. A.2) Model A transformed. Notice that in the first-order equivalent, since  $m < r$ , the number of observation distributions is smaller than the number of states. Notice also that not all state transitions are possible. This is due to (2.21). B.1) Model B: A high-order model with  $m = 2$  and  $r = 1$ , i.e., a 1st order Markov chain generating an observation sequence where each observation depends on both the current and preceding states. B.2) Model B transformed. Notice that since  $r < m$ , the number of transition probabilities is smaller than the number of possible transitions. In general, the number of states in the transformed model is dictated by  $\nu = \max\{r, m\}$ , and the number of parameters is equal to that of the original high-order model.

### The $m = r$ case

Having performed the transformation (only hypothetically of course, because the state sequence is not known yet), we now have a first order hidden state sequence  $\{q_n\}$  with states  $\{0, 1, \dots, S^r - 1\}$ , and first order state dependency of the observation sequence  $\{O_n\}$ . Let us examine our new set of parameters associated with the transformed model, and write the Baum-Welch reestimation formula for each of them [29]. We should expect to have the same number of parameters as that of the original high order model.

*The initial state probabilities.* The first order model has  $S^r$  states (here

$\nu = r = m$ ) and thus  $S^r$  initial state probabilities

$$\pi_i = P(q_1 = i) = \tilde{\pi}_{i_1 \dots i_r} \quad (2.18)$$

where  $i$  is the decimal value of the base  $S$  number  $[i_1, \dots, i_r]$ , or  $i = f([i_1, \dots, i_r])$ . The reestimation formula for  $\pi_i$  is given by

$$\hat{\pi}_i = \gamma_1(i) \quad (2.19)$$

where  $\gamma_n(i) = P(q_n = i | \{O_l\})$  as defined in [29].

*The transition probabilities.* In the first order model we have an  $S^r \times S^r$  transition matrix populated by  $S^{2r}$  transition probabilities

$$a_{ij} = P(q_n = j | q_{n-1} = i) = \tilde{a}_{i_r \dots i_0} \quad (2.20)$$

where  $i = f([i_1, \dots, i_r])$  and  $j = f([i_0, \dots, i_{r-1}])$ . This is obviously more than the  $S^{r+1}$  probabilities of the high order model. That is because there are illegal transitions we have yet to consider. To illustrate the matter of illegal transitions, let us look at the following example in which  $S = 2$  and  $r = 3$ . If, for some  $n$ ,  $\overline{Q}_n^r = [\tilde{q}_n, \tilde{q}_{n-1}, \tilde{q}_{n-2}]^T = [0, 1, 1]^T$  and thus  $q_n = 3$ , it implies that  $\overline{Q}_{n+1}^r = [\tilde{q}_{n+1}, \tilde{q}_n, \tilde{q}_{n-1}]^T = [\tilde{q}_{n+1}, 0, 1]^T$  can only take two possible values:  $[0, 0, 1]^T$  or  $[1, 0, 1]^T$  associated with the two possible values of  $\tilde{q}_{n+1}$ , which means that  $q_{n+1}$  will take either the value 1 or 5. This means that from state 3 we can only go to state 1 or to state 5. There are only two possible transitions for each state and thus only 16 legal transitions in total, whereas the transition matrix has 64 entries.

In general, the legality of the transitions is dictated by the fact that the first  $r - 1$  entries of  $\overline{Q}_n^r = [\tilde{q}_n, \tilde{q}_{n-1}, \dots, \tilde{q}_{n-(r-1)}]^T$  must equal the last  $r - 1$  entries of  $\overline{Q}_{n+1}^r = [\tilde{q}_{n+1}, \tilde{q}_n, \dots, \tilde{q}_{n-(r-2)}]^T$ . Now, since  $q_n, q_{n+1}$  are the decimal values of  $\overline{Q}_n^r, \overline{Q}_{n+1}^r$ , this is equivalent to

$$\left\lfloor \frac{q_n}{S} \right\rfloor = q_{n+1} - \left\lfloor \frac{q_{n+1}}{S^{r-1}} \right\rfloor S^{r-1}. \quad (2.21)$$

A transition from state  $i$  to state  $j$  is legal, then, only if  $i$  and  $j$  satisfy

$$\left\lfloor \frac{i}{S} \right\rfloor = j - \left\lfloor \frac{j}{S^{r-1}} \right\rfloor S^{r-1}. \quad (2.22)$$

Transition probabilities that correspond to illegal transitions must equal zero, Hence  $a_{ij} = 0$  for all  $0 \leq i, j \leq S^r - 1$  that satisfy  $\lfloor i/S \rfloor \neq j - \lfloor j/S^{r-1} \rfloor S^{r-1}$ .

For each possible value of  $q_n$  (out of the  $S^r$  possible ones),  $q_{n+1}$  can take one out of  $S$  possible values (each associated with one of the  $S$  possible values of  $\tilde{q}_{n+1}$ ) and thus, as expected, there is a total of  $S^{r+1}$  transition

probabilities to be estimated. Transition matrix entries that correspond to illegal transitions are set to zero. For the example above, (2.22) becomes

$$\left\lfloor \frac{i}{2} \right\rfloor = j - \left\lfloor \frac{j}{2^2} \right\rfloor 2^2$$

and thus the transition matrix  $\{a_{ij}\}$  is of the form

$$\begin{bmatrix} a_{00} & 0 & 0 & 0 & a_{04} & 0 & 0 & 0 \\ a_{10} & 0 & 0 & 0 & a_{14} & 0 & 0 & 0 \\ 0 & a_{21} & 0 & 0 & 0 & a_{25} & 0 & 0 \\ 0 & a_{31} & 0 & 0 & 0 & a_{35} & 0 & 0 \\ 0 & 0 & a_{42} & 0 & 0 & 0 & a_{46} & 0 \\ 0 & 0 & a_{52} & 0 & 0 & 0 & a_{56} & 0 \\ 0 & 0 & 0 & a_{63} & 0 & 0 & 0 & a_{67} \\ 0 & 0 & 0 & a_{73} & 0 & 0 & 0 & a_{77} \end{bmatrix}.$$

The reestimation formula for  $a_{ij}$  is given by

$$\hat{a}_{ij} = \begin{cases} \frac{\sum_{n=1}^{N-1} \xi_n(i, j)}{N-1} & \text{if } \left\lfloor \frac{i}{S} \right\rfloor = j - \left\lfloor \frac{j}{S^{r-1}} \right\rfloor S^{r-1} \\ \sum_{n=1} \gamma_n(i) & \\ 0 & \text{if } \left\lfloor \frac{i}{S} \right\rfloor \neq j - \left\lfloor \frac{j}{S^{r-1}} \right\rfloor S^{r-1} \end{cases} \quad (2.23)$$

where  $\xi_n(i, j) = P(q_n = i, q_{n+1} = j | \{O_l\})$  as defined in [29].

*The observation probability distributions.* Each of the  $S^r$  states has a corresponding observation probability distribution

$$b_i(O_n) = P(O_n | q_n = i) = \tilde{b}_{i_0 \dots i_{r-1}}(O_n) \quad (2.24)$$

where  $i = f([i_0, \dots, i_{r-1}])$ . We distinguish between two cases:

- *Discrete observation probability distribution.* The reestimation formula for  $b_i(v_k) = P(O_n = v_k | q_n = i)$ , the probability of observing symbol  $v_k$  for state  $i$ , is given by

$$\hat{b}_i(v_k) = \frac{\sum_{n: O_n = v_k} \gamma_n(i)}{N}. \quad (2.25)$$

- *Continuous observation probability distribution.* In the continuous case, the most general representation of the probability density function (pdf) for which a reestimation procedure has been formulated, is a finite mixture of the form

$$b_i(x) = \sum_{k=1}^K c_{ik} g(x; \mu_{ik}, U_{ik}) \quad (2.26)$$

where  $c_{ik}$  is the mixture coefficient for the  $k$ th mixture in state  $i$  and  $g(x; \mu_{ik}, U_{ik})$  is any log-concave or elliptically symmetric pdf (e.g., Gaussian) with mean vector  $\mu_{ik}$  and covariance matrix  $U_{ik}$  for the  $k$ th mixture component in state  $i$  [29]. The reestimation formulas for the pdf parameters are given by

$$\begin{aligned} \hat{c}_{ik} &= \frac{\sum_{n=1}^N \gamma_n(i, k)}{K \sum_{k=1}^K \sum_{n=1}^N \gamma_n(i, k)} \\ \hat{\mu}_{ik} &= \frac{\sum_{n=1}^N \gamma_n(i, k) O_n}{\sum_{n=1}^N \gamma_n(i, k)} \\ \hat{U}_{ik} &= \frac{\sum_{n=1}^N \gamma_n(i, k) (O_n - \hat{\mu}_{ik})(O_n - \hat{\mu}_{ik})^T}{\sum_{n=1}^N \gamma_n(i, k)} \end{aligned} \quad (2.27)$$

where  $\gamma_n(i, k)$  is the probability of being in state  $i$  at time  $n$  with the  $k$ th mixture component accounting for  $O_n$ .

The procedure for computing  $\gamma_n, \xi_n$  is detailed in [29].

An interim summing up: In order to use the E1HMM for our high order model, we make the proper adjustments to the E1HMM (define appropriate states and set illegal transitions to zero), apply it to the observation sequence, and then use the inverse transformation to get the hidden high order state sequence.

### The $m < r$ case

In this case the order of observation dependency on the state is smaller than that of the states sequence ( $\nu = r$ ). The transformation (2.13) becomes

$$q_n = f(\bar{Q}_n^r) = \sum_{l=0}^{r-1} \tilde{q}_{n-l} S^{r-1-l} \quad (2.28)$$

and thus the state sequence  $\{q_n\}$  is a first order Markov process with  $S^r$  states  $\{0, 1, \dots, S^r - 1\}$ . The dependency of order  $m$  assumption means that  $P(O_n | \{O_l\}_{l < n}, \{\tilde{q}_l\}_{l \leq n}) = P(O_n | \{\tilde{q}_l\}_{l=n-(m-1)}^n)$ , and in particular  $P(O_n | \{\tilde{q}_l\}_{l=n-(r-1)}^n) = P(O_n | \{\tilde{q}_l\}_{l=n-(m-1)}^n)$ . In other words, only the first  $m$  entries of  $\bar{Q}_n^r = [\tilde{q}_n, \tilde{q}_{n-1}, \dots, \tilde{q}_{n-(r-1)}]^T$  (the first  $m$  digits of the base  $S$  representation of  $q_n$ ) matter to each state's probability distribution and thus, the observation sequence "sees" only  $S^m$  different states while there are  $S^r$  different ones. Consider the example from section 2.2.2. We apply the transformation to obtain the first order state sequence  $q_n$  which has four states  $\{0, 1, 2, 3\}$ . However, since each observation  $O_n$  is (conditionally) dependent only on the current high order state  $\tilde{q}_n$  we get that

$$\begin{aligned} b_0(O_n) &= P(O_n | \tilde{q}_n = 0, \tilde{q}_{n-1} = 0) \\ &= P(O_n | \tilde{q}_n = 0, \tilde{q}_{n-1} = 1) = b_1(O_n) \end{aligned} \quad (2.29)$$

and

$$\begin{aligned} b_2(O_n) &= P(O_n | \tilde{q}_n = 1, \tilde{q}_{n-1} = 0) \\ &= P(O_n | \tilde{q}_n = 1, \tilde{q}_{n-1} = 1) = b_3(O_n). \end{aligned} \quad (2.30)$$

It follows that (2.25) is not valid in this case because it would yield  $S^r$  different distributions whereas we expect to have only  $S^m$  different ones.

The reestimation formulas are derived by maximizing Baum's auxiliary function

$$Q(\lambda, \hat{\lambda}) = \sum_{q_1, \dots, q_N} P(\{q_l\} | \{O_l\}; \lambda) \log P(\{q_l\}, \{O_l\}; \hat{\lambda}) \quad (2.31)$$

over  $\hat{\lambda}$  where  $\lambda = \{\{\pi_i\}, \{b_i(\cdot)\}, \{a_{ij}\}\}$  is the current set of parameters and  $\hat{\lambda} = \{\{\hat{\pi}_i\}, \{\hat{b}_i(\cdot)\}, \{\hat{a}_{ij}\}\}$  is the reestimated one. It has been proven that maximization of  $Q(\lambda, \hat{\lambda})$  leads to increased likelihood, i.e.,

$$\max_{\hat{\lambda}} Q(\lambda, \hat{\lambda}) \Rightarrow P(\{O_l\}; \hat{\lambda}) \geq P(\{O_l\}; \lambda) \quad (2.32)$$



and that eventually the likelihood function  $P(\{O_l\}; \lambda)$  converges to a critical point. More on the reestimation procedure can be found in [29].

We shall now derive the reestimation formulas under the assumption that  $m < r$  for *the discrete case*. We can write

$$P(\{q_l\}, \{O_l\}; \hat{\lambda}) = \hat{\pi}_{q_1} \left( \prod_{n=1}^N \hat{b}_{q_n}(O_n) \right) \left( \prod_{n=2}^N \hat{a}_{q_{n-1}q_n} \right) \quad (2.33)$$

and thus,

$$\begin{aligned} Q(\lambda, \hat{\lambda}) &= \sum_{q_1, \dots, q_N} P(\{q_l\}|\{O_l\}; \lambda) \log \hat{\pi}_{q_1} + \\ &+ \sum_{q_1, \dots, q_N} P(\{q_l\}|\{O_l\}; \lambda) \sum_{n=1}^N \log \hat{b}_{q_n}(O_n) + \\ &+ \sum_{q_1, \dots, q_N} P(\{q_l\}|\{O_l\}; \lambda) \sum_{n=2}^N \log \hat{a}_{q_{n-1}q_n}. \end{aligned} \quad (2.34)$$

It follows that  $\{\hat{\pi}_i\}$ ,  $\{\hat{b}_i(\cdot)\}$  and  $\{\hat{a}_{ij}\}$  can each be derived separately, so we only have to maximize the second term of the RHS of (2.34) because the assumption  $m < r$  only affects  $\{b_i(\cdot)\}$ . The reestimation formulas for  $\{\pi_i\}$  and  $\{a_{ij}\}$  are as in the case  $m = r$ , and given by (2.19) and (2.23). Let

$$Q_b(\lambda, \{\hat{b}_i\}) = \sum_{q_1, \dots, q_N} P(\{q_l\}|\{O_l\}; \lambda) \sum_{n=1}^N \log \hat{b}_{q_n}(O_n). \quad (2.35)$$

Now,

$$\begin{aligned} Q_b(\lambda, \{\hat{b}_i\}) &= \sum_{n=1}^N \sum_{q_n} \log \hat{b}_{q_n}(O_n) \sum_{q_k: k \neq n} P(\{q_l\}|\{O_l\}; \lambda) \\ &= \sum_{n=1}^N \sum_{q_n} \log \hat{b}_{q_n}(O_n) P(q_n|\{O_l\}; \lambda) \end{aligned} \quad (2.36)$$

and we wish to maximize  $Q_b(\lambda, \{\hat{b}_i\})$  subject to the constraint

$$\sum_k \hat{b}_i(v_k) = 1 \quad \forall i \in \{0, 1, \dots, S^r - 1\} \quad (2.37)$$

where the  $v_k$ 's are all the possible observation symbols.

From the fact that only the first  $m$  digits of the base  $S$  representation of each state matter to the state's probability distribution, it follows that  $b_{i_1}(\cdot) = b_{i_2}(\cdot)$  if

$$\lfloor i_1/S^{r-m} \rfloor = \lfloor i_2/S^{r-m} \rfloor = i'$$

or

$$i_1, i_2 \in \{i'S^{r-m}, i'S^{r-m} + 1, \dots, (i' + 1)S^{r-m} - 1\}$$

where  $i' \in \{0, 1, \dots, S^m - 1\}$ , and thus we can write

$$\begin{aligned} Q_b(\lambda, \{\hat{b}_i\}) &= \sum_{n=1}^N \sum_{i=0}^{S^r-1} \log \hat{b}_i(O_n) \gamma_n(i) \\ &= \sum_{n=1}^N \sum_{i'=0}^{S^m-1} \sum_{i=i_1(i')}^{i_2(i')} \log \hat{b}_i(O_n) \gamma_n(i) \\ &= \sum_{n=1}^N \sum_{i'=0}^{S^m-1} \log \hat{b}_{i_0}(O_n) \sum_{i=i_1(i')}^{i_2(i')} \gamma_n(i) \end{aligned} \quad (2.38)$$

where

$$i_1(i') = i'S^{r-m} \quad (2.39)$$

$$i_2(i') = (i' + 1)S^{r-m} - 1 \quad (2.40)$$

and  $i_0$  is any integer between  $i_1(i')$  and  $i_2(i')$ .

To maximize  $Q_b(\lambda, \{\hat{b}_i\})$  subject to (2.37) we use Lagrange multipliers.

$$\begin{aligned} \frac{\partial Q_b(\lambda, \{\hat{b}_i(v_k)\})}{\partial \hat{b}_j(v_k)} &= \\ &= \sum_{n: O_n=v_k} \frac{\partial}{\partial \hat{b}_j(v_k)} \sum_{i'=0}^{S^m-1} \log \hat{b}_{i_0}(v_k) \sum_{i=i_1(i')}^{i_2(i')} \gamma_n(i) \\ &= \sum_{n: O_n=v_k} \frac{1}{\hat{b}_j(v_k)} \sum_{i=i_1(\lfloor \frac{j}{S^{r-m}} \rfloor)}^{i_2(\lfloor \frac{j}{S^{r-m}} \rfloor)} \gamma_n(i). \end{aligned} \quad (2.41)$$

The derivative of the constraint equals 1 for all  $k$ , so

$$\sum_{n: O_n=v_k} \sum_{i=i_1(\lfloor \frac{j}{S^{r-m}} \rfloor)}^{i_2(\lfloor \frac{j}{S^{r-m}} \rfloor)} \gamma_n(i) = \hat{b}_j(v_k) \eta \quad (2.42)$$

where  $\eta$  is the Lagrange multiplier. Summing both sides over  $k$  and substituting (2.37) yields

$$\eta = \sum_{n=1}^N \sum_{i=i_1(\lfloor \frac{j}{S^{r-m}} \rfloor)}^{i_2(\lfloor \frac{j}{S^{r-m}} \rfloor)} \gamma_n(i) \quad (2.43)$$

and finally we obtain the reestimation formula

$$\hat{b}_i(v_k) = \frac{\sum_{n:O_n=v_k} \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j)}{\sum_{n=1}^N \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j)} \quad (2.44)$$

where

$$j_1(i) = \left\lfloor \frac{i}{S^{r-m}} \right\rfloor S^{r-m} \quad (2.45)$$

$$j_2(i) = \left( \left\lfloor \frac{i}{S^{r-m}} \right\rfloor + 1 \right) S^{r-m} - 1. \quad (2.46)$$

One can see, by comparing (2.25) and (2.44), that this result is very intuitive because, as expected, the estimator treats all states that share the same base  $S$  prefix as equal, summing over them. We conjecture then, that the assumption  $m < r$  will introduce a similar modification to the reestimation formulas of the continuous case (2.27):

$$\begin{aligned}
\hat{c}_{ik} &= \frac{\sum_{n=1}^N \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)}{\sum_{k=1}^K \sum_{n=1}^N \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)} \\
\hat{\mu}_{ik} &= \frac{\sum_{n=1}^N O_n \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)}{\sum_{n=1}^N \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)} \\
\hat{U}_{ik} &= \frac{\sum_{n=1}^N (O_n - \hat{\mu}_{ik})(O_n - \hat{\mu}_{ik})^T \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)}{\sum_{n=1}^N \sum_{j=j_1(i)}^{j_2(i)} \gamma_n(j, k)}.
\end{aligned} \tag{2.47}$$

### The $m > r$ case

Here  $\nu = m$ , and the transformation (2.13) becomes

$$q_n = f(\bar{Q}_n^m) = \sum_{l=0}^{m-1} \tilde{q}_{n-l} S^{m-1-l} \tag{2.48}$$

and thus the state sequence  $\{q_n\}$  is a first order Markov process with  $S^m$  states  $\{0, 1, \dots, S^m - 1\}$ . The fact that  $r < m$  implies that the state sequence  $\{q_n\}$  has sort of a "memory redundancy". It means that at each time step, the current state  $q_n$  has more information than what the transition requires. If we denote  $i = f([i_1, \dots, i_m])$  and  $j = f([i_0, \dots, i_{m-1}])$  then

$$\begin{aligned}
a_{ij} &= P(q_n = j | q_{n-1} = i) \\
&= P(\tilde{q}_n = i_0, \dots, \tilde{q}_{n-(m-1)} = i_{m-1} | \\
&\quad \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-m} = i_m) \\
&= P(\tilde{q}_n = i_0 | \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-m} = i_m) \\
&= P(\tilde{q}_n = i_0 | \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-r} = i_r).
\end{aligned} \tag{2.49}$$

It follows that only the first  $r$  digits of the base  $S$  representation of state  $i$  matter to the probability of the transition to state  $j$ ,  $a_{ij}$  (if the transition

is legal of course. Otherwise  $a_{ij} = 0$ ) and thus there should be only  $S^{r+1}$  transition probabilities to estimate whereas we have  $S^{m+1}$  legal transitions. Note that the condition for a transition to be legal in this case is

$$\left\lfloor \frac{i}{S} \right\rfloor = j - \left\lfloor \frac{j}{S^{m-1}} \right\rfloor S^{m-1}. \quad (2.50)$$

Consider an example in which  $r = 1$ ,  $m = 2$  and  $S = 2$ . In this example  $q_n$  has four states  $\{0, 1, 2, 3\}$ , and according to (2.50), which in this example is given by  $\lfloor i/2 \rfloor = j - \lfloor j/2 \rfloor 2$ , the transition matrix is of the form

$$\begin{bmatrix} a_{00} & 0 & a_{02} & 0 \\ a_{10} & 0 & a_{12} & 0 \\ 0 & a_{21} & 0 & a_{23} \\ 0 & a_{31} & 0 & a_{33} \end{bmatrix}. \quad (2.51)$$

From (2.49) we get that

$$\begin{aligned} a_{00} &= P(\tilde{q}_n = 0 | \tilde{q}_{n-1} = 0, \tilde{q}_{n-2} = 0) \\ &= P(\tilde{q}_n = 0 | \tilde{q}_{n-1} = 0, \tilde{q}_{n-2} = 1) = a_{10}, \end{aligned} \quad (2.52)$$

$$\begin{aligned} a_{21} &= P(\tilde{q}_n = 0 | \tilde{q}_{n-1} = 1, \tilde{q}_{n-2} = 0) \\ &= P(\tilde{q}_n = 0 | \tilde{q}_{n-1} = 1, \tilde{q}_{n-2} = 1) = a_{31} \end{aligned} \quad (2.53)$$

and in a similar manner  $a_{02} = a_{12}$  and  $a_{23} = a_{33}$ . It follows that (2.23) is not valid in this case because it would yield  $S^{m+1}$  different transition probabilities whereas we expect to have only  $S^{r+1}$  different ones.

We have seen in the previous section that  $\{\hat{\pi}_i\}$ ,  $\{\hat{b}_i(\cdot)\}$  and  $\{\hat{a}_{ij}\}$  can each be derived separately, so in this case we only have to maximize the third term of the RHS of (2.34) because the assumption  $m > r$  only affects  $\{a_{ij}\}$ . The reestimation formulas for  $\{\pi_i\}$  and  $\{b_i(\cdot)\}$  are as in the case  $m = r$  and given by (2.19) and (2.25) or (2.27).

The reestimation formula for the transition probabilities for this case can be derived directly by maximizing  $Q(\lambda, \hat{\lambda})$ , but we shall derive it in a shorter, less rigorous way. Let us consider the process  $X_n = f([\tilde{q}_n, \dots, \tilde{q}_{n-(r-1)}])$  with the corresponding probabilities  $\gamma_n^X(i') = P(X_n = i' | \{O_l\})$  and  $\xi_n^X(i', j') = P(X_n = i', X_{n+1} = j' | \{O_l\})$ . According to (2.23), the estimate of the transition probability  $a_{i'j'}^X = P(X_{n+1} = j' | X_n = i')$  is given by

$$\hat{a}_{i'j'}^X = \frac{\sum_{n=1}^{N-1} \xi_n^X(i', j')}{\sum_{n=1}^{N-1} \gamma_n^X(i')}. \quad (2.54)$$

We express now  $a_{ij}$  in terms of  $a_{i'j'}^X$ :

$$\begin{aligned}
a_{ij} &= P(\tilde{q}_n = i_0 | \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-r} = i_r) \\
&= P(\tilde{q}_n = i_0, \dots, \tilde{q}_{n-(r-1)} = i_{r-1} | \\
&\quad \tilde{q}_{n-1} = i_1, \dots, \tilde{q}_{n-r} = i_r) \\
&= P(X_n = f([i_0, \dots, i_{r-1}]) | X_{n-1} = f([i_1, \dots, i_r])) \\
&= P(X_n = \lfloor i/S^{m-r} \rfloor | X_{n-1} = \lfloor j/S^{m-r} \rfloor) = a_{i'j'}^X \\
&\text{where } i' = \lfloor i/S^{m-r} \rfloor \text{ and } j' = \lfloor j/S^{m-r} \rfloor.
\end{aligned}$$

It follows that

$$\hat{a}_{ij} = \frac{\sum_{n=1}^{N-1} \xi_n^X \left( \left\lfloor \frac{i}{S^{m-r}} \right\rfloor, \left\lfloor \frac{j}{S^{m-r}} \right\rfloor \right)}{\sum_{n=1}^{N-1} \gamma_n^X \left( \left\lfloor \frac{i}{S^{m-r}} \right\rfloor \right)}. \quad (2.55)$$

The quantities  $\xi_n^X, \gamma_n^X$  are not part of the E1HMM framework, so, in order to make the estimate computable, we express them in terms of  $\xi_n, \gamma_n$ :

$$\begin{aligned}
\gamma_n^X(i') &= P(X_n = i' | \{O_l\}) \\
&= P\left(q_n = i : i' = \left\lfloor \frac{i}{S^{r-m}} \right\rfloor | \{O_l\}\right) \\
&= \sum_{i=i'S^{m-r}}^{(i'+1)S^{m-r}-1} P(q_n = i | \{O_l\}) = \sum_{i=i'S^{m-r}}^{(i'+1)S^{m-r}-1} \gamma_n(i).
\end{aligned}$$

and

$$\begin{aligned}
\xi_n^X(i', j') &= P(X_n = i', X_{n+1} = j' | \{O_l\}) \\
&= \sum_{k=i'S^{m-r}}^{(i'+1)S^{m-r}-1} P(X_n = i', X_{n+1} = j', q_n = k | \{O_l\}) \\
&= \sum_k P(X_{n+1} = j', q_n = k | \{O_l\}) = \\
&= \sum_k P\left(q_{n+1} = \left\lfloor \frac{j'}{S^{r-1}} \right\rfloor S^{m-1} + \left\lfloor \frac{k}{S} \right\rfloor, q_n = k | \{O_l\}\right) \\
&= \sum_{k=i'S^{m-r}}^{(i'+1)S^{m-r}-1} \xi_n\left(k, \left\lfloor \frac{j'}{S^{r-1}} \right\rfloor S^{m-1} + \left\lfloor \frac{k}{S} \right\rfloor\right).
\end{aligned}$$

We know that the transition probability is zero for illegal transitions so finally we obtain the reestimation formula for  $a_{ij}$ , which is given by :

$$\hat{a}_{ij} = \begin{cases} \frac{\sum_{n=1}^{N-1} \sum_{k=k_1(i)}^{k_2(i)} \xi_n \left( k, \left\lfloor \frac{k}{S} \right\rfloor + \left\lfloor \frac{j}{S^{m-1}} \right\rfloor S^{m-1} \right)}{\sum_{n=1}^{N-1} \sum_{k=k_1(i)}^{k_2(i)} \gamma_n(k)} & \text{if } \left\lfloor \frac{i}{S} \right\rfloor = j - \left\lfloor \frac{j}{S^{m-1}} \right\rfloor S^{m-1} \\ 0 & \text{if } \left\lfloor \frac{i}{S} \right\rfloor \neq j - \left\lfloor \frac{j}{S^{m-1}} \right\rfloor S^{m-1} \end{cases} \quad (2.56)$$

where

$$\begin{aligned} k_1(i) &= \left\lfloor \frac{i}{S^{m-r}} \right\rfloor S^{m-r} \\ k_2(i) &= \left( \left\lfloor \frac{i}{S^{m-r}} \right\rfloor + 1 \right) S^{m-r} - 1 = k_1(i) + S^{m-r} - 1. \end{aligned}$$

Note that in this case the inverse transformation is given by

$$\tilde{q}_n = \left\lfloor \frac{q_n}{S^{m-1}} \right\rfloor. \quad (2.57)$$

## 2.2.4 Incorporating additional information

Usually, additional information on the model is known. Any such information on the high-order model can be modified to fit the first-order one, and be incorporated in the procedure. For example, we might have a training state sequence we would like to use for the initial estimation of the parameters. In this case we apply the transformation (2.13) on the sequence to create its first-order equivalent. Another example might be some prior knowledge we have on the transition probabilities. Say we know that  $\tilde{a}_{i_r \dots i_0} = 0$  for some  $i_0, \dots, i_r$ . In this case we shall set  $a_{ij} = 0$  for  $i, j$  that satisfy  $i = f([i_1, \dots, i_r]), j = f([i_0, \dots, i_{r-1}])$ .

### Summary

A method that transforms high order HMMs into equivalent first order ones was introduced. Using this method, one can solve the three HMM problems for any high order HMM, using the well known first order HMM formulation, by modifying it properly. Given in table 2.1 is a summary of the results

Table 2.1: A summarising table, details the transformation suitable for each of the three cases, and the modifications to be made to fit the E1HMM to the transformed model: the states and the reestimation formulas for each of the model parameters.

	$m = r$	$m < r$	$m > r$
the transformation	$q_n = f(\overline{Q}_n^r)$	$q_n = f(\overline{Q}_n^r)$	$q_n = f(\overline{Q}_n^m)$
the inverse transformation	$\tilde{q}_n = \left\lfloor \frac{q_n}{S^{r-1}} \right\rfloor$	$\tilde{q}_n = \left\lfloor \frac{q_n}{S^{r-1}} \right\rfloor$	$\tilde{q}_n = \left\lfloor \frac{q_n}{S^{m-1}} \right\rfloor$
the states	$0, 1, \dots, S^r - 1$	$0, 1, \dots, S^r - 1$	$0, 1, \dots, S^m - 1$
$\hat{\pi}_i$	Eq.(2.19)	Eq.(2.19)	Eq.(2.19)
$\hat{b}_i(\cdot)$	Eq.(2.27) or (2.25)	Eq.(2.44) or (2.47)	Eq.(2.27) or (2.25)
$\hat{a}_{ij}$	Eq.(2.23)	Eq.(2.23)	Eq.(2.56)

derived in this section. For each of the cases ( $m = r$ ,  $m < r$  and  $m > r$ ), the table details the transformation, its inverse, and the modifications to be made to the E1HMM, i.e., the states of the transformed model and the reestimation formulas for each of the parameters.

## 2.3 Order selection in high-order HMMs

Addressed now is the problem of order determination for high-order hidden Markov modeling. To be determined are the order of the hidden state sequence  $r$ , and the order of observation dependency on the states  $m$ . For the purpose of order selection, we assume that a sample state sequence and a corresponding observation sequence are given (e.g., training data). There are several model selection methods in the literature. Most of which share a similar structure, based on optimizing a cost function comprised of a likelihood term and a penalty term that is a function of the number of independent parameters in the competing model. Two examples are Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) ([1], [33], [9]). In this section we apply AIC for determining the model orders. Any of the other criteria can be chosen to determine the orders, as each of them possesses different properties, depending on the application and the data set given. Most of the other criteria are straightforward to use once AIC is constructed, i.e., once the likelihood and number of parameters have been



calculated.

AIC, which is based on minimizing the expectation of the Kullback-Leibler distance, is given by

$$\text{AIC} = -2 \log \mathcal{L} + 2P \quad (2.58)$$

where  $\mathcal{L}$  is the maximized likelihood, obtained by evaluating the likelihood function at the maximum likelihood estimates of the parameters, and  $P$  is the number of independent parameters of the model. Given a data set and several competing models, the chosen model is the one with the lowest AIC value. The other models can be ranked based on their AIC values [9].

The log-likelihood of the high-order HMM is

$$\log P(\{\tilde{q}_n\}, \{O_n\}) = \log P(\{O_n\}|\{\tilde{q}_n\}) + \log P(\{\tilde{q}_n\}). \quad (2.59)$$

For finding  $r$ , only the second term of the RHS of (2.59) is relevant, whereas for finding  $m$ , only the first term is. We therefore construct AIC for  $r$  and  $m$  separately.

### 2.3.1 Determining $r$

Determining  $r$  means determining the order of a Markov chain. This subject has been studied before, and several methods were proposed, e.g. by [27] and [25]. Here we use AIC for determining  $r$ , as suggested by Tong in [18],[35].

The probability of obtaining the state sequence  $\{\tilde{q}_n\}_{n=1}^N$  is

$$P(\{\tilde{q}_n\}_{n=1}^N) = P(\tilde{q}_1)P(\tilde{q}_2|\tilde{q}_1)P(\tilde{q}_3|\tilde{q}_2, \tilde{q}_1) \cdots P(\tilde{q}_N|\tilde{q}_1, \dots, \tilde{q}_{N-1}). \quad (2.60)$$

For a Markov chain of order  $r$ , (2.60) becomes

$$P(\{\tilde{q}_n\}_{n=1}^N) = P(\tilde{q}_1)P(\tilde{q}_2|\tilde{q}_1) \cdots P(\tilde{q}_r|\tilde{q}_1, \dots, \tilde{q}_{r-1}) \prod_{n=r+1}^N P(\tilde{q}_n|\tilde{q}_{n-r}, \dots, \tilde{q}_{n-1}) \quad (2.61)$$

and thus, the log-likelihood is

$$\log P(\{\tilde{q}_n\}) = \sum_{n=1}^r \log P(\tilde{q}_n|\tilde{q}_1, \dots, \tilde{q}_{n-1}) + \sum_{n=r+1}^N \log P(\tilde{q}_n|\tilde{q}_{n-r}, \dots, \tilde{q}_{n-1}). \quad (2.62)$$

The second term becomes dominant as the sample size  $N$  increases, so we neglect the first term, overlooking initial transients, as they do not provide any information regarding the order of the state sequence. For the same reason, the initial state probabilities will be ignored in the parameter count.

We denote by  $N_{i_r, \dots, i_0}$  the number of transitions  $i_r \rightarrow i_{r-1} \rightarrow \dots \rightarrow i_1 \rightarrow i_0$  in the sample state sequence. Ignoring the first  $r$  terms, the likelihood can be written as

$$P(\{\tilde{q}_n\}) \approx \prod_{i_r, \dots, i_0} \tilde{a}_{i_r, \dots, i_0}^{N_{i_r, \dots, i_0}}. \quad (2.63)$$

The maximum likelihood estimate of the transition probability based on the sample data is

$$\hat{\tilde{a}}_{i_r, \dots, i_0} = \frac{N_{i_r, \dots, i_0}}{N_{i_r, \dots, i_1}} \quad (2.64)$$

where

$$N_{i_r, \dots, i_1} = \sum_{i_0} N_{i_r, \dots, i_0}, \quad (2.65)$$

and thus the maximized log-likelihood is

$$\log \mathcal{L} = \sum_{i_r, \dots, i_0} N_{i_r, \dots, i_0} \log \frac{N_{i_r, \dots, i_0}}{N_{i_r, \dots, i_1}}. \quad (2.66)$$

A Markov chain of order  $r$  with  $S$  states has  $S^{r+1}$  transition probabilities, among which only  $S^{r+1} - S^r$  are independent because the probabilities of transitions terminating in the same state sum to 1, that is,  $\sum_{i_0} \tilde{a}_{i_r, \dots, i_0} = 1$ . Therefore, AIC here is given by

$$\text{AIC}(r) = -2 \sum_{i_r, \dots, i_0} N_{i_r, \dots, i_0} \log \frac{N_{i_r, \dots, i_0}}{N_{i_r, \dots, i_1}} + 2S^r(S-1). \quad (2.67)$$

The value for  $r$  to be chosen is the one that minimizes  $\text{AIC}(r)$ , i.e.,

$$r_{AIC} = \arg \min_r \text{AIC}(r)$$

### 2.3.2 Determining $m$

The relevant likelihood here is the first term of (2.59), and thus, for determining  $m$  we treat  $\{\tilde{q}_n\}$  as a given. The state sequence order  $r$  does not play any role in determining  $m$ . We can write

$$P(\{O_n\}|\{\tilde{q}_n\}) = P(O_1|\{\tilde{q}_n\})P(O_2|O_1, \{\tilde{q}_n\}) \cdots P(O_N|\{O_n\}_{n=1}^{N-1}, \{\tilde{q}_n\}), \quad (2.68)$$

and the order  $m$  dependency on the states (2.8) implies that

$$P(\{O_n\}|\{\tilde{q}_n\}) = P(O_1|\tilde{q}_1)P(O_2|\tilde{q}_1, \tilde{q}_2) \cdots \\ \cdots P(O_{m-1}|\tilde{q}_1, \dots, \tilde{q}_{m-1}) \prod_{n=m}^N P(O_n|\tilde{q}_{n-(m-1)}, \dots, \tilde{q}_n). \quad (2.69)$$

The log-likelihood is then

$$\log P(\{O_n\}|\{\tilde{q}_n\}) = \sum_{n=1}^{m-1} \log P(O_n|\tilde{q}_1, \dots, \tilde{q}_n) + \sum_{n=m}^N \log P(O_n|\tilde{q}_{n-(m-1)}, \dots, \tilde{q}_n). \quad (2.70)$$

As the sample size  $N$  increases, the second term becomes more dominant and the first term negligible. For convenience, we use the transformation (2.13) and define

$$q_n^m = \sum_{l=0}^{m-1} \tilde{q}_{n-l} S^{m-1-l}. \quad (2.71)$$

The likelihood can now be written, ignoring the first  $m - 1$  terms, as

$$P(\{O_n\}|\{\tilde{q}_n\}) \approx \prod_{n=m}^N b_{q_n^m}(O_n) \quad (2.72)$$

where  $b_i(O_n) = P(O_n|q_n^m = i)$ .

We distinguish between the cases of discrete valued observations and continuous valued ones.

## Discrete Observations

We assume here that each observation  $O_n$  is drawn from a finite alphabet  $\{v_1, v_2, \dots, v_K\}$ , according to a probability distribution associated with the last  $m$  states  $\{\tilde{q}_{n-(m-1)}, \dots, \tilde{q}_n\}$ , or equivalently, with the current transformed state  $q_n^m$ . We denote by  $N_{k|i_1, \dots, i_m}$  the number of appearances of the succession  $i_1, \dots, i_m$  in the sample state sequence  $\{\tilde{q}_n\}$ , that generated the observation  $v_k$ , i.e.,

$$N_{k|i_1, \dots, i_m} = |\{n : \tilde{q}_{n-(m-1)} = i_1, \dots, \tilde{q}_n = i_m \text{ and } O_n = v_k\}| \quad (2.73)$$

where  $|\cdot|$  stands for the number of elements in a set. Equivalently, we denote

$$N_{k|i} = |\{n : q_n^m = i \text{ and } O_n = v_k\}|, \quad (2.74)$$

i.e., the number of times that state  $i$  in the transformed sequence  $\{q_n^m\}$  generated the observation  $v_k$ . The likelihood (2.72) can now be written as

$$P(\{O_n\}|\{\tilde{q}_n\}) \approx \prod_{i,k} b_i^{N_{k|i}}(v_k). \quad (2.75)$$

It can be easily verified that the maximum likelihood estimate of  $b_i(v_k)$  in the discrete case is

$$\hat{b}_i(v_k) = \frac{N_{k|i}}{N_i} \quad (2.76)$$

where  $N_i = \sum_k N_{k|i}$ , and thus, the maximized log-likelihood is

$$\log \mathcal{L} = \sum_{i,k} N_{k|i} \log \frac{N_{k|i}}{N_i}. \quad (2.77)$$

Each of the  $S^m$  state combinations corresponds to a probability distribution over the observation alphabet. Each of those distributions consists of  $K$  probabilities that sum to one. Therefore, the total number of independent parameters is  $S^m(K - 1)$ . AIC here is thus given by

$$\text{AIC}(m) = -2 \sum_{i=0}^{S^m-1} \sum_{k=1}^K N_{k|i} \log \frac{N_{k|i}}{N_i} + 2S^m(K - 1), \quad (2.78)$$

and  $m_{AIC} = \arg \min_m \text{AIC}(m)$ .

## Continuous Observations

We assume here that each observation  $O_n$  is a  $D$ -dimensional vector drawn from a Gaussian mixture pdf associated with the last  $m$  states  $\{\tilde{q}_{n-(m-1)}, \dots, \tilde{q}_n\}$ , or equivalently, with the current transformed state  $q_n^m$ . Therefore, the observation pdf for state  $i$  is of the form

$$b_i(x) = \sum_{k=1}^K c_{ik} g(x; \mu_{ik}, U_{ik}) \quad (2.79)$$

where  $c_{ik}$  is the mixture coefficient for the  $k$ th mixture in state  $i$  and  $g(x; \mu_{ik}, U_{ik})$  is the  $D$ -dimensional Gaussian pdf with mean vector  $\mu_{ik}$  and covariance matrix  $U_{ik}$  for the  $k$ th mixture component in state  $i$ :

$$g(x; \mu_{ik}, U_{ik}) = \frac{1}{\sqrt{(2\pi)^D \det U_{ik}}} e^{-\frac{1}{2}(x-\mu_{ik})^T U_{ik}^{-1} (x-\mu_{ik})}. \quad (2.80)$$

In the case of i.i.d. measurements drawn from a GM, the Expectation-Maximization (E-M) algorithm is commonly used to provide the maximum likelihood estimates of the parameters [8]. In our case, the measurements (observations) are independent (given the state sequence) but are *not* identically distributed. Only the observations that were generated from the same state are i.i.d.. Therefore, a proper modification has to be made to the E-M formulas of the i.i.d GM. The complete derivation for the i.i.d case is given in (e.g.) [8]. We shall not completely rederive the E-M equations for this case, but rather follow the derivation in [8], pointing out where our case differs.

First, we define the following:

- $\{y_n\}$  will denote the mixture component indicator sequence, i.e.,

$$P(y_n = k | q_n^m = i) = c_{ik}. \quad (2.81)$$

Given the state sequence  $\{q_n^m\}$ , the  $y_n$ 's are independent and each  $y_n$  depends only on the current state  $q_n^m$ .

- For convenience, we define the following parameter sets

$$\Theta = \{\{c_{ik}\}, \{\mu_{ik}\}, \{U_{ik}\}\} \quad (2.82)$$

$$\theta_{ik} = \{\mu_{ik}, U_{ik}\}. \quad (2.83)$$

- The expected value, with respect to the current set of parameters  $\Theta'$ , of the log-likelihood, given the observed data

$$Q(\Theta, \Theta') = E_{\Theta'} \log P(\{O_n\}, \{y_n\} | \{q_n^m\}; \Theta) | \{O_n\}. \quad (2.84)$$

At each iteration, the E-M algorithm maximizes  $Q(\Theta, \Theta')$  over  $\Theta$  where  $\Theta'$  are the current parameter estimates. The optimized parameters  $\Theta$  are then used as the current estimates in the next iteration.

Expanding  $P(\{O_n\}, \{y_n\} | \{q_n^m\}; \Theta)$  gives

$$\begin{aligned} P(\{O_n\}, \{y_n\} | \{q_n^m\}; \Theta) &= P(\{O_n\} | \{y_n\}, \{q_n^m\}; \Theta) P(\{y_n\} | \{q_n^m\}; \Theta) \\ &= \prod_n c_{q_n^m y_n} g(O_n; \theta_{q_n^m y_n}), \end{aligned} \quad (2.85)$$

and thus  $Q(\Theta, \Theta')$  can be written as

$$\begin{aligned} Q(\Theta, \Theta') &= \sum_{y_1, \dots, y_N} P(\{y_n\} | \{q_n^m\}, \{O_n\}; \Theta') \log P(\{O_n\}, \{y_n\} | \{q_n^m\}; \Theta) \\ &= \sum_{y_1, \dots, y_N} P(\{y_n\} | \{q_n^m\}, \{O_n\}; \Theta') \sum_n \log(c_{q_n^m y_n} g(O_n; \theta_{q_n^m y_n})) \end{aligned}$$

$$\begin{aligned}
&= \sum_n \sum_{y_n} \log(c_{q_n^m y_n} g(O_n; \theta_{q_n^m y_n})) \sum_{y_{n'}: n' \neq n} P(\{y_n\} | \{q_n^m\}, \{O_n\}; \Theta') \\
&= \sum_n \sum_{y_n} \log(c_{q_n^m y_n} g(O_n; \theta_{q_n^m y_n})) P(y_n | \{q_n^m\}, \{O_n\}; \Theta') \\
&= \sum_n \sum_{k=1}^K \log(c_{q_n^m k} g(O_n; \theta_{q_n^m k})) P(y_n = k | q_n^m, O_n; \Theta') \\
&= \sum_{i=0}^{S^m-1} \sum_{n: q_n^m=i} \sum_{k=1}^K \log(c_{ik} g(O_n; \theta_{ik})) P(y_n = k | q_n^m = i, O_n; \Theta'). \quad (2.86)
\end{aligned}$$

The E-M formulas are derived by maximizing  $Q(\Theta, \Theta')$  over  $\Theta$  subject to the stochastic constraint on the mixture coefficients

$$\sum_k c_{ik} = 1 \quad \forall i \in \{0, 1, \dots, S^m - 1\}. \quad (2.87)$$

By expanding (2.86) we get

$$\begin{aligned}
Q(\Theta, \Theta') &= \\
&\sum_{i=0}^{S^m-1} \sum_{n: q_n^m=i} \sum_{k=1}^K \log(c_{ik}) P_i(y_n = k | O_n; \Theta') \\
&+ \sum_{i=0}^{S^m-1} \sum_{n: q_n^m=i} \sum_{k=1}^K \log(g(O_n; \theta_{ik})) P_i(y_n = k | O_n; \Theta') \quad (2.88)
\end{aligned}$$

where  $P_i(y_n = k | O_n; \Theta') = P(y_n = k | q_n^m = i, O_n; \Theta')$ . It follows that the formulas for  $c_{ik}$  and for  $\mu_{ik}, U_{ik}$  can be derived separately. The formula for  $c_{ik}$  is obtained by using Lagrange multipliers:

$$\begin{aligned}
\frac{\partial}{\partial c_{ik}} \left( Q(\Theta, \Theta') - \sum_j \eta_j \left( \sum_l c_{jl} - 1 \right) \right) &= 0 \Rightarrow \\
\sum_{n: q_n^m=i} P_i(y_n = k | O_n; \Theta') &= \eta_i c_{ik} \quad (2.89)
\end{aligned}$$

where the  $\eta_i$ 's are the Lagrange multipliers. Summing both sides over  $k$  yields

$$\eta_i = N_i \quad (2.90)$$

where  $N_i$  is the number of appearances of state  $i$  in the sequence  $\{q_n^m\}$ . The formulas for  $\mu_{ik}$  and  $U_{ik}$  are obtained by unconstrained maximization of

$Q(\Theta, \Theta')$  (the symmetry and positive-definiteness of the covariance matrices are satisfied automatically):

$$\frac{\partial Q(\Theta, \Theta')}{\partial \mu_{ik}} = \frac{\partial}{\partial \mu_{ik}} \sum_{n:q_n^m=i} \log(g(O_n; \theta_{ik})) P_i(y_n = k | O_n; \Theta') = 0 \quad (2.91)$$

$$\frac{\partial Q(\Theta, \Theta')}{\partial U_{ik}} = \frac{\partial}{\partial U_{ik}} \sum_{n:q_n^m=i} \log(g(O_n; \theta_{ik})) P_i(y_n = k | O_n; \Theta') = 0 \quad (2.92)$$

The above equations, excluding the range of the summing index  $n$ , are identical to their equivalents in the i.i.d. case (This is clear since the observations are i.i.d for a given  $i$ ). Therefore, the formulas for  $\mu_{ik}$  and  $U_{ik}$  are obtained by modifying the range of  $n$  in the i.i.d. case equations. We obtain, then, the following procedure:

**E-step**

$$P_i(y_n = k | O_n; \Theta') = \frac{c_{ik} g(O_n; \mu'_{ik}, U'_{ik})}{\sum_{l=1}^K c_{il} g(O_n; \mu'_{il}, U'_{il})} \quad (2.93)$$

**M-step**

$$c_{ik}^{new} = \frac{1}{N_i} \sum_{n:q_n^m=i} P_i(y_n = k | O_n; \Theta') \quad (2.94)$$

$$\mu_{ik}^{new} = \frac{\sum_{n:q_n^m=i} P_i(y_n = k | O_n; \Theta') O_n}{\sum_{n:q_n^m=i} P_i(y_n = k | O_n; \Theta')} \quad (2.95)$$

$$U_{ik}^{new} = \frac{\sum_{n:q_n^m=i} P_i(y_n = k | O_n; \Theta') (O_n - \mu_{ik}^{new})(O_n - \mu_{ik}^{new})^T}{\sum_{n:q_n^m=i} P_i(y_n = k | O_n; \Theta')} \quad (2.96)$$

The likelihood (2.72) in this case becomes

$$L = \prod_i \prod_{n:q_n^m=i} b_i(O_n) \quad (2.97)$$

and is maximized by substituting the parameters obtained by the E-M algorithm.

The number of independent parameters is as follows

- $S^m K$  mixture coefficients that sum to one for each state, and thus there are  $S^m(K - 1)$  independent ones.
- $S^m K$  mean vectors, each contains  $D$  elements. therefore:  $S^m K D$  mean vector parameters.
- $S^m K$  covariance matrices, each contains  $D^2$  entries, among which only  $D(D + 1)/2$  (only upper/lower triangle) are independent due to symmetry.

The number of independent parameters sums to  $S^m(K - 1 + KD + KD(D + 1)/2) = S^m(K(D + 1)(D + 2)/2 - 1)$ , and finally, AIC here is

$$\text{AIC}(m) = -2 \sum_{i=0}^{S^m-1} \sum_{n:q_n^m=i} \log b_i(O_n) + S^m(K(D + 1)(D + 2) - 2). \quad (2.98)$$

### 2.3.3 Summary

A method that transforms high-order HMMs into equivalent first-order ones was introduced. Using this method, one can solve the three HMM problems for any high-order HMM by applying the well known first-order HMM formulation. This is done by using a E1HMM with  $S^\nu$  states and the reestimation formulas given in Fig. 2.1. The high-order hidden state sequence is then obtained by applying the inverse transformation on the first-order one produced by the E1HMM. We assumed in this work that the observation distributions in the discrete case are over a finite alphabet with all probabilities unknown. This can be easily extended to cases of parametric distributions (e.g. Poisson).

Discussed was also the problem of order selection, that naturally arises when approaching a problem with hidden Markov modeling. AIC was applied to determine the model orders  $r$  and  $m$ . Further study can be made on investigating the consistency in estimating  $m$ .

In the next chapter we apply the HMM to detect rain in power measurements from cellular networks.



# Chapter 3

## Using the HMM to detect wet and dry periods

The attenuation is affected by a variety of physical phenomena. Therefore, to estimate rain, we try to separate the rain induced attenuation from the total attenuation. To differentiate between wet and dry periods, given a sequence of RSL measurements from a link, a HMM was employed. In this chapter, we first apply the first-order HMM (E1HMM) to the problem and examine the results. Next, we examine whether the first-order assumptions hold in our problem. Then, we apply the High-order HMM and compare the results to those obtained by the first-order HMM. An attempt to estimate the rain induced attenuation using blind source separation (BSS) techniques was also made, and is detailed in appendix A.

The idea of modeling the temporal occurrence of rain as a Markov chain was proposed and studied in the past (e.g., by [17] and [18]). However, only daily occurrences of rain were studied, and higher temporal resolutions (e.g, of 1 minute, as in our case) were not discussed.

### 3.1 Using the first order HMM

Our interest is in identifying the wet and dry periods. Therefore, we define the hidden state sequence as the rain indicator series:

$$q_n = I_n = \begin{cases} 0 & , \text{ no rain at time } n \\ 1 & , \text{ rain at time } n \end{cases} . \quad (3.1)$$

The RSL measurements are the observations generated by the state sequence, i.e.,  $O_n = P_n$  (see (1.3)). In some of the links, both ends transmit and receive and thus there are two sets of RSL measurements. In that case, where we

have two opposite links,  $O_n$  is a two-dimensional vector containing both RSL measurements.

### 3.1.1 Initializing the HMM

The third HMM problem P3 (see section 2.1.2) is the parameter estimation problem. The procedure for solving this problem is detailed in [29], and is outlined in section 2.2. That procedure is known as the Baum-Welch reestimation algorithm, and it involves iterative update and improvement of the model parameters. The reestimation procedure is guaranteed to converge to a local maximum of  $P(\{O_n\}; \lambda)$  and thus it is sensitive to the initial guess of the model parameters  $\lambda$ .

For a decent initial estimate of the parameters, an initial state sequence (rain indicator series) is needed. One option is to use a nearby rain gauge to identify the wet and dry periods. However, since a rain gauge is not always close at hand, we would rather use a method based on the RSL data. In the case of opposite links, the correlation between attenuations at the two links tends to be high in times of rain because the rain causes well defined attenuation changes in a similar fashion for both links (usually the frequencies of the links are close) [31]. Therefore, in the case of opposite links, the local correlation coefficient between the two links at each time step was computed, and was compared to a threshold to determine the initial wet and dry periods. The threshold, and the segment length (in minutes) on which the correlation coefficient was computed, were determined experimentally. A threshold of 0.6 was used and a segment length of 9 minutes.

In the presence of rain, the attenuation tends to vary more rapidly [20]. Therefore, in the cases where an opposite link was not available, the standard deviation of the RSL was computed at each time step on a segment length of 9 minutes, and was compared to a threshold to determine the wet and dry periods. Since the links vary in their RSL values, a different threshold should be assigned to each link. Therefore, we took the threshold to be the mean of the RSL's standard deviation.

In the case of HMMs with continuous observations, further segmentation is needed for each state to distinguish between the mixture components of the pdf (see section 2.2). In that case, the K-means segmentation algorithm was used [34].

### 3.1.2 Results

Examined here are the results obtained by applying the first-order HMM on a series of RSL measurements from a single link, on a time period of 16

hours. The two ends of the link are on the Ramle switch and in Ramat-Dan. The rain indicator series produced by the HMM is compared to that of a rain gauge located on the Ramle switch. Since the sequences to be compared are binary, a natural selection for a comparison metric is the normalized Hamming distance, denoted here by  $d$ , i.e.,

$$d = \frac{\text{number of entries that are different}}{\text{length of sequence}} \quad (3.2)$$

Shown in Fig 3.1 are the series of RSL measurements, the rain indicator produced by the HMM, the rain gauge measurements and the "real" rain indicator, that is, zero when the rain gauge measures zero rain, and one otherwise.

The distance between the link rain indicator and the rain gauge one is  $d = 129/939 \approx 0.137$ .

Notice that the rain gauges and microwave links provide measurements of different nature. The rain gauges provide a point measurement while the links measure along a path. Therefore, identical results ( $d = 0$ ) are not possible because some of the rain measured by the links cannot be detected by the gauges.

## 3.2 Does the first order assumption hold?

Eventhough the first-order HMM produces decent results in our problem, we examine here whether the first order assumptions actually hold. As suggested in section 2.3, we use Akaike's Information Criterion (AIC) to determine the order of the state sequence  $r$ , and the order of the observation dependency  $m$ . The values for  $r$  and  $m$  are determined by applying AIC to sample data comprised of a state sequence and the corresponding observation sequence.

A sample data of 16 hours was used here to determine the orders. For choosing  $r$ , we applied AIC on the rain indicator of the Ramle switch rain gauge. The resulting minimum AIC estimate (MAICE) here is  $r = 5$ . For choosing  $m$  we used the rain indicator as the state sequence, and the measured RSL (from the Ramle switch - Ramat Dan link) as the observation sequence. The resulting MAICE is  $m = 3$ . Shown in Fig.3.2 are AIC values for both cases. Notice that AIC values for  $r$  are very similar between 2 and 5. This implies that values of  $r$  greater than 2 will not improve model fitness significantly.

AIC was applied to several other data sets of links and gauges to estimate the orders. The MAICE of the Markov order of all the rain gauge sequences ranged between  $r = 3$  and  $r = 5$ . The MAICE of the state dependency orders ranged between  $m = 2$  and  $m = 4$ .

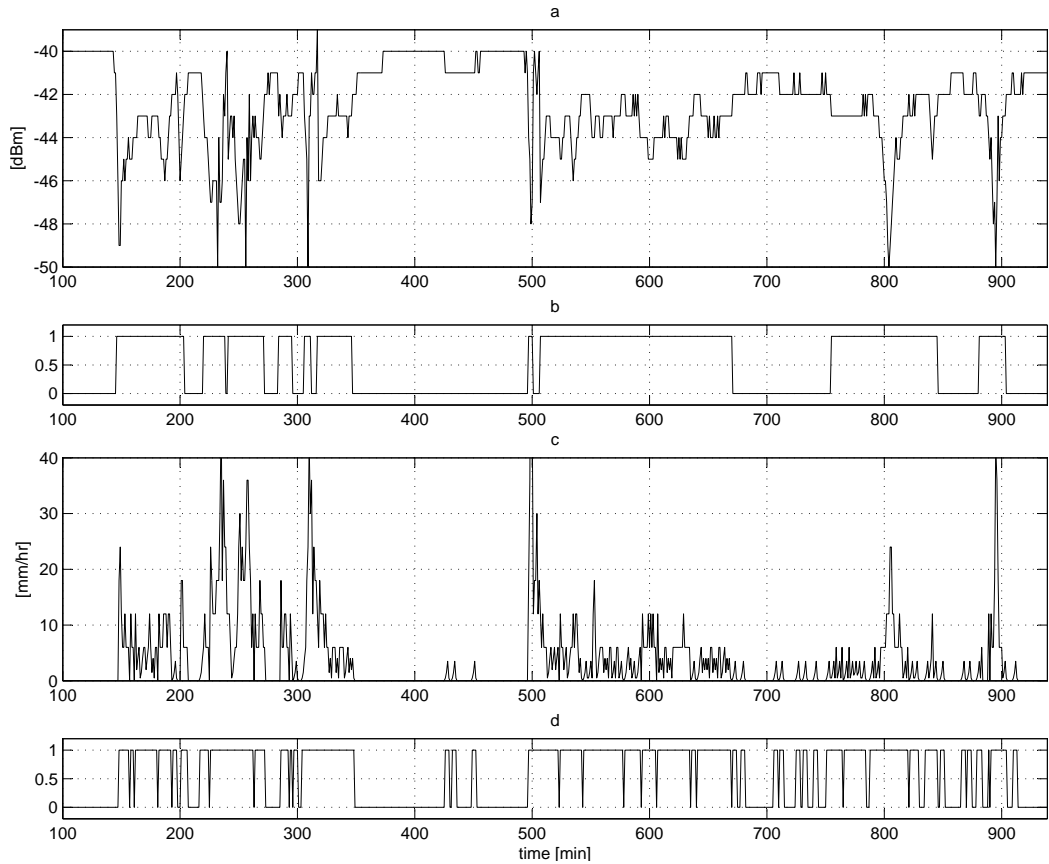


Figure 3.1: Results of a 16 hour long data. a: the RSL measurements. b: the rain indicator produced by the E1HMM. c: the rain measured by the rain gauge. d: rain indicator of the rain gauge.

### 3.3 Applying the higher order HMM

We applied the high-order HMM, which was described in section 2.2, to the data for detecting wet and dry spells. The data used is the same as in the previous section, and again, the observations are the RSL measurements, and the hidden state sequence is the rain indicator. The initialization procedure used is similar to the one described in 3.1.1. In this case we applied the transformation (2.13) to the initial state sequence to obtain the high-order initial state sequence, as suggested in 2.2.4.

Applying HMMs of higher orders introduces problems of different sorts. For example, we noticed that high-order HMMs are very sensitive to the initial parameter estimates, compared with the first-order HMM. This sensitivity increases as we increase the orders. Another two issues are the running

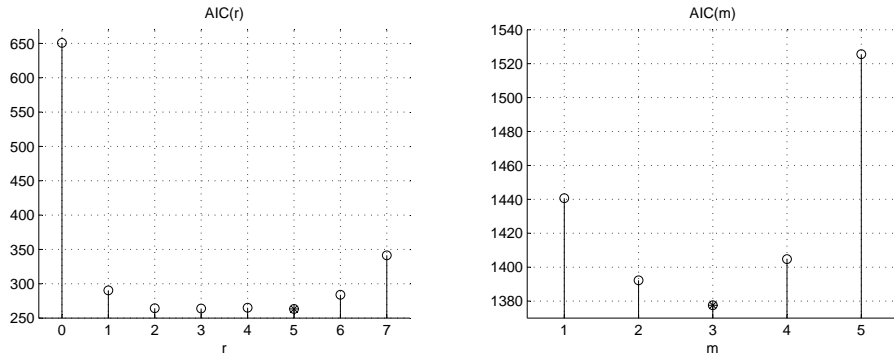


Figure 3.2: AIC values for the order of the state sequence  $r$ , and the order of the observation dependency  $m$ . For choosing  $r$ , AIC was applied on the rain indicator produced by the rain gauge. For choosing  $m$  we used the rain indicator as the state sequence, and the measured RSL as the observation sequence. The minimum AIC estimate (MAICE) in both cases is indicated by \*.

Table 3.1: Hamming distance between the rain indicator of the rain gauge, and the one obtained by applying the high-order HMM to the RSL data, for different values of  $r$  and  $m$ .

$(r, m)$	Hamming distance	$d$
(1,1)	137	0.146
(2,1)	133	0.142
(3,2)	130	0.138

time and the number of parameters to estimate. Both grow large with the order. The latter, however, is already taken into account in AIC, when selecting the orders.

### 3.3.1 Results

Table 3.1 details the results obtained by applying the high-order HMM on the data. Shown, for each of the  $(r, m)$  pairs tested, are both the Hamming distance (number of different entries) and the normalised one, between the resulting rain indicator and that of the rain gauge. Figure 3.3 shows the rain indicators obtained by applying HMMs of different orders, against that of the rain gauge.

Eventhough the "right" order of our model is not first, we see that the improvement obtained by increasing the orders of the HMM is negligible, and does not justify the increased complexity of the algorithm.

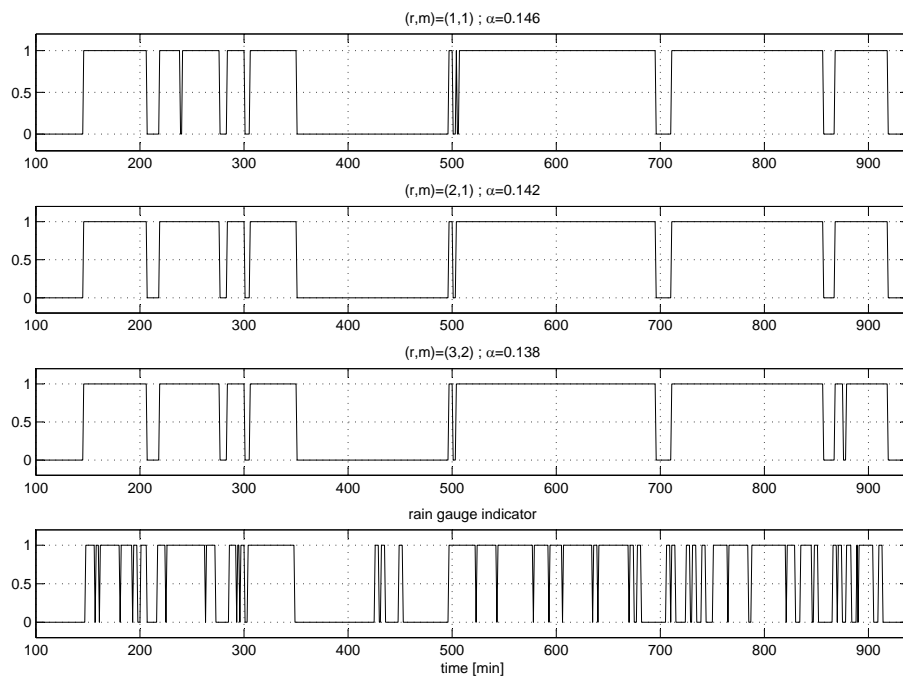


Figure 3.3: Rain indicators obtained by different orders of HMMs against that of the rain gauge.

# Chapter 4

## Estimating rain

### 4.1 Smoothing the Data

The RSL data received by the cellular provider is quantized in a resolution of 1dB (rounded). When converting to rain-rate, this results in an unnatural and irregular behaviour. Therefore, a smoothing procedure, described in this section, was applied to the data.

Let us denote the rounding operation by  $R(\cdot)$ . Assume we are given a quantized version of a signal,  $Y_n = R(X_n)$ . The idea is to optimize some smoothness criterion while keeping the reconstructed signal  $\hat{X}_n$  loyal to the measurements. Two approaches were used:

#### 4.1.1 Smoothing by constrained nonlinear optimization

Given  $Y_n$ , the quantized version of the signal  $X_n$ , The smoothness criterion chosen for the dequantization is of the form

$$\sum_n |\hat{X}_n - \hat{X}_{n-1}|^c \quad (4.1)$$

where  $c > 0$ . The dequantized signal  $\hat{X}_n$  is obtained, then, by minimizing (4.1) subject to the fidelity constraint

$$R(\hat{X}_n) = Y_n \quad \forall n. \quad (4.2)$$

Figure 4.1 shows the result obtained by applying nonlinear smoothing on a 80 minute RSL sequence, with  $c = 2$ . The RSL sequence itself was used as the initial guess for the optimization.

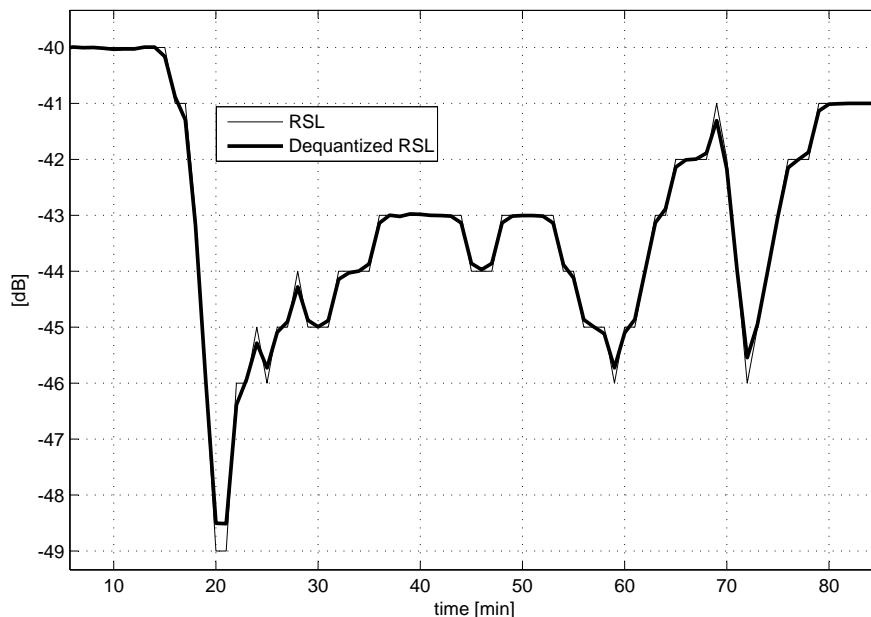


Figure 4.1: Smoothing by constrained nonlinear optimization with  $c = 2$ .

Further study can be made on choosing the right value for  $c$ , and the optimization algorithm to be used.

Results obtained by the nonlinear optimization have perfect fidelity to the measurements. However, the optimization is computationally expensive and sensitive to the initial guess. The next approach trades off fidelity in favor of lower running time and initial guess insensitivity.

### 4.1.2 Smoothing by linear filtering

Here, relaxing the constraint (4.2), we minimize the following functional

$$F = \sum_n (x_n - x_{n-1})^2 + \alpha \sum_n (x_n - Y_n)^2. \quad (4.3)$$

where  $\alpha > 0$  represents the weight given to the fidelity term (the second term of (4.3)), compared to that given to the smoothness term. The dequantized signal  $\hat{X}_n$  is obtained by

$$\{\hat{X}_n\} = \arg \min_{\{x_n\}} F. \quad (4.4)$$

Differentiating  $F$  with respect to  $x_m$  yields

$$\frac{\partial F}{\partial x_m} = 2(2 + \alpha)x_m - 2(x_{m-1} + x_{m+1}) - 2\alpha Y_m \quad (4.5)$$



and setting the derivative to zero

$$\frac{\partial F}{\partial x_m} = 0 \Rightarrow \quad (4.6)$$

$$(2 + \alpha)x_m - (x_{m-1} + x_{m+1}) = \alpha Y_m. \quad (4.7)$$

Applying the  $Z$ -transform gives the dequantization filter

$$H(z) = \frac{\hat{X}(z)}{Y(z)} = \frac{\alpha}{2 + \alpha - (z^{-1} + z)} \quad (4.8)$$

or

$$H(e^{j\omega}) = \frac{\alpha}{\alpha + 2(1 - \cos \omega)}. \quad (4.9)$$

Notice that when  $\alpha = 0$  (zero weight to fidelity) the filter outputs the smoothest signal possible (zero) whatever the measurements. When  $\alpha \rightarrow \infty$  (zero weight to smoothness) the filter equals one, and outputs the measurements. Fig 4.2 shows the frequency response of the dequantization filter  $H(e^{j\omega})$  for different values of  $\alpha$ . Notice that as  $\alpha$  grows larger, less weight is given to smoothness, and the filter changes from 'Low-Pass' to 'All-Pass'. Shown in Fig. 4.3 is the dequantized RSL obtained by linear filtering with  $\alpha = 2$ . Further study can be made on choosing  $\alpha$ .

## 4.2 Converting to rain rate

We denote the rain indicator produced by the HMM by  $I_n$ . The set of dry times is denoted by  $S_0$ , i.e.,

$$S_0 = \{n : I_n = 0\}. \quad (4.10)$$

At dry times, the rain induced attenuation equals zero, i.e.,  $A_n = 0 \quad \forall n \in S_0$ . Therefore, the zero-rain attenuation  $Z_n$  (see eq. (1.4)) at dry times is equal to the total attenuation (which is given by (1.3)), i.e.,

$$Z_n = A_n^{total} \quad \forall n \in S_0 \quad (4.11)$$

At times of rain, i.e., for  $n \notin S_0$ , the values of  $Z_n$  were interpolated from  $\{Z_n, n \in S_0\}$  using cubic interpolation. Then, the rain induced attenuation is computed using (1.4):

$$A_n = A_n^{total} - Z_n, \quad (4.12)$$

and the rain rate by (1.2).

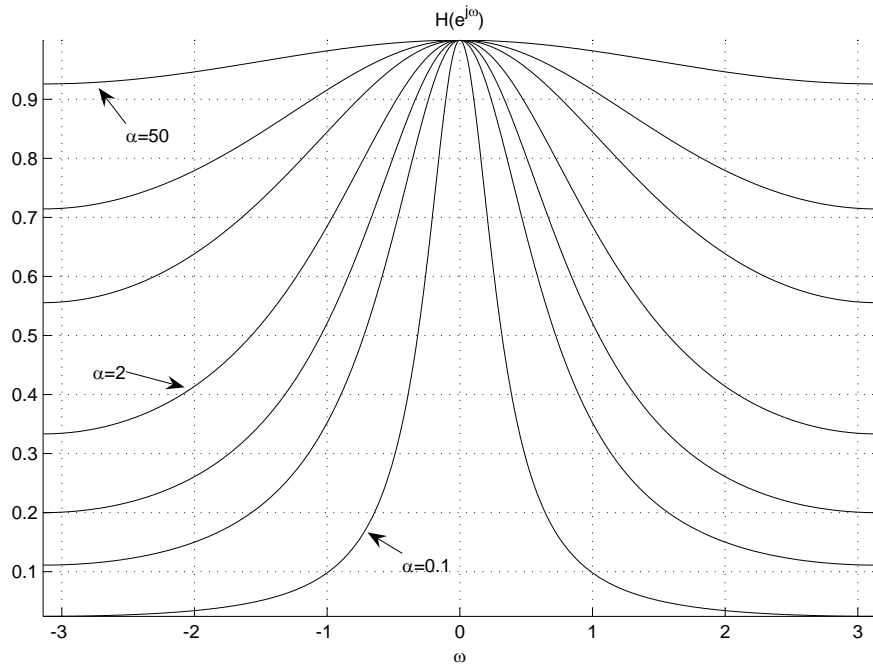


Figure 4.2: Frequency response of dequantization filter  $H(e^{j\omega})$ .

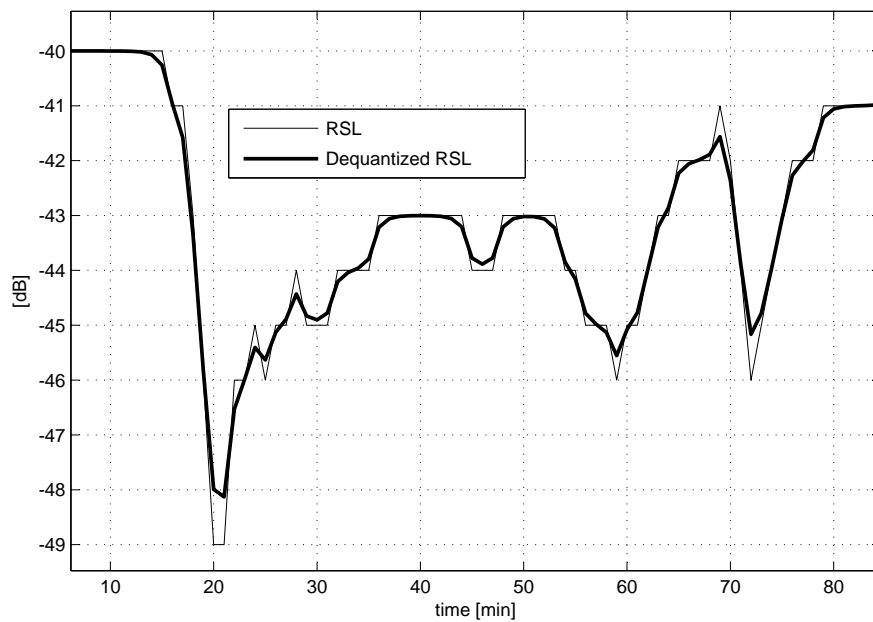


Figure 4.3: Smoothing by linear filtering.

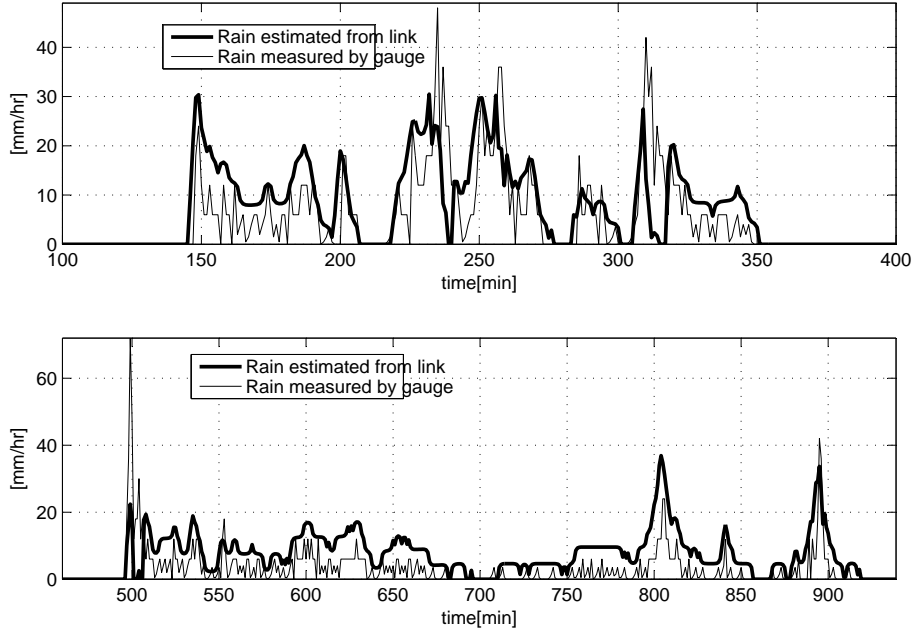


Figure 4.4: Rain estimated from link against measurements from rain gauge.

### 4.3 Results

Shown in Fig. 4.4 is the rain estimated from the link measurements against the rain measured by a rain gauge located on one end of the link. The data used is the same as in section 3.1.2. The rain indicator was created using the first-order HMM, and the RSL data was smoothed (after the HMM) using linear filtering with  $\alpha = 2$ . The parameters used for eq. (1.2) are as follows: the length of the link is  $L = 2.56\text{km}$ . The values for  $a$  and  $b$ , as suggested by [20], were calculated using

$$a = \begin{cases} 4.21 \cdot 10^{-5} f^{2.49} & , \quad 2.9 \leq f \leq 54 \\ 4.09 \cdot 10^{-2} f^{0.699} & , \quad 54 \leq f \leq 180 \end{cases} \quad (4.13)$$

$$b = \begin{cases} 1.41 f^{-0.0779} & , \quad 8.54 \leq f \leq 25 \\ 2.63 f^{-0.272} & , \quad 25 \leq f \leq 164 \end{cases} \quad (4.14)$$

where  $f[\text{GHz}]$  is the frequency of the link, in our case 17.81GHz.

The method above was used for geophysical analysis of several rain events between the years 2006-2008 in different locations across Israel. For example, the average correlation (over the rain events analyzed) between the rain measured by the Ramle switch rain gauge, and the rain estimated from the links is 0.77 for 1min resolution, and 0.85 for 10min. In a rain gauge located

in west Ramle, the correlations are 0.64 for 1min, and 0.83 for 10min. The full geophysical analysis of the results obtained by the method can be found in [30].

# Chapter 5

## Summary

Discussed was high-order hidden Markov modeling. A method was suggested that makes the well known first-order HMM algorithm applicable to models of higher orders, by transforming them into first-order equivalents. Furthermore, the modifications necessary for the first-order HMM algorithm to fit the transformed model were detailed. To complete the discussion, order selection for high-order hidden Markov modeling was addressed. Given a sample set of data, a model selection criterion can be used to determine the optimal orders for modeling the problem as an HMM. The aforementioned method can be then applied to solve the problem using the first-order HMM algorithm.

Suggested also was an HMM based method for rain estimation from RSL measurements in cellular networks. We see that the suggested method produced good results in terms of comparison with rain measured by rain gauges. The method, thus, can provide a mean to measure rainfall with high temporal and spatial resolution, compared with the existing methods of rain gauges and weather radars. Furthermore, since the method relies on data from existing cellular networks, it does not involve additional maintenance, supervision or cost.

By using AIC, we saw that modeling our problem as a first-order HMM introduces model mismatching. Despite that, we saw that the improvement in detecting rain, obtained by applying HMMs of higher orders is negligible compared to the complexity and increased running time they introduce. Further study can be made on expanding the use of the HMM to analyze links with temporal resolution coarser than 1min, and to analyze several links simultaneously.



# Appendix A

## Rainfall estimation using BSS

The problem addressed is the estimation of rainfall intensity using power measurements from an existing network of microwave links. Since the attenuation is affected by a variety of physical phenomena, in order to estimate rainfall intensity, our first goal is to separate the rain induced attenuation from the total attenuation. This chapter presents selected preliminary experimental results of an attempt to apply Blind Source Separation (BSS) [11] to the problem at hand.

### A.1 A brief overview of BSS

Given a set of observed signals, assuming each signal in the set is a mixture, or a linear combination of a set of independent signals (the 'sources'), BSS is an emerging technique aiming at recovering the unobserved signals from the observed mixtures, exploiting only the assumption of mutual independence between the signals. The simplest BSS model assumes the existence of  $N$  independent signals  $s_1(t), \dots, s_N(t)$  and the observation of as many mixtures  $x_1(t), \dots, x_N(t)$ , these mixtures being linear and instantaneous. This is compactly represented by the mixing equation

$$\underline{x}(t) = \mathbf{A}\underline{s}(t) \tag{A.1}$$

where  $\underline{s}(t) = [s_1(t), \dots, s_N(t)]^T$  is the  $N \times 1$  column vector collecting the sources,  $\underline{x}(t)$  similarly collects the observed signals, and  $\mathbf{A}$  is the  $N \times N$  mixing matrix containing the unknown mixture coefficients [11]. The BSS problem consists in recovering the source vector  $\underline{s}(t)$  using only the observed data  $\underline{x}(t)$ , the assumption of independence between the entries of the input vector  $\underline{s}(t)$  and possibly some a priori information about the probability distribution of the inputs. It can be formulated as the computation of an

$N \times N$  'separating matrix'  $\mathbf{B}$  whose output  $\underline{y}(t)$

$$\underline{y}(t) = \mathbf{B}\underline{x}(t) \quad (\text{A.2})$$

is an estimate of the vector  $\underline{s}(t)$  of the source signals.

## A.2 Applying BSS

We denote the RSL of a link, recorded at time  $n$ , by  $P_n$ . Given that the transmitted signal level of the link is a constant denoted by  $P^t$ , the attenuation at time  $n$  is given by

$$A_n[\text{dB}] = -10 \log_{10} \frac{P_n[\text{W}]}{P^t[\text{W}]} = P^t[\text{dBm}] - P_n[\text{dBm}]. \quad (\text{A.3})$$

The data we possess includes only the RSL  $P_n$ . The attenuation is obtained from the measured RSL by subtracting the transmitted power  $P^t$ , which is a constant affecting only the "DC level" of  $A_n$ . Since the transmitted power is not available, we used arbitrarily  $P^t = \max_n P_n$ .

We assume that each attenuating physical phenomenon is realized by a factor multiplying the power level, i.e.  $P_n[\text{W}] = P^t[\text{W}] \prod_i a_i$  where  $a_i$  is the attenuation caused by the corresponding phenomenon. Thus, the total attenuation in dB is given by  $-\sum_i 10 \log_{10} a_i = -\sum_i a_i[\text{dB}]$ . This justifies the use of the BSS model that assumes a linear relation between the sources (contributions of different phenomena to the attenuation) and the observations (RSL or total attenuation in dB of the link). Since the different links vary in their physical properties, e.g., location, direction, carrier frequency and polarization, every phenomenon affects differently on each link. These differences are expressed by the mixing coefficients populating the mixing matrix  $\mathbf{A}$ , which is assumed to be time invariant.

## A.3 Experimental results

A preliminary experiment, addressing the separation of rainfall-induced attenuation from other attenuation sources, was conducted by applying BSS on several links. The results were obtained by using "icalab", a BSS toolbox for MATLAB that features a variety of BSS algorithms [13]. Each input signal used is the attenuation of the corresponding link, normalized by the length of the link, i.e., the input attenuation signal of each link is given by

$$A_n = -\frac{1}{L}(P_n - \max_l \{P_l\}) \quad (\text{A.4})$$



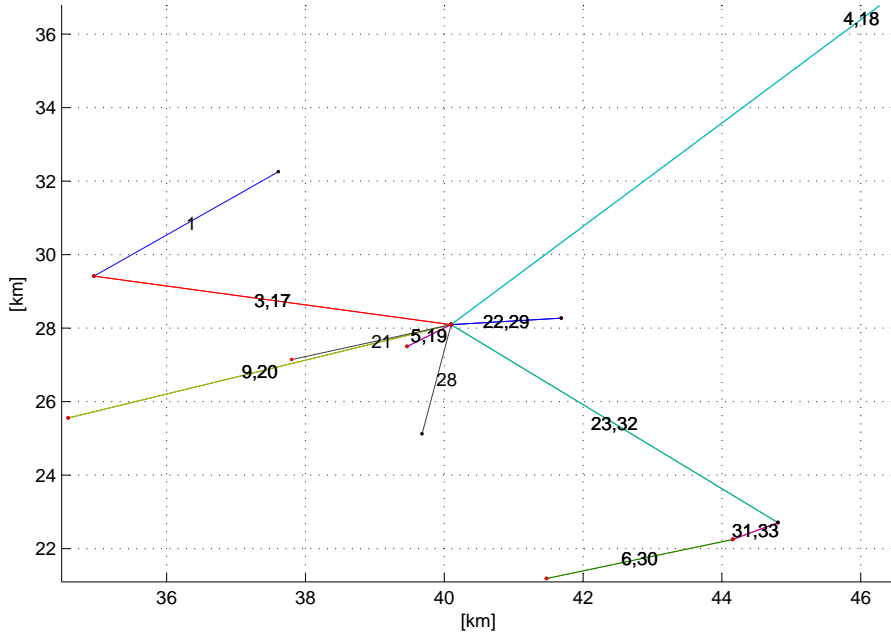


Figure A.1: Map of few of the links in the network.

where  $P_n[\text{dBm}]$  is the recorded RSL of the link at time  $n$ , and  $L[\text{km}]$  is the length of the link.

The data used is from several links in the Ramle-Lod area. Figure A.1 illustrates the relative locations of the links analyzed. The records are in 1min temporal resolution, of a 16 hour long rain event took place on December 26th 2006.

Figure A.2 shows the results obtained by using SOBI (Second Order Blind Identification) algorithm [6] on two opposite links. Figures A.3-A.5 shows further results obtained by applying the JADEop [10] and SOBI algorithms on different sets of links. Looking at the rain measured by a nearby rain gauge (Fig. 4.4), one can see that in each of the figures A.2-A.5, one of the estimated sources captures the behaviour of the rainfall experienced by the link. While the results indicate on the promising potential of BSS to filter out the rain-induced attenuation, further research is still necessary to identify other sources as natural phenomena (humidity, fog, scintillation effects, measurement noises, etc) and to deal with practical problems, as the quantization of the RSL measurements, and scaling.

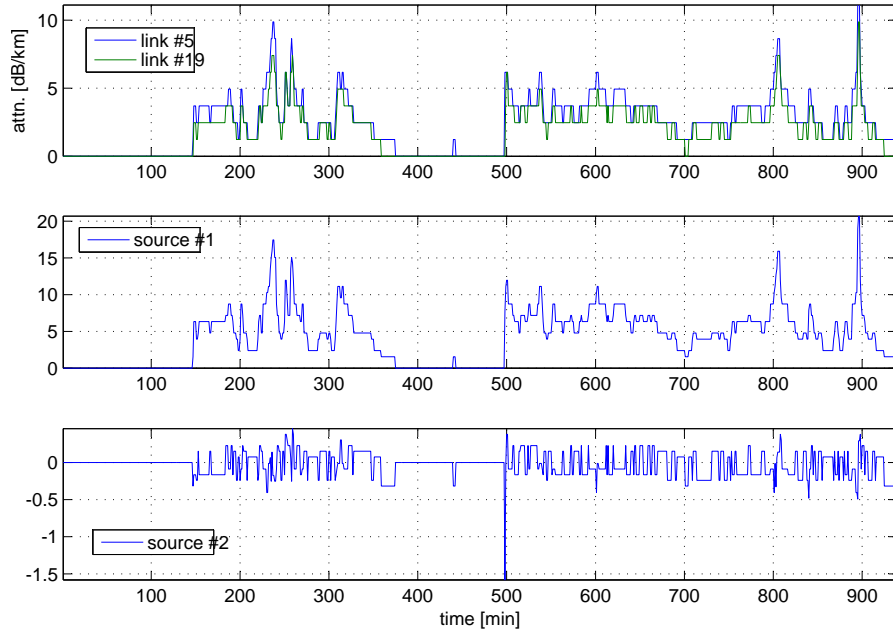


Figure A.2: Two input links. SOBI algorithm.

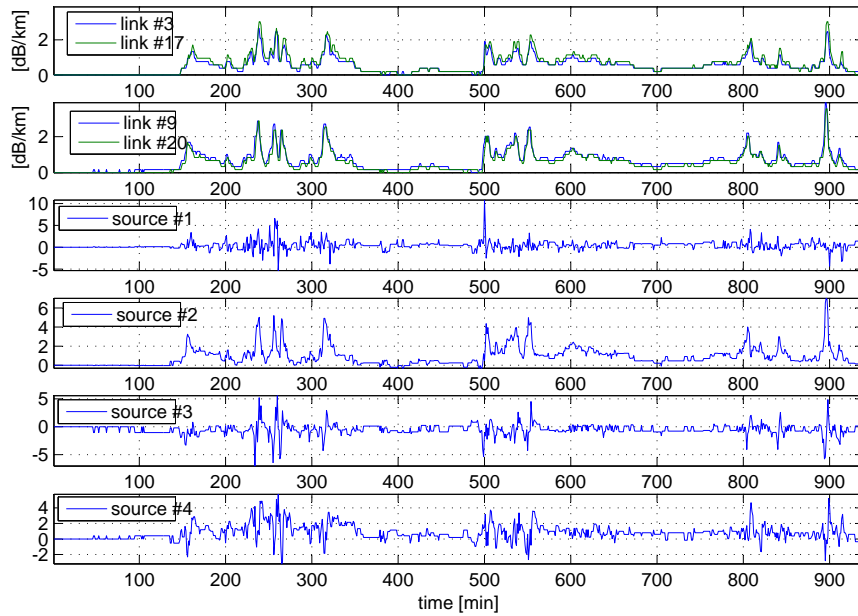


Figure A.3: Four input links. JADE-op algorithm.

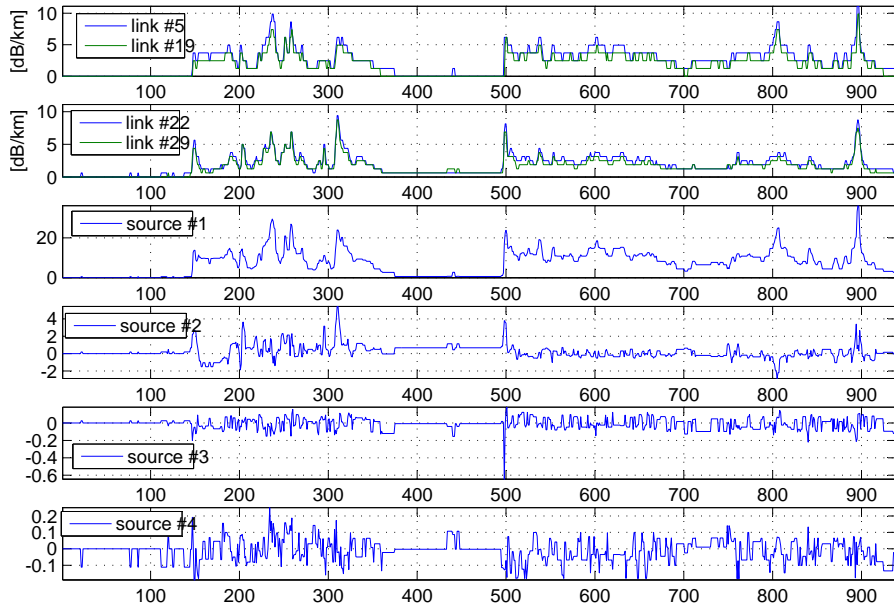


Figure A.4: Four input links. SOBI algorithm.

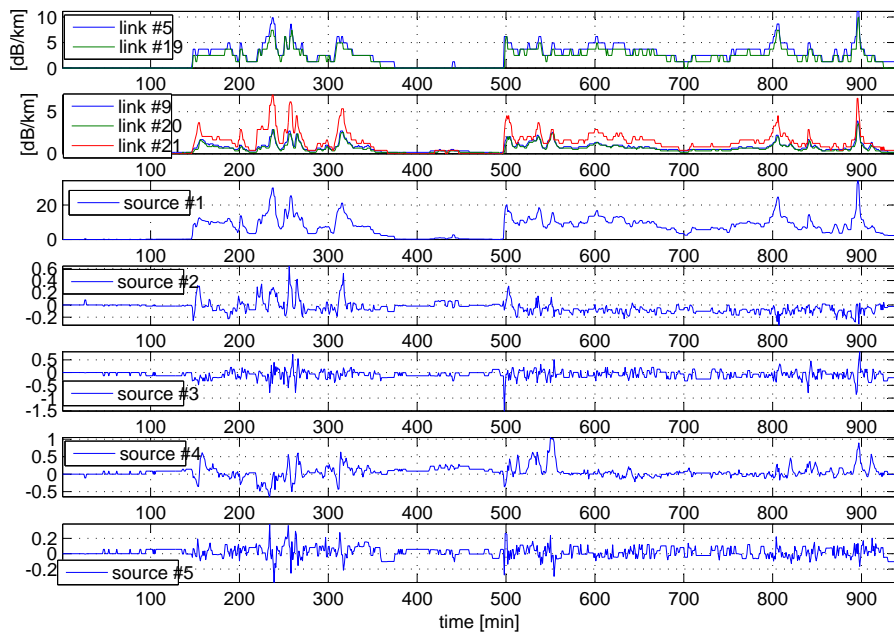


Figure A.5: Five input links. SOBI algorithm.



# Bibliography

- [1] H. Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control 19 (6), 716-723, 1974.
- [2] E. N. Anagnostou, Witold F. Krajewski, James Smith, *Uncertainty Quantification of Mean-Areal Radar-Rainfall Estimates*, Journal of Atmospheric and Oceanic Technology: Vol. 16, No. 2, pp. 206-215.
- [3] L. R. Bahl, F. Jelinek, and R. L. Mercer, *A maximum likelihood approach to continuous speech recognition*, IEEE Trans. Pattern Anal. Machine Intel., vol. PAMI-5, pp. 179-190, Mar. 1983.
- [4] F. G. Ball and J. A. Rice, *Stochastic models for ion channels: Introduction and bibliography*, Math. Biosci., vol. 112, pp. 189-206, 1992.
- [5] L. E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, Ann. Math. Stat., vol. 37, pp. 1554-1563, 1966.
- [6] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso and E. Moulines, *A blind source separation technique using second order statistics*, IEEE Trans. on Signal Processing, vol. 45, No. 2, pp. 434-444, February 1997.
- [7] J. Besag, *On the statistical analysis of dirty pictures*, J. Roy. Statist. Soc. B, vol. 48, no. 3, pp. 259-302, 1986.
- [8] J. Bilmes, *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Technical Report 97-021, Intl. Computer Science Institute, Univ. of California, Berkeley, Apr. 1998.
- [9] K. Burnham and D. Anderson, *Model Selection and Inference - A practical information-theoretic approach*, 2nd ed., Springer, 2000.

- [10] J.F. Cardoso and A. Souloumiac, *Jacobi angles for simultaneous diagonalization*, In *SIAM Journal of Matrix Analysis and Applications*, vole. 17, No 1, pp. 161-164, Jan. 1996.
- [11] J.F. Cardoso, *Blind signal separation: statistical principles*, PROCEEDINGS OF THE IEEE, VOL. 9, NO 10, PP. 2009-2025, OCT. 1998.
- [12] W. K. Ching, E. S. Fung and M. K. Ng, *Higher-Order Hidden Markov Models with Applications to DNA Sequences*, Lecture Notes in Computer Science, ISSU 2690, 2003.
- [13] A. Cichocki, S. Amari, K. Siwek, T. Tanaka , Anh Huy Phan et al., IICALAB Toolboxes, <http://www.bsp.brain.riken.jp/IICALAB>.
- [14] F. Cuccoli, L. Facheris, D. Giuli and L. Meoni, *Spatial rainfall rate estimation through combined use of radar reflectivity and raingauge data*, IEEE 2004, pp. 440-443.
- [15] J. A. du Preez, *Algorithms for high order hidden markov modelling*, In Proc. of the S. African Symposium on Comm. and Sig. Proc., September 1997. pp. 101106.
- [16] T.H. Fan and C.A. Tsai, *A Bayesian Method In Determining the Order of a Finite State Markov Chain*, Communications In Statistics-Theory And Methods, Volume: 28, Issue: 7, Pages: 1711-1730, 1999.
- [17] K. R. Gabriel and J. Neumann, *A Markov chain model for daily rainfall occurrence at Tel Aviv*, Quarterly Journal of the Royal Meteorological Society, vol. 88, issue 375, pp. 90-95, 1962
- [18] P. Gates and H. Tong, *On Markov Chain Modeling to Some Weather Data*, Journal of Applied Meteorology, Volume 15, pp. 1145-1151, 1976
- [19] D. Giuli, A. Toccafondi, G. Biffi Gentili, and A. Freni, *Tomographic reconstruction of rainfall fields through microwave attenuation measurements*, J. Appl. Meteorology, vol 30, no. 9, pp. 1323-1340, 1991.
- [20] O. Goldshtein, *Estimating space-time rainfall intensity using measurements from existing network of fixed terrestrial microwave links*, M.Sc. Thesis, Tel Aviv University, 2006.
- [21] U. Hadar and H. Messer, *High-Order Hidden Markov Models - Implementation, Estimation and Order Selection*. Submitted to IEEE Transactions on Signal Processing.

- [22] M. Hata, S. Doi, *Propagation Tests for 23 GHz and 40GHz*, IEEE Journal on selected areas in communications, Vol. SAC-1. No. 4, 1983 pp. 658-673.
- [23] L. M. Lee and J. C. LEE, *A Study on High-Order Hidden Markov Models and Applications to Speech Recognition*, Lecture Notes in Computer Science, NUMB 4031, 2006.
- [24] J. F. Mari, D. Fohr and J. C. Junqua, *A second-order HMM for high performance word and phoneme-based continuous speech recognition*, In Proc. ICASSP. Atlanta GA, May 1996.
- [25] N. Merhav, M. Gutman and J. Ziv, *On The Estimation of the Order of a Markov-Chain and Universal Data-Compression*, IEEE Transactions On Information Theory, Volume: 35, Issue: 5, Pages: 1014-1019, Sep. 1989.
- [26] H. Messer, A. Zinevich and P. Alpert, *Environmental monitoring by microwave communication networks*, Science, 312, 2006.
- [27] G. Morvai and B. Weiss, *Order estimation of Markov chains*, IEEE Transactions On Information Theory, Volume: 51, Issue: 4, Pages: 1496-1497, Apr. 2005.
- [28] R. L. Olsen, David V. Rogers, Daniel B. Hodge, *The  $aR^b$  Relation of Rain Attenuation*, IEEE Trans. On Antennas and propagation, Vol. AP-26, No. 2, March 1978. pp. 318-328
- [29] L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the IEEE, Vol. 77, No. 2, February 1989.
- [30] A. Rayitsfeld, A. Zinevich, U. Hadar and P. Alpert, *The first two-year results of rainfall monitoring using a commercial microwave communication system*, Submitted to the Bulletin of the American Meteorological Society.
- [31] A.R. Rahimi, A.R. Holt, G.J.G. Upton and R.J. Cummings, *The Use of Dual-Frequency Microwave Links for Measuring Path-Averaged Rainfall*, Department of Mathematics, University of Essex
- [32] T. Ryden, T. Terasvirta, and S. Asbrink, *Stylized facts of daily returns series and the hidden Markov model*, J. Appl. Econ., vol. 13, pp. 217-244, 1998.

- [33] G. Schwarz, *Estimating the Dimensions of a Model*, Annals of Statistics 6, 461-464, 1978.
- [34] G.A.F. Seber, *Multivariate Observations*, Wiley, New York, 1984.
- [35] H. Tong, *Determination of the Order of a Markov chain by Akaike's Information Criterion*, J. Appl. Prob., 12, 488-497, 1975