Symbolic Parameter Estimation of Continuous-Time Markov Chains

Daniel Runge Petersen
dept. of Computer Science
AAU
Aalborg, Denmark
dpet20@student.aau.dk

Lars Emanuel Hansen
dept. of Computer Science
AAU
Aalborg, Denmark
leha20@student.aau.dk

Sebastian Aaholm
dept. of Computer Science
AAU
Aalborg, Denmark
saahol20@student.aau.dk

***Abstract*—This is a placeholder abstract. The whole template is used in semester projects at Aalborg University (AAU).**

I. Introduction

This paper is about improving the runtime of jajapy - a tool for estimating parameters in parametric models.

Markov Chain (MC) - A chain of events described as a sequence of events without knowledge of prior. Hidden Markov Model (HMM) - A markov chain with emission probabilities. Markov Decision Process (MDP) - A markov chain with actions that influence the transitions. Continuous Time Markov Chain - A markov chain with traces that have dwell times as well as label emissions. Baum-Welch algorithm (BW) - Expectation-Maximization algorithm for finding the parameters of a Hidden Markov Model. Algebraic Decision Diagram (ADD) - A data structure of states and binary decisions, also called a Multi-Terminal Binary Decision Diagram (MTBDD).

II. Definitions

Definition 1 (Markov Chain)**.** *A Markov chain is a tuple $\mathcal{M} = (S, \mathcal{L}, \updownarrow, \tau, \pi)$, where:*

- *S is a finite set of states.*
- *$\mathcal{L}$ is a finite set of labels.*
- *$\updownarrow : S \to \mathcal{L}$ is a labeling function, which assigns a label to each state.*
- *$\tau : S \to \mathcal{D}(S)$ is a transition function. The model moves from state s to state s' with probability $\tau(s, s')$.*
- *π: is the initial distribution, the model starts in state s with probability $\pi(s)$.*

Intuitively, a Markov chain is a model that starts in a state s with probability $\pi(s)$, and then transitions to a new state s' with probability $\tau(s, s')$. The model continues to transition between states according to the transition function.

Definition 2 (Hidden Markov Model)**.** *A Hidden Markov Model (HMM) is a tuple $\mathcal{M} = (S, \mathcal{L}, \updownarrow, \tau, \pi)$, where $S, \mathcal{L}, \tau, \pi$ are defined as above, and:*

- *$\updownarrow : S \to D(\mathcal{L})$ is the emission function. The model emits a label l in state s with probability $\updownarrow(s, l)$.*

Intuitively, an HMM is a model that starts in a state s with probability $\pi(s)$, then emits a label l with probability $\updownarrow(s, l)$, and transitions to a new state s' with probability $\tau(s, s')$. The model continues to emit labels and transition between states according to the emission and transition functions.

Definition 3 (Markov Decision Process)**.** *A Markov Decision Process (MDP) is a tuple $\mathcal{M} = (S, \mathcal{L}, \updownarrow, A, \{\tau_a\}_{a \in A}, \pi)$ where $S, \mathcal{L}, \updownarrow, \pi$ are defined as above, and:*

- *A is a finite nonempty set of actions.*
- *$\tau_a : S \to \mathcal{D}(S)$ is a transition function for each action $a \in A$. The model moves from state s to state s' with probability $\tau_a(s, s')$ when action a is taken.*

Intuitively, an MDP is a model that starts in a state s with probability $\pi(s)$, then emits a label $\updownarrow(s)$ and, it can recieve an action $a \in A$ and transition to a new state s' with probability $\tau_a(s, s')$.

A. Continuous-Time

In the previous definitions, the models are discrete-time models, where time advances in fixed, regular steps. For example, in a discrete-time Markov chain, the system transitions between states at each step or tick of a clock, and the probability of moving from one state to another is governed by the transition function $\tau(s, s')$. This means that transitions can only happen at specific time intervals (e.g., after every second, every minute, etc.).

In contrast, continuous-time models allow transitions to occur at any time, rather than at fixed intervals. The time between transitions is variable and follows a continuous distribution. This introduces the concept of transition rates rather than discrete transition probabilities.

Definition 4 (Continuous-Time Markov Chain)**.** *A Continuous-Time Markov Chain (CTMC) is a tuple $\mathcal{M} = (S, \mathcal{L}, \updownarrow, R, \pi)$, where $S, \mathcal{L}, \updownarrow, \pi$ are defined as above, and:*

- *$R : S \times S \to \mathbb{R}_{\geq 0}$ is the rate function. The model transitions from state s to state s' with rate $R(s, s')$.*

For two states s and s', $R(s, s')$ gives the rate at which the system moves from state s to state s'. A higher rate means a faster transition.

A Continuous-Time Markov Chain (CTMC) is a type of Markov model where the time between transitions is not fixed but is governed by exponential distributions. If there are more then one outgoing transition from a state, we get race-conditions, the first transition to occur is the one that will be taken. The time spent in a state before transitioning to a new state is called $dwell - time$. This is exponentially distributed with a rate $E(s) = \sum_{s' \in S} R(s, s')$. The probability of transitioning from state s to state s' is $R(s, s')/E(s)$, the time spent in s is independent from the properbility of transitioning to s'.

B. Matrix Representation

The transition function τ can be represented as a matrix, where each element $\tau(s, s')$ is the probability of transitioning from state s to state s'. The matrix representation of τ is called the transition matrix. The transition matrix is a square matrix with dimensions $|S| \times |S|$, where $|S|$ is the number of states in the model. The transition matrix is a stochastic matrix, meaning that the sum of each row is equal to 1, meaning all the probabilities of transitioning from state s to all other states sum to 1.

If we take an example of a model with two states $S = \{s_1, s_2\}$, the transition matrix τ is defined as:

$$\tau = \begin{bmatrix} \tau(s_1, s_1) & \tau(s_1, s_2) \\ \tau(s_2, s_1) & \tau(s_2, s_2) \end{bmatrix} \tag{1}$$

We can give an example of a transition matrix for a model with two states, where the model transitions from state s_1 to state s_2 with probability 0.4 and transitions from state s_2 to state s_1 with probability 0.5:

$$\tau = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \tag{2}$$

The initial distribution π is a vector that represents the probability of starting in each state. The initial distribution is a stochastic vector, meaning that the sum of all probabilities is equal to 1. The initial distribution π is a vector with dimensions $|S|$, where $|S|$ is the number of states in the model. Each element $\pi(s)$ is the probability of starting in state s.

$$\pi = \begin{bmatrix} 0.6 \\ 0.5 \end{bmatrix} \tag{3}$$

The labeling function $\updownarrow$ can be represented as a matrix, where each element $\updownarrow(s, l)$ is the probability of emitting label l in state s. The matrix representation of $\updownarrow$ is called the emission matrix. The emission matrix is a matrix with dimensions $|S| \times |\mathcal{L}|$, where $|\mathcal{L}|$ is the number of labels in the model. The emission matrix is a stochastic matrix, meaning that the sum of each row is equal to 1, meaning all the probabilities of emitting a label in state s sum to 1.

If we take an example of a model with two states $S = \{s_1, s_2\}$ and two labels $\mathcal{L} = \{l_1, l_2\}$, the emission matrix $\updownarrow$ is defined as:

$$\updownarrow = \begin{bmatrix} \updownarrow(s_1, l_1) & \updownarrow(s_1, l_2) \\ \updownarrow(s_2, l_1) & \updownarrow(s_2, l_2) \end{bmatrix} \tag{4}$$

We can give an example of an emission matrix for a model with two states and two labels, where the model emits label l_1 in state s_1 with probability 0.7 and emits label l_2 in state s_2 with probability 0.6:

$$\updownarrow = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \tag{5}$$

The rate function R can be represented as a matrix, where each element $R(s, s')$ is the rate of transitioning from state s to state s'. The matrix representation of R is called the rate matrix. The rate matrix is a square matrix with dimensions $|S| \times |S|$, where $|S|$ is the number of states in the model. The rate matrix is a non-negative matrix, meaning that all elements are greater than or equal to 0.

$$R = \begin{bmatrix} R(s_1, s_1) & R(s_1, s_2) \\ R(s_2, s_1) & R(s_2, s_2) \end{bmatrix} \tag{6}$$

If we take an example of a model with two states $S = \{s_1, s_2\}$, the rate matrix R is defined as:

$$R = \begin{bmatrix} 0.5 & 0.3 \\ 0.2 & 0.4 \end{bmatrix} \tag{7}$$

III. HMM Example

A. Setup

We have a simple HMM with, two hidden states S_1 and S_2, two observation symbols: O_1 and O_2 and an observation sequence $O = \{O_1, O_2, O_1\}$.

The HMM parameters are:

Transition matrix A (probability of moving from one state to another):

$$A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

Emission matrix B (probability of emitting observation given a state):

$$B = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Initial state probability vector π (probability of starting in each state):

$$\pi = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix}$$

B. Expectation step

In the expectation step we calculate α and β.

1) Forward step α: We first compute the forward probabilities $\alpha_t(i)$, which represent the probability of being in state i at time t after observing the first t symbols.

a) Initialization at $(t = 1)$:

$$\alpha_1 = \pi \circ B_{y1}$$

Where B_{y1} is the first column of the emission matrix, corresponding to observation O_1

(i.e., $B_{y1} = \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix}$) and $\circ$ represents the Hadamard product. So, we get:

$$\alpha_1 = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \circ \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.56 \\ 0.08 \end{bmatrix}$$

b) Induction (for $t = 2, 3, \dots, T$): For subsequent timesteps, we compute:

$$\alpha_{t+1} = B_{yt+1} \circ (A^T \alpha_t)$$

Where A^T is the transpose of the transition matrix. Let's apply this to compute the forward probabilities for $t = 2$ and $t = 3$:

At $t = 2$ (observation O_2):

$$\alpha_2 = B_{y2} \circ (A^T \alpha_1)$$

We have:

$$B(y2) = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}$$

and

$$A^T = \begin{bmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{bmatrix}$$

We get:

$$\alpha_2 = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix} \circ \left(\begin{bmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.56 \\ 0.08 \end{bmatrix} \right) = \begin{bmatrix} 0.1128 \\ 0.1584 \end{bmatrix}$$

At $t = 3$ (observation O_1):

$$\alpha_2 = B_{y1} \circ (A^T \alpha_2)$$

We get:

$$\alpha_3 = \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix} \circ \left(\begin{bmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.1 \\ 0.1584 \end{bmatrix} \right) = \begin{bmatrix} 0.102816 \\ 0.049728 \end{bmatrix}$$

2) Backward step β: The backward probabilities $\beta_t(i)$ represent the probability of observing the rest of the sequence starting from time $t+1$, given that the system is in state i at time t.

Initialization (at $t = T = 3$)

$$\beta_T = \mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

a) Induction (for $t = T-1, T-2, \dots, 1$): For earlier timesteps, we compute:

$$\beta_t = A(\beta_{t+1} \circ B_{yt+1})$$

At $t = 2$ (observation O_1):

$$\beta_2 = A(\beta_3 \circ B_{y1})$$

$$B_{y1} = \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix}, \quad \beta_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

We get:

$$\beta_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \cdot \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix} \right)$$

$$\beta_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.58 \\ 0.55 \end{bmatrix}$$

At $t = 1$ (observation O_2):

$$\beta_1 = A(\beta_2 \circ B_{y2})$$

We have:

$$B_{y2} = \begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix}, \quad \beta_2 = \begin{bmatrix} 0.58 \\ 0.55 \end{bmatrix}$$

$$\beta_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \cdot \left(\begin{bmatrix} 0.3 \\ 0.6 \end{bmatrix} \circ \begin{bmatrix} 0.58 \\ 0.55 \end{bmatrix} \right)$$

$$\beta_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 0.174 \\ 0.33 \end{bmatrix} = \begin{bmatrix} 0.2364 \\ 0.252 \end{bmatrix}$$

C. Step 3: Compute γ and ξ

1) Compute γ: We can compute γ by

$$\gamma_t = (\mathbb{1}^T \bullet \alpha_T)^{-1} \bullet (\alpha_t \circ \beta_t)$$

$$\alpha_T = \begin{bmatrix} 0.089628 \\ 0.053328 \end{bmatrix}$$

$$\mathbb{1}^T \bullet \alpha_T = 0.089628 + 0.053328 = 0.152544$$

This is the total probability of oberserving our sequence $O = \{O_1, O_2, O_1\}$

Now we can compute γ_t for each time stamp.

At t=1: We have

$$\alpha_1 = \begin{bmatrix} 0.56 \\ 0.08 \end{bmatrix}, \quad \beta_1 = \begin{bmatrix} 0.2364 \\ 0.252 \end{bmatrix}$$

We take the Hadamard product of this.

$$\alpha_1 \circ \beta_1 = \begin{bmatrix} 0.56 \cdot 0.2364 \\ 0.08 \cdot 0.252 \end{bmatrix} = \begin{bmatrix} 0.132384 \\ 0.02016 \end{bmatrix}$$

We normalize the first part and take the scalar product.

$$\gamma_1 = \frac{1}{0.152544} \bullet \begin{bmatrix} 0.132384 \\ 0.02016 \end{bmatrix} = \begin{bmatrix} 0.8678414 \\ 0.1321589 \end{bmatrix}$$

At t = 2:

We have:

$$\alpha_2 = \begin{bmatrix} 0.1074 \\ 0.1584 \end{bmatrix}, \quad \beta_2 = \begin{bmatrix} 0.58 \\ 0.55 \end{bmatrix}$$

The Hadamard product is:

$$\alpha_2 \circ \beta_2 = \begin{bmatrix} 0.1074 \cdot 0.58 \\ 0.1584 \cdot 0.55 \end{bmatrix} = \begin{bmatrix} 0.062292 \\ 0.08712 \end{bmatrix}$$

We normalize the first part and take the scalar product.

$$\gamma_2 = \frac{1}{0.152544} \bullet \begin{bmatrix} 0.062292 \\ 0.08712 \end{bmatrix} = \begin{bmatrix} 0.42888609 \\ 0.57111391 \end{bmatrix}$$

At t = 3

We have:

$$\alpha_3 = \begin{bmatrix}0.089628\\0.053328\end{bmatrix}, \quad \beta_3 = \begin{bmatrix}1\\1\end{bmatrix}$$

The Hadamard product is:

$$\alpha_3 \circ \beta_3 = \begin{bmatrix}0.089628\\0.053328\end{bmatrix}$$

We normalize the first part and take the scalar product.

$$\gamma_3 = \frac{1}{0.152544} \bullet \begin{bmatrix}0.089628\\0.053328\end{bmatrix} = \begin{bmatrix}0.67400881\\0.32599119\end{bmatrix}$$

2) Calculating ξ: We calculate ξ by

$$\xi_t = ((\mathbb{1}^T \alpha_T)^{-1} \bullet A) \circ (\alpha_t \otimes (\beta_{t+1} \circ B_{yt+1})^T)$$

We start by calculating $((\mathbb{1}^T \alpha_T)^{-1} \bullet A)$: From before, we have

$$(\mathbb{1}^T \alpha_T)^{-1} = \frac{1}{0.152544} = 6.996$$

We have:

$$A = \begin{bmatrix}0.6 & 0.4\\0.5 & 0.5\end{bmatrix}$$

We get:

$$8.996 \bullet A = \begin{bmatrix}6.996 \cdot 0.6 & 6.996 \cdot 0.4\\6.996 \cdot 0.5 & 6.996 \cdot 0.5\end{bmatrix} = \begin{bmatrix}4.198 & 2.798\\3.498 & 3.498\end{bmatrix}$$

We can now calculate $\alpha_1 \otimes (\beta_2 \circ B_{y2})^T$. We have :

$$\alpha_1 = \begin{bmatrix}0.56\\0.08\end{bmatrix}, \quad \beta_2 = \begin{bmatrix}0.58\\0.55\end{bmatrix}, \quad B_{y2} = \begin{bmatrix}0.3\\0.6\end{bmatrix}$$

We calculate $\beta_2 \circ B_{y2}$:

$$\beta_2 \circ B_{y2} = \begin{bmatrix}0.58\\0.55\end{bmatrix} \circ \begin{bmatrix}0.3\\0.6\end{bmatrix} = \begin{bmatrix}0.174\\0.33\end{bmatrix}$$

Outer product:

$$\alpha_1 \otimes (\beta_2 \circ B_{y2})^T = \begin{bmatrix}0.56\\0.08\end{bmatrix} \otimes \begin{bmatrix}0.174 & 0.33\end{bmatrix}$$

$$= \begin{bmatrix}0.09744 & 0.1848\\0.01392 & 0.0264\end{bmatrix}$$

We can now calculate ξ_1

$$\xi_1 = \begin{bmatrix}4.198 & 2.798\\3.498 & 3.498\end{bmatrix} \circ \begin{bmatrix}0.09744 & 0.1848\\0.01392 & 0.0264\end{bmatrix}$$

$$\xi_1 = \begin{bmatrix}0.38325991 & 0.03650094\\0.60572687 & 0.08653241\end{bmatrix}$$

At t=2:
We have:

$$B_{y1} = \begin{bmatrix}0.7\\0.4\end{bmatrix}, \quad \alpha_2 = \begin{bmatrix}0.1074\\0.1584\end{bmatrix}, \quad \beta_3 = \begin{bmatrix}1\\1\end{bmatrix}$$

Hadamard product for $\beta_3 \circ B_{y1}$

$$\beta_3 \circ B_{y1} = \begin{bmatrix}1\\1\end{bmatrix} \circ \begin{bmatrix}0.7\\0.4\end{bmatrix} = \begin{bmatrix}0.7\\0.4\end{bmatrix}$$

Outer product:

$$\alpha_2 \otimes \begin{bmatrix}0.7 & 0.4\end{bmatrix} = \begin{bmatrix}0.07518 & 0.04296\\0.11088 & 0.06336\end{bmatrix}$$

We can now calculate ξ_2:

$$\xi_2 = \begin{bmatrix}4.198 & 2.798\\3.498 & 3.498\end{bmatrix} \circ \begin{bmatrix}0.07518 & 0.04296\\0.11088 & 0.06336\end{bmatrix}$$

$$\xi_2 = \begin{bmatrix}0.07341938 & 0.06873304\\0.03726872 & 0.0523348\end{bmatrix}$$

At t=3:
We have:

$$B_{y1} = \begin{bmatrix}0.7\\0.4\end{bmatrix}, \quad \alpha_3 = \begin{bmatrix}0.089628\\0.053328\end{bmatrix}, \quad \beta_3 = \begin{bmatrix}1\\1\end{bmatrix}$$

Hadamard product for $\beta_3 \circ B_{y1}$

$$\beta_3 \circ B_{y1} = \begin{bmatrix}1\\1\end{bmatrix} \circ \begin{bmatrix}0.7\\0.4\end{bmatrix} = \begin{bmatrix}0.7\\0.4\end{bmatrix}$$

Outer product:

$$\alpha_3 \otimes \begin{bmatrix}0.7 & 0.4\end{bmatrix} = \begin{bmatrix}0.062740 & 0.035852\\0.037329 & 0.021331\end{bmatrix}$$

We can now calculate ξ_3:

$$\xi_2 = \begin{bmatrix}4.198 & 2.798\\3.498 & 3.498\end{bmatrix} \circ \begin{bmatrix}0.062740 & 0.035852\\0.037329 & 0.021331\end{bmatrix}$$

$$\xi_3 = \begin{bmatrix}0.2839837 & 0.09127753\\0.13480176 & 0.06519824\end{bmatrix}$$

D. Update values

$$\hat{\pi} = \gamma_1 = \begin{bmatrix}0.86784141\\0.1321589\end{bmatrix}$$

$$\hat{A} = (\mathbb{1} \oslash \gamma) \bullet \xi$$

$$\hat{B} = (\mathbb{1} \oslash \gamma) \bullet (\sum_{t=1}^{T} \gamma_t \otimes \mathbb{1}_{yt}^T)$$

When referring to γ, we use the sum of the probabilities:

$$\gamma = \sum_{t=1}^{T} \gamma_t$$

and ξ:

$$\xi = \sum_{t=1}^{T} \xi_t$$

We therefore calculate:

$$\gamma = \begin{bmatrix}0.86784141\\0.1321589\end{bmatrix} + \begin{bmatrix}0.42888609\\0.57111391\end{bmatrix} + \begin{bmatrix}0.67400881\\0.32599119\end{bmatrix} = \begin{bmatrix}1.97073631\\1.02926369\end{bmatrix}$$

And

$$\xi = \begin{bmatrix}0.38325991 & 0.03650094\\0.60572687 & 0.08653241\end{bmatrix} + \begin{bmatrix}0.07341938 & 0.06873304\\0.03726872 & 0.0523348\end{bmatrix}$$

$$+\begin{bmatrix}0.2830837 & 0.09127753\\0.13480176 & 0.06519824\end{bmatrix} = \begin{bmatrix}0.739763 & 0.19651152\\0.77779736 & 0.20406545\end{bmatrix}$$

We can now calculate

$$\mathbb{1} \oslash \gamma = \begin{bmatrix}\frac{1}{2.0923}\\ \frac{1}{1.1352}\end{bmatrix}$$

We can now calculate $\hat{A}$

$$\hat{A} = \begin{bmatrix}\frac{1}{2.0923}\\ \frac{1}{1.1352}\end{bmatrix} \bullet \begin{bmatrix}0.9897 & 0.7370\\0.5670 & 0.3888\end{bmatrix} = \begin{bmatrix}0.37537391 & 0.19092437\\0.39467348 & 0.198226353\end{bmatrix}$$

We calculate $\hat{B}$ We first calculate the sum of the outer products:

$$\sum_{t=1}^{T} \gamma_t \otimes \mathbb{1}_{yt}^T$$

At t = 1:

$$\gamma_1 \otimes \begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}0.86784141\\0.1321589\end{bmatrix} \otimes \begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}0.86784141 & 0.13215859\\0 & 0\end{bmatrix}$$

At t = 2:

$$\gamma_2 \otimes \begin{bmatrix}0 & 1\end{bmatrix} = \begin{bmatrix}0.42888609\\0.57111391\end{bmatrix} \otimes \begin{bmatrix}0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0\\0.42888609 & 0.57111391\end{bmatrix}$$

At t = 3:

$$\gamma_3 \otimes \begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}0.67400881\\0.32599119\end{bmatrix} \otimes \begin{bmatrix}1 & 0\end{bmatrix} = \begin{bmatrix}0.67400881 & 0.32599119\\0 & 0\end{bmatrix}$$

We summearize these to get:

$$\begin{bmatrix}0.86784141 & 0.13215859\\0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\0.42888609 & 0.57111391\end{bmatrix} +$$

$$\begin{bmatrix}0.67400881 & 0.32599119\\0 & 0\end{bmatrix} = \begin{bmatrix}1.54185022 & 0.45814978\\0.42888609 & 0.57111391\end{bmatrix}$$

$$\hat{b} = \begin{bmatrix}\frac{1}{2.0923}\\ \frac{1}{1.1352}\end{bmatrix} \bullet \begin{bmatrix}1.54185022 & 0.45814978\\0.42888609 & 0.57111391\end{bmatrix}$$

$$= \begin{bmatrix}0.78237266 & 0.23247645\\0.41669214 & 0.55487618\end{bmatrix}$$

E. ADD representation

As we only need one bit to represent the the rows and columns with one bit, we only need one variable for the them, as x_1 is the variable for rows and y_1 is the variable for column.

We first make the matrices into ADD representation.

We can now use the ADD representation to calculate α and β.

When using ADD's it is important to remember, if we need to take a row from a matrix, we fix the input to the ADD by setting the x-variables to the desired row. An example is taking the third row of a matrix with 8 rows, we set, $x_1 = 1, x_2 = 1, x_3 = 0$ and $x_4 = 0$. if we need to take the second column, we set $y_1 = 1, y_2 = 0$ and $y_3 = 0, y_4 = 0$. Hadamard product is row-wise multiplication of the matrices. So to calculate the Hadamard product of two matrices, we set the x-variables to the same row in both matrices and multiply the corresponding nodes in the ADDs. To calculate a Hadamard product in ADD, we multiply the corresponding nodes in the ADDs, as shown in the following figure.

Matrix multiplication is done by fixing the input to the first matrix and the output to the second matrix. We then sum the

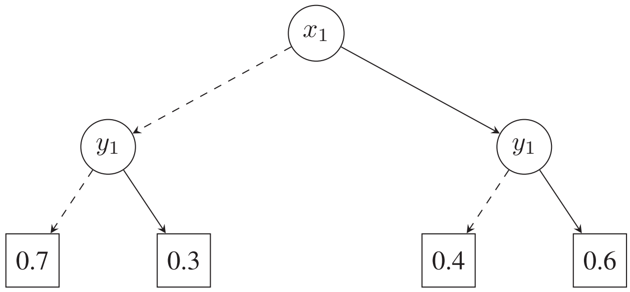

Fig. 1. B-matrix representation in ADD

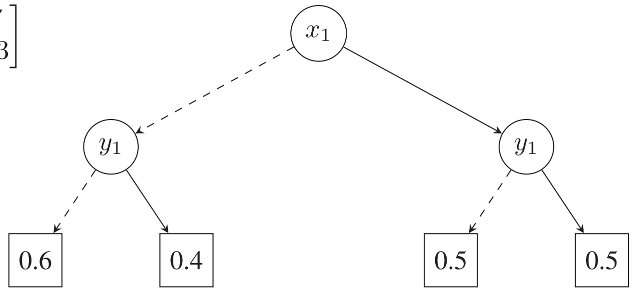

Fig. 2. A-matrix representation in ADD

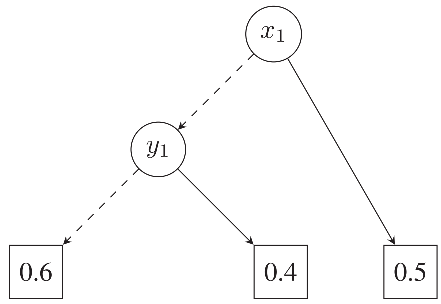

Fig. 3. A-matrix representation in ADD caption reduced

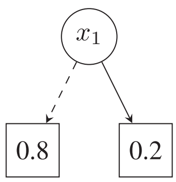

Fig. 4. π-matrix representation in ADD

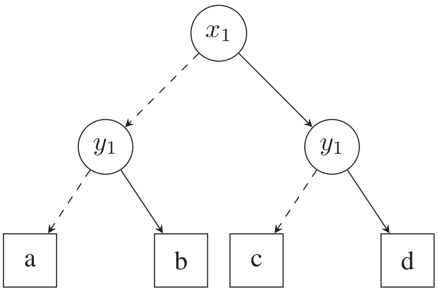

Fig. 5. Matrix A in ADD

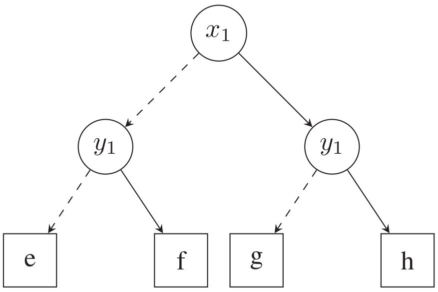

Fig. 6. Matrix B in ADD

result of the Hadamard product of the rows of the first matrix and the columns of the second matrix. This is shown in the following figure.

Acronyms

AAU Aalborg University. 1

Appendix A
Compiling in draft

You can also compile the document in draft mode. This shows todos, and increases the space between lines to make space for your supervisors feedback.

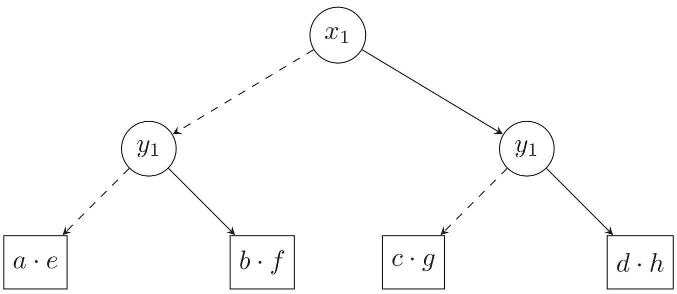

Fig. 7. Hadamard product of A and B in ADD

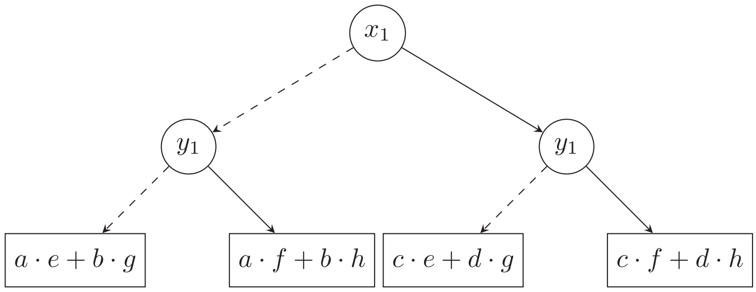

Fig. 8. Matrix multiplication of A and B in ADD

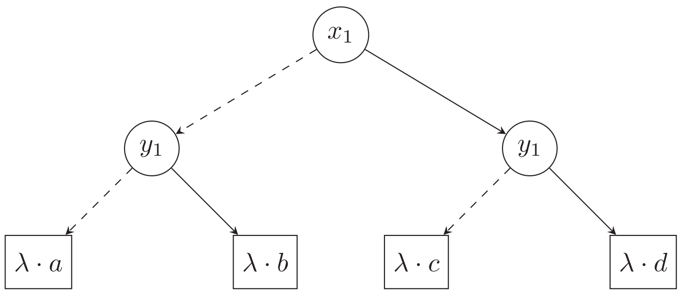

Fig. 9. Scalar product in ADD

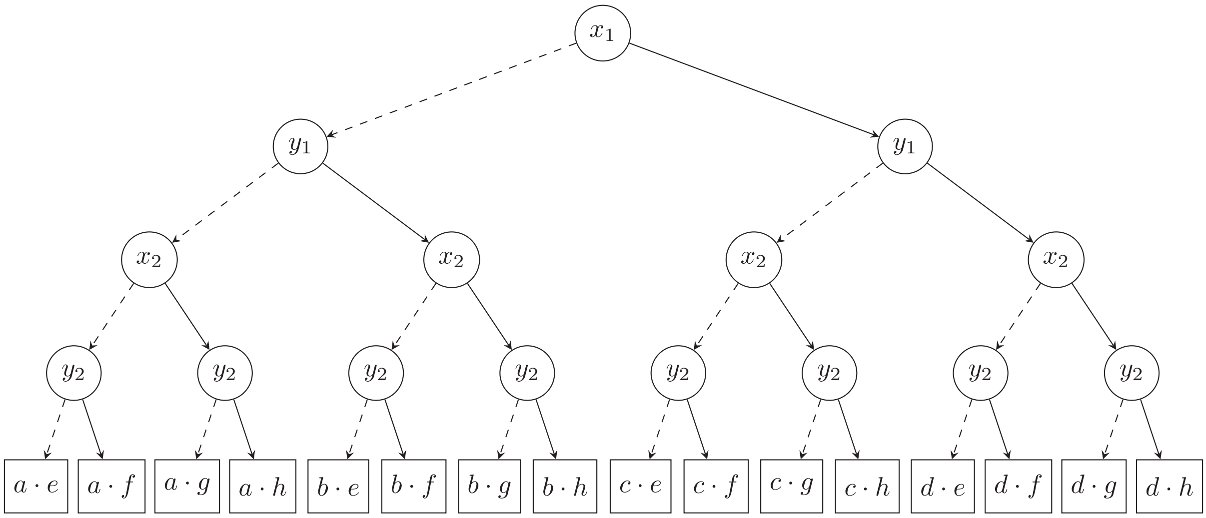

Fig. 10. Kroneker product in ADD

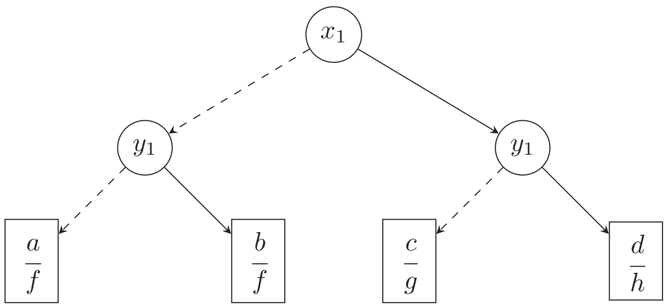

Fig. 11. Hadamard division of A and B in ADD

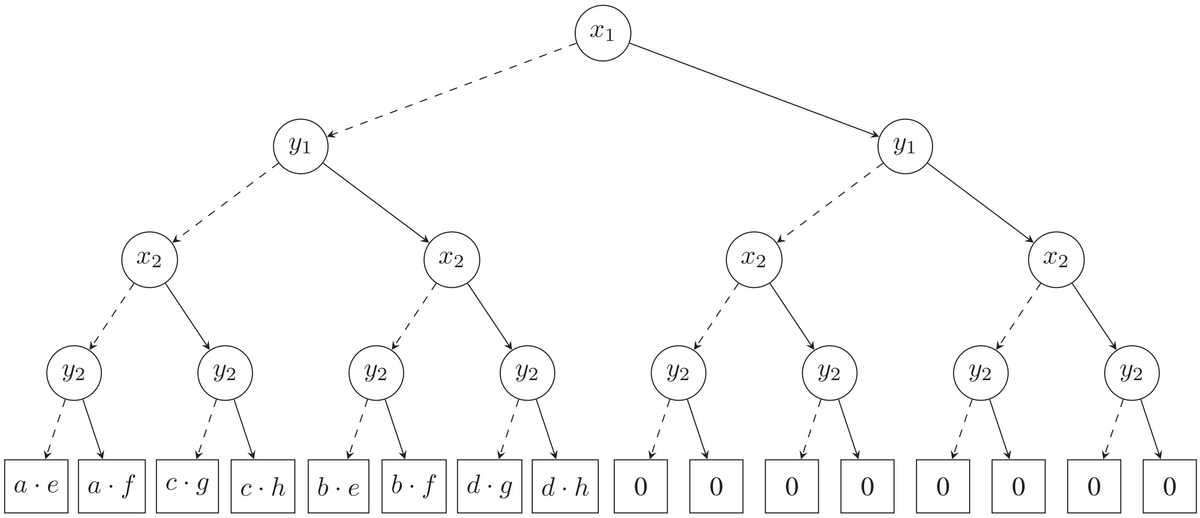

Fig. 12. Katri-Rao in ADD

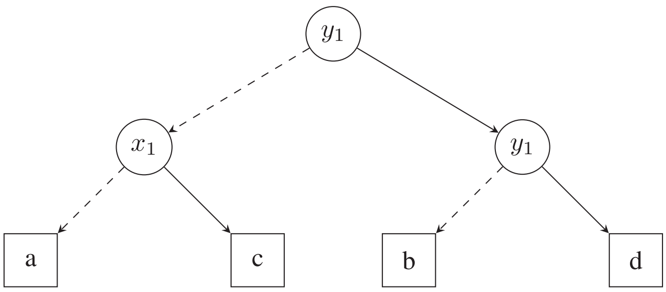

Fig. 13. transpose in ADD