UNIVERSITÉ CÔTE D'AZUR

INTERNSHIP REPORT

Bayesian Latent Variable Model for the Analysis of Multi-modal Longitudinal Biomedical Data

---

Author:

Bhargav Ramudu MANAM

Supervisors:

Dr. Marco LORENZI

Senior Research Scientist

Dr. Irene BALELLI

Research Scientist

An Internship Report submitted in fulfillment of the requirements for the degree of ***M.Sc. Data Science and Artificial Intelligence***

Internship carried out in the

EPIONE Team

of

Centre Inria d'Université Côte d'Azur

Manuscript submitted on July 31, 2023

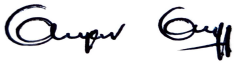
Presented and defended on August 18, 2023

Declaration of Authorship

I, Bhargav Ramudu MANAM, declare that this report titled, "Bayesian Latent Variable Model for the Analysis of Multi-modal Longitudinal Biomedical Data" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a masters' degree at this University.
- Where any part of this report has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the report is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signature:



Date: 31/07/2023

Abstract

This report provides a methodological view to the modeling of disease progression using multi-modal biomedical data, with an emphasis on neurodegenerative diseases (NDDs) such as Alzheimer's and Parkinson's diseases. Given the major impact of these diseases on global health, the report strives to present the literature on Disease Progression Modelling (DPM), in the context of multi-modal longitudinal biomedical data sets such as Alzheimer's Disease Neuroimaging Initiative (ADNI) and Parkinson's Progression Markers Initiative (PPMI).

The principal challenge for DPM lies in the intricate nature of biomedical data, burdened with issues like inter-patient and intra-patient variability, sparse and incomplete data, and high-dimensionality. These factors significantly complicate the effective modeling of disease progression.

To address these challenges, we propose a novel Bayesian Latent Variable Model, inspired by the Probabilistic Principal Component Analysis (PPCA) to accommodate both longitudinal and multi-modality data. This report delves into the theoretical underpinnings of this model, followed by its evaluation against a synthetic dataset. The model's strength lies in its capacity to tackle the complexities of biomedical data and its adaptability to future research directions.

Keywords: Disease Progression Modelling (DPM), Longitudinal Data, Multi-modality, Latent Variable Model, Biomedical Data, PPCA.

Acknowledgements

The successful completion of this internship or report could not have been possible without the support, guidance and encouragement of many individuals who have contributed in one way or another.

I would first like to express my profound gratitude to my internship supervisors, Prof. Marco Lorenzi and Dr. Irene Balelli, for giving me the opportunity to work with them. I thank you for being so warm, welcoming and it is a delight for me everyday to interact and learn from you. I appreciate the guidance and clarity you presented during the course of the internship, especially when I am confused or lost. I am indebted to you for your valuable feedback and constant encouragement throughout this journey. I thank you specifically, for giving me the freedom and time to learn, explore a range of topics. I highly regard and am thankful for the immense help and contribution in shaping the manuscript, without your inputs it would have been near impossible for me to finish it.

I would also like to extend my appreciation to the Epione research team at Inria, whose assistance and intellectual insights have been invaluable. The spirited discussions, shared ideas during the weekly journal club have broadened my understanding and perspective on various subjects. It is very nice to meet several amazing people and special mention to my office mates Lucia and Riccardo for always keeping a good company.

Most importantly, I express my utmost gratefulness to the entire M.Sc DSAI team (academic and staff), specially to Ms. Léa Hollander, for giving me the opportunity to participate in this program. Studying in France gave me a chance to explore the traditions and culture of the nation and they have clearly expanded my horizons and perspectives.

Last but not the least, I thank my parents and family members beyond words for the love, trust and belief in my abilities. I acknowledge the immeasurable personal contributions of all my wonderful friends, class mates, colleagues and especially of those I have met during this last one year.

Contents

List of Figures

List of Tables

To my parents

Chapter 1

Introduction

1.1 About the company

Inria[1], is a French national research institution for digital science and technology. Established in 1967, Inria has its main headquarter at Rocquencourt (Paris) and has 9 research centers distributed across France, as well as one center abroad in Santiago de Chile, Chile. The institute conducts both theoretical and applied research in computer science and has produced many widely used programs [3]. Inria is a leading institute for research in Artificial Intelligence (AI).

Inria research center of Université Côte d'Azur[2] (UCA) was established in 1983 and is located in Sophia Antipolis. It has 37 project teams employing around 700 scientists and research engineers. Its staff represent a rich and diverse cultural heritage with their nationalities spanning across 55 different countries. The theme of the center is interdisciplinary research with computational medicine, neuroscience, biology, AI, geometry, heterogeneous data modeling and collaborative robotics, as the key areas of focuses of the teams. The center also collaborates with reputed institutes such as CNRS, INRAE, Inserm and is a major player in driving the AI innovation in the PACA region through its extensive ecosystem and partnerships comprising of government entities and major companies like Airbus, Orange, Amadeus, Microsoft Research, etc.

[1] Inria: Institut national de recherche en sciences et technologies du numérique, https://www.inria.fr/fr

[2] Inria UCA, https://www.inria.fr/en/inria-centre-universite-cote-azur

The **Epione**[3] team of Inria UCA center, is a biomedical research group with the aim of making advancements for *e-medicine*[4] through the concept of the *e-patient*: a collection of computational models with the ability to simulate (at an individual or population level) the anatomy and physiology of the patient's tissues and organs, at various scales. The team's research revolves around these scientific axes: Biomedical Image Analysis, Imaging Genetics, Computational Anatomy, Computational Physiology, and Computational Cardiology. The organisation of the team's work is guided by a virtuous triangle of academic research, clinical partnerships and industrial partnerships. Where the research directions are chosen so as to solve the challenging issues raised by the clinical or industrial partners.

1.2 Context

Chronic illnesses such as diabetes, arthritis and Neurodegenerative Diseases (NDDs) are diseases that progress over time (often in stages). According to the World Health Organization (WHO) report [5], about 63% of all the deaths in the world can be attributed to chronic diseases. In most cases, the disease progression is often very slow and takes several years to evolve from mild to severe stage. In this report, I emphasize on NDDs such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) for which studies [6, 7] have shown that the incidence rates increase gradually from a young age until about 65 years of age (doubling every 5 years) before exponentially rising up. Therefore, it is crucial to develop models to help our understanding of the evolution of these illnesses (NDDs), which still lack effective treatments despite of decades of clinical trails [8].

The Critical Path Initiative [9], launched by the US Food and Drug Administration (FDA), identified and promoted Disease Progression Modeling (DPM) as a key area of focus [10, 11] when dealing with chronic illnesses so as to prevent, cure and modernize the drug development. DPM [12] consists of data-driven computational methods that processes longitudinal data of patients or subjects to learn the trajectory of the disease (Figure 1.1), where each biomarker is a indicator of one's biological state [13].

[3] Epione team of Inria, https://team.inria.fr/epione/en/

[4] e-medicine (or digital medicine) is defined as the set of computational tools that are applied to the e-patient in order to assist the medical practitioner in several stages of the therapy such as diagnosis/prognosis, planning, control and evaluation [4].

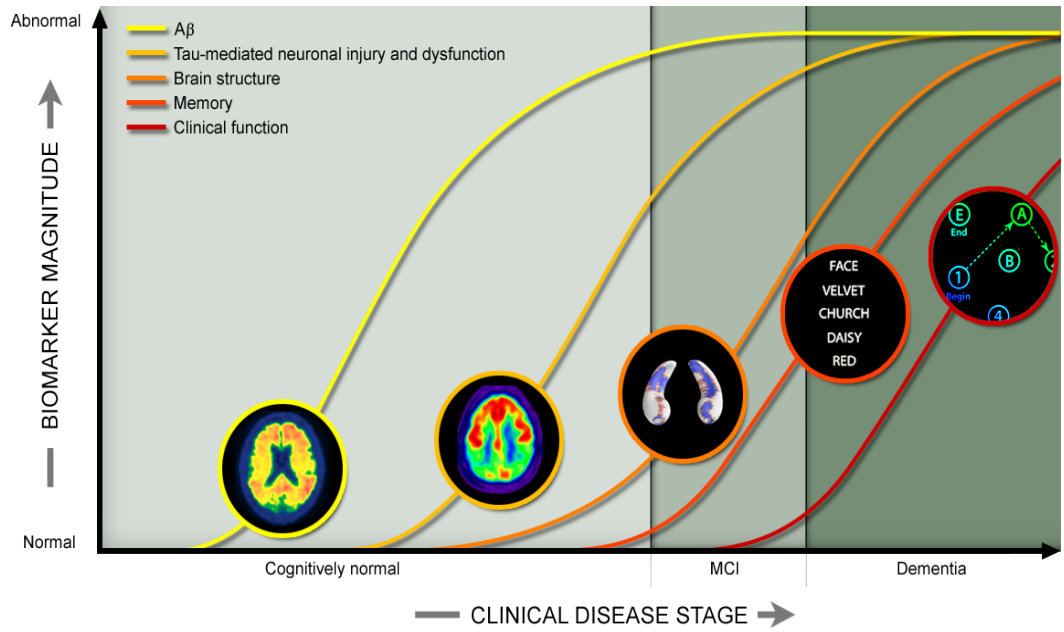


FIGURE 1.1: The evolution of biomarker over the course of AD progression: normal cognition to dementia. Biomarkers tracked (from left to right): $A\beta$ (β – *Amyloid peptide*), Tau protein, Brain atrophy (loss of brain cells), Memory loss (through cognitive tests) and clinical function (indicates cognitive decline). Changes in the last two biomarkers actually diagnose dementia. MCI: Mild Cognitive Impairment. (Figure sourced from [1])

The lack of effective therapies for NDDs such as AD and PD, despite of extensive research [8], is pushing researchers towards quantitative methods based on data like DPM: DPM has become a crucial tool for early diagnosis, therapeutic planning and prognosis due to its ability to visualize and quantify the effect of a drug and its clinical benefits [14, 12].

Traditionally in DPM, disease progression is modelled through a series of patient's data spread over time, also called *longitudinal data*, available through Electronic Health Records (EHRs) [15]. Datasets such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [16] and Parkinson's Progression Markers Initiative (PPMI) [17] are well known among the scientific community for statistical analysis of AD and PD, respectively. These datasets are the result of longitudinal multi-center studies containing multi-modal data such as clinical, advanced medical imaging, genetic and biological data.

1.3 Challenges and Objectives

Challenges:

There are several difficulties that arise while modeling longitudinal multi-modal data in the biomedical context [18, 19], the most prominent ones are listed below:

1. Heterogeneity in disease progression:
   (a) due to inter-patients variability
   (b) due to intra-patient variability
2. Missing and incomplete data: sparse and discrete data due to irregular patient visits and missing modalities
3. High dimensionality and noisy data

Objectives:

The main objectives of the internship are:

1. to theoretically develop and implement a novel Bayesian model framework for the assimilation of longitudinal biomedical data, suited for both single modality and multi-modal dataset,
2. to validate the effectiveness of the proposed model against a synthetically generated dataset,

1.4 Organisation of the Report

The remaining of this report is organized as follows: in Chapter 2, the general setting and challenges of biomedical data in DPM along with relevant State-of-the-Art are discussed. Chapter 3 details the theoretical description of our model followed by Chapter 4, which presents the results and discussion on model evaluation for the synthetic dataset. Additionally, Chapter 4 also mention some potential applications to our proposed model. Finally, Chapter 5 contains the conclusions and some insights into the planned work for the remainder of the internship, highlighting future research directions.

Remark 1. *The terminology patient and subject, multi-views and multi-modalities are used interchangeably throughout this report.*

Chapter 2

Analysis of Multi-modal Longitudinal Biomedical Data

In Section 1.3 of the previous chapter, I have highlighted a number of significant challenges that can arise while dealing with multi-modal biomedical data (e.g. ADNI, PPMI). These complexities in data hinder our understanding and modeling of disease progression. This chapter elaborates on each of these issues by exploring their intricacies which played a key role in the conceptualization of our model, and delves into the analytical methods from the DPM literature.

As discussed earlier, DPM plays a pivotal role to model the progression of chronic illnesses such as NDDs. To analyse the disease progression, the data is collected from the *longitudinal cohorts* over a period of time. At each patient's visits, the collected data consists of multi-modal observations. These repeated measurements allow us to describe the transition of biomarkers from normal to pathological stages (Figure 1.1). To understand the challenges posed by this kind of data it is essential to familiarize with the usual setting of the biomedical data acquisition and its characteristics. Figures, 2.1 and 2.2, are presented in this regard to describe the way in which dataset such as ADNI and PPMI, are collected.

Heterogeneity in disease progression

1. ***Inter-patient variability***: The time of onset of symptoms and the rate of disease progression differs among subjects [19, 20]. This heterogeneity can be seen in the Figure 2.1, where patient's diagnosis occurs and evolves at different ages and rates. This variability can be partially

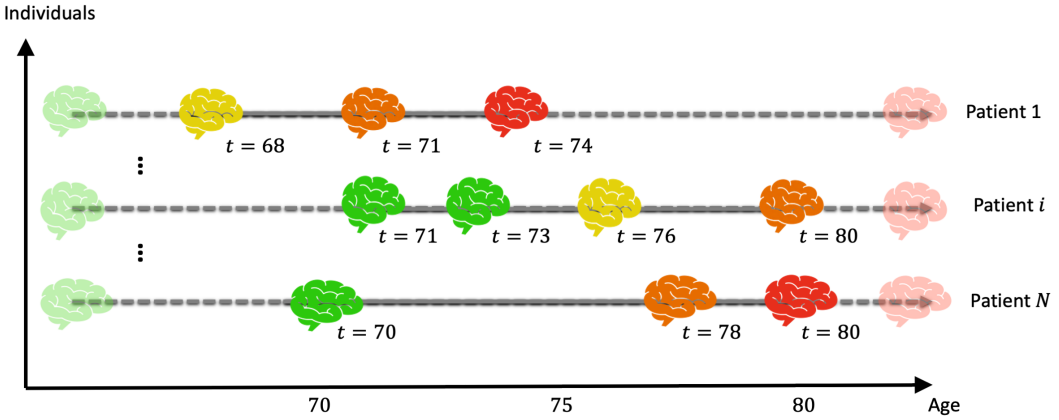


FIGURE 2.1: Schema of longitudinal data acquisition in NDDs. The different stages (four) in NDDs are displayed as unique colors. All the patients are assumed to be healthy at the time of birth (green) and eventually reach abnormality (red). Solid lines indicate the span through which patients are followed up for repetitive measurements. (Figure sourced from [2])

attributed to the subject-specific information such as sex, age, genes, lifestyle choices, etc.

Hence, the inherent timeline of the disease progression, estimated through DPM, requires *temporal alignment* of all the subjects data and should account for the inter-patient variability [2]. The inter-patient variability is figuratively depicted in 2.1 by varying sizes of the brain (height and width).

2. ***Intra-patient variability***: Here, the heterogeneity is observed in the evolution of various biomarkers despite the fact that the disease progression (for a given patient) is same. This is common in studies where biomarkers are analysed by combining different data types, heterogeneous data, coming from multiple data acquisition processes i.e., data from multiple views or modalities.

   For example, in the case of AD (Figure 1.1) PET scans are used to quantify the build up of $\beta-$amyloid ($A\beta$) plaques or to measure tau protein levels (to determine neurodegeneration) whereas MRI scans are used to measure brain atrophy. The remaining two are the clinical functions measured by cognitive tests. It is clear from Figure 1.1 that progression of all five biomarkers is different from one another. The first three biomarkers are called prodromal biomarkers and are invaluable in the early diagnosis of AD. Prodromal biomarkers [21] are those biomarkers that depict prodromal signs thus enabling earlier diagnosis or allowing for the outcome of interest to be determined at a more primitive stage of the disease. In NDDs, identification of such prodromal biomarkers

is still a challenge [8] likely due to the presence of several hidden covariates [19, 18].

Missing and incomplete data

Biomedical data acquisition and processing can be very expensive, hence the data collection is optimized to be cost effective. This is achieved by controlling the frequency (usually in months or years) of subject visits or by only considering specific modalities depending on the severity or speed of disease progression [16, 2, 18].

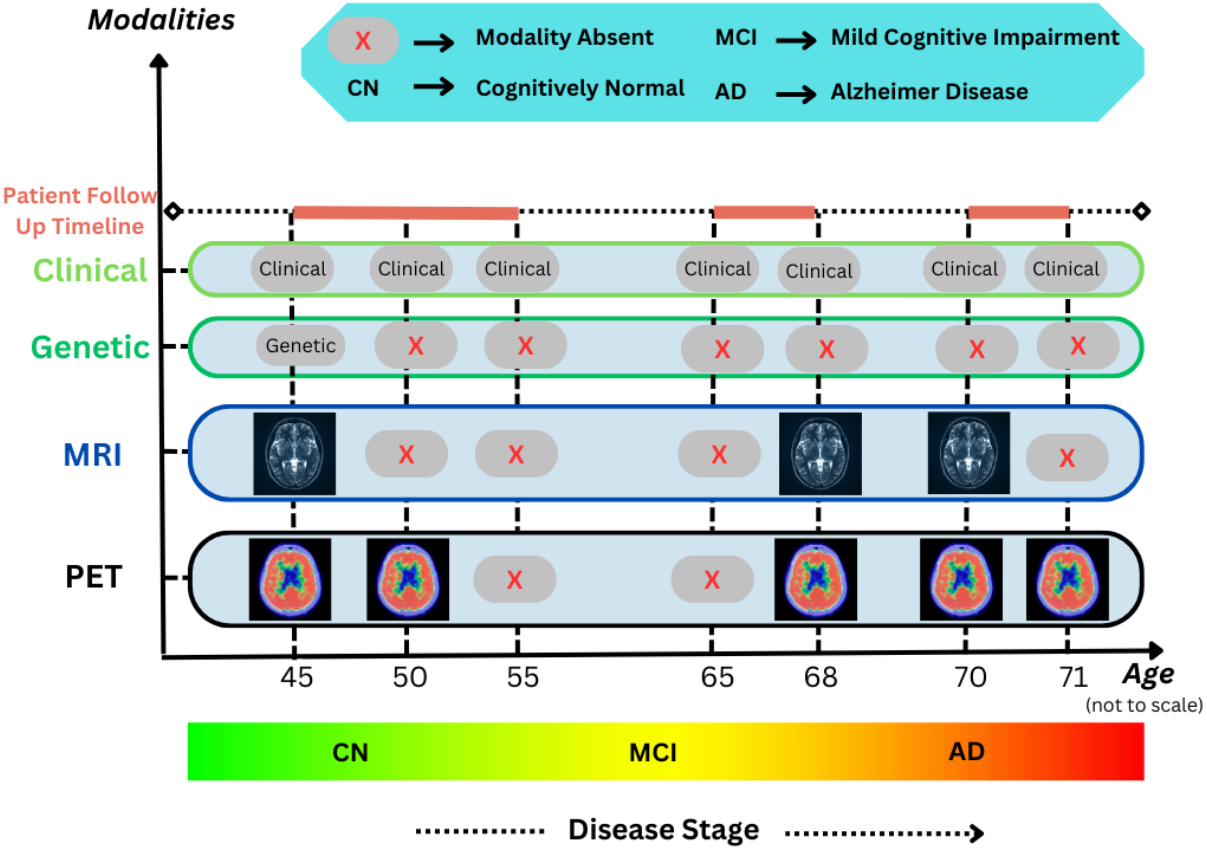


FIGURE 2.2: Multi-modal longitudinal data acquisition in NDDs (for a single patient). Usually, during the first visit, data for all the modalities might be collected or at least most of the modalities are included and further choice of modalities depends on the disease diagnosis and evolution. MRI: Magnetic Resonance Imaging, PET: Positron Emission Tomography

Figure 2.2, presents an hypothetical clinical study of a patient who is affected by AD. During the first visit, the patient is 45 years old and complete clinical (age, height, weight, blood pressure, cholesterol level or some

biomarkers), genetic and imaging data such as MRI, PET is taken. Disease diagnosis is made at each visit by tracking relevant biomarkers based on the clinical information, and the decision is made by the expert whether to include or not, all the other modalities or only a part of them. The frequency of the visits also varies across the different stages of the disease. It is 5 years during *Cognitively Normal* (CN) and 3 and 2 for *MCI* and *AD* stages, respectively. Also, notice that the entire data of the patient is missing until the first visit and for some other intervals (dotted lines in *Patient Follow Up Timeline*).

This incompleteness and missingness caused by the irregularity of the patient visits present substantial hurdles to any analysis, as they can lead to biased estimates effecting the statistical power of the model [19].

High dimensionality and noisy data

Biomedical data, especially imaging and genomics data, are usually very high-dimensional [22, 23, 20] and with relatively small sample size, posing computational as well as statistical challenges: Curse of dimensionality, feature selection. Extracting meaningful and interpretable information from such high-dimensional data requires sophisticated methods like dimension-reduction techniques to address this problem (more details are provided in Section 2.3).

Biomedical data can often be noisy [20, 22] and most likely prone to contain artifacts as well. These can arise from various sources, like the data acquisition process or preprocessing steps. These are not related to the disease process at all and affect the quality of the data, possibly leading to an unreliable analysis. Noise can occur due to some random errors or fluctuations of the equipment during measurement and can be especially problematic to detect in high-dimensional data. On the contrary, artifacts occur due to systematic errors in the equipment or some source of biased deviation from the ideal measurement scenario. Having a model that accounts for these kind of corrupted data by identifying and correcting for such noise and artifacts is an important step in the analysis of biomedical data.

2.1 Disease Progression Modeling (DPM) in NDDs

A first model of NDD's progression

Jack Jr et al. [24] conceptualized one of the first models on AD progression trajectory, in which the authors hypothesized the biomarkers progress during the course of the disease in the form of sigmoids (very similar to the one presented in Figure 1.1). Though this model is an abstract construction coming from practitioners experience, it paved the way for statisticians to theorize and develop data-driven models. These models are roughly classified as discrete and continuous models.

Discrete models

Event-Based Models (EBMs) (Fonteijn et al. [25], Young et al. [26]) are discrete models, in these models the pathological evolution in AD is modeled from biomarkers measurements using "cascade of events" perspective. The progression of the disease is unfolded through a sequence of events characterized by a collection of biological and cognitive changes due to AD. A Gaussian mixture model of two distributions providing normal and abnormal values is used for modeling the biomarkers. Data is assumed to follow a generative model, and Maximum likelihood Estimation (MLE) is used to identify the most probable sequence of events. EBM is successful in estimating the ordering of events over the full course of AD in agreement with the hypothetical sigmoid model of Jack Jr et al. [24]. In addition, the generative capabilities of EBMs are able to quantify the uncertainty in the estimated order of events thereby informing on the confidence of the predicted order. The EBMs are extended to Huntington [27] and Parkinson's diseases [28] as well. However, these models failed to find wider acceptance as the application of these models is limited due to the lack of finer temporal dynamics because they only model *discrete* sequence of events transitioning from normal to abnormal states.

Continuous Models

Attempts for the continuous modeling of disease progression faced difficulties because of the lack of an inherent reference timeline in NDDs due to the

inter-patient variability discussed before. To overcome this, Jedynak et al. [29] proposed an individual *time-warping function* that encodes the individual patient trajectories. Temporal realignment in patients is widely performed via this affine reparameterization function that summarizes the individual progression through a rate of progression (slope) and intercept (at the onset of data collection) for individuals.

Most of the current continuous disease progression models [30, 31, 32, 33, 2, 23] use this realignment strategy and predefined assumptions (e.g. sigmoid) on biomarker's dynamics. Guerrero et al. [31] and Donohue et al. [30] built non-linear mixed effect models [34] from short-term patient data by treating patient parameters as random effects of the model and perceiving the long-term average disease trajectory as fixed effects of the model. The authors in Marinescu et al. [33] suggest a novel approach unlike the preceding models that concentrated on scalar values such as medical examinations or quantifiable metrics derived from imaging modalities. Their strategy involves direct modeling of the trajectories for each voxel extracted from the imaging data. Their method, known as DIVE, approximates a global progression pattern for all patients by categorizing vertices of voxels into distinct clusters, each defined by a sigmoid parameterization.

Additionally, Lorenzi et al. [35], Garbarino et al. [36], Lorenzi and Filippone [37], pursued a non-parametric approach in order to allow for a broader array of temporal profiles other than sigmoid. These models map the dynamic evolution of biomarkers using Gaussian Processes associated with a monotonicity constraint.

Overall, these techniques enable the modeling of long-term dynamics for a wide array of scalar biomarkers throughout a continuous timeline depicting the course of the disease. These models use cases extend to performing individual disease staging. Disease staging [38] involves identifying the subject's disease severity that best represents the pathological progression observed. In essence, they permit a more comprehensive understanding of a disease's progression over time, offering the potential for more precise and personalized diagnosis.

2.2 DPM with Multi-modal Data

Until now, I have mostly discussed models where the disease progression is analysed using data from a single modality. However, we have also seen that the analysis of biomedical datasets such as ADNI and PPMI necessitates the simultaneous modeling of multiple modalities. These include, for instance, clinical scores and medical imaging data such as MRI and PET scans. Each of these views provides unique and complementary insights about the disease under study. Therefore, a comprehensive analysis of all views could improve the diagnostic accuracy, uncovering pathological correlations or projecting the progression of the disease.

Spatio-temporal approaches

Multi-modal imaging data analysis is common in several computational anatomy studies [32, 20, 2, 39, 40], as the spatial information provides richer anatomical description than data coming from scalar biomarkers data such as amyloid level or clinical scores, etc. In Koval et al. [39, 40], the authors used the cortical thickness data to explain the spatio-temporal changes affecting the brain during AD. Abi Nader et al. [41] developed a simulation model for development of drug intervention strategies for AD 'SimulAD' using multi-modal imaging and clinical data from ADNI. Their experiments are found to be compatible with clinical trials and the study recommended that anti-amyloid treatments should be administered a minimum of 7 years earlier than the current clinical practices to achieve a statistically significant enhancement in clinical outcomes.

Moreover, study by Cooper and Chahine [42] concluded that a multi-modal approach is needed for accurate identification of prodromal biomarkers for PD. Hence, building models based on multivariate and multi-modal imaging data is crucial to investigate the dynamics and topographical changes affecting the brain due to NDDs.

However, merging multi-view data while acknowledging their interrelationships and collective variability presents numerous complexities. These challenges are magnified as the multi-modal imaging data is often high-dimensional and noisy [20]. Therefore, efforts are needed towards addressing these hurdles to maximize the potential of multi-view data in advancing healthcare outcomes.

2.3 Latent Variable Models (LVMs) for High Dimensional Data

Latent variable models (LVMs) [43] are based on the assumption that there exists a space, consisting of hidden variables, that can sufficiently represent the underlying data. These are called latent or hidden variables as they cannot be explicitly observed and require to be inferred from the observed variables in the data. The task lies in extracting meaningful relationships via a mathematical mapping function, to transform the observed variables to hidden variables. The goal of the LVMs is to find a lower dimensional representational space that can well explain the correlations and variability among the observed variables. Hence, they are widely used as a form of dimension reductionality techniques [44].

The LVMs are of particular interest in biomedical domain due to the wide prevalence of high-dimensional datasets. LVMs are typically generative and this is an important aspect because when the model is trained it deduces a distribution of latent space for the observed data rather than a single representation. By randomly sampling from this distribution one can generate new data that exhibit same characteristics as the original data. Additionally, the generative capabilities of LVMs can also be utilized to measure the uncertainty and variability in the observed data or for missing data imputation. Hence, LVMs can be invaluable within the healthcare industry due to their ability to simulate hypothetical scenarios [41] such as clinical trials or treatments, thereby contributing significantly to the understanding of diseases and advancement in drug development, while also being cost effective.

The LVMs can be either linear or non-linear depending on the mapping function used to make the transformation between latent and observed spaces. Principal Component Analysis (PCA) [45, 46] is a one of the first known linear LVM, it involves linearly projecting the high-dimensional data onto the lower-dimensional principal axes. The principal axes are chosen such that the projected space retains or captures the maximum variance in the data. Probabilistic Principal Component Analysis (PPCA) is also a linear LVM, which, as the name suggests, is a probabilistic extension of PCA proposed by Tipping and Bishop [47]. PPCA is a generative model based on factor analysis [48]. Coming to non-linear LVMs, Generative Adversarial Networks (GANs) [49] and Variational Autoencoders (VAEs) [50] are two of the well known models. In these, Neural Networks (NNs) are used to

perform complex and non-linear transformations between the latent and observed spaces. While non-linear mappings can capture more intricate patterns and complex structures in the data they fail to maintain a simple and explicit form, often leading to models that are computationally intensive and difficult to interpret.

There are several works found in the literature that have utilized LVMs to analyse biomedical data [51, 52, 22, 23]. Multi-Omics Factor Analysis (MOFA) by Argelaguet et al. [52] generalizes PPCA for the analysis of the multi-omics datasets and is able to adapt to various data types while also integrating multiple noise models. On the other hand, multi-channel Variational Autoencoder (mc-VAE) by Antelmi et al. [51] is an extension of the traditional VAEs by Kingma and Welling [50], designed specifically to handle multi-view data coming from a patient with AD.

Senacheribbe [23] proposed a Latent Slope-Intercept (LSI) model based on PPCA for the analysis of longitudinal data. LSI model applied on the ADNI dataset, in an unsupervised setting, was able to obtain a latent representation where projections of the healthy patients are segregated from the unhealthy ones. Moreover, recent contributions from Balelli et al. [22] have brought a new perspective into this field. They introduced a multi-view latent variable model, which interestingly is constructed on the foundations of the PPCA methodology.

Chapter 3

Bayesian Latent Variable Model for the Analysis of Multi-modal Longitudinal Data

In this chapter, I present our model based on Bayesian framework that extends PPCA [47] for the Analysis of Multi-modal Longitudinal Data. This model attempts to address the challenges laid out in the previous chapter (2). This model resulted by fusing the LSI model, from Senacheribbe [23], with the multi-modal framework of Balelli et al. [22] along with additional flexibilities. Further details regarding the model can be found in appendix A.

3.1 Theoretical Formulation

Problem Setup

Let us consider that we have data coming from a total of **'N'** subjects and it lies in a **d**-dimensional space consisting of **'K'** modalities such that:

$$d = \sum_{k=1}^{K} d_k \tag{3.1}$$

Where, $\forall k \in \{1, \ldots, K\}$, d_k is the number of features defining the k^{th} view i.e., the dimensionality of the k^{th} modality.

For every subject **'n'** and for each modality **'k'**, the data is collected at multiple time instants i.e., longitudinal data, in accordance with the corresponding time vector, $t_{n^{(k)}}$, of the subject. They are given as:

$$t_{n^{(k)}} = [t_{n1^{(k)}}, \ldots, t_{nT_n{}^{(k)}}]^T \in \mathbb{R}^{+T_n{}^{(k)}}; \forall n \in \{1, \ldots, N\}, \forall k \in \{1, \ldots, K\} \quad (3.2)$$

Where, $T_n{}^{(k)}$ is the total number of time stamps in $t_{n^{(k)}}$.

For a given modality **'k'**, the data collected at any time stamp **'t'** $\in t_{n^{(k)}}$, is denoted as $y_{nt}^{(k)} \in \mathbb{R}^{d_k}$, and the data from all the modalities in the concatenated form is:

$$y_{nt} = [y_{nt}^{(1)^T}, \ldots, y_{nt}^{(K)^T}]^T \in \mathbb{R}^d \quad (3.3)$$

Remark 2. *For the sake of simplicity we represented all* **'K'** *views. If the* k^{th}*-view is missing, than it will be simply removed, e.g. one would have:*

$$y_{nt} = [y_{nt}^{(1)^T}, \ldots, y_{nt}^{(k-1)^T}, y_{nt}^{(k+1)^T}, \ldots, y_{nt}^{(K)^T}]^T \in \mathbb{R}^{d-d_k}$$

Same is applied in even in case of more than one missing views.

Overall, the entire data for each patient **'n'** can be summarized by a matrix y_n, given below, of size $T_n \times d$, a **d-**dimensional longitudinal data of length $'T_n'$

$$y_n = \begin{bmatrix} y_{nt_1 1} \cdots y_{nt_1 d} \\ \vdots \\ y_{nt_{T_n} 1} \cdots y_{nt_{T_n} d} \end{bmatrix}_{T_n \times d} \quad (3.4)$$

Remark 3. *Here,* $T_n \neq \sum_{k=1}^{K} T_n{}^{(k)}$*,* T_n *is the number of unique time stamps for subject* **'n'** *across all the modalities.*

Modeling Assumptions

The aim of the model is to provide a common **'q'**-dimensional latent space representation for the longitudinal patient data $'y'_n$ observed in $d-$dimensional space made up of several views.

To model this I made the following critical assumptions:

1. Existence of a $q-$dimensional latent space with $q < min_k(d_k)$.
2. The observed data across the subjects is independent of each other.
3. The observed data follows a linear generative model (refer to 3.5).
4. The observed data evolves linearly over time (refer to 3.6).
5. Given the latent representation of the subject, the subject's data can be observed independently across all the modalities and for all the time stamps. (refer to eq. 3.10)
6. The errors in the observed data due to noise during data collection, in each modality, follows a Isotropic Gaussian distribution.
7. Same prior on the latent variable for all the subjects.
8. The latent space is time independent.

Model Definition

With the assumptions mentioned earlier, consider a patient n at a generic time stamp t i.e., some row of the matrix y_n (refer to 3.4). The observed data for this patient for a specific modality k follows the following generative model, given by 3.5 & 3.6.

$$y_{nt}^{(k)} = (tW^{(k)} + V^{(k)})x_n + t\omega^{(k)} + \mu^{(k)} + \epsilon^{(k)} \\ ,\forall k \in \{1,\ldots,K\}, t \in t_n, n \in \{1,\ldots,N\} \tag{3.5}$$

Equation 3.5 can be rearranged and written as:

$$y_{nt}^{(k)} = t(W^{(k)}x_n + \omega^{(k)}) + V^{(k)}x_n + \mu^{(k)} + \epsilon^{(k)} \\ ,\forall k \in \{1,\dots,K\}, t \in t_n, n \in \{1,\dots,N\} \tag{3.6}$$

Where,

- $W^{(k)}, V^{(k)} \in \mathbb{R}^{d_k \times q}$ & $y_{nt}^{(k)}, \omega^{(k)}, \mu^{(k)}, \epsilon^{(k)} \in \mathbb{R}^{d_k}$.
- $x_n \in \mathbb{R}^q$ is a $q-$dimensional latent variable.
- The term $(tW^{(k)} + V^{(k)})$ in eq. 3.5 provides the linear mapping between the observed and latent variables.
- From eq. 3.5 it is clear that for all the modalities, $\forall k \in \{1,\dots,K\}$, the terms associated with x_n signifies and explains subject specific evolution i.e., inter-patient variability, whereas the independent terms $\omega^{(k)}$, $\mu^{(k)}$ explains the global (average) dynamics of the observed data.
- In eq. 3.6, the terms $(W^{(k)}x_n + \omega^{(k)})$ and $V^{(k)}x_n + \mu^{(k)}$ are the intercept and slope, respectively. These terms characterize the linear time evolution of the $y_{nt}^{(k)}$.
- $\epsilon^{(k)} \sim \mathcal{N}(0, {\sigma^{(k)}}^2 \mathbb{I}_{d_k})$, Isotropic Gaussian Noise.
- The parameters $W^{(k)}, V^{(k)}, \omega^{(k)}, \mu^{(k)}$, *and* $\epsilon^{(k)}$, are all different for different values of k. This explains the intra-patient variability.

Likelihood

From eq. 3.6, we can derive the likelihood of the observable variables conditioned on the value of the latent variable x_n as:

$$y_{nt}^{(k)}|x_n \sim \mathcal{N}((tW^{(k)} + V^{(k)})x_n + t\omega^{(k)} + \mu^{(k)}, {\sigma^{(k)}}^2 \mathbb{I}_{d_k}) \tag{3.7}$$

Now we can use eq. 3.3 to combine all the modalities, as shown below, and the resulting likelihood is given by the eq. 3.9.

$$\begin{aligned} y_{nt} &= [{y_{nt}^{(1)}}^T, \ldots, {y_{nt}^{(K)}}^T]^T \\ &= (tW + V)x_n + t\omega + \mu + \epsilon \end{aligned} \tag{3.8}$$

Where,

- $y_{nt} \in \mathbb{R}^d$
- $W = [{W^{(1)}}^T, \ldots, {W^{(K)}}^T]^T \in \mathbb{R}^{d \times q}$
- $V = [{V^{(1)}}^T, \ldots, {V^{(K)}}^T]^T \in \mathbb{R}^{d \times q}$
- $\omega = [{\omega^{(1)}}^T, \ldots, {\omega^{(K)}}^T]^T \in \mathbb{R}^d$
- $\mu = [{\mu^{(1)}}^T, \ldots, {\mu^{(K)}}^T]^T \in \mathbb{R}^d$
- $\epsilon = [{\epsilon^{(1)}}^T, \ldots, {\epsilon^{(K)}}^T]^T \in \mathbb{R}^d \ \ and \ \ \epsilon \sim \mathcal{N}(0, \psi)$
- $\psi \ \ is \ \ a \ \ diagonal \ \ block - matrix, \ \ \psi = diag({\sigma^{(1)}}^2 \mathbb{I}_{d_1}, \ldots, {\sigma^{(K)}}^2 \mathbb{I}_{d_K})$

Therefore, the likelihood for the observed data of a subject n at a time t conditional to knowing x_n is:

$$y_{nt}|x_n \sim \mathcal{N}((tW + V)x_n + t\omega + \mu, \psi) \tag{3.9}$$

Remark 4. *Notice that inline with our assumption (3), in eq. 3.9 the covariance is a block-diagonal matrix i.e., given x_n there is independence across the modalities. Similarly, for a given subject, the same can be proven for data points across the time instants.*

Prior

For all the subjects, the prior on the latent variable is assumed to be following a normal distribution by assuming independence among the variables in latent space:

$$x_n \sim \mathcal{N}(0, \mathbb{I}_q), \forall \ \ n \in \{1, 2, \ldots, N\} \tag{3.10}$$

Marginal Likelihood

The likelihood of the observed data at the time t marginalised over the prior x_n can be derived through linear transformations of the Gaussian distributions, and the result is given by:

$$y_{nt} \sim \mathcal{N}(t\omega + \mu, \mathcal{C}) \tag{3.11}$$

Where, $\mathcal{C} = \psi + (tW + V)(tW + V)^T$

Remark 5. *1. Notice that eq. 3.11 provides the direct relationship between the model parameters and the observed data. It is also interesting to see that the covariance matrix is not diagonal thus explaining the correlation among the different modalities. Similarly, for a given subject, the same can be proven for data points across the time instants.*

2. Until here, I have only considered the data for a subject n across all the modalities but only at a given time instant t. However, as we know that the subject also has correlated data spread across their time vector. Though, it is interesting to see the complete explicit vectorial formulations for all the subjects at all the time instants they does not offer additional information in order to solve the problem at hand and hence they are omitted. One can refer to Appendix A of Senacheribbe [23] for these formulations.

Posterior Distribution

Similar to PPCA [47], given the data of the subjects one can estimate the latent space i.e., derive the posterior distribution via the Bayes's rule as:

$$\begin{aligned} p(x_n|y_{nt_{n1}}, \ldots, y_{nt_{nT_n}}) &= \frac{p(y_{nt_{n1}}, \ldots, y_{nt_{nT_n}}|x_n)p(x_n)}{p(y_{nt_{n1}}, \ldots, y_{nt_{nT_n}})} \\ &\propto p(y_{nt_{n1}}, \ldots, y_{nt_{nT_n}}|x_n)p(x_n) \end{aligned} \tag{3.12}$$

After solving eq. 3.12, the posterior is given by:

$$x_n|y_{nt_{n1}},\ldots,y_{nt_{nT_n}} \sim \mathcal{N}\left(\Sigma \sum_{k=1}^{K} \frac{1}{{\sigma^{(k)}}^2} \sum_{t=t_{n1}}^{T_n^{(k)}} (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}), \Sigma\right) \tag{3.13}$$

$$\text{With, } \Sigma = \left(\sum_{k=1}^{K} \frac{1}{{\sigma^{(k)}}^2} \sum_{t=t_{n1}}^{T_n^{(k)}} (tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) + \mathbb{I}_q\right)^{-1}$$

The solution to the derivation of posterior distribution is included in the Appendix A.

Predictive Distribution

Suppose that we want predict the data inferred through model parameters and latent space, for a subject n at some future time instant t_*. The likelihood of data observed at the time instant t_* is given as (from eq. 3.9:

$$y_{nt_*}|x_n \sim \mathcal{N}((t_*W + V)x_n + t_*\omega + \mu, \psi) \tag{3.14}$$

Combining the likelihood $y_{nt_*}|x_n$ from eq. 3.14 above with the posterior distribution in eq. 3.13 yields in the following predictive distribution:

$$y_{nt_*}|y_{nt_{n1}},\ldots,y_{nt_{nT_n}} \sim \mathcal{N}\Bigg((t_*W + V)\Big(\Sigma \sum_{k=1}^{K} \frac{1}{{\sigma^{(k)}}^2} \sum_{t=t_{n1}}^{T_n^{(k)}} (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})\Big), \psi + (t_*W + V)\Sigma(t_*W + V)^T\Bigg) \tag{3.15}$$

$$\text{With, } \Sigma = \left(\sum_{k=1}^{K} \frac{1}{{\sigma^{(k)}}^2} \sum_{t=t_{n1}}^{T_n^{(k)}} (tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) + \mathbb{I}_q\right)^{-1}$$

3.2 Optimisation

Our model parameters are characterized by both observed and latent spaces, therefore we can find their joint distribution and maximize their likeliness to occur by tuning over the parameter space i.e., we find the optimal parameters of our proposed model through Maximum Likelihood Estimation (MLE).

Complete Data Log-likelihood

By keeping in mind, the modeling assumptions I discussed earlier, the complete data likelihood can be derived as:

$$\begin{aligned}\text{Likelihood: } &= p(x_1, y_1, \ldots, x_N, y_N) = \prod_{n=1}^{N} p(x_n, y_n) = \prod_{n=1}^{N} p(y_n|x_n) * p(x_n) \\ &= \prod_{n=1}^{N} p(y_{nt_{n1}, \ldots, y_{nt_{nT_n}}}|x_n) * p(x_n) = \prod_{n=1}^{N} \left[\prod_{t=t_{n1}}^{T_n} p(y_{nt}|x_n) \right] * p(x_n)\end{aligned}$$

$$\text{Log-Likelihood} := \mathcal{L} = ln(\text{Likelihood}) = \sum_{n=1}^{N} \left[\sum_{t=t_{n1}}^{T_n} ln\big(p(y_{nt}|x_n)\big) + ln\big(p(x_n)\big) \right]$$

After utilizing the necessary equations from previous section, we will arrive at the following expression for the likelihood:

$$\begin{aligned}\mathcal{L} = -\sum_{n=1}^{N} \Bigg[\Bigg\{ \sum_{k=1}^{K} \sum_{t=t_{n1}}^{T_n^{(k)}} \Bigg(&\frac{d_k}{2} ln(2\pi) + \frac{d_k}{2} ln({\sigma^{(k)}}^2) + \frac{1}{2{\sigma^{(k)}}^2} (||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 \\ &+ x_n^T (tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) x_n \\ &- 2x_n^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})) \Bigg) \Bigg\} + \frac{1}{2} ||x_n||^2 \Bigg] \\ &= f(x, y; W, V, \omega, \mu, \sigma^{(1)}, \ldots, \sigma^{(K)})\end{aligned} \tag{3.16}$$

The steps involved in the derivation of the eq. 3.16 are detailed in the Appendix, A.2.

Computational Complexity

To find the optimal parameters, one possibility is to derive an exact analytical solution for the expression in eq. 3.16, by setting the partial derivatives with respect to each of them to 0. As pointed out in Senacheribbe [23], this can be computationally expensive since it requires to compute the data covariance matrix in time, yielding to an asymptotic complexity of $\mathcal{O}(NT^2d^2)$. An alternative is to use Expectation-Maximization (EM) Algorithm by relying on the optimization framework originally introduced by Tipping and Bishop [47] and this reduces the complexity to $\mathcal{O}(Nq\prod_{k=1}^{K} d_k) \approx \mathcal{O}(N\prod_{k=1}^{K} d_k)$ when $q << min_k(d_k)$, according to Balelli et al. [22] (for subjects with single data point) and hence $\mathcal{O}(NT^2q\prod_{k=1}^{K} d_k) \approx \mathcal{O}(NT^2\prod_{k=1}^{K} d_k)$ in our setting.

Remark 6. *T is the average length of all the time vectors across the subjects and modalities.*

Optimization using EM Algorithm

The EM Algorithm consists of two steps i.e., an Expectation step (E-step) and the Maximization step (M-step). During the E-step, the expected complete log likelihood with respect to the latent space (assumed to be missing) is computed based on the observed data and the current estimates or realizations of the parameters. Subsequently, during the M-step, this likelihood is maximized concerning the model parameters. The EM Algorithm for our model is as follows:

Initialization-**Step**

Initialize the parameters randomly (based on some choice of prior distribution):

$$W, V, \omega, \mu, \sigma^{(1)}, \ldots, \sigma^{(K)}$$

E-**Step**

Evaluate:

$$\mathbb{E}_{x|y,W,V,\omega,\mu,\sigma^{(1)},\ldots,\sigma^{(K)}}(\mathcal{L})$$

$$\mathbb{E}[\mathcal{L}] = \langle \mathcal{L} \rangle = -\sum_{n=1}^{N} \Bigg[\Bigg\{ \sum_{k=1}^{K} \sum_{t=t_{n1}}^{T_n^{(k)}} \Bigg(\frac{d_k}{2} ln(2\pi) + \frac{d_k}{2} ln({\sigma^{(k)}}^2) + \frac{1}{2{\sigma^{(k)}}^2} (||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) \langle x_n {x_n}^T \rangle\big) - 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})) \Bigg) \Bigg\} + \frac{1}{2} Tr(\langle x_n {x_n}^T \rangle) \Bigg] \tag{3.17}$$

Here, $\langle x_n \rangle$ and $\langle x_n {x_n}^T \rangle$ are the first and second order moments of the posterior distribution from eq. 3.13, respectively.

M-Step

Maximize the expectation by searching through the parameter space of:

$$W, V, \omega, \mu, \sigma^{(1)}, \ldots, \sigma^{(K)}$$

For each $k : k \in \{1, \ldots, K\}$ the optimum parameters are:

- $\mu^{(k)}$

$$\frac{\partial \langle \mathcal{L} \rangle_{\mu^{(k)}}}{\partial \mu^{(k)}}\bigg|_{\widetilde{\mu}^{(k)}} = 0 \implies \widetilde{\mu}^{(k)} = \frac{1}{\sum_{n=1}^{N} T_n} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} (y_{nt}^{(k)} - (tW^{(k)} + V^{(k)}) \langle x_n \rangle - t\omega^{(k)}) \tag{3.18}$$

- $\omega^{(k)}$

$$\frac{\partial \langle \mathcal{L} \rangle_{\omega^{(k)}}}{\partial \omega^{(k)}}\bigg|_{\widetilde{\omega}^{(k)}} = 0 \implies \widetilde{\omega}^{(k)} = \frac{1}{\sum_{n=1}^{N} \tau_n^{(k)}} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} t(y_{nt}^{(k)} - (tW^{(k)} + V^{(k)}) \langle x_n \rangle - \mu^{(k)}) \tag{3.19}$$

  Where, $\tau_n^{(k)} = \sum_{t=t_{n1}}^{T_n^{(k)}} t^2$.

- $W^{(k)}$

$$\frac{\partial \langle \mathcal{L} \rangle_{W^{(k)}}}{\partial W^{(k)}}\bigg|_{\widetilde{W}^{(k)}} = 0 \implies \widetilde{W}^{(k)} = \Bigg(\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} t(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T - tV^{(k)} \langle x_n {x_n}^T \rangle \Bigg) \Bigg(\sum_{n=1}^{N} \tau_n^{(k)} \langle x_n {x_n}^T \rangle \Bigg)^{-1} \tag{3.20}$$

Where, $\tau_n^{(k)} = \sum_{t=t_{n1}}^{T_n^{(k)}} t^2$.

- $V^{(k)}$

$$\frac{\partial \langle \mathcal{L} \rangle_{V^{(k)}}}{\partial V^{(k)}}\bigg|_{\widetilde{V}^{(k)}} = 0 \implies \widetilde{V}^{(k)} = \Bigg(\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T - tW^{(k)} \langle x_n {x_n}^T \rangle \Bigg) \Bigg(\sum_{n=1}^{N} T_n^{(k)} \langle x_n {x_n}^T \rangle \Bigg)^{-1} \tag{3.21}$$

- $\sigma^{2(k)}$

$$\frac{\partial \langle \mathcal{L} \rangle_{\sigma^{2(k)}}}{\partial \sigma^{2(k)}}\bigg|_{\widetilde{\sigma^2}^{(k)}} = 0$$

$$\implies \widetilde{\sigma^2}^{(k)} = \frac{1}{d_k \sum_{n=1}^{N} T_n^{(k)}} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \Big(||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) \langle x_n {x_n}^T \rangle\big) - 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \Big) \tag{3.22}$$

Chapter 4

Application to Synthetic Data

In this chapter, I present two experiments, to which I have applied our Bayesian Latent Variable Model from Chapter 3 using synthetically generated datasets.

Remark 7. *Throughout the course of my internship, all the code development was undertaken using the Python programming language. For performing mathematical operations, the NumPy library was extensively utilized [53]. Data visualization was performed using two popular Python libraries: Matplotlib [54] and Seaborn [55]. These tools and libraries facilitated efficient project development and contributed significantly to the quality and clarity of our outcomes.*

4.1 Meta Data

The Meta data is user defined and it is required to generate the synthetic dataset are:

- $n :=$ The number of subjects
- $d := [d_1, \ldots, d_K]$ The vector of dimensions containing the dimension size of all the K modalities
- $q :=$ The size of the latent dimension
- $min\_len :=$ The minimum number of time stamps per subject
- $max\_len :=$ The maximum number of time stamps per subject
- $max\_time :=$ The maximum time until which a subject can have a data

The Meta data for the two experiments are presented in Table 4.1.

#Experiment	n	d	q	min_len	max_len	max_time
1	500	[8, 10]	5	15	20	20
2	1000	[3, 4]	2	10	15	15

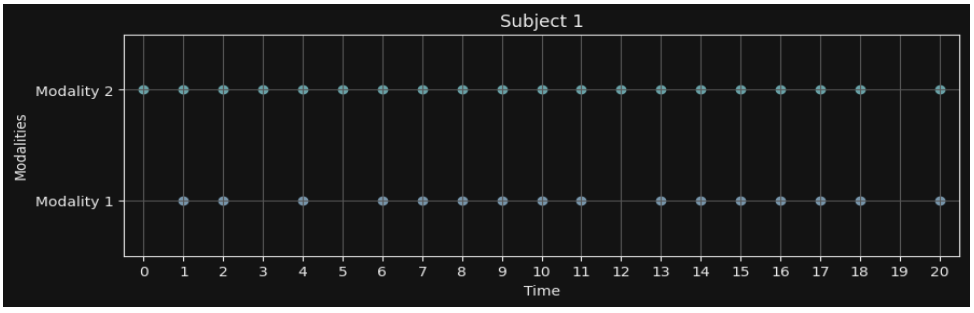
TABLE 4.1: Meta Data for the experiments

4.2 Synthetic Data Generation

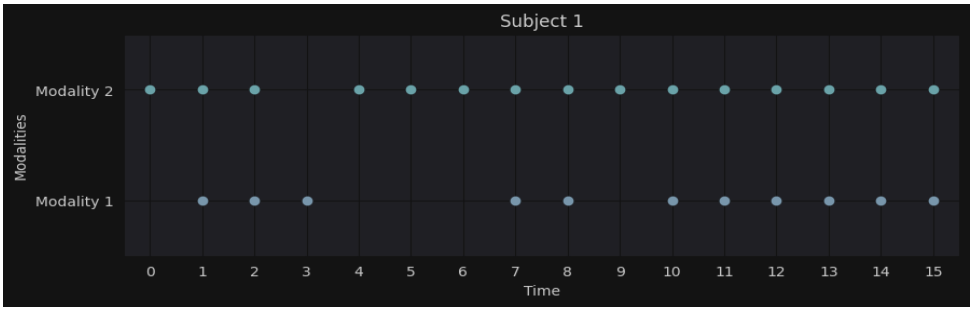
Generating Time Sequences

Once the Meta data (4.1) is known, the time sequences for all the subjects in all the modalities are generated randomly.

First, the length of time sequence i.e., number of time stamps each subject will have for any given modality is chosen randomly between $min\_len :=$ and $max\_len :=$, with uniform probability. After choosing the length, the sequence is then generated, randomly with uniform distribution, containing those many time stamps. All time sequences, for all the subjects and modalities, lie within $0 - max\_time$. In Figure 4.1, it can be noticed that the time sequences are different across the modalities.



(A) Experiment 1



(B) Experiment 2

FIGURE 4.1: Time sequences for subject 1 across modalities

Choosing Model Parameters

All the individual values of the following model parameters:

$$\forall k \in \{1, \ldots, K\} \;\; W^{(k)}, V^{(k)} \in \mathbb{R}^{d_k \times q} \quad \& \quad \omega^{(k)}, \mu^{(k)} \in \mathbb{R}^{d_k},$$

are generated using the uniform distribution: $\mathcal{U}(-1, 1)$

Whereas, $\forall k \in \{1, \ldots, K\} \;\; {\sigma^{(k)}}^2 \in \mathbb{R}^+$ are generated using : $\mathcal{U}(1, 2)$

Sampling in Latent Space

The primary difference between the two experiments is the way in which the latent space is sampled.

Experiment 1

For the experiment 1, all the subjects in the latent space are sampled (refer Figure. 4.2a) from a standard multivariate Gaussian distribution i.e.,

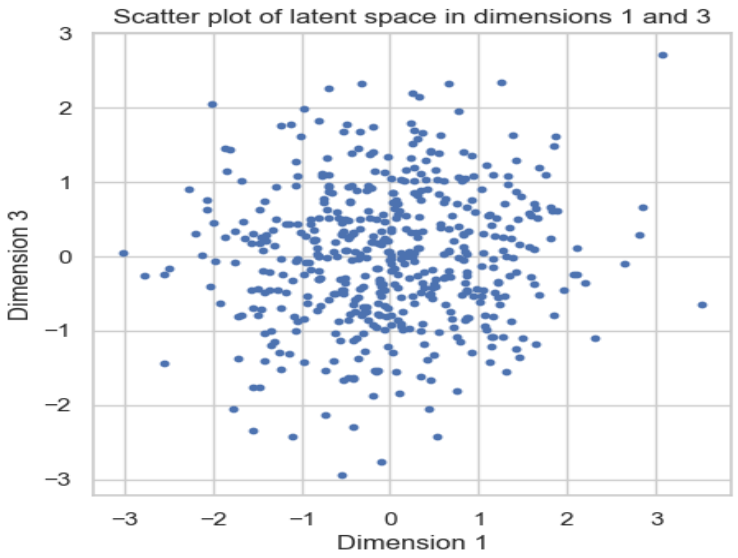
$$x_n \sim \mathcal{N}(0, \mathbb{I}_q), \forall \;\; n \in \{1, 2, \ldots, N\}$$

Experiment 2

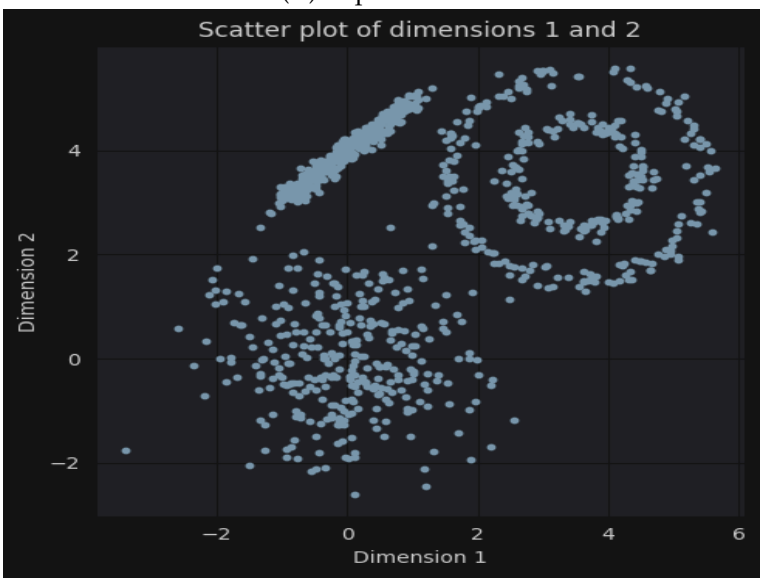
For the experiment 2, the subjects in the latent space are sampled (Figure. 4.2b) randomly from three different distributions out of which one is standard multivariate Gaussian like in experiment 1, whereas the other two are points sampled (with some noise) about a straight line and two concentric circles, respectively.

Generating the Observed Data

After choosing the model parameters and sampling the latent space, the observed data for all the subjects, in all modalities and across all the time stamps, is generated along with noise as per the eq. 3.7. The evolution of the observed data w.r.t the time can be seen in Figures. 4.5, 4.6.



(A) Experiment 1



(B) Experiment 2

FIGURE 4.2: Sampling in Latent Space

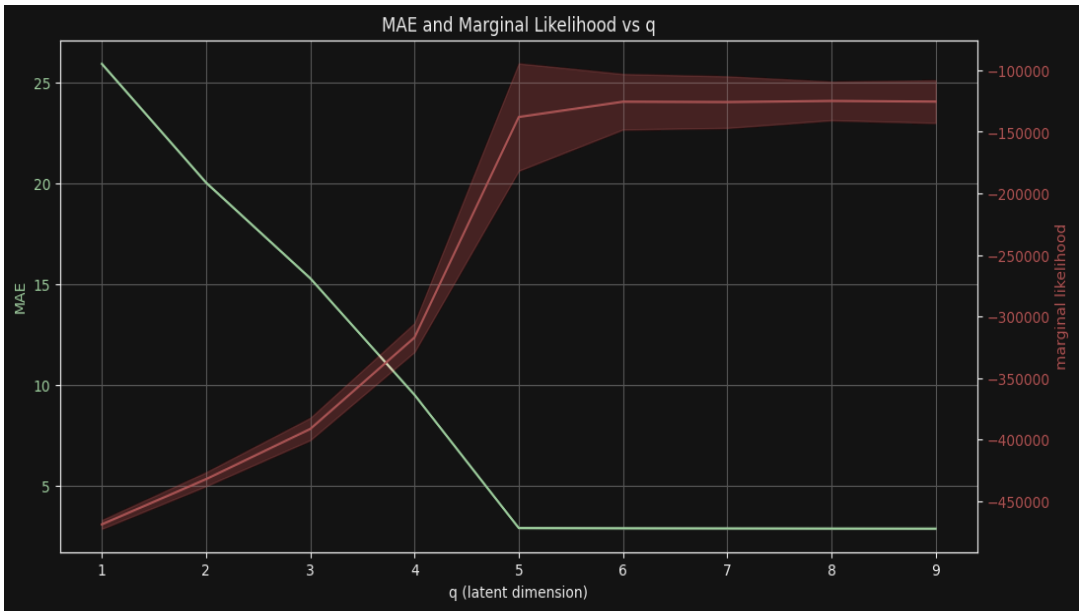
4.3 Model Evaluation

Now assuming that we only have the observed data, we can fit our model to a latent space of dimension q and find the optimal parameters using the $EM-$Algorithm mentioned in 3.2. To find the optimal q that effectively maps the observed data, model evaluation is performed. The models are compared based on the criterion: Marginal Likelihood and Mean Absolute Error (MAE).

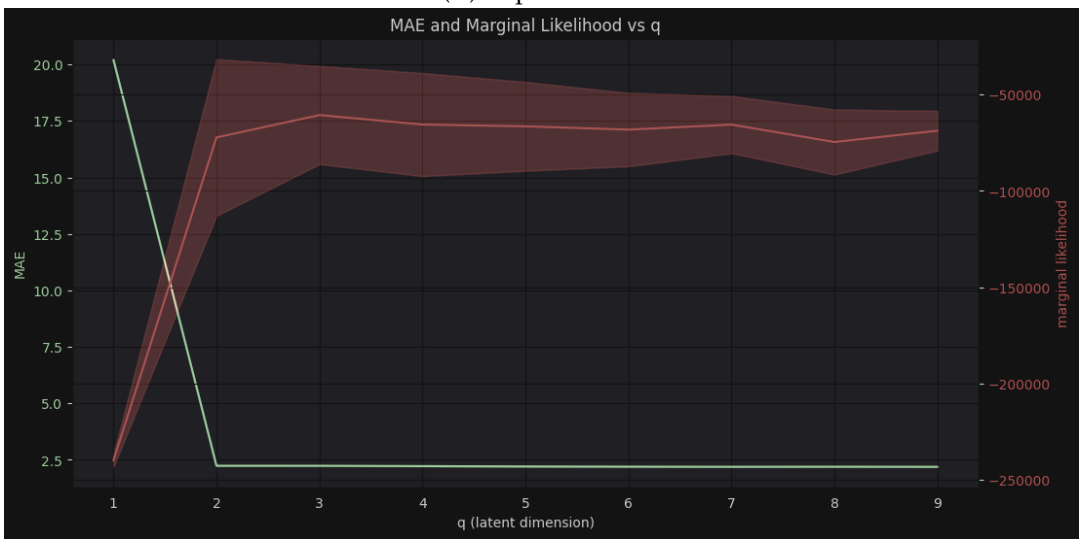
Marginal Likelihood is given by the eq. 3.17 and MAE is as follows:

$$\mathbf{MAE} = \frac{1}{dNKT} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{t=t_{n1}}^{T_n^{(K)}} |y_{nt}^{(k)} - y_{nt,\text{estimated}}^{(k)}|$$

The results of the 5−fold validation is shown in the Figure. 4.3



(A) Experiment 1



(B) Experiment 2

FIGURE 4.3: Model Evaluation with 5-Fold Validation. MAE in Green & Marginal Likelihood in Red (along with the band of standard deviation)

4.4 Discussion

From Figures. 4.3a, 4.3b, one can observe that, starting from $q = 1$ the Likelihood increases at a faster rate until q inches closer to the true latent dimension, which is 5 and 2 for the experiment 1 and 2, respectively. After this the rate of increase in marginal likelihood slows down indicating further increase in the latent dimension does not bring extra information on the observed data, suggesting redundancy or correlated dimensions. Therefore, from Figure. 4.3 it is clear that our model correctly recovered the dimension of the latent space. Similar analysis can be made for MAE as well. Figure

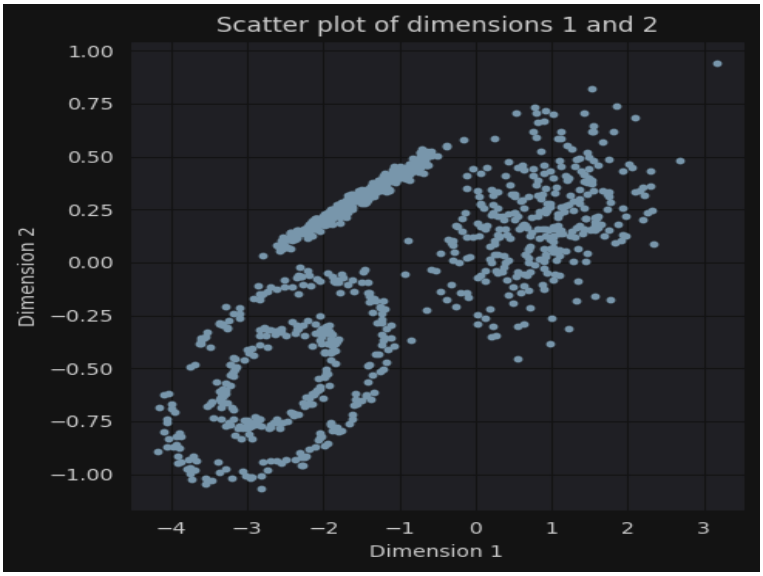
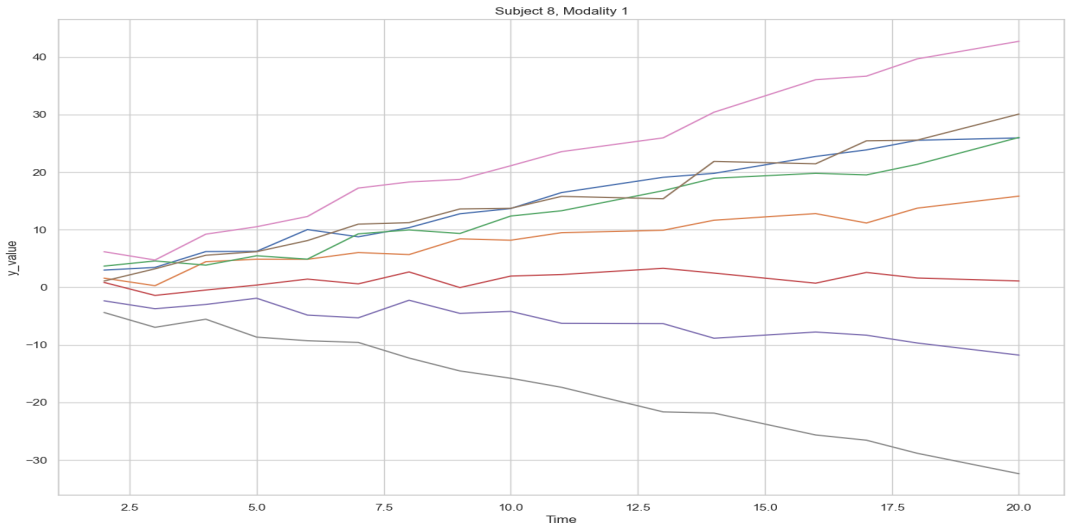
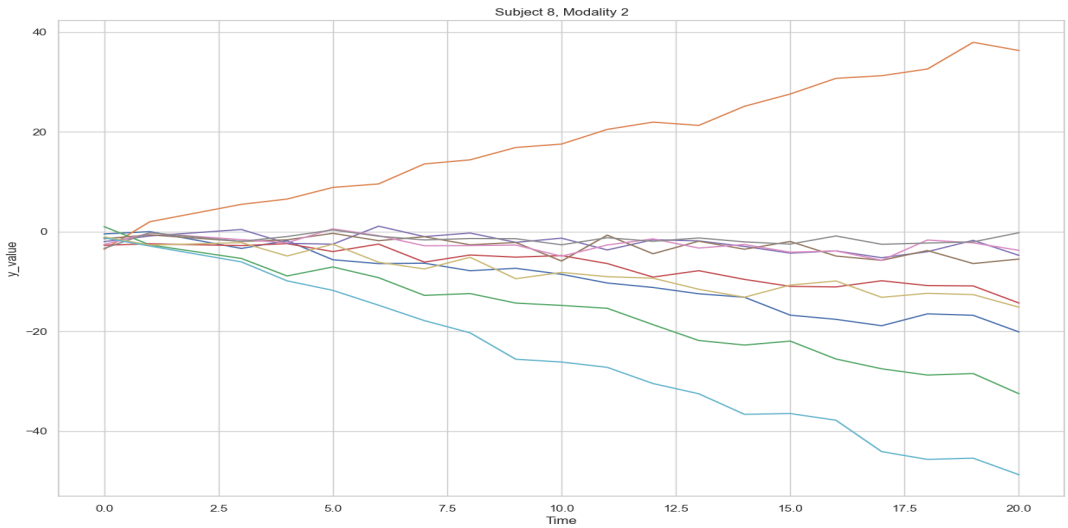


FIGURE 4.4: Experiment 2: Latent Space Reconstruction

4.4, shows the reconstructed latent space from the observed data (after training the model and learning the parameter), which resembles the the original latent space from Figure. 4.2b. This is interesting because we sampled the latent space with a combination of non-standard Gaussian distributions but still able to recover the latent space. This proves the robustness of the proposed latent variable model. This further proves a point that if there is a linear relationship between the observed and latent space then our model is capable of recovering the latent space without any prior knowledge on the latent space distribution. Further, the intra-patient variability, across the modalities, can be observed by comparing the evolution of the observed data within in Figures. 4.5 and 4.6. Additionally, the inter-patient variability can be observed, across the patients from 4.5 and 4.6, by comparing the evolution of the observed data. Figure. 4.7 highlights the latent representation of subjects 8 and 219 about the dimensions 1 and 2, which is one of the many views possible for these two subjects in the latent space. As you can see,



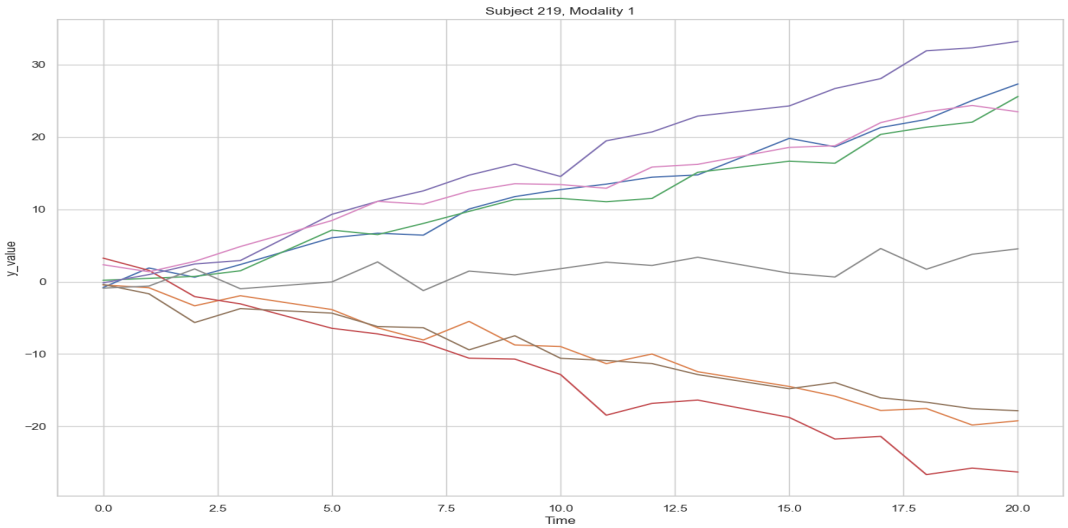
(A) Modality 1



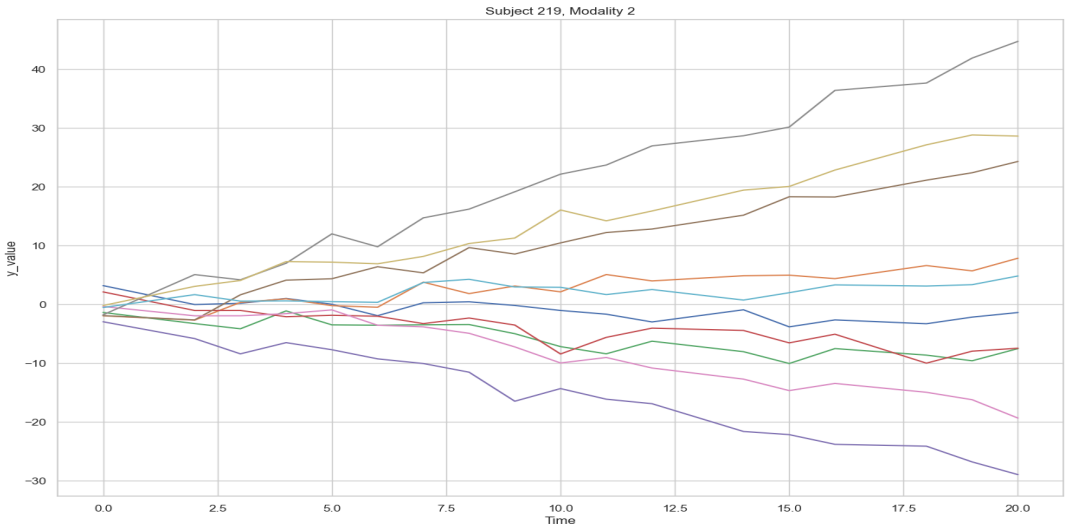
(B) Modality 2

FIGURE 4.5: Experiment 1: Evolution of the observed data for Subject 8 in both modalities. Each colour corresponds to a different dimension. We can observe the almost linear evolution in time, up to the additive noise.

the subjects are placed away in this view which can explain the inter-patient variability in some of the dimensions of the observed data.



(A) Modality 1



(B) Modality 2

FIGURE 4.6: Experiment 1: Evolution of the observed data for Subject 219 in both modalities. Each colour corresponds to a different dimension. We can observe the almost linear evolution in time, up to the additive noise.

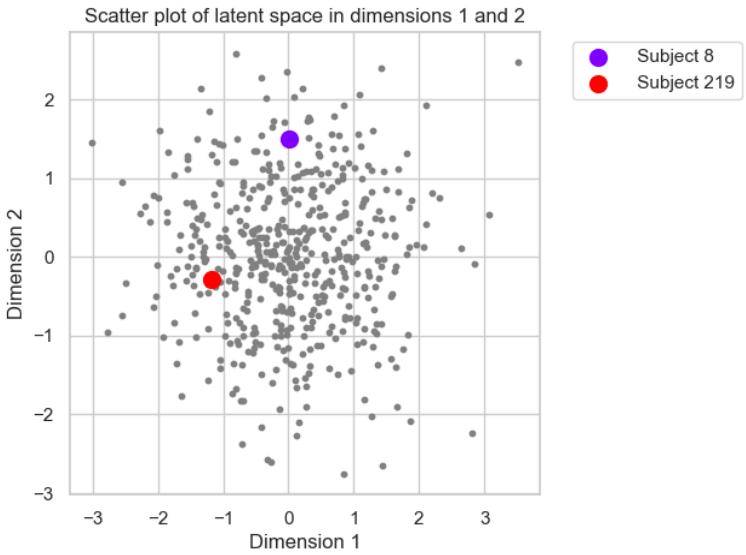


FIGURE 4.7: Experiment 1: Latent Space for Subject 8 and Subject 219

Chapter 5

Conclusions

5.1 Contributions

In this report, the nuances of the disease progression modeling (DPM) within the context of neurogenerative diseases (NDDs) are extensively studied and discussed. We successfully proposed a Bayesian Latent Variable Model based on Probabilistic Principal Component Analysis (PPCA) that successfully extended the work of Senacheribbe [23] to accommodate for multi-modal data while also adding more flexibilities to the model such as allowing for different time lengths across patients and across modalities. The model proposed is linear and hence it is easy to interpret, which is of paramount significance in the medical domain.

The model's effectiveness was confirmed through successful evaluation against synthetic datasets. This demonstrates the model's capability to accurately map the trajectory of disease progression, which can offer valuable insights in healthcare like diagnosis, drug development and therapeutic planning of diseases. Unlike traditional approaches in DPM, our model takes into account the inherent heterogeneity in the disease progression due to inter and intra patient variability. The latent space formulation of the model is crucial in disentangling the heavily correlated and highly dimensional data, while also accounting for the noise. Moreover, the model's robust generative capabilities significantly enhance its utility. Specifically, our model stands well-positioned as an instrumental tool for simulating a wide range of hypothetical scenarios in medical experiments, measuring uncertainties or for missing data imputation.

In conclusion, it is also important to remember that despite of the biomedical inspiration behind, the suggested model is universal in its application.

Its potential application extends to various kinds of datasets, not just those in the biomedical field.

5.2 Limitations

In spite of several advantages or capabilities of the model seen, the Bayesian latent variable model or the LVMs, in general, are based on very strong assumptions (3, Sec. 3.1) and in reality some of those assumptions might not hold. It is very likely that the linearity assumption between the observed and latent space is being used for interpretability reason and there is no evidence that this always exist. Most of the biological processes in nature are not linear. There is also no such guarantee that a latent space exists for all kinds of data or at least it might be difficult to accurately identify such as space. Most importantly, the latent space is assumed to be time independent, though this is reasonable to model if the patient data is not spread over long time spans (several years) but this does not stay valid for data coming from several years apart (Section 5.3). Additionally, LVMs can be sensitive to missing data, data missing not at random (MNAR) can induce bias in the estimates of the latent variables. Therefore, it is very necessary to consider all these limitations before applying the model to a given data.

5.3 Future Work & Perspectives

The authors in Senacheribbe [23] used a Latent Slope Intercept (LSI) model (on ADNI dataset), which is similar to our model but excluding multi-modal feature, to analyse the longitudinal data coming from two groups of sick and healthy patients. With LSI model, they managed to segregate the sick patients from the healthy patients in the latent space.

Now that we have successfully formulated and tested our model on synthetic dataset, the next phase of the internship is focused on applying our model to the longitudinal ADNI dataset similar to [23] along with multiple modality data from PET and MRI scans.

Possible Extensions

i. Global disease progression model

While our Bayesian latent variable model is constructed under the assumption that the observed data evolves linearly in time, this poses some limitations. This presumption is sensible when we monitor the patients within closely spaced follow-ups, but it falters when considering the disease's full trajectory across the time instants spread over years. Over a disease life cycle, a typical patient's condition might be stable in the beginning and then evolve linearly for a few years, during a mild stage, only to undergo a sharp acceleration as the disease progresses into severe stage [6, 7]. To tackle this, Senacheribbe [23] extended their LSI model as a global progression model by allowing the latent variables evolve with time. They used a common disease time variable 's' within the latent space, different from the actual time variable 't', which the patient experiences, and obtained an exponential disease progression model in 's', in latent space, by solving the resultant dynamical systems. This kind of a global disease progression model can also be explored from our model, as it is similar to LSI model w.r.t the latent dynamics.

ii. Extension to multi-centric studies

The Bayesian latent Variable Model that we have developed is a data-driven model that hinges on its ability to learn from an extensive clinical data. This data is vital to ensuring the model's robust performance. However, privacy regulations, such as the European General Data Protection Regulation (GDPR[1]), pose a serious challenge, prohibiting the sharing of raw data across different healthcare institutions and research centers. This scenario necessitates the exploration of novel learning approaches, such as Federated Learning (FL) [56], that can respect data privacy while effectively training our models on data dispersed across various locations.

[1] https://gdpr-info.eu/

Bibliography

[1] Alzheimer's Disease Neuroimaging Initiative. Study Design, 2023. URL https://adni.loni.usc.edu/study-design/#background-container. Accessed: July 23, 2023.

[2] Raphäel Couronné. *Progression models for Parkinson's Disease*. PhD thesis, Sorbonne Université, 2021.

[3] Wikipedia contributors. French institute for research in computer science and automation — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=French_Institute_for_Research_in_Computer_Science_and_Automation&oldid=1156181658. [Online; accessed 17-July-2023].

[4] Inria. Epione research group. URL https://team.inria.fr/epione/en/. [Online; accessed 17-July-2023].

[5] World Health Organization et al. Noncommunicable diseases country profiles 2011 who global report. 2011. *Back to cited text*, (4).

[6] Chengxuan Qiu, Miia Kivipelto, and Eva Von Strauss. Epidemiology of alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues in clinical neuroscience*, 2022.

[7] Gilberto Levy. The relationship of parkinson disease with aging. *Archives of neurology*, 64(9):1242–1246, 2007.

[8] Gabriella MacDougall, Logan Y Brown, Boris Kantor, and Ornit Chiba-Falek. The path to progress preclinical studies of age-related neurodegenerative diseases: A perspective on rodent and hipsc-derived models. *Molecular Therapy*, 29(3):949–972, 2021.

[9] US Food and Drug Administration. Critical path initiative, 2004. URL https://www.fda.gov/science-research/science-and-research-special-topics/critical-path-initiative. Accessed: 2023-07-21.

[10] Janet Woodcock and Raymond Woosley. The fda critical path initiative and its influence on new drug development. *Annu. Rev. Med.*, 59:1–12, 2008.

[11] RL Lalonde, KG Kowalski, MM Hutmacher, W Ewy, DJ Nichols, PA Milligan, BW Corrigan, PA Lockwood, SA Marshall, LJ Benincosa, et al. Model-based drug development. *Clinical Pharmacology & Therapeutics*, 82(1):21–32, 2007.

[12] DR Mould. Models for disease progression: new approaches and uses. *Clinical Pharmacology & Therapeutics*, 92(1):125–131, 2012.

[13] Kyle Strimbu and Jorge A Tavel. What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6):463, 2010.

[14] Jeffrey S Barrett, Tim Nicholas, Karim Azer, and Brian W Corrigan. Role of disease progression models in drug development. *Pharmaceutical Research*, 39(8):1803–1815, 2022.

[15] Wikipedia contributors. Electronic health record — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Electronic_health_record&oldid=1166686025. [Online; accessed 27-July-2023].

[16] ADNI. Alzheimer's disease neuroimaging initiative. URL https://adni.loni.usc.edu.

[17] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

[18] Patrick Heagerty. Longitudinal data analysis. URL https://faculty.washington.edu/heagerty/Courses/VA-longitudinal/private/LDAchapter.pdf.

[19] Xiang Wang, David Sontag, and Fei Wang. Unsupervised learning of disease progression models. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 85–94, 2014.

[20] Clément Abi Nader. *Modelling and simulating the progression of Alzheimer's disease through the analysis of multi-modal neuroimages and clinical data*. PhD thesis, Université Côte d'Azur, 2021.

[21] Creative Diagnostics. What are biomarkers?, 2017. URL https://www.creative-diagnostics.com/blog/index.php/what-are-biomarkers/. [Online; accessed 27-July-2023].

[22] Irene Balelli, Santiago Silva, and Marco Lorenzi. A differentially private probabilistic framework for modeling the variability across federated datasets of heterogeneous multi-view observations. *Machine Learning for Biomedical Imaging*, 1(IPMI 2021):1–36, apr 2022. doi: 10.59275/j.melba.2022-7175. URL https://doi.org/10.59275%2Fj.melba.2022-7175.

[23] Andrea Senacheribbe. Bayesian latent variable model for the analysis of the progression of alzheimer's disease, 2021. URL http://webthesis.biblio.polito.it/id/eprint/17992.

[24] Clifford R Jack Jr, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. Update on hypothetical model of alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207, 2013.

[25] Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sebastien Ourselin, Nick C Fox, et al. An event-based model for disease progression and its application in familial alzheimer's disease and huntington's disease. *NeuroImage*, 60(3):1880–1889, 2012.

[26] Alexandra L Young, Neil P Oxtoby, Pankaj Daga, David M Cash, Nick C Fox, Sebastien Ourselin, Jonathan M Schott, and Daniel C Alexander. A data-driven model of biomarker changes in sporadic alzheimer's disease. *Brain*, 137(9):2564–2577, 2014.

[27] Peter A Wijeratne, Alexandra L Young, Neil P Oxtoby, Razvan V Marinescu, Nicholas C Firth, Eileanoir B Johnson, Amrita Mohan, Cristina Sampaio, Rachael I Scahill, Sarah J Tabrizi, et al. An image-based model of brain volume biomarker changes in huntington's disease. *Annals of clinical and translational neurology*, 5(5):570–582, 2018.

[28] Neil P Oxtoby, Louise-Ann Leyland, Leon M Aksman, George EC Thomas, Emma L Bunting, Peter A Wijeratne, Alexandra L Young, Angelika Zarkali, Manuela MX Tan, Fion D Bremner, et al. Sequence of clinical and neurodegeneration events in parkinson's disease progression. *Brain*, 144(3):975–988, 2021.

[29] Bruno M Jedynak, Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T Wyman, David Raunig, C Pierre Jedynak, Brian Caffo, Jerry L Prince, et al. A computational neurodegenerative disease progression score: method and results with the alzheimer's disease neuroimaging initiative cohort. *Neuroimage*, 63(3):1478–1486, 2012.

[30] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Raman, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-François Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410, 2014.

[31] Ricardo Guerrero, Alexander Schmidt-Richberg, Christian Ledig, Tong Tong, Robin Wolz, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative (ADNI, et al. Instantiated mixed effects modeling of alzheimer's disease markers. *NeuroImage*, 142:113–125, 2016.

[32] Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, and Stanley Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. *Advances in neural information processing systems*, 28, 2015.

[33] Răzvan V Marinescu, Arman Eshaghi, Marco Lorenzi, Alexandra L Young, Neil P Oxtoby, Sara Garbarino, Sebastian J Crutch, Daniel C Alexander, Alzheimer's Disease Neuroimaging Initiative, et al. Dive: A spatiotemporal progression model of brain pathology in neurodegenerative disorders. *NeuroImage*, 192:166–177, 2019.

[34] Marc Lavielle. *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press, 2014.

[35] Marco Lorenzi, Maurizio Filippone, Giovanni B Frisoni, Daniel C Alexander, Sébastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in alzheimer's disease. *NeuroImage*, 190:56–68, 2019.

[36] Sara Garbarino, Marco Lorenzi, Alzheimer's Disease Neuroimaging Initiative, et al. Investigating hypotheses of neurodegeneration by learning dynamical systems of protein propagation in the brain. *Neuroimage*, 235: 117980, 2021.

[37] Marco Lorenzi and Maurizio Filippone. Constraining the dynamics of deep probabilistic models. In *International Conference on Machine Learning*, pages 3227–3236. PMLR, 2018.

[38] Alexandra L Young, Razvan V Marinescu, Neil P Oxtoby, Martina Bocchetta, Keir Yong, Nicholas C Firth, David M Cash, David L Thomas, Katrina M Dick, Jorge Cardoso, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nature communications*, 9(1):4273, 2018.

[39] Igor Koval, J-B Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, Stanley Durrleman, and Alzheimer's Disease Neuroimaging Initiative. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pages 451–459. Springer, 2017.

[40] Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stephanie Allassonniere, and Stanley Durrleman. Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. *Frontiers in neurology*, page 235, 2018.

[41] Clément Abi Nader, Nicholas Ayache, Giovanni B Frisoni, Philippe Robert, Marco Lorenzi, and Alzheimer's Disease Neuroimaging Initiative. Simulating the outcome of amyloid treatments in alzheimer's disease from imaging and clinical data. *Brain communications*, 3(2):fcab091, 2021.

[42] Christine A Cooper and Lama M Chahine. Biomarkers in prodromal parkinson disease: a qualitative review. *Journal of the International Neuropsychological Society*, 22(10):956–967, 2016.

[43] Martin Knott and David J Bartholomew. *Latent variable models and factor analysis*, volume 7. Edward Arnold, 1999.

[44] John P Cunningham and Zoubin Ghahramani. Linear dimensionality reduction: Survey, insights, and generalizations. *The Journal of Machine Learning Research*, 16(1):2859–2900, 2015.

[45] IT Jolliffe. *Principal Component Analysis*. Springer, 1986.

[46] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[47] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[48] Wikipedia contributors. Factor analysis — Wikipedia, the free encyclopedia, 2023. URL https://en.wikipedia.org/w/index.php?title=Factor_analysis&oldid=1160800553. [Online; accessed 29-July-2023].

[49] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[50] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[51] Luigi Antelmi, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data. In *International Conference on Machine Learning*, pages 302–311. PMLR, 2019.

[52] Ricard Argelaguet, Britta Velten, Damien Arnol, Sascha Dietrich, Thorsten Zenz, John C Marioni, Florian Buettner, Wolfgang Huber, and Oliver Stegle. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular systems biology*, 14 (6):e8124, 2018.

[53] Numpy. URL https://numpy.org/. Accessed: 2023-07-31.

[54] Matplotlib. URL https://matplotlib.org/. Accessed: 2023-07-31.

[55] Seaborn. URL https://seaborn.pydata.org/. Accessed: 2023-07-31.

[56] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

Appendix A

Additional details: Bayesian Latent Variable Model Formulation

A.1 Proof of Posterior Distribution

$$\begin{aligned} p(x_n|y_{nt_{n1}},\ldots,y_{nt_{nT_n}}) &= \frac{p(y_{nt_{n1}},\ldots,y_{nt_{nT_n}}|x_n)p(x_n)}{p(y_{nt_{n1}},\ldots,y_{nt_{nT_n}})} \\ &\propto p(y_{nt_{n1}},\ldots,y_{nt_{nT_n}}|x_n)p(x_n) \end{aligned} \tag{A.1}$$

$$\begin{aligned} x_n|y_{nt_{n1}},\ldots,y_{nt_{nT_n}} &\sim exp\{-\frac{1}{2}(x_n-m)^T\Sigma^{-1}(x_n-m)\} \\ &\sim exp\{-\frac{1}{2}{x_n}^T\Sigma^{-1}x_n - \frac{1}{2}m^T\Sigma^{-1}m + x_n^T\Sigma^{-1}m\} \\ &\sim exp\{-\frac{1}{2}x_n^T\Sigma^{-1}x_n + x_n^T\Sigma^{-1}m\} \end{aligned} \tag{A.2}$$

$$\begin{aligned} p(y_{nt_{n1}},\ldots,y_{nt_{nT_n}}|x_n) &= \prod_{t=t_{n1}}^{T_n} p(y_{nt}|x_n) \\ &= \prod_{t=t_{n1}}^{T_n}\prod_{k=1}^{K} p(y_{nt}^{(k)}|x_n) \end{aligned} \tag{A.3}$$

$$y_{nt}^{(k)}|x_n \sim \mathcal{N}((tW^{(k)}+V^{(k)})x_n + t\omega^{(k)} + \mu^{(k)}, {\sigma^{(k)}}^2\mathbb{I}_{d_k})$$

$$p(y_{nt_{n1}},\ldots,y_{nt_{nT_n}}|x_n)p(x_n)$$

$$\sim exp\Big\{ \sum_{t=t_{n1}}^{T_n} \sum_{k=1}^{K} -\frac{1}{2}\Big(y_{nt}^{(k)} - \big((tW^{(k)} + V^{(k)})x_n + t\omega^{(k)} + \mu^{(k)}\big)\Big)^T (\sigma^{(k)^2}\mathbb{I}_{d_k})^{-1}\Big(y_{nt}^{(k)} - \big((tW^{(k)} + V^{(k)})x_n + t\omega^{(k)} + \mu^{(k)}\big)\Big) - \frac{1}{2}x_n^T x_n \Big\}$$

$$\sim exp\Big\{ -\frac{1}{2}\sum_{t=t_{n1}}^{T_n} \sum_{k=1}^{K} \Big(\frac{1}{\sigma^{(k)^2}}\big(x_n^T(tW^{(k)} + V^{(k)})^T(tW^{(k)} + V^{(k)})x_n - 2x_n^T(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})\big)\Big) - \frac{1}{2}x_n^T x_n \Big\}$$

$$\sim exp\Big\{ -\frac{1}{2}x_n^T\Big(\sum_{k=1}^{K}\frac{1}{\sigma^{(k)^2}}\sum_{t=t_{n1}}^{T_n}(tW^{(k)} + V^{(k)})^T(tW^{(k)} + V^{(k)}) + \mathbb{I}_q\Big)x_n + x_n^T\Big(\sum_{k=1}^{K}\frac{1}{\sigma^{(k)^2}}\sum_{t=t_{n1}}^{T_n}(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})\Big)\Big\}$$

$$\Sigma = \Big(\sum_{k=1}^{K}\frac{1}{\sigma^{(k)^2}}\sum_{t=t_{n1}}^{T_n}(tW^{(k)} + V^{(k)})^T(tW^{(k)} + V^{(k)}) + \mathbb{I}_q\Big)^{-1}$$

$$\Sigma^{-1}m = \sum_{k=1}^{K}\frac{1}{\sigma^{(k)^2}}\sum_{t=t_{n1}}^{T_n}(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})$$

$$m = \Sigma\sum_{k=1}^{K}\frac{1}{\sigma^{(k)^2}}\sum_{t=t_{n1}}^{T_n}(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})$$

A.2 Complete Data Log-likelihood

$$\begin{aligned}\text{Likelihood} = p(x_1, y_1, \ldots, x_N, y_N) &= \prod_{n=1}^{N} p(x_n, y_n) = \prod_{n=1}^{N} p(y_n|x_n) * p(x_n) \\ &= \prod_{n=1}^{N} p(y_{nt_{n1},\ldots,y_{nt_{nT_n}}}|x_n) * p(x_n) = \prod_{n=1}^{N}\Big[\prod_{t=t_{n1}}^{T_n} p(y_{nt}|x_n)\Big] * p(x_n)\end{aligned} \tag{A.4}$$

$$\begin{aligned}\mathcal{L} = ln(\text{Likelihood}) &= \sum_{n=1}^{N}\Big[\sum_{t=t_{n1}}^{T_n} ln\big(p(y_{nt}|x_n)\big) + ln\big(p(x_n)\big)\Big] \\ &= \mathcal{L}(x, y; W, V, \omega, \mu, \sigma^{(1)}, \ldots, \sigma^{(K)})\end{aligned} \tag{A.5}$$

$$\begin{aligned}\mathcal{L} = -\sum_{n=1}^{N}\Bigg[\Bigg\{\sum_{k=1}^{K}\sum_{t=t_{n1}}^{T_n^{(k)}}\Bigg(&\frac{d_k}{2}ln(2\pi) + \frac{d_k}{2}ln({\sigma^{(k)}}^2) + \frac{1}{2{\sigma^{(k)}}^2}(||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 \\ &+ x_n^T(tW^{(k)} + V^{(k)})^T(tW^{(k)} + V^{(k)})x_n \\ &- 2x_n^T(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}))\Bigg)\Bigg\} + \frac{1}{2}||x_n||^2\Bigg]\end{aligned} \tag{A.6}$$

A.3 Optimization using EM Algorithm

0-Step

Initialize the parameters: W,V, ω, μ, $\sigma^{(1)}, \ldots, \sigma^{(K)}$

***E*-Step**

Evaluate:

$$\mathbb{E}_{x|y,W,V,\omega,\mu,\sigma^{(1)},\ldots,\sigma^{(K)}}(\mathcal{L})$$

$$\begin{aligned}\mathbb{E}[\mathcal{L}] = \langle\mathcal{L}\rangle = -\sum_{n=1}^{N}\Bigg[\Bigg\{\sum_{k=1}^{K}\sum_{t=t_{n1}}^{T_n^{(k)}}\Bigg(&\frac{d_k}{2}ln(2\pi) + \frac{d_k}{2}ln({\sigma^{(k)}}^2) + \\ &\frac{1}{2{\sigma^{(k)}}^2}(||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T(tW^{(k)} + V^{(k)})\langle x_n {x_n}^T\rangle\big) \\ &- 2\langle x_n\rangle^T(tW^{(k)} + V^{(k)})^T(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}))\Bigg)\Bigg\} + \frac{1}{2}Tr(\langle x_n {x_n}^T\rangle)\Bigg]\end{aligned} \tag{A.7}$$

M-**Step**

Maximize the expectation by searching through the parameter space of:

$$W, V, \omega, \mu, \sigma^{(1)}, \ldots, \sigma^{(K)}$$

Estimation of $\widetilde{\mu} : \widetilde{\mu}^{(1)}, \ldots, \widetilde{\mu}^{(K)}$ For each $k : k \in \{1, \ldots, K\}$

$$\langle \mathcal{L} \rangle_{\mu^{(k)}} = -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big({\mu^{(k)}}^T \mu^{(k)} - 2(y_{nt}^{(k)} - t\omega^{(k)})^T \mu^{(k)} + 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T \mu^{(k)} \big)$$

$$\left. \frac{\partial \langle \mathcal{L} \rangle_{\mu^{(k)}}}{\partial \mu^{(k)}} \right|_{\widetilde{\mu}^{(k)}} = 0 \implies -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(2\widetilde{\mu}^{(k)} - 2(y_{nt}^{(k)} - t\omega^{(k)}) + 2(tW^{(k)} + V^{(k)}) \langle x_n \rangle \big) = 0$$

$$\implies \widetilde{\mu}^{(k)} = \frac{1}{\sum_{n=1}^{N} T_n} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} (y_{nt}^{(k)} - (tW^{(k)} + V^{(k)}) \langle x_n \rangle - t\omega^{(k)})$$

Estimation of $\widetilde{\omega} : \widetilde{\omega}^{(1)}, \ldots, \widetilde{\omega}^{(K)}$ For each $k : k \in \{1, \ldots, K\}$

$$\langle \mathcal{L} \rangle_{\omega^{(k)}} = -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(t^2 {\omega^{(k)}}^T \omega^{(k)} - 2t(y_{nt}^{(k)} - \mu^{(k)})^T \omega^{(k)} + 2t\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T \omega^{(k)} \big)$$

$$\left. \frac{\partial \langle \mathcal{L} \rangle_{\omega^{(k)}}}{\partial \omega^{(k)}} \right|_{\widetilde{\omega}^{(k)}} = 0 \implies -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(2t^2 \widetilde{\omega}^{(k)} - 2t(y_{nt}^{(k)} - \mu^{(k)}) + 2t(tW^{(k)} + V^{(k)}) \langle x_n \rangle \big) = 0$$

$$\implies \widetilde{\omega}^{(k)} = \frac{1}{\sum_{n=1}^{N} \tau_n^{(k)}} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} t(y_{nt}^{(k)} - (tW^{(k)} + V^{(k)})\langle x_n \rangle - \mu^{(k)})$$

$\tau_n^{(k)} = \sum_{t=t_{n1}}^{T_n^{(k)}} t^2$ Estimation of $\widetilde{W} : \widetilde{W}^{(1)}, \ldots, \widetilde{W}^{(K)}$ For each $k : k \in \{1, \ldots, K\}$

$$\begin{aligned}\langle \mathcal{L} \rangle_{W^{(k)}} = -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(& t^2 Tr({W^{(k)}}^T W^{(k)} \langle x_n {x_n}^T \rangle) + tTr({W^{(k)}}^T V^{(k)} \langle x_n {x_n}^T \rangle) \\ & + tTr({V^{(k)}}^T W^{(k)} \langle x_n {x_n}^T \rangle) - 2t \langle x_n \rangle^T {W^{(k)}}^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \big)\end{aligned}$$

$$\begin{aligned}\left. \frac{\partial \langle \mathcal{L} \rangle_{W^{(k)}}}{\partial W^{(k)}} \right|_{\widetilde{W}^{(k)}} = 0 \implies -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(& 2t^2 \widetilde{W}^{(k)} \langle x_n {x_n}^T \rangle + 2tV^{(k)} \langle x_n {x_n}^T \rangle \\ & - 2t(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T \big) = 0\end{aligned}$$

$$\implies \widetilde{W}^{(k)} = \left(\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} t(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T - tV^{(k)} \langle x_n {x_n}^T \rangle \right) \left(\sum_{n=1}^{N} \tau_n^{(k)} \langle x_n {x_n}^T \rangle \right)^{-1}$$

$\tau_n^{(k)} = \sum_{t=t_{n1}}^{T_n^{(k)}} t^2$ Estimation of $\widetilde{V} : \widetilde{V}^{(1)}, \ldots, \widetilde{V}^{(K)}$ For each $k : k \in \{1, \ldots, K\}$

$$\begin{aligned}\langle \mathcal{L} \rangle_{V^{(k)}} = -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(& Tr({V^{(k)}}^T V^{(k)} \langle x_n {x_n}^T \rangle) + tTr({W^{(k)}}^T V^{(k)} \langle x_n {x_n}^T \rangle) \\ & + tTr({V^{(k)}}^T W^{(k)} \langle x_n {x_n}^T \rangle) - 2 \langle x_n \rangle^T {V^{(k)}}^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \big)\end{aligned}$$

$$\begin{aligned}\left. \frac{\partial \langle \mathcal{L} \rangle_{V^{(k)}}}{\partial V^{(k)}} \right|_{\widetilde{V}^{(k)}} = 0 \implies -\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{1}{2{\sigma^{(k)}}^2} \big(& 2\widetilde{V}^{(k)} \langle x_n {x_n}^T \rangle + 2tW^{(k)} \langle x_n {x_n}^T \rangle \\ & - 2(y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T \big) = 0\end{aligned}$$

$$\implies \widetilde{V}^{(k)} = \Bigg(\sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \langle x_n \rangle^T - tW^{(k)} \langle x_n {x_n}^T \rangle \Bigg) \Bigg(\sum_{n=1}^{N} T_n^{(k)} \langle x_n {x_n}^T \rangle \Bigg)^{-1}$$

Estimation of $\widetilde{\sigma^2} : \widetilde{\sigma^2}^{(1)}, \ldots, \widetilde{\sigma^2}^{(K)}$ For each $k : k \in \{1, \ldots, K\}$

$$\langle \mathcal{L} \rangle_{\sigma^{2(k)}} = - \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \Bigg(\frac{d_k}{2} ln(\sigma^{(k)^2}) + \frac{1}{2\sigma^{(k)^2}} (||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) \langle x_n {x_n}^T \rangle\big) - 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)})) \Bigg)$$

$$\frac{\partial \langle \mathcal{L} \rangle_{\sigma^{2(k)}}}{\partial \sigma^{2(k)}} \Bigg|_{\widetilde{\sigma^{2(k)}}} = 0 \implies - \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \frac{d_k}{2\widetilde{\sigma^2}^{(k)}} - \frac{1}{2(\widetilde{\sigma^2}^{(k)})^2} \Bigg(||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) \langle x_n {x_n}^T \rangle\big) - 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \Bigg) = 0$$

$$\implies \widetilde{\sigma^2}^{(k)} = \frac{1}{d_k \sum_{n=1}^{N} T_n^{(k)}} \sum_{n=1}^{N} \sum_{t=t_{n1}}^{T_n^{(k)}} \Bigg(||y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}||^2 + Tr\big((tW^{(k)} + V^{(k)})^T (tW^{(k)} + V^{(k)}) \langle x_n {x_n}^T \rangle\big) - 2\langle x_n \rangle^T (tW^{(k)} + V^{(k)})^T (y_{nt}^{(k)} - t\omega^{(k)} - \mu^{(k)}) \Bigg)$$