Robust parameter estimation of finite mixture models with self-paced learning

by

André Ruben Kleynhans

Submitted in partial fulfillment of the requirements for the degree
Masters of Science (eScience)
in the Faculty of Natural and Agricultural Science
University of Pretoria, Pretoria

November 2022

Robust parameter estimation of finite mixture models with self-paced learning

by

André Ruben Kleynhans
E-mail: kleynhans.andre@up.ac.za

Abstract

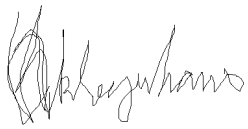
Self-paced learning (SPL) is a training strategy that mitigates the impact of non-typical observations. SPL introduces observations in a meaningful order by considering the likelihood for each observation. The proposed algorithm considers a finite mixture model that includes a distributional structure for non-typical observations in the SPL weight calculation. Two new self-paced learning (SPL) algorithms is proposed for finite mixture models (FMM). This includes self-paced component learning FMMs and a self-paced learning algorithm that includes a distributional structure for non-typical observations. The properties of these algorithms are presented through a simulation study along with an application on real data. A comparison is made with the properties of well known models. The algorithms shows a reduction in parameter estimation bias which indicates an improvement in the estimation accuracy of the parameters.

Keywords: Self-paced learning, finite-mixture models, Gaussian mixture models.

Supervisors : Prof. F.H.J. Kanfer
Prof. S.M. Millard
Department : Department of Statistics
Degree : Master of Science eScience

Decleration

I, Andre Ruben Kleynhans, declare that the dissertation, which I hereby submit for the degree MSc eScience at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution



07/12/2022

Signature

Date

Acknowledgements

I would like to express my gratitude towards the following people and isntitutions:

- The support of the DSI-NICIS National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NEPTTP.
- To thank my supervisors, Prof. Kanfer and Prof. Millard. It was an exciting year full of helpful and awesome sessions. Thank you for the guiding role that you played in my research.
- To my parents for supporting me during this eventful year. Thank you for believing in me and providing me with many words of encouragement. My gratitude can never be expressed enough.
- To my brother, your work and academic career have inspired me tremendously.
- To my girlfriend, thank you for always supporting me and listening to my ramblings about my work. You are truly inspirational and a good role model.
- To my girlfriend's parents for hosting me and my computer during rough loads shedding schedules. Thank you for providing me with a home away from home to work on my studies.

List of Aronyms

CNM	Mixture of contaminated normal distribution
EM	Expectation maximisation
FMM	Finite mixture model
GMM	Gaussian mixture model
MLE	Maximum likelihood estimation
MSE	Mean squared error
MT	Mixture of Student's t distribution
pdf	Probability density function
pmf	Probability mass function
SPFMM	Self-paced finite mixture model
SPCFMM	Self-paced components finite mixture model
SPL	Self-paced learning
RSPGMM	Robust self-paced Gaussian mixture model
SPCGMM	Self-paced component Gaussian mixture model
SPGMM	Self-paced Gaussian mixture model

Contents

List of Figures

List of Algorithms

List of Tables

Chapter 1

Introduction

1.1 Motivation

Finite mixture models (FMM) is a well known probabilistic clustering model [28]. The FMM is a linear combination of probability density functions (pdf) or probability mass functions (pmf) [2]. FMM can uncover hidden structures or groups by simultaneously fitting pdfs (or pmfs) over clusters formed in a set of observations [19].

FMM is a tractable method of modelling random phenomena and as such has received a lot of attention [19]. The parameters of a FMM is estimated using the expectation maximisation (EM) algorithm [15]. The EM algorithm finds the set of parameters that best fits the observations. The EM algorithm maximises the so called complete data likelihood function [7].

The complete-data likelihood function of a FMM tends to have multiple local optima [9]. This can cause major differences in parameter estimation for different initial values [25]. Unfortunately the accuracy of parameter estimation can also be negatively influenced by non-typical observations such as mild and gross outliers [1] [22].

[1]Mild outliers are non-typical observations generated from a different model or a generated far from the assumed distribution [22]. Gross outliers are non-typical observations that can not be modelled by some distribution as they occur unpredictably.

It is a well known fact that data sets often contain non-typical observations [13]. Many research is dedicated to mitigate the impact of these influential observations [22]. We refer to such parameter estimation as robust methods.

If one consider a mixture of heavier tailed distributions one can account for mild outliers [15, 22]. An example of these models includes the mixture of Student's t distribution (MT) and the mixture of Laplace distribution [28, 27]. Unfortunately this approach can not mitigate the influence of observations that are abnormally far away from the mixture distributions [22].

One can soften the influence of these observations by clustering them together [2]. One can do this by considering additional components in the FMM [22]. Observations that are not likely to fit from any component of the FMM is assigned to the additional components [8]. An example of this type of FMM includes the contaminated normal mixture model (CNM).

Self-paced learning (SPL) is a parameter estimation algorithm that has empirically been shown to metigate the influential observations such as gross-outliers [30]. SPL was first introduced in supervised learning problems. The SPL algorithm changes the order in which training observations is introduced to the learning process [10]. It was shown that using SPL with a MT results in overall better clustering [28].

This mini-dissertation will consider the application of the SPL algorithm for FMM parameter estimation. We extend the SPL algorithm for FMM parameter estimation by creating an algorithm that applies SPL on each FMM component individually. We further improve the SPL algorithm by considering a distributional framework for non-typical data.

The properties of this algorithm are presented through a simulation study along with an application on real data. A comparison is made with the properties of frequently used

models. The proposed algorithms show a reduction in parameter estimation bias and generally lower MSE than the other frequently used models.

1.2 Objectives

The overarching aim of this study is to consider and improve the application of SPL on FMMs.

The research objectives are listed below:

- Consider the SPL algorithm and its application on FMM for better parameter estimation.
- Propose a SPL algorithm that applies SPL on each FMM component individually.
- Create a SPL algorithm that uses a distribution for non-typical observations in the SPL weight calculations.
- Compare the proposed SPL algorithm to other parameter estimation methods.
- Apply the proposed robust SPL algorithm to a real data set.

1.3 Contributions

This research will contribute the following:

1. Introducing an algorithm that applies SPL to each component of the FMM.
2. Provides a novel extension to the SPL algorithm by considering a FMM with a distributional framework for non-typical observations and by using it in the SPL weight calculation.
3. Determine and compare the properties of the SPL FMM models to that of other robust FMM models.

1.4 Outline

This mini-dissertation will have the following structure

- **Chapter 2** provide necessary background theory.
- **Chapter 3** give a SPL algorithm that applies SPL on each individual component of the FMM.
- **Chapter 4** propose a robust SPL algorithm that considers the distribution of non-typical observations..
- **Chapter 5** consider a simulation study where we demonstrate the properties of SPL in the FMM environment and compare it to frequently used FMM's.
- **Chapter 6** shows the performance of SPL algorithms on a real data set and studies the model sensitivity.
- **Chapter 7** gives a conclusion to our findings.
- **Appendix A** provide the derivations made in this mini-dissertation.
- **Appendix B** gives the entire simulation study results.

Chapter 2

Background Theory

This chapter briefly describes the required background theory. This chapter can be skipped if the reader wishes to do so. The notation and conventions used in this research are laid out in this chapter.

2.1 Finite Mixture Models

The FMM is a practical tool for exposing hidden structures in data [6, 18, 19]. The FMM is at its core a linear superposition of densities [18]. This allows a convenient way to model complex distributional relationships using simpler densities [6, 18]. This framework also allows the FMM to address clustering problems [6].

2.1.1 Model formulation

The distribution of a random vector $\mathbf{X} \in \mathbb{R}^p$ with a k-component FMM can be written as

$$P(\mathbf{x}|\mathbf{\Psi}) = \sum_{j=1}^{k} \pi_j \cdot f_j(\mathbf{x}|\mathbf{\Theta}_j), \tag{2.1}$$

where π_j is the mixing proportion of the jth component, with $\pi_j > 0$ and $\sum_{j=1}^{k} \pi_j = 1$,
$f_j(\mathbf{x}|\boldsymbol{\Theta}_j)$ is the density of the jth component with parameters $\boldsymbol{\Theta}_j$,
$\boldsymbol{\Psi} = \{\boldsymbol{\Pi}, \boldsymbol{\Theta}\}$, is a set of that contains the FMM parameters,
with $\boldsymbol{\Pi} = \{\pi_j\}_{j=1}^{k}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_j\}_{j=1}^{k}$.

2.1.2 Model-based clustering

Let $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^{n}$ be a p-dimensional random sample of size n from model (2.1). Assuming cluster membership is unknown let the latent indicator variables $\mathbf{z} = \{\mathbf{z}_i\}_{i=1}^{n}$ denote cluster memberships for observation $\mathbf{x}_i$, where $\mathbf{z}_i = (z_{i1}, ..., z_{ik})^T \in \mathbb{R}^k$ and

$$z_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ belongs to cluster } j \\ 0 & \text{otherwise,} \end{cases}$$

and $P(z_{ij} = 1|\boldsymbol{\Psi}) = \pi_j$.

The complete data are

$$(\mathbf{x}, \mathbf{z}) = (\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{z}_1, ..., \mathbf{z}_n)$$

The joint distribution of $\mathbf{x}_i$ and $\mathbf{z}_i$ is

$$\begin{aligned} P(\mathbf{x}, \mathbf{z}|\boldsymbol{\Psi}) &= P(\mathbf{x}|\mathbf{z}, \boldsymbol{\Psi}) \cdot P(\mathbf{z}|\boldsymbol{\Psi}) \\ &= \prod_{i=1}^{n} P(\mathbf{x}_i|\mathbf{z}_i, \boldsymbol{\Psi}) \cdot P(\mathbf{z}_i|\boldsymbol{\Psi}). \end{aligned}$$

Since $P(\mathbf{x}_i|\mathbf{z}_i, \boldsymbol{\Psi}) = \prod_{j=1}^{k} f_j(\mathbf{x}|\boldsymbol{\Theta}_j)^{\pi_j}$ and $P(\mathbf{z}_i|\boldsymbol{\Psi}) = \prod_{j=1}^{k} \pi_j^{z_{ij}}$ we have

$$P(\mathbf{x}, \mathbf{z}|\boldsymbol{\Psi}) = \prod_{i=1}^{n} \prod_{j=1}^{k} \pi_j^{z_{ij}} f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j)^{z_{ij}}. \tag{2.2}$$

The complete data log-likelihood is

$$l(\boldsymbol{\Psi}) = \sum_{i=1}^{n} \sum_{j=1}^{k} \Big\{ z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j) \Big\}, \tag{2.3}$$

Equation (2.3) can be decomposed as the summation of log-likelihood of each observation in $\mathbf{x}$, denoted by $l_i(\mathbf{\Psi}) = \sum_{j=1}^{k} z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\mathbf{\Theta}_j)$ [28]. We see this in equation (2.4). In equation (2.5) each $l_i(\mathbf{\Psi})$ is decomposed as the summation of the log-likelihood for each component of the FMM, represented by $l_{ij}(\Psi_j) = z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\mathbf{\Theta}_j)$ with component parameters $\mathbf{\Psi}_j = \{\pi_j, \mathbf{\Theta}_j\}$.

$$l(\mathbf{\Psi}) = \sum_{i=1}^{n} l_i(\mathbf{\Psi}) \tag{2.4}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} l_{ij}(\mathbf{\Psi}_j) \tag{2.5}$$

2.1.3 Maximum likelihood estimation

The maximum likelihood estimator (MLE) is the set of parameters $\hat{\mathbf{\Psi}}$ such that

$$\hat{\mathbf{\Psi}} = \arg\max_{\mathbf{\Psi}} \ l(\mathbf{\Psi}).$$

FMM use the EM algorithm to calculate MLE estimators [4]. The EM algorithm guarantees convergence to some local optima but it does not necessarily converge to global optima [7, 9].

EM-algorithm formulation

The EM algorithm iterates between the so-called Expectation (E) Step and Maximisation (M) step [14]. This is repeated until the model parameters converged [17].

The EM algorithm is performed as follows

- E-step:
  The E-step finds the expected value of the latent variables. The expected value is used to create the function $Q(\mathbf{\Psi}|\mathbf{\Psi}^{(r)})$ used for maximisation

  $$Q(\mathbf{\Psi}|\mathbf{\Psi}^{(r)}) = \mathbb{E}_{\mathbf{z}|\boldsymbol{x},\mathbf{\Psi}^{(r)}}(l(\mathbf{\Psi}|\mathbf{x},\mathbf{z}))$$

  where $\mathbf{\Psi}^{(r)}$ is the estimated parameter values of the rth iteration.

- M-Step:
  The M-step consist of calculating

$$\mathbf{\Psi}^{(r+1)} = \max_{\mathbf{\Psi}} \left\{ Q(\mathbf{\Psi}|\mathbf{\Psi}^{(r)}) \right\}.$$

2.2 Self-Paced Learning

SPL changes the order in which observations are introduced in the parameter estimation algorithm [10]. SPL has been shown to mitigate the impact of non-typical observations and trained better models in both supervised and unsupervised learning problems [3, 12].

SPL is based on a real human learning strategy. People tend to learn from easier examples and then gradually increase the difficulty of the problem [10]. For example when one learns to build puzzles, one starts off with smaller easier puzzles and then gradually increase the complexity of the puzzles one builds.

SPL uses the target model's (the model being trained) objective function to dictate how difficult an observation is [10]. SPL introduces a set of the easiest observations first and then gradually introduce more difficult observations to the model training algorithm. SPL models assigns a weight for each observation to indicate how difficult an observation is.

By referring back to our puzzle building analogy, the more time one spends on building puzzles, the more likely one is to build more difficult puzzles to increase the challenge. SPL uses a so called age parameter, denoted by λ, to control how much difficult observations is introduced [16]. One might think of it as a way to measure how much time is spent on a set of observations. More difficult observations is introduced to the learning algorithm as the age parameter increases.

The rate at which the age parameter is increased is called the learning rate. If the learning rate is high, then the SPL model quickly uses all the observations, if it is lower,

then the SPL algorithm will take longer to include all te observations [16].

2.2.1 Self-paced learning formulation

Consider a set of observations $\{\mathbf{x}_i\}_{i=1}^n$ with associated labels $\{y_i\}_{i=1}^n$. Let $p(\mathbf{x}_i, \mathbf{\Theta})$ be the target model estimates with model parameters $\mathbf{\Theta}$. Let $L(y_i, p(\mathbf{x}_i, \mathbf{\Theta}))$ represent the loss function that calculates the cost of label y_i and estimated label $p(\mathbf{x}_i, \mathbf{\Theta})$.

SPL consists of a weighted loss term with SPL weights $\mathbf{v} = (v_1, ..., v_n)^T$ and a self-paced regulariser (SP-regulariser) $g(\mathbf{v}, \lambda)$ imposed on sample weights expressed as

$$E(\mathbf{v}, \mathbf{\Theta}|\lambda) = \sum_{i=1}^{n} v_i \cdot L(y_i, p(\mathbf{x}_i, \mathbf{\Theta})) + g(\mathbf{v}, \lambda), \tag{2.6}$$

where λ is the age parameter that controls the amount of difficult of observations included in the training process.

The SP-regulariser plays a crucial role in the SPL process [30]. An axiomatic definition of the SP-regulariser was proposed by Meng et al. [16].

Definition 1. *Let v be a SPL weight variable, let L be a loss value and λ the age parameter. Suppose $v^*(\lambda, L)$ denotes the optimal v such that $v^*(\lambda, L) = \arg\min_v v \cdot L + g(v, \lambda)$, for $v \in [0, 1]$ and $L(\mathbf{y}_i, f(\mathbf{x}_i, \mathbf{\Theta})) = L$. $g(\lambda, v)$ is said to be a SPL-regulariser if*

i. $g(\lambda, v)$ should be convex with respect to $v \in [0, 1]$

ii. $v^*(\lambda; L)$ should be monotonically decreasing with respect to l_i, and the following holds:

(a) $\lim_{L\to 0} v^*(\lambda; L) = 1$.

(b) $\lim_{L\to\infty} v^*(\lambda; L) = 0$.

iii. $v^*(\lambda; L)$ should be monotonically increasing with respect to λ, and it holds that:

(a) $\lim_{\lambda\to 0} v^*(\lambda; L) = 0$.

(b) $\lim_{\lambda \to \infty} v^*(\lambda; L) \leq 1$.

Definition 1 ensures the following:

- i ensures a convex optimisation problem.
- ii ensures that a observations with a lower loss function value is gets a high SPL weight value.
- iii ensures more observations is introduced to the training process as the age parameter increases.

2.2.2 SPL parameter estimation

SPL parameter estimation is an iterative algorithm. It considers an initial set of parameters and age parameter value. The optimal parameter weights are calculated for each age parameter. After each SPL training iteration, the age parameter λ, gets increased. This ensures that more training examples is included in the next SPL iteration process [12, 30]. This gives us our most basic formulation of the SPL algorithm [12, 13]:

Algorithm 1 Basic SPL algorithm.

Require: Initial set of parameters $\mathbf{\Theta}^{(0)}$,
Require: Initial age paramater $\lambda^{(0)} > 0$,
Require: Learning rate $a > 1$.
while (Not all data included) or (Model paramaters not conerged) **do**
Update $\mathbf{v}^{(r+1)} \leftarrow \min_{\mathbf{v}} \left\{ E(\mathbf{\Theta}^{(r)}, \mathbf{v} | \lambda^{(r)}) \right\}$
Update $\mathbf{\Theta}^{(r+1)} \leftarrow \arg\min_{\mathbf{\Theta}} \left\{ E(\mathbf{\Theta}, \mathbf{v}^{(r+1)} | \lambda^{(r)}) \right\}$
Update $\lambda^{(r+1)} \leftarrow a \cdot \lambda^{(r)}$
end while

It is shown that algorithm 1 guarantees convergence if the SP-regulariser meets the conditions stated above [16].

2.2.3 Self-Paced-Learning For Finite Mixture Models

SPL is typically used for a supervised learning problem with latent variables [10, 16]. However, SPL is not just limited to supervised learning [12]. A SPL mixture of Student's t model (SPTMM) was created [28].

SPTMM uses the negative complete log-likelihoods in place of a loss function and proposed using a hard SP-regulariser SPTMM initialises the age parameter λ to be the median of the set $\{l_i(\mathbf{\Psi})\}_{i=1}^{n}$ where l_i is the log-likelihood of the mixture of students t for observation $\mathbf{x}_i$ with model parameters $\mathbf{\Psi}$ [28]. The SPTMM hard regulariser is defined as

$$g^H(\lambda, \mathbf{v}) = -\lambda \sum_{i=1}^{n} |v_i|.$$

with optimal SP-weights $\mathbf{v}^* = (v_1^*, ..., v_n^*)^T \in \mathbb{R}^n$ given as

$$v_i^* = \begin{cases} 1 & \text{if } -l_i < \lambda \\ 0 & \text{if } -l_i \geq \lambda \end{cases}$$

for all $i = 1, 2, ..., n$.
This results in the SPTMM objective function

$$-\sum_{i=1}^{n} v_i \cdot l_i(\mathbf{\Psi}) + g(\lambda, \mathbf{v}). \tag{2.7}$$

Algorithm 2 is used to estimate the SPTMM model's parameters.

Algorithm 2 SPTM parameter estimation algorithm

Require: Initial set of parameters $\mathbf{\Psi}^{(0)}$,

Require: Initial age parameter $\lambda^{(0)}$ as the median of the set $\{l_i(\mathbf{\Psi})\}_{i=1}^{n}$,

Require: Learning rate $a > 1$.

```
while (while not converged) or (all observations included) do
    while (while not converged) or (all observations included) do
        Update v^(r+1) ← min_v {E(Ψ^(r), v|λ^(s))}
        Using fixed v^(r+1) to estimate Ψ using EM algorithm:
        while (while not converged) or (all observations included) do
            E-Step
            M-Step
        end while
        Update Ψ^(r+1)
    end while
    Update λ^(s+1) ← a · λ^(s)
end while
```

By considering the negative complete data log-likelihood of any FMM in equation (2.7), will result in the general self-paced finite mixture model (SPFMM).

2.3 Summary

This chapter was brief discussion of the necessary background theory used in the rest of this research. This chapter also served as an introduction to the notation used in this research.

Chapter 3

Self-paced learning for finite mixture models

In this chapter we consider a self-paced learning finite mixture model (SPFMM). Definition 2.2.1 ii requires a loss function with range of $L \in [0, \infty)$ and an age parameter $\lambda > 0$. Consider the SPTMM model as described in Chapter 2.2, it is possible for a set of initial parameter values to have $l_i(\boldsymbol{\Psi}) < 0$ for all all $i = 1, 2, ..., n$. The SPTMM model would initialise an age parameter $\lambda < 0$ and will therefore by Definition 1 not choose any observation as all the SPL-weights will be equivalent to zero. The opposite can also occur, if any age parameter of $\lambda > 0$ is chosen, then all the observations will be included and therefore the model will just be a regular FMM model.

We propose a generalised definition for the SP-regulariser to account for this and preserve the properties of Definition 1. The SPL weights for the SPFMM is calculated by only considering an observation's likelihood for the entire FMM. This chapter will also extend the SPFMM method by proposing a model that applies SPL on each individual component of the FMM, allowing for a FMM model having self-paced components learning.

3.1 Self-paced components finite mixture model

Let $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n$ be a random p-dimensional sample of size n from model (2.1). Let $\mathbf{z} = \{\mathbf{z_i}\}$ be the set of latent cluster memberships. The complete data is $(\mathbf{x}, \mathbf{z})$ with complete data log-likelihood given as equation (2.5).

Let $\mathbf{V} = (v_{ij})$ denote the collection of SPL weights such that v_{ij} is the SPL weight of observation $\mathbf{x}_i$ for component j. Let $\mathbf{v}_j = (v_{1j}, ..., v_{nj})$ be the set of SPL weights for component j for each individual observation.

Let $g(\mathbf{v}, \boldsymbol{\lambda})$ be the SP-regulariser imposed on SPL weights where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_k)$ is the set of age parameters for each component, where λ_j is the age parameter of component j. The proposed model is expressed as

$$E(\mathbf{V}, \boldsymbol{\Psi}|\boldsymbol{\lambda}) = -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot l_{ij}(\boldsymbol{\Psi}_j) + g(\mathbf{V}, \boldsymbol{\lambda}) \tag{3.1}$$

where $l_{ij}(\boldsymbol{\Psi}) = z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j)$ with component parameters $\boldsymbol{\Psi_j} = \{\pi, \boldsymbol{\Theta}_j\}$.

As stated in the introduction of this chapter, the requirements of Definition 1 will not be met using (3.1). We propose the following definition for a SP-regulariser

Definition 2. *Let v_{ij} be a SPL weight variable for component j, $l_{ij}(\boldsymbol{\Psi}) = z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\theta_j)$ is the be the log-likelihood value for component j for a observation $\mathbf{x}_i$, and λ_j the age parameter for component j. Suppose $v_{ij}^*(\boldsymbol{\lambda}, l_{ij})$ denotes the optimal v_{ij} such that $v_{ij}^*(\boldsymbol{\lambda}, l_{ij}) = \arg\min_v -v_{ij} \cdot l_{ij} + g(\mathbf{V}, \boldsymbol{\lambda})$, for $v \in [0, 1]$. $g(\mathbf{V}, \boldsymbol{\lambda})$ is said to be a SPL-regulariser if*

i. $g(\mathbf{V}, \boldsymbol{\lambda})$ should be convex with respect to $v_{ij} \in [0, 1]$ for all $i = 1, ..., n$ and $j = 1, .., k$.

ii. $v_{ij}^*(\boldsymbol{\lambda}; l_{ij})$ should be monotonically decreasing with respect to $-l_{ij}$, and the following holds:

(a) $\lim_{-l_{ij} \to -\infty} v_{ij}^*(\boldsymbol{\lambda}; l_{ij}) = 1$.

(b) $\lim_{-l_{ij}\to\infty} v_{ij}^*(\boldsymbol{\lambda}; l_{ij}) = 0.$

iii. $v^*(\boldsymbol{\lambda}; -l_{ij})$ *should be monotonically increasing with respect to* λ_j*, and it holds that:*

(a) $\lim_{\lambda_j\to-\infty} v_{ij}^*(\boldsymbol{\lambda}; l_{ij}) = 0.$

(b) $\lim_{\lambda_j\to\infty} v_{ij}^*(\boldsymbol{\lambda}; l_{ij}) \leq 1.$

Definition 2 ensures the following properties still holds,

- i ensures a convex optimisation problem.
- ii ensures that a observations with a lower negative log-likelihood obtains a higher SPL weight value.
- iii ensures more observations is introduced to the training process as the age parameter increases.

Model (3.1) is called the self-paced components finite mixture model (SPCFMM).

3.2 Model estimation

The SPL parameter estimation algorithm iterates between the following two steps, 1) minimise (3.1) for a particular age parameter and, 2) increase the age parameters. This is repeated until all the observations are included or until the parameters convergences.

For the SPFMM model in particular we required to minimise (3.1)

$$(\hat{\mathbf{V}}, \hat{\boldsymbol{\Psi}}) = \arg\min_{\mathbf{V},\boldsymbol{\Psi}} \left\{ -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot l_{ij}(\boldsymbol{\Psi}_j) + g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}) \right\}.$$

This is equivalent to finding the parameter values that maximises the negative of model (3.1) i.e.

$$(\hat{\mathbf{V}}, \hat{\boldsymbol{\Psi}}) = \arg\max_{\mathbf{V},\boldsymbol{\Psi}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot l_{ij}(\boldsymbol{\Psi}_j) - g(\mathbf{V}, \boldsymbol{\lambda}) \right\}. \tag{3.2}$$

We use the EM-algorithm to do this.

Expectation-step

For the sth age parameter on the $(r+1)$th iteration of the EM algorithm, denoted by $(s, r+1)$, requires the calculation of $Q(\boldsymbol{\Psi}, \boldsymbol{V}|\boldsymbol{\lambda}^{(s)})$ using the fixed set of parameters $\boldsymbol{\Psi}^{(s,r)}$.

$$\begin{aligned} Q(\boldsymbol{\Psi}, \boldsymbol{V}|\boldsymbol{\Psi}^{(s,r)}, \boldsymbol{\lambda}^{(s)}) &= \mathbb{E}_{\mathbf{z}|\boldsymbol{x}, \boldsymbol{\Psi}^{(r)}, \boldsymbol{\lambda}^{(s)}}(E(\mathbf{V}, \boldsymbol{\Psi}^{(s,r)}|\boldsymbol{\lambda}^{(s)})) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot \mathbb{E}_{z_{ij}|\boldsymbol{x}, \boldsymbol{\Psi}^{(r)}, \boldsymbol{\lambda}^{(s)}}(l_{ij}(\boldsymbol{\Psi}_j^{(s,r)})) - g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot \mathbb{E}_{z_{ij}|\boldsymbol{x}, \boldsymbol{\Psi}^{(r)}, \boldsymbol{\lambda}^{(s)}}(z_{ij}) \cdot \left\{ \log {\pi_j}^{(s,r)} + \log f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j^{(s,r)}) \right\} - g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}). \end{aligned} \tag{3.3}$$

We simply calculate $\mathbb{E}_{z_{ij}|\boldsymbol{x}, \boldsymbol{\Psi}^{(r)}, \boldsymbol{\lambda}^{(s)}}(z_{ij})$ and replace the value in equation (3.3)

$$\begin{aligned} z_{ij}^{(s,r+1)} &= \mathbb{E}_{(z_{ij}|\mathbf{x}, \boldsymbol{\Psi}^{(r)})}(z_{ij}) \\ &= \frac{\pi_j^{(s,r)} \cdot f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j^{(s,r)})}{\sum_{o=1}^{k} \pi_o^{(s,r)} \cdot f_o(\mathbf{x}_i|\boldsymbol{\Theta}_o^{(s,r)})} \end{aligned} \tag{3.4}$$

for all $i = 1, ..., n$ and $j = 1, .., k$.

$\mathbf{z}^{(s,r+1)} = (z_{ij}^{(s,r+1)})$ is the set of estimated values. The derivation of $z_{ij}^{(s,r+1)}$ is given in Appendix A.0.1.

This results in

$$Q(\boldsymbol{\Psi}, \mathbf{V}|\boldsymbol{\lambda}^{(s)}) = -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot z_{ij}^{(s,r+1)} \cdot \left\{ \log {\pi_j}^{(s,r)} + \log f_j(\mathbf{x}_i|\boldsymbol{\Theta}^{(s,r)}) \right\} + g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}). \tag{3.5}$$

Note that we do not use the SPL weights in the calculation of $\mathbf{z}^{(s,r+1)}$

3.2.1 Maximisation-Step

On the $(s, r+1)$th iteration, we calculate $\mathbf{V}^{(s,r+1)}$ by using the fixed parameters $\boldsymbol{\Psi}^{(s,r)}$ and then calculate $\boldsymbol{\Psi}^{(s,r+1)}$ by using $\mathbf{V}^{(s,r+1)}$.

Optimisation of V

We use the following updating equations

$$\begin{aligned}\mathbf{V}^{(s,r+1)} &= \arg\max_{\mathbf{V}} Q(\mathbf{V}|\mathbf{\Psi}^{(s,r)}, \boldsymbol{\lambda}^{(s)}) \qquad (3.6)\\ &= \arg\max_{\mathbf{V}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot z_{ij}^{(s,r+1)} \cdot \left\{ \log \pi_j^{(s,r)} + \log f_j(\mathbf{x}_i|\mathbf{\Theta}_j^{(s,r)}) \right\} - g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}) \right\}.\end{aligned}$$

If $g(\mathbf{V}, \boldsymbol{\lambda}^{(s)}) = -\sum_{i=1}^{n}\sum_{j=1}^{k} \lambda_j \cdot |v_{ij}|$ then the optimal weights $\mathbf{V}^*$ is given by:

$$v_{ij}^* = \begin{cases} 1 & \text{if } -l_{ij}(\mathbf{\Psi}_j|\mathbf{z}^{(s,r+1)}) < \lambda_j^{(s)} \\ 0 & \text{if } -l_{ij}(\mathbf{\Psi}_j|\mathbf{z}^{(s,r+1)}) \geq \lambda_j^{(s)} \end{cases}. \qquad (3.7)$$

Optimisation of Ψ

We estimate $\mathbf{\Psi}^{(s,r+1)}$ by

$$\begin{aligned}\mathbf{\Psi}^{(s,r+1)} &= \arg\max_{\mathbf{\Psi}} Q(\mathbf{\Psi}|\mathbf{V}^{(s,r+1)}, \mathbf{\Psi}^{(s,r)}, \boldsymbol{\lambda}^{(s)}) \qquad (3.8)\\ &= \arg\max_{\mathbf{\Psi}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}^{(s,r+1)} \cdot l_{ij}(\mathbf{\Psi}|\boldsymbol{z}^{(s,r+1)}) - g(\mathbf{V}^{(s,r+1)}, \boldsymbol{\lambda}) \right\}\\ &= \arg\max_{\mathbf{\Psi}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}^{(s,r+1)} \cdot l_{ij}(\mathbf{\Psi}|\boldsymbol{z}^{(s,r+1)}) \right\} \qquad (3.9)\\ &= \arg\max_{\mathbf{\Psi}} \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)} \cdot \left\{ \log \pi_j{}^{(s,r+1)} + \log f_j(\mathbf{x}_i|\mathbf{\Theta}_j)^{(s,r+1)} \right\}.\end{aligned}$$

This results in Algortihm 3.

Algorithm 3 SPCFMM parameter estimation algorithm

Require: Initial set of parameters $\boldsymbol{\Psi}^{(0)}$

Require: Initialise $\lambda_j^{(0)}$ as the median of $\{l_{ij}(\boldsymbol{\Psi}_j^{(0)})\}_{i=1}^k$ for $j = 1, ..., k$.

Require: $a > 1$

while (not converged) or $(\sum_{i=1}^n \sum_{j=1}^k v_{ij} = n \cdot k)$ **do**

while (not converged) or $(\sum_{i=1}^n \sum_{j=1}^n v_{ij} = n \cdot k)$ **do**

E-step:

Calculate: $z_{ij}^{(s,r+1)}$ using (3.4)

M-step:

Calculate: $v_{ij}^{(s,r+1)}$ using (3.7)

Calculate $\boldsymbol{\Psi}^{(s,r+1)}$ using (3.8)

end while

Update $\boldsymbol{\Psi}^{(s+1)} \leftarrow \boldsymbol{\Psi}^{(\mathbf{s})}$

Update $\boldsymbol{\lambda}^{(s+1)} \leftarrow \boldsymbol{\lambda}^{(s)} \cdot a$

end while

3.3 The self-paced components Gaussian mixture model

We propose the self-paced components Gaussian mixture model (SPCGMM). It is a SPCFMM model with a k-component Gaussian mixture model and a hard SP-regulariser. Consider the GMM complete data log-likelihood function

$$l(\boldsymbol{\Psi}) = \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \big\{ \log \pi_j + \log \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \big\} \tag{3.10}$$

where $\phi()$ indicates the pdf of a Gaussian with unknown parameters for the jth component $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ and where $\boldsymbol{\Psi} = \{\boldsymbol{\mu}, \boldsymbol{\Pi}, \boldsymbol{\Sigma}\}$ with $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k\}$, $\boldsymbol{\Pi} = \{\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_k\}$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k\}$.

Equation (3.10) is of the same form as equation (2.5).

3.3.1 Model formulation

SPCGMM has objective function

$$E(\mathbf{V}, \boldsymbol{\Psi}) = \left\{ -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot z_{ij}\big\{ \log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\big\} - \lambda_j |v_{ij}| \right\}, \tag{3.11}$$

where $z_{ij}\big\{ \log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\big\}$ is the log-likelihood for component j and observation $\mathbf{x}_i$, $\phi()$ indicates the pdf of a Gaussian distribution, where $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}_j\}_{j=1}^{k}$ is the unknown parameters for component j with $\boldsymbol{\Psi}_j = \{\pi_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}$.

3.3.2 Model estimation

We use the EM-algorithm is used to maximise the negative of (3.11).

Expectation Step

On the $(s, r+1)$th iteration of the EM-algorithm we calculate

$$\begin{aligned} \mathbb{E}_{(z_{ij}|\mathbf{X}, \boldsymbol{\Psi}^{(s,r)})}[z_{ij}] &= \frac{\pi_j^{(s,r)} \cdot \phi_j(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})}{\sum_{o=1}^{k} \pi_o^{(s,r)} \cdot \phi_o(\mathbf{x}_i|\boldsymbol{\mu}_o^{(s,r)}, \boldsymbol{\Sigma}_o^{(s,r)})} \\ &= z_{ij}^{(s,r+1)} \end{aligned} \tag{3.12}$$

We find (3.12) by simply substituting $\phi()$ in equation (3.4).

Therefore,

$$Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \boldsymbol{\lambda}^{(s)}) = \left\{ -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot z_{ij}^{(s,r+1)}\big\{ \log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\big\} - \lambda_j^{(s)} |v_{ij}| \right\} \tag{3.13}$$

Maximisation step

Estimation of V

We calculate the optimal SPL weight, $\mathbf{V}^{(s,r+1)}$, by

$$v_{ij}^{(s,r+1)} = \begin{cases} 1 & \text{if } -z_{ij}^{(s,r+1)}\{\log \pi_j^{(s,r)} + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})\} < \lambda \\ 0 & \text{if } -z_{ij}^{(s,r+1)}\{\log \pi_j^{(s,r)} + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})\} \geq \lambda \end{cases}$$

for all $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$.

The proof of this optimal weight is given in Appendix A.1.1.

Estimation of $\boldsymbol{\Psi}$

The following equations are used to update $\boldsymbol{\Psi}$:

$$\pi_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}}{\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}}. \tag{3.14}$$

$$\boldsymbol{\mu}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)} \cdot \mathbf{x}_i^T}{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}.}. \tag{3.15}$$

$$\boldsymbol{\Sigma}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T(\mathbf{x}_i - \boldsymbol{\mu}_j)}{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}}. \tag{3.16}$$

The derivations of the updating steps are given in Appendix A.1.

3.4 Notes on robustness of the self-paced component Gaussian mixture model

The updating equation (3.15) is a weighted sum of the observations with weight $v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}$. If $v_{ij}^{(s,r+1)} = 0$ then the observation is completely removed in the component's mean parameter calculation. The same occurs for the calculation of (3.16) and (3.14).

SPCGMM is different to the standard SPL since it only excludes an observations using the likelihood of the component instead of using the whole target GMM's likelihood value. Therefore if an observation has a high likelihood for some component j but not for other components, then it is not outright excluded in the parameter estimation process but still included in component j's parameter estimation calculations.

3.5 SPCFMM relationship with SPFMM

We show that SPFMM is a special case of the SPCFMM. Suppose we have a SPCFMM model with the following conditions:

- $\lambda_1 = \lambda_2 = ... = \lambda_k = \lambda$.
- $v_{i1} = v_{i2} = ... = v_{ik} = v_i$ for all $i = 1, 2, ..., n$.
- $\mathbf{v}$ is the collection of each v_i.

This leads to the following objective function:

$$\begin{aligned} E(\mathbf{V}, \mathbf{\Psi}) &= -\sum_{i=1}^{n}\sum_{j=1}^{k} v_i \cdot \log l_{ij}(\mathbf{\Psi}) + g(\lambda, \mathbf{v}) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{k} v_i \cdot \left[z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\theta_j) \right] + g(\lambda, \mathbf{v}) \\ &= -\sum_{i=1}^{n} v_i \cdot \left[\sum_{j=1}^{k} z_{ij} \cdot \log \pi_j + z_{ij} \cdot \log f_j(\mathbf{x}_i|\theta_j) \right] + g(\lambda, \mathbf{v}) \\ &= -\sum_{i=1}^{n} v_i \cdot l_i(\mathbf{\Psi}) + g(\lambda, \mathbf{v}). \end{aligned}$$

This is the SPFMM objective function. This shows that SPFMM is a special case of SPCFMM.

3.6 Summary

This chapter proposed a SPCFMM model that provides SPL for each individual components of the FMM. We also proposed a general definition for a SP-regulariser that preserves the properties of a SP-regulariser and that will work for a FMM's log-likelihood range. We discussed and showed how the SPCGMM minimises the impact of non-typical observations.

Chapter 4

Robust Self-Paced Finite Mixture Model Learning

The SPL-FMM model described in Chapter 3 only considers the likelihood of the FMM to assign SPL weight values. We propose a SPL model that also takes into account a distributional framework for non-typical observations.

We propose a FMM with a component with bad observations as members. This method introduces a new set of latent indicator variables namely whether the observation is a good or bad observation.

Instead of using expectation to determine whether an observation is good or bad, we propose using a SPL weight calculation to determine this. This chapter will describe the underlying theory for the Robust SPL-FMM model.

4.1 FMM with bad component formulation

Consider a FMM with a component for bad observations. This is a mixture model that consists of the target FMM and a distribution for bad data. Let the distribution of a random vector $\mathbf{X} \in \mathbb{R}^p$ for such a mixture model be written as

$$P(\mathbf{x}|\mathbf{\Psi}) = \alpha \cdot u(\mathbf{x}|\mathbf{\Psi}_1) + (1-\alpha)h(\mathbf{x}|\mathbf{\Psi}_2), \tag{4.1}$$

where α is the mixing proportion of good observations, with $\alpha \in [0,1]$,
$u(\mathbf{x}, \mathbf{\Psi}_1)$ is the density of good observations with parameters $\mathbf{\Psi}_1$,
$h(\mathbf{x}|\mathbf{\Psi}_2)$ is the density of the bad observations with paramaters $\mathbf{\Psi}_2$.
with $\mathbf{\Psi} = \{\mathbf{\Psi}_1, \mathbf{\Psi}_2\}$.

4.1.1 Model based clustering

Consider a set of n p-variate observations $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n$ from model (4.1), unknown cluster memberships for density $u(\mathbf{x}, \mathbf{\Psi}_1)$ indicator vectors $\mathbf{z} = \{\mathbf{z}_i\}_{i=1}^n$ and let $\mathbf{v} = \{v_1, ..., v_2\}$ be the set of indicator variables that indicate if an observation is good, with

$$v_i = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ is a good observation} \\ 0 & \text{otherwise} \end{cases}$$

for all $i = 1, 2, .., n$.
The complete data set are $(\mathbf{x}, \mathbf{z}, \mathbf{v})$, with joint distribution

$$P(\mathbf{x}, \mathbf{z}, \mathbf{v}|\mathbf{\Psi}) = \prod_{i=1}^{n} \left\{ \alpha \cdot \left\{ \prod_{j=1}^{k} \pi_j^{z_{ij}} f_j(\mathbf{x}_i|\mathbf{\Theta}_j)^{z_{ij}} \right\} \right\}^{v_i} \cdot \left\{ (1-\alpha) \cdot h(\mathbf{x}_i|\mathbf{\Psi}_2) \right\}^{(1-v_i)}$$

and with complete data log-likelihood given by

$$\begin{aligned} l(\mathbf{\Psi}) \quad = \quad & \sum_{i=1}^{n} v_i \log \alpha \\ & + \sum_{i=1}^{n} (1-v_i) \cdot \log(1-\alpha) \\ & + \sum_{i=1}^{n} \sum_{j=1}^{k} v_i \cdot z_{ij} \log \pi_j + \sum_{i=1}^{n} \sum_{j=1}^{k} v_i \cdot z_{ij} \log f_j(\mathbf{x}_i|\mathbf{\Theta}_j) \\ & + \sum_{i=1}^{n} (1-v_i) \cdot \log h(\mathbf{x}_i|\mathbf{\Psi}_2). \end{aligned} \tag{4.2}$$

This can be rewritten as

$$\begin{aligned} l(\boldsymbol{\Psi}) &= \sum_{i=1}^{n} v_i \cdot l^G(\boldsymbol{\Psi}_1) + (1 - v_i) \cdot l^B(\boldsymbol{\Psi}_2) \\ &= \sum_{i=1}^{n} v_i \cdot \Big(l^G(\boldsymbol{\Psi}_1) - l^B(\boldsymbol{\Psi}_2) \Big) + l^B(\boldsymbol{\Psi}_2) \end{aligned} \tag{4.3}$$

where $l^G(\boldsymbol{\Psi}_1) = \log \alpha + \sum_{j=1}^{k} z_{ij} \log \pi_j + \sum_{i=1}^{n} \sum_{j=1}^{k} \cdot z_{ij} \log f_j(\mathbf{x}_i|\boldsymbol{\Theta}_j)$ is the log-likelihood of good observations component for observation x_i, with parameters $\boldsymbol{\Psi}_1 = \{\alpha, \boldsymbol{\Pi}\}$ and, $l^B(\boldsymbol{\Psi}_2) = \log(1 - \alpha) + \log h(\mathbf{x}_i|\boldsymbol{\Psi}_2)$ is the log-likelihood of the bad component distribution for observation $\mathbf{x}_i$, with parameters $\boldsymbol{\Psi}_2 = \{\alpha, \boldsymbol{\Psi}\}$.

4.2 Robust self-paced learning formulation

Consider a set of observations $\{\mathbf{x}_i\}_{i=1}^n$ from model (4.1) with $\mathbf{x}_i \in \mathbb{R}^p$ for all $i = 1, 2, ..., n$. Let the missing values $\mathbf{v} = \{v_i\}_{i=1}^n$ denote SPL weights. By adding a SP-regulariser to the negative complete data log-likelihood of model (4.1) results in the robust SPL FMM model

$$E(\mathbf{v}, \boldsymbol{\Psi}|\lambda) = \sum_{i=1}^{n} -\Big\{ v_i \cdot l^G(\boldsymbol{\Psi}_1) + (1 - v_i) \cdot l^B(\boldsymbol{\Psi}_2) \Big\} + g(\mathbf{v}, \lambda) \tag{4.4}$$

$$= \sum_{i=1}^{n} v_i \cdot \Big(l^B(\boldsymbol{\Psi}_2) - l^G(\boldsymbol{\Psi}_1) \Big) - l^B(\boldsymbol{\Psi}_2) + g(\mathbf{v}, \lambda) \tag{4.5}$$

where λ is the age parameter of the SPL model.

The proposed robust SPL model is different than the models described in Chapter 2 and 3, since the FMM has SPL weights embedded in the complete log-likelihood. Instead of estimating the missing values $\{v_i\}_{i=1}^n$ by using expectation, we propose to treat them as SPL-weights and estimate them by minimising the objective function (4.4) like in a traditional SPL model.

4.3 Robust self-paced finite mixture model estimation

RSPFMM uses a similar parameter estimation algorithm as Algorithm 3. Consider the following updating equations used to calculate the SPL weights and model parameters

4.3.1 Expectation step

RSPLMN uses the exact same formula to estimate the latent cluster memberships as equation (3.3). We therefore calculate $Q(\boldsymbol{\Psi}, \boldsymbol{v}|\lambda^{(s)})$ by first calculating

$$z_{ij}^{(s,r+1)} = \mathbb{E}_{z_{ij}|\boldsymbol{x},\boldsymbol{\Psi}^{(s,r)},\lambda^{(s)}}(z_{ij})$$

for all $i = 1, 2, .., n$ and $j = 1, 2, ..., k$.

We then find Q by substituting the missing values in $E(\boldsymbol{\Psi}, \mathbf{v})$ with $\mathbf{z}^{(s,r+1)}$

$$Q(\boldsymbol{\Psi}, \boldsymbol{v}|\lambda^{(s)}) = \sum_{i=1}^{n} v_i \cdot \Big(l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) - l^G(\boldsymbol{\Psi}_1|\mathbf{z}^{(s,r+1)})\Big) - l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) + g(\mathbf{v}, \lambda). \quad (4.6)$$

4.3.2 Minimisation Step

Estimation of v

On the $(s, r+1)$, we calculate $\mathbf{v}^{(s,r+1)}$ by

$$\begin{aligned}
\mathbf{v}^{(s,r+1)} &= \arg\min_{\mathbf{v}} \left\{ \sum_{i=1}^{n} v_i \cdot \Big(l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) - l^G(\boldsymbol{\Psi}_1|\mathbf{z}^{(s,r+1)})\Big) - l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) + g(\mathbf{v}, \lambda) \right\} \\
&= \arg\min_{\mathbf{v}} \left\{ \sum_{i=1}^{n} v_i \cdot \Big(l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) - l^G(\boldsymbol{\Psi}_1|\mathbf{z}^{(s,r+1)})\Big) + g(\mathbf{v}, \lambda) \right\} \quad (4.7)
\end{aligned}$$

If $g(\mathbf{v}, \lambda) = -\lambda \sum_{i=1}^{n} |v_i|$ then the optimal weights $\mathbf{v}^*$ is given by:

$$v_i^* = \begin{cases} 1 & \text{if } l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) - l^G(\boldsymbol{\Psi}_1|\mathbf{z}^{(s,r+1)}) < \lambda^{(s)} \\ 0 & \text{if } l^B(\boldsymbol{\Psi}_2|\mathbf{z}^{(s,r+1)}) - l^G(\boldsymbol{\Psi}_1|\mathbf{z}^{(s,r+1)}) \geq \lambda^{(s)} \end{cases} \quad (4.8)$$

RSPFMM considers the difference of log-likelihood between an observation being a good or bad observation. If an observation has a higher likelihood of being a good observation, then it is included. If an observation has a higher likelihood of being a bad observation then it is excluded unless difference in likelihood is below the threshold λ. As λ increases, the more observations are included.

Estimation of $\boldsymbol{\Psi}$

On the $(s, r+1)$th iteration estimate $\boldsymbol{\Psi}$ by using updating equations

$$\begin{aligned} \boldsymbol{\Psi}^{(s,r+1)} &= \arg\min_{\boldsymbol{\Psi}} Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \mathbf{v}^{(s,r+1)}) \\ &= \arg\min_{\boldsymbol{\Psi}} \left\{ l(\boldsymbol{\Psi}|\mathbf{z}^{(s,r+1)}, \mathbf{v}^{(s,r+1)}) + g(\mathbf{v}^{(\mathbf{s},\mathbf{r+1})}, \lambda^{(s)}) \right\} \\ &= \arg\min_{\boldsymbol{\Psi}} \; l(\boldsymbol{\Psi}|\mathbf{z}^{(s,r+1)}, \mathbf{v}^{(s,r+1)}) \end{aligned} \tag{4.9}$$

consider the following RSPLM algorithm

Algorithm 4 RSPFMM algorithm

Require: Initial set of parameters $\boldsymbol{\Psi}^{(0)}$
Require: Initial set of age parameters λ
Require: $a > 1$
  while $(\sum_{i=1}^{n} v_i = n)$ or (paramaters converged) **do**
    while $(\sum_{i=1}^{n} v_i = n)$ or (paramaters converged) **do**
      Calculate $Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)})$ **by**:
        Update: $z_{ij}^{(s,r+1)}$ using (3.4)
      Perform minimisation step by:
        Update: $v_{ij}^{(s,r+1)}$ using (3.6)
        Update $\boldsymbol{\Psi}^{(s,r+1)}$ using (3.8)
    end while
    Update $\boldsymbol{\Psi}^{(s+1)} \leftarrow \boldsymbol{\Psi}^{(\mathbf{s})}$
    Update $\lambda^{(s+1)} \leftarrow \lambda^{(s)} \cdot a$
  end while

4.4 Robust self-paced Gaussian mixture model

Here we describe a robust self-paced Gaussian mixture model (RSPGMM).

4.4.1 Model formulation

Consider a RSPFMM model with a k-component GMM with complete data log-likelihood for the good observations with

$$l(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij}^{u} \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\} \tag{4.10}$$

where $\mathbf{z^u} = (z^u(z_{ij}))$ is the set of cluster memberships for the good points, with unknown model parameters, $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^{k}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_j\}_{j=1}^{k}$, $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_j\}_{j=1}^{k}$.

Let the distribution of the bad observations, h, be a k-component GMM with the following complete data log-likelihood

$$l(\boldsymbol{\pi}, \boldsymbol{\mu}, \eta, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij}^{h} \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \eta \cdot \boldsymbol{\Sigma}_j)\} \tag{4.11}$$

where $\mathbf{z^h} = (z^h(z_{ij}))$ is the set of cluster memberships for the bad observations, with unknown model parameters, $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^{k}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_j\}_{j=1}^{k}$, η, $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_j\}_{j=1}^{k}$. where $\eta \in (1, \infty)$.

h shares common parameters $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^{k}$ and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_j\}_{j=1}^{k}$ with the target GMM. The covariance parameters of h is the scaled covariance matrices of the target GMM, $\eta \cdot \boldsymbol{\Sigma}_j$ for all $j = 1, 2, .., k$. This means each component of the non-typical distribution has a wider spread.

η plays a similar role as the degree of contamination parameter of the CNM and will therefore be referred to as the degree of contamination parameter for the RSPGMM model.

The proportion of good observations is $\alpha \in (0, 1)$. RSPGMM has a hard SP-regulariser

$$g(\mathbf{v}, \lambda) = -\lambda \sum_{i=1}^{n} |v_i|.$$

Therefore, RSPGMM has objective function

$$E(\mathbf{v}, \boldsymbol{\Psi}, \lambda) \quad = \quad \sum_{i=1}^{n} v_i \cdot \Big(l^B(\boldsymbol{\Psi}_2) - l^G(\boldsymbol{\Psi}_1) \Big) - l^B(\boldsymbol{\Psi}_2) + g(\mathbf{v}|\lambda)$$

with $l(\boldsymbol{\Psi}_1) = \log(\alpha) + \sum_{j=1}^{k} z_{ij}^u \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\}$ the log-likelihood of the good observations with parameters $\boldsymbol{\Psi}_1 = \{\alpha, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$,
$l(\boldsymbol{\Psi}_2) = \log(1-\alpha) + \sum_{j=1}^{k} z_{ij}^h \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \eta \cdot \boldsymbol{\Sigma}_j)\}$ the log-likelihood of the bad observations with parameters $\boldsymbol{\Psi}_2 = \{(1-\alpha), \boldsymbol{\pi}, \boldsymbol{\mu}, \eta, \boldsymbol{\Sigma}\}$.

4.4.2 Model estimation

Expectation step

On iteration $(s, r+1)$ of the E-step we need to calculate two latent variables, $\mathbf{z}^u$ and $\mathbf{z}^h$. The E-step uses the following updating equations:

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}^u|\mathbf{x}_i, \mathbf{v}, \boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)}}[E(\mathbf{v}, \boldsymbol{\Psi}^{(r)}, \lambda^{(s)})] &= \frac{\pi_j^{(s,r)} \cdot \phi_j(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})}{\sum_{o=1}^{k} \pi_o^{(s,r)} \cdot \phi_o(\mathbf{x}_i|\boldsymbol{\mu}_o^{(s,r)}, \boldsymbol{\Sigma}_o^{(s,r)})} \\ &= z_{ij}^{u(s,r+1)} \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{Z}^h|\mathbf{x}_i, \mathbf{v}, \boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)}}[E(\mathbf{v}, \boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)})] &= \frac{\pi_j^{(s,r)} \cdot \phi_j(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \eta^{(s,r)} \cdot \boldsymbol{\Sigma}_j^{(s,r)})}{\sum_{o=1}^{k} \pi_o^{(s,r)} \cdot \phi_o(\mathbf{x}_i|\boldsymbol{\mu}_o^{(s,r)}, \eta^{(s,r)} \cdot \boldsymbol{\Sigma}_o^{(s,r)})} \\ &= z_{ij}^{h(s,r+1)}. \end{aligned}$$

This is derived in a similar fashion to the E-step equations in Chapter 3. The RSPGMM is capable of assigning a cluster membership to the non-typical observation.

This results in $Q(\mathbf{v}, \boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)})$,

$$\begin{aligned} Q(\mathbf{v}, \boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)}) &= \log(\alpha) + \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}^{u(s,r+1)} \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\} \\ &\quad + \log(1-\alpha) + \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}^{h(s,r+1)} \cdot \{\log \pi_j + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \eta \cdot \boldsymbol{\Sigma}_j)\} \\ &\quad - \lambda^{(s)} \sum_{i=1}^{n} |v_i|. \end{aligned}$$

Maximisation Step

Estimating v

By using equation (4.8), $\mathbf{v}^{(s,r+1)}$ is calculated by using:

$$v_i^{(s,r+1)} = \begin{cases} 1 & \text{if } l^B(\boldsymbol{\Psi_2}|z^{h(s,r+1)}) - l^G(\boldsymbol{\Psi_1}|z^{u(s,r+1)}) < \lambda \\ 0 & \text{if } l^B(\boldsymbol{\Psi_2}|z^{h(s,r+1)}) - l^G(\boldsymbol{\Psi_1}|z^{u(s,r+1)}) \geq \lambda \end{cases} \tag{4.12}$$

for all $i = 1, 2, ..., n$

Estimation of $\boldsymbol{\Psi}$

The minimisation step simply consist of:

$$\boldsymbol{\Psi}^{(s,r+1)} = \arg\min_{\boldsymbol{\Psi}} \left\{ Q(\boldsymbol{\Psi}|\mathbf{v}^{(s,r+1)}\boldsymbol{\Psi}^{(s,r)}, \lambda^{(s)}) \right\}$$

The following equations is used to update $\boldsymbol{\Psi}$:

$$\alpha^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_i^{(s,r+1)}}{n},$$

$$\pi_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} S_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{k} S_{ij}}, \tag{4.13}$$

$$\boldsymbol{\mu}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} N_{ij} \cdot \mathbf{x}_i}{\sum_{i=1}^{n} N_{ij}}, \tag{4.14}$$

$$\boldsymbol{\Sigma}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} N_{ij} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j^{(s,r+1)}) \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j^{(s,r+1)})^T}{\sum_{i=1}^{n} S_{ij}}, \tag{4.15}$$

where,

$S_{ij} = \left[z_{ij}^{u(s,r+1)} \cdot v_i^{(s,r+1)} + z_{ij}^{h(s,r+1)} \cdot (1 - v_i^{(s,r+1)}) \right]$
and
$N_{ij} = \left[z_{ij}^{u(s,r+1)} \cdot v_i^{(s,r+1)} + \frac{z_{ij}^{h(s,r+1)} \cdot (1 - v_i^{(s,r+1)})}{\eta} \right]$,
for all $i = 1, 2, ..., n$ and $j = 1, 2, ..., k$.

The derivations of the updating steps is provided in section A.2.1.

We update η by using some numerical optimisation algorithm.

$$\eta^{(r+1)} = \arg\max_{\eta} \left\{ \begin{array}{c} \sum_{i=1}^{n} \sum_{j=1}^{k} \left[z_{ij}^{u(s,r+1)} \{ \log \pi_j + \log \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \} \right] \\ + \sum_{i=1}^{n} \sum_{j=1}^{k} \left[z_{ij}^{h(s,r+1)} \{ \log \pi_j + \log \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \eta \cdot \boldsymbol{\Sigma}_j) \} \right] \end{array} \right\}$$

4.5 Notes on the robustness of the robust self-paced Gaussian mixture model

Consider equation (4.14). Clearly the component mean parameter is the weighted sum of the observations with weight $N_{ij} = \left[z_{ij}^{u(s,r+1)} \cdot v_i^{(s,r+1)} + \frac{z_{ij}^{h(s,r+1)} \cdot (1 - v_i^{(s,r+1)})}{\eta} \right]$. There are two main mechanisms that mitigate the impact of non-typical observations:

1. The first mechanism uses the RSPGMM's ability to estimate a cluster membership for non-typical observations and uses it in the parameter estimation.

   $$\text{If } v_i = 1 \text{ then } N_{ij} = \frac{z_{ij}^{h(s,r+1)}}{\eta}.$$

   RSPGMM uses the non-typical distribution's estimation of cluster membership in the weight calculation. It frequently produce a better cluster membership estimate as it considers a more spread out target mixture model to model outliers.

2. The second mechanism RSPFMM used to mitigate the impact of non-typical observations is the use of the degree of contamination parameter η. Since $\eta > 1$ we

know that N_{ij} is a decreasing function in terms of η. This means the more spread out the non-typical observations the less it contributes to the calculation of (4.14). RSPGMM can adapt the degree of penalisation based on the observations since η is estimated using the observations and not a fixed value.

The same two mechanisms occur in the calculation of (4.4.2). This differs from the standard SPFMM and SPCGMM as it does not completely exclude any of the observations labeled as non-typical in the parameter estimation calculations, but it does reduce the impact of these observations.

The mixing proportions of the RSPGMM is similar to that of SPGCMM but it does not completely exclude the non-typical observations. It uses the estimated non-typical observation's cluster membership in the calculation of (4.13).

4.6 Summary

In this chapter we proposed a robust self-pace learning algorithm that includes a distributional framework for non-typical observations. We provided a general RSPFMM model in equation (4.4) and Algorithm 4 provides the means to estimate the model.

We described a specific RSPFMM model called the RSPGMM. RSPGMM is a GMM model using the RSPFMM framework for model estimation. We provided the updating equations used to estimate the RSPGMM model. We discussed the robustness of the RSPGMM model and also discussed how it mitigates the impact of non-typical observations.

Chapter 5

Simulation Study

The simulation study demonstrates the properties of the proposed algorithms and compare them with well known models. We consider data simulated from five different mixture models, where the data sets covers different types of scenarios. This study will measure each fitted model's parameter estimation bias and mean squared error (MSE) under the different scenarios. The study will keep the component structure simple by only considering two dimensional data with two components.

5.1 Simulation design

We use five data generation processes with two dimensions $p = 2$ and two clusters $k = 2$:

(a) CNM distribution, with parameters $\alpha_1 = 0.9$ and $\alpha_2 = 0.8$, $\eta_1 = 20$, $\eta_2 = 30$;

(b) GMM;

(c) GMM with 1% points randomly substituted with the non-typical points with coordinates $(0, x_{i2})$ where x_{i2} is generated from a uniform distributed over the interval $(10, 15)$;

(d) GMM with 5% of observations randomly substituted with uniform noise over the interval $(-10, 10)$ for each dimension x_1 and x_2;

(e) MT with $\nu_1 = 4$ and $\nu_2 = 10$ degrees of freedom;

The processes share the following common parameters:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \ \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \ \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \ \pi_1 = 0.3.$$

We consider two scenarios for the mean of the first component. The first scenario with well separated components $\mu_1 = [0, -3]^T$ and the second with partially overlapped components with $\mu_1 = [0, -1]^T$.

The data sets covers a wide range of data typically encountered in practice [22]:

- Scenarios (a) and (e) covers heavy-tailed clusters
- Scenario (b) represents a data set with no non-typical data.
- Scenarios (c) and (d) represents data sets with two different types of non-typical observations.

The data sets are generated from each scenario by taking into consideration the degree of overlap and the size of the sample ($n = 500, 200, 100$). We generated a 1000 data sets from each scenario for each model we fit. A total of 150 000 data sets are generated for this simulation study.

5.2 Model Fitting

We compare SPCGMM and RSPGMM with the well known CNM, GMM and MT. We also compare them with the SPGMM, a self-paced learning algorithm applied on a Gaussian mixture model. The simulation study was conducted using **R** [23]. All models used the output of a k-means model as fitted model's initial mean, covariance and proportion parameters values. We refer to the literature for reasonable initial model specific parameters values such as the degree of contamination or age parameter.

The initial values of the model specific parameters are specified below.

1. RSPGMM is implemented in **R**. The model will uses the following model specific initial values: Proportion of good observations, $\alpha = 0.5$, degree of contamination $\eta = 10$, learning rate $a = 1.05$.

2. CNM with proportion of good observations for each component $\alpha_1, \alpha_2 = 0.5$ and degree of contamination $\eta_1, \eta_2 = 10$.

3. GMM implemented using the gpcm() function of the mixture package [20].With learning rate $a = 1.05$

4. SPGMM implemented in **R** with learning rate $a = 1.05$

5. MT using the teigen() function of the teigen() package [1]. We use the unconstrained MT model.

For learning rates, we considered a moderate value of $a = 1.05$ for all the SPL models. We do this to ensure a fair comparison between the performance of the SPL models. Similarly we use a conservative degree of contamination parameter $\eta = 10$ for each component of the CNM and RSPGMM.

We initialise the age parameters by using the median of the likelihood values of each observation [28]:

1. For SPGMM
$$\lambda = median\left(\left\{\sum_{j=1}^{k} -l(\mathbf{\Psi}|\mathbf{z}^{(0)})\right\}_{i=1}^{n}\right).$$

2. For SPCGMM
$$\lambda_j = median\left(\left\{ - l_{ij}(\mathbf{\Psi}_j|\mathbf{z}^{(0)})\right\}_{i=1}^{n}\right),$$
for all $j = 1, 2, ..., k$.

3. For RSPGMM
$$\lambda = median\left(\left\{l^B(\mathbf{\Psi_2}|z^{h(0)}) - l^G(\mathbf{\Psi_1}|z^{F(0)})\right\}_{i=1}^{n}\right).$$

Label switching is common in FMMs [22]. We will choose a label by using the labels with the closest geometric distance from the true labels.

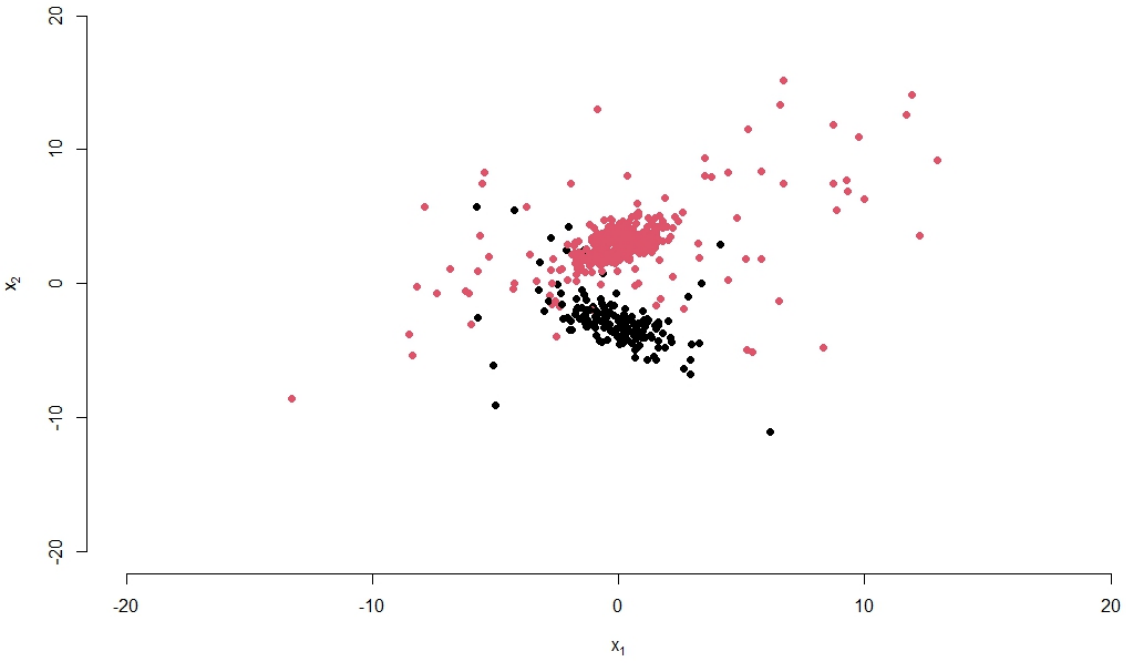


Figure 5.1: Scenario (e) with well separated components

5.3 Simulation study results

The parameter bias and MSE for μ_{11}, μ_{12}, μ_{21}, μ_{22}, $\sigma_{11}^{(1)}$, $\sigma_{22}^{(1)}$, $\sigma_{11}^{(2)}$, $\sigma_{22}^{(2)}$ and π_1 are used as a comparative measure of each methods estimation abilities[1]. The bias is calculated by taking the average estimated parameter and subtracting the true value. The MSE is calculated by taking the average of the squared estimated parameter subtracted with the true value. Bias and MSE values closer to 0 is considered to be a better.

All the simulation study results can be found in Appendix B

[1] $\sigma_{do}^{(j)}$ represents the jth component's covariance matrix entry d, o.

Tables B.1-B.6 and Tables B.25-B.30 reports on scenario (a) and (e) respectively. It is clear the RSPGMM and CNM performs better than the GMM model. RSPGMM did overall the best and did particularly well for component variance estimations. For scenario (a) SPCGMM generally performed better than CNM and was slightly out performed by RSPGMM. SPGMM performed comparably to the GMM model. For scenario (e) both SPGMM and SPCGMM performed worse in the estimation of μ_{21}. The rest of the estimated parameters are similar to GMM. MT did the worst in these scenario.

Tables B.7-B.12 reports on scenario (b). In the well separated component scenarios we find that RSPGMM, SPCGMM and SPGMM did comparably well. CNM and GMM was also comparable to each other but out performed the SPL models. This was to be expected as SPL model tend to find slightly worse parameter estimates in clean observations [16]. For the partially separated components we find RSPGMM, CNM and GMM did comparable well with RSPGMM slightly outperforming the other models. SPGMM and SPCGMM did once again worse in the estimation of μ_{21}. The rest of the parameters estimated was also worse when comparing CNM, GMM and RSPGMM. MT had overall the worst performance.

Tables B.13-B.18 and Tables B.19-B.24 reports on scenario (c) and (d) respectively. RSPGMM and CNM did comparably well however, RSPGMM performed significantly better in the component variance estimation. For scenario (d) SPCGMM did comparably well with RSPGMM and CNM for well separated components. SPCGMM and SPGMM does however struggle with close components to calculate μ_{21} since it has a higher MSE and bias values than compared to other models.

Overall RSPGMM did the best to mitigate the impact of non-typical observations by obtaining in general a lower bias and MSE values than compared to the other models. RSPGMM does not perform as well as CNM and GMM when working with a data set with no non-typical observations. SPCGMM performs overall better than SPGMM, but both models did not yield the best component estimates using the same learning rate as RSPGMM. The results does show SPGMM and SPCGMM does mitigate the impact of

non-typical observations in some cases, but does not always perform better than GMM. MT did the worst in all results. This may be due to the package used.

5.4 Summary

We have compared the parameter estimation ability between a few robust methods fitted to five simulated scenarios with data sets consisting of 500, 200 and 100 observations. We fitted each model on a 1000 data sets for each scenario. The bias and MSE values for each parameter estimates were recorded. RSPGMM which uses the RSPFMM strategy showed an overall better fit for data with noisy or non-typical observations when compared to the CNM, SPCGMM, SPGMM, GMM and MT models. We also find SPCGMM outperformed SPGMM.

Chapter 6

Real Data Application

In this chapter we apply the RSPGMM, SPCGMM, SPGMM and CNM approaches on the well known crabs data set [5, 22]. Each algorithm is applied on the original data set and on an altered data set. The goal of this application is to investigate how close each model can calculate parameters to that of the observed summary statistics. We also wish to observe how sensitive the models are to outliers by observing how much the parameters estimates change on the altered data set. We once again compare the proposed models to that of the CNM that serves as a baseline performance.

6.1 Data set description

The data set contains morphological measurements of blue crabs or the species Leptograpsus variegatus. The data set was collected at Fremantle, W. Australia. The data sets contains a 100 observations in total of which 50 observations is male and female. Each observation contains 5 measurements, frontal lobe size (FL), rear width (RW), carapace length (CL), carapace width (CW) and body depth (BD) [5]. Each observation is measured in millimetres.

A scatter plot matrix of the crab data set can be seen in figure 6.1. We note that the CW and BD variables indicates little separation of clusters. Therefore we will only consider the CL and RW data set depicted in figure 6.2

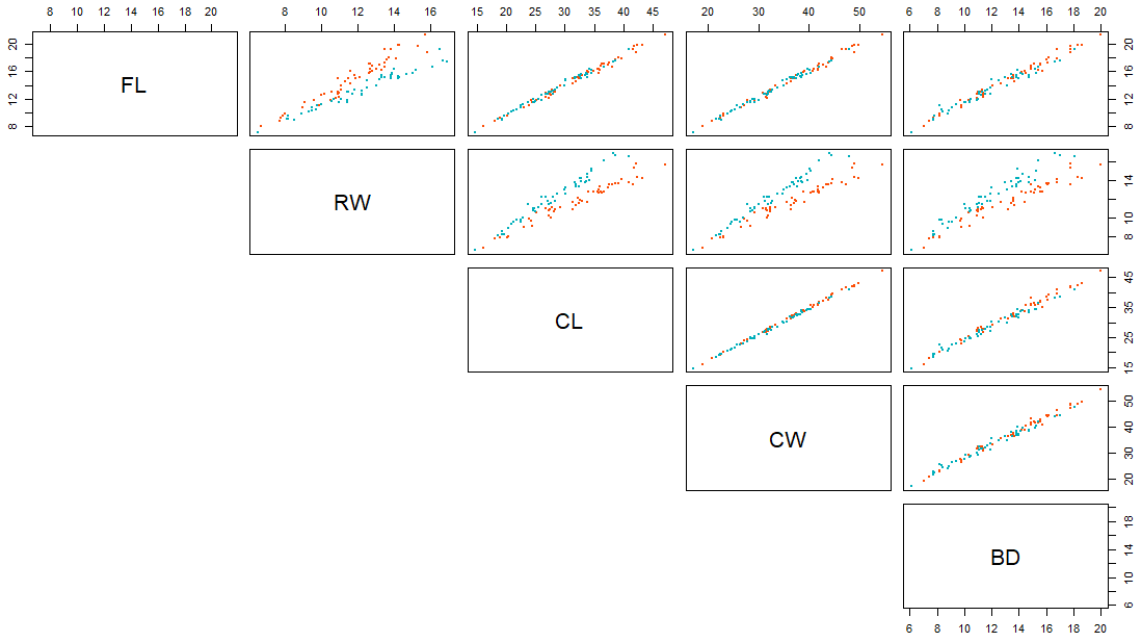


Figure 6.1: Scatter plot matrix of Blue crabs data set (Orange points denotes male and blue points denote females)

We alter the CL-RW crabs data set by replacing 5% of the observations with five outliers. We do this by changing the CL values of five observations:

- $\mathbf{x}_{1,1} = 0$
- $\mathbf{x}_{1,2} = 1$
- $\mathbf{x}_{1,50} = 30$
- $\mathbf{x}_{1,95} = 25$
- $\mathbf{x}_{1,100} = 27$

Figure 6.2 shows the original data set and the altered data set next to each other.

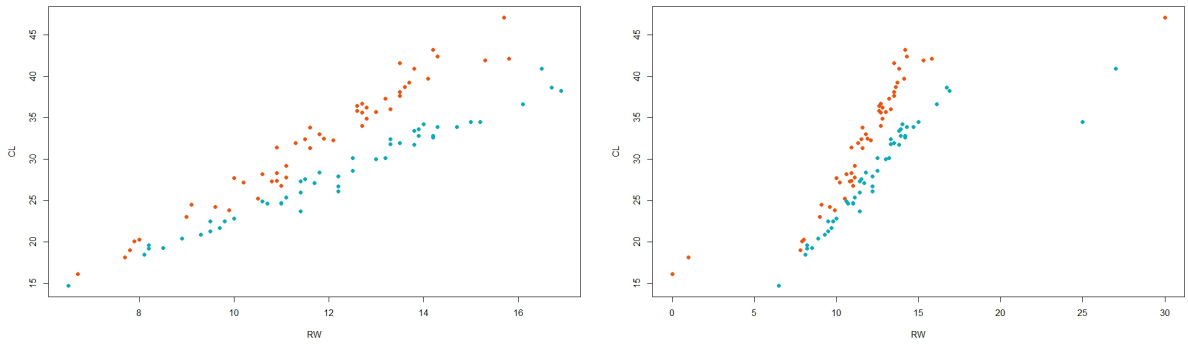


Figure 6.2: Scatter plots of original and altered CL-CW blue crabs data sets (Orange points denotes male and blue points denote females)

6.2 Study design

We compare the RSPGMM, SPCGMM and SPGMM models with the CNM model, by documenting the parameters estimated from the original data set and the parameters estimated from the altered data set. We calculate the sample mean for each sex and use it as a baseline measure to see how accurate the models estimate the parameters.

For this study let the following symbols represent:

- μ_{11} is the mean RW value for the male blue crab;
- μ_{12} is the mean CL value for the male blue crab;
- μ_{21} is the mean RW value for the female blue crab;
- μ_{22} is the mean CL value for the female blue crab;
- π_1 is the proportion of male blue crab;

6.2.1 Model fitting

Each model is initialised by using the same k-means output value. This was done for both data sets. Each methods uses the following model specific initial values and hyper

parameters:

- RSPGMM with $\alpha = 0.5$, $\eta = 10$,$a = 1.005$.
- CNM with $\boldsymbol{\alpha} = (0.5, 0.5)^T$, $\boldsymbol{\eta} = (10, 10)^T$.
- SPCGMM with $a = 1.005$.
- SPGMM with $a = 1.005$.

Each age parameter for the SPL models are initialised using the same strategy as in Chapter 5.

6.2.2 Results

	Summary statistics	RSPGMM		CNM		SPCGMM		SPGMM	
		Original	Altered	Original	Altered	Original	Altered	Original	Altered
μ_{11}	11.7	11.6597864	11.5978385	10.1068777	7.8026215	11.3713138	11.5654454	11.2898369	11.4730881
μ_{12}	32.0	34.2669874	34.4816216	31.5756940	21.3615323	34.4669094	34.3497564	27.8055413	27.7908770
μ_{21}	13.3	12.3274335	12.3794837	11.2996200	7.9693041	12.3676475	12.3176423	10.7550773	10.7658385
μ_{22}	28.1	27.0940652	27.0932739	23.4986901	18.2937461	26.7020267	27.0940652	26.4699461	26.8665332
π_1	0.5	0.4017175	0.3891028	0.4089935	0.4851704	0.3194348	0.3440598	0.1408499	0.1443908

Table 6.1: Parameter estimates for crabs data set

Table 6.1 contains the results of this study. On the original data set we clearly see RSPGMM and CNM is comparable. SPCGMM had good estimates similar to RSPGMM, however the proportion of male crabs is significantly worse. SPGMM was the least accurate on all parameter estimates.

When looking at the altered data set we observe that the RSPGMM and SPCGMM had the best estimates. All the SPL models was comparable in mitigating the impact of the added outliers but the RSPFMM and SPCGMM had the more accurate parameter estimates. RSPGMM did however get worse proportion of male crabs estimation when compared to CNM. CNM was the most sensitive to the non-typical observations and resulted in significantly worse estimates for all component, the only exception is for the proportion of male crabs was better on the altered data set.

6.3 Summary

We fitted the RSPLMN, SPCLMN, SPLMN and CNM models on the crabs data set. We compared the parameter estimates of the component means and proportion of male crabs with the estimated parameters of the same models trained on a artificially contaminated data set. We observe that all the SPL approaches mitigated the influences of the non-typical observations, however RSPLMN appeared to have performed the best. CNM struggled to capture the components well with and without outliers.

Chapter 7

Conclusions

This chapter summarises the findings of this mini-dissertation.

7.1 Summary of Conclusions

This mini-dissertation focused on applying SPL on FMMs for accurate model parameter estimation. We proposed a generalised SP-regulariser used for FMMs that preserves the properties of a well defined SP-regulariser used in supervised learning problems. A generalised SPFMM model description was provided and the theory was extended to create a SPCFMM model that performs SPL on each component of the FMM individually. We described the SPCGMM model which applies the SPCFMM method on a GMM. We discussed the model estimation equations and the how they mitigate the impact of non-typical observations and how it differs with a standard SPL algorithm

We also proposed an extension to the SPFMM frame work by considering a distributional frame work for non-typical observations in the SPL weight calculation. This propsed method is called RSPFMM and we provided a general model and algorithm for this model. We described the the estimation process of the RSPGMM model which is a specific application of RSPFMM on a GMM. We similarly discussed the mechanism that mitigates the influence of non-typical observations in the updating equations and compared the differences of RSPGMM and SPCGMM in terms of how it mitigates the

impact of non-typical observations.

We performed a simulation study to observe and compare the properties of the prosed methods with that of well known robust mixture models. We found that the RSPGMM using the RSPFMM model resulted in the lowest parameter estimation bias and MSE values for data with non-typical observations.

We also applied the proposed models to the blue crabs data set. The estimated parameters of the data set was compared to that of the to the observed summary statics of the data set and to the parameters estimated using the CNM model. We also fit the models on an altered blue crabs data set to investigate how much the estimated parameters changed in the presence of non-typical observations.

We find that the SPL models did comparably and mitigated the effects of non-typical observations and was less sensitive to these influential points than the CNM model. This research confirms that the proposed algorithms and in particular RSPLFMM mitigates the impact of outliers and is thus a highly robust parameter estimation algorithm. For future work we suggest exploring the consistency and asymptotic normality of the resulting estimators from our proposed methods.

Bibliography

[1] Jeffrey L. Andrews, Jaymeson R. Wickins, Nicholas M. Boers, and Paul D. McNicholas. teigen: An R package for model-based clustering and classification via the multivariate t distribution. *Journal of Statistical Software*, 83(7):1–32, 2018.

[2] Jeffrey D Banfield and Adrian E Raftery. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, pages 803–821, 1993.

[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[4] Christopher M Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, tenth edition edition, 2006.

[5] NA Campbell and RJ Mahon. A multivariate study of variation in two species of rock crab of the genus leptograpsus. *Australian Journal of Zoology*, 22(3):417–425, 1974.

[6] Alberto Maria Cozzini. *Supervised And Unsupervised Model-Based Clustering With Variable Selection*. PhD thesis, Imperial College London, 2012.

[7] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[8] Christian Hennig. Breakdown points for maximum likelihood estimators of location–scale mixtures. *The Annals of Statistics*, 32(4):1313–1340, 2004.

[9] John R Hipp and Daniel J Bauer. Local solutions in the estimation of growth mixture models. *Psychological methods*, 11(1):36, 2006.

[10] M Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. *Advances in neural information processing systems*, 23, 2010.

[11] Kenneth Lange, David R Hunter, and Ilsoon Yang. Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, 9(1):1–20, 2000.

[12] Weiyi Li, Hongmei Chen, Tianrui Li, Jihong Wan, and Binbin Sang. Unsupervised feature selection via self-paced learning and low-redundant regularization. *Knowledge-Based Systems*, 240:108150, 2022.

[13] Bo Liu, Junrui Liu, Yanshan Xiao, Qihang Chen, Kai Wang, Ruiguang Huang, and Liangjiao Li. A new self-paced learning method for privilege-based positive and unlabeled learning. *Information Sciences*, 609:996–1009, 2022.

[14] Geoffrey J McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, second edition edition, 2008.

[15] Geoffrey J McLachlan and David Peel. Robust cluster analysis via mixtures of multivariate t-distributions. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 658–666. Springer, 1998.

[16] Deyu Meng, Qian Zhao, and Lu Jiang. A theoretical understanding of self-paced learning. *Information Sciences*, 414:319–328, 2017.

[17] Xiao-Li Meng and David Van Dyk. The em algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):511–567, 1997.

[18] D Peel and G MacLahlan. Finite mixture models. *John & Sons*, 2000.

[19] David Peel and Geoffrey J McLachlan. Robust mixture modelling using the t distribution. *Statistics and computing*, 10(4):339–348, 2000.

[20] Nik Pocuca, Ryan P. Browne, and Paul D. McNicholas. *mixture: Mixture Models for Clustering and Classification*, 2021. R package version 2.0.4.

[21] Antonio Punzo, Angelo Mazza, and Paul D. McNicholas. ContaminatedMixt: An R package for fitting parsimonious mixtures of multivariate contaminated normal distributions. *Journal of Statistical Software*, 85(10):1–25, 2018.

[22] Antonio Punzo and Paul D McNicholas. Parsimonious mixtures of multivariate contaminated normal distributions. *Biometrical Journal*, 58(6):1506–1537, 2016.

[23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[24] Luca Scrucca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1):289–317, 2016.

[25] Emilie Shireman, Douglas Steinley, and Michael J Brusco. Examining the effect of initialization strategies on the performance of gaussian mixture modeling. *Behavior research methods*, 49(1):282–293, 2017.

[26] James Stewart. *Single Variable Calculus : Early Transcendentals*. Cengage Learning, eighth edition metric version. edition, 2016.

[27] Salvatore D Tomarchio, Luca Bagnato, and Antonio Punzo. Model-based clustering via new parsimonious mixtures of heavy-tailed distributions. *AStA Advances in Statistical Analysis*, 106(2):315–347, 2022.

[28] Yang Zhang, Qingtao Tang, Li Niu, Tao Dai, Xi Xiao, and Shu-Tao Xia. Self-paced mixture of t distribution model. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2796–2800. IEEE, 2018.

[29] Hua Zhou, Liuyi Hu, Jin Zhou, and Kenneth Lange. MM algorithms for variance components models. *Journal of Computational and Graphical Statistics*, 28(2):350–361, 2019.

[30] Hua Zhou and Yiwen Zhang. EM vs MM: A case study. *Computational statistics & data analysis*, 56(12):3909–3920, 2012.

Appendix A

Proofs

This Appendix will provide the necessary proofs and derivations used in this mini-dissertation.

A.0.1 Derivation of $\mathbb{E}_{(z_{ij}|\mathbf{X},\boldsymbol{\Psi}^{(r)})}[z_{ij}]$

Consider the following expected value:

$$\begin{aligned}
\mathbb{E}_{(z_{ij}|\mathbf{X},\boldsymbol{\Psi}^{(r)})}[z_{ij}] &= 0 \cdot P(z_{ij}=0|\boldsymbol{\Psi}^{(r)}) + 1 \cdot P(z_{ij}=1|\boldsymbol{\Psi}^{(r)}) \\
&= P(z_{ij}=1|\boldsymbol{\Psi}^{(r)}) \\
&= \frac{P(z_{ij}=1)\cdot P(\mathbf{X},\boldsymbol{\Psi}^{(r)}|z_{ij}=1)}{P(\mathbf{X},\boldsymbol{\Psi}^{(r)})} \\
&= \frac{\pi_j \cdot f_j(\mathbf{x}_i|\boldsymbol{\mu}_j^{r+1},\boldsymbol{\Sigma}_j)}{\sum_{o=1}^{k}\pi_o \cdot f_o(\mathbf{x}_i|\boldsymbol{\mu}_o,\boldsymbol{\Sigma}_o)} \\
&= z_{ij}^{r}
\end{aligned}$$

for all $i = 1, 2..., n$ and $j = 1, 2, ..., k$.

A.1 Derivations for SPCGMM

A.1.1 Derivation of optimal V estimation for SPCGMM

Consider objective function (3.11)

$$E(\mathbf{V},\mathbf{\Psi}) = \left\{ -\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} - \lambda_j|v_{ij}|\right\},$$

then the optimal weights $\mathbf{v}^{(s,r+1)}$ is given by:

$$v_{ij}^{*} = \begin{cases} 1 & \text{if } -z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j)^{(s,r)}\big\} < -\lambda_j^{(s)} \\ 0 & \text{if } -z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} \geq -\lambda_j^{(s)} \end{cases}$$

for all $i = 1, 2..., n$ and $j = 1, 2, .., k$.

Proof. Consider,

$$\begin{aligned} Q(\mathbf{V}|\mathbf{\Psi}^{(s,r)},\boldsymbol{\lambda}^{(s)}) &= \sum_{i=1}^{n}\sum_{j=1}^{k} -v_{ij}\cdot z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} - \lambda_j^{(s)}|v_{ij}| \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} -v_{ij}\cdot z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} - \lambda_j^{(s)}\cdot v_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\left\{ -z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} - \lambda_j^{(s)}\right\}. \end{aligned}$$

If $\left(-z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j)^{(s,r)}\big\}\right) < -\lambda_j^{(s)}$ then

$$\left(-z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j)^{(s,r)}\big\} - \lambda_j^{(s)}\right) < 0,$$

for all $i = 1, 2, .., n$ and $j = 1, 2, ..., k$.

These terms will decrease the objective function if included therefore $v_{ij}^{(s,r+1)} = 1$ if $-z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} < -\lambda_j^{(s)}$.

Similarly if $-z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} \geq -\lambda_j^{(s)}$ then

$$\left(-z_{ij}^{(s,r+1)}\big\{\log {\pi_j}^{(s,r)} + \log\phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)},\boldsymbol{\Sigma}_j^{(s,r)})\big\} - \lambda_j^{(s)}\right) \geq 0,$$

for all $i = 1, 2, .., n$ and $j = 1, 2, ..., k$

This will increase the objective function if included and thus,

$$v_{ij}^{(s,r+1)} = 0 \text{ if } -z_{ij}^{(s,r+1)}\{\log {\pi_j}^{(s,r)} + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})\} \geq -\lambda_j^{(s)}.$$

We therefore have shown the optimal weights $V^{(s,r+1)}$ is

$$v_{ij}^* = \begin{cases} 1 & \text{if } -z_{ij}^{(s,r+1)}\{\log {\pi_j}^{(s,r)} + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})\} < -\lambda_j^{(s)} \\ 0 & \text{if } -z_{ij}^{(s,r+1)}\{\log {\pi_j}^{(s,r)} + \log \phi(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s,r)}, \boldsymbol{\Sigma}_j^{(s,r)})\} \geq -\lambda_j^{(s)} \end{cases}$$

for all $i = 1, 2..., n$ and $j = 1, 2, .., k$. □

Derivation of π_j

The parameters $\boldsymbol{\pi} = \{\pi_1, \pi_2, ..., \pi_k\}$ has the following restriction that must be satisfied: $\sum_{j=1}^{k} \pi_j = 1$. This requires a Lagrangian Multiplier $\mathcal{L}$ in order to estimate $\pi_j^{(r+1)}$:

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij} \cdot z_{ij} \cdot \ln \pi_j - \lambda\Big[\sum_{j=1}^{k} \pi_j - 1\Big]$$

Therefore, the following equations must be solved:

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \frac{v_{ij} \cdot z_{ij}}{\pi_j} - \lambda = 0 \tag{A.1}$$

for all $j = 1, 2, .., k$.

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{k} \pi_j - 1 = 0 \tag{A.2}$$

Using equation (A.7) to solve for π_j yields:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{v_{ij} \cdot z_{ij}}{\pi_j} - \lambda &= 0 \\
\iff \sum_{i=1}^{n} \frac{v_{ij} \cdot z_{ij}}{\pi_j} &= \lambda \\
\iff \lambda \cdot \pi_j &= v_{ij} \cdot z_{ij} \\
\iff \sum_{i=1}^{n} \frac{v_{ij} \cdot z_{ij}}{\lambda} &= \pi_j.
\end{aligned}$$

This results in:

$$\pi_j = \sum_{i=1}^{n} \frac{v_{ij} \cdot z_{ij}}{\lambda} \tag{A.3}$$

Now by using equation (A.14) we solve for λ:

$$\begin{aligned}
\sum_{j=1}^{k} \pi_j - 1 &= 0 \\
\iff \sum_{j=1}^{k} \Big\{ \sum_{i=1}^{n} \frac{v_{ij} \cdot z_{ij}}{\lambda} \Big\} &= 1 \\
\iff \lambda &= \sum_{j=1}^{k} \sum_{i=1}^{n} \Big\{ v_{ij} \cdot z_{ij} \Big\}
\end{aligned} \tag{A.4}$$

By substituting λ in equation (A.15) we obtain the final equation for $\hat{\pi}_j$:

$$\pi_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}}{\sum_{i=1}^{n} \sum_{j=1}^{k} v_{ij}^{(s,r+1)} \cdot z_{ij}^{(s,r+1)}}$$

Derivation of $\boldsymbol{\mu}_j$

For the sake of simplicity let $Q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the function that consists of the terms in $Q(\mathbf{V}|\boldsymbol{\Psi}^{(s,r)}, \boldsymbol{\lambda}^{(s)})$ that contains the parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \sum_{j=1}^{k} -\frac{1}{2} v_{ij} \cdot z_{ij} \cdot \ln |\boldsymbol{\Sigma}_j| - \frac{1}{2} v_{ij} \cdot z_{ij} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j),$$

consider

$$\begin{aligned} Q(\boldsymbol{\mu}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}^{(s,r+1)}\cdot z_{ij}^{(s,r+1)}\cdot\left\{\boldsymbol{\mu}_j^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j-2\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j+\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1}\mathbf{x}_i\right\} \end{aligned}$$

$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j}$ is given by:

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} = -\frac{1}{2}\sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\left\{2\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j-2\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1}\right\}$$

By solving $\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j}=0$ for μ_j:

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} &= 0 \\ \iff -\frac{1}{2}\sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\left\{2\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j-2\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1}\right\} &= 0 \\ \iff -\sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j+\sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} &= 0 \\ \iff \sum_{i=1}^{n} N_{ij}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j &= \sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} \\ \iff \sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\boldsymbol{\mu}_j &= \sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\mathbf{x}_i^T \\ \iff \boldsymbol{\mu}_j &= \frac{\sum_{i=1}^{n} v_{ij}\cdot z_{ij}\cdot\mathbf{x}_i^T}{\sum_{i=1}^{n} v_{ij}\cdot z_{ij}} \end{aligned} \tag{A.5}$$

This implies that:

$$\boldsymbol{\mu}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)}\cdot z_{ij}^{(s,r+1)}\cdot\mathbf{x}_i^T}{\sum_{i=1}^{n} v_{ij}^{(s,r+1)}\cdot z_{ij}^{(s,r+1)}}.$$

Derivation of $\boldsymbol{\Sigma}_j$

By considering $Q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, one only uses the terms that contain $\boldsymbol{\Sigma}_j$:

$$\begin{aligned} Q(\boldsymbol{\Sigma}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\cdot \ln|\boldsymbol{\Sigma}_j| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\cdot (\mathbf{x}_i-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\cdot \ln|\boldsymbol{\Sigma}_j| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} v_{ij}\cdot z_{ij}\cdot tr(\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T) \end{aligned}$$

let $s_{ij} = v_{ij}\cdot z_{ij}$.
By solving $\frac{\partial Q(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}_j} = 0$ for Σ_j:

$$\frac{\partial Q(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}_j} = 0$$

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} s_{ij}\cdot \ln|\boldsymbol{\Sigma}_j| - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} s_{ij}\cdot tr(\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T) = 0$$

$$\begin{aligned} \sum_{i=1}^{n} s_{ij}\cdot\boldsymbol{\Sigma}_j^T &= \sum_{i=1}^{n} s_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T \\ \boldsymbol{\Sigma}_j^T &= \sum_{i=1}^{n}\frac{s_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T}{\sum_{i=1}^{n} s_{ij}} \\ \boldsymbol{\Sigma}_j &= \frac{\sum_{i=1}^{n} s_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)^T(\mathbf{x}_i-\boldsymbol{\mu}_j)}{\sum_{i=1}^{n} s_{ij}} \end{aligned} \tag{A.6}$$

This results in:

$$\boldsymbol{\Sigma}_j^{(s,r+1)} = \frac{\sum_{i=1}^{n} v_{ij}^{(s,r+1)}\cdot z_{ij}^{(s,r+1)}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)^T(\mathbf{x}_i-\boldsymbol{\mu}_j)}{\sum_{i=1}^{n} v_{ij}^{(s,r+1)}\cdot z_{ij}^{(s,r+1)}}$$

A.1.2 Derivations for the self-paced component Gaussian mixture model estimation

Derivation of π_j

From chapter 3, the SPL objective function (3.11) will be used.

The parameters $\mathbf{\Pi} = \{\pi_1, \pi_2, ..., \pi_k\}$ has the following restriction that must be satisfied: $\sum_{j=1}^{k} \pi_j = 1$. This requires a Lagrangian Multiplier $\mathcal{L}$ in order to estimate $\pi_j^{(r+1)}$:

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{k} v_i \cdot z_{ij} \cdot \ln \pi_j - \lambda\Big[\sum_{j=1}^{k} \pi_j - 1\Big]$$

Therefore, the following equations must be solved:

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \frac{v_i \cdot z_{ij}}{\pi_j} - \lambda = 0 \tag{A.7}$$

for all $j = 1, 2, .., k$.

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{k} \pi_j - 1 = 0 \tag{A.8}$$

Using equation (A.7) to solve for π_j yields:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{v_i \cdot z_{ij}}{\pi_j} - \lambda &= 0 \\
\iff \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}}{\pi_j} &= \lambda \\
\iff \lambda \cdot \pi_j &= v_i \cdot z_{ij} \\
\iff \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}}{\lambda} &= \pi_j.
\end{aligned}$$

This results in:

$$\pi_j = \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}}{\lambda} \tag{A.9}$$

Now by using equation (A.14) to solve for λ yields:

$$\begin{aligned} \sum_{j=1}^{k} \pi_j - 1 &= 0 \\ \iff \sum_{j=1}^{k} \left\{ \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}}{\lambda} \right\} &= 1 \\ \iff \lambda &= \sum_{j=1}^{k} \sum_{i=1}^{n} \left\{ v_i \cdot z_{ij} \right\} \end{aligned} \tag{A.10}$$

By substituting λ in equation (A.15) we obtain the final equation for $\hat{\pi}_j$:

$$\pi_j^{(r+1)} = \frac{\sum_{i=1}^{n} v_i \cdot z_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{k} v_i \cdot z_{ij}}$$

Derivation of $\boldsymbol{\mu}_j$

By using equation:

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \sum_{j=1}^{k} -\frac{1}{2} v_i \cdot z_{ij} \cdot \ln |\boldsymbol{\Sigma}_j| - \frac{1}{2} v_i \cdot z_{ij} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j)$$

In order to optimise with respect to μ_j we need only consider:

$$\begin{aligned} Q(\boldsymbol{\mu}) &= -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} v_i \cdot z_{ij} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) \\ &= -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} v_i \cdot z_{ij} \cdot \left\{ \boldsymbol{\mu}_j^T \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j - 2\mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j + \mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1} \mathbf{x}_i \right\} \end{aligned}$$

$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j}$ is given by:

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} = -\frac{1}{2} \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \left\{ 2\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j - 2\mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1} \right\}$$

By solving $\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} = 0$ for μ_j:

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} &= 0 \\
\iff -\frac{1}{2}\sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \left\{ 2\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j - 2\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} \right\} &= 0 \\
\iff -\sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} &= 0 \\
\iff \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j &= \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} \\
\iff \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \boldsymbol{\mu}_j &= \sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \mathbf{x}_i^T \\
\iff \boldsymbol{\mu}_j &= \frac{\sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \mathbf{x}_i^T}{\sum_{i=1}^{n} v_i \cdot z_{ij}}
\end{aligned} \tag{A.11}$$

This implies that:

$$\boldsymbol{\mu}_j^{(r+1)} = \frac{\sum_{i=1}^{n} v_i \cdot z_{ij} \cdot \mathbf{x}_i^T}{\sum_{i=1}^{n} v_i \cdot z_{ij}.}$$

A.2 Derivations for the robust self-paced Gaussian mixture model estimation

A.2.1 Derivations of the minimisation step for RSPLMN

Consider a RSPGMM with function $Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \boldsymbol{\lambda})$ given in equation (3.13)

Derivation of α

Consider a RSPGMM with function Q given in equation (3.13) Only terms that contain α in function Q is considered to estimate α. For the sake of simplicity we denote these terms as the function $Q(\alpha)$.

$$Q(\alpha) = \sum_{i=1}^{n} v_i \cdot \ln \alpha + (1 - v_i) \cdot \ln (1 - \alpha) \tag{A.12}$$

$\frac{\partial Q(\alpha)}{\partial \alpha}$ is given by:

$$\begin{aligned}\frac{\partial Q(\alpha)}{\partial \alpha} &= \sum_{i=1}^{n} \frac{v_i}{\alpha} + \frac{1-v_i}{1-\alpha} \\ &= \sum_{i=1}^{n} \frac{v_i - \alpha}{\alpha \cdot (1-\alpha)}\end{aligned}$$

We find α where $\frac{\partial Q(\alpha)}{\partial \alpha} = 0$:

Only the numerator of $\frac{\partial Q(\alpha)}{\partial \alpha}$ is considered therefore:

$$\begin{aligned}\sum_{i=1}^{n} v_i - \hat{\alpha} &= 0 \\ \iff \sum_{i=1}^{n} v_i - \sum_{i=1}^{n} \hat{\alpha} &= 0 \\ \iff -\sum_{i=1}^{n} \hat{\alpha} &= -\sum_{i=1}^{n} v_i \\ \iff n \cdot \hat{\alpha} &= \sum_{i=1}^{n} v_i \\ \iff \hat{\alpha} &= \frac{\sum_{i=1}^{n} v_i}{n}\end{aligned}$$

$\hat{\alpha} \leq 1$ since $\sum_{i=1}^{n} v_i \leq n$.
If $\sum_{i=1}^{n} v_i = 0$ α then is not defined.

Note that $\alpha^{(r+1)} = \hat{\alpha}$

Derivation of π_j

The paramaters $\mathbf{\Pi} = \{\pi_1, \pi_2, ..., \pi_k\}$ has the following restriction that must be satisfied: $\sum_{j=1}^{k} \pi_j = 1$. This requres a Lagrangian Multiplier $\mathcal{L}$ in order to estimate $\pi_j^{(r+1)}$:

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left\{ \left[v_i \cdot z_{ij}^{u} + (1-v_i) \cdot z_{ij}^{h} \right] \cdot \ln \pi_j \right\} - \lambda \left[\sum_{j=1}^{k} \pi_j - 1 \right]$$

Therefore, the following equations must be solved:

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\pi_j} - \lambda = 0 \tag{A.13}$$

for all $j = 1, 2, .., k$.

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=1}^{k} \pi_j - 1 = 0 \tag{A.14}$$

Using equation (A.13) to solve for π_j yields:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\pi_j} - \lambda &= 0 \\
\iff \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\pi_j} &= \lambda \\
\iff \lambda \cdot \pi_j &= v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h \\
\iff \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\lambda} &= \pi_j.
\end{aligned}$$

This results in:

$$\pi_j = \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\lambda} \tag{A.15}$$

Now by using equation (A.14) to solve for λ yields:

$$\begin{aligned}
\sum_{j=1}^{k} \pi_j - 1 &= 0 \\
\iff \sum_{j=1}^{k} \left\{ \sum_{i=1}^{n} \frac{v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\lambda} \right\} &= 1 \\
\iff \lambda &= \sum_{j=1}^{k} \sum_{i=1}^{n} \left\{ v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h \right\}
\end{aligned} \tag{A.16}$$

By substituting λ in equation (A.15) we obtain the final equation for $\hat{\pi_j}$:

$$\hat{\pi_j} = \frac{\sum_{i=1}^{n} v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h}{\sum_{i=1}^{n} \sum_{j=1}^{k} \{ v_i \cdot z_{ij}^u + (1 - v_i) \cdot z_{ij}^h \}}$$

Derivation of $\boldsymbol{\mu}_j$

Let the function $Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ denote the function of the terms that only contain the parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta$ in $Q(\boldsymbol{\Psi}|\boldsymbol{\Psi}^{(s,r)}, \lambda)$

$$\begin{aligned} Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}(1-v_i)\cdot z_{ij}^{h}\cdot \ln \eta \\ &\quad -\left[v_i \cdot z_{ij}^{u} + (1-v_i)\cdot z_{ij}^{h}\right]\cdot \ln|\boldsymbol{\Sigma}_j| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^{u} + \frac{(1-v_i)\cdot z_{ij}^{h}}{\eta}\right]\cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j) \end{aligned}$$

In orrder to optimise with respect to μ_j we need only consider:

$$\begin{aligned} Q(\boldsymbol{\mu}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^{u} + \frac{(1-v_i)\cdot z_{ij}^{h}}{\eta}\right]\cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^{u} + \frac{(1-v_i)\cdot z_{ij}^{h}}{\eta}\right]\left\{\boldsymbol{\mu}_j^T \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j - 2\mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1}\mathbf{x}_i\right\} \end{aligned}$$

$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j}$ is given by:

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} = -\frac{1}{2}\sum_{i=1}^{n}\left[v_i \cdot z_{ij}^{u} + \frac{(1-v_i)\cdot z_{ij}^{h}}{\eta}\right]\left\{2\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j - 2\mathbf{x}_i^T \boldsymbol{\Sigma}_j^{-1}\right\}$$

Let $N_{ij} = \left[v_i \cdot z_{ij}^{u} + \frac{(1-v_i)\cdot z_{ij}^{h}}{\eta}\right]$.

By solving $\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} = 0$ for μ_j:

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}_j} &= 0 \\
\iff -\frac{1}{2}\sum_{i=1}^{n} N_{ij}\left\{2\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j - 2\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1}\right\} &= 0 \\
\iff -\sum_{i=1}^{n} N_{ij}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \sum_{i=1}^{n} N_{ij}\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} &= 0 \\
\iff \sum_{i=1}^{n} N_{ij}\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j &= \sum_{i=1}^{n} N_{ij}\mathbf{x}_i^T\boldsymbol{\Sigma}_j^{-1} \\
\iff \sum_{i=1}^{n} N_{ij}\boldsymbol{\mu}_j &= \sum_{i=1}^{n} N_{ij}\mathbf{x}_i^T \\
\iff \boldsymbol{\mu}_j &= \frac{\sum_{i=1}^{n} N_{ij}\mathbf{x}_i^T}{\sum_{i=1}^{n} N_{ij}}
\end{aligned} \tag{A.17}$$

This implies that:

$$\boldsymbol{\mu}_j^{(r+1)} = \frac{\sum_{i=1}^{n}\left[v_i^{(s,r+1)}\cdot z_{ij}^{u(s,r+1)} + \frac{(1-v_i^{(s,r+1)})\cdot z_{ij}^{h(s,r+1)}}{\eta}\right]\mathbf{x}_i^T}{\sum_{i=1}^{n}\left[v_i^{(s,r+1)}\cdot z_{ij}^{u(s,r+1)} + \frac{(1-v_i^{(s,r+1)})\cdot z_{ij}^{h(s,r+1)}}{\eta^{(s,r)}}\right]}$$

Derivation of $\boldsymbol{\Sigma}_j$

By considering $Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$, one only uses the terms that contain $\boldsymbol{\Sigma}_j$:

$$\begin{aligned} Q(\boldsymbol{\Sigma}) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^u + (1-v_i)\cdot z_{ij}^h\right]\cdot \ln|\boldsymbol{\Sigma}_j| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^u + \frac{(1-v_i)\cdot z_{ij}^h}{\eta}\right]\cdot (\mathbf{x}_i-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^u + (1-v_i)\cdot z_{ij}^h\right]\cdot \ln|\boldsymbol{\Sigma}_j| \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^u + \frac{(1-v_i)\cdot z_{ij}^h}{\eta}\right]\cdot tr(\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T) \end{aligned}$$

Let $S_{ij} = \left[v_i \cdot z_{ij}^u + (1-v_i)\cdot z_{ij}^h\right]$ By solving $\frac{\partial Q(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}_j} = 0$ for Σ_j:

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}_j} &= 0 \\ -\frac{1}{2}\sum_{i=1}^{n} z_{ij}\cdot \boldsymbol{\Sigma}_j^T - \frac{1}{2}\sum_{i=1}^{n} N_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T &= 0 \\ \sum_{i=1}^{n} z_{ij}\cdot\boldsymbol{\Sigma}_j^T &= N_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T \\ \boldsymbol{\Sigma}_j^T &= \frac{N_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)(\mathbf{x}_i-\boldsymbol{\mu}_j)^T}{\sum_{i=1}^{n} z_{ij}} \\ \boldsymbol{\Sigma}_j &= \frac{N_{ij}\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j)^T(\mathbf{x}_i-\boldsymbol{\mu}_j)}{\sum_{i=1}^{n} z_{ij}} \end{aligned} \tag{A.18}$$

This results in:

$$\boldsymbol{\Sigma}_j^{(s,r+1)} = \frac{\left[v_i^{(s,r+1)}\cdot z_{ij}^{u(s,r+1)} + \frac{(1-v_i^{(s,r+1)})\cdot z_{ij}^{h(s,r+1)}}{\eta^{(s,r)}}\right]\cdot(\mathbf{x}_i-\boldsymbol{\mu}_j^{(s,r+1)})^T(\mathbf{x}_i-\boldsymbol{\mu}_j^{(s,r+1)})}{\sum_{i=1}^{n}\left[v_i^{(s,r+1)}\cdot z_{ij}^{u(s,r+1)} + (1-v_i^{(s,r+1)})\cdot z_{ij}^{h(s,r+1)}\right]}$$

calculation of η

The only terms in $Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$ that contains η will be considered:

$$\begin{aligned} Q(\eta) &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}(1 - v_i^{(s,r+1)}) \cdot z_{ij}^{h(s,r+1)} \cdot \ln \eta \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[v_i \cdot z_{ij}^{u(s,r+1)} + \frac{(1 - v_i^{(s,r+1)}) \cdot z_{ij}^{h(s,r+1)}}{\eta}\right] \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j^{(s,r+1)})^T \boldsymbol{\Sigma}_j^{(s,r+1)})^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j^{(s,r+1)}) \end{aligned}$$

An optimisation method like Newton-Rhapson will be used to calculate the optimal value of η that will minimise $Q(\eta)$ since η can not be written in a explicit form.

Appendix B

Simulation study results

B.0.1 Results for scenario (a)

Results for well separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0061021	0.0105914	0.0254399	0.1885539	-0.0473522	0.0843514	-0.0737568	0.0813617	-0.0910230	0.1034642	0.2532486	0.1268887
$\mu_{21} = 0$	-0.0005462	0.0048128	0.0079054	0.0417993	0.0538479	0.1339499	0.0032560	0.0074210	0.0446842	0.0405023	-0.5802644	0.6780362
$\mu_{12} = -3$	0.0008598	0.0101267	0.6527035	0.8973212	0.6178948	2.0190661	0.1955068	0.3991035	2.8352342	16.1890997	2.4132606	6.1638681
$\mu_{22} = 3$	-0.0007651	0.0046966	-0.1566502	0.0776801	-0.0046153	0.1580686	0.0013636	0.0079333	0.0625244	0.0343743	-2.7539579	7.6470940
$\sigma_{11}^{(1)} = 1$	-0.2786190	0.0966082	1.2176976	6.2897091	4.9197898	46.4575141	0.4259788	0.8746912	-0.4279292	0.2149050	-0.8238095	0.6994703
$\sigma_{22}^{(1)} = 1$	-0.2045193	0.0511432	0.8920367	12.8702715	4.4061749	51.3705657	-0.3386820	0.1416242	-0.3162584	0.1245046	-0.8126871	0.6612806
$\sigma_{11}^{(2)} = 1$	-0.2534063	0.0833873	2.3174009	14.8120424	5.9917861	94.9258200	0.7255592	5.5407341	-0.4505647	0.2435259	-0.9295621	0.8680894
$\sigma_{22}^{(2)} = 1$	-0.1977278	0.0475680	0.9042283	10.7083033	4.8679081	69.5697189	-0.3435630	0.1463243	-0.3086352	0.1209451	-0.9218930	0.8503797
$\pi_1 = 0.3$	0.0126118	0.0006934	0.0747576	0.0194334	0.0620598	0.0212522	0.0941726	0.0136283	-0.0769981	0.0318370	0.0021985	0.0008114

Table B.1: Simulation results for scenario (a) with well separated components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0395374	0.0817266	-0.0231607	0.4832805	-0.1768133	0.9673572	-0.1446826	0.6594844	-0.1032496	0.1640173	0.2886791	0.2373050
$\mu_{21} = 0$	0.0416832	0.0509247	0.1594906	1.0167003	0.1265721	1.3484386	0.0231143	0.1017922	0.0376815	0.0519065	-0.5342551	0.6894632
$\mu_{12} = -3$	0.3749850	2.1129203	0.8172670	1.6929788	0.5586007	2.3377687	0.4036138	1.3181589	2.8006063	15.6385016	2.4391827	6.3341808
$\mu_{22} = 3$	0.0164190	0.0199970	-0.0816909	0.5956597	0.0861548	0.8610068	0.0306590	0.1473990	0.0896708	0.0571915	-2.7405066	7.6044799
$\sigma_{11}^{(1)} = 1$	-0.3083510	0.1601489	1.5894317	12.4087968	5.1384774	60.3803675	0.9231039	4.0723888	-0.4716285	0.2870289	-0.7437148	0.8445001
$\sigma_{22}^{(1)} = 1$	-0.2338256	0.0950454	1.1536539	21.2873192	5.4257687	86.1945629	-0.3387699	0.3475011	-0.3352546	0.1497062	-0.7598093	0.8610679
$\sigma_{11}^{(2)} = 1$	-0.2705701	0.1335236	2.8210953	22.7873662	5.3803574	100.3130953	1.7025230	18.2761071	-0.5119307	0.3238035	-0.8585325	0.9257289
$\sigma_{22}^{(2)} = 1$	-0.2093825	0.0731886	1.2462112	16.6731973	6.2585351	126.5031056	-0.3224662	0.6785015	-0.3387752	0.1564892	-0.8924273	0.8357126
$\pi_1 = 0.3$	0.0302198	0.0061724	0.0943104	0.0324999	0.0677582	0.0335767	0.1127432	0.0214978	-0.0539225	0.0324204	0.0066266	0.0060542

Table B.2: Simulation results for scenario (a) with well separated components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0041325	0.0535254	-0.0824589	0.7822125	-0.2884425	1.3522157	-0.1963315	0.8798790	-0.0996886	0.2059567	0.3352902	0.4013598
$\mu_{21} = 0$	0.0108103	0.0243034	0.1879409	1.2687050	0.1909383	1.4872379	0.0562391	0.9817053	0.0596058	0.1153524	-0.4533453	0.7617172
$\mu_{12} = -3$	0.0272135	0.1671829	0.7991876	2.0887988	0.3569542	1.7950629	0.5829168	2.0903330	2.9293863	16.3627691	2.4997943	6.7874435
$\mu_{22} = 3$	-0.0041452	0.0247694	-0.0366616	0.9290418	0.2242026	1.2164697	0.1287342	0.8813220	0.1111648	0.0980046	-2.6709318	7.4304515
$\sigma_{11}^{(1)} = 1$	-0.2469456	0.1383582	1.5682932	13.3486262	5.3735378	75.0728470	1.3399721	9.1009279	-0.4685336	0.3190576	-0.6171896	0.9945345
$\sigma_{22}^{(1)} = 1$	-0.1914298	0.0772335	1.3458442	31.4605113	5.2138949	88.2104708	-0.1998371	3.5906152	-0.3955882	0.2218978	-0.7336922	0.8529997
$\sigma_{11}^{(2)} = 1$	-0.1912358	0.1393324	2.7542126	27.5115759	4.4026736	69.2890834	1.8810944	21.8064291	-0.4873966	0.3880183	-0.8004259	0.8759356
$\sigma_{22}^{(2)} = 1$	-0.1762000	0.0713533	1.6021138	31.5862892	5.2756312	97.5456944	-0.2092953	2.7753060	-0.3464476	0.2250561	-0.8596806	0.8309824
$\pi_1 = 0.3$	0.0131158	0.0030196	0.0911992	0.0351708	0.0573726	0.0286143	0.1193373	0.0275957	-0.0094567	0.0426556	0.0057826	0.0110617

Table B.3: Simulation results for scenario (a) with well separated components and sample size n = 100.

Results for partially overlapped components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0424117	0.0370221	0.1051676	0.2245852	-0.0460531	0.3096113	-0.1236195	0.0361789	-0.1984278	0.1798965	0.1773176	0.1174302
$\mu_{21} = 0$	0.0079576	0.0134059	0.0368978	0.1469863	0.1322288	0.4084846	0.0582494	0.0331812	0.0570930	0.0543937	-0.3223126	0.3431143
$\mu_{12} = -1$	0.1239648	0.2637266	0.8069544	1.2516055	2.5355392	6.6834962	0.5139554	0.8734463	1.8335961	6.2318139	0.6786182	0.6672640
$\mu_{22} = 3$	-0.0050162	0.0102136	-0.5403176	0.4315766	-0.7161834	0.8436621	0.0745000	0.0282889	0.0923107	0.0638483	-2.8117533	7.9671874
$\sigma_{11}^{(1)} = 1$	-0.2860096	0.1195602	3.3438809	43.2138719	8.5636019	238.3406453	0.1010152	0.1187301	-0.5086561	0.3013826	-0.0892457	2.6082111
$\sigma_{22}^{(1)} = 1$	-0.1911643	0.0507184	1.5862061	32.8960305	18.7835023	556.5820056	-0.4071860	0.2182765	-0.3504029	0.1488776	-0.3463394	2.1613538
$\sigma_{11}^{(2)} = 1$	-0.2300115	0.0986439	3.7431714	43.0024026	11.5252382	273.1946482	0.8321976	2.5605001	-0.4460883	0.2718411	-0.3048542	1.3120134
$\sigma_{22}^{(2)} = 1$	-0.1653914	0.0545280	2.1351734	34.0686535	20.8194797	617.0326643	-0.4228749	0.2282076	-0.3285712	0.1506753	-0.5141782	1.1519307
$\pi_1 = 0.3$	0.0195003	0.0018904	0.0095798	0.0188341	0.3168588	0.1964637	0.1622371	0.0547702	-0.0374643	0.0280762	0.0895554	0.0414498

Table B.4: Simulation results for scenario (a) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.1373625	1.2334699	0.0221596	0.7480685	-0.1824584	1.2959743	-0.1369975	0.2492206	-0.1917299	0.2326567	0.2551775	0.3900352
$\mu_{21} = 0$	0.1759528	2.5256329	0.1003426	0.7628399	0.1399818	1.4761554	0.1153560	0.4885637	0.0797435	0.1033508	-0.2302945	0.4751183
$\mu_{12} = -1$	0.2824853	0.8339904	0.7642523	1.3575059	2.0543326	5.4000661	0.5806254	1.1507100	1.9155972	6.5467891	0.7408273	0.8991908
$\mu_{22} = 3$	0.0968680	0.7748858	-0.4305431	0.7529002	-0.3989046	1.3822834	0.1189747	0.4572794	0.1207349	0.1032353	-2.7495069	7.8070551
$\sigma_{11}^{(1)} = 1$	-0.3308766	0.2379312	2.9449870	36.4390940	10.2012602	293.0190824	0.2430847	1.3693376	-0.5385568	0.3491913	0.1410629	3.1905795
$\sigma_{22}^{(1)} = 1$	-0.2300183	0.1011397	1.5833053	33.9359810	14.2051349	435.4443722	-0.3480274	0.9929475	-0.3880477	0.1941794	-0.2139721	2.6557871
$\sigma_{11}^{(2)} = 1$	-0.1784617	0.9656910	3.4948112	41.9125394	12.3600230	314.8116836	1.0721352	6.8156329	-0.4344549	0.3408894	-0.0978007	1.6444604
$\sigma_{22}^{(2)} = 1$	-0.1782934	0.1419368	2.1405921	32.9002003	16.6258358	509.8128096	-0.2784224	5.0828975	-0.3538142	0.2005093	-0.4095947	1.2337453
$\pi_1 = 0.3$	0.0332436	0.0119772	0.0376222	0.0289326	0.2354280	0.1585767	0.1653977	0.0562683	-0.0130978	0.0406157	0.0709602	0.0516311

Table B.5: Simulation results for scenario (a) with partially overlapped components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.1217327	0.3424093	-0.0848093	1.7570912	-0.4906590	3.3682445	-0.2016235	0.5648811	-0.1757096	0.2486751	-0.1353760	0.1953724
$\mu_{21} = 0$	0.0150624	0.2150206	0.2501916	1.6218933	0.3480364	3.4420354	0.1173630	0.9002749	0.0532905	0.1217878	0.0290429	0.0900410
$\mu_{12} = -1$	0.2430972	5.6123614	0.7049292	1.9296563	1.3111672	4.3678248	0.6921974	1.6021225	1.9709080	6.7110000	1.9170979	6.7954106
$\mu_{22} = 3$	0.0117352	0.0754955	-0.2970192	1.1433462	0.0554118	3.3949696	0.1908549	1.1161805	0.1070227	0.1205184	0.0752922	0.1000488
$\sigma_{11}^{(1)} = 1$	-0.3689784	0.2503617	3.3469621	99.0409465	10.1704657	301.9445944	0.5450771	7.6415904	-0.5557620	0.3911848	-0.7099795	0.5592844
$\sigma_{22}^{(1)} = 1$	-0.2158844	0.1118284	2.1233408	50.2963406	10.6454680	336.6016887	-0.2821301	2.0588171	-0.4162402	0.2404544	-0.6388686	0.4440544
$\sigma_{11}^{(2)} = 1$	-0.2537603	0.2926127	3.6062624	59.4579246	10.7111371	281.6722659	1.3474430	9.9347551	-0.3963694	0.4524477	-0.6697443	0.5397530
$\sigma_{22}^{(2)} = 1$	-0.1689193	0.1422386	2.5618490	47.0941013	11.6867835	328.3603332	-0.3075801	0.9512385	-0.3432691	0.2445652	-0.5268288	0.3332719
$\pi_1 = 0.3$	0.0274095	0.0082815	0.0587818	0.0440277	0.1683148	0.1299882	0.1719826	0.0602972	-0.0017699	0.0410446	-0.0411630	0.0403783

Table B.6: Simulation results for scenario (a) with partially overlapped components and sample size n = 100.

B.1 Results for scenario (b)

Results for well separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0523904	0.0175912	0.0081322	0.0068030	-0.0010393	0.0064822	-0.0272350	0.0104453	-0.0396750	0.0261212	0.3001022	0.1881578
$\mu_{21} = 0$	0.0117577	0.0046756	0.0018170	0.0032129	-0.0006972	0.0029395	0.0016514	0.0117386	-0.0010831	0.0091387	-0.7298004	1.0486684
$\mu_{12} = -3$	-0.0585745	0.0169097	0.0009698	0.0066430	0.0010388	0.0063869	0.1595639	0.4127467	0.1657163	0.6405706	2.2939622	5.7846632
$\mu_{22} = 3$	0.0086786	0.0041262	0.0000079	0.0029180	0.0017731	0.0029438	0.0018376	0.0113467	-0.0083445	0.0240661	-2.6849733	7.3060727
$\sigma_{11}^{(1)} = 1$	-0.5251326	0.3024161	-0.2491629	0.0911683	-0.0084125	0.0136248	-0.0337846	0.0303442	-0.0497125	0.0282443	-0.0411751	0.0117111
$\sigma_{22}^{(1)} = 1$	-0.3935743	0.1681852	-0.1660932	0.0404020	-0.0023100	0.0061394	-0.5692947	0.3491047	-0.0171691	0.0164963	-0.0349273	0.0032820
$\sigma_{11}^{(2)} = 1$	-0.4875619	0.2598381	-0.2595159	0.0921189	-0.0074552	0.0154606	0.3601734	2.5206562	0.0869563	1.0872276	-0.8885049	0.7896422
$\sigma_{22}^{(2)} = 1$	-0.3505499	0.1339561	-0.1726553	0.0417569	-0.0028531	0.0054768	-0.5643197	0.3440210	0.0027559	0.1327281	-0.8871348	0.7871045
$\pi_1 = 0.3$	-0.0172722	0.0008719	0.0014646	0.0004183	-0.0008366	0.0004010	0.1202638	0.0203726	0.0011789	0.0085798	0.0003015	0.0004618

Table B.7: Simulation results for scenario (b) with well separated components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0455729	0.0377378	0.0037019	0.0169957	0.0009363	0.0186999	-0.0339249	0.0268527	-0.0316485	0.0557496	0.3200235	0.2020764
$\mu_{21} = 0$	0.0104079	0.0104353	0.0030929	0.0081809	0.0017658	0.0079981	0.0012861	0.0252661	-0.0119759	0.1700182	-0.7022162	1.0230076
$\mu_{12} = -3$	-0.0493157	0.0389216	0.0066693	0.0170638	0.0083988	0.0177755	0.3081342	0.9227010	0.9297510	4.0347304	2.2592544	5.6257306
$\mu_{22} = 3$	0.0121153	0.0101556	0.0085627	0.0075606	-0.0009593	0.0068874	0.0069429	0.0276770	-0.0363564	0.1783975	-2.7047626	7.4171690
$\sigma_{11}^{(1)} = 1$	-0.5450935	0.3291987	-0.2398150	0.1104222	-0.0184718	0.0374325	-0.0610047	0.0676710	-0.0667421	0.0955050	-0.0534927	0.0244499
$\sigma_{22}^{(1)} = 1$	-0.3427160	0.1419901	-0.1630059	0.0521968	-0.0101484	0.0147175	-0.5917637	0.3773129	-0.1353438	0.1577996	-0.0394358	0.0060218
$\sigma_{11}^{(2)} = 1$	-0.5431028	0.3182883	-0.2509139	0.1082647	-0.0155524	0.0380804	0.7085469	5.1943487	1.1464255	9.3742029	-0.8883005	0.7896333
$\sigma_{22}^{(2)} = 1$	-0.2840550	0.1022861	-0.1768103	0.0543394	-0.0105650	0.0147806	-0.5629846	0.3520183	-0.0476952	0.7187962	-0.8877503	0.7883253
$\pi_1 = 0.3$	-0.0194158	0.0017859	-0.0015006	0.0011256	-0.0005175	0.0010008	0.1323124	0.0267621	0.0842837	0.0669623	-0.0007083	0.0010663

Table B.8: Simulation results for scenario (b) with well separated components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0582866	0.0674123	0.0084256	0.0337989	-0.0027792	0.0380174	-0.0507057	0.0560028	-0.0671884	0.1417398	0.2992541	0.1976778
$\mu_{21} = 0$	0.0044360	0.0207349	0.0070702	0.0143312	-0.0048816	0.0143228	-0.0069422	0.0495167	-0.0132220	0.2009707	-0.7531042	1.1040302
$\mu_{12} = -3$	-0.0500111	0.0640024	-0.0002153	0.0348312	0.0082122	0.0357170	0.3911641	1.1433486	1.9266445	9.1789833	2.3095192	5.8604633
$\mu_{22} = 3$	0.0036630	0.0196810	0.0017599	0.0147682	-0.0048640	0.0149518	0.0055019	0.0493550	0.0114686	0.2669684	-2.6814861	7.2973929
$\sigma_{11}^{(1)} = 1$	-0.4671766	0.2826484	-0.2310131	0.1309515	-0.0362605	0.0733687	-0.0928261	0.1132361	-0.2275759	0.2053591	-0.0630655	0.0507171
$\sigma_{22}^{(1)} = 1$	-0.3278169	0.1474229	-0.1750314	0.0705382	-0.0175223	0.0267358	-0.5972663	0.3908425	-0.3041015	0.3009060	-0.0538185	0.0122528
$\sigma_{11}^{(2)} = 1$	-0.4584398	0.2783689	-0.2423612	0.1271893	-0.0256011	0.0835042	0.8955284	6.4516388	1.4467501	12.9893216	-0.8898399	0.7929122
$\sigma_{22}^{(2)} = 1$	-0.2981149	0.1251144	-0.1763062	0.0678424	-0.0137265	0.0293658	-0.5467591	0.3429725	-0.1597559	0.9565928	-0.8881575	0.7893093
$\pi_1 = 0.3$	-0.0145615	0.0027384	-0.0019064	0.0018915	-0.0014898	0.0019666	0.1273638	0.0260902	0.1337906	0.1051307	-0.0000879	0.0019729

Table B.9: Simulation results for scenario (b) with well separated components and sample size n = 100.

Results for partially overlapped components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0106755	0.0089174	0.0780211	0.0142909	-0.0041326	0.0095585	-0.0054390	0.0055033	-0.0021023	0.0100254	0.2837641	0.1698810
$\mu_{21} = 0$	0.0035001	0.0034588	0.0078961	0.0032561	0.0026695	0.0028415	0.1386510	0.1084468	0.0011081	0.0035253	-0.6699872	0.8980480
$\mu_{12} = -1$	0.0144661	0.0112483	-0.0671442	0.0132989	0.0092944	0.0126937	1.5523025	4.2980573	0.0061350	0.0131385	0.3431154	0.5655761
$\mu_{22} = 3$	0.0054232	0.0040738	-0.0394639	0.0054729	0.0037994	0.0039136	0.1377501	0.0857361	-0.0008200	0.0055556	-2.7117457	7.4434819
$\sigma_{11}^{(1)} = 1$	-0.0037534	0.0182359	-0.5411190	0.2996691	0.0011809	0.0167533	0.0122423	0.0112428	-0.0043919	0.0183641	-0.0352154	0.0148811
$\sigma_{22}^{(1)} = 1$	-0.0085514	0.0060539	-0.3122559	0.1034130	-0.0025190	0.0066116	-0.5134664	0.4767323	-0.0059768	0.0066295	-0.0443459	0.0046889
$\sigma_{11}^{(2)} = 1$	0.0153139	0.0281876	-0.5919774	0.3546627	0.0080365	0.0296165	1.9103749	6.5221865	0.0038662	0.0309693	-0.7719776	0.5974512
$\sigma_{22}^{(2)} = 1$	-0.0135670	0.0095890	-0.3266021	0.1125159	-0.0030334	0.0089684	-0.5213739	0.4904526	-0.0003855	0.0224480	-0.7798408	0.6086173
$\pi_1 = 0.3$	0.0011139	0.0005132	-0.0078418	0.0005799	0.0025955	0.0004925	0.3759716	0.2487907	0.0007716	0.0005961	0.0032392	0.0005613

Table B.10: Simulation results for scenario (b) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0031572	0.0209632	0.0747542	0.0269011	0.0017379	0.0230802	-0.0158808	0.0152328	-0.0242013	0.0564460	0.2801950	0.1786643
$\mu_{21} = 0$	0.0019843	0.0077844	0.0135509	0.0079194	0.0057299	0.0084582	0.1030375	0.1105263	0.0095068	0.0310872	-0.6585969	0.8883574
$\mu_{12} = -1$	0.0060818	0.0287462	-0.0635027	0.0273749	0.0005417	0.0321041	1.3098469	3.5671855	0.1412619	0.2998001	0.3378631	0.5747554
$\mu_{22} = 3$	0.0024464	0.0094895	-0.0346547	0.0105104	0.0041287	0.0094155	0.1163228	0.0816585	-0.0270661	0.0516521	-2.7157978	7.4760902
$\sigma_{11}^{(1)} = 1$	-0.0143307	0.0430150	-0.5476304	0.3201459	-0.0150640	0.0447073	0.0210347	0.0318910	-0.0232155	0.0611935	-0.0473897	0.0333908
$\sigma_{22}^{(1)} = 1$	-0.0123656	0.0166080	-0.3244714	0.1190775	-0.0080048	0.0169458	-0.4016957	0.3593049	-0.0356297	0.0317205	-0.0531372	0.0093524
$\sigma_{11}^{(2)} = 1$	-0.0174309	0.0688119	-0.6024462	0.3751095	0.0086232	0.0785699	1.6385791	5.6285979	0.1872048	0.7304868	-0.7723854	0.6008558
$\sigma_{22}^{(2)} = 1$	-0.0056566	0.0228862	-0.3402426	0.1284850	0.0010865	0.0228800	-0.4183479	0.3780218	0.0663827	0.3220906	-0.7825495	0.6136584
$\pi_1 = 0.3$	0.0011584	0.0012101	-0.0072650	0.0012577	0.0007413	0.0014918	0.3017615	0.1844474	0.0135494	0.0127207	0.0044042	0.0014225

Table B.11: Simulation results for scenario (b) with partially overlapped components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0111212	0.0493723	0.0691507	0.0518827	-0.0058708	0.0485664	-0.0284915	0.0402491	-0.0556713	0.1025105	-0.1353760	0.1953724
$\mu_{21} = 0$	0.0174831	0.0164469	0.0194223	0.0183416	0.0085916	0.0180038	0.0936533	0.0970549	-0.0102130	0.1378755	0.0290429	0.0900410
$\mu_{12} = -1$	0.0355405	0.0697176	-0.0492389	0.0559699	0.0104602	0.0749108	1.0019830	2.5972729	0.6205737	1.4809538	1.9170979	6.7954106
$\mu_{22} = 3$	0.0191623	0.0233022	-0.0318386	0.0221104	-0.0015731	0.0206914	0.1065512	0.0862036	-0.0424659	0.1572173	0.0752922	0.1000488
$\sigma_{11}^{(1)} = 1$	-0.0248113	0.0783372	-0.5439787	0.3374480	-0.0221809	0.0840875	0.0124859	0.0659315	-0.0605605	0.1211860	-0.7099795	0.5592844
$\sigma_{22}^{(1)} = 1$	-0.0273691	0.0336066	-0.3446079	0.1489623	-0.0270350	0.0336352	-0.3008118	0.2582892	-0.1226669	0.1204448	-0.6388686	0.4440544
$\sigma_{11}^{(2)} = 1$	0.0488965	0.1757681	-0.5920872	0.3773998	0.0105114	0.1825663	1.3457496	4.6993678	0.7895446	3.0735431	-0.6697443	0.5397530
$\sigma_{22}^{(2)} = 1$	-0.0230507	0.0594265	-0.3437113	0.1557023	-0.0148131	0.0510909	-0.3144878	0.2764307	0.0797694	0.8258204	-0.5268288	0.3332719
$\pi_1 = 0.3$	0.0080310	0.0026880	-0.0035453	0.0026481	0.0038629	0.0028110	0.2180591	0.1148964	0.0908617	0.0597930	-0.0411630	0.0403783

Table B.12: Simulation results for scenario (b) with partially overlapped components and sample size n = 100.

B.2 Results for scenario (c)

Results for well separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0260428	0.0139754	0.0077422	0.0076116	0.0254538	0.0074407	-0.0233925	0.0104423	-0.0165433	0.0356129	0.2785859	0.1489655
$\mu_{21} = 0$	0.0098411	0.0068644	0.0290170	0.0049677	-0.0116893	0.0029316	-0.0014049	0.0115343	0.0109956	0.0140059	-0.6966691	0.9694092
$\mu_{12} = -3$	0.0061889	0.2120334	0.2381494	0.1590184	-0.0518485	0.0093237	0.1024493	0.2352612	2.9404117	17.4829644	2.3060170	0.5749706
$\mu_{22} = 3$	0.0072221	0.0040823	-0.0333902	0.0955529	0.0937980	0.0121939	-0.0013690	0.0120203	0.0205685	0.0184770	-2.7240235	7.4928166
$\sigma_{11}^{(1)} = 1$	-0.4108217	0.1964771	-0.4325642	0.2028285	-0.0287911	0.0153598	-0.0406540	0.0304410	-0.6634545	0.4613750	-0.0597183	0.0136966
$\sigma_{22}^{(1)} = 1$	-0.2550936	0.0788221	-0.2383151	0.0769714	-0.0035891	0.0058473	-0.5574052	0.3358351	-0.5157174	0.2780593	-0.2320579	0.0560750
$\sigma_{11}^{(2)} = 1$	-0.3724357	0.1637375	0.6706892	1.7207269	-0.0954160	0.0218118	0.2255135	1.5909490	-0.6559079	0.4539656	-0.9065110	0.8219016
$\sigma_{22}^{(2)} = 1$	-0.1870412	0.0430421	-0.1914099	0.1144589	1.4362620	2.1168495	-0.5420591	0.3174045	-0.4887539	0.2510987	-0.9116352	0.8311589
$\pi_1 = 0.3$	-0.0159279	0.0013134	0.0231602	0.0031327	-0.0119292	0.0005486	0.1063652	0.0156478	-0.1780186	0.0437827	-0.0065460	0.0004675

Table B.13: Simulation results for scenario (c) with well separated components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0308868	0.0324344	0.0061113	0.0186932	0.2555147	0.1094040	-0.1539296	0.0571333	-0.1207730	0.1587481	0.2629659	0.1483199
$\mu_{21} = 0$	0.0055092	0.0150999	0.0272731	0.0090512	-0.0631240	0.0142519	0.1405245	0.1336555	0.0624581	0.0565042	-0.7260002	1.0114762
$\mu_{12} = -3$	0.0337635	0.3874107	0.2230368	0.1441150	1.6825445	0.1552453	3.5711036	3.8148958	3.8999027	6.8051034	2.3418282	0.6019301
$\mu_{22} = 3$	0.0050658	0.0094208	-0.0253172	0.1384808	-0.1754172	0.1487526	0.1836123	0.3700283	0.1042152	0.0645465	-2.7084957	7.4129110
$\sigma_{11}^{(1)} = 1$	-0.4357989	0.2364255	-0.4280828	0.2136043	-0.1780483	0.1020275	-0.0576847	0.0593778	-0.7285366	0.5621754	-0.0695713	0.0310590
$\sigma_{22}^{(1)} = 1$	-0.2772914	0.1057515	-0.2507775	0.0869654	-0.0100421	0.0158693	-0.7450800	0.5883909	-0.6131186	0.3981044	-0.2273048	0.0576560
$\sigma_{11}^{(2)} = 1$	-0.3781389	0.1909639	0.6397038	1.6648532	-0.2754995	0.1604744	1.9082012	5.7446222	-0.7046223	0.5601073	-0.9069928	0.8230093
$\sigma_{22}^{(2)} = 1$	-0.2005722	0.0614958	-0.1867153	0.1330812	2.0757813	4.7528042	-0.7413390	0.5830453	-0.5532054	0.3367873	-0.9109526	0.8300383
$\pi_1 = 0.3$	-0.0153014	0.0028573	0.0199281	0.0027959	-0.0788485	0.0089187	0.3844052	0.1967171	-0.0775875	0.0306094	-0.0040197	0.0010065

Table B.14: Simulation results for scenario (c) with well separated components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0465735	0.0589915	0.0218378	0.0313172	0.2079795	0.1447606	-0.0427568	0.0472282	-0.0697722	0.1593906	0.2607050	0.1569600
$\mu_{21} = 0$	0.0065921	0.0221059	0.0245362	0.0187909	-0.0463382	0.0190976	0.0041092	0.0480527	0.0358333	0.0706825	-0.7310007	1.0265958
$\mu_{12} = -3$	-0.0339072	0.1514083	0.2097463	0.1604919	1.7275389	3.0780420	0.4275214	1.4702309	2.7295025	15.5581290	2.3473692	5.9991110
$\mu_{22} = 3$	0.0056403	0.0194462	-0.0269431	0.1069069	-0.1709625	0.0768588	0.0800055	0.7053710	0.0640894	0.0644206	-2.7077586	7.4140483
$\sigma_{11}^{(1)} = 1$	-0.4310498	0.2524704	-0.4168720	0.2236804	-0.1359957	0.1959003	-0.1115518	0.1257562	-0.6558799	0.4790587	-0.0932820	0.0664353
$\sigma_{22}^{(1)} = 1$	-0.2812622	0.1190111	-0.2581960	0.0998044	-0.0344439	0.0305249	-0.6054827	0.4028328	-0.6542023	0.4548867	-0.2192016	0.0616372
$\sigma_{11}^{(2)} = 1$	-0.3720718	0.2119943	0.6655420	1.9506709	-0.2466144	0.2721298	0.9134271	8.2238711	-0.6620876	0.4868165	-0.8940632	0.8174377
$\sigma_{22}^{(2)} = 1$	-0.2163547	0.0781019	-0.1880995	0.1735091	2.0235275	4.6840880	-0.5266942	0.3210884	-0.5104381	0.2981506	-0.9092124	0.8278134
$\pi_1 = 0.3$	-0.0159575	0.0032242	0.0182443	0.0623401	-0.0764857	0.0090310	0.1243452	0.0283581	-0.0644798	0.0321608	-0.0037619	0.0019734

Table B.15: Simulation results for scenario (c) with well separated components and sample size n = 100.

Results for partially separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0050064	0.0105713	0.1951078	0.0463683	0.2556306	0.0794778	-0.1034187	0.0286861	-0.0626956	0.0750688	0.2540658	0.1161696
$\mu_{21} = 0$	0.0011538	0.0034929	-0.0148497	0.0044812	-0.0704784	0.0078155	0.2212208	0.1694466	0.0237840	0.0239447	-0.6041864	0.7609487
$\mu_{12} = -1$	0.0129716	0.0120704	0.0422964	0.0238115	-0.3324842	0.1211984	2.2002607	5.5544430	1.8265868	6.8265138	0.3408842	0.5100485
$\mu_{22} = 3$	-0.0006981	0.0042782	-0.2232120	0.0861617	-0.1858181	0.0435836	0.2151450	9.6313171	0.0507315	0.0316584	-2.7669749	7.7072766
$\sigma_{11}^{(1)} = 1$	-0.0858593	0.0538281	-0.7726900	0.6018803	-0.1842383	0.0559836	-0.0698368	0.0309342	-0.7232990	0.5484814	-0.0983138	0.0233996
$\sigma_{22}^{(1)} = 1$	-0.0551203	0.0200092	-0.4032176	0.1668991	0.0086991	0.0055625	-0.8499805	0.7484600	-0.5626771	0.3277626	-0.2170418	0.0514850
$\sigma_{11}^{(2)} = 1$	-0.0542676	0.0454571	-0.0273345	0.6231634	-0.2997707	0.1045900	2.4361194	7.3028933	-0.7092960	0.5387288	-0.8281386	0.6867939
$\sigma_{22}^{(2)} = 1$	-0.0351582	0.0140596	-0.1893121	0.0488364	2.0358350	4.2610151	-0.8544223	0.7738874	-0.5336169	0.2993700	-0.8362320	0.6995904
$\pi_1 = 0.3$	-0.0044453	0.0006829	-0.0101196	0.0013891	-0.0815029	0.0072110	0.5323645	0.1942151	-0.1383659	0.0353259	-0.0127042	0.0006557

Table B.16: Simulation results for scenario (c) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0007457	0.0473278	0.1875568	0.0553022	0.2555147	0.1094040	-0.1539296	0.0571333	-0.1207730	0.1587481	0.2379499	0.1241031
$\mu_{21} = 0$	0.0052598	0.0134580	-0.0136229	0.0105858	-0.0631240	0.0142519	0.1405245	0.1336555	0.0624581	0.0565042	-0.6184958	0.7772891
$\mu_{12} = -1$	0.0693351	4.4101577	0.0463612	0.0520940	-0.3174555	0.1552453	1.5711036	3.8148958	1.8999027	6.8051034	0.3816927	0.5433389
$\mu_{22} = 3$	-0.0079315	0.0158799	-0.2146315	0.1893088	-0.1754172	0.1487526	0.1836123	0.3700283	0.1042152	0.0645465	-2.7537815	7.6456545
$\sigma_{11}^{(1)} = 1$	-0.4786085	0.2927100	-0.7644609	0.5960355	-0.1780483	0.1020275	-0.0576847	0.0593778	-0.7285366	0.5621754	-0.1371543	0.0530398
$\sigma_{22}^{(1)} = 1$	-0.2872298	0.1141352	-0.4030729	0.1768544	-0.0100421	0.0158693	-0.7450800	0.5883909	-0.6131186	0.3981044	-0.2049912	0.0527396
$\sigma_{11}^{(2)} = 1$	-0.3268201	0.2101050	-0.0241063	0.7727319	-0.2754995	0.1604744	1.9082012	5.7446222	-0.7046223	0.5601073	-0.8021667	0.7281557
$\sigma_{22}^{(2)} = 1$	-0.1786598	0.0593833	-0.1780742	0.0583799	2.0757813	4.7528042	-0.7413390	0.5830453	-0.5532054	0.3367873	-0.8322031	0.6946053
$\pi_1 = 0.3$	-0.0197413	0.0029626	-0.0085607	0.0025710	-0.0788485	0.0089187	0.3844052	0.1967171	-0.0775875	0.0306094	-0.0056393	0.0019438

Table B.17: Simulation results for scenario (c) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.1217327	0.3424093	0.1694477	0.0735184	0.2079795	0.1447606	-0.1671211	0.0923151	-0.1282267	0.2111417	0.2349404	0.1589278
$\mu_{21} = 0$	0.0150624	0.2150206	-0.0052920	0.0185982	-0.0463382	0.0190976	0.1055003	0.1145325	0.0439545	0.0855435	-0.6130445	0.7627412
$\mu_{12} = -1$	0.2430972	5.6123614	0.0607839	0.0781125	-0.2724611	0.1678865	1.1666014	2.5906847	1.8715418	6.6171551	0.4516950	0.6023874
$\mu_{22} = 3$	0.0117352	0.0754955	-0.2133121	0.1112804	-0.1709625	0.0768588	0.2009732	0.9133126	0.1050799	0.0859778	-2.7322620	7.5534300
$\sigma_{11}^{(1)} = 1$	-0.3689784	0.2503617	-0.7340855	0.5600108	-0.1359957	0.1959003	-0.1244170	0.1258824	-0.7198621	0.5624173	-0.1833282	0.1023982
$\sigma_{22}^{(1)} = 1$	-0.2158844	0.1118284	-0.4082106	0.1904212	-0.0344439	0.0305249	-0.6928704	0.5154260	-0.6663228	0.4724465	-0.2252852	0.0751256
$\sigma_{11}^{(2)} = 1$	-0.2537603	0.2926127	0.0452558	1.0730323	-0.2466144	0.2721298	1.4262727	4.3581118	-0.6765046	0.5542579	-0.7205595	0.7472357
$\sigma_{22}^{(2)} = 1$	-0.1689193	0.1422386	-0.1798496	0.0854610	2.0235275	4.6840880	-0.6569214	0.4721772	-0.5308087	0.3289451	-0.8154226	0.6747523
$\pi_1 = 0.3$	0.0274095	0.0082815	-0.0087817	0.0035286	-0.0764857	0.0090310	0.2829616	0.1192610	-0.0358169	0.0279864	0.0080474	0.0052697

Table B.18: Simulation results for scenario (c) with partially overlapped components and sample size n = 100.

B.3 Results for scenario (d)

B.3.1 Results for well separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0210139	0.3740896	0.0963055	0.1988129	-0.0724993	0.4445397	-0.1776877	0.2486331	-0.1370771	0.1203035	0.2614873	0.1465963
$\mu_{21} = 0$	-0.0023015	0.1668076	-0.0025194	0.2166969	0.0351004	0.1862772	0.1605836	0.1440432	0.1846574	0.2502978	-0.5376732	0.5967841
$\mu_{12} = -3$	2.0487159	0.2315514	2.0177060	0.0610856	2.7686202	0.8374607	3.6090077	4.0708474	3.6022516	4.0906072	2.4600724	0.5172285
$\mu_{22} = 3$	-0.0163129	0.0221396	-0.1596680	0.0686085	0.0142246	0.0846006	0.1752517	0.0964531	0.1686227	0.1177871	-2.7373684	7.5717983
$\sigma_{11}^{(1)} = 1$	-0.2128782	2.7268934	0.8263929	6.8203606	4.4724025	31.0649947	0.0009992	0.0427987	0.0155108	0.0640331	-0.6558897	0.4325649
$\sigma_{22}^{(1)} = 1$	-0.1720220	0.0417675	-0.2670580	0.1347234	0.0269321	2.6960427	-0.6951581	0.5536820	-0.6930684	0.5596335	-0.6382926	0.4100375
$\sigma_{11}^{(2)} = 1$	-0.1448088	3.3844125	0.8470228	7.7438461	6.2840685	54.7351720	1.9982187	8.4002770	1.9132680	5.3252856	-0.8330814	0.6947570
$\sigma_{22}^{(2)} = 1$	-0.1154482	0.0827906	-0.1372394	0.5534311	0.1863750	3.0208575	-0.7037965	0.5627381	-0.7009686	0.5618837	-0.8383335	0.7032107
$\pi_1 = 0.3$	-0.0024474	0.0025337	0.0191747	0.0042977	0.1020248	0.0171178	0.3816330	0.2210282	0.3838840	0.2236962	0.0454264	0.0025621

Table B.19: Simulation results for scenario (d) with well separated components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0219499	0.2109601	0.0058742	0.2948392	-0.0501578	0.4943390	-0.0187564	0.1289489	-0.0424182	0.3498993	0.3045998	1.845446e-01
$\mu_{21} = 0$	0.0048591	0.0096174	-0.0198427	0.3021911	-0.0005583	0.2415569	0.0031194	0.1429256	-0.0055904	0.2157599	-0.6364462	8.434147e-01
$\mu_{12} = -3$	-0.0008454	0.1291718	0.2764906	0.4367348	0.0699355	0.2758241	0.1650754	0.4583010	0.2189478	0.6587435	2.3277332	5.855984e+00
$\mu_{22} = 3$	0.0031454	0.0084196	-0.0422119	0.0972438	0.0674940	0.0951926	0.0061570	0.0528422	0.0151340	0.0376999	496.4619986	3.298078e+05
$\sigma_{11}^{(1)} = 1$	-0.3138324	0.1394170	1.0597694	8.6430104	3.1657065	14.4781408	0.3742312	1.3647371	0.3489268	1.2884606	-0.6580995	4.356703e-01
$\sigma_{22}^{(1)} = 1$	-0.2141798	0.0661497	-0.0609073	2.2735947	0.7829054	1.2033415	-0.5098141	0.3074774	-0.5118486	0.3005148	-0.6406984	4.130624e-01
$\sigma_{11}^{(2)} = 1$	-0.3004276	0.1896887	1.3164268	10.8689185	2.0720374	15.3584344	0.6306956	4.1451650	0.6985601	5.0719223	-0.9085772	8.257287e-01
$\sigma_{22}^{(2)} = 1$	-0.2060830	0.0615536	0.0496417	2.4865398	0.4369835	1.5668574	-0.4867574	0.5044550	-0.4614835	0.9309619	-0.9065194	8.218816e-01
$\pi_1 = 0.3$	-0.0060228	0.0015637	0.0323916	0.0081249	0.0216684	0.0046322	0.1005689	0.0157907	0.1042110	0.0188642	0.0298800	1.468300e-03

Table B.20: Simulation results for scenario (d) with well separated components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0258370	0.1843475	0.0138815	0.2811421	-0.0214366	0.2388048	-0.0538915	0.4338850	-0.0565158	0.4284314	0.2993924	0.1877535
$\mu_{21} = 0$	0.0042724	0.0660930	0.0116096	0.1817466	0.0027457	0.0306944	0.0080715	0.1946304	0.0007288	0.2902689	-0.6629062	0.8975190
$\mu_{12} = -3$	-0.0084907	0.2563048	0.1948396	0.3093027	-0.0674117	0.1266359	0.2998881	0.8644226	0.3163721	1.1627490	2.3404064	5.9354253
$\mu_{22} = 3$	0.0087638	0.0427556	-0.0381950	0.0709461	0.0709091	0.0335278	0.0243220	0.0596031	0.0272592	0.1322402	-2.6965814	7.3686339
$\sigma_{11}^{(1)} = 1$	-0.2960430	0.1560360	0.7290186	5.7455011	2.6700947	11.8152817	0.6140585	3.0554444	0.7160472	3.3517457	-0.6127886	0.4131690
$\sigma_{22}^{(1)} = 1$	-0.2195444	0.0873230	-0.0118928	4.6010379	0.9826721	1.9044996	-0.5268502	0.3187241	-0.5168514	0.3091389	-0.6337343	0.4160776
$\sigma_{11}^{(2)} = 1$	-0.2910664	0.2429983	0.8896350	6.3088540	1.3361658	6.3731744	1.0691981	7.9172798	1.1170074	8.0426709	-0.9090767	0.8280591
$\sigma_{22}^{(2)} = 1$	-0.2027648	0.1291049	-0.0511654	1.0735128	0.4380011	0.5103282	-0.4951348	0.2984964	-0.4949052	0.3011944	-0.9100358	0.8286883
$\pi_1 = 0.3$	-0.0022187	0.0042265	0.0275589	0.0068791	0.0146825	0.0027714	0.1059653	0.0187060	0.1062130	0.0202234	0.0162304	0.0026742

Table B.21: Simulation results for scenario (d) with well separated components and sample size n = 100.

Results for partially separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0210139	0.3740896	0.0963055	0.1988129	-0.0724993	0.4445397	-0.1776877	0.2486331	-0.1370771	0.1203035	0.2614873	0.1465963
$\mu_{21} = 0$	-0.0023015	0.1668076	-0.0025194	0.2166969	0.0351004	0.1862772	0.1605836	0.1440432	0.1846574	0.2502978	-0.5376732	0.5967841
$\mu_{12} = -1$	0.0487159	0.2315514	0.0177060	0.0610856	0.7686202	0.8374607	1.6090077	4.0708474	1.6022516	4.0906072	0.4600724	0.5172285
$\mu_{22} = 3$	-0.0163129	0.0221396	-0.1596680	0.0686085	0.0142246	0.0846006	0.1752517	0.0964531	0.1686227	0.1177871	-2.7373684	7.5717983
$\sigma_{11}^{(1)} = 1$	-0.2128782	2.7268934	0.8263929	6.8203606	4.4724025	31.0649947	0.0009992	0.0427987	0.0155108	0.0640331	-0.6558897	0.4325649
$\sigma_{22}^{(1)} = 1$	-0.1720220	0.0417675	-0.2670580	0.1347234	0.0269321	2.6960427	-0.6951581	0.5536820	-0.6930684	0.5596335	-0.6382926	0.4100375
$\sigma_{11}^{(2)} = 1$	-0.1448088	3.3844125	0.8470228	7.7438461	6.2840685	54.7351720	1.9982187	8.4002770	1.9132680	5.3252856	-0.8330814	0.6947570
$\sigma_{22}^{(2)} = 1$	-0.1154482	0.0827906	-0.1372394	0.5534311	0.1863750	3.0208575	-0.7037965	0.5627381	-0.7009686	0.5618837	-0.8383335	0.7032107
$\pi_1 = 0.3$	-0.0024474	0.0025337	0.0191747	0.0042977	0.1020248	0.0171178	0.3816330	0.2210282	0.3838840	0.2236962	0.0454264	0.0025621

Table B.22: Simulation results for scenario (d) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0369317	0.4650493	0.0834215	0.4619042	-0.1136500	0.8253892	-0.1839537	0.2066609	-0.1966533	0.3510159	0.2779668	0.1627126
$\mu_{21} = 0$	-0.0191546	0.3356610	-0.0092065	0.1271325	0.0219925	0.2467959	0.1121369	0.3650270	0.1517230	0.3607011	-0.5296189	0.6017390
$\mu_{12} = -1$	0.1269186	0.4887688	-0.0006669	0.1900428	0.4601102	0.7288414	1.3247692	3.1277468	1.3321180	3.2244933	0.4301596	0.5441808
$\mu_{22} = 3$	0.0104925	0.1173855	-0.1932359	0.0918804	-0.0000414	0.1313689	0.1509985	0.1114732	0.1646370	0.1876276	-2.7386379	7.6231756
$\sigma_{11}^{(1)} = 1$	-0.1636192	8.7223799	0.9811589	10.8026412	4.6070054	48.3826871	0.0212394	0.1086416	0.0345508	0.3131001	-0.6127420	0.7053472
$\sigma_{22}^{(1)} = 1$	-0.1600104	3.8855462	-0.2112005	0.5167836	0.9080375	23.6319008	-0.6541804	0.5337273	-0.6678442	0.4960648	-0.6349073	0.4106212
$\sigma_{11}^{(2)} = 1$	-0.0735321	5.6692272	0.9854440	11.0272156	5.2394444	50.7041820	1.6832085	4.6320340	1.7742903	5.2875727	-0.8156065	0.6933946
$\sigma_{22}^{(2)} = 1$	-0.1204357	1.2490516	-0.0236779	1.2661828	0.6727319	8.6268129	-0.6271134	0.8692261	-0.6718058	0.5012934	-0.8305614	0.7303644
$\pi_1 = 0.3$	0.0001612	0.0064032	0.0078897	0.0061958	0.0708816	0.0217632	0.3181738	0.1562014	0.3205392	0.1582784	0.0374787	0.0038954

Table B.23: Simulation results for scenario (d) with partially overlapped components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0287354	0.5646453	0.1176565	0.3337819	-0.0540947	1.1249191	-0.2022527	0.4866851	-0.1559768	0.6461948	0.2829646	0.3683158
$\mu_{21} = 0$	-0.0040346	0.1937367	0.0281257	0.2827156	0.0021178	0.5671911	0.0774680	0.5302680	0.0406859	0.5427673	-0.5533686	0.7419038
$\mu_{12} = -1$	0.1585430	0.6660500	-0.0526656	0.4861615	0.1663243	1.1598684	1.0412025	2.5465180	0.9710964	2.5098122	0.3558672	0.6929170
$\mu_{22} = 3$	0.0057648	0.1048312	-0.1796223	0.1368956	0.0110899	0.3055844	0.1400780	0.2502754	0.1165376	0.1963951	-2.7332169	7.7847024
$\sigma_{11}^{(1)} = 1$	-0.0133667	19.8150488	0.8502892	9.6173282	4.6935148	73.2055720	-0.0035054	0.3968943	0.0384610	0.5951526	-0.5664173	0.6099264
$\sigma_{22}^{(1)} = 1$	-0.1459812	6.7250312	-0.1903233	2.0313383	1.6186234	39.3100916	-0.5985096	0.5111043	-0.6139848	0.4305498	-0.6261340	0.4225411
$\sigma_{11}^{(2)} = 1$	0.0361966	9.9818909	0.8039370	8.9243216	4.4948964	60.6548189	1.3473701	3.9278648	1.3207514	3.9908705	-0.7364586	0.9252704
$\sigma_{22}^{(2)} = 1$	-0.1500297	0.1938260	-0.0544542	0.7774401	0.9305212	9.5733200	-0.6114835	0.4427536	-0.5930152	0.4195381	-0.7617572	0.9459782
$\pi_1 = 0.3$	0.0006141	0.0072694	0.0034630	0.0089361	0.0512567	0.0328395	0.2527826	0.1037175	0.2387884	0.0967542	0.0354057	0.0142314

Table B.24: Simulation results for scenario (d) with partially overlapped components and sample size n = 100.

B.4 Results for scenario (e)

Results for well separated components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0060471	0.0344949	0.0983043	0.0211450	-0.1029291	0.0315755	-0.0837747	0.0199954	-0.0797996	0.0665444	0.2616083	0.1506494
$\mu_{21} = 0$	0.0132220	0.0063877	0.0245221	0.0046476	0.0574335	0.0083315	0.1963183	0.1465490	0.0147389	0.0166834	-0.6512996	0.8302804
$\mu_{12} = -3$	2.0293298	0.0411247	1.9158909	0.0177659	2.1498537	0.0717232	4.6283696	7.4386725	3.9685491	7.4577328	2.3901306	0.5605601
$\mu_{22} = 3$	0.0012373	0.0071818	-0.0574924	0.0079205	0.0724607	0.0125606	0.2076667	0.1271959	0.0156979	0.0179984	-2.7188935	7.4747636
$\sigma_{11}^{(1)} = 1$	-0.7514158	0.5998033	-0.4593465	0.2394055	1.1467305	10.7792138	0.0869935	0.0636674	-0.7076403	0.5270730	-0.2764628	0.1051452
$\sigma_{22}^{(1)} = 1$	-0.3993173	0.2182821	-0.3558723	0.1363719	0.1571238	0.0551478	-0.8949246	0.8188710	-0.5414400	0.3027939	-0.2970577	0.0974401
$\sigma_{11}^{(2)} = 1$	-0.6660005	0.5124189	-0.5421991	0.3110750	1.1453735	2.7054359	2.9295256	9.3959560	-0.6852136	0.4997322	-0.7811970	0.6127788
$\sigma_{22}^{(2)} = 1$	-0.3601700	0.1697975	-0.3710533	0.1461810	0.1294598	0.0622467	-0.9047942	0.8343070	-0.5218269	0.2863080	-0.7886417	0.6227380
$\pi_1 = 0.3$	-0.0269684	0.0017357	-0.0096972	0.0007223	0.0247586	0.0018654	0.6065043	0.3897739	-0.1756965	0.0415565	0.0040496	0.0006733

Table B.25: Simulation results for scenario (e) with well separated components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0540133	0.0400524	0.0236826	0.0249990	-0.0058303	0.0349378	-0.0718465	0.0406449	-0.0312346	0.0774533	0.2904864	0.1812209
$\mu_{21} = 0$	0.0106118	0.0125447	0.0120752	0.0084101	0.0086461	0.0102640	0.0132337	0.0339214	0.0181739	0.0376563	-0.7156718	1.0147333
$\mu_{12} = -3$	-0.0516878	0.0397607	0.0059685	0.0290415	-0.0040223	0.0396970	0.9716312	3.1136868	3.0644057	18.0144832	2.3142903	5.8574780
$\mu_{22} = 3$	-0.0009863	0.0139178	-0.0067722	0.0105461	0.0168960	0.0099386	0.0282085	0.0307681	0.0348319	0.0341701	-2.6998103	7.3897672
$\sigma_{11}^{(1)} = 1$	-0.5386905	0.3392090	-0.2933823	0.1643186	0.9378852	1.8447644	0.3356138	0.4669462	-0.6380661	0.4379524	-0.1939022	0.1173581
$\sigma_{22}^{(1)} = 1$	-0.3593544	0.1627394	-0.2720269	0.1055809	0.2456680	0.1349933	-0.6111046	0.4024386	-0.5461805	0.3197566	-0.2625288	0.0891913
$\sigma_{11}^{(2)} = 1$	-0.5075315	0.3021147	-0.3188959	0.1946761	0.8984405	3.1709920	2.6728542	20.9233847	-0.6395537	0.4383434	-0.8723019	0.7628030
$\sigma_{22}^{(2)} = 1$	-0.3062220	0.1593731	-0.2778537	0.1280865	0.1950701	0.0758922	-0.6067797	0.3982640	-0.4881230	0.2654358	-0.8819644	0.7783013
$\pi_1 = 0.3$	-0.0240638	0.0019937	-0.0000375	0.0011625	0.0025499	0.0011499	0.1880286	0.0565536	-0.1463409	0.0437796	-0.0006554	0.0010782

Table B.26: Simulation results for scenario (e) with well separated components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0654854	0.0760471	0.0291276	0.0504505	0.0041291	0.0629525	-0.0741240	0.0707887	-0.0664427	0.1419253	0.2854007	0.1909059
$\mu_{21} = 0$	0.0089269	0.0231069	0.0154098	0.0195728	0.0015865	0.0217014	0.0193146	0.0561817	0.0273420	0.0946584	-0.7201518	1.0308830
$\mu_{12} = -3$	-0.0634003	0.1001592	0.0072635	0.0542123	-0.0028976	0.0776103	0.9292826	2.9457795	2.9409250	16.8984414	2.3115253	5.8592984
$\mu_{22} = 3$	-0.0017477	0.0216603	0.0019144	0.0181135	0.0078786	0.0203152	0.0362637	0.0550056	0.0865268	0.1001363	-2.7003601	7.4017641
$\sigma_{11}^{(1)} = 1$	-0.4635620	0.3031651	-0.2289377	0.2544521	0.8512597	1.8350923	0.3409249	0.7055566	-0.6393042	0.4566218	-0.0806335	0.1501102
$\sigma_{22}^{(1)} = 1$	-0.3230002	0.1577617	-0.2422657	0.1153371	0.2571878	0.1969924	-0.5815293	0.3792404	-0.6151954	0.4158007	-0.2365160	0.0886920
$\sigma_{11}^{(2)} = 1$	-0.4280108	0.2755257	-0.2606702	0.2668199	0.8023023	2.1790315	2.5777973	19.2382574	-0.6470624	0.4718023	-0.8480423	0.7279720
$\sigma_{22}^{(2)} = 1$	-0.2754677	0.1247892	-0.2686309	0.1200411	0.1829756	0.1119165	-0.5591455	0.3589983	-0.5091334	0.3076259	-0.8798671	0.7751422
$\pi_1 = 0.3$	-0.0221787	0.0032747	0.0019382	0.0023612	0.0005988	0.0023112	0.1713397	0.0493995	-0.0602888	0.0507415	-0.0014204	0.0022210

Table B.27: Simulation results for scenario (e) with well separated components and sample size n = 100.

Results for partially overlapped components

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0060471	0.0344949	0.0983043	0.0211450	-0.1029291	0.0315755	-0.0837747	0.0199954	-0.0797996	0.0665444	0.2616083	0.1506494
$\mu_{21} = 0$	0.0132220	0.0063877	0.0245221	0.0046476	0.0574335	0.0083315	0.1963183	0.1465490	0.0147389	0.0166834	-0.6512996	0.8302804
$\mu_{12} = -1$	0.0293298	0.0411247	-0.0841091	0.0177659	0.1498537	0.0717232	2.6283696	7.4386725	1.9685491	7.4577328	0.3901306	0.5605601
$\mu_{22} = 3$	0.0012373	0.0071818	-0.0574924	0.0079205	0.0724607	0.0125606	0.2076667	0.1271959	0.0156979	0.0179984	-2.7188935	7.4747636
$\sigma_{11}^{(1)} = 1$	-0.7514158	0.5998033	-0.4593465	0.2394055	1.1467305	10.7792138	0.0869935	0.0636674	-0.7076403	0.5270730	-0.2764628	0.1051452
$\sigma_{22}^{(1)} = 1$	-0.3993173	0.2182821	-0.3558723	0.1363719	0.1571238	0.0551478	-0.8949246	0.8188710	-0.5414400	0.3027939	-0.2970577	0.0974401
$\sigma_{11}^{(2)} = 1$	-0.6660005	0.5124189	-0.5421991	0.3110750	1.1453735	2.7054359	2.9295256	9.3959560	-0.6852136	0.4997322	-0.7811970	0.6127788
$\sigma_{22}^{(2)} = 1$	-0.3601700	0.1697975	-0.3710533	0.1461810	0.1294598	0.0622467	-0.9047942	0.8343070	-0.5218269	0.2863080	-0.7886417	0.6227380
$\pi_1 = 0.3$	-0.0269684	0.0017357	-0.0096972	0.0007223	0.0247586	0.0018654	0.6065043	0.3897739	-0.1756965	0.0415565	0.0040496	0.0006733

Table B.28: Simulation results for scenario (e) with partially overlapped components and sample size n = 500.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	0.0019489	0.0520772	0.1043447	0.0397729	-0.0782659	0.0579529	-0.1527953	0.0622532	-0.1246923	0.1334549	0.2708891	0.1665980
$\mu_{21} = 0$	0.0112435	0.0138172	0.0210965	0.0100677	0.0370151	0.0617837	0.1228177	0.1214299	0.0278260	0.0494748	-0.6250334	0.8155103
$\mu_{12} = -1$	0.0421815	0.0751604	-0.0858658	0.0365632	0.1053786	0.1110366	1.9743915	5.0857062	1.8627651	6.6428555	0.3632743	0.5506701
$\mu_{22} = 3$	-0.0096788	0.0146613	-0.0549085	0.0140329	0.0526455	0.0163000	0.1361475	0.1041958	0.0595668	0.0486378	-2.7381732	7.5911181
$\sigma_{11}^{(1)} = 1$	-0.6125207	0.4271949	-0.4773855	0.2871713	0.9537432	1.5987956	0.1158339	0.1261761	-0.7225514	0.5558014	-0.1629830	0.1154533
$\sigma_{22}^{(1)} = 1$	-0.4224130	0.2123577	-0.3598558	0.1533634	0.2890520	10.3252122	-0.7705836	0.6299126	-0.5938462	0.3748171	-0.2654831	0.0903572
$\sigma_{11}^{(2)} = 1$	-0.4877528	0.3554450	-0.5593725	0.3490985	0.9698035	1.9408717	2.4493288	7.9031647	-0.6743248	0.5278856	-0.7453552	0.5670658
$\sigma_{22}^{(2)} = 1$	-0.3635215	0.1670572	-0.3769458	0.1621409	0.1588380	0.3695681	-0.7875331	0.6518294	-0.5363923	0.3171555	-0.7839895	0.6164469
$\pi_1 = 0.3$	-0.0278783	0.0026523	-0.0105223	0.0016213	0.0198755	0.0031174	0.4524068	0.2488730	-0.1008510	0.0355022	0.0036236	0.0017304

Table B.29: Simulation results for scenario (e) with partially overlapped components and sample size n = 200.

	RSPGMM		CNM		GMM		SPCGMM		SPGMM		MT	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\mu_{11} = 0$	-0.0033527	0.1043697	0.0735965	0.0956291	-0.0542638	0.0957797	-0.1727151	0.1016749	-0.1353760	0.1953724	0.2807926	0.1945358
$\mu_{21} = 0$	0.0042753	0.0279047	0.0246708	0.0236560	0.0336944	0.0264631	0.0936083	0.1078511	0.0290429	0.0900410	-0.5743128	0.7526875
$\mu_{12} = -1$	0.0556037	0.1630369	-0.0617532	0.0911228	0.0649814	0.1470506	1.4552682	3.4461454	1.9170979	6.7954106	0.3464670	0.5476960
$\mu_{22} = 3$	-0.0089672	0.0292058	-0.0633122	0.0335379	0.0499986	0.0316324	0.1081746	0.0958297	0.0752922	0.1000488	-2.7516950	7.6858069
$\sigma_{11}^{(1)} = 1$	-0.5445853	0.3860574	-0.4392144	0.3557475	0.8277314	1.8089644	0.1118904	0.2609287	-0.7099795	0.5592844	-0.0556095	0.1627080
$\sigma_{22}^{(1)} = 1$	-0.3804203	0.1941473	-0.3647610	0.1742323	0.2282898	0.2131957	-0.6973972	0.5255043	-0.6388686	0.4440544	-0.2530784	0.0986142
$\sigma_{11}^{(2)} = 1$	-0.3935344	0.3373814	-0.5078070	0.4937007	0.8635268	2.3823580	1.9475219	6.3941232	-0.6697443	0.5397530	-0.6962241	0.5148469
$\sigma_{22}^{(2)} = 1$	-0.3246795	0.1585607	-0.3556969	0.1814564	0.1612450	0.1922791	-0.6799338	0.5150042	-0.5268288	0.3332719	-0.7812435	0.6144969
$\pi_1 = 0.3$	-0.0211242	0.0040642	-0.0081736	0.0030121	0.0152877	0.0044242	0.3274992	0.1488794	-0.0411630	0.0403783	0.0123459	0.0040631

Table B.30: Simulation results for scenario (e) with partially overlapped components and sample size n = 100.