



Swiss Federal Institute of Technology Zurich

Seminar for
Statistics

Department of Mathematics

Bachelor Thesis

Winter 2019

Nicolas Trutmann

Comparison of EM-algorithm and MLE using
Cholesky decomposition

Submission Date: placeholder

Advisor: placeholder

Abstract

The intent of this work is to compare The EM algorithm to a MLE approach in the case of multivariate normal mixture models using the Cholesky decomposition. The EM algorithm is widely used in statistics and is proven to converge, however in pathological cases convergence slows down considerably.

methods(not done)

results(not done)

Contents

1	Introduction to normal mixture models	1
1.1	choice of notation	1
1.2	problems of EM	4
2	The norMmix Package	7
2.1	Theoretical Groundwork	7
2.2	finer details of norMmix package	7
3	Comparing Algorithms	9
4	Coclusions	11
	Bibliography	12

List of Figures

List of Tables

1.1	Table of Parameters	3
-----	-------------------------------	-------------------

Chapter 1

Introduction to normal mixture models

1.1 Definitions

here intro to normal mixtures

A good and thorough introductory book is the work of McLachlan and Peel 2000 and the reader is encouraged to study that to learn in depth about normal mixtures. We will here give a short overview of normal mixtures to fix notation and nomenclature.

Let $\mu \in \mathbb{R}^p$, $\Sigma \in \mathbb{R}^{p \times p}$ and $\phi(\mu, \Sigma)$ be the normal distribution with mean μ and covariance matrix Σ .

Normal mixture models are designed for situations where we assume that a given dataset originates from more than one population of explaining variables.

$\mathbf{Y}_1, \dots, \mathbf{Y}_n$

Definition 1.1.0.1. Suppose we have a random sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ with probability density function $\mathbf{Y}_j \sim f(y_j)$ on \mathbb{R}^p . We assume that the density $f(y_j)$ of \mathbf{Y}_j can be written in the form

$$f(y_j) = \sum_{i=1}^K \pi_i \phi_i(y_j)$$

The π_i are called the component densities of the mixture.

explain in sketch EM algo

EM has desirable qualities like proven convergence, (give reference to Dempster 1977 paper)

explain idea to use parameter optimizer instead, EM has pathological insufficiencies, like 'getting stuck' for many iterations. we hope we need less iterations, and as consequence less time. 'special' idea: using Cholesky decomp.

1.2 choice of notation

describe difference in notation between ceuleux & govaert and our covariance matrix decomposition.

The classification of models in this paper relies heavily on the work of Celeux and Grovaert, however, out of necessity for clarity, we break with their notation. So as to not confuse the reader we describe here in depth the differences in notation between Celeux and Govaert and ours.

explanation for the volume, shape and orientation descriptors

The basis of classification in CnG is the decomposition of a symmetric matrix into an orthogonal and a diagonal component. A symmetric positive definite matrix Σ can be decomposed as follows

$$\Sigma = \lambda \mathbf{D} \mathbf{A} \mathbf{D}^\top$$

with \mathbf{D} an orthogonal matrix and \mathbf{A} a diagonal matrix and $\lambda = \sqrt[p]{\det(\Sigma)}$ the p -th root of the determinant of Σ .

This decomposition has an appealing geometric interpretation, with \mathbf{D} as the *orientation* of the distribution, \mathbf{A} the *shape*, and λ the *volume*. The problem of notation comes from standard conventions in linear algebra, where the letters A and D are usually occupied by arbitrary and diagonal matrices respectively. Furthermore, we intend to apply a variant of the Cholesky decomposition to Σ , the \mathbf{LDL}^\top decomposition. This obviously raises some conflicts in notation.

Therefore we, from here on, when referring to the decomposition as described by cng, will use the following modification of notation:

$$\begin{aligned} \mathbf{D} &\mapsto \mathbf{Q} \\ \mathbf{A} &\mapsto \mathbf{\Lambda} \\ \lambda &\mapsto \alpha \\ \Sigma = \lambda \mathbf{D} \mathbf{A} \mathbf{D}^\top &= \alpha \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top \end{aligned}$$

These were chosen according to general conventions of linear algebra. \mathbf{Q} is usually chosen for orthonormal matrices; $\mathbf{\Lambda}$ is often a choice for eigen vectors and α was somewhat arbitrarily chosen.

make clear that the models can not be translated one to one to ldlt model

make nice table(maybe sideways to account for parameter list)

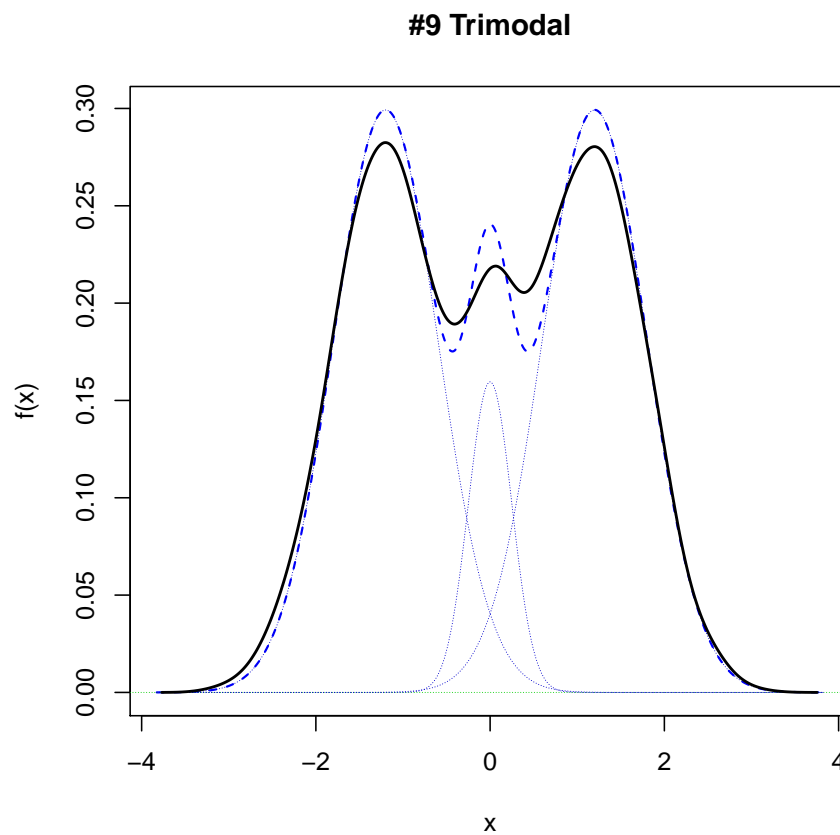
Model	Σ_k C&G	volume	shape	orientation	parameters	count	LDL^\top same as C&G	parameters	count
EII	αI	equal	equal	-	α	1	same as C&G		
VII	$\alpha_k I$	variable	equal	-	α_k	K			
E EI	$\alpha \Lambda$	equal	equal	coordinate axes	α, λ_i	$1 + p$			
VEI	$\alpha_k \Lambda$	variable	equal	coordinate axes	α_k, λ_i	$K + p$			
EVI	$\alpha \Lambda_k$	equal	variable	coordinate axes	$\alpha, \lambda_{i,k}$	$1 + pK$			
VVI	$\alpha_k \Lambda_k$	variable	variable	coordinate axes	$\alpha_k, \lambda_{i,k}$	$K + pK$			
EEE	$\alpha Q \Lambda Q^\top$	equal	equal	equal	$\alpha, \lambda_i, q_{i,j}$	$1 + p + p^2$	αLDL^\top		
EVE	$\alpha Q \Lambda_k Q^\top$	equal	variable	equal	$\alpha, \lambda_{i,k}, q_{i,j}$	$1 + pK + p^2$	doesn't exist		
VEE	$\alpha_k Q \Lambda Q^\top$	variable	equal	equal	$\alpha_k, \lambda_i, q_{i,j}$	$K + p + p^2$	$\alpha_k LDL^\top$		
VVE	$\alpha_k Q \Lambda_k Q^\top$	variable	variable	equal	$\alpha_k, \lambda_{i,k}, q_{i,j}$	$K + pK + p^2$			
EEV	$\alpha Q_k \Lambda Q_k^\top$	equal	equal	variable	$\alpha, \lambda_i, q_{i,j,k}$	$1 + p + Kp^2$			
VEV	$\alpha_k Q_k \Lambda Q_k^\top$	variable	equal	variable	$\alpha_k, \lambda_i, q_{i,j,k}$	$K + p + Kp^2$			
EVV	$\alpha Q_k \Lambda_k Q_k^\top$	equal	variable	variable	$\alpha, \lambda_i, q_{i,j,k}$	$1 + pK + Kp^2$	$\alpha L_k D_k L_k^\top$	$\lambda, d_{i,k}, l_{i,j,k} \ j > i$	$1 + pK + K \frac{p(p-1)}{2}$
VVV	$\alpha_k Q_k \Lambda_k Q_k^\top$	variable	variable	variable	$\alpha_k, \lambda_i, q_{i,j,k}$	$K + pK + Kp^2$	$\alpha_k L_k D_k L_k^\top$	$\lambda_k, d_{i,k}, l_{i,j,k} \ j > i$	$K + pK + K \frac{p(p-1)}{2}$

1.3 problems of EM

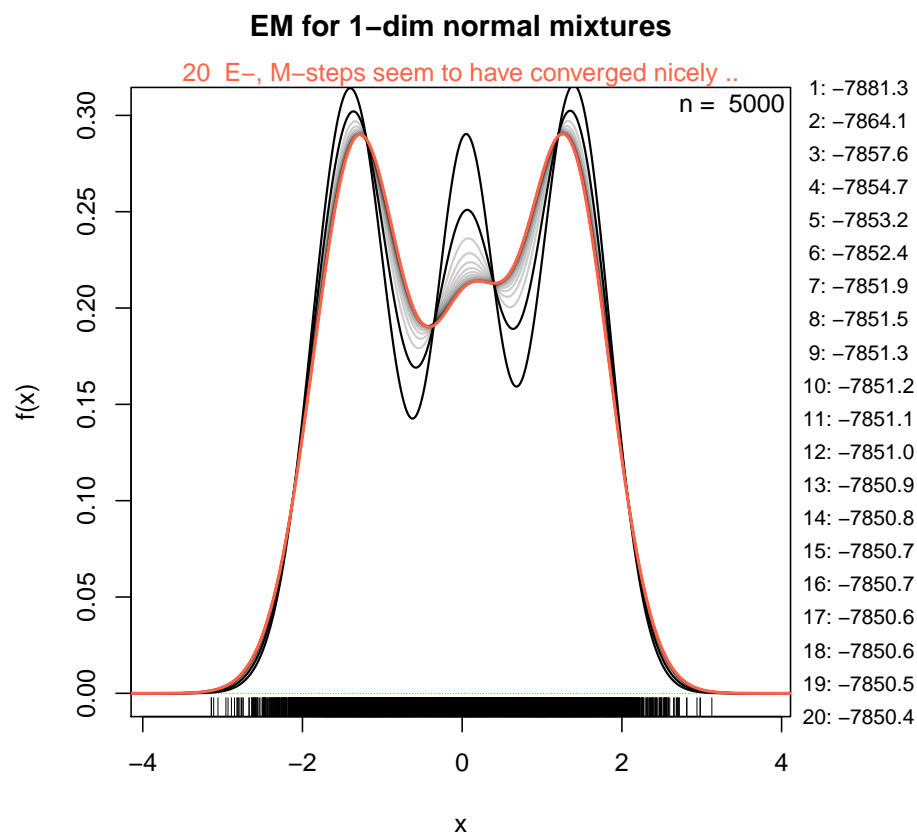
the EM algo has stalling problems especially close to a local optimum

show an example using `nor1mix`

```
> library("nor1mix")
> plot(MW.nm9, lty=2, col = "blue", p.norm=FALSE, p.comp=TRUE)
> set.seed(2019)
> x9 <- rnorMix(5000, MW.nm9)
> lines(density(x9), lwd=1.8) # "clearly" 3 components
```



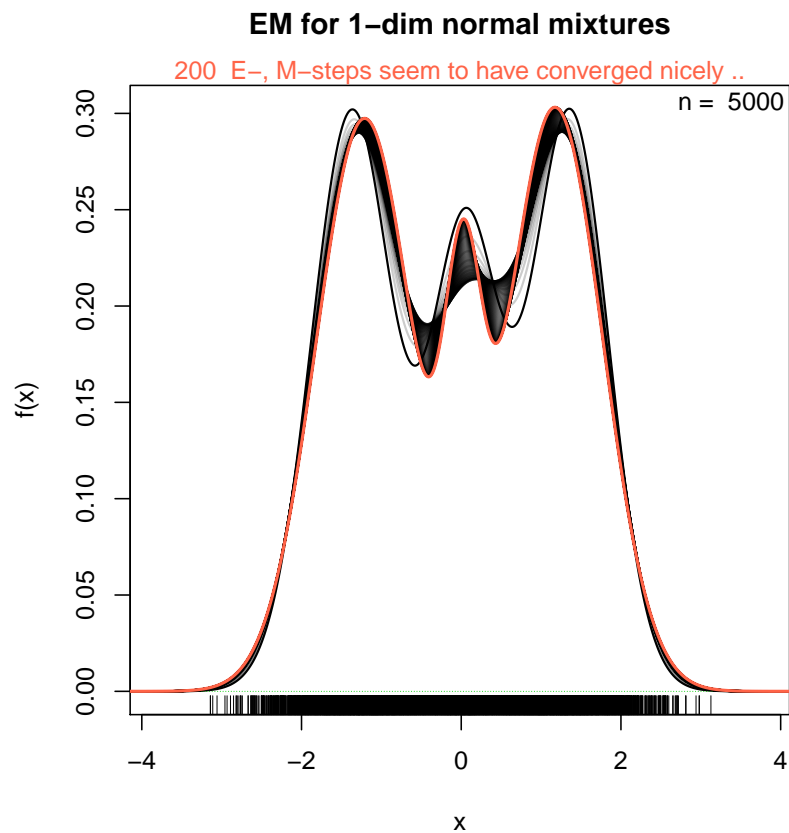
then an illustration of MW examples of pathological cases



yay, got figure to print. solution was use of `fig=TRUE`, instead of various mutations like `figure=true`.

here we see how change in loglik seems to stagnate. However, this does not stay that way, if we let EM run a bit further.

```
> r <- p.EMsteps(200, x9, nm1)
```



to conclude example show part of mixest that shows it takes 1200 iterations to converge

In fact, it seems that the previous solution is a saddle point in the likelihood function, where EM has chronic problems continuing improvements.

should include animations?? like mix_est_1d.R line 249+24 lines

maybe show Marr Wand's examples of 'difficult' mixtures

give conclusion recapping the just demonstrated, and lead in for next chapter

Chapter 2

The norMmix Package

explain, that this package was written purposefully for this paper.

The norMmix package is constructed around the `norMmix` object, that codifies a **normal** Multivariate mixture model, and the `llnorMmix()` function.

quickly list contents of norMmix object

relies on `optim()` generic optimizer. maximizes `llnormix` by varying model parameters.

since `mclust` is one of the more popular packages implementing the EM algo, we employ a lot of functions from `mclust`, to keep things around EM as similar as possible.

2.1 Theoretical Groundwork

(this Section maybe one chapter earlier)

about Cholesky decomp as `ldlt`. has advantages: fast, parametrically parsimonious, can easily compute loglikelihood

maybe reread section in McLachlan about accelerating EM algo

not possible to sensibly compare normal mixtures except maybe a strange sorting algorithm using mahalanobis distance or Kullback-Leibler distance or similar(Hellinger), but not numerically sensible to integrate over potentially high-dimensional spaces.

So caomparison of algos done through throwing difficult mixtures and non-mixtures at it and hoping that norMmix finds better solutions than EM. So the criteria for "better fit" are 1. better log-likelihood 2. correct model, where EM fails.

2.2 finer details of norMmix package

Chapter 3

Comparing Algorithms

maybe apply to MW[0-9] objects?

not sure

Chapter 4

Coclusions

Bibliography

- Bar, F. and H. Meier (2001). Title of the article. *Journal where the article has been published volume of the journal*, 12–77.
- Hampel, F. R. (1985). The breakdown points of the mean combined with some rejection rules. *Technometrics* 27(2), 95–107.
- Stahel, W. and S. Weisberg (1991). *Directions in Robust Statistics and Diagnostics*, 2 vol. N. Y.: Springer-Verlag.

Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor .

Title of work (in block letters):

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

First name(s):

Muster.....Student.....
.....
.....
.....

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the Citation etiquette information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work .
- I am aware that the work may be screened electronically for plagiarism.
- I have understood and followed the guidelines in the document *Scientific Works in Mathematics*.

Place, date:

Signature(s):

Zurich August 19th 2009.....bla.....
.....
.....
.....

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.