Thesis for the Degree of Licentiate of Engineering

Estimating Diffusion Coefficients in Colloidal Particle Systems

Mats Kvarnström

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology
and Göteborg University
SE-412 96 Göteborg, Sweden

Göteborg, October 2002

Estimating Diffusion Coefficients in Colloidal Particle Systems
MATS KVARNSTRÖM

ISSN 0347-2809/NO 2002:79
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Göteborg, Sweden 2002

Estimating Diffusion Coefficients in Colloidal Particle Systems
MATS KVARNSTRÖM
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University

Abstract

This licentiate thesis deals with the estimation of the diffusion variance (or equivalently, the diffusion coefficient) of colloidal particles. Particle positions have been observed and estimated from a series of images recorded with a video microscope, using more or less standard image processing algorithms and tools. In particular, to estimate the position of the particles in each image we used a modification of the Hough Transform.

The particles are assumed to perform independent Brownian motions in three dimensions. A complicating fact is that some of the observed particles are not moving, but are instead either particles adsorbed on the objective or cover glass of the specimen, or correspond to defects in the optics of the microscope.

A model is introduced with two kinds of particles, diffusing and fixed. To each particle position estimate we assume an additive measurement error. The parameter of the model consists of the diffusion variance, the measurement error variance, and the proportion of diffusing particles. The problem can now be considered as an incomplete data problem since we do not know *a priori* which particles are really diffusing. The maximum likelihood estimator is computed via the EM algorithm and as a side-effect we also get the classification variables of the particles, i.e. if they are diffusing or fixed. The estimator is shown to be strongly consistent and asymptotically normal, as the number of particles approaches infinity, under a reasonable restriction on the parameter space.

Key words: Discretely observed diffusion, measurement error, mixture distribution, asymptotic normality, strong consistency, curved exponential family, EM algorithm, image processing, Hough Transform

Preface

The report

This licentiate thesis is part of an ongoing project in particle trajectory estimation in image processing.

The work has been financially supported by the Swedish Research Council for Engineering Sciences (TFR) for the years 2000 and 2001, and by the Swedish Research Council (VR) for the year 2002.

Acknowledgements

My personal thanks goes to:

My supervisor Mats Rudemo who has been supporting and encouraging me through the entire process of writing this thesis.

Chris Glasbey for great advice and interesting talks and for letting me visit him in Edinburgh last autumn.

Lennart Lindfors of AstraZeneca R&D for producing the images and for introducing the project, first to Mats, and then later to me. Thanks also for keeping me in touch with the application.

Holger Rootzén and Ib Skovgaard for interesting comments, fruitful discussions and good intuition regarding the asymptotics.

Olle Häggström who read through the whole manuscript and supplied valuable comments.

Last but not least, I thank my family for always supporting and believing in me. Especially you, Jessica.

Mats Kvarnström, October 2002

Contents

Chapter 1

Introduction

The understanding of the behaviour of colloidal suspensions are of crucial importance in a vast number of different areas. Examples of common, everyday life colloidal systems are milk and paint. In milk, various interactions between the small (100 nm to 1 μm in diameter) colloidal milk fat particles and proteins suspended in the fluid, decide whether it coagulates into cheese or yoghurt. These interactions depend on how the milk was treated before the coagulation. For the second example, the pigments in the paint must stay suspended in the liquid in a can for years, yet, as they are spread on a wall, be able to coagulate fast.

The goal of the ongoing project is to develop techniques for the quantitative study of the diffusion and interaction of particles in a colloidal system, with subsequent application to pharmacy. Here the possible modifications of the colloidal particles have a large impact on modern therapies such as oral vaccines and gene therapy.

The standard theory for the interactions of colloidal particles, the DLVO-theory (see for example [5]), is merely an approximation and experiments have shown that it fails to predict the behaviour of several important suspensions, see [1] and [8]. Therefore, observations on the microscopic level are needed.

The idea is to make inference of particle diffusion and interaction from a series of light microscope images of moving particles. Figure 1.1 illustrates an example of what an image from such a sequence might look like. The particles in these image are spherical, made of latex (polystyrene), and have all a diameter of 494 nm. The experiment is constructed in such a way, that we can assume that the moving particles are performing a pure Brownian motion in three dimensions. That is, no drift and no interaction between different particles.

The apparent differences in size and brightness variations of the particles is

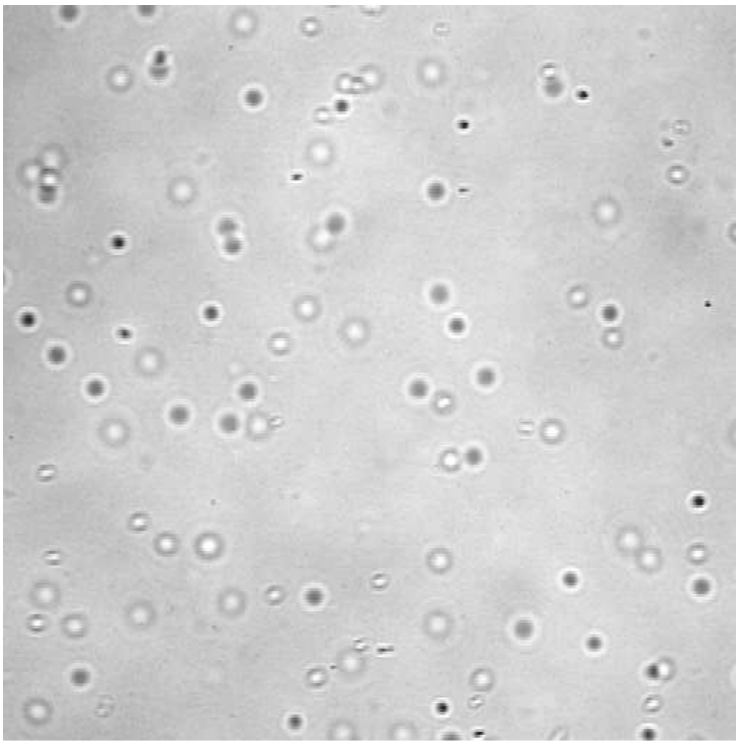

Figure 1.1: A single microscope image in the sequence.

due to off-focus placement. Particles in the focal plane are depicted as small, distinct, black spots, while particles above or below the focal plane, are either light or dark in middle, respectively. Also, the further away from the focal plane a particle is, the larger and more blurred it appears. This effect will give us a method to, at least in principle, estimate the third coordinate of a particle position. It should be mentioned that the light is coherent, which is the reason for this optical effect. If light would have been incoherent, particles off-focus would simply be blurred.

This is the first, introductory, part of the project and as a first study we have tracked some of the particles in 12 consecutive images like the one in Figure 1.1, using basic image processing tools. The estimated trajectories can be found in Figure 2.1. From the trajectories of the particles we then estimated the diffusion coefficient. For this, we used a model which incorporated two kinds of particles, diffusing and fixed. This was necessary since some of the particles we tracked were essentially fixed in position. These seemed to be either particles adsorbed by the surrounding glass surfaces, or not particles at all, but rather defects

in the optics. We write "essentially fixed" since their estimated position was still fluctuating a bit between different images. This observation also made us introduce a a measurement error of the observed position versus the real position.

We have not estimated the interaction potential between particles. To be able to do that, we need position estimates in all three dimensions and the technique for this has not been developed yet.

1.1 Related work

Crocker and Grier have written a number of articles where they use image analysis tools to analyse the pairwise interaction and diffusion coefficients of particles in colloid suspensions. In [2], they locate particles in face-centered cubic colloid crystal of polystyrene spheres of diameter 326 nm. After suitable pre-processing, candidates for particle positions are found using grey-scale dilation (a technique that identifies local brightness maxima in the image). They refine these estimates in the plane by fitting a brightness-weighted centroid to the pixels within a region around the first candidate. They also use a variant of this idea for estimating the placement in depth. The linking of particle locations in each of the images to trajectories are done by a nearest-neighbour technique, like the one we used, see Chapter 3.

The major reason for the non-applicability of their image analysis methods to our problem, is first of all that their depth of focus is comparable to the diameters of the particles. This means that they do not have to deal with the problem associated with particles looking different depending on their location in depth. Secondly, since their particles have formed a crystal, they are not moving very rapidly, which is typically the case in our problem.

1.2 Outline of the report

In Chapter 2, we introduce the model, with two types of particles and measurement error on the position estimates that we used for the observed particle trajectories. Then we describe two common estimators of the diffusion variance, and compare them with the maximum likelihood estimator, under our model assumption. This chapter is rather sketchy and serves primarily as motivation for our model and to the paper *"Estimation of the diffusion coefficient in a mixture model with diffusing and fixed particles"*, written by the author of this thesis and supplemented to it.

Chapter 3 deals with the image processing we used, to get the estimated trajectories in Figure 2.1. A three stage approach was used for each image. First we did appropriate pre-processing consisting of interpolation and smoothing. Then we looked for circles in the images by using a modification of a method called Circular Hough Transform. To actually get trajectories, we finally matched the position estimates of the particles from the image sequence. For each particle in one image, we looked for the nearest particle in the next image. Problem arises in this step when we have different number of particle position estimates from one image to another. This is typically the case in our application, first of all since our method to find particle candidates is far from perfect, and secondly because our particles are actually moving and may thus become occluded by other particles, disappear out of the region covered by the image, come in again, etc.

Here, we got around this problem by manually discarding the particles not found in all the images in a sequence. Because this is a time-consuming task to do manually, we only considered a subset of the image and just 12 images.

In Chapter 4 we present some of the subjects that need to be looked at further in the continuation of the project.

The main part of the work can be found in the supplemented paper. We estimated the parameters of the model of Section 2.1, using maximum likelihood estimation. The EM algorithm (see [4] and [14]) was a natural choice of computational procedure in this problem, since we can regard the classification variables for each particle (indicating whether it is a diffusing or fixed particle) as missing, or unobserved, data. The estimated diffusion variance seems reasonable when compared to the theoretical one. We also show that the estimator exists for any fixed number of observed particles n, and that it is strongly consistent and asymptotically normal as the number of particles goes to infinity.

We write [P], to denote this supplemented paper, which has been submitted for journal publication.

Chapter 2

Estimating the diffusion coefficient

Assume we have observed n particles diffusing over a time period corresponding to N discrete increments, and that we want to estimate the diffusion variance. Which is "the best" estimator depends obviously on the assumed underlying diffusion model, but also on ease of computation. Here we compare two more or less *ad hoc* estimators, commonly used and easy to compute, with the maximum likelihood estimator dealt with in [P]. The entire chapter serves as a complement, or rather a motivation, to this paper and in order not to repeat to much from it here, we recommend that it is read in close conjunction.

In Figure 2.1 we can see the inspiration for the model with two kinds of particles, that we are going to introduce. Clearly, some of the particles in the figure do not seem to be diffusing.

2.1 The model

Denote the true and observed position of a generic particle at time k, where $k = 0, \ldots, N$, by R_k and S_k, respectively.

For diffusing particles we assume the following state-space model:

$$\begin{aligned} R_k &= R_{k-1} + w_k \\ S_k &= R_k + e_k \end{aligned}$$

for $k = 1, \ldots, N$, where the increments $\{w_k\}$ are i.i.d. normally distributed random variables with variance σ^2. The measurement errors $\{e_k\}$ are i.i.d. zero

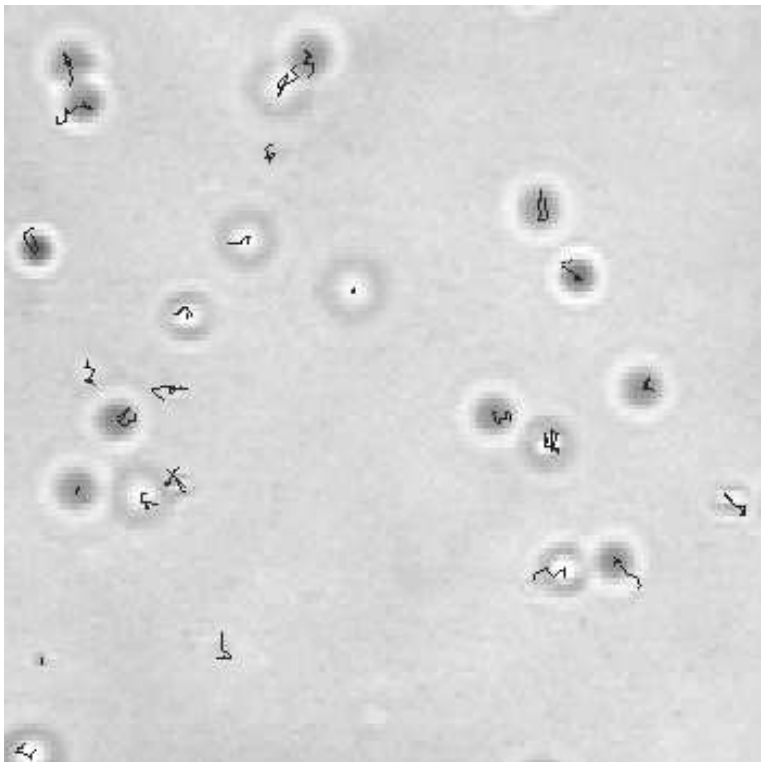

Figure 2.1: The 26 trajectories estimated in a sequence of 12 images together with the first image in the sequence. Notice that three of the particles seem to be fixed.

mean normal variables with variance σ_e^2, and independent of the increments $\{w_k\}$.

We allow for some particles to be fixed in position. For these we have the same state-space model as above but with $w_k = 0$, for all k. We can also think of them as being degenerate zero mean random variables with $\sigma^2 = 0$.

We let each particle be diffusing with probability p independently of each other. We associate a classification variable Z_i for $i = 1, \ldots, n$, to be one if the i:th particle is diffusing and zero otherwise. A particle is either diffusing or fixed for the entire sequence of $N+1$ observations.

This full model, with two kinds of particles, and imperfect position measurement, is denoted $\mathcal{M}$. The unknown parameters are

- σ^2: the diffusion variance, i.e. the variance of an increment

- σ_e^2: the variance of the measurement error
- p: the proportion of particles which are diffusing

For ease of notation we only consider the 1-dimensional case. Everything extends naturally to the general d-dimensional case if we assume the measurement error to be i.i.d. in all dimensions. Note however, that this assumption is probably a rather coarse simplification of what the truth might be. At least one would think that the estimates of the depth relative to the focal plane (the z coordinate) are more uncertain than the estimates in the plane (the x and y coordinates). It is also reasonable to believe that the degree of uncertainty in the z coordinate estimates might depend on the true position in depth. However, in two dimensions, and with the image processing tools used, we believe that the assumption of independent and identically distributed measurement error is plausible.

2.2 Notation

The index i, $i = 1, \ldots, n$, is used to distinguish between the n particles and a subindex i as in Z_i, means that the entity belongs to the i:th particle. If the index is neglected, we mean a generic particle. The index k, $k = 1, \ldots, N$, is used for a generic particle only, and correspond to the discrete time k in the state-space model.

We are going to work with the observed increments $Y_k = S_k - S_{k-1}$ for $k = 1, \ldots, N$. The observed increment vector is defined as $Y = [Y_1, \ldots, Y_N]^T$.

The covariance matrix of the observed increment vector is denoted $\Sigma_1 = \Sigma_1(\sigma^2, \sigma_e^2)$ for a diffusing particle, and $\Sigma_0 = \Sigma_0(\sigma_e^2)$ for a fixed. See Section 2.1 in [P].

Since the covariance matrices are non-diagonal, we work with the transformed increment vector $\tilde{Y} = U^T Y$ instead for easier notation, where U is the matrix of eigenvectors to Σ_1 as columns. It turns out that $\mathbf{Var}\{\tilde{Y}\} = \mathrm{diag}\{\sigma^2 + \lambda_k \sigma_e^2, 1 \leq k \leq N\}$, where λ_k are the eigenvalues to a tri-diagonal matrix T. See Section 2.2 in [P].

We classify data into two categories, observed and complete. The observed data is simply the observed increment vectors Y_i, $i = 1, \ldots, n$, of the n particles. The complete data consists of the observed data together with the the the corresponding classification variables Z_i. The latter are unobserved and is therefore called the unobserved data. The complete data is of curved exponential type (see Section 3 and the Appendix in [P]) and the observed data is a mixture distribution g, consisting of two N-variate normal components, both with zero mean but with different covariance matrices, Σ_1 and Σ_0:

$$g(y; \sigma^2, \sigma_e^2, p) = p f(y\,; \Sigma_1) + (1-p) f(y\,; \Sigma_0)$$

where $f(\,\cdot\,;\Sigma)$ is a N-variate normal density with mean zero and variance Σ.

2.3 Comparing three estimators

In the analysis of this Section, we assume that $p = 1$ and that the measurement error σ_e^2 is known, but different from zero.

The sum of squares estimator

Given a increment vector Y from an observed Brownian motion, the first estimator of σ^2 one would come to think of would probably be

$$\hat{\sigma}_S^2 = \frac{1}{N}\sum_{k=1}^{N} Y_k^2 = \frac{1}{N}\sum_{k=1}^{N} \tilde{Y}_k^2$$

This is the maximum likelihood estimator of σ^2 under the assumption that $p = 1$ and $\sigma_e^2 = 0$, i.e. if we have perfect observation of diffusing particles only.

Under our more complicated model assumption from Section 2.1, this will however be a biased estimator of σ^2:

$$\mathbf{E}\{\hat{\sigma}_S^2\} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{E}Y_k^2 = \sigma^2 + 2\sigma_e^2$$

where we used that $\mathbf{E}Y_k^2 = \sigma^2 + 2\sigma_e^2$. For the example in [P], this corresponds to a relative bias of roughly 25 percent.

The variance of this estimator is

$$\mathbf{Var}\{\hat{\sigma}_S^2\} = \mathbf{Var}\Big\{\frac{1}{N}\sum_{k=1}^{N} \tilde{Y}_k^2\Big\} = \frac{1}{N^2}\sum_{k=1}^{N} 2(\sigma^2 + \lambda_k\sigma_e^2)^2 \qquad (2.1)$$

due to the independence of the components of the transformed increment vector. To get a closed form expression for (2.1), one can use the identities $\sum \lambda_k = \mathrm{trace}\{T\} = 2N$, and $\sum \lambda_k^2 = \mathrm{trace}\{T^2\} = 6N-2$.

Mean square displacement

The second estimator is based on the "mean-square-displacement" quantity; for each τ, $\tau = 1, \ldots, N$, take the mean of the squared displacement up to time τ over all particles. For one particle this becomes

$$\mathrm{MSD}(\tau) = \Big(\sum_{k=1}^{\tau} Y_k\Big)^2$$

Since $\sum_{k=1}^{\tau} Y_k$ is a zero mean normal variable with variance $\tau\sigma^2 + 2\sigma_e^2$, the expected value of this is

$$\mathbf{E}\{\mathrm{MSD}(\tau)\} = \mathbf{Var}\Big\{\sum_{k=1}^{\tau} Y_k\Big\} = \tau\sigma^2 + 2\sigma_e^2 \tag{2.2}$$

This quantity is usually plotted for increasing values of τ and mainly used as a graphical method for detecting a drift in the diffusion. If there is a drift present in the diffusion, the bias in (2.2) will then be, in addition to the $2\sigma_e^2$ term above, the square of the expected drift up to τ.

Define the estimator of σ^2 based on mean square displacement, $\hat{\sigma}^2_{MSD}$, as

$$\hat{\sigma}^2_{MSD} = \frac{1}{N}\Big(\sum_{k=1}^{N} Y_k\Big)^2 = \frac{\mathrm{MSD}(N)}{N}$$

When used as an estimator for σ^2, this has very poor properties. If we take a look at the variance of the estimator

$$\mathbf{Var}\{\hat{\sigma}^2_{MSD}\} = \frac{2}{N^2}(N\sigma^2 + 2\sigma_e^2)^2 = 2(\sigma^2)^2 + o(1)$$

we see that this does not go to zero as N increases, which typically is a property that we would like our estimator to have. In fact, the only good thing about it, is that the bias induced by σ_e^2 decreases as we increase N:

$$\mathbf{E}\{\hat{\sigma}^2_{MSD}\} = \sigma^2 + \frac{2}{N}\sigma_e^2$$

which is a rather small comfort.

Maximum likelihood estimation

The density for the increment Y is (see Section 3 in [P])

$$\begin{aligned} f_1(y|\sigma^2) &= \frac{1}{(2\pi)^{N/2}|\Sigma_1|^{1/2}} \exp\Big\{-\frac{y^T\Sigma_1^{-1}y}{2}\Big\} \\ &= \frac{1}{(2\pi)^{N/2}\prod_{k=1}^{N}(\sigma^2 + \lambda_k\sigma_e^2)^{1/2}} \exp\Big\{-\frac{1}{2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{\sigma^2 + \lambda_k\sigma_e^2}\Big\} \end{aligned}$$

Hence, disregarding an additive constant, the log-likelihood for σ^2 can be written

$$l(\sigma^2) = -\frac{1}{2}\sum_{k=1}^{N}\log(\sigma^2 + \lambda_k\sigma_e^2) - \frac{1}{2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{\sigma^2 + \lambda_k\sigma_e^2}$$

and the score equation becomes

$$\dot{l}(\sigma^2) = -\frac{1}{2}\sum_{k=1}^{N}\frac{1}{\sigma^2+\lambda_k\sigma_e^2} + \frac{1}{2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{(\sigma^2+\lambda_k\sigma_e^2)^2} = 0$$

where the dot indicates the derivative with respect to σ^2. This equation has no closed form solution, except for case $N = 1$.

The expected information of σ^2 is

$$\begin{aligned}\mathbf{E}\left(-\ddot{l}(\sigma^2)\right) &= -\mathbf{E}\Big(\frac{1}{2}\sum_{k=1}^{N}\frac{1}{(\sigma^2+\lambda_k\sigma_e^2)^2} - \sum_{k=1}^{N}\frac{\tilde{Y}_k^2}{(\sigma^2+\lambda_k\sigma_e^2)^3}\Big)\\ &= \sum_{k=1}^{N}\frac{\mathbf{E}\tilde{Y}_k^2}{(\sigma^2+\lambda_k\sigma_e^2)^3} - \frac{1}{2}\sum_{k=1}^{N}\frac{1}{(\sigma^2+\lambda_k\sigma_e^2)^2} = \frac{1}{2}\sum_{k=1}^{N}\frac{1}{(\sigma^2+\lambda_k\sigma_e^2)^2}\end{aligned}$$

This estimator is consistent and asymptotically normal as $n \to \infty$. The asymptotic variance of $\sqrt{n}(\hat{\sigma}^2_{ML} - \sigma^2)$ becomes

$$\mathbf{Var}\{\sqrt{n}(\hat{\sigma}^2_{ML} - \sigma^2)\} \to \frac{1}{\frac{1}{2}\sum_{k=1}^{N}\frac{1}{(\sigma^2+\lambda_k\sigma_e^2)^2}} \tag{2.3}$$

which can be compared with the variance of $\sqrt{n}\hat{\sigma}^2$ from (2.1) for different N and "signal-to-noise", σ^2/σ_e^2, ratios. Since (2.3) is the asymptotical variance of the maximum likelihood estimator, we should have that the variance of $\hat{\sigma}^2_{ML}$ is smaller than the variance of $\hat{\sigma}^2_S$, because maximum likelihood estimators are asymptotically efficient. This is also verified by the Cauchy-Schwarz inequality:

$$\frac{\mathbf{Var}\{\hat{\sigma}^2_{MLas}\}}{\mathbf{Var}\{\hat{\sigma}^2_S\}} = \frac{N^2}{\Big(\sum_{k=1}^{N}(q+\lambda_k)^2\Big)\Big(\sum_{k=1}^{N}\frac{1}{(q+\lambda_k)^2}\Big)} \le \frac{N^2}{\Big(\sum_{k=1}^{N}\frac{q+\lambda_k}{q+\lambda_k}\Big)^2} = 1$$

where $q = \sigma^2/\sigma_e^2$ and by $\mathbf{Var}\{\hat{\sigma}^2_{MLas}\}$ we mean the asymptotical variance as shown to right in (2.3). It is a fact that the maximal element of the eigenvalues λ_k is less than 4, so we see that the ratio goes to one when q grows large, and N is kept fixed. This is what we should expect, since then the measurement error σ_e^2 can be neglected.

2.4 The proposed estimator

If we look at the 26 trajectories in Figure 2.1 we notice that three of them seem to correspond to fixed particles. The fact that the trajectories are not just

a single point was our motivation for introducing the measurement error; the apparent movement is probably due to that error.

The first idea would be to discard the three fixed particles manually as outliers and then estimate σ^2 and σ_e^2 using the maximum likelihood estimator under the assumption of $p=1$ for the remaining 23 particle trajectories. Since the fixed particle "trajectories" contain information about σ_e^2, this seems like a slightly wasteful approach, though.

Our next alternative would be to, rather than just discard the particles we manually labelled as fixed above, incorporate these into the analysis by using that the corresponding observed increment vectors come from an N-variate normal density with covariance matrix Σ_0. This is the full model assumption $\mathcal{M}$ but with complete data.

The last alternative is to use the observed data only, that is, the observed increment vectors from the 26 particles. Since we do not classify the data, the classification variables can be thought of as missing, or unobserved, data. So, we still assume the full model, but in this case the data comes from a mixture model with two N-variate normal components.

In [P], we mainly consider the last alternative of maximum likelihood estimation using observed data. The EM algorithm [4] is then a natural method to use when maximizing the likelihood, since it also gives us estimates of the classification variables as a side-effect.

Chapter 3

Image processing

In this chapter we describe the method we used to get the estimated trajectories of the 26 particles in Figure 2.1. The method consists of three parts. First is a pre-processing stage of interpolating and smoothing. Then we find the locations of the particles by looking for circles in the image. After the particle positions have been estimated, we link these to form the actual trajectories.

3.1 Instrumentation setup

Latex particles made of polystyrene with a diameter of 494nm were placed between an objective and a cover glass and sealed. The illumination consisted of coherent light. The sample was studied in a Zeiss Axiovert 135 TV microscope equipped with a Newicon video camera.

The video signal was then digitized and stored as TIF files of size 512 times 512 pixels. The digitized sequence consists of 50 images (or frames) per second. In practice however, only half of them are useful. The camera records only half of the rows at each scan, alternating between the even and odd rows (also called the even and the odd fields) and duplicates this information to the rows which were not scanned.

In other words, if we decide to use the frames corresponding to the even fields we get 25 pictures every second consisting of either the even or the odd rows only. Another possibility is to interlace two consecutive frames (one even plus one odd) into a single frame, using the even rows from the even frame and the odd from the odd frame. Since there is an interval of a 1/50 of a second between the even and the odd frame, this will cause problems when observing rapidly moving particles. This is the case in this study, so this method does not

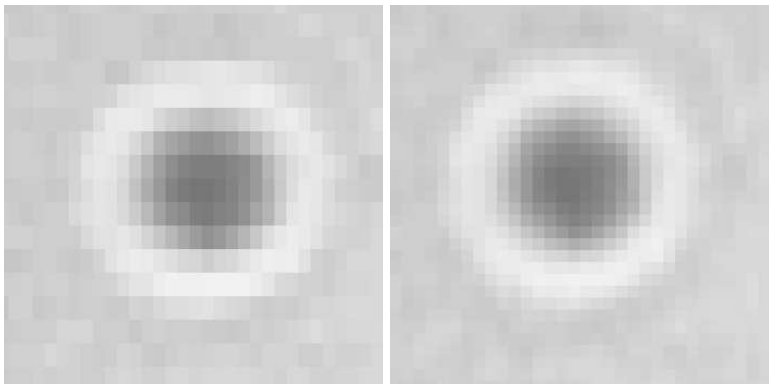

Figure 3.1: *Left*: Part of an even frame. Note that the pixels look rectangular due to the duplication of rows. *Right*: The same part of the image after reconstruction by interpolation.

work very well here. We use the first method, where we use the even frames in combination with interpolation (see section 3.2).

The domain in which the particles are confined, and are available for our inspection through the images, is a box with equal length of the sides of about 90 μm, and a depth of 26 μm. The focal plane is set to a depth of 13 μm.

3.2 Estimation of the location of particles

Determining the location of the particles in an image is a somewhat tricky issue due to the fact of the range of the focus. The method proposed by Crocker and Grier [2] is not applicable to our problem, the main reason being that our particles are both dark and bright depending on the z-coordinate of the current particle; we cannot simply search for black objects standing out of an brighter background.

If we are unable to search with respect to brightness, we search with respect to shape. The method described below involves a Circular Hough Transform (CHT) applied to an gradient image. The problem with the ordinary CHT [7] is that the parameter space becomes three dimensional (and it is computationally heavier to search in three dimensions) if we are searching for circles of different radii. This is the case here because the particles are depicted as concentric circles of different radii. Instead, we use the idea of *coherent CHT* [10], where the image (see below) is convolved with a single, but complex valued mask.

Pre-processing

The raw data from the digitizer is a sequence of even field images. Since each image consists of the even rows only, with the odd rows as a duplicate of the even row above, we interpolated the pixel values of the even rows to the odd to get a smoother image. For each odd row, we simply took the mean of the two neighbouring even rows.

To get rid of spurious errors due to digitization and other imperfections caused in the image formation process, we smoothed the interpolated image I^{int} by convolving this with square gaussian mask H of sidelength $2w+1$

$$H_{k,l} = \frac{1}{2\pi s^2} \exp\left(-\frac{k^2+l^2}{2s^2}\right)$$

where $-w \leq k,l \leq w$ and s^2 is a smoothing parameter. We used $w = 3$ and $s^2 = 1$.

Denote the pre-processed image by I^{pre} and the individual pixels in I^{pre} by $I^{pre}_{i,j}$. Then

$$I^{pre}_{i,j} = \sum_{-3 \leq k,l \leq 3} H_{k,l} I^{int}_{i+k,j+l}$$

Finding estimates of the x and y coordinates

To get candidates for the position of particles, we first estimated the modulus of the gradient at each point in the image I^{pre}. Then we used the coherent Circle Hough Transform to this gradient image to find the centers of the circles in the image. These centers correspond (hopefully) to the x and y coordinates of the particles.

Prewitt's gradient filter

A simple way of detecting edges in an image is to calculate the modulus of the discrete gradient at each pixel in an image. This method is called Prewitt's gradient filter, see Glasbey and Horgan [6]. First we estimate the derivative in the x direction by

$$E^x_{i,j} = \sum_{-1 \leq k,l \leq 1} H^x_{k,l} I^{pre}_{i+k,j+l} \tag{3.1}$$

where

$$H^x = \frac{1}{6}\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

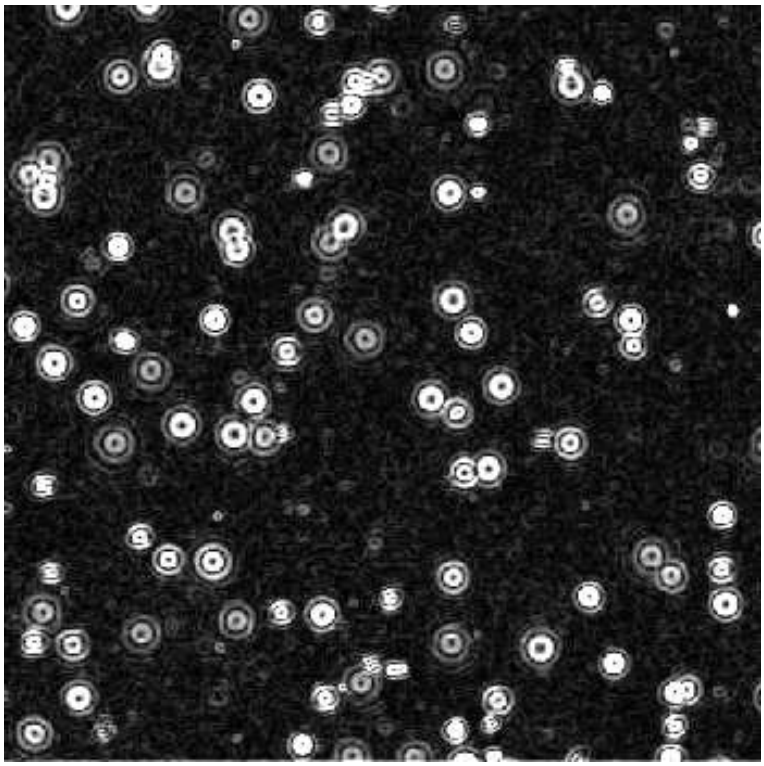

Figure 3.2: The gradient image using Prewitt's filter.

The derivative in the y direction E^y is estimated similarly using

$$H^y = \frac{1}{6}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

in (3.1) instead of H^x. The modulus of the gradient, E, was calculated according to

$$E_{i,j} = \sqrt{(E^x_{i,j})^2 + (E^y_{i,j})^2} \tag{3.2}$$

Figure 3.2 shows the output for the image in Figure 1.1.

The Coherent Circle Hough Transform

The Hough Transform is used to find circles in an image. This is done by convolving the edge image by a mask W containing a discrete approximation

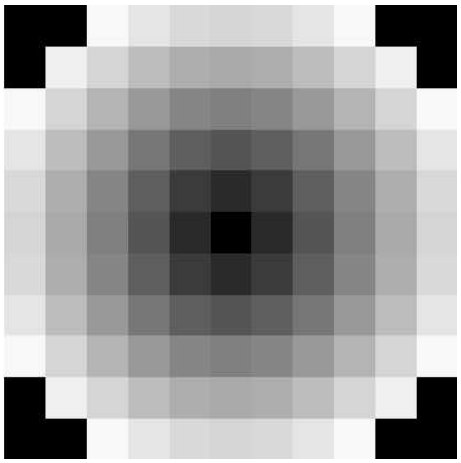

Figure 3.3: The phase coded annulus with $R_{min} = 0$ and $R_{max} = 5$. The grey scale represent the phase in radians with zero as black and 2π as white.

to a circle. The maxima of the output of this convolution is an estimate of the centers of the circles in the image. For example, the mask

$$W = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

could be used to find the center of circles of radius 4 in an image. The problem with this standard CHT is that, if the task is to find circles of radii between 3 and 6, then the resulting parameter space is three dimensional (the x and y coordinates and the radius), making the search for maxima harder. In our problem we are searching for particles which are "smeared out" differently according to the distance from the focal plane of the camera, so we have to look for circles of different radii.

The Coherent Circle Hough Transform [10] uses the concept of *Phase Coded Annulus*. This is a mask consisting of complex numbers, all with magnitude unity, but with the phase coded according to the distance away from the center of the mask. When convolved with a gradient image of a circle, the mask will integrate *coherently* (with equal phase), resulting in an output of large

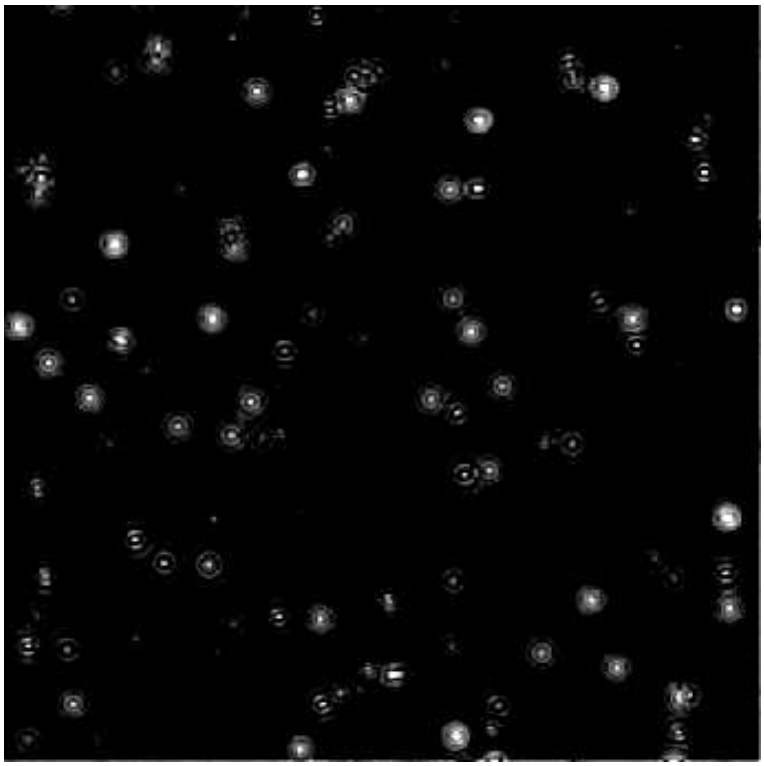

Figure 3.4: The coherent CHT with a chirp coded phase applied to the gradient image in Figure 3.2. Notice how the white dots correspond to centres of the circles in the same image.

magnitude when located right "above" the center of the circle.

The mask W^{pca} is defined as

$$W_{k,l}^{pca} = \begin{cases} e^{i\varphi_{k,l}} & \text{if } R_{min}^2 \leq k^2 + l^2 \leq R_{max}^2 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where R_{min} and R_{max} specify the range in which to detect circles. There are two proposed ways of coding the phase $\varphi_{k,l}$; a linear or a ramped. For a linear coding, $\varphi_{k,l}$ is given by

$$\varphi_{k,l}^{lin} = 2\pi \left(\frac{\sqrt{k^2 + l^2}}{R_{max} - R_{min}} \right)$$

whereas for the ramped (chirp) coding,

$$\varphi_{k,l}^{ram} = 2\pi\Big(\frac{\sqrt{k^2+l^2}}{R_{max}-R_{min}}\Big)\Big(\frac{R_{min}}{R_{max}} + \frac{(1-\frac{R_{min}}{R_{max}})\left(\sqrt{k^2+l^2}-R_{min}\right)}{R_{max}-R_{min}}\Big) \quad (3.4)$$

The phase of a linearly coded mask with $R_{min} = 0$ and $R_{max} = 5$ is illustrated in Figure 3.3.

In our application, we used the ramped coding style, so from now on, W^{pca} means the mask specified by (3.3) together with (3.4).

Let Q be the result after convolving W^{pca} with the gradient image E,

$$Q = W^{pca} \star E$$

where $\star$ denotes the convolution operator. In Figure 3.4 we can see the result after applying the W^{pca} to the gradient image shown in Figure 3.2.

Finding candidates

By locating local maxima in Q, we find suitable candidates for particle locations in the image. A location (i, j) is set as a candidate if the center pixel $Q_{i,j}$ is the maximum of all the pixels within a pre-determined distance of m from (i, j). In practice, this gives too many candidates so we require furthermore that $Q_{i,j}$ is above a pre-specified threshold value T.

Let C be the matrix of candidates. For every pixel (i, j) in Q do the following:

1. If $Q_{i,j} < T$, then $C_{i,j} = 0$ and go to step 3, otherwise proceed.
2. Let $S_{i,j} = \{k, l : (k-i)^2 + (l-j)^2 \leq m^2\}$.
   If $Q_{i,j} = \max_{k,l \in \mathcal{S}_{i,j}} Q_{k,l}$ then $C_{i,j} = 1$. If not, then $C_{i,j} = 0$.
3. Move to the next pixel and go to 1.

The position estimates in the image corresponds to the locations (i, j) where $Q_{i,j} = 1$. In Figure 3.5, the result of the procedure is displayed with crosses at the positions where we have a particle candidate.

Comments regarding the estimation of location

We see that we tend to get candidates on the surge of the particles which are located largely off-focus. In order to get the very faint particles (which actually correspond to particles located in the focal plane!) accepted as candidates

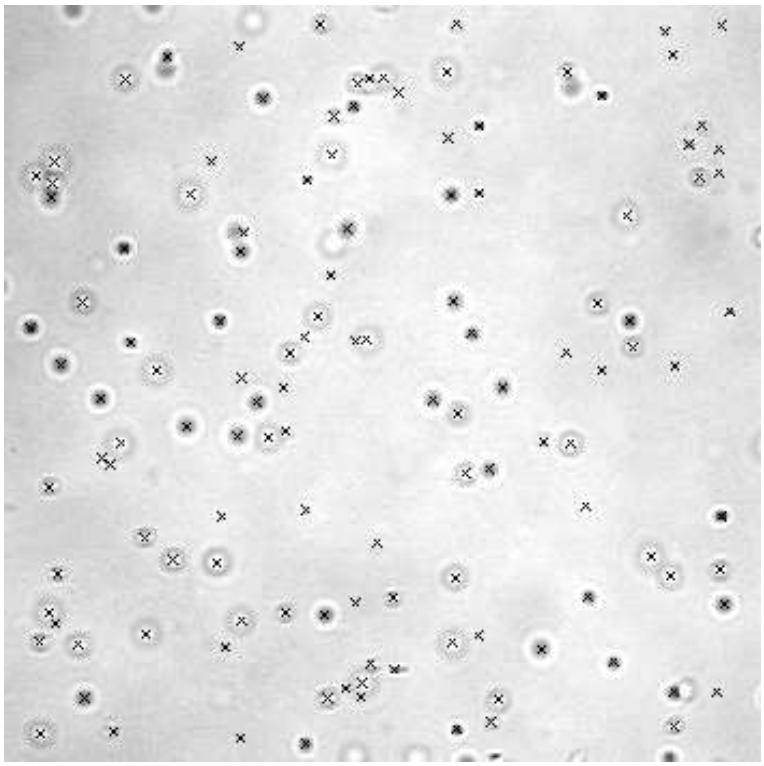

Figure 3.5: The resulting 127 candidate positions, marked with crosses, using the approach proposed in Section 3.2.

the only thing we could do was to lower the threshold, resulting in even more candidates on the surges.

A problem with the algorithm described above, is the quite large number of parameters (R_{min}, R_{max}, m and T). In the example above they have been chosen on a trial and error basis. For another image sequence, we probably have to choose different parameters, again by trial and error. The algorithm is fast though, and the result for our image sequence must be considered as satisfactory.

Since our goal is to estimate all three coordinates of the particles, we probably have to refine our method, from just looking for circles in the edge image, to a more elaborate shape search. One method is to use template matching where one has one template for what a particle looks like at each depth with respect to the focal plane. This is discussed further in Chapter 4.

3.3 Linking the estimates

To actually get trajectories, we have to link the position estimates from all the images in the sequence. The main problem of this step is that the number of particle candidates vary quite a lot between the images in the sequence. This is due to the fact that the image processing step described above is far from perfect; some particles are not found at all and some of the candidates do not correspond to real particles. New particles may also come in from the sides of the image (the image just covers a subset of the volume in which the particles are confined). So, generally, we have a matching problem with different number of points in each image. This is something which will be exploited in the future, see Chapter 4.

To get the trajectories shown in Figure 2.1, we manually removed particles in each of the 12 images which were not considered as real particles. Since this is quite a time-consuming task, we only looked at a subset of the images. Also, we just looked at 12 consecutive images.

Assume that we have observed n non-interacting particles, all performing a Brownian motion, in two consecutive images. Denote the distance between the i:th particle in the first image and the j:th particle in the consecutive image by $R_{i,j}$. The maximum likelihood estimator of the matching, when just considering two subsequent images, then becomes the permutation J of $\{1,\ldots,n\}$ that minimizes

$$\sum_{i=1}^{n} R_{i,J(i)}^2 \tag{3.5}$$

where the minimum is over all permutations J.

This is a classical problem in integer optimization called the Assignment Problem. Solving the problem directly, without modification, would involve $n!$ evaluations of the sum (3.5). However, there exists an algorithm which has complexity $O(n^4)$, see for example [16]. In the first step of this algorithm, we do what probably would be considered to be the "natural" thing: for each particle in one image look for the nearest particle in the consecutive image. Now, if this establishes a 1-1 correspondence between the particles in the two images, this must be the minimizer J of (3.5), and we are done. Otherwise proceed as in [16].

For the particles in Figure 2.1 we used this method and as one almost might expect from the figure, we could stop the algorithm after the first step for all eleven consecutive image pairs in the sequence.

Chapter 4

Future work

In this chapter we present a small selection of future research topics within this project.

4.1 Template matching

In the example illustrated in Figure 2.1, we only had particle position estimates in two dimensions. By using the fact that particles look different depending on their depth relative to the focal plane, we should in principle be able to estimate the depth, or z, coordinate. One way to do this is by template matching, where one, for a known prototype, or template, look for its best agreement, or matching, in an image (see for example [7]). Here we would need several templates, each corresponding to the appearance of a particle at certain depth of focus. Then the goodness-of-fit of each template is then evaluated at all possible positions in the image. A common matching criterion is the mean square difference between the pixel values of the image and the template.

The construction of templates will probably be based on empirical data. We have images of particles at different depths relative the focal plane so we know what a particle "looks like" at a certain distance away from the focal plane. An alternative is to build the templates mathematically. However, attempts in this direction has so far been fruitless since the optics involved in the image formation seems to be hard to handle. The reason is that the wavelength of light is of the same size as the diameters of the particles, so neither of the two standard approximations of optics, geometric and fourier optics (see for example the standard reference [9]), work for our situation.

Examples of where template matching successfully have been used are, Young

et al. [17] who used it to automatically identify and measure yeast cells in DIC microscopy, and Larsen and Rudemo [12] who used it to estimate the position of trees from aerial photographs.

4.2 Matching position estimates

The general solution to the matching of the position estimates in the images to trajectories, have to be able to handle the situation of different number of candidate particles in each step. What we saw in Chapter 3, is that it is unrealistic to think that an automatic image processing algorithm can find all our particles in all of the images. And even if this would be the case, we still have that particles can move "in and out of the image" or be occluded by other particles.

One approach is to follow the work of Walker [15] who used an idea based on Cross and Hancock [3]. They matched a point set in one image, to points in another, under the assumption that the relevant points where the same up to an affine transformation and a Gaussian i.i.d. measurement error term. The irrelevant points (the "noise") were assumed to distributed according to a two-dimensional Poisson process. Which points that where relevant or not was considered as unobserved data and they used the EM algorithm to find the maximum likelihood estimate of the affine transform parameters and the Gaussian noise variance.

Lund and Rudemo used a similar, but more advanced model in [13], where they matched tree-top position estimates from one image, with the a set of "true" positions.

Common for both methods is the fact that they match points from two images. In full generality, our matching procedure should work with the $N+1$ point sets, one from each image. Also, as we do it now, we consider the linking as a procedure separated from the analysis of the diffusion; we first estimate the trajectories, and then analyse these. At least in theory, we should be able to work with everything in one single step, taking care of the matching of the trajectories of the diffusing and fixed particle, while at the same time estimating the diffusion model parameters (since these are needed in the matching), together with ability to take into account the possibility for spurious particles ("noise particles") due to the imperfect image processing.

4.3 Diffusion coefficients from a distribution

One way to generalize the model in Chapter 2, would be to allow the diffusion coefficients of the moving particles to come from an arbitrary distribution as mentioned in Section 5.4 in [P]. Instead of assuming that the diffusing particles all have the same diffusion variance, we just assume that they have random diffusion variances, taken from the same distribution. Our goal then becomes to estimate this distribution.

Kiefer and Wolfowitz [11] show that under certain assumptions on this distribution of coefficients, the maximum likelihood estimate of the distribution, is consistent for all continuity points, as the number of observed particles goes to infinity.

Apart from being a more theoretically advanced model, it also lies closer to the truth in applications, where one usually works with so called poly-disperse solutions. That is, solutions with particles of different sizes, as opposed to mono-disperse, where they all have the same size.

Bibliography

[1] J.C. Crocker and D.G. Grier. Microscopic measurements of the pair interaction potential of charge-stabilized colloid. *Physical Review Letters*, 73:352–355, 1994.

[2] J.C. Crocker and D.G. Grier. Methods of digital video microscopy for colloidal studies. *Journal of Colloid and Interface Science*, 179:298–310, 1996.

[3] A.D.J. Cross and E.R. Hancock. Graph matching with a dual-step EM algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1263–1253, 1998.

[4] A.P. Dempster, N.M Laird, and D.B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, 39:1–38, 1977.

[5] D.F. Evans and H. Wennerström. *The colloidal domain. Where Physics, Chemistry, Biology, and Technology meet.* Wiley-VCH, New York, second edition, 1999.

[6] C.A. Glasbey and G.W. Horgan. *Image Analysis for the Biological Sciences.* Wiley, Chichester, UK, 1995.

[7] R.C. Gonzales and R.E Woods. *Digital Image Processing.* Addison-Wesley, Reading, Massachusetts, 1992.

[8] D.G. Grier. Colloids: A surprisingly attractive couple. *Nature*, 393:621, 1998.

[9] E. Hecht. *Optics.* Addison-Wesley, Reading, Massachusetts, third edition, 1998.

[10] D.J. Kerbyson and T.J. Atherton. Circle detection using Hough Transform Filters. In *Fifth International Conference on Image Processing and its applications.* IEEE Conference Publications, 4-6 July 1995.

[11] J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, 27:887–906, 1956.

[12] M. Larsen and M. Rudemo. Optimizing templates for finding trees in aerial photographs. *Pattern Recognition Letters*, 19:1153–1162, 1998.

[13] J. Lund and M. Rudemo. Models for point processes observed with noise. *Biometrika*, 87:235–249, 2000.

[14] G. McLachlan and T. Krishnan. *The EM algorithm and Extensions*. Wiley-Interscience, New York, first edition, 1997.

[15] G. Walker. *Robust, non-parametric and automatic methods for mathcing spatial point patterns*. PhD thesis, Department of Statistics, University of Leeds, June 1999.

[16] L.A. Wolsey. *Integer Programming*. Wiley-Interscience, New York, first edition, 1998.

[17] D. Young, C.A. Glasbey, A.J. Gray, and Martin N.J. Towards automatic cell identification in dic microscopy. *Journal of Microscopy*, 192:186–193, 1998.

Estimation of the diffusion coefficient in a mixture model with diffusing and fixed particles

Mats Kvarnström
Chalmers University of Technology and Göteborg University

October 22, 2002

Abstract

Particle positions have been observed and estimated in a series of images. The particles are assumed to perform a Brownian motion, however some of them seem to be fixed. A model is introduced with two kinds of particles, diffusing and fixed. To each particle position estimate we assume an additive normal measurement error. The parameter of the model consists of the diffusion variance, the measurement error variance, and the proportion of diffusing particles. The problem can be considered as an incomplete data problem since we do not know *a priori* which particles are really diffusing. The complete data is of curved exponential type and the observed data is a mixture of two normal components. The maximum likelihood estimator is computed via the EM algorithm. The estimator is shown to be strongly consistent and asymptotically normal, as the number of particles approaches infinity, under a reasonable restriction on the parameter space.

Key words: asymptotic normality, curved exponential family, discretely observed diffusion, EM algorithm, measurement error, mixture distribution, strong consistency

1 Introduction

This article deals with the estimation of the diffusion variance (or equivalently, the diffusion coefficient) of colloidal particles. Particle positions have been observed and estimated in a series of images (frames) recorded on a video microscope using more or less standard image processing algorithms and tools. The

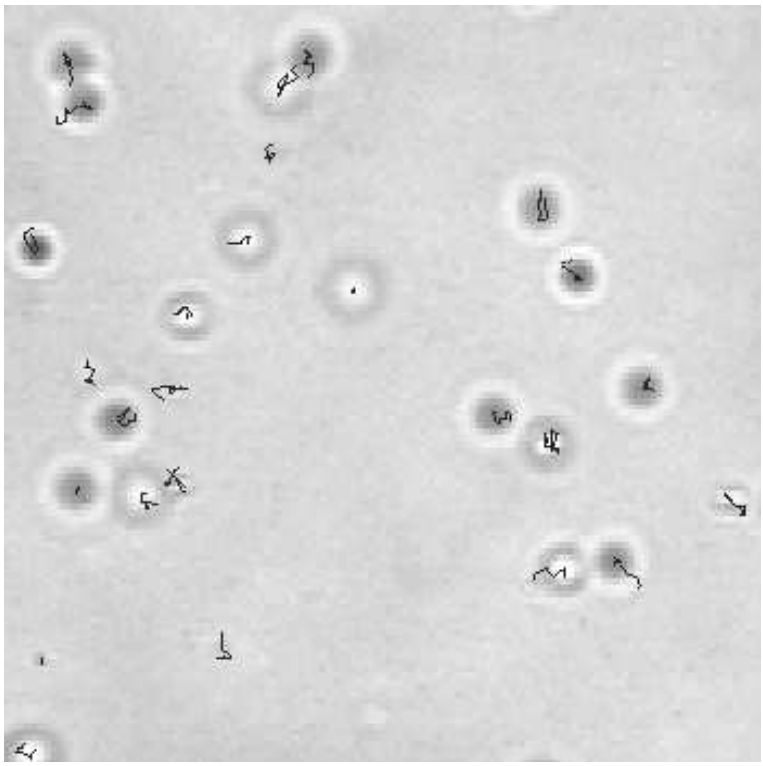

Figure 1: The 26 trajectories estimated in a sequence of 12 images together with the first image in the sequence. Notice that three of the particles seem to be fixed.

position estimates of the particles are then linked so that we get a trajectory for each particle in the sequence.

The particles are performing a Brownian motion in three dimensions and they move independently of each other. The naive estimate of the diffusion variance is the mean of the squared increments. If we, however, assume that the position estimates of the particles are imperfect, i.e. if we assume a measurement error, this estimate of the diffusion variance will be biased.

Another complicating fact is that some of the observed particles are not moving but are instead either particles adsorbed on the objective or cover glass of the specimen, or "false" particles which are due to for example defects in the optics of the microscope. One solution to this problem is of course to remove these false particles manually. This is not satisfactory from a statistical point of view, first because we should be able to do it using statistical methods, second because

these particles actually gives us information on the measurement error.

The proposed method takes care of both of these problems by assuming a mixture distribution of diffusing and fixed particles and then estimate the relevant parameters using maximum likelihood estimation. For fixed number of observed increments the estimator is shown to be strongly consistent and asymptotically normally distributed, as we let the number of particles go to infinity.

An example of what the situation may look like, can be seen in Figure 1. The figure shows the initial image in a sequence of 12 images, together with the position estimates of a major part of the particles in the subsequent 11 images, thereby forming the estimated trajectories of the particles. By manual inspection, we made sure that no change of the identities of the particles occurred in the process of converting the position estimates in the images into trajectories. The time interval between two images is 40 milliseconds. The particles are spherical, made of polystyrene and are all equal in size, 494 nm in diameter. The apparent difference in size and brightness are due to an out-of-focus effect and depend on their placement in depth relative to the focal plane. Particles above the focal plane are bright in the middle and dark on the circumference and vice versa for the particles below the focal plane. The depicted size of a particle is also increased the further away from the focus plane it is. Here three of particles seem to be fixed; one adsorbed on the cover glass, one on the objective glass, and one which probably correspond to a defect in the optics.

1.1 Outline of the paper

The article is organized as follows.

In Section 2 we introduce a model with two kinds of particles, diffusing and fixed, both observed with additive measurement error on the position estimates. The observation length is $N+1$ frames. We have three parameters in the model, σ^2 is the diffusion variance, σ_ϵ^2 is the variance of the measurement error on the position estimates, and p is the proportion of diffusing particles. The problem can now be considered as a missing data problem since the only way to infer whether a particle is diffusing or not is by the observed movement of the particle. In this section we also look at the structure of the covariance matrices for the two kinds of particles.

In Section 3 we introduce two concepts of data, observed and complete data. The observed data is the observed increments of each particle and the complete data is the observed data together with the classification variable of each particle (indicating whether it is diffusing or fixed). We also look at the different densities these two kinds of data correspond to. In particular, the observed data is of finite mixture type.

The likelihood is discussed in Section 4 together with the EM algorithm (see

Dempster et al. (1977) and McLachlan and Krishnan (1997)) in Section 4.1 and some basic theory regarding this method of finding the maximum likelihood estimate.

In Section 5 we study the asymptotic properties of the estimator when we keep the observation length and and let the number of particles go to infinity. We show that the estimator of the triple $(\sigma^2, \sigma_e^2, p)$ using only the observed increments is strongly consistent and asymptotically normally distributed under a small but reasonable restriction on the parameter space.

Finally, in Section 6, we use the model assumption and estimate the diffusion variance for the data corresponding to the trajectories seen in Figure 1.

2 The model

Denote the true and the observed position of a particle at time $k = 0, \ldots, N$ by R_k and S_k, respectively, where S_k is the true position with measurement noise added to it. We arrive at the following state-space model

$$\begin{array}{rclcl} R_k & = & R_{k-1} & + & w_k \\ S_k & = & R_k & + & e_k \end{array} \tag{1}$$

where w is the position increment of the motion of the particle and e the measurement error of the position.

The particle is performing a Brownian motion so the w_k:s are i.i.d. zero mean normally distributed random variables with variance σ^2. For a fixed particle, w_k is zero for all k (an alternative is to think about this as σ^2 being equal to zero for fixed particles). The errors e_k are also assumed to be i.i.d. zero mean normally distributed random variables with variance σ_e^2, independent of the true position of the particle, of other particles, and of the increments w_k.

The initial value, R_0 is assumed to be a constant.

Let n be the number of observed particles and introduce an indicator Z_i to each particle to be one if the i:th particle is diffusing (performing a Brownian motion, $\sigma^2 > 0$) and zero if fixed ($\sigma^2 = 0$). Let $\{Z_i\}_{i=1}^n$ be i.i.d. and introduce a third parameter p, defined as

$$p = \mathbf{P}(Z_i = 1).$$

The model can easily be extended to noisy observations of a Brownian motion in d dimensions if we assume the measurement error in each dimension to be distributed as w_k above and independent of each other. Then a particle follows same the state-space model (1) in each dimension independently of each other.

For ease of notation, we will assume that d is one. The only exception from this is in Section 6, which deals with the analysis on the trajectories plotted in Figure 1.

2.1 Covariance matrix of the observed increment vector

We define the observed increments for a particle as $Y_k = S_k - S_{k-1}$, $k = 1, \ldots, N$. The covariance matrix of the increment vector, $Y = [Y_1, \ldots, Y_N]^T$, is

$$\Sigma_1 = \sigma^2 I + \sigma_e^2 T \tag{2}$$

for a diffusing particle and

$$\Sigma_0 = \sigma_e^2 T$$

for a fixed particle, where I is the N times N identity matrix and T is the tri-diagonal matrix defined as

$$T = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}.$$

We see from the covariance matrix above that the measurement noise in the position induces a dependence between the increments, which originally, by definition of a Brownian motion, were independent.

2.2 Transformation of the increment vector

To make our formulas look cleaner in the subsequent sections, we use some basic linear algebra to transform the increment vector so that the transformed vectors become uncorrelated.

In (2), Σ_1 has the same eigenvectors as T since every vector is an eigenvector to I. If we denote the eigenvalues of T as λ_k, $k = 1, \ldots, N$, the eigenvalues of Σ_1 are

$$\gamma_k = \sigma^2 + \sigma_e^2 \lambda_k,\ k = 1, \ldots, N.$$

Let U have the eigenvectors of T as columns. Then we can write, by the spectral decomposition theorem, $T = U \Lambda U^T$, where $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_N\}$. So if

$$\tilde{Y} = U^T Y \tag{3}$$

is the transformed increment vector, its covariance matrix will be diagonal:

$$\begin{aligned} \mathbf{Var}\{\tilde{Y}\} &= U^T \mathbf{Var}\{Y\} U \\ &= U^T(\sigma^2 I + \sigma_e^2 U \Lambda U^T) U \\ &= \sigma^2 I + \sigma_e^2 \Lambda \\ &= \mathrm{diag}\{\gamma_1, \ldots, \gamma_N\} \end{aligned} \tag{4}$$

for a diffusing particle and likewise, with $\sigma^2 = 0$, for a fixed particle. The dependence between the increments is now "hidden" in U and Λ, which do not depend on σ^2 or σ_e^2, but only on the length of the increment vector N, which of course is known.

3 Data and densities

The observed data consists of the vectors of noise corrupted increments Y_i while the classification variables Z_i are unobserved. Together, they make up the complete data, denoted $X_i = (Y_i, Z_i)$, $i = 1, \ldots, n$.

The probability density function of the complete data X is

$$g_c(x\,;\sigma^2, \sigma_e^2, p) = [p f_1(y\,;\sigma^2, \sigma_e^2)]^z \, [(1-p) f_0(y\,;\sigma_e^2)]^{1-z} \tag{5}$$

where f_1 and f_0 are the pdf of a zeros mean N-variate normally distributed random variable with covariance matrices Σ_1 and Σ_0, respectively. Using the transformed increment vector $\tilde{Y}$, we write

$$f_1(y\,;\sigma^2, \sigma_e^2) = \frac{1}{(2\pi)^{N/2} \prod_{k=1}^{N} (\sigma^2 + \lambda_k \sigma_e^2)^{1/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{N} \frac{\tilde{y}_k^2}{\sigma^2 + \lambda_k \sigma_e^2} \right\} \tag{6}$$

and

$$f_0(y\,;\sigma_e^2) = \frac{1}{(2\pi)^{N/2} \prod_{k=1}^{N} (\lambda_k \sigma_e^2)^{1/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{N} \frac{\tilde{y}_k^2}{\lambda_k \sigma_e^2} \right\} \tag{7}$$

$$= \frac{1}{(2\pi\sigma_e^2)^{N/2} (N+1)^{1/2}} \exp\left\{ -\frac{1}{2\sigma_e^2} \sum_{k=1}^{N} \frac{\tilde{y}_k^2}{\lambda_k} \right\} \tag{8}$$

where $\tilde{y} = U^T y$ from (3). The second equality in the last expression, comes from the fact that $\prod \lambda_k = |\Lambda| = |T| = N + 1$.

In the d dimensional case, f_i will be a dN-variate normal density with d independent parts, one in each in each dimension, since each coordinate process is independent of the others.

The probability density of the observed data, Y, we get by integrating (5) over Z

$$g(y\,;\sigma^2,\sigma_e^2,p) = pf_1(y\,;\sigma^2,\sigma_e^2) + (1-p)f_0(y\,;\sigma_e^2). \tag{9}$$

Our observed data is a finite mixture of two normal components. For a thorough account on finite mixture models and their applications, we refer to McLachlan and Peel (2000).

4 Likelihood

Denote our parameter $\theta = [\sigma^2, \sigma_e^2, p]^T$.

The complete likelihood L_c induced by the complete data (increments and classification variables) from n observed particles is

$$L_c(\theta) = \prod_{i=1}^{n} [pf_1(y_i\,;\sigma^2,\sigma_e^2)]^{z_i}\,[(1-p)\,f_0(y_i\,;\sigma_e^2)]^{1-z_i} \tag{10}$$

However, our observed data consists of only the increments so the observed likelihood becomes

$$L(\theta) = \prod_{i=1}^{n} pf_1(y_i\,;\sigma^2,\sigma_e^2) + (1-p)\,f_0(y_i\,;\sigma_e^2) \tag{11}$$

4.1 The EM algorithm

A intuitive method to get the maximum likelihood estimate from our observed data is to use a method whose name, the EM algorithm, comes from the article by Dempster et al. (1977), but whose essence actually was introduced and used, for the special case of finite mixtures of distributions from the exponential family, by Hasselblad (1969). Further examples of its use, before it was actually called the EM algorithm, can be found in Day (1969), Behboodian (1970) and Sundberg (1976). For an overview of the theory and applications of the method we refer to McLachlan and Krishnan (1997).

The method uses the simple structure of the complete likelihood together with estimates of the unobserved data in a iterative scheme.

4.1.1 Notation

Let k be the conditional density of the unobserved data Z, given the observed Y. Then

$$k(z|\,y;\theta) = \frac{g_c(x;\theta)}{g(y;\theta)}$$

Taking the logarithm and re-arranging, we get

$$\log g(y;\theta) = \log g_c(x;\theta) - \log k(z|\, y;\theta) \tag{12}$$

Denote by $L(\theta)$ and $L_c(\theta)$, the observed and the complete data likelihoods

$$L(\theta) = g(y;\theta)$$

$$L_c(\theta) = g_c(x;\theta)$$

and take the conditional expectation of (12) given Y, at the parameter θ'

$$\log L(\theta) = \mathbf{E}_{\theta'}\{\log g_c(x;\theta)|y\} - \mathbf{E}_{\theta'}\{\log k(z|\, y;\theta)|y\}$$

and denote the first term $Q(\theta|\theta')$ and the second $H(\theta|\theta')$.

Let furthermore

$$S(y;\theta) = \partial \log L(\theta)/\partial\theta$$

and

$$S_c(x;\theta) = \partial \log L_c(\theta)/\partial\theta$$

be the score functions.

4.1.2 Method

The EM algorithm consists of two steps at each iteration. Assume $\theta^{(k)}$ is the estimate of θ from the k:th iteration step. Then we do:

- E-step: Compute $Q(\theta|\theta^{(k)})$
- M-step: Choose $\theta^{(k+1)} \in \text{argmax}\, Q(\theta|\theta^{(k)})$

Since $H(\theta|\theta^{(k)}) \leq H(\theta^{(k)}|\theta^{(k)})$ for all θ by Jensen's inequality, the rule of choosing $\theta^{(k+1)}$ as a maximizer of $Q(\theta|\theta^{(k)})$ gives us that

$$L(\theta^{(k+1)}) \geq L(\theta^{(k)})$$

guaranteeing that we approach a local maximum of the likelihood. In practice, we iterate until some sort of convergence criterion is met.

Notice that there is no guarantee that we converge to the *global* maximum of the likelihood function, and thus at the actual maximum likelihood estimate.

The EM algorithm should simply be thought of a numerical method for maximizing the likelihood. Often, it suffers from painstakingly slow convergence, and then a Newton-Raphson approach usually does better. However, when the data is considered to have missing values, it is very appealing to use it since we also get estimates of the missing values. We write "is considered" because the missing values may be a theoretical construction only. In our problem, though, it is natural to think of the classification variables as being missing data.

4.1.3 Finite mixtures

When the data comes from a mixture, the E-step consists of estimating the unobserved data, i.e. the classification variables. In the M-step we maximize the complete likelihood (10) using the estimated classification variables, $\hat{Z}_i$, from the E-step together with our data Y_i:

- E-step: For each $i = 1, \ldots, n$, compute

$$\hat{Z}_i = \mathbf{E}_{\theta^{(k)}}\{Z_i|Y_i\} = \frac{p^{(k)} f_1(y_i; \Sigma_1^{(k)})}{p^{(k)} f_1(y_i; \Sigma_1^{(k)}) + (1-p^{(k)}) f_0(y_i; \Sigma_0^{(k)})}$$

- M-step: Maximize $\mathbf{E}_{\theta^{(k)}}\{\log L_c(\theta)|y\} =$

$$= \sum_{i=1}^{n} \hat{Z}_i \log\{p f_1(y_i\,; \sigma^2, \sigma_e^2)\} + (1-\hat{Z}_i) \log\{(1-p) f_0(y_i\,; \sigma_e^2)\}$$

  with respect to $\theta = (\sigma^2, \sigma_e^2, p)$.

In this application of the EM algorithm, each of the two steps has a probabilistic meaning; in the E-step we classify each particle using a quadratic discriminant rule, and in the M-step we use these classifications as if we had the complete data. Note however that the classifications are not just zero or one, but any number in between.

A fast, Newton-Raphson based, computational method for the M-step can be found in (30) of Appendix C.

4.1.4 Information matrix

Now we are going to explore how the information matrix from our observed data relates to the information matrix from the complete data. This will also give us a computationally efficient way of calculating the observed information when using the EM algorithm. Appropriate regularity conditions allowing us to differentiate under the integral are assumed in the following. In our application this is true since we are dealing with exponential families, see for example van der Vaart (1999).

Let

$$I(\theta; y) = -\partial^2 \log L(\theta)/\partial\theta\partial\theta^T$$

and

$$I_c(\theta; x) = -\partial^2 \log L_c(\theta)/\partial\theta\partial\theta^T$$

Using the following version of (12)

$$\log L(\theta) = \log L_c(\theta) - \log k(z|y;\theta),$$

and differentiating twice and taking conditional expectation of z given y, we get

$$I(\theta;y) = \mathcal{I}_c(\theta;y) - \mathcal{I}_m(\theta;y) \tag{13}$$

where

$$\mathcal{I}_c(\theta;y) = \mathbf{E}_\theta\{I_c(\theta;x)|y\}$$

and

$$\mathcal{I}_m(\theta;y) = -\mathbf{E}_\theta\{\partial^2 \log k(z|y;\theta)/\partial\theta\partial\theta^T|y\}$$

corresponding to the conditional expectation of the information matrix of the complete data given y, and the missing information, respectively.

In Louis (1982), it is shown that $\mathcal{I}_m$ can be expressed as

$$\mathcal{I}_m(\theta;y) = \mathbf{E}_\theta\{S_c(X;\theta)S_c^T(X;\theta)|y)\} - S(y;\theta)S^T(y;\theta). \tag{14}$$

This is nice, first since $S(y;\theta) = 0$ at the MLE $\hat{\theta}$ and secondly because now the observed information matrix at $\hat{\theta}$ is

$$I(\hat{\theta};y) = \mathcal{I}_c(\hat{\theta};y) - [\mathbf{E}_\theta\{S_c(X;\theta)S_c^T(X;\theta)|y)\}]_{\theta=\hat{\theta}} \tag{15}$$

where both terms easily can be computed in the last M-step in the EM algorithm since the first term is actually the negative of the Hessian of the function to maximize in the M-step, and this is often used in the actual maximization.

Denote the expected information matrix by $\mathcal{I}(\theta)$ which can be expressed as

$$\mathcal{I}(\theta) = \mathcal{I}_c(\theta) - \mathbf{E}_\theta\{\mathcal{I}_m(\theta;Y)\} \tag{16}$$

by taking expectation of (13) over the distribution of Y.

5 Asymptotics

Is this section we are going to study the asymptotic properties of the estimator as the number of particles n grows large. As it turns out, our maximum likelihood estimator is both strongly consistent and asymptotically normal. First we address some important issues regarding the data and the parameter space.

The complete data comes from the exponential family of distributions, see for example Lindsey (1996). If $N \neq 1$ however, it is non-regular or curved, since the parameter space is 3-dimensional and the dimension of the sufficient statistics is $N + 2$ (see the Appendix for a derivation of this). The case $N = 1$ is not very

interesting though since we think of our problem as studying a video sequence of images of particles.

Let Ω be the parameter space consisting of those θ defining valid finite mixture densities (9). $\Omega = \{\theta = [\sigma^2, \sigma_e^2, p]^T : p \in [0,1], \sigma^2 > 0, \sigma_e^2 > 0\}$. The true parameter point θ_0 is assumed to lie in the interior of Ω, denoted int(Ω).

Often when one deals with finite mixtures, there is a problem of identifiability, i.e. that a permutation of the parameters in the model yields the same distribution. In our model, and as long as the true parameter θ_0 lies in the interior of Ω, we do not have this problem since the two distributions in the mixture are not interchangable.

The asymptotics when using complete data is covered in Appendix D.

5.1 Existence of a maximum likelihood estimator

To guarantee that the likelihood has a global maximizer of for each n, we restrict the parameter space Ω, by using an idea from Hathaway (1985). For fixed $c \in (0,1)$, define Ω_c to be the subset of Ω such that

$$0 < c \leq \frac{\sigma^2}{\sigma_e^2} \leq c^{-1} < \infty \tag{17}$$

This restriction means that we do not allow the "signal-to-noise" ratio to be too small, neither too big.

Lemma 1. *Let $\{Y_1, \ldots, Y_n\}$ be a set of observations from the finite mixture specified by the density (9). Then, with probability one, there exists a global constrained maximizer of $L(\theta)$ in Ω_c.*

Proof. The idea is to show that

$$\sup_{\theta \in \Omega_c} L(\theta) = \sup_{\theta \in K} L(\theta)$$

for some appropriate, compact $K \subset \Omega$.

With probability one, the increment vectors will all be different from zero. Therefore all the terms in the likelihood will stay bounded. Also, it will go to zero if both σ^2 and σ_e^2 either go to zero or to infinity. By condition (17) above however, it is enough that one of the two variances goes to zero or infinity; the other variance "will follow".

So, there exists constants a_i and b_i such that $K = \{\theta \in \Omega_c : a_1 \leq \sigma_e^2 \leq a_2,\ b_1 \leq \sigma^2 \leq b_2\}$ gives the desired result. □

Remark. *Without the condition (17), our trouble spots are*

- $L \to \prod_{i=1}^{n} f_0(y_i\,;\sigma_e^2)$ *as* $\sigma^2 \to 0$
- $L \to p^n \prod_{i=1}^{n} f_1(y_i \,|\, \sigma^2, 0)$ *as* $\sigma_e^2 \to 0$
- $L \to (1-p)^n \prod_{i=1}^{n} f_0(y_i\,;\sigma_e^2)$ *as* $\sigma^2 \to \infty$

A maximum hence exists, but it does not necessarily have to be unique for finite n: If $\hat{p} = 0$, we see that σ^2 is "free". Likewise, if $\hat{p} = 1$ and $N = 1$, all values of σ^2 and σ_e^2 satisfying $\sigma^2 + 2\sigma_e^2 = c$ for some constant c, are maximum likelihood estimators.

5.2 Consistency

5.2.1 Special case, $N = 1$

When $N = 1$, the complete data is of regular exponential type. Sundberg (1974) gives the consistency and asymptotic normality of the maximum likelihood estimator $\hat{\theta}_n$, under the single condition that the information matrix $\mathcal{I}(\theta)$ is positive definite at the true parameter point θ_0. Since Lemma 2 below says that this is true for all $\theta_0 \in \text{int}(\Omega)$, we are actually done for $N = 1$, both with the consistency and the asymptotic normality.

5.2.2 Generally, $N \geq 1$

To prove consistency of the maximum likelihood estimator for general N, we verify that Wald's classical conditions for the mixture density g in (9) are satisfied when the true parameter is in Ω_c. In the process, we use results from Redner (1981).

Theorem 1. *Let the true parameter point θ_0 be in Ω_c and let $\hat{\theta}_n$ be the global maximizer of $L(\theta)$ over Ω_c, for each n. Then*

$$\mathbf{P}\{\hat{\theta}_n \to \theta_0 \ \textit{as}\ n \to \infty\} = 1$$

Proof. Wald's conditions are enumerated as in Redner (1981) to 1 through 6. We refer the reader to that article.

Conditions 1,2,4' and 5 are satisfied for Ω and the mixture component densities f_1 and f_0. The proof of Redner's Theorem 5 shows that Conditions 2 and 4 are satisfied for the mixture density $g = pf_1 + (1-p)f_0$. If we restrict Ω to Ω_c as defined above (17), then also Conditions 3 and 6 are satisfied, giving us the result by applying Theorems 1 and 2 from Wald (1949). □

Remark 1. *The extra condition* (17) *helps us in the process of first to prove that an maximum likelihood estimator exists for all n and second, to prove that Condition 3 of Redner (1981), $L(\theta_i) \to 0$ if $d(\theta_0, \theta_i) \to \infty$, where d means Euclidean distance.*

Remark 2. *The restriction* (17) *of the parameter space also gives us consistency under an expanded model with a drift term in the diffusion together with systematic position measurement errors, that is, if the mixture components have non-zero expected value and we need to estimate these as well. Also, the conclusion of Lemma 1 holds if the number of observations n is larger than three (one more than the number of mixture components).*

5.3 Asymptotic normality

Sufficient conditions for the asymptotic normality of the maximum likelihood estimate $\hat{\theta}_n$ can be found in for example Theorem 5.23 in van der Vaart (1999). Since we have consistency and that $\log g(y;\theta)$ is smooth, what remains is to be proven is that the map $\theta \mapsto \mathbf{E}_{\theta_0} \log g(Y;\theta)$ admits a second order Taylor expansion around $\theta_0 \in \text{int}(\Omega)$ with non-singular second derivative matrix. In other words, what we have to prove is that the expected information matrix $\mathcal{I}(\theta_0)$ is positive definite.

Theorem 2. *Let $\theta_0 \in int(\Omega_c)$ be the true parameter point. Then the maximum likelihood estimator $\hat{\theta}_n$ is asymptotically normal, i.e.*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, \mathcal{I}(\theta_0)^{-1}) \tag{18}$$

as $n \to \infty$.

By the discussion above, the result follows from the next lemma.

Lemma 2. *The information matrix $\mathcal{I}(\theta)$ is positive definite for all $\theta \in int(\Omega)$.*

Proof. Positive definiteness means that $a^T \mathcal{I}(\theta) a > 0$, for all $a \in \mathbb{R}^3 \setminus 0$.

Now, since $\mathcal{I}(\theta)$ is the variance of the score function $\partial \log g(Y;\theta)/\partial\theta$, $a^T \mathcal{I}(\theta) a$ is the variance of the linear combination $a^T \partial \log g(Y;\theta)/\partial\theta$.

So, what we have to prove is that

$$\mathbf{Var}\{a^T \frac{\partial \log g(Y;\theta)}{\partial \theta}\} > 0$$

for all $a \in \mathbb{R}^3 \setminus 0$.

Assume the opposite. Then we have, with probability one, that

$$a^T \frac{\partial \log g(Y;\theta)}{\partial \theta} = 0 \tag{19}$$

for some $a \in \mathbb{R}^3 \setminus 0$ since the mean of the score is zero.

Writing out the components of the score function, we have

$$\begin{aligned}
\frac{\partial \log g}{\partial \sigma^2} &= \frac{p\frac{\partial f_1}{\partial \sigma^2}}{pf_1 + (1-p)f_0} \\
\frac{\partial \log g}{\partial \sigma_e^2} &= \frac{p\frac{\partial f_1}{\partial \sigma_e^2} + (1-p)\frac{\partial f_0}{\partial \sigma_e^2}}{pf_1 + (1-p)f_0} \\
\frac{\partial \log g}{\partial p} &= \frac{f_1 - f_0}{pf_1 + (1-p)f_0}
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial f_1}{\partial \sigma^2} &= \frac{1}{2}\sum_{k=1}^{N}\left(\frac{\tilde{y}_k^2}{(\sigma^2 + \lambda_k\sigma_e^2)^2} - \frac{1}{\sigma^2 + \lambda_k\sigma_e^2}\right) f_1(y\,;\sigma^2,\sigma_e^2) = k_1(y) f_1(y\,;\sigma^2,\sigma_e^2) \\
\frac{\partial f_1}{\partial \sigma_e^2} &= \frac{1}{2}\sum_{k=1}^{N}\left(\frac{\lambda_k\tilde{y}_k^2}{(\sigma^2 + \lambda_k\sigma_e^2)^2} - \frac{\lambda_k}{\sigma^2 + \lambda_k\sigma_e^2}\right) f_1(y\,;\sigma^2,\sigma_e^2) = k_2(y) f_1(y\,;\sigma^2,\sigma_e^2) \\
\frac{\partial f_0}{\partial \sigma_e^2} &= \left(\frac{1}{2(\sigma_e^2)^2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{\lambda_k} - \frac{N}{2}\frac{1}{\sigma_e^2}\right) f_0(y\,;\sigma_e^2) = k_3(y) f_0(y\,;\sigma_e^2)
\end{aligned}$$

We write (19) as

$$a_1 p\frac{\partial f_1}{\partial \sigma^2} + a_2\Big[p\frac{\partial f_1}{\partial \sigma_e^2} + (1-p)\frac{\partial f_0}{\partial \sigma_e^2}\Big] + a_3\Big[f_1 - f_0\Big] = 0$$

Re-arranging and noticing that $f_1(y) > 0$ and $f_0(y) > 0$ for all y, we see that this is equivalent to saying that

$$\begin{cases} a_1 p\, k_1(Y) + a_2 p\, k_2(Y) + a_3 = & 0 \\ a_2(1-p)k_3(Y) - a_3 = & 0 \end{cases} \tag{20}$$

Since $k_1(Y)$, $k_2(Y)$, and $k_3(Y)$ are non-zero with probability one, we have a contradiction because (20) is satisfied only if a is zero. □

Remark. *Notice that* (20) *is satisfied for non-zero* a *if* $p = 0$. *This is also what we would expect since then we have no information on* σ^2. *Also, if* $N = 1$, *then* $k_2(Y) = \lambda_1 k_1(Y)$, *so if* $p = 1$, (20) *is satisfied as long as* $a_1 + \lambda_1 a_2 = 0$ *and* $a_3 = 0$.

5.4 Note on a further generalization

An interesting article with relevance to our problem, is Kiefer and Wolfowitz (1956). It deals with the consistency of a maximum likelihood estimator when there are infinitely many incidental parameters present. These incidental parameters could be, in a generalization of our problem, the variance of the Brownian motion σ^2 if all diffusing particle have different diffusion coefficients. This corresponds to a so called poly-disperse solution in contrast to our present problem, which is mono-disperse (every particle has the same diffusion coefficient).

Assume that for each $i = 1, \ldots, n$, we have that Y_i is N-variate normally distributed random variable with mean zero and covariance matrix $\Sigma_i = I\sigma_i^2 + T\sigma_e^2$. Then, in the language of Kiefer and Wolfowitz (1956), the σ_i^2:s are the incidental parameters and σ_e^2 the parameter (even though, in our context, these names are misleading since we consider it to be the other way around). Notice that if the σ_i^2:s are constants and different for each i we only observe one increment vector Y_i for each σ_i^2. Obviously the estimates of the σ_i^2:s can not be consistent. It turns out however, that if we consider σ_i^2, $i = 1, \ldots, n$ to independent random variables with common (but unknown) distribution function F, and under certain assumption on F, the maximum likelihood estimator of F converges to F at every point of continuity of F, almost surely. Also, the maximum likelihood estimator of σ_e^2 is strongly consistent.

The case discussed in this section is of course a special case of these σ_i^2 coming from an unknown distribution function with only two values: let

$$F(x) = \begin{cases} 0 & \text{when } x < 0 \\ 1 - p & \text{when } 0 \leq x < \sigma^2 \\ 1 & \text{when } \sigma^2 \leq x \end{cases}$$

6 Application

The data from the example in the introduction were analysed with the EM algorithm. The positions of the 26 particles were estimated in two dimensions in each image using a circle detection algorithm. The total number of frames were 12, so $N = 11$.

By manual inspection, we concluded that three particles in Figure 1 seem to be fixed, and refer to them as particle 1 to 3, where 1 is the big white in the middle, 2 the big black to the left, and 3 the seemingly "false" particle, probably due to an optics defect, in the lower left corner. The remaining 23 are considered as diffusing particles.

6.1 Results

We applied the EM algorithm to the observed data with initial value $\theta^0 = [1,\ 1,\ 0.5]^T$. We stopped when the change of the Z_i:s between two consecutive E-steps was smaller than 10^{-6}. This criterion was satisfied after 3 steps with the resulting estimates

$$\begin{aligned} \hat{\sigma}^2 &= 2.2058 \\ \hat{\sigma}_e^2 &= 0.3172 \\ \hat{p} &= 0.8847 \end{aligned} \tag{21}$$

where the unit for the first two is pixel^2. The estimated classification variables $\hat{Z}_i$ were

$$\begin{aligned} \hat{Z}_1 &= 1.049 \cdot 10^{-5} \\ \hat{Z}_2 &= 1.528 \cdot 10^{-5} \\ \hat{Z}_3 &= 2.473 \cdot 10^{-3} \\ \hat{Z}_i &= 1.000 \quad \text{for} \quad i = 4, \ldots 26 \end{aligned} \tag{22}$$

in good correspondence with our manual classification.

6.2 Observed information matrix

Using the result (15) to compute the observed information matrix at the MLE

$$\begin{aligned} I(\hat{\sigma}^2, \hat{\sigma}_e^2, \hat{p}\,; Y) &= \begin{bmatrix} 33.75 & 52.75 & 0 \\ 52.75 & 476.6 & 0 \\ 0 & 0 & 254.9 \end{bmatrix} - \begin{bmatrix} 0.034 & 0.153 & -0.090 \\ 0.153 & 0.691 & -0.405 \\ -0.090 & -0.405 & 0.240 \end{bmatrix} \\ &= \begin{bmatrix} 33.72 & 52.59 & 0.090 \\ 52.59 & 475.9 & 0.405 \\ 0.090 & 0.405 & 254.7 \end{bmatrix} \end{aligned}$$

were the second term of the upper row corresponds to the missing information due to lack of the unobserved classification variables.

The inverse of this is

$$I^{-1}(\hat{\sigma}^2, \hat{\sigma}_e^2, \hat{p}\,; Y) = \begin{bmatrix} 0.0358 & -0.0040 & 0.0000 \\ -0.0040 & 0.0025 & 0.0000 \\ 0.0000 & 0.0000 & 0.0039 \end{bmatrix} \tag{23}$$

which gives us an approximate variance of the estimate of $\hat{\sigma}^2$ to

$$\mathbf{Var}\{\hat{\sigma}^2\} \simeq 0.0358 \tag{24}$$

6.3 Comparison with the theoretical diffusion coefficient

The estimated $\hat{\sigma}^2$ above corresponds to an estimated diffusion coefficient of

$$\hat{D} = 0.893\,\mu\text{m}^2/s$$

using the relationship between diffusion variance and diffusion coefficient, $\sigma^2 = 2D\tau$ and scaling to μm. Here, τ=0.040 s is the time interval between observations, and each pixel corresponds to a square with side M=180μm.

The asymptotic normality result from Section 5.3 can be used to give an approximate 95%-confidence interval of D:

$$D = \hat{D} \pm 1.96 \cdot \frac{M^2}{2\tau}\sqrt{.0358} = .893 \pm .150\,\mu\text{m}^2/s \tag{25}$$

The theoretical diffusion coefficient is given by Stoke-Einstein's relation (see for example Evans and Wennerström (1999) pages 370-372)

$$D = \frac{k_B T}{6\pi\eta R_H}$$

where k_B is Bolzmann's constant, η the viscosity of the solution, T the temperature and R_H the hydrological radius of the particle.

The appropriate values for the viscosity and temperature are η=0.9 mPa and T=298 K. The geometric radius of the particles are 247 nm and we used this as the hydrological radius, even if the latter is often a bit larger than the former. Plugging this into (6.3) gives us

$$D = 0.982\,\mu\text{m}^2/s$$

Comparing with the confidence interval in (25), we see that the theoretical diffusion coefficient is within this interval.

6.4 Simulation of the approximate distribution of the estimates

We simulated 1000 time series with 26 particles, of which 3 were fixed, over 12 frames in two dimension, using the estimated values of $\sigma^2 = 2.2058$ and $\sigma_e^2 = 0.3172$ from (21) as the true diffusion variance and error variance. For each time series, we used the EM algorithm to estimate σ^2 and σ_e^2.

The histograms of the estimated values are displayed in Figure 2. The mean and empirical covariance of the 1000 estimates of σ^2 and σ_e^2 were

$$\bar{\hat{\sigma}^2} = 2.2054$$
$$\bar{\hat{\sigma}_e^2} = 0.3185$$

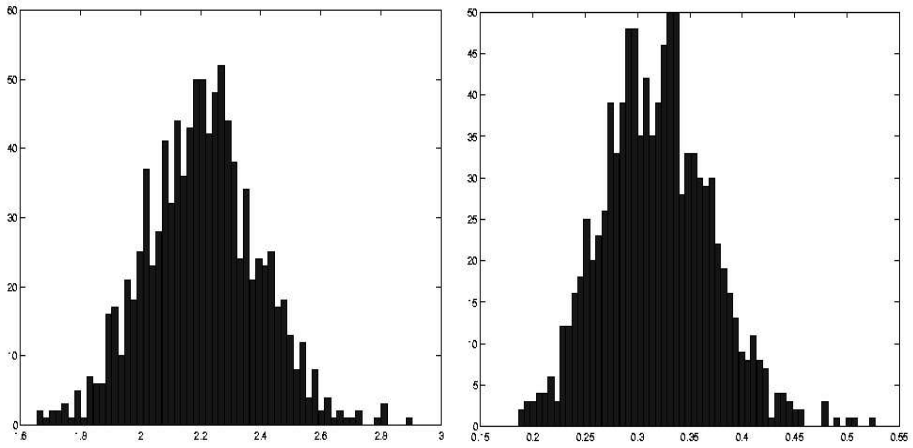

Figure 2: The histograms of the estimated $\hat{\sigma}^2$ and $\hat{\sigma}_e^2$ using the EM algorithm from 1000 simulations using 2.2058 and 0.3172 as true values.

and

$$\begin{bmatrix} .0348 & -.0040 \\ -.0040 & .0027 \end{bmatrix}$$

in good agreement with the true values of $\sigma^2 = 2.2058$ and $\sigma_e^2 = 0.3172$ and the inverse of the observed information matrix in (23).

Acknowledgements

This work was financially supported by the Swedish Research Council for Engineering Sciences (TFR) and the Swedish Research Council (VR). The author wishes to thank Mats Rudemo for his supervision and numerous invaluable comments, Ib Skovgaard for interesting and helpful discussions regarding the asymptotics, Lennart Lindfors of AstraZeneca R&D for producing the images and keeping me in touch with the application, and Olle Häggström for reading through the draft.

References

S.-I. Amari, O. Barndorff-Nielsen, R. Kass, S. Lauritzen, and C. Rao. *Differential geometry in Statistical Inference.* Lecture Notes - Monograph series. Institute of Mathematical Statistics, first edition, 1987.

O. Barndorff-Nielsen and D. Cox. *Inference and Asymptotics.* Chapman & Hall, London, first edition, 1994.

J. Behboodian. On a mixture of normal distributions. *Biometrika*, 57:215–217, 1970.

N. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56:463–474, 1969.

A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, 39:1–38, 1977.

B. Efron. Defining the curvature of a statistical problem (with applications to second order efficiency)(with discussion). *The Annals of Statistics*, 3:1189–1242, 1975.

B. Efron. The geometry of exponential families. *The Annals of Statistics*, 6: 362–376, 1978.

D. Evans and H. Wennerström. *The colloidal domain. Where Physics, Chemistry, Biology, and Technology meet.* Wiley-VCH, New York, second edition, 1999.

V. Hasselblad. Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64:1459–1471, 1969.

R. Hathaway. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13:795–800, 1985.

J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, 27:887–906, 1956.

J. Lindsey. *Parametric Statistical Inference.* Oxford University Press, Oxford, first edition, 1996.

T. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological.*, 44: 226–233, 1982.

P. McCullagh. *Tensor Methods in Statistics.* Chapman & Hall, London, 1987.

G. McLachlan and T. Krishnan. *The EM algorithm and Extensions.* Wiley-Interscience, New York, first edition, 1997.

G. McLachlan and D. Peel. *Finite mixture models.* Wiley-Interscience, New York, first edition, 2000.

R. Redner. Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *The Annals of Statistics*, 9:225–228, 1981.

R. Sundberg. Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, 1:49–58, 1974.

R. Sundberg. An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Communications in Statistics. Part B. Simulation and Computation*, 5:55–64, 1976.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, U.K., first edition, 1999.

A. Wald. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 20:595–601, 1949.

B. Wei. *Exponential Family Nonlinear Models*. Springer, Singapore, first edition, 1998.

Mats Kvarnström, Department of Mathematical Statistics, Chalmers University of Technology and Göteborg University, Göteborg, SE-412 96, Sweden
Email: matskv@math.chalmers.se

Appendix A: Sufficient statistics

Consider the complete data density (5). Take the logarithm and use the transformed increment vectors $\tilde{Y}$ (see section 2.2) for easier notation

$$\begin{aligned}
\log g_c &= z\log p - \frac{z}{2}\sum_{k=1}^{N}\log(\sigma^2+\lambda_k\sigma_e^2) - \frac{z}{2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{\sigma^2+\lambda_k\sigma_e^2} + \\
&+ (1-z)\log(1-p) - \frac{1-z}{2}\sum_{k=1}^{N}\log(\sigma_e^2\lambda_k) - \frac{1-z}{2}\sum_{k=1}^{N}\frac{\tilde{y}_k^2}{\sigma_e^2\lambda_k} \\
&= \sum_{k=1}^{N} z\tilde{y}_k^2\left(-\frac{1}{2}\frac{1}{\sigma^2+\lambda_k\sigma_e^2}\right) - \frac{1}{2\sigma_e^2}\sum_{k=1}^{N}\frac{(1-z)\tilde{y}_k^2}{\lambda_k} \\
&+ z\left(\log(\frac{p}{1-p}) - \frac{1}{2}\sum_{k=1}^{N}\log(\frac{\sigma^2+\lambda_k\sigma_e^2}{\lambda_k\sigma_e^2})\right) - \left(\frac{1}{2}\sum_{k=1}^{N}\log(\lambda_k\sigma_e^2) - \log(1-p)\right)
\end{aligned}$$

and we see that a minimal sufficient statistic can be chosen to be

$$\begin{aligned} t_1 &= z\tilde{y}_1^2 \\ &\vdots \\ t_N &= z\tilde{y}_N^2 \\ t_{N+1} &= \sum_{k=1}^{N} \frac{(1-z)\tilde{y}_k^2}{\lambda_k} \\ t_{N+2} &= z \end{aligned}$$

with the corresponding canonical parameter α

$$\begin{aligned} \alpha_1 &= -\frac{1}{2}\frac{1}{\sigma^2 + \lambda_1\sigma_e^2} \\ &\vdots \\ \alpha_N &= -\frac{1}{2}\frac{1}{\sigma^2 + \lambda_N\sigma_e^2} \\ \alpha_{N+1} &= -\frac{1}{2\sigma_e^2} \\ \alpha_{N+2} &= \log(\frac{p}{1-p}) - \frac{1}{2}\sum_{k=1}^{N}\log(\frac{\sigma^2 + \lambda_k\sigma_e^2}{\lambda_k\sigma_e^2}) \end{aligned}$$

which is a function of our parameter θ. Since this is 3-dimensional and the sufficient statistics is $(N+2)$-dimensional, we say that the complete data belongs to a curved exponential family or, with the terminology of Barndorff-Nielsen and Cox (1994), a $(N+2,3)$-exponential model.

Solving for p in the expression for α_{N+2} above, we get

$$p = \frac{e^{\alpha_{N+2}}\prod_{k=1}^{N}\left(\frac{\sigma^2+\lambda_k\sigma_e^2}{\lambda_k\sigma_e^2}\right)^{1/2}}{1 + e^{\alpha_{N+2}}\prod_{k=1}^{N}\left(\frac{\sigma^2+\lambda_k\sigma_e^2}{\lambda_k\sigma_e^2}\right)^{1/2}} = \frac{e^{\alpha_{N+2}}\prod_{k=1}^{N}\left(\frac{\alpha_{N+1}}{\lambda_k\alpha_k}\right)^{1/2}}{1 + e^{\alpha_{N+2}}\prod_{k=1}^{N}\left(\frac{\alpha_{N+1}}{\lambda_k\alpha_k}\right)^{1/2}}$$

and we can write the complete data density as

$$\log g_c = \alpha^T t - k(\alpha) \tag{26}$$

where $\alpha = \alpha(\theta)$ and k is

$$k(\alpha) = \frac{1}{2}\log(N+1) - \frac{N}{2}\log(-2\alpha_{N+1}) + \log\Big(1 + e^{\alpha_{N+2}}\prod_{k=1}^{N}\Big(\frac{\alpha_{N+1}}{\lambda_k\alpha_k}\Big)^{1/2}\Big) \tag{27}$$

From standard theory of exponential families, we get the cumulants of the sufficient statistics by differentiating $k(\alpha)$. In particular, we have $\mathbf{E}T = \frac{\partial k}{\partial \alpha}$ and $\mathbf{Var}\{T\} = \frac{\partial^2 k}{\partial \alpha \partial \alpha^T}$, which we denote μ and V, respectively.

The expectation of the sufficient statistics

$$\mathbf{E}T = \begin{bmatrix} p(\sigma^2 + \lambda_1 \sigma_e^2) \\ \vdots \\ p(\sigma^2 + \lambda_N \sigma_e^2) \\ (1-p)N\sigma_e^2 \\ p \end{bmatrix}$$

Appendix B: Geometry of the complete data

Differentiating the complete data likelihood once, we get the score function

$$\frac{\partial \log g_c}{\partial \theta} = (\frac{\partial \alpha}{\partial \theta^T})^T T - \frac{\partial k}{\partial \theta} = (\frac{\partial \alpha}{\partial \theta^T})^T (T - \mathbf{E}T)$$

where we used

$$\frac{\partial k}{\partial \theta} = (\frac{\partial \alpha}{\partial \theta^T})^T \frac{\partial k}{\partial \alpha}$$

and $\partial k / \partial \alpha = \mathbf{E}T$.

Hence at the MLE $\hat{\theta}$, the difference vector $T - \mu$ is orthogonal to the derivative of the canonical parameter α with respect to the parameter θ. This relation and other geometrical interpretations of the maximum likelihood estimate in a curved exponential family, were, to the author's knowledge, first made by Efron in two groundbreaking articles, Efron (1975, 1978). Since then a lot of research has been made in this area with fruitful connections between statistics and differential geometry, see for example Amari et al. (1987), McCullagh (1987) and Barndorff-Nielsen and Cox (1994)

Appendix C: Iterative scheme for complete data MLE

Exploring the geometrical relations of the curved exponential further, brings us to a Newton-Raphson style of iterative method of finding the maximum likelihood estimate of θ, given the complete data. Even if the unobserved classification variables Z_i are not available to us, the M-step in the EM algorithm (see section 4.1.3) maximizes the complete data likelihood using the estimated $\hat{Z}_i$:s from the E-step. The idea comes from Wei (1998).

We adopt the notation of Wei:

$$\begin{aligned}\mu(\theta) &= \mathbf{E}_\theta T \\ V(\theta) &= \mathbf{Var}_\theta\{T\} \\ D_\alpha &= \frac{\partial \alpha}{\partial \theta^T} \\ D &= \frac{\partial \mu}{\partial \theta^T}\end{aligned}$$

and derive the following identity

$$D = \frac{\partial \mu}{\partial \theta^T} = \frac{\partial \mu}{\partial \alpha^T}\frac{\partial \alpha}{\partial \theta^T} = V D_\alpha$$

Write $l_c = \log L_c$ and let $\dot{l}_c$ and $\ddot{l}_c$ denote the first and second derivative of l_c with respect to θ.

Expressing the score function $\dot{l}_c$ as

$$\begin{aligned}\dot{l}_c = \frac{\partial l_c}{\partial \theta} &= \Big(\frac{\partial \alpha}{\partial \theta}\Big)^T \frac{\partial l_c}{\partial \alpha} = \Big(\frac{\partial \alpha}{\partial \theta}\Big)^T (T-\mu) \\ &= D_\alpha (T-\mu) = D^T V^{-1}(T-\mu)\end{aligned} \tag{28}$$

and the score equation at $\hat{\theta}$ can be written

$$0 = D^T V^{-1}(T-\mu)$$

where D, V, and μ are evaluated at $\hat{\theta}$.

Differentiate $\dot{l}_c$ once again and we get

$$\begin{aligned}\ddot{l}_c &= \Big(\frac{\partial l_c}{\partial \alpha}\Big)^T \Big[\frac{\partial^2 \alpha}{\partial \theta \partial \theta^T}\Big] - D_\alpha^T V D_\alpha \\ &= (T-\mu)^T \Big[\frac{\partial^2 \alpha}{\partial \theta \partial \theta^T}\Big] - D_\alpha^T V D_\alpha\end{aligned} \tag{29}$$

Now, Newton's classical iterative scheme can be written

$$\theta_{i+1} = \theta_i + [-\ddot{l}_c(\theta_i)]^{-1}\dot{l}_c(\theta_i)$$

Using the expressions in (28) for $\dot{l}_c$ together with $\mathbf{E}_\theta\{-\ddot{l}_c\} = D_\alpha^T V D_\alpha$ instead of $\ddot{l}_c$ we get

$$\begin{aligned}\theta_{i+1} &= \theta_i + \big[D_\alpha^T V D_\alpha\big]^{-1} D_\alpha (T-\mu) \\ &= \theta_i + \big[D^T V^{-1} D\big]^{-1} V^{-1} D (T-\mu)\end{aligned} \tag{30}$$

We find V by differentiating k twice with respect to α. The matrix turns out to be quite complicated with all its elements different from zero. In the iteration scheme (30), it is the inverse V^{-1} we need, and since this is in fact much less complicated, we present it here

$$V^{-1} = \begin{bmatrix} \frac{1}{2p(\sigma^2+\lambda_1\sigma_e^2)^2} & \cdots & 0 & 0 & -\frac{1}{2p(\sigma^2+\lambda_1\sigma_e^2)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{2p(\sigma^2+\lambda_N\sigma_e^2)^2} & 0 & -\frac{1}{2p(\sigma^2+\lambda_N\sigma_e^2)} \\ 0 & \cdots & 0 & \frac{1}{2(1-p)N(\sigma_e^2)^2} & \frac{1}{2(1-p)\sigma_e^2} \\ -\frac{1}{2p(\sigma^2+\lambda_1\sigma_e^2)} & \cdots & -\frac{1}{2p(\sigma^2+\lambda_N\sigma_e^2)} & \frac{1}{2(1-p)\sigma_e^2} & \frac{2+N}{2p(1-p)} \end{bmatrix}$$

Appendix D: Complete data asymptotics

In applications it may happen that you label the particles manually as diffusing or fixed or defect particles and want to estimate the parameters. Then our problem becomes easier, mainly because the likelihood is composed of a product.

From (29), we get the expected information matrix to the complete data

$$\begin{aligned} \mathcal{I}_c(\theta) &= D^T V^{-1} D = D^T D_\alpha \\ &= \begin{bmatrix} \frac{p}{2}\sum_{k=1}^N \frac{1}{(\sigma^2+\lambda_k\sigma_e^2)^2} & \frac{p}{2}\sum_{k=1}^N \frac{\lambda_k}{(\sigma^2+\lambda_k\sigma_e^2)^2} & 0 \\ \frac{p}{2}\sum_{k=1}^N \frac{\lambda_k}{(\sigma^2+\lambda_k\sigma_e^2)^2} & \frac{p}{2}\sum_{k=1}^N \frac{\lambda_k^2}{(\sigma^2+\lambda_k\sigma_e^2)^2} + \frac{N(1-p)}{2(\sigma_e^2)^2} & 0 \\ 0 & 0 & \frac{1}{p(1-p)} \end{bmatrix} \end{aligned}$$

which is positive definite for all $\theta \in \text{int}(\Omega)$. To see this, apply the Cauchy-Schwarz inequality on the upperleft 2 by 2 matrix elements.

For $\theta_0 \in \text{int}(\Omega_c)$ we get strong consistency from Wald (1949), and since also $\mathcal{I}_c(\theta_0)$ is positive definite, all conditions for asymptotic normality are satisfied.