Parameter estimation in a nonstationary Markov model for copolymer propagation

by

Longhow Lam

Supervisors:

Dr. M.C.M. de Gunst
(Vrije Universiteit Amsterdam, Department of Mathematics & Computer Science)

Dr.ir. H.J.A.F. Tulleken
(Koninklijke/Shell-Laboratorium, Amsterdam, Department MCA)

Dr. M.C.J. van Pul
(Koninklijke/Shell-Laboratorium, Amsterdam, Department MCA)

July, 1995

Preface

This thesis is the final result of my graduation project in Mathematics at the Vrije Universiteit Amsterdam, performed at the Koninklijke/Shell-Laboratorium, Amsterdam (KSLA) from January 1995 until July 1995. I worked at the department Measurement and Computational Applications (MCA) under supervision of dr.ir. H.J.A.F. Tulleken and dr. M.C.J. van Pul.

I would like to thank Shell for offering me the opportunity to gain the experience of working on a math project. Furthermore, Herbert Tulleken and Mark van Pul are acknowledged for their great support during the project, their criticism, their advise on the report and feedback on both mathematical and non-mathematical issues. Problem owner M.A. van Dijk (KSLA) is acknowledged for his feedback on chemical issues and for making data available.

I would like to thank my university supervisor Dr. M.C.M. de Gunst for her outstanding advise and support. Last but not least, I would like to thank my roommates for their practical hints and all the other stagiaires for those nice coffee breaks, lunches and evenings.

Amsterdam, July, 1995 Longhow Lam

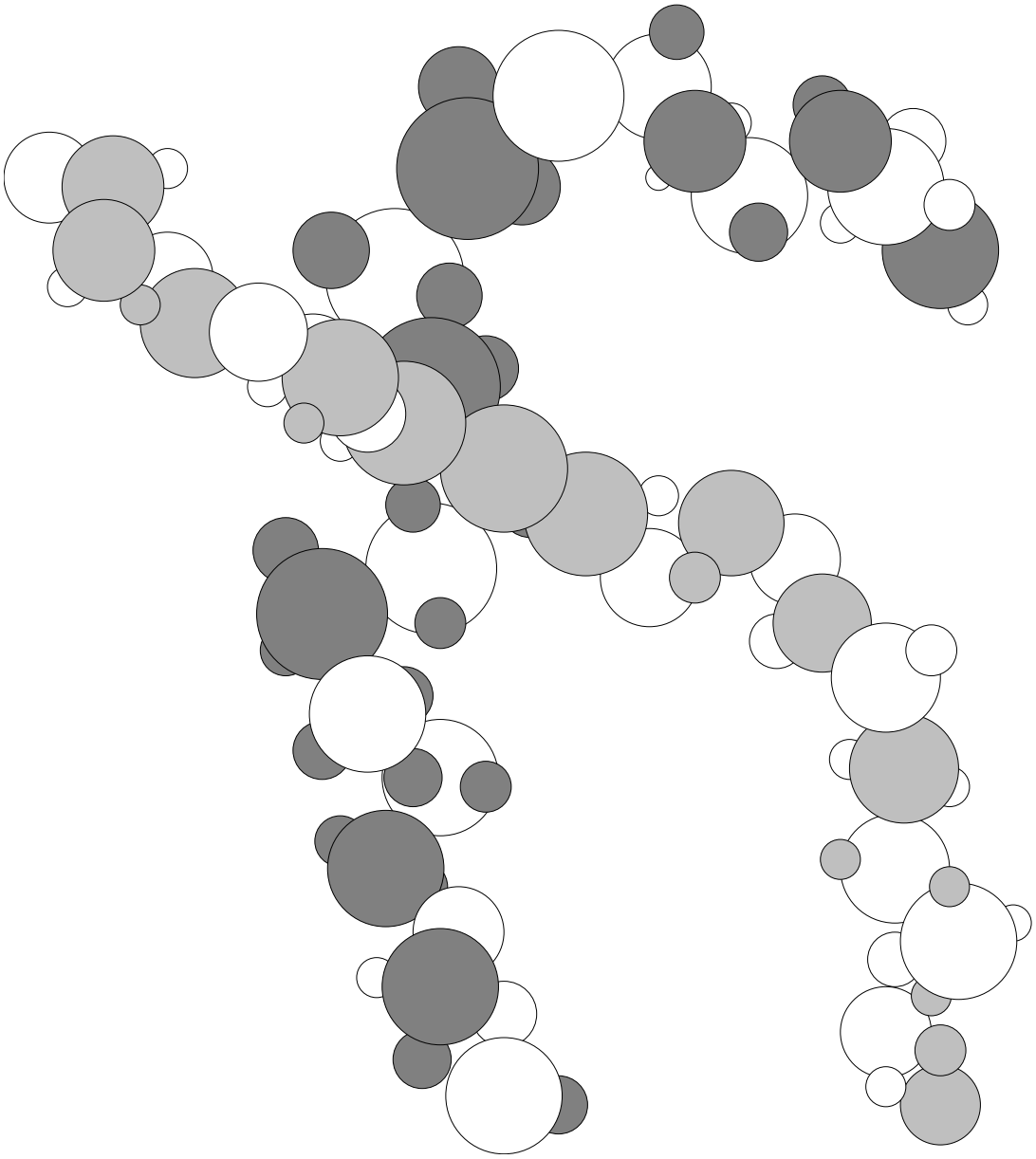

Figure 1: An artistic illustration of two polymer chains

Contents

Chapter 1

Introduction

1.1 What is a polymer?

In daily life we can hardly imagine a world without polymers anymore, just think of the plastics and rubbers that are being used in the electronic and automative industries and household/consumer products. Polymers are very large molecules which are made by repeated chemical reactions between the growing polymer and monomer molecules. Important for the physical (material) properties of polymers are the number, the types and positions of monomers build in a polymer chain. Therefore several polymer types can be distinguished. Firstly there are the *homopolymers*. These polymers consist of only one type of monomer. The formation of polymers with two or more types of monomers results in *copolymers*. The relative position of the different monomers will strongly influence physical properties. For example, in the case two types of monomers (say A and B) the place of the monomers could be structured like:

Diblock copolymer	$AAAAAAAAAAAABBBBBBBBBBBBB$
Alternating copolymer	$ABABABABABABABABABABABABA$
Random copolymer	$BABAAABBBAABABBABABABBABABB$

Each structure has a specific effect on the physical properties. These representations are highly idealized, in practice one always finds a distribution of structures.

1.2 Problem description

In this thesis we will discuss a mathematical model for the description of polymer propagation, i.e. the formation of polymer chains from subsequent attachment of monomers. In particular we are interested in a tri-block polymer, that is build-up of two types of monomers, A and B in the following structured way

$$AAA\cdots AAABBB\cdots BBBAAA\cdots AAA\ .$$

In practice such tri-blocks are produced in three phases in a chemical batch process. First an amount of A monomers is put in a reaction vessel which does

not contain any B monomers. After a while the first blocks $AAAA\cdots AAAA$ are formed. Then B monomers are added to these blocks, which results in diblocks $AAA\cdots AAABBB\cdots BBB$. Finally A monomers are added again, to form complete tri-blocks, such as
$AAAAAABBBBBBBBAAAAAAA$. However, for various reasons (related to product properties and quality) A-type monomers are added in the third phase, while there are still reacting B monomers present. So in the reaction vessel there arises a sort of competition between A and B monomers to react with the 'living' chain. A random occurrence of A and B in the chain will result. This phenomenon is also known as *tapering*. This tapering influences product properties. In case it is more likely that B monomers will react to the 'living' chain, rather than A monomers the B monomers will run out first. So eventually there will be an A block at the end of the chain. For example a complete chain could look like:

$$\underbrace{AAAAA\cdots AAA}_{\text{phase 1}}\underbrace{BBBBB\cdots BBB}_{\text{phase 2}}\underbrace{BBAB\cdots BAABBAB}_{\text{tapering}}AAA\cdots AAAAA \quad \underbrace{\qquad\qquad\qquad\qquad}_{\text{phase 3}}$$

Our model will focus on this special case and study the consequences of tapering, as a result of the events in phase 3. In particular, we shall deal with the description of the statistics of the structure of patterns along the tapering part of the chains.

Not every 'living' chain takes part in the reaction. A complication that may arise is that chains that end with an A monomer and B monomer respectively, can (reversely) aggregate. These aggregated chains are excluded from the reaction: single A and B monomers can not attach to them. This phenomenon will also be modelled to find out to which extent this occurs.

1.3 Organization of the thesis

The structure of this thesis is as follows. After this introduction we shall discuss, in chapter 2, some basic chemical aspects that are relevant in this project. In chapter 3 we shall derive the mathematical model for the polymer propagation. In chapter 4 we give some theory about parameter estimation. Then chapter 5 consists of numerical methods that are used and chapter 6 consists of tuning the model by comparing real data with model predictions. Finally, a summary and conclusions are presented.

Chapter 2

Chemical aspects

In this chapter we shall give some background on the chemical aspects used in this thesis. Precise chemical details are beyond the reach of this thesis; the interested reader is referred to standard chemical textbooks.

2.1 A possible classification of polymer types

A way to distinguish polymers is to characterize them by the mobility of their polymer chains. We have the following types:
Thermoplastic: material that softens (plasticizes) with increasing temperature.
Thermoset: material that tends to harden with increasing temperature because more and more crosslinks are formed. Another possibility to give polymeric material mechanical properties, is to link (the ends of) the polymer chains. These links are called 'crosslinks' and can be created in several ways
Elastomer: material which can be stretched to more than twice their length and then retract to less then 50% permanent deformation.
The structures of these three types are schematically illustrated in Fig. 2.1.

2.2 MWD

An important quantity by which a polymer is characterized and which determines the physical properties is the molecular weight. Molecular weight (MW) is the mass of one mole particles, so dim(MW)=kg/mol. One mole consists of about $6.0 * 10^{23}$ particles, the number of Avogadro. Since every sample of polymers consists of a variety polymer molecules with different length, one often speaks of the average molecular weight (in that sample) and about the MW distribution. A complete characteristic of a sample requires the knowledge of the molecular weight distribution (MWD), thus the number n_i of molecules with a MW of m_i for all possible weights $m_i,\ i = 1, 2, \ldots$.

To illustrate the relatively large size of a typical polymer chain, we take a chemical-industry polypropene grade. The average molecular weight is 400 kg/mol, whereas oxygen, for example, only weighs 16 g/mol. If one would

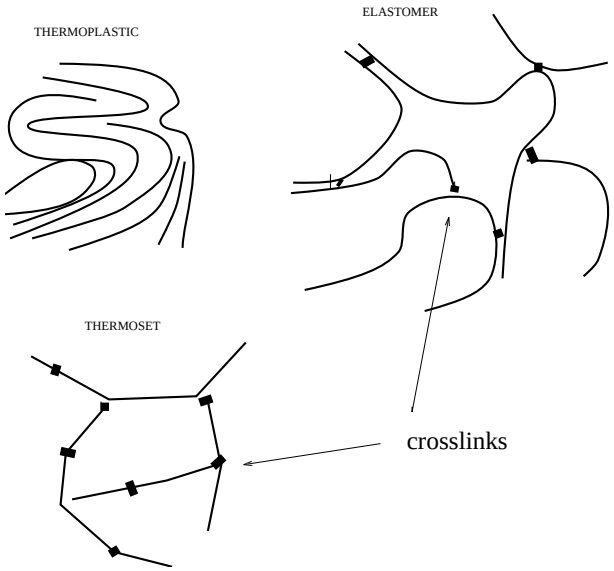

Figure 2.1: Different types

consider a scale model of such a polymer as a tube with the diameter of a spaghetti string, it would be about 30 meters long.

2.3 Manufacture of Polymers

The manufacture of polymers involves a chemical reaction of one or more reactive molecules. Most polymerization reactions consist of three stages:

- Initiation (the start of reactive chains is formed)
- Propagation (monomers are added to the reactive chain ends),
- Termination (reactive chain ends are killed).

For completeness, we shall briefly mention some types of polymerization:

• Anionic polymerization
The reaction is started by addition of an organometallic material (e.g. Butyl-Lithium, $C_4H_9^-Li^+$). The initiator attacks monomers and forms a new anion, then other monomers are subsequently inserted. The termination is carried out by adding a proton donor, such as water, to kill the very stable anions. This is an important propagation mechanism for the practical reactions. In fact, the class of polymers studied in this report (i.e. triblock) are being build in this way.

• Free-radical polymerization
Decomposition of peroxide generates free radicals, which start the reaction by attacking the double bonds of the monomers. Successively other double bonds

are attacked and monomers are added to the growing chain.

• **Ziegler-Natta polymerization**
Stereospecific structures can be produced by this kind of polymerization.

• **Poly-addition/condensation**
These reactions do not consist of the three stages as the previous reaction types. Monomers just simply react with each other to from a (long) chain.

2.4 Reaction rates and equilibrium constants

The *reaction speed*, s, is the total amount of matter transformed or formed per second. This speed depends on the concentration and type of the matter as well as temperature. We can measure this speed by observing the change of concentration during the reaction, for example, the change in gasvolume or change in color spectrum. For a second-order reaction

$$A + B \to C \ ,$$

the reaction speed depends on two concentrations $[A]$ and $[B]$ and is given by

$$s = \frac{dC}{dt} = k * [A] * [B] \ .$$

The *reaction rate* k expresses information about the average fraction of molecules A and of molecules B that reacts into the product equals $k\Delta t$ in a small time interval $[t, t + \Delta t)$.

In a *chemical (dynamic) equilibrium* the reactions still go on but no change in concentration can be observed directly. This typically happens in the case of reversible reactions, e.g. $A \rightleftharpoons B$. In a short time interval molecules B are being formed, but on the average as many other molecules B disintegrate into A's again. The equilibrium is described by the so-called *equilibrium constant* K, which is a fraction of concentrations. For example, for the equilibrium reaction $nA \rightleftharpoons mB$ we have

$$K = \frac{[B]^m}{[A]^n}.$$

Hence large values of K indicate that the position of the equilibrium is more to the right-hand side, and that on the average more B molecules than A molecules will be present.

2.5 An introduction to the analysis of polymer structures

2.5.1 Ozonolysis

Ozonolysis is a method of analysis which separates certain blocks of molecules through the impact of ozone reagents. For example, with this method the B

blocks can be degraded by ozone while the A blocks remain intact. Thus if we e.g. have the molecule:

$$BBBBBAAAABBBAAABBBAAAAAA$$

then after ozonolysis the B's are cut out, and the three blocks $AAAA$, AAA, and $AAAAAA$ remain. Further investigation of the remaining patterns can, for instance, be done by GPC, which method of analysis is discussed below.

2.5.2 Gel Permeation Chromatography (GPC) [4]

GPC is used to determine molecular weights. It is a method to separate soluble molecules by size; small ones with a MW less then 100 as well as large ones with a MW of several millions. The separation is usually carried out in columns that are tightly packed within a gel or some other porous material and completely filled with solvent. The same solvent is used to dissolve the sample before introducing it into the column. Small sample molecules can diffuse into the pores of the gel; large ones are excluded, others of intermediate size can penetrate into some of the larger pores. The molecules are constantly diffusing back and forth between the pores and the interstices. Solvent pumped through the column flows only in the interstices, sweeping along all sample molecules being present there. The molecules in the pores stay behind for a longer time until they are released. The large molecules which are always or mostly excluded from the pores are therefore eluded first, the small ones which are mostly inside the pores come out last. The instrumentation for GPC is relatively simple in concept, but somewhat involved in execution. A passage through the instruments contain the following functions: (a) sample injection, (b) pumping, (c) chromatographic column, (d) sample detection, (e) eluent volume detection, (f) data recording. Like any other type of chromatography, the GPC chromatogram of a monomeric compound appears as a curve of finite width. The position of the peak of the curve, that is the mode of the distribution, depends on the molecular weight of the compound; the area under the curve is proportional to the amount of the compound in the total sample; and the width of the curve depends on various band spreading mechanisms in the GPC instrument, both within and outside of the columns. For a polydispersed (i.e. a sample with molecules with multiple MW's) sample, such as those normally encountered in high polymers, the chromatogram is a composite of curves of all its components. The total area under the curve is still proportional to the amount of the entire sample but the height of the curve does not reflect the relative abundance of the components at the corresponding elusion volumes, as it depends also on the abundance of the neighbouring components. At the ends of the chromatogram there are curve portions representing components which may not even exist. See Fig 2.2 for an example curve.

2.5.3 Nuclear Magnetic Resonance (NMR) [3]

NMR spectrography is the study of the magnetic properties of nuclei. It is a useful technique to study the structural characteristics of polymers. Certain nuclei have angular momentum and thus as a spinning electric charge, a magnetic

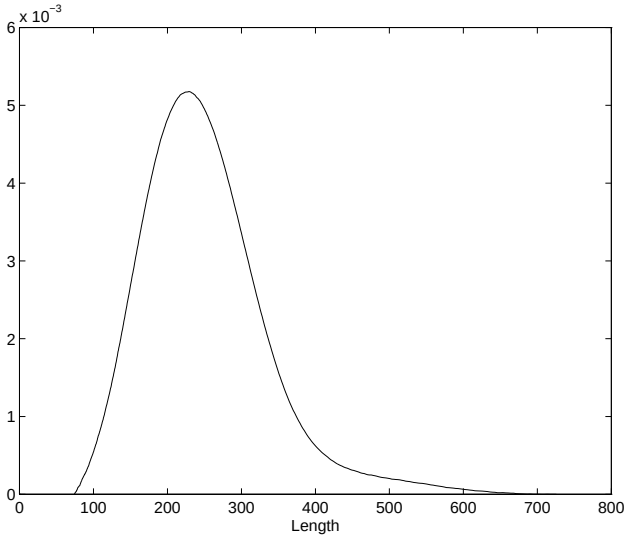

Figure 2.2: Example of GPC curve

moment. These nuclear magnetic moments were detected in the twenties with molecular beam experiments. A beam of molecules with nuclei containing angular momentum is first passed through a strong static homogeneous magnetic field, and then subjected to an oscillating magnetic field. Consequently the nuclei can undergo resonance, resulting in absorption and subsequent emission of energy. The analysis of polymer structures using NMR techniques is based on the fact that nuclei of the same species will resonate at slightly different frequencies depending on their chemical environment. These separations of frequencies are called chemical shifts; the magnitudes of the shifts are given in terms of parts per million changes in field strength, specified by the NMR-spectrum, see Fig. 2.3 for an example spectrum.

The NMR-spectrum is a result of accumulation of responses from very many individual polymer molecules, thus we see and interpret results from a final summation that represents directly "the average polymer chain". Each peak in the spectrum corresponds to a certain structure of monomers in the chain and the area under a peak is proportional to the fraction of the corresponding structure of monomers. For example if we measure triad structures, which are sequences of three monomers, then the maximum number of peaks that can be observed in a NMR-spectrum, equals the number of different structures.

2.5.4 The error structure of the NMR data [7]

As explained above the NMR spectrum shows a continuous graph that consist of several peaks. Each area under a peak, corresponding to a certain structure, is proportional to the mol fraction of that structure. However it is not unusual that peaks overlap, so in the spectrum cutting points have to be chosen to determine the area under a peak. Therefore an error will be introduced. If we

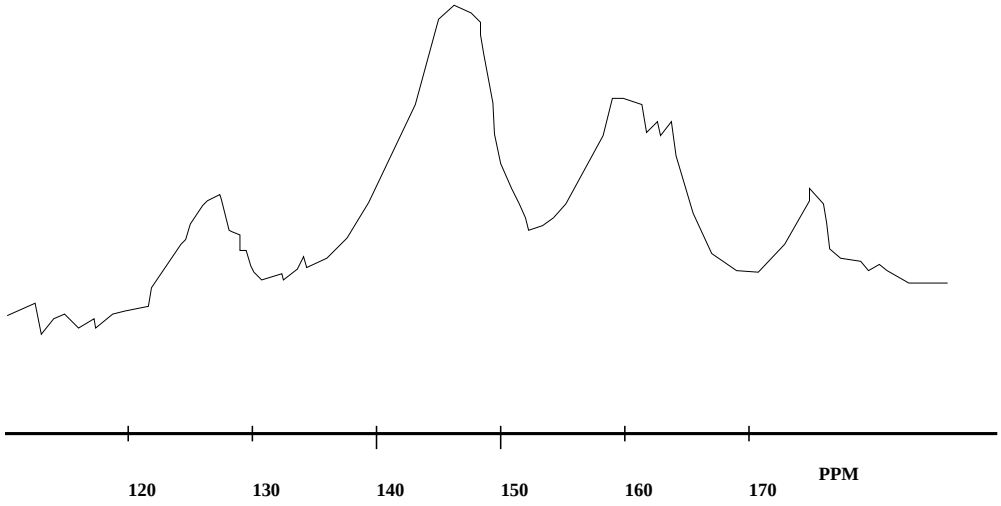

Figure 2.3: Example of NMR spectrum

look at the NMR spectrum in Fig 2.4 we see that for a specific peak the error in the area under the peak is only influenced by the choice of the two cutting points on both sides of that peak. This is why we assume that the error e_i, made in the determination of the area under the i-th peak, satisfies

$$e_i = \alpha_i g(c_i) - \alpha_{i-1} g(c_{i-1}) \ ,$$

here the α_i are independent and identically distributed (i.i.d.) $N(0, \rho)$, c_i and c_{i-1} are the right- and left hand cutting points, respectively, and $g(c_i)$'s are the heights of the cutting points in the spectrum. Thus, if $e = (e_1, \ldots, e_n)$ denotes the vector of area errors, the dependency of the errors can be obtained from the covariance matrix, $\Sigma_{\text{NMR}} = \mathbb{E}(ee')$ which has the form

$$\rho \begin{pmatrix} g(c_0)^2 + g(c_1)^2 & -g(c_1)^2 & & & \\ -g(c_1)^2 & g(c_1)^2 + g(c_2)^2 & -g(c_2)^2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & -g(c_{n-1})^2 & \\ & & -g(c_{n-1})^2 & g(c_{n-1})^2 + g(c_n)^2 & \end{pmatrix} \quad (2.1)$$

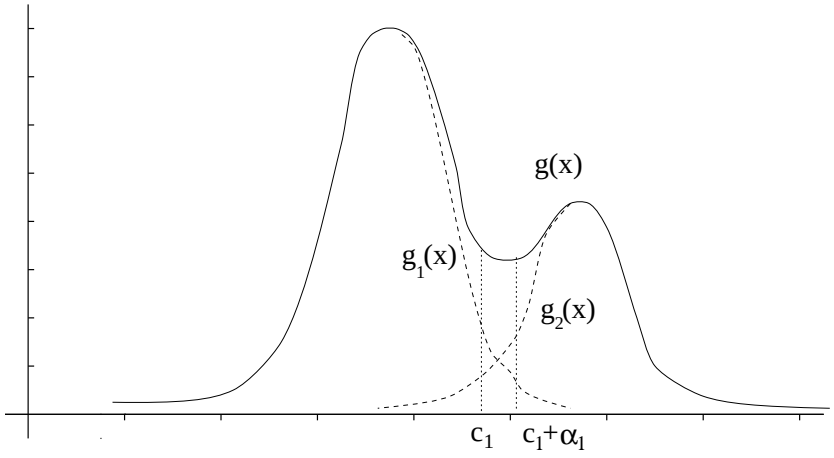

Figure 2.4: Overlapping peaks and the choice of their cutting point.

Chapter 3

Mathematical model building

3.1 Markov model

The propagation of a (living) chain, can be described with Markovian stochastics. The Markov state could be the kind of reactive head, i.e. the chain end. For a copolymerization this would be either A or B. So everytime when a monomer is added to the chain a transition takes place. There are four different transitions possible:

$$\begin{aligned} \text{—}A + A &\xrightarrow{k_{aa}} \text{—}AA \\ \text{—}A + B &\xrightarrow{k_{ab}} \text{—}AB \\ \text{—}B + A &\xrightarrow{k_{ba}} \text{—}BA \\ \text{—}B + B &\xrightarrow{k_{bb}} \text{—}BB \end{aligned}$$

where the k's are the reaction rates of the different kind of reactions. We can illustrate the different transitions between the states in a Markov diagram, Fig. 3.1.

In this figure $p(j|--i)$ represents the probability that monomer of type j is added to a chain ending with monomer i. We assume exponentially distributed sojourn times [1] in the two states. So if the current state is A, i.e. the chain is ending with an A monomer, it has to 'wait' an exponentially distributed time T_{aA} (with rate $k_{aa}a(t)$), to react with an A monomer and an exponentially distributed time T_{bA} (with rate $k_{ab}b(t)$), to react with a B monomer, where $a(t)$ and $b(t)$ are the concentrations of A and B monomers at time t. Therefore the probability that an A monomer is added to the A chain equals $\mathbb{P}(T_{aA} < T_{bA})$ and the probability of adding a B monomer equals $\mathbb{P}(T_{aA} > T_{bA})$. These probabilities can be calculated using the following well-known result:

$$\begin{aligned} \mathbb{P}(T_{aA} < T_{bA}) &= \int_0^\infty \mathbb{P}(T_{aA} < t | T_{bA} > t)\ d\mathbb{P}(T_{bA} > t) \\ &= \int_0^\infty (1 - e^{-\mu_{aA}t})\ de^{-\mu_{bA}t} \end{aligned}$$

[1] A (sojourn) time T is exponentially distributed with rate μ if $\mathbb{P}(T < t) = 1 - e^{-\mu t}$

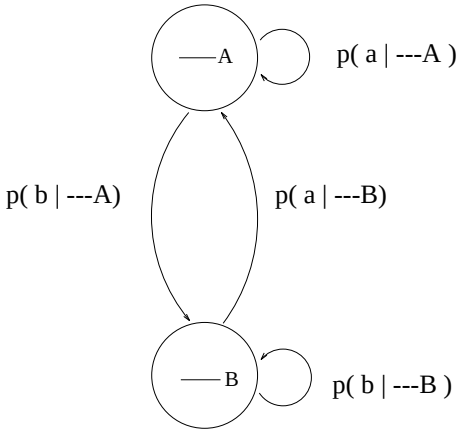

Figure 3.1: Transition diagram

$$
\begin{aligned}
&= \mu_{bA} \int_0^\infty \left(e^{-\mu_{bA} t} - e^{-(\mu_{bA}+\mu_{aA})} \right) \, dt \\
&= \mu_{bA} \left(\frac{-e^{\mu_{bA} t}}{\mu_{bA}} \Big|_0^\infty + \frac{e^{-(\mu_{aA}+\mu_{bA})t}}{\mu_{bA}+\mu_{aA}} \Big|_0^\infty \right) \\
&= \frac{\mu_{aA}}{\mu_{aA}+\mu_{bA}}
\end{aligned}
$$

if T_{aA} and T_{bA} are independently exponentially distributed with rates μ_{aA} and μ_{bA} respectively. Consequently

$$p(A|A) = \frac{k_{aa} a(t)}{k_{ab} b(t) + k_{aa} a(t)} , \tag{3.1}$$

$$p(B|A) = \frac{k_{ab} b(t)}{k_{ab} b(t) + k_{aa} a(t)} , \tag{3.2}$$

as T_{aA} and T_{bA} are indeed independently exponentially distributed by definition of the Markov chain.
The same idea holds for a B chain, then we have sojourn times T_{aB} and T_{bB}:

$$p(B|B) = \frac{k_{bb} b(t)}{k_{ba} a(t) + k_{bb} b(t)} , \tag{3.3}$$

$$p(A|B) = \frac{k_{ba} a(t)}{k_{ba} a(t) + k_{bb} b(t)} , \tag{3.4}$$

where $a(t)$ and $b(t)$ mean the concentration of monomer A and monomer B.

We could also create a second-order Markov model, i.e. a model in which the transition probabilities depend on the last two monomers of the chain end. See [7] for a full description of such models. We could estimate the transition probabilities in the same way as in [7]. However, if we look at equations (3.1) - (3.4)

then we see that the transition probabilities depend on the concentration of the monomers; this concentration is not kept constant, but varies stochastically. So we are dealing with a *nonstationary* Markov model, as the transition probabilities are depending on the amounts of A and B monomers. These amounts will decrease in time, because more and more monomers are added to the living chain. This is why instead of describing chain ends, we shall model the amounts (concentrations) of A and B monomers, at some moment t. In the next sections we shall derive differential equations that model these concentrations.

Thus we have $a(t)$ as the amount of A monomers and $b(t)$ as the amount of B monomers, both at time t. The Markov model assumption (Fig. 3.1) then implies exponentially distributed sojourn times in a state, so that we can calculate the probability $q^A(t, \Delta t)$ that in a time interval $[t, t + \Delta t)$ no reaction takes place if the chain has an A-head. This means that the minimum of the sojourn times, T_{aA} and T_{bA}, in state A has to be larger than Δt. Thus we have $\mathbb{P}(\min(T_{aA}, T_{bA}) > \Delta t)$, which can be calculated by using some properties of the exponential distribution. We find

$$\begin{aligned} q^A = q^A(t,\Delta t) &= \text{no reaction of a given } A\text{-chain in time period } [t, t+\Delta t] \\ &= \mathbb{P}(\min(T_{aA}, T_{bA}) > \Delta t) = \mathbb{P}(T_{aA} > \Delta t \text{ and } T_{bA} > \Delta t) \\ &= \mathbb{P}(T_{aA} > \Delta t)\mathbb{P}(T_{bA} > \Delta t) = e^{-k_{aa}a(t)\Delta t} \cdot e^{-k_{ab}b(t)\Delta t} \\ &= e^{-[k_{ab}b(t)+k_{aa}a(t)]\Delta t}\,, \end{aligned} \tag{3.5}$$

as T_{aA} and T_{bA} are independent.
Now define $p^{Ab}(t, \Delta t)$ as the probability that in the time interval $[t, t + \Delta t)$ an A chain reacts with a B monomer. Then using (3.2) and (3.5) we obtain

$$p^{Ab} = p^{Ab}(t,\Delta t) = (1-q^A)p(B|A) = \frac{(1-q^A)k_{ab}b(t)}{k_{ab}b(t)+k_{aa}a(t)}\,. \tag{3.6}$$

The probabilities p^{Aa}, q^B, p^{Bb} and p^{Ba}, are defined analogously and are given by

$$p^{Aa} = p^{Aa}(t,\Delta t) = \frac{(1-q^A)k_{aa}a(t)}{k_{ab}b(t)+k_{aa}a(t)}\,, \tag{3.7}$$

$$q^B = q^B(t,\Delta t) = e^{-[k_{bb}b(t)+k_{ba}a(t)]\Delta t}\,, \tag{3.8}$$

$$p^{Bb} = p^{Bb}(t,\Delta t) = \frac{(1-q^B)k_{bb}b(t)}{k_{bb}b(t)+k_{ba}a(t)}\,, \tag{3.9}$$

$$p^{Ba} = p^{Ba}(t,\Delta t) = \frac{(1-q^B)k_{ba}a(t)}{k_{bb}b(t)+k_{ba}a(t)}\,. \tag{3.10}$$

3.2 An averaged propagation model

We are now in a position to derive a model for the average monomer concentrations $a(t), b(t)$ and the active chain concentrations $A(t)$ and $B(t)$ of which

$A^{\text{eff}}(t)$ and $B^{\text{eff}}(t)$ are the effective (i.e. available for immediate propagation) portions of the active chains. We assume that we have 1 mole of active polymer that is able to propagate. According to the principle of conservation of mass, we have for Δt small

$$\begin{aligned} a(t+\Delta t) &= a(t) - B^{\text{eff}}(t)p^{Ba} - A^{\text{eff}}(t)p^{Aa} \ , && (3.11)\\ b(t+\Delta t) &= b(t) - B^{\text{eff}}(t)p^{Bb} - A^{\text{eff}}(t)p^{Ab} \ , && (3.12)\\ B(t+\Delta t) &= B(t) - B^{\text{eff}}(t)p^{Ba} - A^{\text{eff}}(t)p^{Ab} \ , && (3.13)\\ A(t)+B(t) &\equiv 1 \ . && (3.14) \end{aligned}$$

Thus, by rearranging (3.11), (3.12) and (3.13) into difference quotients and letting Δt tend to zero we get the following differential equations

$$\begin{aligned} \dot{a}(t) &= -\left[k_{ba}B^{\text{eff}}(t) + k_{aa}A^{\text{eff}}(t)\right] a(t) \ , && (3.15)\\ \dot{b}(t) &= -\left[k_{bb}B^{\text{eff}}(t) + k_{ab}A^{\text{eff}}(t)\right] b(t) \ , && (3.16)\\ \dot{B}(t) &= -k_{ba}B^{\text{eff}}(t)a(t) + k_{ab}A^{\text{eff}}(t)b(t) \ , && (3.17) \end{aligned}$$

where we have used the definitions (3.5) - (3.10) of the probabilities. Together with the equations (3.18) and (3.19) for $A^{\text{eff}}(t)$ and $B^{\text{eff}}(t)$ which are explained in the next section, we get a complete description of the monomer and active chain ends amounts.

We note that this model describes the concentrations in a deterministic way, i.e. the solutions of (3.15)-(3.17) will always be the same if we recalculate them. However if one could measure the precise concentrations of the monomers during the reaction one would always find random fluctuations around the curves. These fluctuations could be modelled with *stochastic differential equations*, but this is beyond the scope of this thesis. Thus we shall concentrate on the characteristics of the expected population that will obey the above deterministic differential equations.

3.3 The formation of aggregates

Aggregates are clustered chains. They are in equilibrium with the free (reactive) chains; only the latter can propagate. A- and B-aggregates are in equilibrium with free A- and B-chains respectively. There is strong evidence that A-aggregates are a cluster of two chains, whereas B-aggregates consist of four chains with the active heads linked together. This is represented as

$$\begin{aligned} 2 \times \text{—}A &\rightleftharpoons A_2 \\ 4 \times \text{—}B &\rightleftharpoons B_4 \ , \end{aligned}$$

and is illustrated in Fig. 3.2. The equilibrium constants are K_A and K_B respectively.
We define $A^{\text{eff}}(t)$ and $B^{\text{eff}}(t)$ as the amount of effective (i.e. reactive) A- and B-chains and $A(t)$ and $B(t)$ as the total amount of A- and B-chains at some

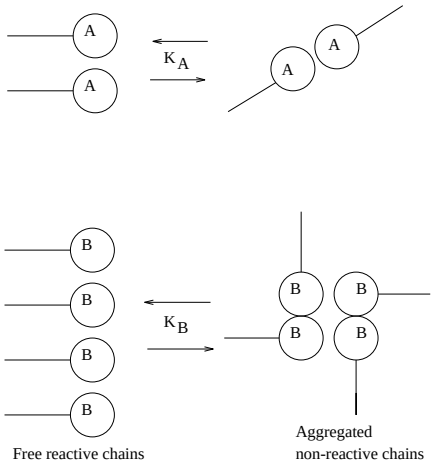

Figure 3.2: the aggregated A- and B-chains can not react with monomers

moment t. We shall call $A(t)$ and $B(t)$ the active chain concentrations. Then we can calculate A^{eff} and B^{eff} by using the equilibrium equation and mass equation. These are

$$\begin{aligned} \text{equilibrium equation} \quad &: \quad K_A = \frac{[A^{\text{eff}}(t)]^2}{[A_2]} \ , \\ \text{mass equation} \quad &: \quad 2[A_2] + [A^{\text{eff}}(t)] = A(t) \ . \end{aligned}$$

Thus, $A^{\text{eff}}(t)$ is the positive solution of

$$x^2 + \frac{K_A}{2}x - \frac{K_A}{2}A(t) \quad = \quad 0 \ ,$$

which is

$$A^{\text{eff}}(t) \quad = \quad \frac{-\frac{1}{2}K_A + \sqrt{K_A^2/4 - 2K_A A(t)}}{2} \ . \tag{3.18}$$

Similarly for $B^{\text{eff}}(t)$ we find

$$x^4 + \frac{K_B}{4}x - \frac{K_B}{4}B(t) \quad = \quad 0 \ ,$$

with solution

$$B^{\text{eff}}(t) \quad = \quad \frac{-f_1 + \sqrt{f_1^2 - 4f_0}}{2} \ , \tag{3.19}$$

where

$$f_1 \quad = \quad \sqrt{2\left(\frac{d_1}{3z} - z\right)} \ ,$$

$$
\begin{aligned}
f_0 &= \frac{d_1}{3z} - z - \sqrt{\left(\frac{d_1}{3z} - z\right)^2 - B(t)K_B/4}\ , \\
z &= \sqrt[3]{\frac{-K_B/128 + \sqrt{d_0^2 - 4e_0}}{2}}\ , \\
e_0 &= -d_1^3/27\ , \\
d_0 &= -K_B^2/128\ , \\
d_1 &= B(t)^2/4\ .
\end{aligned}
$$

3.4 The parameters and initial conditions of the model equations

Here our time t starts to run at the beginning of phase 3. So at that time there are only B-chains present. The starting conditions for the differential equations (3.15) - (3.17) are $a(0) = a_0$, the amount of A monomers added, and $b(0) = b_0$ the amount of B monomers that is still remaining after phase 2. Since we normalized the amount of chains so that we have a total of one mol chains, at the beginning of phase 3 ($t = 0$), $A(0) = 0$ and $B(0) = 1$. The unknown parameters in our model are

$$\theta = (K_A, K_B, k_{aa}, k_{ab}, k_{ba}, k_{bb})$$

In practice one has a fairly good estimate, through measurements, about the values of the reaction rates see Table 3.1 for an example taken from kinetic data. For the equilibrium constants, K_A and K_B, reliable values are much

reaction rate	value ($\text{mol}^{-1}\text{sec}^{-1}$)
k_{aa}	0.0275
k_{ab}	1.272
k_{ba}	0.0006
k_{bb}	0.006

Table 3.1: Measured values

harder to obtain.

If one sets $K_A = K_B = 1$ and solves the differential equation (3.15), (3.16) and (3.17), with starting conditions $a_0 = b_0 = 10$ mol and $B_0 = 1$ mol, the monomer consumptions can be read from the solutions of $a(t)$ and $b(t)$.
In Fig. (3.3) where the consumption of $a(t)$ and $b(t)$ is depicted, we can clearly see a two-stage reaction. B monomers are consumed first, because the reaction rate k_{bb} is much larger than k_{ba}. Moreover if a chain ends with an A monomer it will react much faster with a B monomer than with an A monomer, unless there are no B monomers available. Note that A monomers will react at this stage and therefore *tapering* occurs. Then the second stage begins with the formation of the A end blocks.

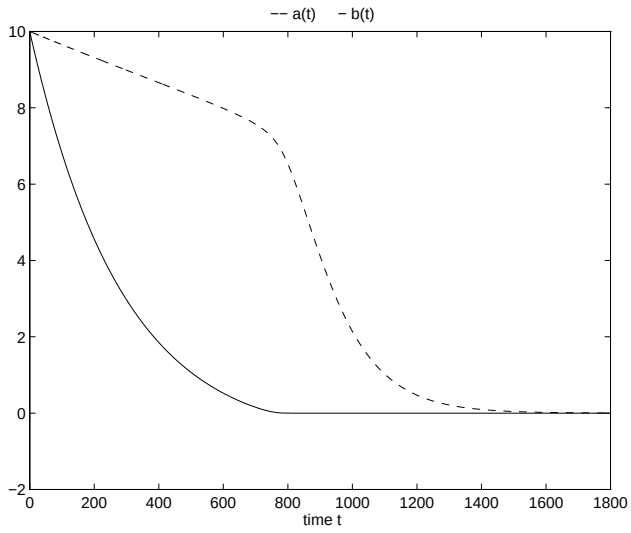

Figure 3.3: consumption of $a(t)$ and $b(t)$ for $a_0 = b_0 = 10$ mol

Also active chain concentrations $A(t)$ and $B(t)$ are drawn, see Fig. (3.4). When the B monomers are almost consumed, it becomes more likely that the B-chains switch heads, i.e. will become an A-chain. As we can see, at the end of the reactive phase (end of phase 3), which we shall call time t_e, a part of the chains still ends with a B monomer; in this example $B(t_e) = 0.56$, i.e. 56%. The quantity $\lim_{t\to\infty} B(t)$ is often called *long diblock content* or just *diblock content*. Hence, these chains have not formed a complete tri-block and therefore they spoil the quality properties of products, which strongly depend on this tri-block structure.

3.5 Length and weight distribution

In this section we shall derive the length and weight distribution of the chains at the end of the reaction, in practice one uses GPC methods to determine length and weight distributions of samples. So we extend the model for concentrations with differential equations describing length and weight. Let us start with the length distribution. Define

$$l_j(t) = \text{probability of a chain with length } j \text{ at moment } t \quad , \tag{3.20}$$

then we can split up $l_j(t)$ as $l_j(t) = l_j^A(t) + l_j^B(t)$ with

$$\begin{aligned} l_j^A(t) &= \text{fraction of chains with length } j \text{ and ending with an } A \text{ monomer} \ , \\ l_j^B(t) &= \text{fraction of chains with length } j \text{ and ending with a } B \text{ monomer} \ , \\ &\quad \text{for } j = 1, 2, 3, \ldots \ . \end{aligned}$$

Note that we have to scale $l_j^A(t)$ and $l_j^B(t)$ by $A(t)$ and $B(t)$ respectively to get probability distributions for the length of the subpopulations of A- and B chains, respectively.

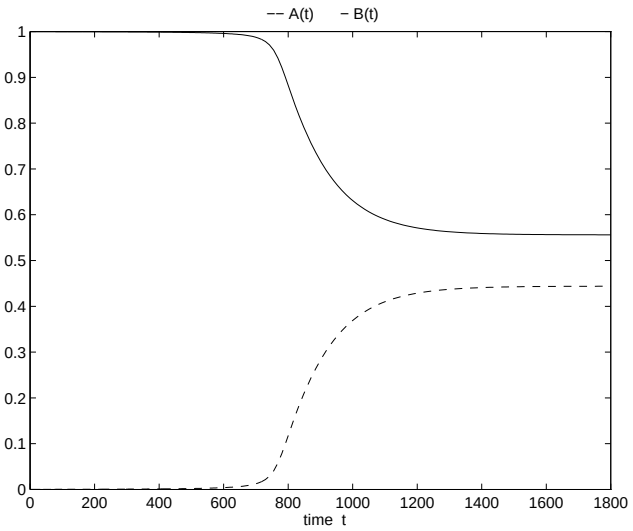

Figure 3.4: evolution of $A(t)$ and $B(t)$

Thus

$$\begin{aligned} l_j^A(t+\Delta t) &= (1-\frac{A^{\text{eff}}(t)}{A(t)})l_j^A(t) + \frac{A^{\text{eff}}(t)}{A(t)} l_j^A q^A + \frac{A^{\text{eff}}(t)}{A(t)} l_{j-1}^A(t) p^{Bb} + \\ &\quad \frac{B^{\text{eff}}(t)}{B(t)} l_{j-1}^B(t) p^{Ba} \ , \qquad (3.21) \\ l_j^B(t+\Delta t) &= (1-\frac{B^{\text{eff}}(t)}{B(t)})l_j^B(t) + \frac{B^{\text{eff}}(t)}{B(t)} l_j^B q^B + \frac{B^{\text{eff}}(t)}{B(t)} l_{j-1}^B(t) p^{Bb} + \\ &\quad \frac{A^{\text{eff}}(t)}{A(t)} l_{j-1}^A(t) p^{Ab} \ , \qquad (3.22) \\ &\qquad \text{for } j = 1, 2, 3, \ldots \ . \end{aligned}$$

Now by rearranging (3.22) and (3.21) into difference quotients and letting Δt tend to zero we get

$$\begin{aligned} \frac{d}{dt} l_j^A(t) &= -l_j^A(t) \frac{A^{\text{eff}}(t)}{A(t)} \left[k_{ab} b(t) + k_{aa} a(t)\right] + \frac{A^{\text{eff}}(t)}{A(t)} l_{j-1}^A(t) k_{aa} a(t) + \\ &\quad \frac{B^{\text{eff}}(t)}{B(t)} l_{j-1}^B k_{ba} a(t) \ , \qquad (3.23) \\ \frac{d}{dt} l_j^B(t) &= -l_j^B(t) \frac{B^{\text{eff}}(t)}{B(t)} \left[k_{bb} b(t) + k_{ba} a(t)\right] + \frac{B^{\text{eff}}(t)}{B(t)} l_{j-1}^B(t) k_{bb} b(t) + \\ &\quad \frac{A^{\text{eff}}(t)}{A(t)} l_{j-1}^A(t) k_{ab} b(t) \ , \qquad (3.24) \\ &\qquad \text{for } j = 1, 2, 3, \ldots \ . \end{aligned}$$

For $j = 0$ we have $l_0^A(t) = 0$ and

$$\frac{d}{dt} l_0^B(t) = -\frac{B^{\text{eff}}(t)}{B(t)} l_0^B(t) \left[k_{bb} b(t) + k_{ba} a(t) \right].$$

Here we used once more the definitions (3.5) - (3.10) of the probabilities. The differential equations (3.24) and (3.23) can be best explained in words, stating that a positive change in the j-th length class is caused by growth of chains with length $j-1$ at a certain rate and negative change is caused by reacting chains of length j. The starting conditions for these differential equations are: $l_0^B(0) = 1, \quad l_j^B(0) = l_j^A(0) = 0$. At time t_e, the end of the reactive phase we have a complete length distribution: $l_j(t_e)$ for $j = 1, 2, 3, \ldots$, as well as 'sub' distributions for B-chains and A-chains: $l^B(t_e)$ and $l_j^A(t_e)$ for $j = 1, 2, \ldots$. For example, we have solved the differential equations starting with 10 mol A and 10 mol B monomers, see Fig.3.5

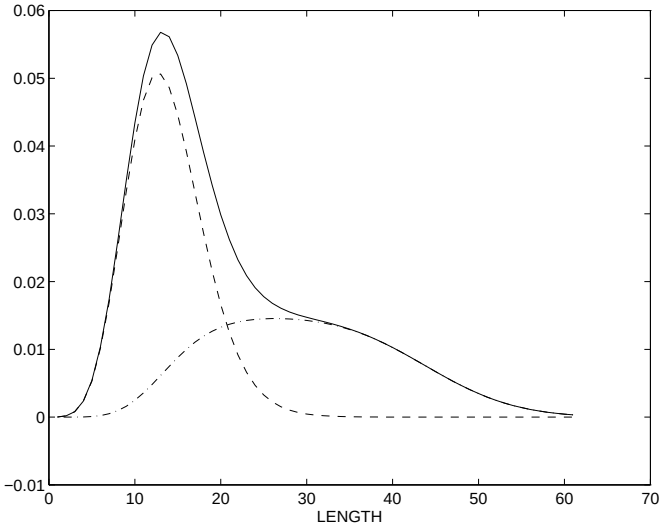

Figure 3.5: - chain length distribution, -. A subpopulation, - - B subpopulation

If the starting conditions are large, i.e. if a lot of monomers are added, then obviously the chains become longer. In order to get the length distribution we have to solve a large system of differential equations, since this system becomes larger when the chains become longer. To avoid this we can describe the length distribution with a few moments. These moments evolve in time and we are specially interested in their values at time t_e. So with (3.20), we define the k-th moment of the length distribution as

$$\mu_k(t) = \mathbb{E}(L^k) = \sum_{j=0}^{\infty} l_j(t) j^k, \qquad \text{for } j = 1, 2, \ldots .$$

We can split up $\mu_k(t)$ as we did with $l_j(t)$, thus $\mu_k(t) = \mu_k^A(t) + \mu_k^B(t)$ with

$$\begin{aligned}\mu_k^A(t) &= \sum_{j=0}^{\infty} l_j^A(t) j^k \ , \\ \mu_k^B(t) &= \sum_{j=0}^{\infty} l_j^B(t) j^k \ .\end{aligned}$$

Using the differential equations (3.23) and (3.24) we can derive differential equations for $\mu_k^A(t)$ and $\mu_k^B(t)$

$$\begin{aligned}\dot{\mu}_k^A(t) &= \frac{d}{dt}\sum_{j=0}^{\infty} l_j^A(t) j^k = \sum_{j=1}^{\infty} \dot{l_j^A}(t) \\ &= -\frac{A^{\text{eff}}(t)}{A(t)}[k_{aa}a(t) + k_{ab}b(t)]\sum_{j=0}^{\infty} l_j^A(t) j^k + \frac{A^{\text{eff}}(t)}{A(t)} k_{aa}a(t)\sum_{j=1}^{\infty} l_{j-1}^A(t) j^k + \\ &\quad \frac{B^{\text{eff}}(t)}{B(t)} k_{ba}a(t)\sum_{j=1}^{\infty} l_{j-1}^B(t) j^k \\ &= -\frac{A^{\text{eff}}(t)}{A(t)}[k_{aa}a(t) + k_{ab}b(t)]\mu_k^A(t) + \frac{A^{\text{eff}}(t)}{A(t)} k_{aa}a(t)\sum_{l=0}^{k}\binom{k}{l}\mu_{k-l}^A(t) + \\ &\quad \frac{B^{\text{eff}}(t)}{B(t)} k_{ba}a(t)\sum_{l=0}^{k}\binom{k}{l}\mu_{k-l}^B(t) \qquad (3.25) \\ &\quad \text{for } k = 1, 2, 3, \ldots \ .\end{aligned}$$

For $\dot{\mu}_k^B(t)$ we have a similar expression as (3.25).

$$\begin{aligned}\dot{\mu}_k^B(t) &= -\frac{B^{\text{eff}}(t)}{B(t)}[k_{bb}b(t) + k_{ba}a(t)]\mu_k^B(t) + \frac{B^{\text{eff}}(t)}{B(t)} k_{bb}b(t)\sum_{l=0}^{k}\binom{k}{l}\mu_{k-l}^B(t) + \\ &\quad \frac{A^{\text{eff}}(t)}{A(t)} k_{ab}b(t)\sum_{l=0}^{k}\binom{k}{l}\mu_{k-l}^A(t) \qquad (3.26) \\ &\quad \text{for } k = 1, 2, 3, \ldots \ .\end{aligned}$$

In the above we used that $\sum_{j=1}^{\infty} l_{j-1}^A(t) j^k = \sum_{l=0}^{k}\binom{k}{l}\mu_{k-l}^A(t)$, which is easily verified.

Starting with 10 mol of A and B monomers we solved the differential equations (3.26) and (3.25) for $k = 1$, this is the expected length, for the solutions of $\mu_1(t)$, $\mu_1^A(t)$ and $\mu_1^B(t)$ see fig 3.6. We notice that the expected chain length $\mu_1(t)$ converges to 20, when t tends to infinity, which is necessary since we added 20 mol monomers on one mol chains.

We now consider the weight distribution. In case a polymer consists of only one kind of monomer the weight, W, and length, L, are proportional: $W = M_0 L$ where M_0 is the molecular weight of the monomer. In our case, where we have two kinds of monomers, we can derive the weight distribution by splitting the chains into A- and B-chains and then keep account of how many A and B monomers accumulate on a chain during the reactive phase. We define

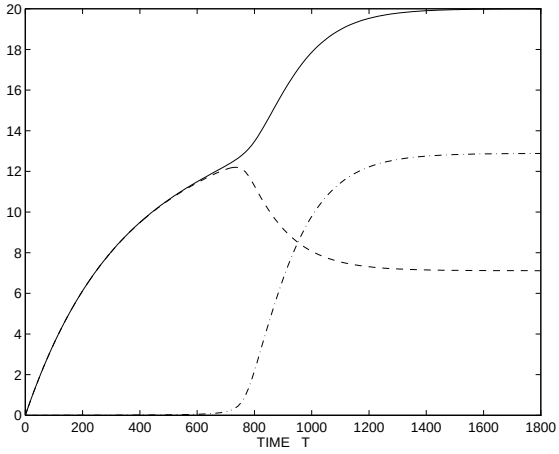

Figure 3.6: - Length expectation - . ditto for chain ending at A - - ditto for chain ending with B

$$
\begin{aligned}
A_{k,l}(t) &= \text{fraction } A\text{-chains with } k \text{ monomers of } B \text{ and } l \text{ monomers of } A \\
B_{k,l}(t) &= \text{fraction } B\text{-chains with } k \text{ monomers of } B \text{ and } l \text{ monomers of } A \\
&\quad \text{for } k, l = 1, 2, 3, \ldots
\end{aligned}
$$

Let M_A and M_B be the molecular weights of the A and B monomers, respectively, then the fraction of chains with weight $kM_A + lM_B$ is equal to $A_{k,l}(t) + B_{k,l}(t)$. In a similar way as for (3.23) and (3.24) we can derive differential equations for $A_{k,l}(t)$ and $B_{k,l}(t)$

$$
\begin{aligned}
\dot{A}_{k,l}(t) &= -A_{k,l}(t)\frac{A^{\text{eff}}(t)}{A(t)}\left[k_{ab}b(t) + k_{aa}a(t)\right] + \frac{A^{\text{eff}}(t)}{A(t)}A_{k,l-1}(t)k_{aa}a(t) + \\
&\quad \frac{B^{\text{eff}}(t)}{B(t)}B_{k,l-1}k_{ba}a(t) \qquad (3.27) \\
\dot{B}_{k,l}(t) &= -B_{k,l}(t)\frac{B^{\text{eff}}(t)}{B(t)}\left[k_{bb}b(t) + k_{ba}a(t)\right] + \frac{B^{\text{eff}}(t)}{B(t)}B_{k-1,l}(t)k_{bb}b(t) + \\
&\quad \frac{A^{\text{eff}}(t)}{A(t)}A_{k,l-1}k_{ab}b(t) \qquad (3.28) \\
&\quad \text{for } k, l = 1, 2, 3, \ldots
\end{aligned}
$$

Notice the analogy with (3.23) and (3.24). It is also possible to calculate the moments of the weight distribution. Let these be defined by

$$
\begin{aligned}
\nu_m(t) &= \sum_k \sum_l (A_{k,l}(t) + B_{k,l}(t))(kM_A + lM_B)^m \\
&= \sum_k \sum_l A_{k,l}(t)(kM_A + lM_B)^m + \sum_k \sum_l B_{k,l}(t)(kM_A + lM_B)^m
\end{aligned}
$$

$$= \nu_m^A(t) + \nu_m^B(t) \ , \ m = 1, 2, \ldots$$

Where $\nu_m^A(t) = \sum_k \sum_l A_{k,l}(t)(kM_A + lM_B)^m$ and $\nu_m^B(t) = \sum_k \sum_l B_{k,l}(t)(kM_A + lM_B)^m$. Hence, with equations (3.28) and (3.27) we get

$$\begin{aligned} \dot{\nu}_m^A(t) &= -\frac{A^{\text{eff}}(t)}{A(t)}[k_{aa}a(t) + k_{ab}b(t)]\nu_m^A(t) + \frac{A^{\text{eff}}(t)}{A(t)}k_{aa}a(t)\sum_{l=0}^{m}\binom{k}{l}M_A^{m-l}\nu_l^A(t) + \\ &\quad \frac{B^{\text{eff}}(t)}{B(t)}k_{ab}b(t)\sum_{l=0}^{m}\binom{k}{l}M_A^{m-l}\nu_l^B(t) \ , \end{aligned} \tag{3.29}$$

$$\begin{aligned} \dot{\nu}_m^B(t) &= -\frac{B^{\text{eff}}(t)}{B(t)}[k_{bb}b(t) + k_{ba}a(t)]\nu_m^B(t) + \frac{B^{\text{eff}}(t)}{B(t)}k_{bb}b(t)\sum_{l=0}^{m}\binom{k}{l}M_B^{m-l}\nu_l^B(t) + \\ &\quad \frac{A^{\text{eff}}(t)}{A(t)}k_{ba}a(t)\sum_{l=0}^{m}\binom{k}{l}M_B^{m-l}\nu_l^A(t) \ . \end{aligned} \tag{3.30}$$

3.6 Triads

Triads are little strings that consists of three monomers within a chain. For example if we have a polymer that looks like $ABBAABA$ then the triads are: ABB, BBA, BAA, AAB and ABA. In practice one can measure the amounts of the different triads by NMR methods (see chapter 2). In this section we shall derive differential equations for the amounts of the eight possible triads.
Define

$$H_{BB}(t) = \text{amount of chains that ends with } BB, \text{ at moment } t \ , \quad (3.31)$$
$$H_{BA}(t) = \text{amount of chains that ends with } BA, \text{ at moment } t \ , \quad (3.32)$$
$$H_{AA}(t) = \text{amount of chains that ends with } AA, \text{ at moment } t \ , \quad (3.33)$$
$$H_{AB}(t) = \text{amount of chains that ends with } AB, \text{ at moment } t \ . \quad (3.34)$$

Then we have

$$\begin{aligned}
H_{BB}(t+\Delta t) &= H_{BB}(t) + \frac{H_{AB}(t)B^{\text{eff}}(t)p^{Bb} - H_{BB}B^{\text{eff}}(t)p^{Ba}}{H_{AB}(t)+H_{BB}(t)} \\
H_{BA}(t+\Delta t) &= H_{BA}(t) + B^{\text{eff}}p^{Ba} - \frac{H_{BA}(t)}{H_{BA}(t)+H_{AA}(t)}(p^{Ab}+p^{Aa})A^{\text{eff}}(t) \ , \\
H_{AA}(t+\Delta t) &= H_{AA}(t) + \frac{H_{BA}(t)A^{\text{eff}}(t)p^{Aa} - H_{AA}(t)A^{\text{eff}}(t)p^{Ab}}{H_{BA}(t)+H_{AA}(t)} \\
H_{AB}(t+\Delta t) &= H_{AB}(t) + A^{\text{eff}}(t)p^{Ab} - \frac{H_{AB}(t)}{H_{AB}(t)+H_{BB}(t)}(p^{Ba}+p^{Bb})B^{\text{eff}}(t) \ .
\end{aligned}$$

So by rearranging the above given equations into difference quotients and letting Δt tend to zero we obtain

$$\dot{H}_{BB}(t) = \frac{H_{AB}(t)B^{\text{eff}}(t)k_{bb}b(t) - H_{BB}(t)B^{\text{eff}}(t)k_{ba}a(t)}{H_{AB}(t)+H_{BB}(t)} \quad (3.35)$$
$$\dot{H}_{BA}(t) = B^{\text{eff}}(t)k_{ba}a(t) - \frac{H_{BA}(t)}{H_{BA}(t)+H_{AA}(t)}A^{\text{eff}}(t)(k_{ab}b(t)+k_{aa}a(t)) \quad (3.36)$$
$$\dot{H}_{AA}(t) = \frac{H_{BA}(t)A^{\text{eff}}(t)k_{aa}a(t) - H_{AA}(t)A^{\text{eff}}(t)k_{ab}b(t)}{H_{BA}(t)+H_{AA}(t)} \quad (3.37)$$
$$\dot{H}_{AB}(t) = A^{\text{eff}}(t)k_{ab}b(t) - \frac{H_{AB}(t)}{H_{AB}(t)+H_{BB}(t)}B^{\text{eff}}(t)(k_{ba}a(t)+k_{bb}b(t)) \quad (3.38)$$

With start conditions $H_{BA} = H_{AB} = H_{AA} = 0$ and $H_{BB} = 1$. So with the equations (3.35) - (3.38) we can keep account of the amount of chains with a certain ending combination of A and B over time. Because a reaction of a monomer with such a chain will immediately form a triad, we are able to add up the number of the triads formed during the reactive period. Hereto define $T_{BBB}(t)$ as the total amount of BBB triad (i.e. over the full chain length) at moment t then we have

$$T_{BBB}(t+\Delta t) = \underbrace{\frac{H_{BB}(t)}{H_{BB}(t)+H_{AB}(t)}}_{\text{fraction of chains that end with } BB} B^{\text{eff}}(t)p^{Bb} + T_{BBB}(t) \ ,$$

hence

$$\dot{T}_{BBB}(t) = \frac{H_{BB}(t)}{H_{BB}(t)+H_{AB}(t)} B^{\text{eff}}(t) k_{bb} b(t) \ . \tag{3.39}$$

Similarly for the other triads we find

$$\dot{T}_{BBA}(t) = \frac{H_{BB}(t)}{H_{BB}(t)+H_{AB}(t)} B^{\text{eff}}(t) k_{bb} b(t) \ , \tag{3.40}$$

$$\dot{T}_{BAA}(t) = \frac{H_{BA}(t)}{H_{BA}(t)+H_{AA}(t)} A^{\text{eff}}(t) k_{aa} a(t) \ , \tag{3.41}$$

$$\dot{T}_{AAA}(t) = \frac{H_{AA}(t)}{H_{AA}(t)+H_{AB}(t)} A^{\text{eff}}(t) k_{aa} a(t) \ , \tag{3.42}$$

$$\dot{T}_{AAB}(t) = \frac{H_{AA}(t)}{H_{AA}(t)+H_{AB}(t)} A^{\text{eff}}(t) k_{bb} b(t) \ , \tag{3.43}$$

$$\dot{T}_{ABB}(t) = \frac{H_{AB}(t)}{H_{BB}(t)+H_{AB}(t)} B^{\text{eff}}(t) k_{bb} b(t) \ , \tag{3.44}$$

$$\dot{T}_{ABA}(t) = \frac{H_{AB}(t)}{H_{BB}(t)+H_{AB}(t)} B^{\text{eff}}(t) k_{ba} a(t) \ , \tag{3.45}$$

$$\dot{T}_{BAB}(t) = \frac{H_{BA}(t)}{H_{BA}(t)+H_{AA}(t)} A^{\text{eff}}(t) k_{ab} b(t) \ . \tag{3.46}$$

At the beginning of the reaction no triads are formed, therefore the starting conditions are zero for the equations (3.39) - (3.46). If we start with 10 mol A and 10 mol B monomers and use the values for the reaction rates given in table 3.1 then we get the following (normalized) triad fractions, see Fig. 3.7. We see that BBB and AAA form the majority, these triads correspond with the B-blocks which are formed first and the final A-blocks. The fraction of BAB gives us an idea of how much single A monomers are tapered in. For example, if we start with a different amount of B monomers, say $b_0 = 1$ then we get a totally different triad structure, see Fig. 3.8. We note that the triad amounts are sample averaged amounts, i.e. the triad fractions given in Fig. 3.7 can be interpreted as the probability of picking a specific triad structure, if one randomly grabs a triad structure from the entire sample. To get a insight into the variation of these amounts within a specific chain, one then could simulate each individual reaction of a large amount of chains.

3.7 Run length distribution

If B monomers are cut out from the chains by ozonolysis (see section 2.4.1) then the remaining parts consist only of A monomers. These are called A-runs. These A-runs can be examined by GPC methods (see section 2.4.2), in order to derive their length distribution. We can build differential equations to describe the evolution in time of these A-run lengths. To this end define the auxiliary variables $A_i(t)$

$A_i(t)$ = the amount of chains with i ($i = 1, 2, 3, \ldots$) times an A monomer at the end, at moment t, $i = 1, 2, \ldots$.

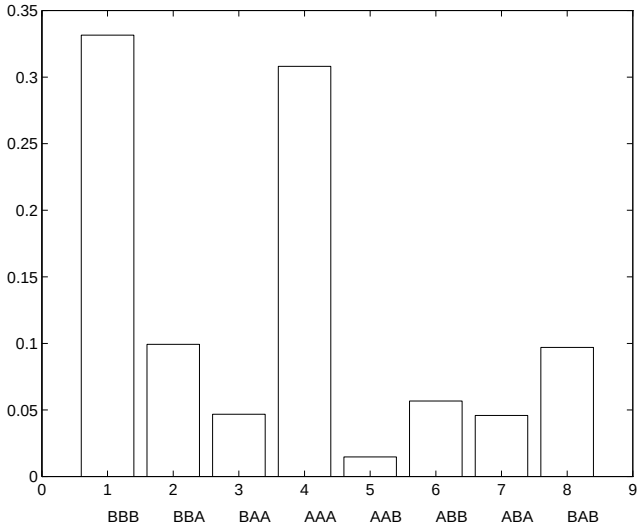

Figure 3.7: triad fractions for the case $a_0 = 10$ and $b_0 = 10$

Hence

$$\begin{aligned}
\dot{A}_1(t) &= -A_1(t)\frac{A^{\text{eff}}(t)}{A(t)}(k_{ab}b(t) + k_{aa}a(t)) + B^{\text{eff}}(t)k_{ba}a(t) \ , \\
\dot{A}_i(t) &= -A_i(t)\frac{A^{\text{eff}}(t)}{A(t)}(k_{ab}b(t) + k_{aa}a(t)) + A_{i-1}\frac{A^{\text{eff}}(t)}{A(t)}k_{aa}a(t) \ , \\
&\quad \text{for } i = 2, 3, \ldots \ . \qquad (3.47)
\end{aligned}$$

Now define

$$\begin{aligned}
R_i^A(t) &= \text{the amount of } A\text{-runs with length } i \\
&\quad \text{for } i = 1, 2, 3, \ldots \ .
\end{aligned}$$

With the aid of (3.47) we get

$$R_1^A(t + \Delta t) = R_1^A(t)B^{\text{eff}}(t)p^{Ba} - A_1(t)\frac{A^{\text{eff}}(t)}{A(t)}p^{Aa} \ , \qquad (3.48)$$

$$\begin{aligned}
R_i^A(t + \Delta t) &= R_i^A(t) + A_{i-1}(t)\frac{A^{\text{eff}}(t)}{A(t)}p^{Aa} - A_i(t)\frac{A^{\text{eff}}(t)}{A(t)}p^{Aa} \qquad (3.49) \\
&\quad \text{for } i = 2, 3, 4, \ldots \ .
\end{aligned}$$

Thus

$$\begin{aligned}
\dot{R}_1^A(t) &= B^{\text{eff}}(t)k_{ba}a(t) - A_1(t)\frac{A^{\text{eff}}(t)}{A(t)}k_{aa}a(t) \ , \\
\dot{R}_i^A(t) &= A_{i-1}(t)\frac{A^{\text{eff}}(t)}{A(t)}k_{aa}a(t) - A_i(t)\frac{A^{\text{eff}}(t)}{A(t)}k_{aa}a(t) \ , \\
&\quad \text{for } i = 2, 3, \ldots \ . \qquad (3.50)
\end{aligned}$$

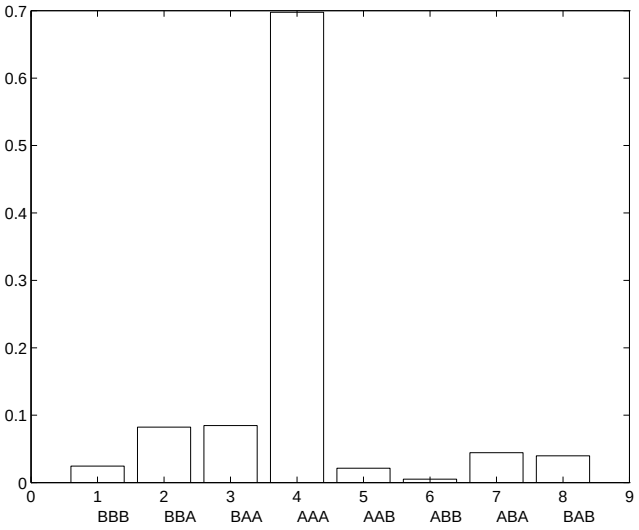

Figure 3.8: triad fractions for the case $a_0 = 10$ and $b_0 = 1$

Obviously equations (3.47) keep account of the actual state of the chain ends at some moment t, i.e. how much A monomers are at the end of a chain, while equations (3.50) accumulate the amounts of a particular run length i. Furthermore if we normalize $R_i^A(t)$, then we get a probability distribution of the A-run length. As an example we solved the equations (3.50) with start conditions $b_0 = 1.5$ and $a_0 = 188$, see Fig. 3.9. We notice a somewhat strange shape. This can be explained by the two stage reaction. First while B monomers react, tapering of A monomers occurs, which accounts for the smaller run lengths. Next when all B monomers are consumed, the final A-blocks are formed, which results in the larger run lengths.

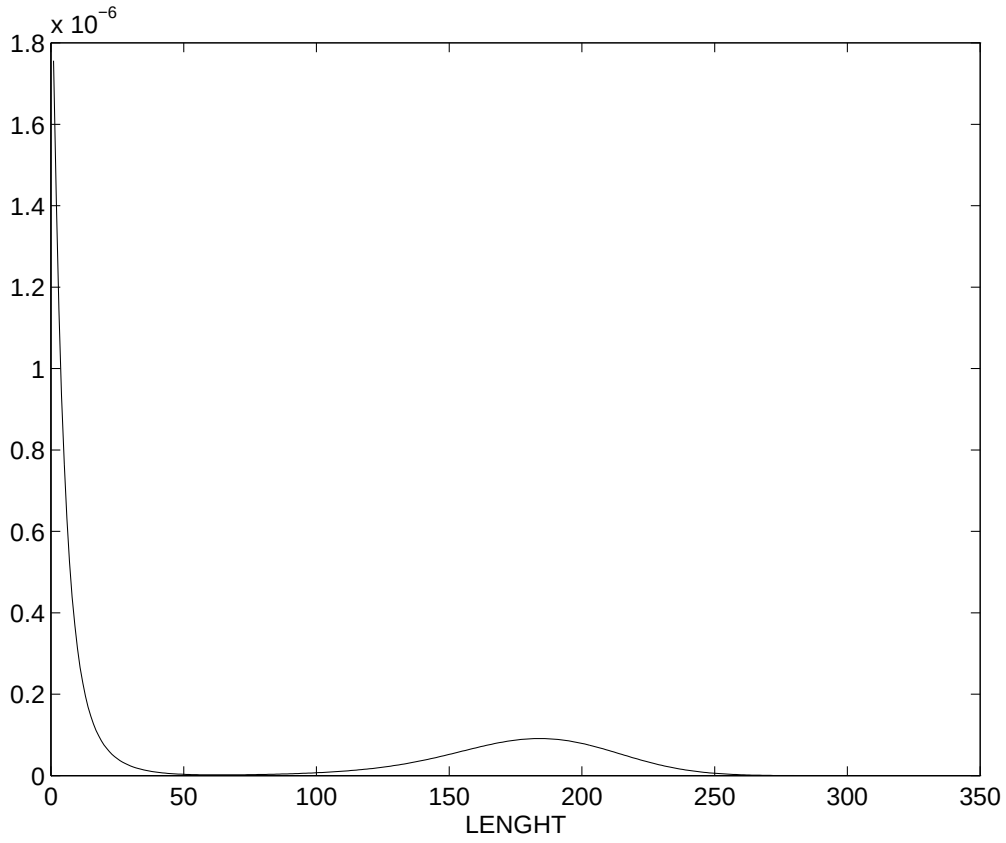

Figure 3.9: probability distribution of A-run length, $a_0 = 188$ and $b_0 = 1.5$

Chapter 4

Parameter estimation

In this chapter we shall discuss briefly the method of Maximum Likelihood Estimation (MLE) and how this applies for general non-linear models. Further we shall show how to identify combinations of parameters, that are hard or impossible to estimate, using singular value decomposition. For a more thorough treatment of these topics see [1].

4.1 Maximum Likelihood Estimation

Roughly speaking one can describe MLE as follows. Given a set of data x, we assume this set x is generated by a known distribution dependent on a parameter vector $\theta = (\theta_1, \ldots, \theta_s)$ which we do not know. Then we maximize with respect to the parameter θ, the probability of observing the given data set under the assumed distribution. In case of discrete distributions this makes sense, but for continuous distributions it does not, since a particular data set then has probability zero. However, the principle is the same. Indeed, if the distribution has density function $f(x, \theta)$ (this may be univariate or multivariate), and we have p observations, jointly independent, then the joint density is the product of the p individual densities. This joint density is known as the *likelihood function* $L(\theta)$, defined as a function of the unknown parameter(vector) θ.

$$L(\theta) = L(\theta|X) = \prod_i f(x_i, \theta)$$

The value of θ which maximizes this function is the maximum likelihood estimator, usually denoted by $\hat{\theta}_{\mathrm{ML}}$ or $\hat{\theta}$. Without proof we give the following properties of $\hat{\theta}_{\mathrm{ML}}$.

Under regularity conditions on f

- $\hat{\theta}$ is consistent: $\hat{\theta}$ converges to the true value θ_0 of θ when p tends to infinity.
- It is asymptotically normally distributed:

$$\sqrt{p}(\hat{\theta}_{\mathrm{ML}} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta)^{-1})$$

when p tends to infinity.

The variance of $\hat{\theta}$ is for large values of p approximately equal to $\mathcal{I}(\theta)^{-1}/p$ and will vanish for $p \to \infty$. $\mathcal{I}(\theta)$ is known as the Fisher Information matrix. It is defined as the $s \times s$ matrix, with i, j-th element $\mathcal{I}_{ij}(\theta)$ given by

$$\mathcal{I}_{ij}(\theta) = \mathbb{E}\left[-\frac{\partial^2 \ln L(\theta)}{\partial\theta_i \partial\theta_j}\right]$$

which can be evaluated at $\hat{\theta}$ to estimate the variance matrix for the MLE. However the exact expected value will rarely be available in practice, because of the complicated non-linear second derivatives. Instead we may estimate the i, j-th element of $\mathcal{I}(\theta_0)$ by

$$\hat{\mathcal{I}}_{ij}(\theta_0) = \left[-\frac{\partial^2 \ln L(\theta)}{\partial\theta_i \partial\theta_j}\right]_{\theta=\hat{\theta}}$$

This is computed simply by evaluating the second derivative matrix of the log-likelihood function at the maximum likelihood estimates, this is the so-called 'observed information matrix' not the expected. Hence it is possible to calculate (asymptotic) confidence intervals for the estimated parameters θ_i with confidence $1 - \alpha$

$$\left(\hat{\theta}_i - \xi_{1-\alpha/2}\sqrt{\frac{\hat{\mathcal{I}}_{ii}(\theta_0)^{-1}}{p}}\ ,\ \hat{\theta}_i + \xi_{1-\alpha/2}\sqrt{\frac{\hat{\mathcal{I}}_{ii}(\theta_0)^{-1}}{p}}\right) \tag{4.1}$$

where ξ_α is the α percentile of the standard normal distribution. In practice one uses also t-intervals to correct for the fact that one has a (small) finite number of datapoints, and the interval given by (4.1) may be somewhat optimistic. It is then replaced by

$$\left(\hat{\theta}_i - t^{\nu}_{1-\alpha/2}\sqrt{\frac{\hat{\mathcal{I}}_{ii}(\theta_0)^{-1}}{\nu}}\ ,\ \hat{\theta}_i + t^{\nu}_{1-\alpha/2}\sqrt{\frac{\hat{\mathcal{I}}_{ii}(\theta_0)^{-1}}{\nu}}\right) \tag{4.2}$$

Here ν is the degrees of freedom, i.e. the number of observations minus the number of parameters, and t^{ν}_{α} is the α percentile of the t distribution with ν degrees of freedom. Note that (4.2) is an exact confidence interval for parameters from a linear model
$y = X\theta + \varepsilon$, where ε is normally distributed.

4.2 General non-linear models

4.2.1 Non-linear models

We now derive the likelihood functions for general non-linear models, so let us assume that for one observation vector y^{obs}, we have the following model:

$$y^{\text{obs}} = y(\theta) + \varepsilon$$

So the observation vector consists of a non-random model prediction $y(\theta)$ and a measurement error ε. Where $y(\theta)$ could be any functional form of θ, ε could have any distribution, and θ is the unknown parameter vector. Usually one assumes normally distributed measurement errors, thus $\varepsilon \sim N(0, \Sigma^*)$. Note that if $x \sim N(\mu, \Sigma^*)$, then the density function is given by:

$$f(x) = (2\pi)^{n/2}(\det \Sigma)^{1/2} e^{(-1/2)(x-\mu)'\Sigma^{-1}(x-\mu)}.$$

Now suppose we have N independent and normally distributed observation vectors and the dimension of one observation vector is n, thus

$$y_i^{\text{obs}} = y(\theta) + \varepsilon_i, \qquad i = 1, \ldots, N.$$

Then:

$$\begin{aligned} L(\theta, \Sigma) &= \mathbb{P}(y_1^{\text{obs}}, y_2^{\text{obs}}, \ldots, y_N^{\text{obs}} | \theta, \Sigma) \\ &= [(2\pi)^n \det(\Sigma)]^{-N/2} \, e^{-\frac{1}{2}\sum_{k=1}^{N}(y_k^{\text{obs}} - y(\theta))'\Sigma^{-1}(y_k^{\text{obs}} - y(\theta))} \, . \end{aligned}$$

4.2.2 Σ Known

Let us assume that Σ is completely known. Thus in order to estimate the parameter θ we have to maximize the likelihood function. Define

$$J(\theta) = -\ln(L(\theta)).$$

Then maximizing L is equavalent to minimizing J. We have:

$$J(\theta) = \frac{1}{2} nN \log(2\pi) + \frac{1}{2} N \log \det \Sigma + \frac{1}{2} \sum_{k=1}^{N} r_k(\theta)' \Sigma^{-1} r_k(\theta) \, , \tag{4.3}$$

where

$$r_k(\theta) = y_k^{\text{obs}} - y(\theta) \, ,$$

the k-th residual.
We see that minimizing (4.3) is equivalent with the familiar 'weighted least squares'. First order conditions for θ can be calculated as follows

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{2} \sum_{k=1}^{N} \frac{\partial r_k(\theta)}{\partial \theta} \Sigma^{-1} r_k(\theta).$$

Here $\frac{\partial r_k(\theta)}{\partial \theta}$ is the Jacobian of the residual, thus

$$\frac{\partial r_k(\theta)}{\partial \theta} = -\frac{\partial y(\theta)}{\partial \theta}.$$

Suppose $y(\theta)$ is a transformed solution of a differential equation at the final time t_e, thus $y(\theta) = g(x(\theta, t_e))$, where g is some transformation and $x(\theta, t_e)$ is the solution of the differential equation

$$\dot{x}(\theta, t) = f(x(\theta, t), \theta) \tag{4.4}$$

at some moment t_e. Then we can compute the Jacobian of the residual as follows. Denote

$$J_y(\theta, t) = -\frac{\partial r_k(\theta)}{\partial \theta} .$$

Then

$$J_y(\theta, t) = \frac{\partial y(\theta)}{\partial \theta} = \frac{\partial g(x(\theta, t))}{\partial \theta}.$$

If we define

$$J_x(\theta, t) = \frac{\partial x(\theta, t)}{\partial \theta},$$

then

$$J_y(\theta, t) = \frac{\partial g}{\partial x} J_x(\theta, t).$$

Using (4.4) and the chainrule we can obtain $J_x(\theta, t_e)$ from the following matrix differential equation:

$$\begin{cases} \dot{J}_x(\theta, t) = \frac{\partial f}{\partial x} J_x(\theta, t) + \frac{\partial f(x,\theta)}{\partial \theta} \\ J_x(\theta, t_0) = \frac{\partial x(\theta, t_0)}{\partial \theta} \end{cases} \tag{4.5}$$

4.2.3 Σ Unknown

Let us assume that Σ is completely unknown, since Σ is symmetric, it contains $n(n+1)/2$ parameters which have to be estimated. In order to estimate these parameters and θ simultaneously we may maximize the likelihood function. Define

$$J(\theta, \Sigma) = -\ln(L(\theta, \Sigma))$$

Then maximizing L is equivalent to minimizing J. We have

$$J(\theta, \Sigma) = \frac{1}{2} nN \log(2\pi) + \frac{1}{2} N \log\det\Sigma + \frac{1}{2} \sum_{k=1}^{N} r_k(\theta)' \Sigma^{-1} r_k(\theta) , \tag{4.6}$$

where $r_k(\theta) = y_k^{\text{obs}} - y(\theta)$, the k-th residual.

Using matrix differentiation rules, given in appendix B, we can obtain the first-order conditions for Σ :

$$\frac{\partial J(\theta, \Sigma)}{\partial \Sigma} = \frac{1}{2} N \Sigma^{-1} - \frac{1}{2} \Sigma^{-1} \left\{ \sum_{k=1}^{N} r_k(\theta) r_k(\theta)' \right\} \Sigma^{-1} = 0$$

Hence we can estimate Σ by $\hat{\Sigma} = D(\hat{\theta})$, where

$$D(\hat{\theta}) = \frac{1}{N}\sum_{k=1}^{N} r_k(\theta) r_k(\theta)' \tag{4.7}$$

for some estimate $\hat{\theta}$ of θ.

And with (4.7) we are able to reduce the number of variables of our minimization problem. In order to find the value $\hat{\theta}$ which minimizes $J(\theta, \Sigma)$ we substitute $D(\theta)$ in $J(\theta, \Sigma)$ we obtain the 'concentrated' likelihood, which only depends on θ,

$$\begin{aligned} J(\theta) &= \frac{1}{2}nN\log(2\pi) + \frac{1}{2}N\log\det D(\theta) + \frac{1}{2}\sum_{k=1}^{N} r_k' D(\theta)^{-1} r_k \\ &= \frac{1}{2}nN\log(2\pi) + \frac{1}{2}N\log\det D(\theta) + \frac{1}{2}\text{trace}\left\{\left(\sum_{k=1}^{N} r_k r_k'\right) D(\theta)^{-1}\right\} \\ &= \frac{1}{2}nN(\log(2\pi)+1) + \frac{1}{2}N\log\det D(\theta), \end{aligned}$$

using $r_k' w = \text{trace}(w r_k')$, since $w = D(\theta)^{-1} r_k$ and r_k are vectors of the same length. Hence minimization of $J(\theta)$ is equal to minimization of $\frac{1}{2}N\log\det D(\theta)$. Using again matrix differentiation rules we can get the first order derivatives of $J(\theta)$,

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_i} &= \sum_k \sum_j \left(\frac{\partial \log\det D(\theta)}{\partial D(\theta)}\right)_{kj} \frac{\partial D_{kj}(\theta)}{\partial \theta_i} \\ &= \text{trace}\left\{D(\theta)^{-1}\frac{2}{N}\sum_{k=1}^{N}\frac{\partial r_k(\theta)}{\partial \theta_i} r_k(\theta)'\right\} \\ &= \frac{2}{N}\sum_{k=1}^{N} r_k(\theta)' D(\theta)^{-1}\frac{\partial r_k(\theta)}{\partial \theta_i} \ . \end{aligned}$$

Where $\frac{\partial r_k(\theta)}{\partial \theta}$ is the Jacobian of the residual, which can be calculated using the matrix differential equation (4.5).

4.2.4 Σ Structured

In the previous subsection we described how to estimate all elements of Σ and θ simultaneously. This only works if the number of observations N is large enough. The problem of too many parameters can be seen in the equation (4.7). For example if $N = 1$ we would require the rank one matrix $r_1 r_1'$ to equal

the positive definite matrix Σ, of rank n which is a contradiction, if $n > 1$. We reduce the number of parameters in Σ by structuring Σ as follows:

$$\Sigma(\rho) = \begin{pmatrix} \rho_1\Sigma_1 & & & \\ & \rho_2\Sigma_2 & & \\ & & \ddots & \\ & & & \rho_p\Sigma_p \end{pmatrix} .$$

Where the ρ_i are unknown parameters and the Σ_i are known. This structure usually appears when the observation vector y_i^{obs} consists of p different data domains and all domains are independent of each other. So y_i^{obs} is also partitioned:

$$y_i^{\text{obs}} = \begin{pmatrix} y_{i1}^{\text{obs}} \\ y_{i2}^{\text{obs}} \\ \vdots \\ y_{ip}^{\text{obs}} \end{pmatrix} , i = 1, \ldots, N .$$

Where y_{ij}^{obs} is the j-th 'sub' observation, i.e. belonging to data domain j. We use the same notation for the residuals, $r_{ij}(\theta) = y_{ij}^{\text{obs}} - y_j(\theta)$ and $r_i(\theta) = y_i^{\text{obs}} - y(\theta)$.

So we have (4.6) again, but now with a structured (i.e. parametrized with ρ) $\Sigma(\rho)$, which enables us to work out the last two terms of (4.6). Indeed, because of the block diagonal structure of $\Sigma(\rho)$ we know that

$$\det \Sigma(\rho) = \prod_{j=1}^{p} \det \rho_j\Sigma_j = \prod_{j=1}^{p} \rho_j^{d_j} \det \Sigma_j.$$

Here d_j is the dimension of the j-th sub-observation vector, thus

$$\frac{1}{2}N \log\det\Sigma = \frac{1}{2}N(\sum_{j=1}^{p} d_j \log\rho_j + \sum_{j=1}^{p} \log\det\Sigma_j) . \tag{4.8}$$

Also,

$$\frac{1}{2}\sum_{k=1}^{N} r_k(\theta)'\Sigma^{-1}r_k(\theta) = \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{p} \rho_j^{-1} r_{kj}(\theta)'\Sigma_j^{-1}r_{kj}(\theta) . \tag{4.9}$$

From (4.8) and (4.9) we get

$$J(\theta, \Sigma(\rho)) = \frac{1}{2}N\sum_{j=1}^{p} d_j \log\rho_j + \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{p} \rho_j^{-1} r_{kj}(\theta)'\Sigma_i^{-1}r_{kj}(\theta) + C,$$

where $C = \frac{1}{2}N\sum_{j=1}^{p} \log\det\Sigma_j + \frac{1}{2}nN\log(2\pi)$ not depending on ρ_j or θ. So we are now able to calculate the first order condition for ρ_i. We find

$$\frac{\partial J(\theta, \Sigma(\rho))}{\partial \rho_j} = d_j\rho_j^{-1}\frac{1}{2}N - \frac{1}{2}\frac{1}{\rho_i^2}\sum_{k=1}^{N} r_{kj}(\theta)'\Sigma_j^{-1}r_{kj}(\theta) = 0$$

which yields the estimator $\hat{\rho}_j$ for ρ_j given by

$$\hat{\rho}_j = \hat{\rho}_j(\theta) = \frac{\frac{1}{N}\sum_{k=1}^{N} r_{kj}(\theta)'\Sigma_j^{-1} r_{kj}(\theta)}{d_j}\ , \tag{4.10}$$

for some value $\hat{\theta}$ of θ.
Expression (4.10) tells us that we can estimate the 'overall' variance of a domain j by taking the weighted squared residues of that domain, corrected by the dimension d_j. In order to find the value $\hat{\theta}$ which minimizes $J(\theta, \Sigma(\rho))$, we can substitute (4.10) in $J(\theta, \Sigma(\rho))$. Note that we then again reduce the dimension of the optimization problem, since $J(\theta)$ becomes independent of ρ. After some tedious algebra we get

$$J(\theta) = \frac{1}{2} N \sum_{i=j}^{p} d_j \log \text{trace}\left\{\Sigma_j^{-1} D_j(\theta)\right\} + c, \tag{4.11}$$

where c is a constant that does not depend on the parameters and $D_j(\theta) = \frac{1}{N}\sum_{k=1}^{N} r_{kj}(\theta) r'_{kj}(\theta)$. If we rewrite $\text{trace}\left\{\Sigma_j^{-1} D_j(\theta)\right\}$ as $\frac{1}{N}\sum_{k=1}^{N} r_{kj}(\theta)'\Sigma_j^{-1} r_{kj}(\theta)$ and use the matrix differentiation rules we can as before calculate the first order derivatives of $J(\theta)$. We find

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{2} N \sum_{j=1}^{p} d_j \left\{ \frac{1}{\text{trace}\left\{\Sigma_j^{-1} D_j(\theta)\right\}} \frac{2}{N} \sum_{k=1}^{N} \left(\frac{\partial r_{kj}(\theta)}{\partial \theta}\right)' \Sigma_j^{-1} r_{kj}(\theta) \right\} \tag{4.12}$$

where we again can compute the Jacobian of the residual $\frac{\partial r_{kj}(\theta)}{\partial \theta}$ by using the matrix differential equation (4.5).

4.3 Singular value decomposition

We shall describe a useful method to find out if certain parameters or combination of parameters are estimable, i.e. whether they influence model predictions enough, or whether they will be obscured by measurement noise. Let us assume we have a (non linear) model

$$y_{\text{pred}} = y(\theta).$$

We will need the so-called sensitivity matrix $S(\theta)$, which is defined by

$$S(\theta) = \frac{\partial y_{\text{pred}}}{\partial \log(\theta)}.$$

So S can be calculated by:

$$(S(\theta))_{ij} = \left(\frac{\partial y_{\text{pred}}}{\partial \log(\theta)}\right)_{ij} = \theta_j \frac{\partial y_i^{\text{pred}}}{\partial \theta_j}.$$

Note that we have taken the derivatives with respect to $\log(\theta)$. This is done because the dependency of $y(\theta)$ on θ through products or quotients of parameters

is more likely than that of linear combinations. This method will rank the importance with respect to the influence on y_{pred} of the linear combinations of the logarithms of parameters. Thereto we perform a singular value decomposition of S

$$S = U \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} V \ ,$$

where U and V are unitary, and the σ_i are called the singular values of $S(\theta)$. This can also be interpreted as

$$\Delta(y_{\text{pred}}) \simeq U \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} V \cdot \Delta(\log(\theta)) \ .$$

The i-th singular value σ_i shows the effect of changes of the parameters in the direction given in the i-th row of V. To see this, suppose this row consists of only zeros except for the j-th element, which must then be a one because V is unitary. A perturbation of 0.01 in $\log(\theta_j)$, say, will cause a change in the measurement space of magnitude $0.01\sigma_i$. So by the size of the singular values we can see how large the impact of a change in a certain parameter is. If a singular value drops below a certain critical value σ_{crit} it is not advisable to try to estimate the corresponding parameter, it is usually better to fix it at a prior known value. This is obvious from the fact that, if a singular value is (nearly) zero, a small change in the parameter will have no effect on the measurement space.
The same reasoning holds when a row of V has more than one non-zero entries. Then the corresponding singular value tells you how well or badly a linear combination of the logarithms of the parameters is determined. In practice the following situation frequently occurs. Two rows of V one corresponding with a large and one with a small singular value, have the following structure

$$\begin{array}{ll} \text{large } \sigma & (0\ldots0 \quad a \quad 0\ldots0 \quad b \quad 0\ldots0) \ , \\ \text{small } \sigma & (0\ldots0 \; -b \quad 0\ldots0 \quad a \quad 0\ldots0) \ . \end{array}$$

Now $a\log(\theta_i) + b\log(\theta_j)$ is well determined but the combination in the orthogonal direction is not. So $\theta_i^a\theta_j^b$ is well determined and θ_j^a/θ_i^b is not.

Chapter 5

Numerical methods

In this chapter we shall describe some numerical methods which we used to solve the differential equations and optimize the likelihood function. Only the basic ideas are given, more details of the methods can be found in [9] and [10]

5.1 The Runge-Kutta method

Consider the differential equation $\dot{x} = f(x,t)$ with starting condition $x(t_0) = x_0$. A first simple numerical solution is based on the following observation:

$$\frac{x(t+h) - x(t)}{h} \approx f(x,t).$$

Therefore

$$x(t+h) \approx x(t) + hf(x,t)$$

which advances a solution from t to $t+h$. Hence if we let η_i be the approximation of the exact solution $x_i = x(t_i)$, then we get the approximations at equidistant points $t_i = t_0 + hi$, $i = 1, 2, \ldots$, according to

$$\begin{aligned} \eta_0 &= x_0 \ , \\ \eta_{i+1} &= \eta_i + hf(\eta_i, t_i), \quad t_{i+1} = t_i + h \ , \\ & \quad i = 1, 2, 3, \ldots \ . \end{aligned} \tag{5.1}$$

Method (5.1) is called the polygon method of Euler, see fig. 5.1. Eulers method is a typical *onestep method*. In general such methods are given by a function $\Phi(x,t,h,f)$ and the approximations η_i are defined by

$$\begin{aligned} \eta_0 &= x_0, \\ \eta_{i+1} &= \eta_i + h\Phi(\eta_i, t_i, h, f) \ , \\ & \quad i = 1, 2, 3, \ldots \ . \end{aligned}$$

In case of the Euler method $\Phi(x,t,h,f) = f(x,t)$, which is independent of the stepsize h. In practice the Euler method is not recommended because the method is not very accurate, and also not very stable. However, consider the

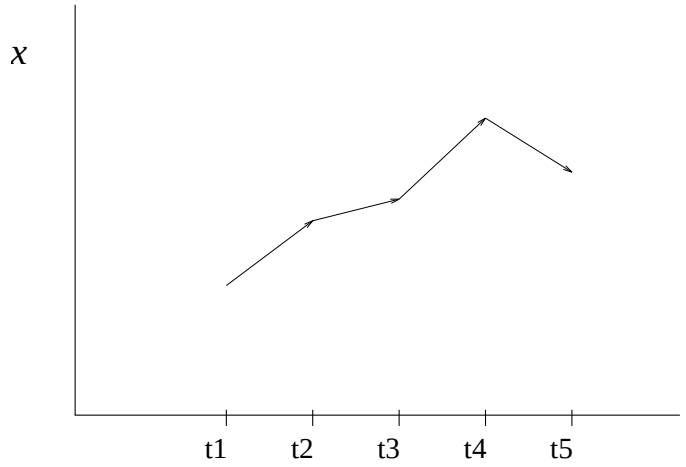

Figure 5.1: Eulers method

use of a 'trial step' to the midpoint of an interval, then use the value of t and x at that midpoint to compute the real step across the whole interval. This leads to the modified Euler rule, or midpoint rule, where $\Phi(x,t,h,f) = f(x+h/2(f(x,t)),t+h/2)$. It is illustrated in Fig. 5.2.

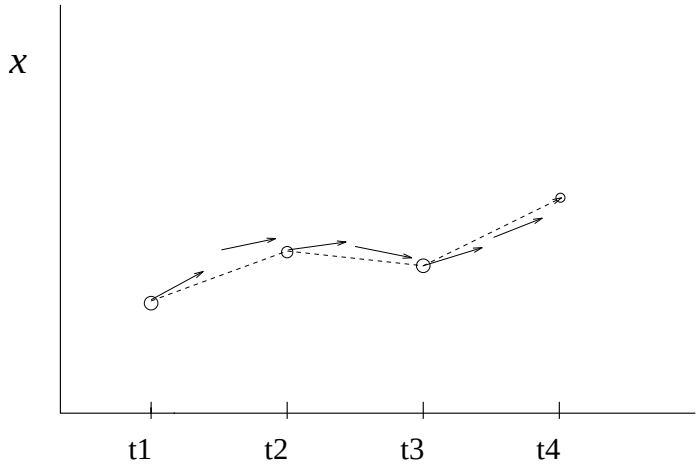

Figure 5.2: midpoint rule

An approximation method is said to be of order p if the error term is $O(h^{p+1})$. One can show by expansion in power series that the Euler method is of order one and the modified Euler method of order two. But we need not to stop here: there are more ways to evaluate the right-hand side $f(x,t)$, that all agree in their first-order terms but differ in their higher-order error terms. Adding up the right combination of these, we can eliminate the error terms order by order. A frequently used method is the classical *fourth-order Runge-Kutta* method. It

has the form

$$\Phi(x,t,h,f) = \frac{1}{6}[k_1 + 2k_2 + 2k_3 + k_4] \tag{5.2}$$

with

$$\begin{aligned} k_1 &= f(x,t) \ , \\ k_2 &= f(x + 1/2hk_1, t + 1/2h) \ , \\ k_3 &= f(x + 1/2hk_2, t + 1/2h) \ , \\ k_4 &= f(x + hk_3, t + h) \ . \end{aligned}$$

Consider the one dimensional case. If $f(x,t)$ does not depend on x, the solution of $\dot{x} = f(t),\ x(t_0) = x_0$ is just the integral $x(t) = x_0 + \int_{t_0}^{t} f(\tau)d\tau$ and the Runge-Kutta method corresponds with the familiar Simpson's rule for approximating integrals. In our problem we first used the Runge-Kutta method, which is a standard procedure, implemented in 'MatLab'. Computing times up to half an hour were not exceptional. This is mainly caused by *stiffness* of the differential equations, see next section.

5.2 Stiff differential equations

Typical for chemical kinetics are the widely different time scales of the various reactions. This results in coefficients in the differential equations of very different orders of magnitude. A linearization of the differential equations therefore gives a large spread of eigenvalues. Integration can give serious problems. The following example will illustrate this. Consider the following linear system:

$$\begin{aligned} \dot{y}_1 &= \frac{\lambda_1 + \lambda_2}{2} y_1 + \frac{\lambda_1 - \lambda_2}{2} y_2 \ , \\ \dot{y}_2 &= \frac{\lambda_1 - \lambda_2}{2} y_1 + \frac{\lambda_1 - \lambda_2}{2} y_2 \end{aligned}$$

with $\lambda_1, \lambda_2 < 0$ and exact solutions

$$\begin{aligned} y_1(t) &= c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 x} \ , \\ y_2(t) &= c_1 e^{\lambda_1 x} - c_2 e^{\lambda_2 x} \ , \end{aligned}$$

for some integration constants c_1 and c_2. If we integrate this system,without round-off errors, using Eulers method, the numerical solutions can be represented as follows

$$\begin{aligned} \eta_{1i} &= c_1(1 + h\lambda_1)^i + c_2(1 + h\lambda_2)^i \ , \\ \eta_{2i} &= c_1(1 + h\lambda_1)^i - c_2(1 + h\lambda_2)^i \ . \end{aligned}$$

These approximations only converge to zero if the steplength h is chosen small enough, so that $|1+h\lambda_1| < 1$ and $|1+h\lambda_2| < 1$ hold. For instance if $\lambda_1 = -1$ and

$\lambda_2 = -1000$, we must have $h < 0.002$. Thus even though e^{-1000x} contributes practically nothing to the solution the factor 1000 in the exponent severely limits the stepsize. This behaviour in the numerical solution is called *stiffness*. Appropriate methods for integrating stiff equations are the so called implicit methods [9]. Since we did not used these methods, we shall not discuss them here. We chose for the simpler explicit method of Bulirsch & Stoer, which yields good results in practice, even for stiff systems.

5.3 The Bulirsch-Stoer method

First we give the *modified midpoint rule*, which advances a vector of dependent variables $x(t)$ from t to $t + H$ by a sequence of n substeps, each of stepsize $h = H/n$. The formulas for the method are

$$\begin{aligned} \eta_0 &= x(t_0) \ , \\ \eta_1 &= \eta_0 + hf(\eta_0, t_0) \ , \\ \eta_{m+1} &= \eta_{m-1} + 2hf(\eta_m, t_0 + mh) \ , \\ &\quad \text{for } m = 1, 2, \ldots, n-1 \ , \\ x(t+H) \approx x_n &= 1/2[\eta_n + \eta_{n-1} + hf(\eta_n, t+H)]. \end{aligned}$$

Here the η's are the intermediate approximations while x_n is the final approximation of $x(t + H)$. The method consists basically of (many) midpoint rules, except for the first and last points. which explains the term 'modified'. In principle this method alone could be used to integrate differential equations, but in practice this method is part of a more advanced technique, *the Bulirsch-Stoer* method. The key idea is extrapolation, a single B.S. step takes us from x to $x + H$. The interval $(x, x + H)$ is spanned by different sequences of finer and finer substeps, integration is performed by using the modified midpoint rule. After each sequence of substeps a polynomial extrapolation is carried out. This extrapolation returns both extrapolated values and error estimates. If we find that the error is not satisfactory we take finer substeps. See Fig. 5.3.

5.4 Practical considerations

An useful way to speed up computing time is adaptive stepsize control. Roughly speaking stepsize control takes big steps to speed through smooth uninteresting country side, while small steps are taken to pass through treacherous terrain. Therefore the algorithm needs additional information about its performance, which uses computing time, but the resulting gain can sometimes be very high.

Another important aspect for successful integration is scaling. Problems can occur if the dependent variable $x(t)$ has components of different magnitude. It is important to scale the errors obtained in every step. In the computer codes, given in appendix A, the user is asked to supply a vector y_{scal}, which scales the errors. For constant fractional errors one should y_{scal} equal to the dependent variable $x(t)$. In case of the different magnitudes of the $x(t)$ components, the

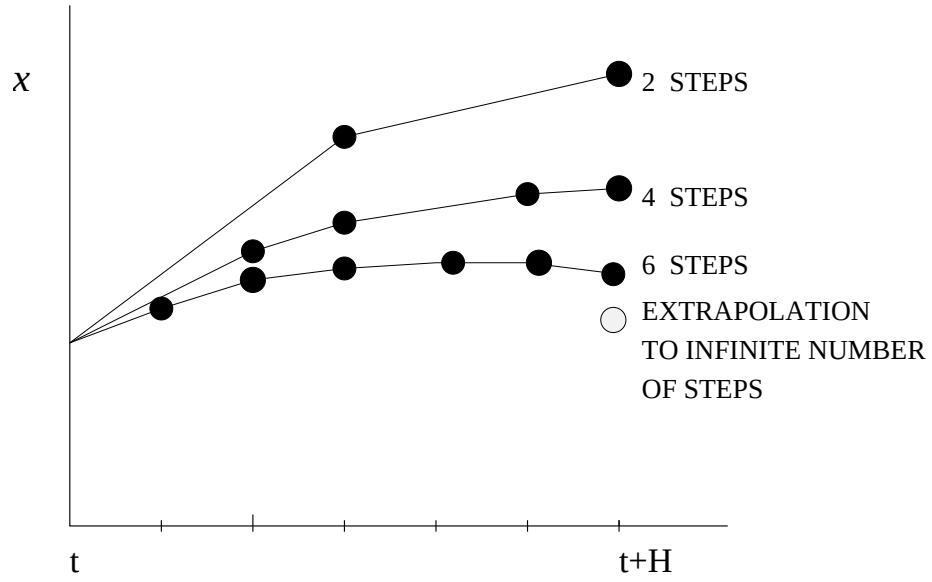

Figure 5.3: extrapolation

algorithm will take small steps where $|x(t)_i|$ takes large values, in order to keep the error per step below some level ε. Of course this increases computing time. Another option is to control the relative errors above some threshold c. To this end set $y_{\text{scal}}(i) = \max(c, |x(t)_i|)$. This could lead to precision loss, but a gain in computing time is obtained.

5.5 Downhill Simplex method

Normally one uses iterative gradient methods to optimize a certain function. These methods use first order derivatives. But we saw in the previous chapter that one evaluation of the gradient of J involves solving a matrix differential equation (see (4.5)), which can demand a lot of computing time. One option would be to switch to numerical derivatives, another to use the Downhill Simplex Method. This method uses only function evaluations.

A simplex is a geometrical figure consisting, in N dimensions, of $N + 1$ points and all their interconnecting line segments. So in two dimensions this is just a triangle and in three dimensions it is a tetrahedron. We are only interested in simplexes that are nondegenerate, that is they enclose a finite inner N-dimensional volume. In order to get the method going one needs a starting simplex. One can take a starting point P_0 and create a simplex by defining N other points, for example $P_i = P_0 + e_i \quad i = 1, \ldots, N$, where the e_i's are N unit vectors. At every point the function is evaluated. Then the downhill simplex method performs a series of steps, most steps consisting of transformation of the simplex by moving the point of the simplex where the function value is largest. See Fig. 5.4 for the possible transformations: one

can almost see the simplex 'walking' through the multi-dimensional terrain. A stopping-criterion could be for example that the decrease in the function value is fractionally smaller than some tolerance ϵ.

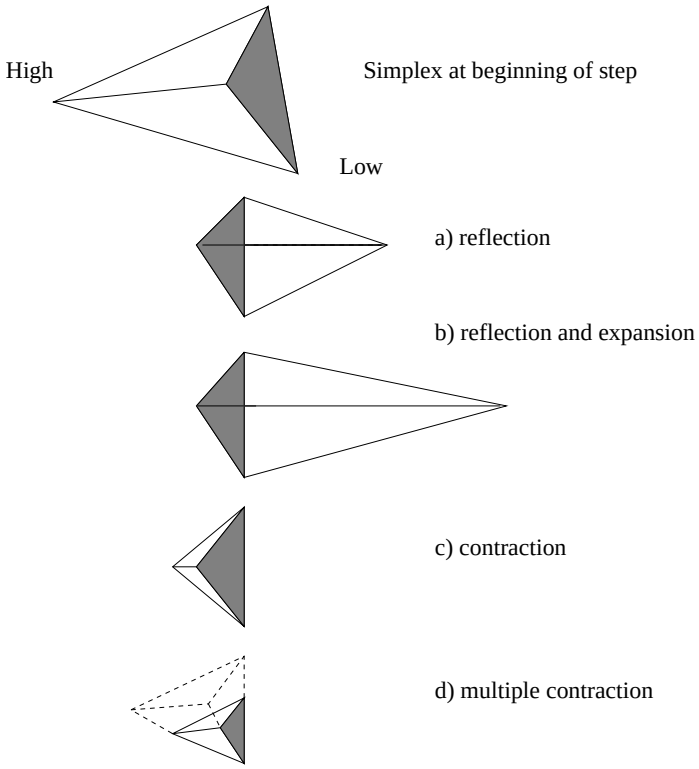

Figure 5.4: Possible outcomes for a step in the downhill simplex method

Chapter 6

Practical results

6.1 Estimation of K_A and K_B from diblock content data

In this section we describe how we estimated the values of K_A and K_B (the equilibrium constants for the A_2 and B_4 complex formation) by 'matching' the observed diblocks (chains that eventually end with a B monomer) as closely as possible to our model diblocks. We have eight measurements of diblock contents corresponding to eight different starting conditions (a_0, b_0), in this case b_0 is systematically zero, see Table 6.1. This means that we are starting with B-chains, and add only A monomers.

start condition (a_0, b_0)	measured diblock contents	
(6.40, 0)	46.5%	50.8%
(10.3, 0)	22.2%	20.6%
(12.2, 0)	18.9%	19.7%
(21.2, 0)	11.3%	13.2%
(23.9, 0)	9.1%	5.7%
(49.9, 0)	3.4%	2.8%
(75.5, 0)	1.0%	1.1%
(123.0, 0)	0.15%	0.90%

Table 6.1: Diblock content data, two measurements per starting condition

Because we have a fairly good idea of the values of the reaction rates (k_{ij}'s), we kept them fixed at the kinetically 'measured' values mentioned in Table 3.1.

If we assume that the measurement errors are standard normally distributed and independent, then we are in the case of known Σ, namely $\Sigma = I$, and maximum-likelihood estimation is equivalent with ordinary least-squares estimation, see section (4.2.2). So we have to minimize (cf. (4.3))

$$J(\theta) \quad = \quad r(\theta)' r(\theta) \tag{6.1}$$

$$= \sum_{k=1}^{8} \left((\text{observed diblock})_k - (\text{model diblock}(\theta))_k\right)^2 \ ,$$

where $r(\theta)$ is the vector of residuals and $\theta = (K_A, K_B)$.
However, this approach is based on the absolute error. So the large diblocks, which will occur when a_0 is small, will dominate since the minimization procedure will try to keep the distance between large diblocks small. To overcome this problem we could scale each residual. Because with each starting condition two measurements were performed, we have a rough idea of the variation of each measured diblock. Hence, we can estimate the diblock variance based on these two points. The squared residuals are scaled by the corresponding estimated diblock variance. Thus (6.1) becomes

$$J(\theta) = r(\theta)' \Sigma_{db}^{-1} r(\theta) \tag{6.2}$$
$$\tag{6.3}$$

where Σ_{db} is a diagonal matrix containing the estimated variances.
Since K_A and K_B could vary from zero to infinity, we scaled them for numerical convenience, introducing the $[0, 1]$ variables

$$\tilde{K}_A = \frac{K_A}{1 + K_A}$$
$$\tilde{K}_B = \frac{K_B}{1 + K_B} \ .$$

If we look at the plot of $J(\theta)$ in Fig.6.1, we see that J decreases in $\tilde{K}_B$. This means that our estimate of K_B is

$$K_B = \infty \ .$$

This implies that the model in which there are no B_4 aggregates, fits the observed data best.
Thus by substituting $\hat{K}_B = \infty$ in our model we can subsequently minimize with respect to K_A. With the Simplex Search algorithm we found the estimate $\hat{K}_A = 0.1649$ with the 95% confidence interval (0.1171 ; 0.2127) using (4.2). In Fig. (6.2) we plotted measured values with the predicted values of the diblock content as function of a_0 for the estimated values of K_A and K_B, together with some other values for K_A and K_B, to illustrate the quality of the fit. In addition the 95% confidence intervals in the prediction domain are plotted, see Fig. 6.3 indicating a very good explanation of the measured dependencies of diblock on a_0. To illustrate what a value of 0.1649 for K_A means for the ratio between effective and active chains, see Fig.6.1. The time t runs along the curve, since at $t = 0$ there are no A-chains and gradually A chains are formed.

A model with B_2 aggregates instead of B_4 aggregates was also evaluated. The result was that the data was, again, best explained by a equilibrium constant (for B_2 aggregates) with a value of infinity.

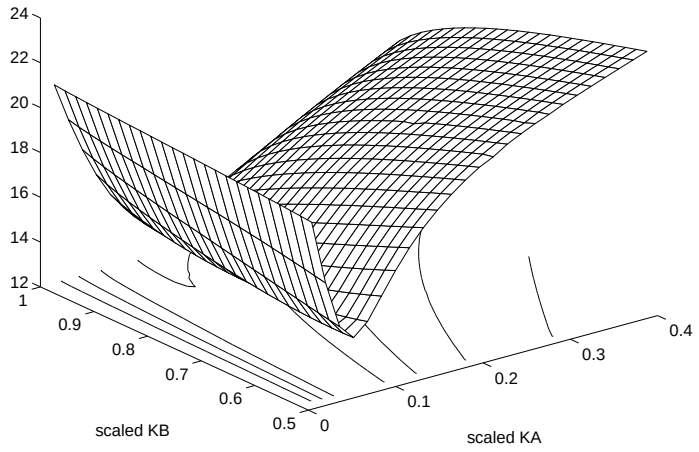

Figure 6.1: plot of $J(\theta)$

6.2 Estimation of reaction rates with NMR data

The next step is to estimate the four reaction rates. In practice one is only interested in the ratio of the reaction rates. For, multiplying the four rates with a fixed number has only effect on the absolute time scale of the reactions. Thus we can normalize the reaction rates with one of them. We then have three remaining parameters, r_1, r_2 and r_3, to be estimated. Their kinetically measured values are shown in Table 6.2, cf. Table 3.1.

ratio	(measured) value
$r_1 = \frac{k_{aa}}{k_{ba}}$	45.83
$r_2 = \frac{k_{ab}}{k_{ba}}$	2120
$r_3 = \frac{k_{bb}}{k_{ba}}$	10

Table 6.2: reaction ratios

The measured diblock contents were obtained through a homopolymerization, i.e. only A monomers were added to one mol B-chains. Thus only parameter r_1 plays a role in this particular case. This is why we tried to estimate K_A (K_B is pinned down at ∞) and r_1 using diblock data only. If the $-\log$ likelihood function is plotted (cf. fig 6.5), a banana-shaped (non-convex) valley is observed where no minimum could be found. Therefore, the use of diblock data only, is insufficient to estimate these parameters.
Hence we need additional information to be able to say anything about the reaction rates.

Fortunately there are measurements of triad fractions available. Three tri-

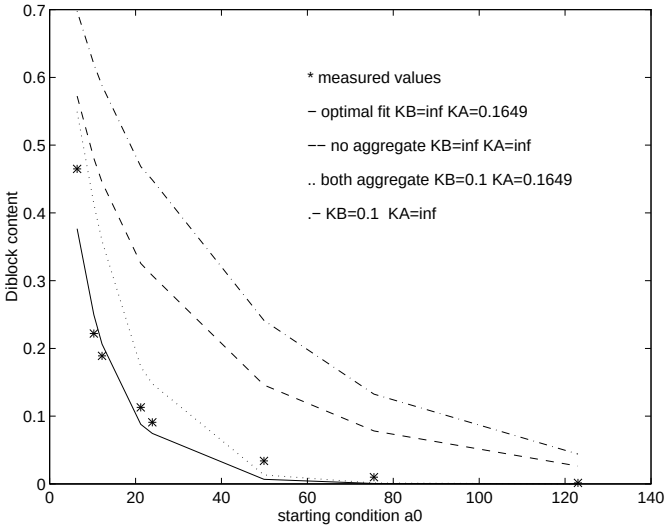

Figure 6.2: Diblock content as function of a_0 for several values of K_A and K_B

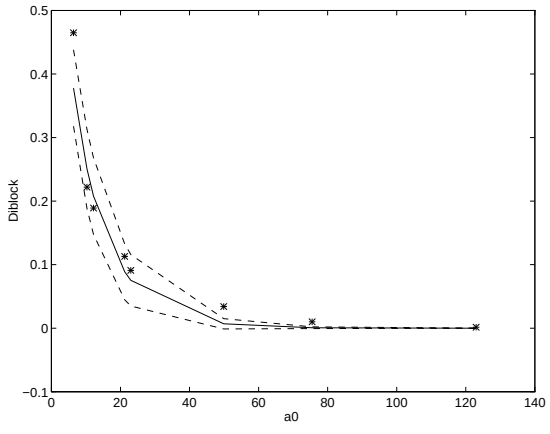

Figure 6.3: 95% confidence intervals around model-predicted diblock contents

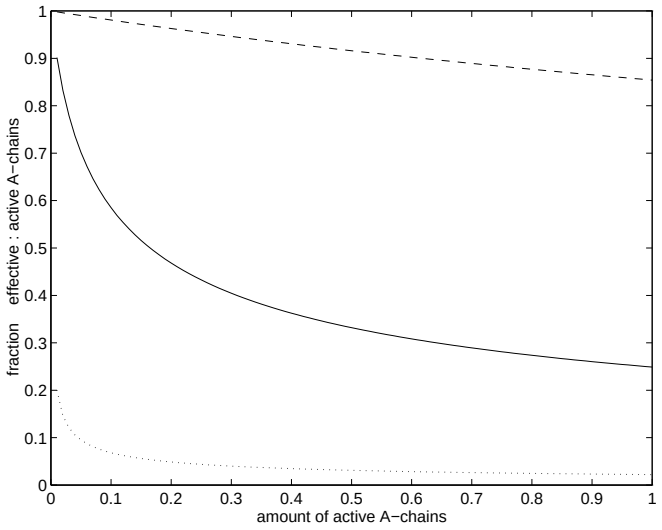

Figure 6.4: fraction of the effective A chains
- - - $K_A = 10$ — $K_A = 0.1649$ $\cdots K_A = 0.001$

ads were measured, namely AAA, BAA and BAB, cf. Table 6.3. In order to use these triad data for an analysis we have to use our model with care. First, the triads are measured for the complete chains, while our model triad predictions are only based on their occurrence in the final block, i.e. propagation in phase 3. But this is easily dealt with, since phase 1 and 2 consist only of a pure A block and a pure B block, with known length. Secondly, the NMR machine is not able to distinguish the position of a polymer chain. Therefore the measured BAA triad consists not only of BAA but of AAB as well, and they must be combined in the model calculations. Finally the measured triads are only proportions, while our model deals with absolute amounts of triads, this requires the computation of normalized triads under the model.

Normalized triad fractions			
starting condition (a_0, b_0)	AAA	BAA	BAB
(106, 9.81)	0.8786	0.1040	0.0173
(106, 39.2)	0.8284	0.1001	0.0710
(106, 108)	0.8278	0.1055	0.0666

Table 6.3: NMR data

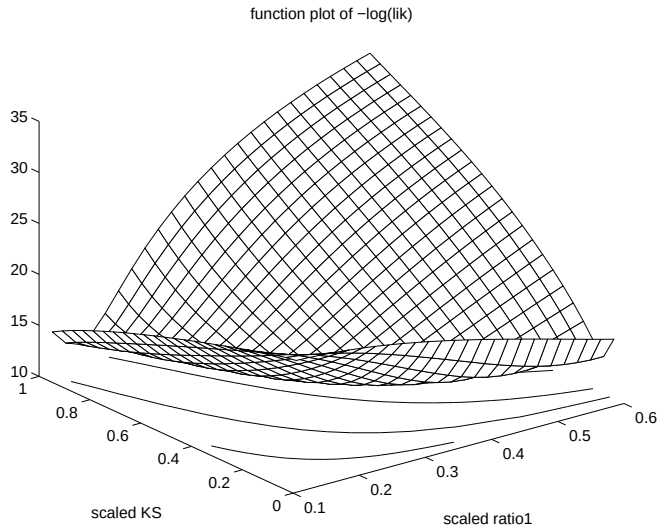

Figure 6.5: plot of J

6.2.1 Identifiability check

The NMR spectrum clearly shows three separate peaks, so that we can assume that the errors of the areas under the three peaks are independent and (2.1) becomes a diagonal matrix. Before we tried to estimate the three ratios r_1, r_2 and r_3, we investigated their identifiability by simulation and singular value decomposition. For the simulation we had to choose three r-values. We took the measured values and then simulated 20 repetitions of (AAA, BAA and BAB) triad distributions with our model. When we successively fixed one r at the measured value and calculated the two dimensional 'sub' likelihood for the two others we got the global results as summarized in table 6.4.

fixed ratio	shape 'sub' likelihood
r_1	convex, with minimum values of r_2, r_3 close to their chosen values
r_2	convex, with minimum values of r_1, r_3 close to their chosen values
r_3	not convex, no minimum could be found

Table 6.4:

These results indicates that a problem would arise if we would estimate the three r's simultaneously. This conjecture was confirmed by singular value decomposition analyses. We have the three model predicted triads

$$y^{\mathrm{pred}}(\theta) = \begin{pmatrix} AAA \\ BAA \\ BAB \end{pmatrix} ,$$

where $\theta = (r_1, r_2, r_3)$.
The sensitivity matrix $S(\theta)$, with respect to changes in $\ln(\theta)$, is

$$S(\theta) = \frac{\partial y^{\mathrm{pred}}(\theta)}{\partial \ln \theta} \ .$$

Using the ideas of section (4.3) we performed a singular value decomposition on $S(\theta)$, and found

$$\begin{aligned} S &= UDV \\ U &= \begin{pmatrix} -0.5372 & -0.6150 & 0.5772 \\ 0.8012 & -0.1584 & 0.5770 \\ -0.2634 & 0.7725 & 0.5778 \end{pmatrix} , \qquad (6.4) \\ D &= \begin{pmatrix} 0.0469 & 0 & 0 \\ 0 & 0.0240 & 0 \\ 0 & 0 & 0.0000 \end{pmatrix} , \qquad (6.5) \\ V &= \begin{pmatrix} 0.6677 & -0.6403 & -0.3796 \\ -0.2683 & 0.2686 & -0.9251 \\ 0.6944 & 0.7169 & 0.0075 \end{pmatrix} . \qquad (6.6) \end{aligned}$$

From this we see that the third singular value is zero (at least numerically) and therefore at least one combination of the r's is not identifiable. The corresponding (third) row of V shows which r's are involved, it indicates that the combination of parameter perturbations

$$0.6944 \ln(r_1) + 0.7169 \ln(r_2) + 0.0075 \ln(r_3)$$

is ill conditioned in the sense that any change in the parameter space in the direction of that combination will not induce any change in the model predicted values. Consequently the product $r_1 r_2$ is ill conditioned and can not be identified from the given triad data, as the coefficient 0.0075 for r_3 is negligible relative to the other coefficients.

So we had to fix either r_1 or r_2 on their measured value, we chose to fix r_2. Because the equilibrium constants K_A and K_B have very little impact on the triads in our model, we kept them fixed at the earlier estimated values. Using both diblock data and NMR data, for estimating our parameters r_1 and r_3 we now have to minimize

$$\begin{aligned} J(\theta) &= r_{db}(\theta)' \Sigma_{db}^{-1} r_{db}(\theta) + r_{\mathrm{nmr}}(\theta)' \Sigma_{\mathrm{nmr}}^{-1} r_{\mathrm{nmr}}(\theta) \qquad (6.7) \\ &= J_1 + J_2 \ , \qquad (6.8) \\ & \qquad (6.9) \end{aligned}$$

where $\theta = (r_1, r_3)$. A plot of the likelihood function is shown in Fig. 6.6. We obtained estimated values $\hat{r}_1$ and $\hat{r}_3$ as given in tabel 6.5 .

As in the case with the diblock data, we scaled squared residuals by the estimated variances of the diblocks. However, with NMR data we have no precise

ratio	estimated value (95% confidence)	kinetically measured value
$\hat{r}_1 = \frac{k_{aa}}{k_{ba}}$	45.48 ($\pm$ 7.77)	45.833
$\hat{r}_3 = \frac{k_{bb}}{k_{ba}}$	4.94 ($\pm$ 8.20)	10

Table 6.5: estimated values

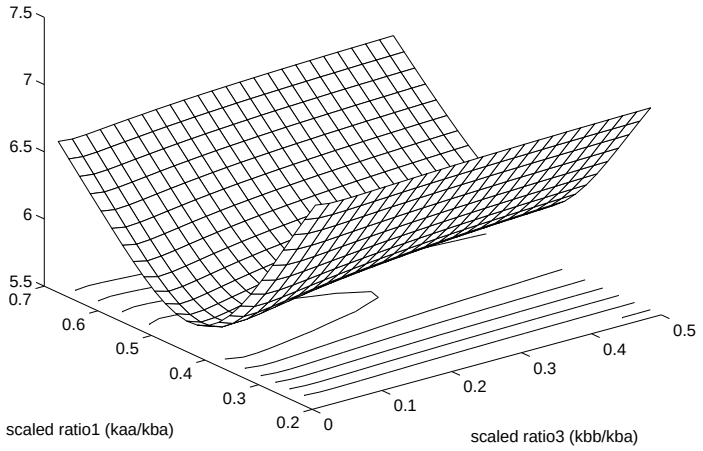

Figure 6.6: plot of -log(likelihood)

idea of the absolute accuracy of the NMR measurements, since only one data set per start condition is available. Usually one then scales the squared residuals 'psychological', i.e. both squared residuals are made to be of the same order of magnitude, $J_1 \approx J_2$. If we plot the NMR data together with the predicted values, we get an indication of the quality of the fit, the dotted bars are the 95% intervals around model-predicted triads, see Fig. 6.7 and 6.8 for the two different sets of values of the starting conditions a_0 and b_0. We notice that the 95% intervals around the predictions only covers the measured triad fraction BAB, indicating a lesser fit than in the diblock domain. A reason for this could be the extreme sensitivity of the triads on the starting conditions. In practice the amounts of reagent are not determined before the reactions, they are determined afterwards. Hence, it is likely that the obtained starting conditions are subjected to inaccuracies

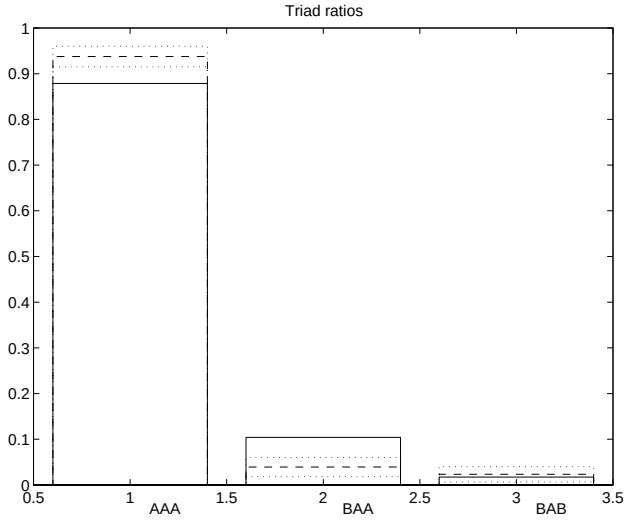

Figure 6.7: - - - model predicted, — measured, start conditions $a_0 = 106$ $b_0 = 9.8$

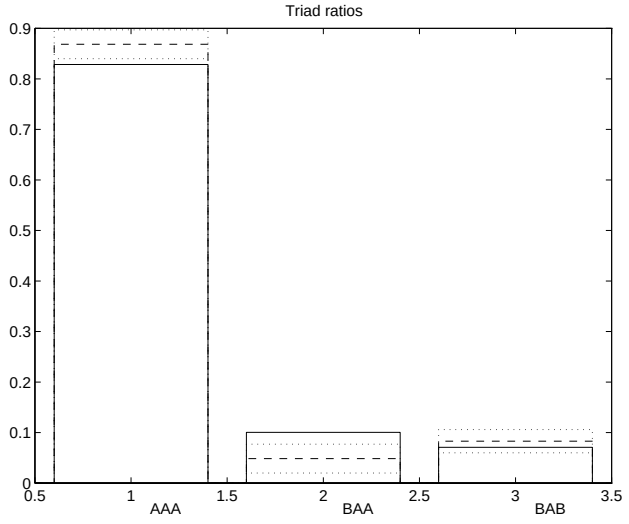

Figure 6.8: - - - model predicted, — measured, start conditions $a_0 = 106$ $b_0 = 39.2$

6.3 GPC data

One ratio remained non-identifiable, thus we needed an additional data domain. The measured distribution of A run length, using GPC methods (see section 2.4) would serve as such. However, GPC is a more qualitative technique of measurement. Therefore 'real' data is not directly comparable with model predictions, and we have to make a few assumptions to rearrange the data in a more useful state. Usually GPC chromatograms show right tails representing components which do not even exist; therefore one has to cut of these tails at certain point. Another shortcoming of GPC is that small A run lengths are not measured at all, for instance the left peak of Fig. 3.9 would not show up in a GPC chromatogram. An assumption would be that the measured A blocks are the end A blocks. However, this is not exactly true, since GPC is able to measure A blocks formed at the tapering section. With equations (3.47) with the empirically obtained values for the reaction rates we can calculate the length distribution of the final A blocks. However, by comparing the length distribution of the final A blocks with the measured GPC curve, see Fig. 6.9, we observe that the two curves vary on a different length scale. The GPC data even shows an average length that is inconsistent with the mass balance. In this experiment there were 188 mol A monomers added to one mol chains. Because of tapering a certain amount of A monomers is built in the B part. With the already estimated parameters we found that approximately 30 mol of the 188 mol A monomers is tapered in. This means that approximately 158 mol A monomers are left for the formation of the A end blocks. As we can see in Fig 6.9 the average length of the A end blocks based on the GPC curve is far too high. We believe that the GPC machine systematically overestimates the long run lengths. A correction model for the GPC data could be used to correct for this overestimation. However, such a model would be rather subjective and therefore it is doubtful whether these corrected GPC data are suitable for parameter estimation.

6.4 Overall parameter estimation

In our model two groups of parameters are present, the equilibrium constants and the reaction rates (reaction ratios). So far the parameters were estimated by fixing one group and estimate the others. A valid reason for fixing the reaction ratios is that they are known fairly well from external experiments. An overall parameter estimation for all the parameters, i.e. all parameters (except K_B) are now treated as free variables, resulted in very 'strange' values for the parameters, with wide confidence intervals. For example r_1 is totally different from the earlier estimated and measured r_1. A singular value decomposition analyses on the parameters r_1, r_3 and K_A resulted in a small value for one of the singular values which indicates that a certain combination of parameters is not identifiable. We have:

$$S = UDV$$

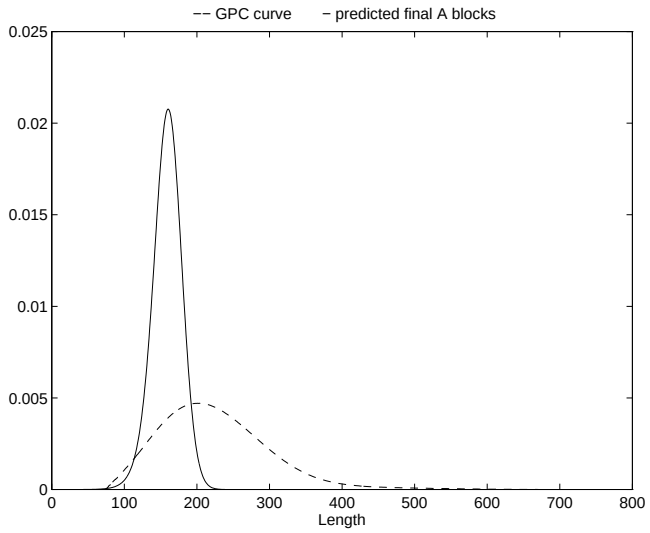

Figure 6.9: comparison of GPC curve and predicted length of final A blocks

$$D = \begin{pmatrix} 0.5099 & 0 & 0 \\ 0 & 0.0113 & 0 \\ 0 & 0 & 0.0076 \end{pmatrix} \tag{6.10}$$

$$V = \begin{pmatrix} 0.9111 & -0.0003 & 0.4123 \\ -0.2434 & 0.8067 & 0.5386 \\ 0.3327 & 0.5910 & -0.7348 \end{pmatrix} . \tag{6.11}$$

U is an 11 by 11 matrix and not relevant to give here. The third row of V (it corresponds with the small singular value) indicates that the parameter combination

$$0.3327 \log(r_1) + 0.5910 \log(r_3) - 0.7348 \log(K_S)$$

is badly conditioned. One reason could be that the two data domains (diblock content and triad fractions) are 'separate' domains. That is, the equilibrium constants have almost no effect on the triad structures, while on the other hand the diblocks are formed by homopolymerization. Thus only one reaction ratio ($r_1 = \frac{k_a a}{k_b a}$) has effect on the diblocks. This insensitivity of certain parameters on the prediction causes (near) invariance of the likelihood in some sub-spaces. Therefore a proper estimation is very difficult or even impossible. We note that the available data were never collected for the aim of parameter estimation. The variation in the data is merely caused by the difference in starting conditions.

6.5 Amount of tapering

As we mentioned in the introduction, tapering strongly influences product properties. With the derived model for copolymerization, we are now in a position

to predict the 'amount of tapering', i.e. the amount of A monomers that has reacted with the living chain before exhaustion of B monomers (cf. Fig. 3.3). Also interesting is the single tapered A monomers, which correspond with the triad structure BAB (cf. eq. (3.46)). Although our model describes only phase 3 (see section 1.2), a complete chain consists of three phases. The first phase has a_0 A monomers, and therefore the total amount of A monomers for a complete triblock is $2a_0$. We compared the tapered A monomers with this quantity, The results are summarized in tabel 6.6 and Fig. 6.10.

starting condition (a_0, b_0)	amount of tapering (mol)	percentage of $2a_0$	percentage single tapered A monomers
(106, 9.81)	18.33	8.65	2.31
(106, 39.2)	32.23	15.2	8.23
(106, 108)	46.13	21.8	14.8
(106, 304)	57.16	26.9	21.9
(106, 520)	63.78	30.1	25.2
(106, 981)	69.23	32.6	28.7

Table 6.6: Amount of tapering

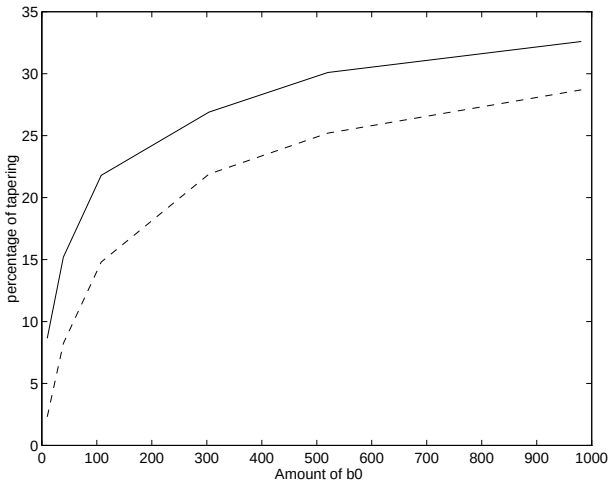

Figure 6.10: — total tapered A - - - single tapered (= isolated) A

Chapter 7

Summary and Conclusions

In this report a model is derived, in order to estimate parameters of a nonstationary Markov model for copolymer propagation with two types of monomers, A and B. This model can be used to predict the chain characteristics. Several characteristics of the polymer chains, length, weight, triad structures and block (run) length distribution are modelled with a set of differential equations. We have numerically checked model consistency, with (probability) mass-balance checks. For example the probability of triad structure BAB must be equal to the probability of run length one of A. Also the expected length of the chains must equal to the amount of added monomers.

Algebraic equations are added to the model, to describe the phenomenon that active chains can aggregate to non-reactive clusters.

We note that this model can be extended for more than two types of monomers. Extra differential equations for the other types are then added to the model, these equations are derived at the same way.

Differential equations are derived that describe the sample-averaged monomer statistics, i.e. the 'average' molecule will obey these equations. To obtain information about the effect of the random fluctuations, one e.g. has to simulate each individual reaction of a large amount of chains. However these fluctuations are neglictable for the relative changes in the concentration of monomers because of the large numbers of molecules.

Runge-Kutta methods were initially used to solve the model equations; however numerical problems related to stiffness suggested the use of implicit methods. However the explicit; relative simple Bulirsch-Stoer extrapolation method yielded quit good results in conjunction with the parameter estimation routines that require the ODE solutions.

The comparison of modelled with measured diblock fractions indicates that the data is best explained by aggregated A-chains and totally free B-chains. We found that the aggregate forming had very little effect on the triad statistics. Using NMR and diblock data together it was possible to get a value for two of the three reaction ratio's, $r_1 = \frac{k_{aa}}{k_{ba}}$ and $r_3 = \frac{k_{bb}}{k_{ba}}$. The values correspond with measured values, which were obtained in a totally different way. To get a better insight into ratio $r_2 = \frac{k_{ab}}{k_{ba}}$ we are in need of an additional data domain. However it appeared that the A run length GPC data, was too unreliable to be

used for parameter estimation.

In the current experiments, data has been collected once for every (batch) operating condition. One has to scale 'psychologically' to balance the information in diblock content and NMR-data more or less equally. However, with repetitive data, one could obtain a better understanding of the variances, allowing the residuals to be weighted according to their corresponding (estimated) variances.

A recommendation would be the use of a totally different data domain, such as data related to the *kinetics* of the chemical reactions. One measures the concentration of the monomers, or other chain characteristics, at some (fixed) times t_i. These quantities could then be compared with model-predicted quantities at time t_i. It is obvious that these data reveal more information than only the characteristics of a polymer chain at the end of the reaction. For example, an expected chain length at the end of the reaction tells us nothing about the parameters, since due to mass balance this expected length must equal to the amount of monomers added, while the progress of the expected length during the reaction is very informative. However there are practical limitations to obtain chain characteristics during the reactions and therefore no data concerning the kinetics were available yet.

Bibliography

[1] W.H. Greene, Econometric Analysis, Macmillan 1990.

[2] G.C. Goodwin, R.L. Payne, Dynamical System Identification Academic Press 1977.

[3] J.C. Randall, Polymer Sequence Determination, Academic Press 1977.

[4] K.H. Altgelt, L. Segal, Gel Permeation Chromatography, Marcel Dekker, Inc. 1971.

[5] L. Ameraal, C++, Academic service 1992.

[6] A.J. Staverman, macromoleculen, Leidse uitgeverij 1981.

[7] M.W. van der Burg, J.C. Chadwick, O. Sudmeijer, H.J.A.F. Tulleken, Makromol. Chem., Theory Simul. **2**, 385 (1993)

[8] P.E. Gill, M.H. Wright, Practical Optimization, Academic Press 1981.

[9] W.H. Press, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, Cambridge University Press 1992.

[10] R. Bulirsch, J. Stoer, Introduction to Numerical Analysis, Springer-Verlag 1980.

[11] The Mathworks, Inc, South Nattick, Massachusetts, MATLAB user's Guide, August 1992.

Appendix A

Computer program

The solver for the differential equations is programmed in C, and can be found in [9]. Two adjustments are made. First the stopping criterion is changed: the solver now stops when the amounts of monomers drop below a certain level, i.e. the reaction is finished, instead of stopped at a fixed time. Secondly a MATLAB gateway has been written to create a so called 'MEX file', see [11]. The great advantage of this is that once the gateway has been created, the C codes can be activated within MATLAB, and therefore can be combined with standard MATLAB commands. The codes given below, consist of the gateway function and a function `odeint` that arranges the Bulirsch-Stoer steps and checks whether the reaction is finished.

```
#include <math.h>
#include "nrutil.h"
#include "mex.h"
#include "nr.h"
#define NRANSI
#define MAXSTP 100000
#define TINY 1.0e-30

/*****************************************/
/*                                       */
/*        COMPUTATIONAL ROUTINE          */
/*           odeABA.c                    */
/*                                       */
/*  This routine solves the ode y'=f(y)  */
/*  using Bulirsch-Stoer method, output  */
/*  xp = time were the reaction has      */
/*       finished.                       */
/*  yp = state vector at time xp         */
/*  required input are:                  */
/*  ystart := initial conditions         */
/*  para   := vector of parameters       */
/*  nvar   := number of variables        */
```

```
/*  x1     := begin time of integration   */
/*  eps    := required accuracy           */
/*  h1     := initial stepsize            */
/*  hmin   := minimum stepsize            */
/*                                        */
/****************************************/


void odeint(double ystart[], double para[],int nvar, double x1,
            double eps, double h1,
    double hmin,
            double *xp, double yp[])

/* This function uses derivs.c and bsstep.c, in derivs.c the diff. equations */
/* are programmed. In bsstep.c one Bulirsch-Stoer step is carried out.       */

{
int nstp,i;
double xsav,x,hnext,hdid,h;
double *yscal,*y,*dydx;

yscal=dvector(1,nvar);
y=dvector(1,nvar);
dydx=dvector(1,nvar);
x=x1;
h=h1;
for (i=1;i<=nvar;i++)  y[i]=ystart[i-1];
        for (nstp=1;nstp<=MAXSTP;nstp++){
         derivs(x,y,dydx,para);
                for (i=1;i<=nvar;i++) yscal[i]=fabs(y[i])+fabs(dydx[i]*h)+TINY;
                bsstep(y,dydx,nvar,&x,h,eps,yscal,&hdid,&hnext,para);
                if (y[2] <= ystart[1])*1e-4){         /* Is the reaction completed?
                  *xp=x;
          for (i=1;i<=nvar;i++){ yp[i-1]=y[i];}
          free_dvector(dydx,1,nvar);
          free_dvector(y,1,nvar);
          free_dvector(yscal,1,nvar);
          return; }
                h=hnext;  }
        nrerror("Too many steps");
}
#undef MAXSTP
#undef TINY
#undef NRANSI


/*************************************/
```

```
/*                                   */
/*           GATEWAY ROUTINE         */
/*              odeABA.c             */
/*                                   */
/*************************************/

void mexFunction(int nlhs,Matrix *plhs[],int nrhs,Matrix *prhs[])

{
  double *ystart;
  double *xp, *yp;
  int nvar;
  double *para;
  double x1,eps;
  double h1,hmin;
  unsigned int k,l;
  int m;

/* check for proper number of arguments */

  if (nrhs != 8)
  { mexErrMsgTxt(" 8 input arguments required.");}
  else if (nlhs !=2)
  { mexErrMsgTxt(" 2 output arguments required.");}

/* ystart must be of size 19x1 */

  k = mxGetM(prhs[0]);
  l = mxGetN(prhs[0]);
  if (!mxIsNumeric(prhs[0]) || mxIsComplex(prhs[0]) ||
      !mxIsDouble(prhs[0]) ||
      !(k==19 && l==1))
  { mexErrMsgTxt(" Ystart must be a vector of length 19");}

/* nvar must be of size 1x1 */

  k = mxGetM(prhs[1]);
  l = mxGetN(prhs[1]);
  if (!mxIsNumeric(prhs[1]) || !mxIsDouble(prhs[1]) ||
      !(k==1 && l==1))
  { mexErrMsgTxt(" nvar must be scalair ");}

/* para must be of size 6x1  */

  k = mxGetM(prhs[2]);
  l = mxGetN(prhs[2]);
  if (!mxIsNumeric(prhs[2]) || !mxIsDouble(prhs[2]) ||
```

```
     !mxIsFull(prhs[2]) || !mxIsDouble(prhs[2]) ||
     !(k==6 && l==1))
  { mexErrMsgTxt(" para must be a vector of length 6 ");}

/* x1 must be of size 1x1 */

  k = mxGetM(prhs[3]);
  l = mxGetN(prhs[3]);
  if (!mxIsNumeric(prhs[3]) || !mxIsDouble(prhs[3]) ||
      !(k==1 && l==1))
  { mexErrMsgTxt(" x1 must be scalair ");}

/* eps must be of size 1x1 */

  k = mxGetM(prhs[4]);
  l = mxGetN(prhs[4]);
  if (!mxIsNumeric(prhs[4]) || !mxIsDouble(prhs[4]) ||
      !(k==1 && l==1))
  { mexErrMsgTxt(" eps must be scalair ");}

/* h1 must be of size 1x1 */

  k = mxGetM(prhs[5]);
  l = mxGetN(prhs[5]);
  if (!mxIsNumeric(prhs[5]) || !mxIsDouble(prhs[5]) ||
      !(k==1 && l==1))
  { mexErrMsgTxt(" h1 must be scalair ");}

/* hmin must be of size 1x1 */

  k = mxGetM(prhs[6]);
  l = mxGetN(prhs[6]);
  if (!mxIsNumeric(prhs[6]) || !mxIsDouble(prhs[6]) ||
      !(k==1 && l==1))
  { mexErrMsgTxt(" hmin must be scalair ");}


/* create matrix for the return argument. */
/* nvar is the number of state variables  */
/* and should be given a specific value   */
/* when using this routine                */

  plhs[0] = mxCreateFull(1,1, REAL);
  plhs[1] = mxCreateFull(nvar,1, REAL);


/* dereference arguments and call the computational routine */
```

```
  xp     =  mxGetPr(plhs[0]);
  yp     =  mxGetPr(plhs[1]);
  ystart =  mxGetPr(prhs[0]);
  nvar   = *mxGetPr(prhs[1]);
  para   =  mxGetPr(prhs[2]);
  x1     = *mxGetPr(prhs[3]);
  eps    = *mxGetPr(prhs[4]);
  h1     = *mxGetPr(prhs[5]);
  hmin   = *mxGetPr(prhs[6]);
  odeint( ystart,para, nvar,x1,eps,h1,hmin, xp, yp);


}
```

Appendix B

Matrix differentiation rules

In this appendix we shall give some useful matrix differentiation rules. These results are used in maximum likelihood estimation (see chapter 4).

Result 1 for a *symmetric* matrix A $\quad \frac{\partial}{\partial x} x'Ax = 2Ax.$

proof:
Let $x = (x_1, \ldots, x_n)$ and let the i,m-th element of A defined by a_{im} then it is sufficient to show that

$$\frac{\partial}{\partial x_i}(x'Ax) = (2Ax)_i = 2\sum_m a_{im}x_m \ .$$

We find

$$\begin{aligned}
\frac{\partial}{\partial x_i}(x'Ax) &= \frac{\partial}{\partial x_i}(\sum_k \sum_m x_k a_{km} x_m) \\
&= \sum_{k=i}\sum_{m\neq i} \frac{\partial}{\partial x_i} x_k a_{km} x_m + \sum_{k\neq i}\sum_{m=i} \frac{\partial}{\partial x_i} x_k a_{km} x_m + \frac{\partial}{\partial x_i} x_i a_{ii} x_i \\
&= 2\sum_{m\neq i} \frac{\partial}{\partial x_i}(x_i a_{im} x_m) + \frac{\partial}{\partial x_i} a_{ii} x_i^2 \\
&= 2\sum_{m\neq i} a_{im} x_m + 2a_{ii}x_i \\
&= 2\sum_m a_{im}x_m \ = \ (2Ax)_i \qquad \square
\end{aligned}$$

Derivatives involving determinants and traces can also be calculated.

Result 2 For any invertible matrix A,

$$\frac{\partial \ln |A|}{\partial A} = (A^{-1})'$$

Proof:
From cofactorization of a matrix A we know that

$$|A| = \sum_{j=1}^{k} a_{ij}(-1)^{i+j} |A_{ij}| ,$$

where A_{ij} is the matrix obtained by deleting the i-th column and j-th row and k is the dimension of A. When the correct sign $(-1)^{i+j}$ is added, it becomes a cofactor.
Hence

$$\frac{\partial |A|}{\partial a_{ij}} = (-1)^{i+j} |C_{ji}|$$

where $|C_{ij}|$ is the ij-th cofactor in A. Notice the reversal in subscripts.

The inverse of A can be computed using

$$a_{ij}^{-1} = \frac{(-1)^{i+j} |C_{ij}|}{|A|},$$

which implies:

$$\frac{\partial \ln |A|}{\partial a_{ij}} = \frac{(-1)^{i+j} |C_{ji}|}{|A|} .$$

Collecting terms we get

$$\frac{\partial \ln |A|}{\partial A} = (A^{-1})' \qquad \square$$

Result 3 For any matrix W and invertible matrix A,

$$\frac{\partial}{\partial A}\text{trace}(WA^{-1}) = -(A^{-1}WA^{-1})'$$

Proof:
Since

$$AA^{-1} = I ,$$

we have

$$\frac{\partial A}{\partial a_{ij}}A^{-1} + A\frac{\partial A^{-1}}{\partial a_{ij}} = 0 .$$

Hence

$$\frac{\partial A^{-1}}{\partial a_{ij}} = -A^{-1}\frac{\partial A}{\partial a_{ij}}A^{-1} = -A^{-1}e_i e_j' A^{-1} ,$$

and therefore

$$\frac{\partial}{\partial a_{ij}}\text{trace}(WA^{-1}) = \text{trace}\left(W\frac{\partial A^{-1}}{\partial a_{ij}}\right) = -\text{trace}\left(WA^{-1}e_i e_j' A^{-1}\right)$$

$$= -e_j' A^{-1}WA^{-1}e_i = -e_i\left(A^{-1}WA^{-1}\right)' e_j .$$

By collecting terms we get

$$\frac{\partial}{\partial A}\text{trace}(WA^{-1}) = -(A^{-1}WA^{-1})' \qquad \square$$