



universität
wien

MASTERARBEIT

Titel der Masterarbeit

„Parameter Estimation of a Kinetic Model of Cytochrome
c Oxidase from a Bayesian Viewpoint“

verfasst von

David Hildenbrandt, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt:

A 066 876

Studienrichtung lt. Studienblatt:

Masterstudium Physik

Betreut von:

Ao. Univ.-Prof. Dr. Martin Neumann



Die vorliegende Masterarbeit wurde zum Teil am Austrian Institute of Technology (AIT) in Wien im Zeitraum zwischen Februar 2012 und Dezember 2013 verfasst.

Contents

1. Introduction	1
2. Probability and Likelihood	3
2.1. Probability	4
2.1.1. Conditional Probability	4
2.1.2. Random Variables	4
2.1.3. Cumulative Probability and Probability Density	5
2.1.4. Expectation Value and Variance	5
2.2. Likelihood Function of a Continuous Probability Distribution	6
2.2.1. Maximum Likelihood Estimation	8
2.3. Interpretations of Probability	9
2.3.1. Classical Probability	9
2.3.2. Frequentist Probability	10
2.3.3. Bayesian Probability	11
3. Bayesian Inference	13
3.1. Statistical Inference	13
3.1.1. Frequentist Confidence Intervals	14
3.2. The Bayesian Paradigm	19
3.3. Drawing Conclusions from the Posterior	21
3.3.1. Marginal Posterior Density	21
3.3.2. Marginal Maximum A Posteriori Estimate	22
3.3.3. Credible Intervals	23
4. Markov Chain Monte Carlo Methods	27
4.1. Metropolis Algorithm	33
4.2. Delayed Rejection	33
4.3. Convergence Assessment	37
4.3.1. Overdispersed Starting Points	37
4.3.2. Between-Within Variance	38
5. Kinetic Model of CcO	49
5.1. Basic Experimental Setup	49
5.2. Kinetic Model	51
5.2.1. Master Equation	51

5.2.2.	Redox Reactions and Protonation	52
5.2.3.	Electrode Reactions	54
5.2.4.	Absorbance Changes Associated with Changes in Redox States . .	58
5.2.5.	Caveats of the Kinetic Model	60
5.3.	Measurement Methods	65
5.3.1.	Cyclic Voltammetry	65
5.3.2.	Fourier-Transform Infrared Spectroscopy	68
5.3.3.	Electrochemical Excitation of CcO and tr-ATR-SEIRAS Measure- ments	73
6.	Implementation	79
6.1.	Function Declarations	79
6.2.	Solving the Master Equation	82
6.3.	Bayesian Inference and MCMC Simulation	88
6.3.1.	Bayesian Parameter Estimation	88
7.	Results	99
7.1.	Conclusion	99
7.1.1.	Fit Results: Tables and Histograms	101
	Bibliography	i
	A. Abstract	ix
	B. Zusammenfassung	xi
	C. Acknowledgements	xiii
	D. Curriculum Vitae	xv

List of Figures

3.1. 68 % confidence limit confidence interval construction for a Gaussian probability density function	15
3.2. Confidence intervals obtained by increase of $-2 \ln \mathcal{L}$ by $\Delta = 1$	18
3.3. Density Quantile Approach	25
5.1. Schematic view of the model system of Cytochrome <i>c</i> Oxidase.	50
5.2. Reaction scheme of an individual redox center <i>m</i> including protonations of oxidized and reduced species.	52
5.3. Reaction scheme for a second-order reaction of total species	54
5.4. Schematic view of the sequential electron transfer mechanism in CcO under anaerobic and aerobic conditions.	57
5.5. Transition State Theory	61
5.6. Applied and effective Potentials	63
5.7. Applied potential sweep in cyclic voltammetry.	65
5.8. Cyclic Voltammogram, schematically	66
5.9. Measured CV of CcO under anaerobic conditions in the non-activated state.	67
5.10. Fourier-Transform Interferometer.	68
5.11. Polychromatic Interference.	69
5.12. Sampling techniques.	72
5.13. Step Scan Technique.	73
5.14. Principle of electrochemically modulated tr-ATR-SEIRAS.	74
5.15. Applied and effective Potentials	76
5.16. Absorbance spectra obtained from tr-ATR-SEIRAS measurements.	77
6.1. Transient interval of the effective potential.	81
6.2. Solutions to the Master Equation for the 8th cycle of the applied potential.	87
7.1. Measured absorbance change plots vs. computer-generated plots according to our model.	104
7.2. Time evolution of the scale reduction factor $\hat{R}$ for the seven parameters of interest.	105
7.3. Histogram and Box and Whisker Plot: E_0^1	106
7.4. Histogram and Box and Whisker Plot: E_0^2	107
7.5. Histogram and Box and Whisker Plot: E_0^3	108
7.6. Histogram and Box and Whisker Plot: E_0^4	109
7.7. Histogram and Box and Whisker Plot: τ	110

List of Figures

7.8. Histogram and Box and Whisker Plot: pK_4^r	111
7.9. Histograms of nuisance Parameters.	112
7.10. Histograms of nuisance Parameters.	113
7.11. Visual representation of the second half of the sample generated by the MCMC procedure showing the sample draws of the four standard redox potentials E_i^0	114

List of Tables

4.1. Simulation results of m different Markov chains thought of as m different treatments	40
6.1. Fixed parameters	89
7.1. Parameter boundaries and random walk step sizes	101
7.2. Starting values	102
7.3. Results	103

1. Introduction

During the second half of the 20th century a paradigm in mathematical statistics, which dates back to Thomas Bayes¹, gained more and more attention among statisticians and other scientists. This so-called Bayesian branch of statistics was inspired by Bayes' famous Theorem and offers some interesting applications in the sciences. The central method that will be used within the scope of this thesis is a form of statistical inference founded on the Bayes Theorem and influenced by the works of Jeffreys [33] and de Finetti [18, 19], among many others. Bayesian Inference is intrinsically tied to a different concept of probability, compared to classical and frequentist interpretations. In the literature this concept is often called subjective or Bayesian probability, a measure for the *degree of belief* in a specific outcome, based on subjective information (see for example D'Agostini [16], p. 3). Bernardo describes this paradigm as "*rational, conditional measure of uncertainty*" (Bernardo [7], p. 1) and it is this subjective trait wherein its strength lies: With the Bayes' Theorem we gain a rule on how to update initial estimations of uncertainty in the light of new information. Bayesian Inference therefore proves its merit when dealing with theoretical models of real physical systems that incorporate parameters not easily accessible through direct measurement: Given a plausible probability density for those parameters, one might update and refine initial approximations of these parameters with help of new experimental evidence.

On behalf of the Biosensors Technologies (BST) group of Austrian Institute of Technology, the author of this thesis aspires to use Bayesian Inference in order to conduct estimations on some of the free parameters of a computational model of an important enzyme of the respiratory chain, called Cytochrome *c* Oxidase (CcO). Based on the work of Walz [67] and Schach [57], this kinetic model of CcO was proposed in order to investigate the coupling of proton to electron transfer steps between the four centers of the enzyme.

Within the scope of this thesis, contributions were made with regard to the calculation of proton association/dissociation probabilities as well as to the implementation of the model on the computer, which were incorporated into a publication by Srajer et al. [63]. Subsequently the implementation of this model was extended to simulate infrared absorbance measurements conducted by means of an experimental technique called time-resolved attenuated total reflection surface-enhanced IR-absorption spectroscopy (tr-ATR-SEIRAS). With the help of the IR absorbance data, a Bayesian parameter estimation was then conducted.

The central aim of this thesis is to examine the applicability of Bayesian Inference in order to gain physically meaningful point estimates as well as probable regions for as many of the models' free parameters as possible. Furthermore, an improvement upon initial estimations for some of the parameters (e.g. pK^s values), deduced from previous experimental evidence

¹Thomas Bayes (1702–1761): English mathematician and Presbyterian minister

1. Introduction

such as fast-scan cyclic voltammetry measurements (see Srajer et al. [63]), was intended. The output of our kinetic model, generated with help of these updated, fitted parameters was then compared to IR absorbance measurement data of three of the four redox centers of CcO. The motivation behind this parameter fit is to gain a better understanding of the coupling between electron and proton transfer processes within the CcO molecule. Ultimately, these efforts were published in cooperation with Schwaighofer et al. [59].

The present work will be divided into three parts: In the first section we will develop the theoretical framework needed to perform Bayesian inference. We will begin this section with the axioms of probability, introduce the concept of the likelihood function and briefly discuss three of the main philosophical stands on the concept of probability itself (chapter 2). Next, an introduction to statistical inference will be given and central differences between frequentist and Bayesian inference will be examined (chapter 3). The statistical concepts of point and interval estimation will be discussed in both views, albeit the focus lies on the Bayesian part. In the subsequent section, practical questions in performing Bayesian inference and calculating multidimensional posterior densities on the computer will be discussed. We will then introduce Markov chain Monte Carlo simulation as a tool to generate samples from these posterior densities, which can be used for Bayesian interval and point estimations of the parameters of interest (chapter 4).

In the second part, the kinetic model of CcO will be discussed in detail. Experimental techniques used to test the model will be explained with a main focus on IR absorbance measurements and tr-ATR-SEIRAS (chapter 5).

Finally, the implementation of the model in MATLAB2010b will be commented and all the technical details of inference on the parameters of interest with help of MCMC simulation will be provided (chapter 6). The results of the Simulation as well as maximum likelihood estimations and credible intervals for each parameter of interest separately will be presented and discussed (chapter 7).

2. Probability and Likelihood

We begin this thesis with the axioms of probability theory, introduced by Andrey Nikolaevich Kolmogorov in 1933 [37, 38]. First, we want to briefly recall some facts from the theory of measurable spaces, following Gut [27]. For a more detailed account we want to refer the reader to Gut's introduction to the theory of probability (Gut [27], pp. 1–15).

It is useful to define a set of all possible outcomes of a random experiment, called the sample space Ω , before devising axioms of probability. The set Ω can be a quite general space and is not necessarily a countable set. To be able to define a meaningful measure of probability on it, one therefore introduces collections of subsets of Ω which abide to certain rules. These so-called σ -algebras $\mathcal{Z}$ are defined by the following properties:

$$\Omega \in \mathcal{Z} \tag{2.1}$$

$$Z \in \mathcal{Z} \Rightarrow Z^c \in \mathcal{Z} \tag{2.2}$$

$$Z_n \in \mathcal{Z}, n \geq 1 \Rightarrow \bigcup_{n=1}^{\infty} Z_n \in \mathcal{Z}. \tag{2.3}$$

One calls a σ -algebra *generated* by Ω if it is the smallest possible σ -algebra that incorporates Ω and writes $\mathcal{Z} = \sigma\{\Omega\}$. An important example is the σ -algebra $\mathcal{R}$ which is generated by Borel sets on the real line. Borel sets are collections of open (or closed) intervals:

$\mathcal{R} = \sigma\{(a, b), -\infty \leq a \leq b \leq \infty\}$ is one of several different possibilities to define $\mathcal{R}$ (see Gut [27], p. 15).

The pair $(\Omega, \mathcal{Z})$ is called measurable space and we can define mappings $\mu : \mathcal{Z} \mapsto (-\infty, \infty)$, called measures, on it. One central property of μ is called countable additivity:

$$\mu\left(\bigcup_{n=1}^{\infty} Z_n\right) = \sum_{n=1}^{\infty} \mu(Z_n) \quad \text{for disjoint } \{Z_n, n \geq 1\}. \tag{2.4}$$

We can now finally introduce the probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is the σ -algebra of measurable subsets of Ω , also called events. Isolated outcomes ω that build up Ω , on the other hand, are usually referred to as *elementary events* (see, e.g., Gut [27], p. 10).

2. Probability and Likelihood

2.1. Probability

Definition 1 Let the sample space Ω be a set consisting of outcomes ω (e.g. the result of a single execution of an experiment) and let $\mathcal{F}$ be a σ -algebra on it. Then a measure $P : \mathcal{F} \rightarrow [0, 1]$, called probability measure, can be defined on $\mathcal{F}$ with the following properties (Kolmogorov's axioms):

1. $0 \leq P(E \in \mathcal{F}) \leq 1$
2. $P(\Omega) = 1$
3. $P(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n)$ for disjoint $\{E_n, n \geq 1\}$.

2.1.1. Conditional Probability

Let $A \in \mathcal{F}$ be a set with non-zero probability $P(A) > 0$, then

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

defines another probability measure. It is interpreted as the probability that B happens, provided that A has occurred before.

2.1.2. Random Variables

In order to describe random events in the language of mathematics one usually defines functions from the probability space to another measurable space. These functions are generally called random variables or elements (cf. Gut [27], p. 25).

Definition 2 A real-valued random variable is defined as a mapping $X : \Omega \rightarrow \mathbb{R}$, assigning a real number $X(\omega)$ to every possible outcome $\omega \in \Omega$. For the inverse function this means that

$$X^{-1}(M) = \{\omega : X(\omega) \in M\} \in \mathcal{F} \quad \forall M \in \mathcal{R}.$$

Hence, the inverse image of every set $M \in \mathcal{R}$ lies in $\mathcal{F}$. Except when explicitly stated, throughout this work we will use the convention of denoting random variables (=the mapping) with upper case letters, while writing a distinct $x = X(\bar{\omega})$, $\bar{\omega} \in \Omega$ —called a realization of X —in lower case letters. One can associate an *induced* probability measure $\mathbb{P}$ to a particular random variable X (see Gut [27], p. 26):

$$\mathbb{P}(M \in \mathcal{R}) = P(X^{-1}(M)) = P(\{\omega : X(\omega) \in M\}).$$

Hence, the probability for a random variable X to take on a value within a subset M of the Borel set $\mathcal{R}$ can be gained from the probability of the image of the inverse function on $(\Omega, \mathcal{F})$. In that manner, the measurable space $(\mathbb{R}, \mathcal{R})$ can be equipped with a probability

measure $\mathbb{P}$, yielding the *induced probability space* $(\mathbb{R}, \mathcal{R}, \mathbb{P})$. When working with random variables, the specific form of the underlying probability space $(\Omega, \mathcal{F}, P)$ is therefore not necessarily a major concern.

2.1.3. Cumulative Probability and Probability Density

Let X be a real valued random variable. We define for a subset $A \subseteq \mathbb{R}$ (in particular left-open intervals like $A = (-\infty, a]$)

$$F(a) = P(X \leq a),$$

the probability for X to be smaller or equal to $a \in \mathbb{R}$ and call F the cumulative probability distribution function of X . The integrand of the function

$$F(a) = \int_{-\infty}^a dx f(x)$$

is called probability density function and satisfies

$$f(x) \geq 0,$$

$$\int_{-\infty}^{\infty} dx f(x) = 1.$$

2.1.4. Expectation Value and Variance

The expectation of a random variable X is the center of (probability) mass of the probability density distribution

$$E(X) := \int_{-\infty}^{\infty} dx f(x)x$$

(Note: for discrete probabilities p_i the integral is a sum over all possible outcomes x_i weighted with p_i). Furthermore, the expectation of a function g of X is

$$E(g(X)) = \int_{-\infty}^{\infty} dx f(x)g(x).$$

The *variance* of X , for example, is defined as expectation of the function

$$g(X) = (X - E(X))^2,$$

hence

$$\text{Var}(X) = E[(X - E(X))^2].$$

2.2. Likelihood Function of a Continuous Probability Distribution

Definition 3 “A parametric statistical model consists of the observation of a random variable x , distributed according to $f(x|\theta)$, where only the parameter θ is unknown and belongs to a vector space Θ of finite dimension. ” (Robert [55], p. 7).

With help of definition 3, in this form given by Robert, we obtain a set of probability density functions $f(x|\theta)$ for the random variable X , where each one is determined by the value of θ .¹ Defining a statistical model in such way is often motivated by an ambition to gain knowledge about quantities that cannot directly be measured by a random experiment in question. Instead, within the context of an investigation, those quantities θ are somehow implicated in the measured outcomes x of a random experiment: Often, the specific form of the statistical model is based on assumptions on the general process leading to the observed outcomes. Limitations in knowledge of this process are then expressed with help of parameters θ . We therefore seek to use statistical analysis to *infer* on those unknown parameters using all available experimental information. In other words, we aspire to expand our knowledge on the unknown parameters θ based on what we measure in order to gain reasonable estimates of θ , functions of θ and future events with probability distributions that depend on θ (see Robert [55], p. 7f).

A useful concept to make inferences about parameters is the notion of *likelihood*. In it's continuous version one can define it in the following way (see Pawitan [51], p. 22):

Definition 4 Given a parametric statistical model with observations $\mathbf{m}$ and a continuous probability density function f depending on a vector of parameters θ , a function

$$\mathcal{L}(\theta|\mathbf{m}) = f(\mathbf{m}|\theta), \quad \theta \mapsto f(\mathbf{m}|\theta) \quad (2.5)$$

is called *likelihood-function*. For a fixed θ , this likelihood $\mathcal{L}(\theta|\mathbf{m})$ is the probability of the recorded data vector $\mathbf{m}$ interpreted as a function of θ .

Thus, $\mathcal{L}$ evaluated at a point θ_0 is nothing but the value of f for a measurement $\mathbf{m}$ (one realization of a random experiment which can yield a scalar as well as a data vector) at this point and can be used for comparing different sets of parameters to find the one that “fits best”, given the available data.

For continuous random variables one has to keep in mind, however, that the probability of a point value $\mathbf{m}$ would be zero. As Pawitan suggests, however, this is not a real problem, since in reality every observation is made with finite precision. Hence, observing $\mathbf{m}$ stands for observing $\mathbf{m} \in (\mathbf{m} - \epsilon/2, \mathbf{m} + \epsilon/2)$, with precision limit ϵ .

¹As Robert [55] points out (p. 6), the statement “the observations $x_1, \dots, x_n$ are distributed according to f ” serves as a shorter form of saying that $x_1, \dots, x_n$ are realizations of the random variable X and f is the probability density function of the distribution function of X .

2.2. Likelihood Function of a Continuous Probability Distribution

For a sufficiently small ϵ , on observing $\mathbf{m}$, the likelihood for $\boldsymbol{\theta}$ amounts to (see Pawitan [51], p. 23):

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|\mathbf{m}) &= P(\mathbf{M} \in (\mathbf{m} - \epsilon/2, \mathbf{m} + \epsilon/2)|\boldsymbol{\theta}) \\ &= \int_{\mathbf{m}-\epsilon/2}^{\mathbf{m}+\epsilon/2} f(\mathbf{m}'|\boldsymbol{\theta})d\mathbf{m}' \approx \epsilon f(\mathbf{m}|\boldsymbol{\theta}).\end{aligned}\tag{2.6}$$

One usually compares likelihoods of different values of $\boldsymbol{\theta}$ with help of likelihood ratios since those are invariant under one-to-one transformations of the measured data $\mathbf{m}$ (see Pawitan [51], p. 29): Let $\mathbf{n} = \boldsymbol{\tau}(\mathbf{m})$ be such a transformation with $\mathbf{n} \in V \subseteq \mathbb{R}^k$, $\mathbf{m} \in \boldsymbol{\tau}^{-1}(V) \subseteq \mathbb{R}^k$ with $\boldsymbol{\tau}^{-1}$ denoting the inverse transformation. For the probability density f it follows that

$$\int_V f(\mathbf{n}|\boldsymbol{\theta})d\mathbf{n} = \int_V f(\boldsymbol{\tau}^{-1}(\mathbf{n})|\boldsymbol{\theta}) \left| \frac{\partial \boldsymbol{\tau}^{-1}(\mathbf{n})}{\partial \mathbf{n}} \right| d\mathbf{n}.$$

Hence—using a somewhat shorter notation—one can write

$$f(\mathbf{n}|\boldsymbol{\theta}) = f(\mathbf{m}(\mathbf{n})|\boldsymbol{\theta}) \left| \frac{\partial \mathbf{m}(\mathbf{n})}{\partial \mathbf{n}} \right|$$

and for the likelihood function it follows that

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{n}) = \mathcal{L}(\boldsymbol{\theta}|\mathbf{m}) \left| \frac{\partial \mathbf{m}}{\partial \mathbf{n}} \right|.$$

Since $\boldsymbol{\tau}$ is a bijective transformation, the original and the transformed data vector contain the same amount of information about $\boldsymbol{\theta}$ and should not lead to different inferences. Hence, one uses the ratio of likelihoods to compare two different values $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ because this function is invariant under such transformations [51]:

$$\frac{\mathcal{L}(\boldsymbol{\theta}_2|\mathbf{n})}{\mathcal{L}(\boldsymbol{\theta}_1|\mathbf{n})} = \frac{\mathcal{L}(\boldsymbol{\theta}_2|\mathbf{m})}{\mathcal{L}(\boldsymbol{\theta}_1|\mathbf{m})}.$$

It follows that the likelihood function can only be reasonably determined up to an arbitrary, constant² factor and the small precision limit ϵ in eq. 2.6 can therefore be ignored.

It has to be noted that despite $f(\mathbf{m}|\boldsymbol{\theta})$ being often written as a conditional probability density, it should not generally be interpreted in that way. Likelihood functions are commonly used both in Bayesian as well as classical (frequentist) statistical inference, however, they do not share the same meaning in both pictures. Within the frequentist paradigm, $\boldsymbol{\theta}$ is seen as a (vector of) parameter(s), with an unknown but distinct, “true” value, as opposed to a random variable with unknown probability distribution (Calvetti and Somersalo [11], p. 22).

In fact, writing $f(\mathbf{m}|\boldsymbol{\theta})$ can be interpreted as hinting at a Bayesian viewpoint and in the literature (e.g. Pawitan [51]) the likelihood function is sometimes expressed as

²i.e. not depending on $\boldsymbol{\theta}$

2. Probability and Likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(\mathbf{m}).$$

The important conceptual differences between the frequentist and the Bayesian picture become more obvious when one tries to define error or confidence intervals, which will be discussed in chapter 3. We will, however, mainly adopt the Bayesian view on likelihood throughout this thesis.

Generally, the likelihood function should be considered in its entirety when we want to obtain rational estimates of the unknown parameters based on the data. It can, however, be sensible to ask which specific parameter values yield the highest probability given our statistical model and the observed data. The *maximum likelihood estimate* provides such a point estimate.

2.2.1. Maximum Likelihood Estimation

Assuming that the random variable M has the probability density function $f(\mathbf{m}|\boldsymbol{\theta})$, one can define the following estimator of $\boldsymbol{\theta}$ (see Robert [55], p. 21):

$$\hat{\boldsymbol{\theta}} = \arg \sup_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathbf{m}). \quad (2.7)$$

This so-called *maximum likelihood estimate* (MLE) gives the value of the parameter vector $\boldsymbol{\theta}$ that maximizes the probability density at $\mathbf{m}$, $f(\mathbf{m}|\boldsymbol{\theta})$ [55]. Roughly spoken, it chooses the one density $f(\mathbf{m}|\hat{\boldsymbol{\theta}})$ which yields the highest probability of observing the given outcome $\mathbf{m}$. Unfortunately, such a maximization is not always feasible using conventional maximization algorithms, especially when a complex, multi-parameter likelihood model is given. Furthermore, there can be several maxima of similar magnitude and giving a single point estimate for a multimodal likelihood function might over-simplify or even misrepresent the information gained from observations.

Keeping this in mind, let us now assume we have a data point vector consisting of measurements recorded at distinct points in time $t_i \in (t_0, t_1, \dots, t_n)$, with likelihood function

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{m}) = f(\mathbf{m}|\boldsymbol{\theta}) \quad (2.8)$$

where $\mathbf{m} = (m_{t=0}, \dots, m_{t=T})$ and $\boldsymbol{\theta}$ is the vector of the assembled parameters of interest. Depending on the problem one has to choose a suitable conditional probability distribution f . Assuming that this conditional probability for each point in time is statistically independent from each other, the likelihood function is of the form ($m_{t_i} := m_i$):

$$f(\mathbf{m}|\boldsymbol{\theta}) = \prod_{i=1}^N f(m_i|\boldsymbol{\theta}). \quad (2.9)$$

A very common choice for the conditional probability distribution f is the Gaussian distri-

bution. During the course of this thesis we will be confronted with the problem of comparing the computer-generated output of a theoretical model depending on unknown parameters to a set of measured data points $\mathbf{m}$. A corresponding likelihood model that can be useful for this task is of the following form:

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{m}) \propto \prod_{i=1}^N \exp\left(-\frac{1}{2\sigma_i^2}(m_i - s_i(\boldsymbol{\theta}))^2\right). \quad (2.10)$$

Where $s(\boldsymbol{\theta})$ is the vector of simulated (computer-generated) points depending on a specific set of parameters. The set of parameters that maximizes $\mathcal{L}$ yields the highest probability to produce the measured outcomes according to the (likelihood) model. For practical reasons this is commonly done by minimizing the double negative logarithm of the likelihood function χ^2 :

$$\chi^2 = -2 \ln(\mathcal{L}) = \sum_{i=1}^N \frac{1}{\sigma_i^2} (m_i - s_i(\boldsymbol{\theta}))^2 + \text{const.} \quad (2.11)$$

2.3. Interpretations of Probability

Kolmogorov's probability calculus, discussed in chapter 2 and theoretical concepts like outcome, event, sample space and random variable are the basis of a purely mathematical theory of probability. Apart from that, there are different views on how to interpret this notion of probability in itself, especially when applied to problems arising in the sciences.

One of the central questions hereby is whether the concept of probability represents a real, physical inclination for an event to occur or merely a measure of the degree of belief an observer might hold in the occurrence of a specific event.

We conclude the present section with a brief examination of some of the philosophical ideas behind the classical, the frequentist and the Bayesian interpretation of probability.

2.3.1. Classical Probability

The classical concept of probability, as it was coined by Laplace [40], considered simple systems (games of chance) with a finite number of clearly determined possible outcomes. All outcomes are considered "equally possible" (Laplace [40], p. 6f) and the probability of one such event is simply the ratio of the number of cases favourable to it to the total number of outcomes. A prominent example for such a system is a six-sided, fair dice and the probability of rolling a prime number, for example, is $\frac{1}{2}$.

One of the main objections to this view centers on the term "equally possible": Comparing degrees of possibility would give rise to a circular definition of probability (since possibility would just be another word for probability).

According to the *The Principle of Indifference*, a term coined by Keynes², in cases where there is an absence of evidence suggesting a preference of one outcome over another, both

²As Keynes points out, this concept goes back to Jacob Bernoulli (see Keynes [36], p. 44)

2. Probability and Likelihood

outcomes can be considered of equal probability (see Keynes [36], p. 44f). For the n -sided fair dice, symmetry arguments convince us that no side can be favoured over any other and the principle of indifference thus enables us to assign a probability of $\frac{1}{n}$ to each one, without having to determine this probability empirically (through repeated trials) (see Keynes [36], p. 90; Hájek [29]).

It is however argued, for example by Hájek, that this notion does not fully resolve the circularity issue: In the case of symmetrically balanced evidence it is not entirely clear how, precisely, one can relate the evidence (e.g. the symmetrical geometry of the dice) to the set of possible outcomes $O_1, O_2, \dots, O_n$ without having to rely on the concept of probability itself again. Hájek states that the “the most obvious characterization of symmetrically balanced evidence” can be given with help of conditional probabilities, as “[...] given evidence E and possible outcomes $O_1, O_2, \dots, O_n$, the evidence is symmetrically balanced iff $P(O_1|E) = P(O_2|E) = \dots = P(O_n|E)$.” (see Hájek [29]).

The above argument is by no means the only objection to the classical view on probability (and its extensions) and there seems to be an ongoing debate, whose full extent is beyond the scope of this thesis.

2.3.2. Frequentist Probability

Another way of looking at probability is the frequentist philosophy whose fundamentals can be, according to Keynes, traced back to Aristotle, who famously said that “the probable is that which for the most part happens” (Aristotle [1], as cited in Keynes [36], p. 87). According to Keynes, the theory itself was first systematically developed by Venn [66] in his work *Logic of chance* (Keynes [36], p. 103; see also: Hájek [29]).

Along this line of reasoning the probability of an event is motivated as a limit of the relative frequency of its occurrence in a large number of (identically prepared) random experiments. Similarly to the classical interpretation of probability, equal weight is attributed to every realization of a set of such experiments and an approximation of the “real” probability of a certain event is merely a matter of counting its realizations proportional to the number of all *recorded* outcomes. Herein lies a key difference to the classical concept, as the probability of an event is constructed on an empirical basis through relating to the total number of realized outcomes instead to all the theoretically *possible* ones [29].

Keynes argues that this property can be seen as one of the main advantages of frequentism, as it is based on measurement of occurrences without the necessity of any new *a priori* conceptions of *possibilities* of a particular system in question or “appeals to intuition” (Keynes [36], p. 105). He notes, that in Venn’s view probability is fundamentally connected with an *ensemble* of events and can be seen merely as a practicable way of summarizing a set of outcomes. It is pointed out, however, that even Venn himself was conscious of the fact that not every phenomenon, which is usually deemed probable (or improbable) could be assigned such a statistical frequency (p. 106).

2.3.3. Bayesian Probability

The Bayesian concept of probability contrasts frequentist ideas in a way that is more compatible with an intuitive meaning of the word *probability* as it is often used in the English language: If, for example, the winning team of the FIFA World Cup would play against a team composed of regional amateur players, one would assert a high probability to the outcome of the World Cup winner triumphing over the amateurs' club³. We can assign this probability even if the game is never to be repeated again and we do not even theoretically understand this probability as the long-run limit to the frequency of victories in a large ensemble of repetitions of this match. Instead we mean to say that *given* the evidence and knowledge we have on the situation, there is a very good chance (=greatly reduced uncertainty), that the World Cup winner will be victorious in *this very* game. In the words of Bernardo, this concept of probability is “[...] a *conditional measure of uncertainty* associated with the occurrence of a particular event, given the available information and the accepted assumptions.” (Bernardo [7], p. 4).

Generally, we can thus use the conditional probability $P(E|C)$ as a way to quantify the degree of belief we hold in the incidence of event E given conditions C . For practical purposes, Bernardo divides C further into the three categories of measured data D , assumptions A about details of the measurement process and contextual knowledge K . The so-defined probability $P(E|D, A, K)$, as is noted in Bernardo [7] (p. 4), may in some cases even be interpreted as a relative frequency of occurrence of comparable events under comparable (“similar”) conditions. The key principle, that enables us to actually calculate such probabilities is the so-called *Bayes theorem*, which we will introduce in section 3.2.

³See Calvetti and Somersalo [11], p. 3 for a similar example.

2. *Probability and Likelihood*

3. Bayesian Inference

The definition of a parametric statistical model (see definition 3) is motivated by the intention to interpret and, in a way, understand the mechanisms that might lie behind a random process such as the outcome of a physical experiment. The parametric approach hereby presupposes the assertion that reasonable assumptions can be made on the specific functional form of the parameterized probability distribution intrinsic to the random variate. A set of observations of such a random process (a sample) can then be used to *infer* on the unknown parameters of its assumed probability density.

3.1. Statistical Inference

As explained in Robert [55], *inference* stands for a mere characterization of the parameter vector θ through deduction from *previously* obtained observations (p. 8). *Probabilistic modeling*, on the other hand, “characterizes the behavior of future observations *conditional* on θ ” (Robert [55], p. 8) and is often also called the “forward modeling problem” (Tarantola [64], p. 492).

In the literature, the extraction of unobserved quantities is therefore usually referred to as *inverse problem* or *inversion* (see for example Calvetti and Somersalo [11], p. 1). As Robert points out (p. 8), this idea of statistical inversion is being reflected in the very definition of the likelihood function

$$\mathcal{L}(\theta|\mathbf{m}) = f(\mathbf{m}|\theta),$$

which is merely the interpretation of the density f as function of the unknown parameter(s) θ . While Bayesians as well as frequentists make use of the likelihood function, both groups interpret this tool in a very different way. In Bayesian analysis we consider the unknown parameters as *random variables* of an n -dimensional parameter space Θ and therefore the likelihood function can be seen as a probability density function in θ . Frequentists, on the other hand use $\mathcal{L}$ as a mere tool to compare ratios between the likelihood function evaluated at a specific parameter value and a certain optimal value $\theta_{\max}$, which maximizes $\mathcal{L}$. In doing so, regions which cover the “true” value of θ in a specific fraction of all cases can be determined (see section 3.1.1). We can sum up the key difference between Bayesian and frequentist inference strategies as follows (see Calvetti and Somersalo [11], p. 39):

In the frequentist’s point of view, unknown parameters of a statistical model are treated as intrinsically *deterministic* quantities. Therefore, parameters have underlying “true” values which can be estimated with help of statistical inference. The Bayesian way of thinking, on the other hand, promotes the idea that “[...] *any* quantity that is not known exactly is regarded as a random variable.” (Calvetti and Somersalo [11], p. 39).

3. Bayesian Inference

According to this notion, a priori information on the parameters as well as further knowledge that can be obtained through measurement is inherently incomplete. Because of this fundamental uncertainty we need to think of parameters as random variables with probability distributions, even if the unknown model parameters account for real physical quantities with fixed and deterministic values. A joint probability distribution of all parameters is merely a convenient way of condensing all available information *conditional* on basic assumptions, prior knowledge and obtainable data¹(see Bernardo [7], p. 2).

Apart from gaining *point estimates* for the underlying “true” value of some parameter θ from measurements, an experimenter often seeks to map out regions wherein the true value might be found with a certain (high) probability. In the following, we will try to elaborate on the main conceptual difference between the Bayesian and the frequentist view on these so-called *confidence regions*.

3.1.1. Frequentist Confidence Intervals

Following the description given by Cousins [14], we consider the case of the measurement of some scalar physical quantity (e.g. the mass of a particle) expressed in form of the random variable X . We denote the expected value of this quantity by $E(X) = \theta$ and the outcome of a specific measurement by x_0 . Assuming for example that the measuring device produces data points that are normally distributed with mean θ —the unknown *parameter* in this case—and a known variance σ_x^2 , we can specify a probability density for measuring the value x given the true value θ :

$$P(x|\theta) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-(x-\theta)^2/2\sigma_x^2}. \quad (3.1)$$

Frequentist confidence intervals (x_1, x_2) , as discussed by Neyman²[49], can then be defined at a specific confidence level, for example 68%. The defining property of these intervals is that θ will be included in 68% of the intervals calculated from a large ensemble of trial measurements, a principle also known as “frequentist coverage” (Cousins [14], p. 400). In a truly frequentist sense this concept therefore represents the relative frequency of the inclusion of θ in arbitrary, $(1 - \alpha)100\%$ confidence level confidence intervals, which tends to the desired percentage value in the long run. Cousins notes, however, that since θ is fixed while x_1 and x_2 are probabilistic variables, we cannot think of a frequentist confidence interval as a measure for a *degree of belief* that θ will be found within this interval in a particular trial experiment. This is emphasized by the fact that $P(X|\theta)$ is a probability density in x and not in θ .

A key point hereby is whether $P(X|\theta)$ can plausibly be determined or at least approximated for a particular experiment and usually such questions need to be considered in the design of these experiments. Here, however, we will confine the discussion to the case of normally distributed data with unknown mean.

¹cf. the Bayesian concept of probability discussed in sec. 2.3.3

²As cited in Cousins [14], p. 399

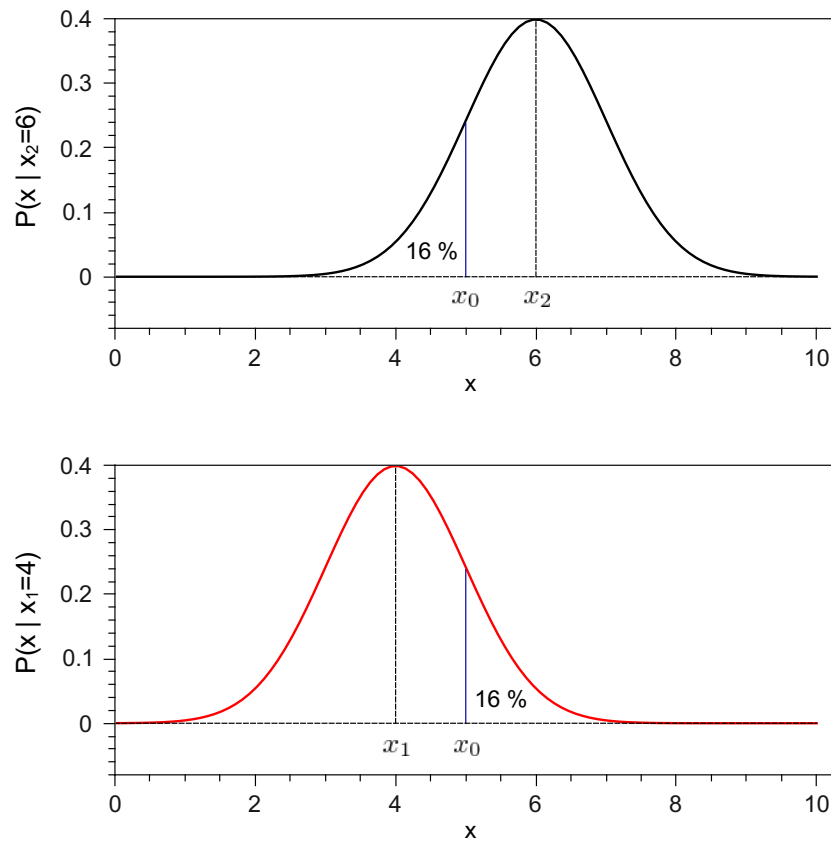


Figure 3.1.: 68 % confidence limit confidence interval construction for a Gaussian probability density function with known $\sigma_x = 1$ for a single measurement $x_0 = 5$: Construction of the left and right interval boundaries x_1 and x_2 . This illustration was redrawn and adapted from Cousins [14], p. 399.

3. Bayesian Inference

One way to construct a classical, 68 % confidence interval (x_1, x_2) for a single measured point x_0 is to find a value $x_1 < x_0$ such that 16 % of the area under $N(x_1, \sigma_x)$ is reserved for values greater than x_0 , hence

$$P(X \geq x_0 | x_1) = \int_{x_0}^{\infty} N(x_1, \sigma_x) dx = 16 \% \quad (3.2)$$

and a value $x_2 > x_0$ such that 16 % of the area underneath $N(x_2, \sigma_x)$ covers values of x smaller than x_0 ,

$$P(X \leq x_0 | x_2) = \int_{-\infty}^{x_0} N(x_2, \sigma_x) dx = 16 \%. \quad (3.3)$$

Thus, since $P(X|\theta)$ is symmetric, we get the interval boundaries $x_1 = x_0 - \lambda_0$ and $x_2 = x_0 + \lambda_0$. One can state the result of a single measurement as $x_m \pm \lambda_m$ and if we repeat this procedure for a large number of measurements, then in 68 % of all cases the true value can be found within such a region (see Cousins [14], p. 399).

Consider now the case of an ensemble of sample measurements $\{x_0^k, x_1^k, \dots, x_n^k\}$, each sample k measuring n outcomes of a random experiment (denoted by a random variable X), assumed to be normally distributed with known variance σ but unknown expectation $\mu = E(X)$. The sample mean $\bar{X}$ is a function of X and an appropriate, unbiased estimator of μ . Since the mean $\bar{X}$ is also a random variable with Gaussian probability density function and expectation μ ($\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$), we can construct confidence intervals for $\bar{X}$ analogously to the above example of a single measurement. We begin with defining central intervals $(\bar{x}_1, \bar{x}_2)$ for the random variable $\bar{X}$, which—in the long run—tend to cover $\bar{x}^k = \frac{1}{n} \sum_{i=1}^n x_i^k$ (realizations of $\bar{X}$) with probability $1 - \alpha$:

$$P(\bar{x}_1 \leq \bar{X} \leq \bar{x}_2) = 1 - \alpha.$$

For a specific sample with mean $\bar{x}^k$, the above condition could alternatively be written as (cf. equations 3.2 and 3.3)

$$P(\bar{X} \geq \bar{x}^k | \bar{x}_1) = \int_{\bar{x}^k}^{\infty} \mathcal{N}(\bar{x}_1, \sigma^2/n) d\bar{x} = \frac{\alpha}{2}$$

and

$$P(\bar{X} \leq \bar{x}^k | \bar{x}_2) = \int_{-\infty}^{\bar{x}^k} \mathcal{N}(\bar{x}_2, \sigma^2/n) d\bar{x} = \frac{\alpha}{2}.$$

We now define the standardized random variable

$$T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

with probability density function $\mathcal{N}(0, 1)$. Random variables like T , with probability densities that do not depend on any unknown parameters, are also often called *pivotal quantities*. Confidence intervals constructed from pivots yield the desired coverage probability for all possible values of the unknown parameters (see for example Pawitan [51], p. 140). One can

find a central interval which covers T in $(1 - \alpha)100\%$ of all cases:

$$P\left(-t_{(1-\alpha/2)} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq t_{(1-\alpha/2)}\right) = 1 - \alpha.$$

Here $t_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ -percentile of $\mathcal{N}(0, 1)$. We can now reformulate this expression to get

$$P\left(\bar{X} - t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \quad (3.4)$$

Hence, a confidence interval with confidence level $(1 - \alpha)100\%$ can be defined as

$$\left(\bar{X} - t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}\right) \quad (3.5)$$

and a specific realization of such an interval for the mean of a specific sample $\bar{x}^k$ amounts to

$$\left(\bar{x}^k - t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}, \bar{x}^k + t_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}\right). \quad (3.6)$$

We therefore get distinct interval boundaries for each specific sample mean $\bar{x}^k$, however, an exact portion of $(1 - \alpha)100\%$ of these intervals will cover the expected value μ .

Another way of obtaining confidence intervals in the frequentist picture is to center the normal distribution on the measured value x_0 instead of the unknown parameter θ . Thus, we make use of the concept of likelihood as we have already discussed in chapter 2 and—in a nutshell—use $\mathcal{L}(\theta|x_0)$ (instead of $P(X|\theta)$) to compare different values of θ :

$$\mathcal{L}(\theta|x_0) = \frac{1}{\sqrt{2\pi}\sigma_x^2} e^{-(x_0-\theta)^2/2\sigma_x^2}. \quad (3.7)$$

As emphasized in Cousins [14] (p. 399f), there is a key difference between the Bayesian and the frequentist philosophies when it comes to extracting error intervals from the likelihood function: In the frequentist picture approximate confidence intervals can be determined by comparison of different values of the negative log-likelihood function to its minimum (see fig. 3.2). A “one sigma” confidence interval with a confidence level (C.L.) of approximately 68.3 % can for example be achieved by choosing points θ_1, θ_2 where $\ln(\mathcal{L}(\theta|x_0))$ is $1/2$ unit lower than its maximum (or equivalently points that lie one unit above the minimum of $-2\ln\mathcal{L}$, see fig 3.2): For the likelihood ratio

$$\ln \frac{\mathcal{L}(\theta|x_0)}{\mathcal{L}(\theta = x_0|x_0)} = -\frac{1}{2}, \quad (3.8)$$

3. Bayesian Inference

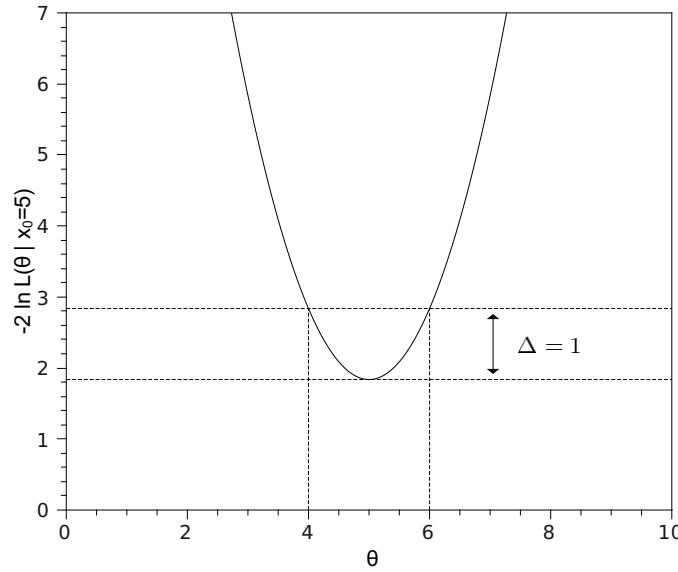


Figure 3.2.: Confidence intervals obtained by increase of $-2 \ln \mathcal{L}$ by $\Delta = 1$. This illustration was redrawn and adapted from Cousins [14], p. 399.

it follows that

$$\begin{aligned} \frac{(x_0 - \theta)^2}{2\sigma_x^2} &= \frac{1}{2} \\ \Rightarrow \theta_{1,2} &= x_0 \pm \sigma_x. \end{aligned}$$

We can see that both the likelihood ratio approach and directly using $P(X|\theta)$ (eqns. 3.2 and 3.3) would lead to the same confidence intervals.

To understand how these regions fulfill the desired property of coverage of the sought-for parameter, one may consider the following explanation: Suppose we are interested in the expected value of X , again assuming $X \sim P(X|\theta)$ with unknown θ . If a measurement would by chance yield the numerical value of the expected value $x_0 = E(X) = \theta$, a 68.3 % confidence level confidence interval derived from a likelihood function constructed around x_0 would be identical to the one sigma region, $\theta \pm \sigma_x$, of this density. If, on the other hand, a measurement would strike a value that is one standard deviation σ_x below or above θ , a 68.3 % C.L. confidence interval would still contain θ as boundary value. All other outcomes x_0 that fall within the region $\theta \pm \sigma_x$ account for 68.3 % of all possible outcomes. Thus, 68.3 % of the 68.3 % C.L. confidence intervals derived in such a manner will cover θ .

It is important to note that, from a frequentist point of view, the likelihood function is a useful tool to determine confidence intervals as long as one does not treat $\mathcal{L}(\theta|x_0)$ as an actual probability density in θ . According to the frequentist philosophy it would be nonsensical to integrate $\mathcal{L}$ and calculate *areas* under this function to gain an estimate for the probability of the parameter θ to fall into this region. This ambivalent nature of $\mathcal{L}$, which is widely used in

both frequentist as well as Bayesian inference, can be a source of confusion.

3.2. The Bayesian Paradigm

One of the most obvious differences between frequentist and Bayesian statistical inference is that in the latter unknown parameters of a statistical model are itself treated as random variables. In Bayesian inference incorporating the *Bayes Theorem* leads us to the construction of a probability density function for unknown parameters. Ultimately, the deduction of unknown parameters from observed quantities can be conducted through the methods discussed in the following (cf. Robert [55], p. 8f.):

Let A and E be events such that $P(E) \neq 0$. From the definition of conditional probability

$$P(A|E) = \frac{P(A \cap E)}{P(E)},$$

$$P(E|A) = \frac{P(A \cap E)}{P(A)}$$

follows

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)}. \quad (3.9)$$

For an arbitrary partition $\{A_i\}$ of the event space one can write

$$P(E) = \sum_j P(E|A_j)P(A_j)$$

and insert it into 3.9, e.g. in the case of a binary partition:

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)}. \quad (3.10)$$

Thomas Bayes himself gave a continuous version of this result, e.g. in the case of two random variables Y and X , with conditional distribution $f(x|y)$ and marginal distribution $g(y)$, one can write the conditional distribution of y given x as (see Bayes [6]³):

$$h(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}.$$

In Bayesian analysis, the uncertainty of the parameters θ of a statistical model prior to a planned measurement can be modeled through a probability distribution π_p on Θ , the so-called *prior distribution*. Within this paradigm of statistical modeling, there is no fundamental difference between the parameters and the observation, since they both are treated as random variables governed by probability distributions. Although no generally applicable

³Cited after Robert [55] p. 9

3. Bayesian Inference

recipe for constructing priors (as well as likelihoods) can be given, it is reasonable to incorporate any information available in advance to an experiment. An obvious starting point would be considerations of physical meaningfulness, e. g. parameter values that would lead to contradictory or nonsensical results if used in the model should be excluded.

In real world problems a statistician will have incomplete or only very general information about unknown parameters, which makes it difficult to propose an exact and unique prior distribution for the problem at hand. It is quite clear that the choice of a prior can significantly change the inference process. Therefore, objections to the Bayesian approach might be justified if available information is ignored for sake of computational simplification. Also, an unjustified use of well-known parameterized distributions (e.g. Gaussian distribution), without proper support from the particular problem at hand, can lead to criticism.

As Robert stresses, the real relevance of the prior distribution lies not in the idea that “the parameter of interest θ can (or cannot) be perceived as generated from π or even as a random variable, but rather [therein] that the use of a prior distribution is the best way to summarize the available information [...], thus allowing for incorporation of this imperfect information in the decision process.” (Robert [55], p. 10). If no prior information is available, however, a *non-informative prior* can be chosen, which is of the form

$$\pi_p(\theta) = \begin{cases} 1 & \text{for } \theta_{\min} \leq \theta \leq \theta_{\max} \\ 0 & \text{else} \end{cases}$$

with arbitrary boundaries $\theta_{\min}, \theta_{\max}$, depending on the subspace of Θ one is interested in. Using the Bayes theorem, a distribution of the parameter random variable Θ (with realizations θ) depending on the data random variable M (with realizations m) can finally be constructed:

$$\pi(\theta|m) = \frac{f(m|\theta)\pi_p(\theta)}{\int f(m|\theta)\pi_p(\theta)d\theta}. \quad (3.11)$$

This function is called the *posterior distribution* and we can identify $f(m|\theta)$ with the likelihood function (see Robert [55], p. 9):

Definition 5 A Bayesian statistical model is comprised of a parametric statistical model, $f(m|\theta)$, and of a probability density of the parameters, $\pi_p(\theta)$, called prior distribution.

The main difference to a pure likelihood model lies therefore in the consideration of prior knowledge about the parameters. The posterior distribution of the unknown quantities is conditional on the observations and is in some sense an update on the information available on θ , based on the information condensed in $f(\theta|m)$, the likelihood.

Just as in our description of the Bayesian concept of probability (see 2.3.3), we now follow Bernardo [7] (p. 9) to more precisely assert the dependence of our statistical model on three crucial types of “information”: Assumptions on the model itself A , prior knowledge on the parameters K and the measured data M .

3.3. Drawing Conclusions from the Posterior

A one-dimensional posterior distribution given the measured point $M = m$, for example, should therefore better be written as

$$\pi(\theta|M = m, A, K) = \frac{f(M = m|\theta)\pi_p(\theta|K)}{\int f(M = m|\theta)\pi_p(\theta|K)d\theta}.$$

It is important to remember that the posterior distribution provides us with a measure for the uncertainty of the parameters, *conditional* on those three types of information. Although this dependence is usually not explicitly stated, one should always keep in mind that inferences based on the posterior density are in the same way conditional on A and K . For sake of a concise notation we will however stick to the form presented in eq. 3.11 within this thesis.

3.3. Drawing Conclusions from the Posterior

Evaluating posterior densities, especially in the case of multidimensional joint posterior functions, might not be trivial and is often not viable using standard integration techniques. In practice, one often encounters the following situation (see Robert, p. 285 [55]):

1. An explicit, analytical expression for the posterior density $\pi(\theta|\mathbf{m})$ is not known and the numerical calculation of the expression $\int f(\mathbf{m}|\theta)\pi_p(\theta)d\theta$ proves to be difficult, especially for higher dimensional parameter spaces.
2. If an analytical form of $\pi(\theta|\mathbf{m})$ is not given, the numerical computation of global maxima and minima of this function can be difficult or even impossible.

To perform some kind of statistical inference on the parameters of interest, for example in form of point estimates, it is however often desirable to find at least local maxima of $\pi(\theta|\mathbf{m})$. For Gelman et al., the main advantage that the Bayesian frame of reference has over other methods of inference lies in its treatment of statistical models with multiple unknown parameters. Even if the model includes a large number of parameters, it is a common practice to draw conclusions via statistical inference about only a small subset of all parameters at the same time (Gelman et al. [23], p. 75). In case of interval estimations, regions of high posterior probability are often calculated for each parameter of interest (see section 3.3.3). Within the Bayesian paradigm, this can naturally be achieved with help of marginal posterior densities.

3.3.1. Marginal Posterior Density

Given a posterior of a multitude of parameters, it is possible to look at the so-called *marginal* posterior distribution of the particular parameter(s) an experimenter is most interested in. Unknown parameters that are necessary to construct a model, but are generally not the aim of any kind of statistical inference, are also known as *nuisance parameters*.

We get the marginal posterior density of a specific parameter when we “average” (integrate)

3. Bayesian Inference

over all the other parameters:

$$\pi_{\theta_i}(\theta_i|\mathbf{m}) = \int_{\theta_{\min}}^{\theta_{\max}} d\theta_1 d\theta_2 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_n \pi(\boldsymbol{\theta}|\mathbf{m}). \quad (3.12)$$

Consider for example a two-dimensional posterior (see Gelman et al. [23], p. 76)

$$\pi(\theta_1, \theta_2|\mathbf{m}) \propto f(\mathbf{m}|\theta_1, \theta_2)\pi_p(\theta_1, \theta_2).$$

Averaging over θ_2 yields

$$\pi_{\theta_1}(\theta_1|\mathbf{m}) = \int \pi(\theta_1, \theta_2|\mathbf{m}) d\theta_2.$$

Alternatively, the joint posterior can be factored to get

$$\pi_{\theta_1}(\theta_1|\mathbf{m}) = \int \tilde{\pi}(\theta_1|\theta_2, \mathbf{m}) \pi_{\theta_2}(\theta_2|\mathbf{m}) d\theta_2 \quad (3.13)$$

with

$$\tilde{\pi}(\theta_1 | \theta_2, \mathbf{m}) = \frac{\pi(\theta_1, \theta_2|\mathbf{m})}{\pi_{\theta_2}(\theta_2|\mathbf{m})}. \quad (3.14)$$

As suggested in Gelman et al. [23], $\pi_{\theta_1}(\theta_1|\mathbf{m})$ can therefore also be seen as weighted average of the conditional posterior distributions (eq. 3.14) given the different possible values of θ_2 , where the weights $\pi_{\theta_2}(\theta_2|\mathbf{m})$ are obtained from the marginal posterior of θ_2 .

3.3.2. Marginal Maximum A Posteriori Estimate

In this thesis we will primarily base statistical inference on one-dimensional marginal posterior densities for each parameter of interest. More specifically we will make use of marginal maximum a posteriori (MMAP) estimates for the marginal posterior distributions $\pi_{\theta_k}(\theta_k|\mathbf{m})$ of each parameter separately (Doucet et al. [20], p. 77):

$$\theta_k^{MMAP} = \arg \max_{\Theta} \pi_{\theta_k}(\theta_k|\mathbf{m}). \quad (3.15)$$

Assuming that an explicit, analytical form of the joint posterior density is not known, one possible strategy to perform such a maximum a posteriori estimation relies on the ability to generate sample vectors that are distributed according to $\pi(\boldsymbol{\theta}|\mathbf{m})$:

The general idea is to draw samples from the joint posterior (e.g. one parameter-vector $\boldsymbol{\theta}_i$ at a time) and notate the frequency of occurrence of the parameter of interest for a large number of points in a defined interval, while ignoring the values of the other unknowns. The resulting histograms approximate the particular marginal posterior densities in the examined regions. The mode of such a histogram can therefore be used as an empirical approximation of the MMAP. Point estimates, such as the MMAP, however, can be mis-representative for a given marginal posterior distribution if they are isolated from the majority of the “probability mass”

of the density. In such cases, interval estimators such as Bayesian credible regions can be favoured.

3.3.3. Credible Intervals

Credible intervals are the Bayesian analogs to the frequentist confidence intervals we discussed before. In contrast to the frequentist concept of coverage, where we define intervals in a way that a fraction of $(1 - \alpha)100\%$ of them contains the parameter, in this picture a parameter θ actually has a defined probability of $(1 - \alpha)$ to be found in a fixed region (see Robert, p. 260 [55]).

In this work we will focus on credible intervals of scalar parameters θ_i derived from one-dimensional marginal posterior densities $\pi_{\theta_i}(\theta_i|\mathbf{m})$. Let us look at a specific parameter of interest which we will denote by $\theta_i := \gamma$.

A Bayesian $100(1 - \alpha)\%$ credible interval (for $0 \leq \alpha \leq 1$) can simply be defined as the region where $100(1 - \alpha)\%$ of the parameter “population” is located. We can obtain such a region (γ_1, γ_2) by integrating the posterior:

$$\int_{\gamma_1}^{\gamma_2} \pi_{\gamma}(\gamma|\mathbf{m})d\gamma = 1 - \alpha. \quad (3.16)$$

There are various ways to satisfy this condition. Analogous to the frequentist picture we can define *central* intervals as

$$\int_{-\infty}^{\gamma_1} \pi_{\gamma}(\gamma|\mathbf{m})d\gamma = \int_{\gamma_2}^{\infty} \pi_{\gamma}(\gamma|\mathbf{m})d\gamma = \frac{\alpha}{2}. \quad (3.17)$$

Highest Density Regions

Alternatively, Bayesian minimum length intervals (also called highest posterior density intervals or regions in more dimensions) may be defined: It has been argued that it seems to be a rational choice to postulate intervals which comply with the following two properties, especially if the posterior density in question is not symmetrical (See Hyndman [32], p. 120):

1. The numerical value of $\pi_{\gamma}(\gamma_{\text{out}}|\mathbf{m})$ for any γ_{out} outside of the interval is equal or smaller than $\pi_{\gamma}(\gamma_{\text{in}}|\mathbf{m})$ for all γ_{in} inside of the interval
2. The interval (or region) takes up the shortest possible length (or volume) for a given probability level $(1 - \alpha)$

As stressed by Hyndman, referring to the discussion of these criteria by Box and Tiao [8], both properties are equivalent and can be further specified by the following definition of highest density regions: Given a random variable X with probability density function $g(x)$, Hyndman defines a $100(1 - \alpha)\%$ highest density interval $R(g_{\alpha})$ as a subset of the sample space of X so that

$$R(g_{\alpha}) = \{x : g(x) \geq g_{\alpha}\}, \quad (3.18)$$

3. Bayesian Inference

where g_α is defined as the largest possible constant for which

$$P(X \in R(g_\alpha)) \geq 1 - \alpha.$$

If $g(x)$ is a normal distribution (or some other unimodal, symmetric probability distribution), this definition gives an symmetric interval around the mean, located between the $\alpha/2$ and $1 - \alpha/2$ quantiles. In contrast to that, in the case of multimodal densities, a highest density interval can even include a number of disjoint regions of parameter space (see fig. 3.3).

Hyndman provides a method to calculate such highest posterior regions which tends to be less computationally expensive than any method relying on standard numerical integration procedures (at least for higher-dimensional regions):

The so-called *density quantile approach* (Hyndman [32], p. 124). Given a random variable X with density $g(x)$, one can create a new random variable Y by applying $g(x)$ to X : $Y := g(X)$. Then g_α is the α -quantile of Y , since

$$P(g(X) \geq g_\alpha) = 1 - \alpha.$$

As discussed by Hyndman, we can estimate g_α from the sample quantiles of a set of independently conducted realizations of Y :

Assume $\{x_1, \dots, x_n\}$ represents a set of independent realizations from $g(x)$. From that we can generate a set of independent observations of Y : $\{g(x_1), \dots, g(x_n)\}$. We denote the j th smallest (largest) member of $\{g(x_i)\}$ by $g_{(j)}$ which is the (j/n) sample quantile of Y . Setting $\hat{g}_\alpha = g_{(j)}$, where $j = \lfloor \alpha n \rfloor$, we have gained an estimator of g_α :

$$\begin{aligned} \hat{g}_\alpha &\rightarrow g_\alpha \text{ for } n \rightarrow \infty \\ R(\hat{g}_\alpha) &\rightarrow R(g_\alpha) \text{ for } n \rightarrow \infty. \end{aligned}$$

Hereby, the brackets $\lfloor \cdot \rfloor$ denote the floor function (also known as Gaussian bracket), i.e. the largest preceding integer of a given real number.

In many practical problems an explicit, analytical expression for the density $g(x)$ is not given, a large number of realizations (or measurements), however, is available. As a consequence, methods of deriving empirical highest posterior density regions from (ideally) i.i.d. observations of $g(x)$ have been developed.

Chen and Shao [12], for example, proposed a method of computing one-dimensional highest marginal posterior density intervals which uses samples from the posterior density generated by means of a MCMC simulation procedure (see chapter 4). Given an unimodal marginal posterior density $\pi_\xi(\xi|\mathbf{m})$ one can formulate it the following way:

Let $\{\xi_1, \dots, \xi_n\}$ be a sample of realizations produced by a long-running MCMC sampling process so that the statistical properties of $\pi_\xi(\xi|\mathbf{m})$ are represented reasonably well by this sample. Sorting the set $\{\xi_i\}$ by size so that we reassign the index 1 to the smallest member, the index 2 to the second smallest member and so forth yields

$$\xi_{(1)} < \xi_{(2)} < \dots < \xi_{(\lfloor (\alpha/2)n \rfloor)} < \dots < \xi_{(\lfloor (1-\alpha/2)n \rfloor)} < \dots < \xi_{(n-1)} < \xi_{(n)}. \quad (3.19)$$

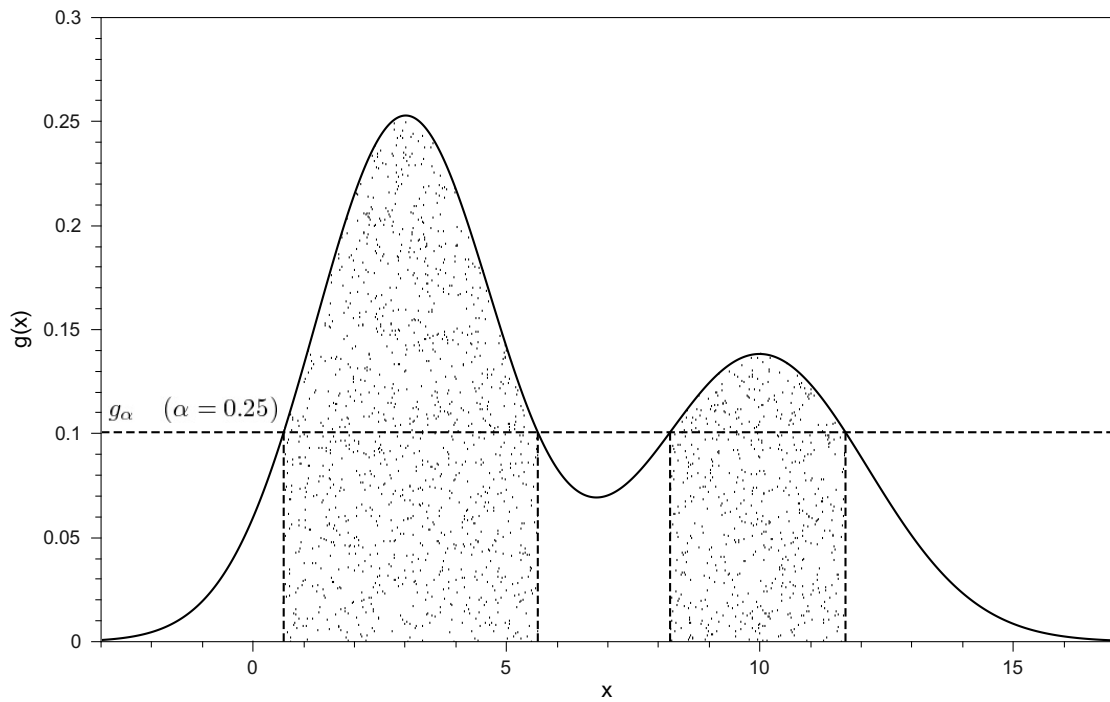


Figure 3.3.: Density Quantile Approach: The 75 % Highest density region is indicated by the dotted areas. A sample of size $n=500000$ was generated by an acceptance-rejection sampling procedure. Subsequently, an estimation for $g_{0.25}$ was obtained from this sample. This Illustration was redrawn and adapted from Hyndman [32], p. 120, Figure 1.

3. Bayesian Inference

Thus, we can see that an interval defined with help of the $\lfloor (\alpha/2)n \rfloor$ th and $\lfloor (1 - \alpha/2)n \rfloor$ th smallest members of $\{\xi_i\}$, namely

$$\left(\xi_{(\lfloor (\alpha/2)n \rfloor)}, \xi_{(\lfloor (1-\alpha/2)n \rfloor)} \right),$$

approximates a $100(1 - \alpha)\%$ Bayesian credible interval since $\xi_{(\lfloor (\alpha/2)n \rfloor)}$ and $\xi_{(\lfloor (1-\alpha/2)n \rfloor)}$ enclose a proportion of

$$\frac{(1 - \alpha/2 - \alpha/2)n}{n} = 1 - \alpha$$

of the sample $\{\xi_i\}$ (see Chen and Shao [12], p. 73f).

Along this line of reasoning, if one denotes (ξ_k) as the k th smallest member of $\{\xi_i\}$,

$$R_k(n) = (\xi_{(k)}, \xi_{(k+\lfloor (1-\alpha)n \rfloor)}) \quad (3.20)$$

determines $100(1 - \alpha)\%$ Bayesian credible intervals for $k = 1, 2, \dots, n - \lfloor (1 - \alpha)n \rfloor$.

To obtain an empirical highest posterior density interval, one can therefore choose the interval of minimal width from all possible $R_k(n)$, denoted by the index k^* :

$$\xi_{(k^*+\lfloor (1-\alpha)n \rfloor)} - \xi_{(k^*)} = \min_{1 \leq k \leq n - \lfloor (1-\alpha)n \rfloor} (\xi_{(k+\lfloor (1-\alpha)n \rfloor)} - \xi_{(k)}). \quad (3.21)$$

According to Chen and Shao (p. 73), for a k^* selected in such manner and

$$R_{k^*}(n) = (\xi_{(k^*)}, \xi_{(k^*+\lfloor (1-\alpha)n \rfloor)}),$$

it follows that

$$R_{k^*}(n) \rightarrow R(g_\alpha) \text{ a.s. as } n \rightarrow \infty,$$

for a unimodal marginal posterior density $\pi_\phi(\phi|\mathbf{m})$.

4. Markov Chain Monte Carlo Methods

In most cases it is difficult to directly draw samples from the posterior distribution $\pi(\theta|\mathbf{M})$. One can, however, use a generally applicable technique to sample from the posterior, developed by Metropolis et al. [45]. The idea behind this algorithm is to generate a Markov-chain whose unique stationary distribution is equal to $\pi(\theta|\mathbf{M})$. As a result, the distribution of realized states of this stochastic process approaches the desired posterior distribution in the long run. How this can be achieved will in the following be shown by a short introduction and definition of discrete Markov chains, whereas we complete the section with a derivation of the Metropolis-Hastings algorithm, taken in this concise form from the original publication by W. K. Hastings [30].

To define discrete-time Markov chains we follow Norris [50] (pp. 1–5) as well as Karlin and Taylor [34] (pp. 45–47). We recall that we operate in a probability space $(\Omega, \mathcal{F}, P)$ (see section 2). Furthermore, let S be a countable state space and X_i random variables that take values in S :

$$X_i : \mathcal{F} \mapsto S.$$

Discrete probability distributions $\rho : \mathcal{F} \mapsto S$ will be written as

$$\rho_s = P(\{\omega \in \Omega : X_i(\omega) = s \in S\}) \quad \text{if } \sum_{s \in S} \rho_s = 1. \quad (4.1)$$

Definition 6 *A stochastic process in form of a sequence of random variables $\{X_n\}_{n \geq 0}$ on a state space S is called a discrete time Markov chain with initial distribution ρ if*

1. $P(X_0 = s_0) = \rho_{s_0}$
2. $P(X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = P(X_n = s_n | X_{n-1} = s_{n-1})$

with $s_n \in S$ denoting a realization of a random variable at time n .

We abbreviate the conditional probability of $X_n = s_j$ given $X_{n-1} = s_i$ by

$$p_{ij}^{n,n-1} = P(X_n = s_j | X_{n-1} = s_i).$$

4. Markov Chain Monte Carlo Methods

These so-called *transition probabilities* are generally not only dependent on the specific states s_i and s_j but also on the time n at which the transition occurs. If the transition probabilities do not depend on time, e.g. if

$$P(X_{n+1} = s_j | X_n = s_i) = P(X_n = s_j | X_{n-1} = s_i),$$

we speak of *stationary* or *time-homogeneous* Markov chains. For these conditional probabilities collected in a transition matrix $p_{ij} := P(X_n = s_j | X_{n-1} = s_i)$ the following is true:

$$\begin{aligned} 0 &\leq p_{ij} \leq 1, \\ \sum_j p_{ij} &= 1. \end{aligned} \tag{4.2}$$

A discrete-time, stationary Markov process is completely determined through conditions (1) and (2) of definition 6. To show this we follow Karlin and Taylor [34] and calculate the probability of realization of a chain of length n (see Karlin and Taylor [34], pp. 46–47):

$$\begin{aligned} P(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) \\ = P(X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_1 = s_1, X_0 = s_0) \\ \cdot P(X_0 = s_0, X_1 = s_1, \dots, X_{n-1} = s_{n-1}). \end{aligned} \tag{4.3}$$

From condition (2) of def. 6 follows:

$$P(X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = P(X_n = s_n | X_{n-1} = s_{n-1}) = p_{n-1,n}. \tag{4.4}$$

Combining 4.3 and 4.4 yields

$$P(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = p_{n-1,n} P(X_0 = s_0, X_1 = s_1, \dots, X_{n-1} = s_{n-1}). \tag{4.5}$$

Through complete induction we finally get

$$P(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = p_{n-1,n} p_{n-2,n-1} \dots p_{0,1} \rho_{s_0}. \tag{4.6}$$

The *total* probability for the system to be in a specific state s_i at time n , can be expressed as a sum over all probabilities for the system to be in some other state s_j at time $(n - 1)$ AND in state s_i at time n

$$\begin{aligned} P(X_n = s_i) &= \sum_j P(\{X_n = s_i\} \cap \{X_{n-1} = s_j\}) \\ &= \sum_j P(X_n = s_i | X_{n-1} = s_j) P(X_{n-1} = s_j). \end{aligned} \tag{4.7}$$

The main idea of the Metropolis Monte Carlo method is to construct a Markov chain with time-invariant transition probabilities p_{ij} in such a way, that after a certain initial phase, the resulting absolute probabilities $P(X_n = s_i)$ themselves become time-independent and equal to a set of given probabilities v_i . For systems with a discrete set of possible states, these would be the probabilities to take on each of those states. Analogously, in the continuous case, one would hope to create a sample from a probability density $\pi(x)$ in such a way. Rewriting Equation 4.7 one can derive an expression for the change of $P(X_n = s_i)$ as time progresses (see for example Penrose [52], pp. 63–64):

$$\begin{aligned} P(X_n = s_i) &= \sum_{j \neq i} P(X_{n-1} = s_j) p_{ji} + P(X_{n-1} = s_i) p_{ii} \\ &= \sum_{j \neq i} P(X_{n-1} = s_j) p_{ji} + P(X_{n-1} = s_i) \\ &\quad - \sum_{j \neq i} P(X_{n-1} = s_i) p_{ij} \end{aligned} \quad (4.8)$$

where the identity $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$ was used. It follows that

$$P(X_n = s_i) - P(X_{n-1} = s_i) = \sum_{j \neq i} P(X_{n-1} = s_j) p_{ji} - \sum_{j \neq i} P(X_{n-1} = s_i) p_{ij}. \quad (4.9)$$

Adding $P(X_{n-1} = s_i) p_{ii}$ to both sums leads to

$$P(X_n = s_i) - P(X_{n-1} = s_i) = \sum_j P(X_{n-1} = s_j) p_{ji} - \sum_j P(X_{n-1} = s_i) p_{ij}. \quad (4.10)$$

Equation 4.10 is a *master equation* as it describes the time evolution of the probabilities of the states the system can be in. For time-invariant absolute probabilities $P(X = s_i) = v_i$ it follows that

$$\sum_j (v_j p_{ji} - v_i p_{ij}) = 0 \quad (4.11)$$

and we call those v_i the unique stationary distribution of the Markov chain. One way to meet this condition is to require that

$$v_j p_{ji} = v_i p_{ij}. \quad (4.12)$$

This is called the *detailed balance* condition. It means that the probability to be in state j AND advance to state i must be just the same as being in state i AND continuing to j .

4. Markov Chain Monte Carlo Methods

Consider now the given discrete probability distribution

$$\boldsymbol{\pi} = \{\pi_0, \pi_1, \dots, \pi_k\}, \quad \pi_i > 0 \quad \forall i. \quad (4.13)$$

Following Hastings [30] (p. 99), the task is now to construct a Markov chain with transition matrix $P = \{p_{ij}\}$ in a way that $\boldsymbol{\pi}$ is its unique stationary distribution, in other words that

$$\sum_i \pi_i p_{ij} = \pi_j,$$

ensured by the detailed balance condition 4.12. Generally such a p_{ij} is not easy to find, but Hastings makes the following ansatz:

$$\begin{aligned} p_{ij} &= q_{ij} \alpha_{ij} \quad (i \neq j) \\ p_{ii} &= 1 - \sum_{i \neq j} p_{ij}. \end{aligned} \quad (4.14)$$

The matrix $Q = \{q_{ij}\}$ is a transition matrix of an arbitrary Markov chain on the states $0, 1, 2, \dots, k$ and

$$\alpha_{ij} = \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}}, \quad (4.15)$$

where the function s_{ij} is symmetric in i and j and chosen in a way to guarantee $0 \leq \alpha_{ij} \leq 1$. Inserting ansatz 4.14 into the detailed balance condition 4.12 leads to

$$\pi_i \alpha_{ij} q_{ij} = \pi_j \alpha_{ji} q_{ji} \quad (4.16)$$

and from that we obtain the ratio

$$\frac{\alpha_{ij}}{\alpha_{ji}} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, \quad (4.17)$$

which we want to hold at all times. The proposal or trial matrix q_{ij} stands for the conditional probability to draw j given i while α_{ij} can be interpreted as the probability to accept this trial move. Thus, through the substitution 4.14 we split up the transition kernel p_{ij} into two statistically independent probability densities, a proposal density (preferably easy to create samples from) and a “correction factor” α_{ij} .

Hastings mentions two possible choices for s_{ij} , the first of which was proposed by Metropolis et al. [45] (marked by an M), and the second one was suggested by Barker [4] (marked by a B):

$$\begin{aligned} s_{ij}^{(M)} &= \begin{cases} 1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}} & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \geq 1 \right), \\ 1 + \frac{\pi_j q_{ji}}{\pi_i q_{ij}} & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \leq 1 \right), \end{cases} \\ s_{ij}^{(B)} &= 1. \end{aligned}$$

To satisfy condition 4.17 and specify α_{ij} , we want to focus on the first possibility, which leads to an asymmetric α_{ij} :

$$\alpha_{ij} = \begin{cases} 1 & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \geq 1 \right) \\ \frac{\pi_j q_{ji}}{\pi_i q_{ij}} & \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} < 1 \right) \end{cases}. \quad (4.18)$$

From this arbitrary definition of α_{ij} follows an acceptance probability of 1 for a move $i \rightarrow j$, if $\pi_j q_{ji} \geq \pi_i q_{ij}$.

As Chib and Greenberg [13] have put it (p. 328), one might say that if q_{ij} by itself does not satisfy the detailed balance condition 4.12 but we rather find that $\pi_i q_{ij} > \pi_j q_{ji}$ for some i, j , the process moves more often from $i \rightarrow j$ than vice versa. We therefore compensate this “distortion” of equilibrium by introducing a probability

$$\alpha_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}} < 1$$

that this move is made, while a step from $j \rightarrow i$ is accepted with the highest possible probability $\alpha_{ji} = 1$. For $\pi_j q_{ji} > \pi_i q_{ij}$ one can argue analogously by setting $\alpha_{ij} = 1$ and deriving the correct expression for α_{ji} from the detailed balance condition 4.16.

In the case of a symmetric proposal density $q_{ij} = q_{ji}$, the expression for α_{ij} simplifies to

$$\alpha_{ij} = \begin{cases} 1 & \left(\frac{\pi_j}{\pi_i} \geq 1 \right) \\ \frac{\pi_j}{\pi_i} & \left(\frac{\pi_j}{\pi_i} < 1 \right) \end{cases}. \quad (4.19)$$

We can see that if step $i \rightarrow j$ leads to a state of equal or higher probability it will always be accepted. If, on the other hand, $\pi_j < \pi_i$ the move will only be accepted with probability π_j/π_i . It is evident that n moves from $i \rightarrow j$ will be accepted n times if $\pi_j > \pi_i$, while of n moves from $j \rightarrow i$ only a proportion of $n(\pi_i/\pi_j)$ will be accepted¹. Hence, the ratio of moves arriving at j to moves ending in i amounts to $n/n(\pi_i/\pi_j) = \pi_j/\pi_i$, correctly depicting the ratio of probabilities of those two points.

We sum up the Metropolis-Hastings algorithm in the form discussed by Hastings [30] (p. 100):

1. Given state $X(t) = i$, select a state j according to the probability distribution given by the i -th row of the matrix Q .
2. Accept $X(t+1) = j$ with probability

$$\alpha_{ij} = \min \left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right),$$

else set $X(t+1) = i$ with probability $1 - \alpha_{ij}$.

¹The acceptance probability $\alpha_{ji} = \pi_i/\pi_j$ follows from the detailed balance condition 4.16

4. Markov Chain Monte Carlo Methods

The involved probabilities α_{ij} and q_{ij} can also be defined as continuous probability densities on a continuous sample space². According to Hastings [30] (pp. 100–101), in case of a continuous state space with continuous probability density π we may substitute the discrete π_i, P and Q for $\pi(x)d\mu(x)$, $p(x, x')d\mu(x)$ and $q(x, x')d\mu(x)$. Here $d\mu(x)$ is the Lebesgue measure of a real and continuous random variable X , often also simply written as dx . The density π is in that case the density of random variable X with respect to the reference measure μ on a measure space $(\mathbb{R}^n, \mathcal{A})$, where $\mathcal{A}$ is a Borel set. For the probability that X is part of a subset $A \in \mathcal{A}$ follows (cf. chapter 2):

$$P[X \in A] = \int_A \pi d\mu.$$

A representation of x or x' on the computer, however, is a finite set of values $x_0, x_1, \dots, x_s$. As discussed by Hastings, we can therefore approximate the above probability elements by the discrete quantities $\pi(x_i)\delta\mu(x_i)$, $p(x_i, x_j)\delta\mu(x_j)$ and $q(x_i, x_j)\delta\mu(x_j)$, respectively. It follows that if we replace π_i by $\pi(x_i)\delta\mu(x_i)$ etc., we get

$$p(x_i, x_j) = q(x_i, x_j) \alpha_{ij}$$

and $(\pi_i q_{ij})/(\pi_j q_{ji})$ can be substituted by

$$\frac{\pi(x_i)q(x_i, x_j)}{\pi(x_j)q(x_j, x_i)}.$$

Keeping this approximation in mind, we may in the following center on a description in terms of continuous probability densities and write $\alpha(x, x')$, $\pi(x)$ etc. for the respective functions. We conclude with summing up the symmetric Metropolis-Hastings Algorithm, which will primarily be used for the rest of this thesis.

²See for example Calvetti et al. [11] for a continuous description.

4.1. Metropolis Algorithm

Given the posterior density $\pi(\theta|\mathbf{m})$ (known up to a normalizing factor) and a symmetric conditional proposal density $q(\theta'|\theta)$ such a Markov chain is generated by:

1. Choose an arbitrary initial value $\theta_{m=0}$
2. For $m=0,1,2, \dots$
 - a) Sample a new proposal θ^* from the proposal distribution $q(\theta^*|\theta_m)$
 - b) Calculate the ratio of the posteriors $\frac{\pi_{new}}{\pi_{old}}$:

$$\rho = \frac{\pi(\theta^*|\mathbf{m})}{\pi(\theta_m|\mathbf{m})}$$

- c) Set

$$\theta_{m+1} = \begin{cases} \theta^* & \text{with probability } \alpha = \min(\rho, 1) \\ \theta_m & \text{with probability } 1 - \alpha \end{cases}$$

In practice:

If $\pi(\theta^*|\mathbf{m}) \geq \pi(\theta_m|\mathbf{m})$, accept and set $\theta_{m+1} = \theta^*$.

Else draw a uniformly distributed random number $\xi \in [0, 1)$ and set

$$\begin{aligned} \theta_{m+1} &= \theta^* & \text{if } \rho \geq \xi \\ \theta_{m+1} &= \theta_m & \text{if } \rho < \xi. \end{aligned}$$

Given a sufficiently long simulation time we expect the generated set of θ -vectors to be approximately distributed according to the target density. After discarding a certain number of draws generated during the initial phase of the algorithm we thus regard the remaining set $\{\theta_m\}$ as a sample generated from $\pi(\theta|\mathbf{m})$.

4.2. Delayed Rejection

Various mechanisms have been proposed in order to improve the efficiency of the Metropolis algorithm. If newly proposed sample candidates are repeatedly being rejected over a longer period of time, the resulting Markov chain is caught in the same position for too long. As a result sufficient exploration of the state space can therefore require an unfeasible amount of simulation time. We briefly want to introduce an algorithm which aims to increase efficiency of the traditional Metropolis-Hastings procedure in a certain sense and follow the description

4. Markov Chain Monte Carlo Methods

given in Haario et al. [28] (p. 340). If we want to estimate the expectation $E_\pi(g)$ of a function g under a probability distribution π , for example, we can use the estimator

$$\hat{E}_\pi(g) = \frac{1}{N} \sum_{k=1}^N g(X_k), \quad (4.20)$$

where $\{X_k\}$ is a Markov chain with stationary distribution π (see Haario et al. [28], p. 339). Haario et al. discuss a modification to the Metropolis-Hastings algorithm, developed by Tierney and Mira [65], Green and Mira [25] and Mira [46, 47]³, called *delayed rejection algorithm*, which can reduce asymptotic variance in the sense described by Peskun [54] for those types of estimators. Peskun hereby refers to asymptotic variance v of the estimate $\hat{E}_\pi(g)$ (p. 608) as it is defined through

$$v(g, \pi, P) = \lim_{N \rightarrow \infty} \text{var} \left(\frac{1}{N} \sum_{k=1}^N g(X_k) \right),$$

where P is the transition matrix of the Markov chain.

When using the delayed rejection algorithm, we do not keep the position and progress time when a newly proposed candidate is rejected, instead we rather propose a second candidate. The acceptance probability of this newly proposed move, however, has to be altered to ensure that the desired distribution $\pi(x)$ is the unique stationary distribution of the resulting Markov chain in the long run. In case of further rejection of the second proposed candidate, the dice can be rolled a third time and so on. If none of the new proposal moves has been accepted after a predetermined number of trials, the delaying procedure can be aborted and we keep the original candidate.

In such a way a single iteration can be split into several different stages and the proposal of a new candidate can be based on previously rejected ones within this iteration.

We follow the description in Haario et al. (p. 340) and suppose we have an arbitrary target distribution $\pi(x)$ and the Markov chain is at position $X_t = x$ at time t . A new candidate y_1 is generated from $q_1(x, y_1)$. The probability of acceptance is

$$\alpha_1(x, y_1) = \min \left(1, \frac{\pi(y_1)q_1(y_1, x)}{\pi(x)q_1(x, y_1)} \right) = \min \left(1, \frac{N_1}{D_1} \right). \quad (4.21)$$

If, after the rejection of y_1 , we want to construct a better candidate y_2 via a new proposal $q_2(x, y_1, y_2)$ while upholding the stationary distribution of our chain, one way to do so is to require detailed balance individually for each stage within one iteration. From that one can then obtain an expression for an acceptance probability that complies with this newly derived detailed balance condition. One can begin with asserting that the probability of a rejection of the move $x \rightarrow y_1$ can be written as $P(x \text{ rejected}) = 1 - \alpha_1(x, y_1)$. After rejection, a new candidate y_2 is drawn from a proposal distribution $q(x, y_1, y_2)$. Note that, in all generality, the new proposal distribution is allowed to depend on all previously rejected moves as well

³all four cited after Haario et al. [28], p. 339

4.2. Delayed Rejection

as on x . One can now construct the detailed balance condition for the case of one rejected move y_1 and a subsequent newly proposed move y_2 : The detailed balance condition in this case takes on the following form

$P[\text{starting from } x]$ $\times$ $P[y_1 \text{ is drawn from } q_1]$ $\times$ $P[y_1 \text{ is rejected}]$ $\times$ $P[y_2 \text{ is drawn from } q_2]$ $\times$ $P[y_2 \text{ is accepted}]$	$=$	$P[\text{starting from } y_2]$ $\times$ $P[y_1 \text{ is drawn from } q_1]$ $\times$ $P[y_1 \text{ is rejected}]$ $\times$ $P[x \text{ is drawn from } q_2]$ $\times$ $P[x \text{ is accepted}]$
--	-----	--

Hence, we have

$$\begin{aligned} & \pi(x)q_1(x, y_1)[1 - \alpha_1(x, y_1)]q_2(x, y_1, y_2)\alpha_2(x, y_1, y_2) \\ &= \pi(y_2)q_1(y_2, y_1)[1 - \alpha_1(y_2, y_1)]q_2(y_2, y_1, x)\alpha_2(y_2, y_1, x), \end{aligned}$$

leading to the ratio

$$\frac{\alpha_2(x, y_1, y_2)}{\alpha_2(y_2, y_1, x)} = \frac{\pi(y_2)q_1(y_2, y_1)q_2(y_2, y_1, x)[1 - \alpha_1(y_2, y_1)]}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2)[1 - \alpha_1(x, y_1)]}. \quad (4.22)$$

Analogously to the standard Metropolis-Hastings procedure one can therefore define an acceptance probability for y_2 (Haario et al. [28], p. 340):

$$\begin{aligned} \alpha_2(x, y_1, y_2) &= \min \left(1, \frac{\pi(y_2)q_1(y_2, y_1)q_2(y_2, y_1, x)[1 - \alpha_1(y_2, y_1)]}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2)[1 - \alpha_1(x, y_1)]} \right), \\ \alpha_2(x, y_1, y_2) &= \min \left(1, \frac{N_2}{D_2} \right). \end{aligned} \quad (4.23)$$

From the rejection of the first proposal y_1 follows $\alpha_1(x, y_1) = \frac{N_1}{D_1}$ with $N_1 < D_1$. Inserting this into 4.23 leads to

$$\begin{aligned} \alpha_2(x, y_1, y_2) &= \min \left(1, \frac{N_2}{q_2(x, y_1, y_2)[\pi(x)q_1(x, y_1) - \pi(y_1)q_1(y_1, x)]} \right) \\ &= \min \left(1, \frac{N_2}{q_2(x, y_1, y_2)[D_1 - N_1]} \right). \end{aligned} \quad (4.24)$$

4. Markov Chain Monte Carlo Methods

Following the argumentation path above, one can derive the i -th stage of the delaying algorithm (Mira⁴ [46]):

$$\begin{aligned}\alpha_i(x, y_1, \dots, y_i) &= \min \left(1, \frac{\pi(y_i)q_1(y_i, y_{i-1})q_2(y_i, y_{i-1}, y_{i-2})\dots q_i(y_i, y_{i-1}, \dots, x)}{\pi(x)q_1(x, y_1)q_2(x, y_1, y_2)\dots q_i(x, y_1, \dots, y_i)} \right. \\ &\quad \left. \frac{[1 - \alpha_1(y_i, y_{i-1})][1 - \alpha_2(y_i, y_{i-1}, y_{i-2})]\dots[1 - \alpha_{i-1}(y_i, \dots, y_1)]}{[1 - \alpha_1(x, y_1)][1 - \alpha_1(x, y_1, y_2)]\dots[1 - \alpha_1(x, y_1, \dots, y_{i-1})]} \right) \\ &= \min \left(1, \frac{N_i}{D_i} \right).\end{aligned}\tag{4.25}$$

Since reversibility is retained, the process can be interrupted at any stage. Having reached the i -th stage means that $N_j < D_j$ and $\alpha_j(x, y_1, \dots, y_j) = N_j/D_j$ for $j = 1, \dots, i - 1$. Thus, a recursive formula for D_i can be derived:

$$\begin{aligned}D_i &= q_i(x, \dots, y_i)[q_{i-1}(x, \dots, y_{i-1})][q_{i-2}(x, \dots, y_{i-2})\dots \\ &\quad [q_2(x, y_1, y_2)[q_1(x, y_1)\pi(x) - N_1] - N_2] - N_3]\dots - N_{i-1}].\end{aligned}\tag{4.26}$$

According to Haario et al. (p. 340), we thus gain a procedure that reduces asymptotic variance of MCMC estimators such as defined in equation 4.20, a proof of which can be found in Tierney and Mira [65]. We can conclude that it is possible to extend and improve on the relatively versatile and generally applicable Metropolis-Hastings algorithm when specific quantities need to be calculated in an efficient way.

⁴cited after Haario et al. [28], p. 340

4.3. Convergence Assessment

In order to make use of the output of a long-running MCMC Simulation procedure one is in need of some kind of diagnostic tool that helps to decide whether resulting Markov chain has converged to its stationary distribution or not. In other words, what criteria can we use to justify the treatment of our simulation draws as an approximately identical and independently distributed sample generated from a given posterior density?

According to Gelman et al. [23] (p. 301), two main problems stand out: Early iterations will be influenced by the arbitrary initial distribution of the Markov chain and generally some time will pass, until the algorithm “forgets” an ill-chosen starting point and starts sampling from the actual target distribution (burn-in time).

Thus, not excluding the burn-in period will lead to a distortion of calculated quantities, e.g. the mean or variance of the marginal density of a specific parameter. Secondly, when using MCMC methods there will always be within-sequence correlations. Statistical inference from correlated samples is in general less accurate than from independently drawn samples of the same sample size. At convergence of the chain, however, the simulation draws are identically distributed according to $\pi(\boldsymbol{\theta}|\mathbf{m})$ and for inference on parameters the order of the draws is not important. Since high correlation between the iterations results from slow movement of the random walk procedure, leaving vast areas of $\pi(\boldsymbol{\theta}|\mathbf{m})$ “unexplored” for a long time, it is however an efficiency issue.

In a comparative review of a number of popular techniques of convergence assessment by Cowles and Carlin [15], one basic problem of every method applying convergence diagnostics to the empirical output of MCMC methods is discussed: At least in cases where the stationary distribution of the MCMC sampling procedure (our posterior $\pi(\boldsymbol{\theta}|\mathbf{m})$) is not known a priori, there is no way to prove convergence has been reached at a given number of iterations from the output of the MCMC sampling chains alone. According to Cowles and Carlin, it is because of that reason that many researchers in the various areas where MCMC methods have widely been used have come to the conclusion that those diagnostics are “fundamentally unsound” (Cowles and Carlin [15], p. 884). Cowles and Carlin however remark that a pragmatic reply to this position is the often held view that any kind of well-defined diagnostic (even if flawed) is still better than no convergence assessment strategy at all.

4.3.1. Overdispersed Starting Points

In this work we will use methods of assessing convergence recommended by Gelman et al. [23]. The recommended approach involves comparing m multiple sequences, which are run from different starting points. It is generally required that these starting points were generated from a starting distribution which is *overdispersed* with respect to the target $\pi(\boldsymbol{\theta}|\mathbf{m})$. Overdispersion in this context means that the starting distribution should have greater variance than the target distribution. Ideally, it should (generously) cover all of the target distribution, much like “[...] an approximate distribution for rejection sampling should cover the exact distribution” (Gelman and Rubin [24], p. 459).

Such a starting distribution, as it is also discussed in Gelman et al. [23] (pp. 322ff.), is for

4. Markov Chain Monte Carlo Methods

example some approximation based on previously found modes⁵ of the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{m})$. These modes can then be used to construct some normal-mixture approximation to the posterior. Suppose K modes $\boldsymbol{\mu}_k$ of the $\pi(\boldsymbol{\theta}|\mathbf{m})$ have been found: In Gelman and Rubin [24] (p. 459), a recommended approach is the approximation of the target distribution through the sum of K multivariate normals with center $\boldsymbol{\mu}_k$ and variance matrix Σ_k (determined with help of the second derivative matrix at each mode). Such a starting distribution would then be of the form

$$\widehat{\Pi}(\boldsymbol{\theta}) = \sum_{k=1}^K \rho_k (2\pi)^{(-d/2)} |\Sigma_k|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_k)\right), \quad (4.27)$$

where ρ_k represents a relative mass of the k th component of this sum of normal densities and d is the dimension of $\boldsymbol{\theta}$. The masses can be obtained by comparing $\widehat{\Pi}(\boldsymbol{\mu}_k)$ with the target density $\pi(\boldsymbol{\mu}_k|\mathbf{m})$ at $\boldsymbol{\mu}_k$. We seek to generate a relatively small sample of initial parameter vectors $\boldsymbol{\theta}$, each one to be used as starting point for a separate MCMC sampling procedure. In this case the authors recommend a more conservative approximation than the normal mixture to achieve greater coverage of the whole of the parameter space. This can be achieved by replacing each normal density in eq. 4.27 by a multivariate Student's t -distribution (p. 459), with a small number of degrees of freedom η , resulting in a mixture density function of the form

$$\tilde{\Pi}(\boldsymbol{\theta}) \propto \sum_{k=1}^K \rho_k |\Sigma_k|^{-1/2} (\eta + (\boldsymbol{\theta} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_k))^{-(d+\eta)/2}.$$

The authors consider a Cauchy mixture ($\eta = 1$) as a conservative first choice, while higher values for η might be preferable for a specific density at hand. The idea behind this is the fact that t distributions with η near 1 have broader tails than normal distributions and thus cover a larger part of the parameter space, a property that can be advantageous for posterior densities with significant probability mass (e.g. secondary maxima) in the vicinity of modes. It has to be noted, however, that in high-dimensional problems normal or Student's t -distributions may not be suited at all as overdispersed approximation for a given joint posterior. Draws from such a density may either stray to far from the mode to be still considered a good approximation of the target density or the mixed t density itself may on the other hand not be overdispersed at all. Herein lies, according to Cowles and Carlin [15] (p. 885), one of the main problems with Gelman and Rubins method: It is crucial to verify overdispersion of the chosen starting distribution, however, in practical problems knowledge of the posterior might be too limited to succeed in this endeavor.

4.3.2. Between-Within Variance

After having obtained an overdispersed starting distribution, the core of the Gelman and Rubin [24] method consists of drawing a number of points from it and use them as starting

⁵The term “modes” hereby refers to local maxima of the target distribution, several of which may have been found with the help of one or more optimization algorithms starting from different initial values.

4.3. Convergence Assessment

points for a parallel-running, iterative simulation procedure, e.g. a number of Markov chains generated by the Metropolis-Hastings algorithm. This approach is then based on the marginal densities derived from the joint posterior and monitors convergence for each parameter independently. The distribution of each scalar parameter θ_k to be estimated can be obtained through marginalizing over all the other parameters:

$$\pi_{\theta_k}(\theta_k|\mathbf{m}) = \int \pi(\boldsymbol{\theta}|\mathbf{m}) d\theta_1 d\theta_2 \dots d\theta_{k-1} d\theta_{k+1} d\theta_N.$$

The main idea behind Gelman and Rubin's approach in the form that is described in Gelman et al. [23] consists in the definition of estimators for the two statistics $E[\theta_k] := \mu = \int \theta_k \pi_{\theta_k}(\theta_k|\mathbf{m}) d\theta_k$ and $(\text{var}(\theta_k|\mathbf{m}))^2 = \int (\theta_k - \mu)^2 \pi_{\theta_k}(\theta_k|\mathbf{m}) d\theta_k$. The goal is to seek an estimator $(\widehat{\text{var}}(\theta_k|\mathbf{m}))^2$ with the following two characteristics:

1. Given overdispersion of the starting distribution, $(\widehat{\text{var}}(\theta_k|\mathbf{m}))^2$ overestimates $(\text{var}(\theta_k|\mathbf{m}))^2$.
2. Given strong stationarity (the starting distribution already equals the target distribution) or a long simulation time (number of iterations $n \rightarrow \infty$), $(\widehat{\text{var}}(\theta_k|\mathbf{m}))^2$ is unbiased: $E[(\widehat{\text{var}}(\theta_k|\mathbf{m}))^2] = (\text{var}(\theta_k|\mathbf{m}))^2$.

The monitoring strategy is focused on this estimator which can be constructed as follows: For m parallel chains, each of length n and each scalar estimand $\theta_k := \vartheta$, one assembles the simulation draws into a matrix ϑ_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) and defines the following quantities (Gelman et al. [23], p. 303):

$$\begin{aligned} B &:= \frac{n}{m-1} \sum_{j=1}^m (\bar{\vartheta}_{\cdot j} - \bar{\vartheta}_{\cdot\cdot})^2, \quad \text{where} \quad \bar{\vartheta}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \vartheta_{ij}, \quad \bar{\vartheta}_{\cdot\cdot} = \frac{1}{m} \sum_{j=1}^m \bar{\vartheta}_{\cdot j} \\ W &:= \frac{1}{m} \sum_{j=1}^m s_j^2, \quad \text{where} \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\vartheta_{ij} - \bar{\vartheta}_{\cdot j})^2. \end{aligned} \quad (4.28)$$

To estimate the marginal posterior variance of ϑ a weighted average of W and B is used:

$$\widehat{\text{var}}^+(\vartheta|\mathbf{m}) = \frac{n-1}{n} W + \frac{1}{n} B. \quad (4.29)$$

The superscripted “+” hereby indicates that, given overdispersion, this estimator tends to overestimate the marginal posterior variance (see eq. 4.54). The authors then recommend to monitor the convergence of the iterative simulation by estimating the scale reduction factor (p. 304) of the current distribution for $n \rightarrow \infty$. This scale reduction is estimated by

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\vartheta|\mathbf{m})}{W}}, \quad (4.30)$$

which tends to 1 from above for $n \rightarrow \infty$, given appropriate overdispersion of the starting values.

4. Markov Chain Monte Carlo Methods

The motivation behind this approach becomes clearer if we follow Gelman and Rubin [24] (p. 462–464) and think about it in the following way: The authors see the iteratively generated data ϑ_{ij} as a set of m treatments with n realizations each sampled from a population, in a similar sense as it is defined in one-way analysis of variance (ANOVA).

The term “treatment” hereby refers to the sequence of states generated by a Markov pro-

Chain 1	Chain 2	Chain 3	...	Chain m
ϑ_{11}	ϑ_{12}	ϑ_{13}	...	ϑ_{1m}
ϑ_{21}	ϑ_{22}	ϑ_{23}	...	ϑ_{2m}
ϑ_{31}	ϑ_{32}	ϑ_{33}	...	ϑ_{3m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
ϑ_{n1}	ϑ_{n2}	ϑ_{n3}	...	ϑ_{nm}

Table 4.1.: Simulation results of m different Markov chains thought of as m different treatments

cess, in other words to one particular Markov chain with specific initial state. Two different choices of an initial state are therefore seen as two different preparations—or treatments—of the same “experiment”. As described in Hoff [31] (p. 62), a basic model (“treatment effects model”) for fixed effects one-way ANOVA of ϑ_{ij} can be stated as

$$\begin{aligned}\vartheta_{ij} &= \mu + \gamma_j + \epsilon_{ij} \\ E[\epsilon_{ij}] &= 0 \\ \text{Var}[\epsilon_{ij}] &= \sigma_i^2,\end{aligned}\tag{4.31}$$

where μ is the grand mean, γ_j is an individual effect (a constant) of treatment j and ϵ_{ij} is a random error confined to the i -th observation of the j -th treatment. Individual treatment effects fulfill

$$\sum_j \gamma_j = 0$$

and we make the additional assumptions that the ϵ_{ij} are independent and

$$\begin{aligned}\epsilon_{ij} &\sim N(0, \sigma_i^2) \\ \sigma_i^2 &= \sigma.\end{aligned}\tag{4.32}$$

Hypothesis Testing

As discussed in Hoff, an important application of standard ANOVA is the testing of the null hypothesis $H_0 : \gamma_j = 0$ for all j , hence the absence of variation associated with treatment. Hence for the treatment means $\mu_j = \mu + \gamma_j$, it follows, that

$$\mu = \mu_1 = \dots = \mu_m.$$

4.3. Convergence Assessment

Under the alternative hypothesis H_1 one assumes that $\{\mu_1, \dots, \mu_m\}$ are not all equal. In order to estimate treatment means, the least squares method is often used to compare possible values of $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_m\}$:

$$SS_E(\boldsymbol{\mu}) = \sum_{i=1}^n \sum_{j=1}^m (\vartheta_{ij} - \mu_j)^2. \quad (4.33)$$

The global minimizer of $SS_E(\boldsymbol{\mu})$, denoted by $\hat{\boldsymbol{\mu}}$ turns out to be (see Hoff [31], pp. 63f.):

$$SS_E(\hat{\boldsymbol{\mu}}) = \sum_{i=1}^n \sum_{j=1}^m (\vartheta_{ij} - \bar{\vartheta}_{.j})^2.$$

It is evident that the pooled variance s^2 , a weighted average of the sample variances s_j^2 , is the within-variance W defined in equation 4.28:

$$MS_W := s^2 = \frac{\sum_{j=1}^m (n-1)s_j^2}{\sum_{j=1}^m (n-1)} = \frac{\sum \sum (\vartheta_{ij} - \bar{\vartheta}_{.j})^2}{m(n-1)} = \frac{SS_E(\hat{\boldsymbol{\mu}})}{m(n-1)}. \quad (4.34)$$

Analogously, between treatment variability is defined as

$$SS_T = \sum_{i=1}^n \sum_{j=1}^m (\bar{\vartheta}_{.j} - \bar{\vartheta}_{..})^2 = n \sum_{j=1}^m (\bar{\vartheta}_{.j} - \bar{\vartheta}_{..})^2, \quad (4.35)$$

from which we get the between-variance $MS_B = SS_T/(m-1)$ which is equivalent to B (eq. 4.28):

$$MS_B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\vartheta}_{.j} - \bar{\vartheta}_{..})^2. \quad (4.36)$$

Put together one can express the total variability of the data compared to the grand mean of the sample as sum of the between and within treatment variability (see for example Hoff [31], pp. 70f.):

$$SS_{\text{tot}} := \sum_{i=1}^n \sum_{j=1}^m (\vartheta_{ij} - \bar{\vartheta}_{..})^2 = SS_E + SS_T. \quad (4.37)$$

Given the assumptions stated in 4.32, one can calculate the expectations of MS_B and MS_W to obtain estimates for σ . First we want to assert that

$$\begin{aligned} E[\bar{\vartheta}_{.j}] &= E[\vartheta_{ij}] = \mu_j = \bar{\mu} + \gamma_j \\ \text{Var}[\bar{\vartheta}_{.j}] &= \sigma^2/n \\ \text{Var}[\bar{\vartheta}_{..}] &= \sigma^2/(nm) \end{aligned} \quad (4.38)$$

4. Markov Chain Monte Carlo Methods

with $\bar{\mu} := (1/m) \sum_{j=1}^m \mu_j$ and conclude that

$$E[\text{MS}_W] = \sigma^2 \quad (4.39)$$

$$E[\text{MS}_B] = \sigma^2 + \frac{n \sum_{j=1}^m \gamma_j^2}{m-1}. \quad (4.40)$$

One can see that 4.39 follows from

$$\begin{aligned} E[\text{MS}_W] &= \frac{1}{m(n-1)} \sum_{j=1}^m E \left[\sum_{i=1}^n (\vartheta_{ij} - \bar{\vartheta}_{.j})^2 \right] \\ &= \frac{1}{m(n-1)} \sum_{j=1}^m E[(n-1)s_j^2] \\ &= \frac{1}{m(n-1)} \sum_{j=1}^m (n-1)\sigma^2 \\ &= \sigma^2. \end{aligned} \quad (4.41)$$

The expected value 4.40, on the other hand, can be proven in the following way:

$$\begin{aligned} E \left[\frac{n}{m-1} \sum_{j=1}^m (\bar{\vartheta}_{.j} - \bar{\vartheta}_{..})^2 \right] &= E \left[\frac{n}{m-1} \left(\sum_{j=1}^m \bar{\vartheta}_{.j}^2 - 2 \bar{\vartheta}_{..} \sum_{j=1}^m \bar{\vartheta}_{.j} + \sum_{j=1}^m \bar{\vartheta}_{..}^2 \right) \right] \\ &= E \left[\frac{n}{m-1} \left(\sum_{j=1}^m \bar{\vartheta}_{.j}^2 - 2m \bar{\vartheta}_{..}^2 + m \bar{\vartheta}_{..}^2 \right) \right] \\ &= E \left[\frac{n}{m-1} \left(\sum_{j=1}^m \bar{\vartheta}_{.j}^2 - m \bar{\vartheta}_{..}^2 \right) \right] \\ &= \frac{n}{m-1} \left(\sum_{j=1}^m E[\bar{\vartheta}_{.j}^2] - m E[\bar{\vartheta}_{..}^2] \right). \end{aligned} \quad (4.42)$$

Using the identity

$$\begin{aligned} \text{Var}(\bar{\vartheta}_{.j}^2) &= E[\bar{\vartheta}_{.j}^2] - \mu_j^2 \\ \text{Var}(\bar{\vartheta}_{..}^2) &= E[\bar{\vartheta}_{..}^2] - \bar{\mu}^2, \end{aligned}$$

one can then express eq. 4.42 as

$$\begin{aligned}
 E[\text{MS}_B] &= \frac{n}{m-1} \left[\sum_{j=1}^m (\text{Var}(\vartheta_{.j}) + \mu_j^2) - m (\text{Var}(\vartheta_{..}) + \bar{\mu}^2) \right] \\
 &= \frac{n}{m-1} \left[\sum_{j=1}^m \mu_j^2 + \frac{m \sigma^2}{n} - \frac{m \sigma^2}{m n} - m \bar{\mu}^2 \right] \\
 &= \sigma^2 + \frac{n}{m-1} \left[\sum_{j=1}^m (\mu_j^2 - \bar{\mu}^2) \right] \\
 &= \sigma^2 + \frac{n}{m-1} \left[\sum_{j=1}^m (\mu_j - \bar{\mu})^2 \right] \\
 &= \sigma^2 + \frac{n}{m-1} \sum_{j=1}^m \gamma_j^2 \\
 &= \sigma^2 + n\varepsilon_\gamma,
 \end{aligned} \tag{4.43}$$

with

$$\varepsilon_\gamma = \frac{1}{m-1} \sum_{j=1}^m \gamma_j^2.$$

From these results, we can see that the expectations of both estimators under the null hypothesis $H_0 : \mu_j = \mu$ amount to

$$\begin{aligned}
 E[\text{MS}_W | H_0] &= \sigma^2 \\
 E[\text{MS}_B | H_0] &= \sigma^2,
 \end{aligned} \tag{4.44}$$

while under the alternative H_1 we get

$$\begin{aligned}
 E[\text{MS}_W | H_1] &= \sigma^2 \\
 E[\text{MS}_B | H_1] &= \sigma^2 + n\varepsilon_\gamma > \sigma^2.
 \end{aligned} \tag{4.45}$$

Thus one can see that the relation MS_B/MS_W should be close to one under the null hypothesis H_0 , while it will be > 1 under the alternative H_1 . According to Hoff [31] (p. 68), such a test statistic F is introduced within the framework of ANOVA, to detect deviations from the null hypothesis:

$$F = \frac{\text{MS}_B}{\text{MS}_W}. \tag{4.46}$$

To justify the ANOVA approach, Gelman and Rubin initially assume (p. 462) that the target distribution is normally distributed

$$\pi_\vartheta(\vartheta | \mathbf{m}) = N(\vartheta | \mu, \sigma^2)$$

4. Markov Chain Monte Carlo Methods

with unknown μ and σ . The second assumption is that of strong stationarity, hence $\pi_{\theta_k}(\theta_k|\mathbf{m}) = \rho_0(\vartheta) = \rho_t(\vartheta)$, where ρ_0 denotes the starting distribution and ρ_t the resulting distribution for any iteration t of the simulation process. These assumption both represent a “best-case” scenario. The author of this thesis interprets Gelman and Rubin’s approach as a matter of initially assuming the null hypothesis H_0 as discussed above. A similar statistic like the F-statistic in eq. 4.46 is then defined and examined for the ideal case.

As discussed by Gelman and Rubin [24] (p. 463), the following identity then holds for each of the m sequences $(\vartheta_{i1}, \vartheta_{i2}, \dots, \vartheta_{im})$, generated by the Metropolis-Hastings algorithm:

$$\frac{1}{n} \sum_{i=1}^n (\vartheta_{ij} - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (\vartheta_{ij} - \bar{\vartheta}_{.j})^2 + (\bar{\vartheta}_{.j} - \mu)^2. \quad (4.47)$$

We can see that identity 4.47 holds if we expand the right-hand side:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n ((\vartheta_{ij})^2 - 2 \vartheta_{ij} \bar{\vartheta}_{.j} + (\bar{\vartheta}_{.j})^2) + ((\bar{\vartheta}_{.j})^2 - 2 \bar{\vartheta}_{.j} \mu + \mu^2) \\ &= \frac{1}{n} \sum_{i=1}^n (\vartheta_{ij})^2 - \frac{2}{n} \sum_{i=1}^n \vartheta_{ij} \bar{\vartheta}_{.j} + \frac{1}{n} \sum_{i=1}^n (\bar{\vartheta}_{.j})^2 + (\bar{\vartheta}_{.j})^2 - 2 \bar{\vartheta}_{.j} \mu + \mu^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\vartheta_{ij})^2 - 2 (\bar{\vartheta}_{.j})^2 + (\bar{\vartheta}_{.j})^2 + (\bar{\vartheta}_{.j})^2 - 2 \bar{\vartheta}_{.j} \mu + \mu^2 \\ &= \frac{1}{n} \sum_{i=1}^n ((\vartheta_{ij})^2 - 2 \vartheta_{ij} \mu + \mu^2) \\ &= \frac{1}{n} \sum_{i=1}^n (\vartheta_{ij} - \mu)^2. \end{aligned} \quad (4.48)$$

If we take expectations of both sides of eq. 4.47, we get

$$E \left[\frac{1}{n} \sum_{i=1}^n (\vartheta_{ij} - \mu)^2 \right] = \sigma^2 = E \left[\frac{n-1}{n} s_j^2 \right] + \text{Var} [\bar{\vartheta}_{.j}]. \quad (4.49)$$

One can then, according to Gelman and Rubin, express the right-hand side in terms of B and W :

$$E \left[\frac{n-1}{n} W \right] = E \left[\frac{n-1}{n} s_i^2 \right] = \frac{n-1}{n} \sigma^2 \quad (4.50)$$

$$E \left[\frac{B}{n} \right] = \text{Var} [\bar{\vartheta}_{.j}] = \frac{\sigma^2}{n}. \quad (4.51)$$

We have already verified these identities in our treatment of one-way ANOVA between and within variances, eqns. 4.39 and 4.40. It follows that (see eq 4.49)

$$E[\widehat{\text{var}}^+(\vartheta|M)] = E \left[\frac{n-1}{n} W \right] + E \left[\frac{B}{n} \right] = \sigma^2 \quad (4.52)$$

4.3. Convergence Assessment

and thus

$$\widehat{\text{var}}^+(\vartheta|M) = \frac{n-1}{n}W + \frac{1}{n}B \quad (4.53)$$

is an unbiased estimator of the true variance σ^2 (Gelman and Rubin [24], p. 463).

This statistic can then be used to study the non-ideal case, i.e. the initial phase of the MCMC procedure, the output of which should more likely fall under the H_1 hypothesis: One can show that, given overdispersion, $\widehat{\text{var}}^+(\vartheta|M)$ overestimates σ^2 if the MCMC procedure has not yet “converged” and the Metropolis-Hastings algorithm does not yet sample from its stationary distribution. Expressing $E[\widehat{\text{var}}^+(\vartheta|M)]$ by means of the statistics of a specific sample chain with the help of eqns. 4.50, 4.51 and 4.53 yields

$$E[\widehat{\text{var}}^+(\vartheta|M)] = E\left[\frac{n-1}{n}s_i^2\right] + \text{Var}(\bar{\vartheta}_{\cdot j}).$$

Gelman and Rubin (p. 464) then use the right-hand side of the algebraic identity 4.47 and its expectations 4.49 to write

$$E[\widehat{\text{var}}^+(\vartheta|M)] = E\left[\frac{1}{n}\sum_{i=1}^n(\vartheta_{ij} - E[\bar{\vartheta}_{\cdot j}])^2\right],$$

using $\mu = E[\bar{\vartheta}_{\cdot j}]$.

Expanding and regrouping in order to express the right-hand side in terms of the mean μ_i and variance σ_i at iteration i yields:

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n E[\vartheta_{ij} - E[\bar{\vartheta}_{\cdot j}]]^2 \\ &= \frac{1}{n}\sum_{i=1}^n E[\vartheta_{ij}^2 - 2\vartheta_{ij}E[\bar{\vartheta}_{\cdot j}] + E[\bar{\vartheta}_{\cdot j}]^2] \\ &= \frac{1}{n}\sum_{i=1}^n (E[\vartheta_{ij}^2] - 2E[\vartheta_{ij}]E[\bar{\vartheta}_{\cdot j}] + E[\bar{\vartheta}_{\cdot j}]^2) \\ &= \frac{1}{n}\sum_{i=1}^n \{E[\vartheta_{ij}^2] - 2E[\vartheta_{ij}]^2 + E[\vartheta_{ij}]^2 + (E[\vartheta_{ij}]^2 - 2E[\vartheta_{ij}]E[\bar{\vartheta}_{\cdot j}] + E[\bar{\vartheta}_{\cdot j}]^2)\} \\ &= \frac{1}{n}\sum_{i=1}^n \{E[\vartheta_{ij} - E[\vartheta_{ij}]]^2 + (E[\vartheta_{ij}] - E[\bar{\vartheta}_{\cdot j}])^2\}. \end{aligned}$$

Hence, the expected value for $\widehat{\text{var}}^+(\vartheta|M)$ at iteration n finally amounts to:

$$E[\widehat{\text{var}}^+(\vartheta|M)] = \frac{1}{n}\left(\sum_{i=1}^n (\sigma_i^2 + (\mu_i - E[\bar{\vartheta}_{\cdot j}])^2)\right) \geq \sigma^2, \quad (4.54)$$

4. Markov Chain Monte Carlo Methods

if $\sigma_i^2 \geq \sigma^2$ for all i . Since, given appropriate overdispersion of the initial distribution, $\widehat{\text{var}}^+(\vartheta|M)$ overestimates the (true) variance of the marginal posterior density of ϑ , the ratio

$$R = \frac{\widehat{\text{var}}^+(\vartheta|M)}{\sigma^2}$$

will be > 1 ,⁶ as long as the Metropolis-Hastings algorithm is still heavily influenced by our chosen initial distribution rather than its stationary distribution. Generally, the denominator of R is unknown. One can, however, underestimate σ^2 by inserting W instead and thus gain an overestimate for R (Gelman and Rubin [24], p. 465). This can be motivated by the argument, that in the initial phase of the MCMC sampling procedure, the individual random walkers have not yet “wandered” across the whole of the target distribution, resulting in within-sequence variances smaller than σ^2 (Gelman et al. [23], p. 304).

Summing up this monitoring procedure, Gelman et al. recommend to monitor the convergence process of the iterative simulation by estimating the so-called “potential scale reduction factor” $\hat{R}$ of the current distribution (p. 304). This factor estimates the reduction in scale of the currently approximated posterior distribution of ϑ for a hypothetically infinite simulation period $n \rightarrow \infty$:

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\vartheta|M)}{W}}. \quad (4.55)$$

Given overdispersion of the starting distribution, this factor tends to 1 from above. Gelman et al. [23] argue (p. 304) that if this potential scale reduction is large, there is some evidence that continuing the simulation will improve the accuracy of inferences on the target distribution of the scalar parameter of interest.

⁶ R therefore reminds us of the test statistic F , defined in eq. 4.46

4.3. Convergence Assessment

The recommended step-by-step strategy for the simulation of an arbitrary posterior density given by Gelman et al. will in the following be summed up for the purposes of this thesis (see Gelman et al. [23], pp. 315-316):

1. Start with an approximation of the posterior density, for example with the help of previously obtained modes of the posterior (retrieved with the aid of some optimization procedure). Draw the initial values of the iterative procedure from such mode-based approximations and add noise to ensure overdispersion in comparison to the target distribution $\pi(\boldsymbol{\theta}|\mathbf{m})$.
2. In the most general case, if the conditional posterior densities of each scalar parameter are “non-standard” or unknown use the Metropolis-Hastings algorithm.
3. Monitor $\widehat{R}$ during the simulation. If $\widehat{R}$ is not near 1, continue the MCMC simulation procedure.
4. As soon as the ratio $\widehat{R}$ is near 1 and smaller than some predefined tolerance level (e.g. smaller than 1.1), collect the second halves of all m sequences and treat them as a representative sample from the respective marginal posterior distribution.
5. Check for inconsistencies, for example through comparison of inferences on parameters of interest with their respective starting distributions. Be skeptical of simulation results that suggest a radically different marginal posterior distribution, compared to the initial approximation and search for possible errors in the simulation and inference process.

In this context, Gelman et al. note (p. 304), that the arbitrary boundary within which $\widehat{R}$ is considered to be close enough to 1 generally depends on the specific problem to be solved. According to them, $\widehat{R} < 1.1$ is “acceptable” for most problems, however, concerning the more delicate cases, a smaller tolerance level might be advisable to inspire confidence in our inferences. Dropping the first halves of all sequences is hereby regarded as a “conservative choice” by Gelman et al. (p. 302), to minimize the influence that the initial distribution might still have on the sample draws of the random-walk algorithm (burn-in time).

Gelman et al. however remark, that even if $\widehat{R}$ is already very close to 1, there is still no definitive guarantee that convergence is reached: The starting distribution for example might misrepresent certain regions of the target distribution, and these areas might be difficult to reach with a given iterative simulation procedure. In addition to that, the univariate nature of this method can be seen as a disadvantage, although multivariate extensions have been proposed (See Brooks and Gelman [9]).

4. *Markov Chain Monte Carlo Methods*

5. Kinetic Model of CcO

Cytochrome *c* Oxidase (CcO), the final complex of the respiratory chain, has an important role in not only the human body. The enzyme reduces oxygen to water and subsequently uses the energy gained by this reduction to pump protons against an electrochemical potential gradient, across the mitochondrial inner membrane. This proton pump mechanism basically results from the coupling of a number of electron (ET) and proton transfer (PT) steps. CcO itself consists of four redox centres or ET sites, which are Cu_A, heme *a*, heme *a*₃ and Cu_B. In the following sections we will discuss a kinetic model of this important respiratory chain protein, which incorporates a large number of unknown parameters not easily accessible through direct measurement. As has been stated before, this model is based on the works of Walz [67] and Schach [57]. It is necessary to note that the description of the kinetic model given in the following sections was mainly adopted from Srajer et al. [63] as well as Schwaighofer et al. [59], to both of which the author of this thesis is one of the co-authors. Concerning Srajer et al. [63], contributions were made with regard to the calculation of proton association/dissociation probabilities as well as to the numerical solution of the master equation 5.1 for different parameter combinations via MATLAB, resulting in the computer-simulated cyclic voltammograms displayed in Figure 6 of Srajer et al. [63] (p. 44). In the case of Schwaighofer et al., the author of this thesis contributed significantly to the section “Estimation of kinetic model parameters from tr-SEIRAS data”, originated Figure 4 and obtained the maximum likelihood estimates listed in Table 2. Furthermore, the author of this present work composed parts of the section “Kinetic Model” of the supporting material. The sections “Parameter Estimation” and “MCMC Simulation Details” as well as the Figures S3–S11 and Tables S1–S4 of the supporting material were solely contributed by the author of this thesis. Parts of Srajer et al. [63], Schwaighofer et al. [59] and the Supporting Material will therefore be used word-by-word in the sections 5.1, 5.2, 5.2.1, 5.2.2, 5.2.3, 5.2.4, 5.3.3, 6.3.1 and 7 of this thesis without quotation marks.

5.1. Basic Experimental Setup

In the basic experimental setup described in Srajer et al., CcO from *R. sphaeroides*¹ is tethered to an electrode through a His-tag attached to subunit (SU) II, followed by reconstitution into a protein-tethered bilayer lipid membrane (ptBLM). According to previous studies, an ET path from the electrode directly to Cu_A is plausible [57] (see fig. 5.1). Above all, this system guarantees the accessibility of all four redox centers, among other advantages: A

¹*Rhodobacter sphaeroides* is a widely studied, gram-negative, purple photosynthetic bacterium.

5. Kinetic Model of CcO

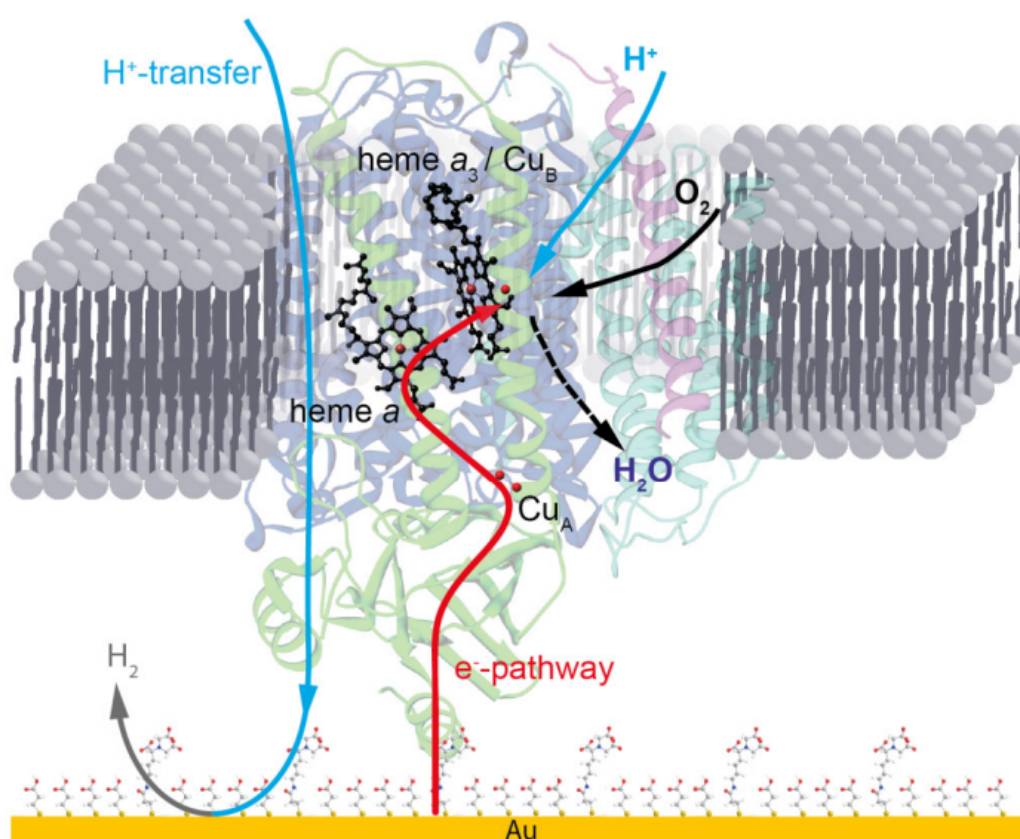


Figure 5.1.: Schematic view of the model system of Cytochrome *c* Oxidase from *R. sphaeroides*. The enzyme was immobilized on a gold electrode through a His-tag that was attached to subunit II and subsequently reconstituted into a lipid bilayer membrane (from: Srajer et al. [63], p. 34).

number of experimental methods can be applied to such a system to gain knowledge about the inner mechanisms of the enzyme and to infer on unknown parameters used in models which are describing its behaviour.

5.2. Kinetic Model

In this work a sequential four-electron transfer model will be used, which was previously developed by Schach [57]. This model postulates that an electron passes the four centres in the following order: Cu_A, heme *a*, heme *a*₃ and Cu_B. Furthermore, in Srajer et al. [63] an extension of the four-electron model is proposed in two steps: First of all, protonation equilibria of the oxidised and reduced species for all four centres are examined. Secondly, oxygen/H₂O is added as fifth redox couple, taking into account protonation of reduced oxygen to water. The proton and electron currents predicted by this kinetic model of five redox couples can then be compared to cyclic voltammetry data of the CcO in the absence and presence of oxygen. When oxygen is present, CcO exhibits enzymatic activity, demonstrated by amplified electron and proton currents [63]. For the evaluation of the tr-SEIRAS data studied in this thesis, however, the kinetic model under anaerobic conditions will be sufficient.

In summary chemical reaction kinetics based on a modified Nernst equation is applied, altogether describing the mode of operation of the enzyme in terms of standard redox potentials, rate constants and acid dissociation constants. We will now proceed to describe the details of this model.

5.2.1. Master Equation

Each of the four centres of the enzyme can individually be in either a reduced (**r**) or oxidised (**o**) state, adding up to a total of $2^4 = 16$ possible redox states of the enzyme (cf. fig. 5.4) and these states are called *conformations*.

In Srajer et al. [63] a probabilistic description of the dynamic processes within a single complex is chosen: A probability P_i for every single redox state i ($i = 1, \dots, 16$) of the enzyme can be defined and the time-evolution of this system is then expressed with help of a master equation

$$\frac{d}{dt}P_i = \sum_{j \neq i} (k_{ji}P_j - k_{ij}P_i), \quad (5.1)$$

where k_{ij} and k_{ji} are the forward and backward reaction rates for the transition from redox state i into state j . Regarding the relation between these transition rates the authors define the equilibrium constant $K = \frac{k_{ij}}{k_{ji}}$. For a single complex, the sum of the probabilities of all possible states must add up to unity at all times

$$\sum_{i=1}^{16} P_i = 1. \quad (5.2)$$

5. Kinetic Model of CcO

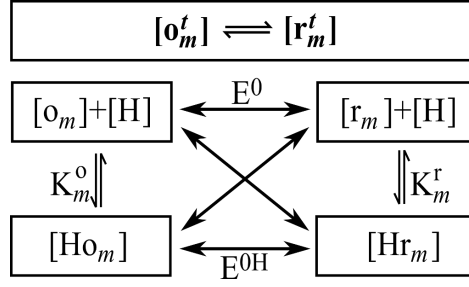


Figure 5.2.: Reaction scheme of an individual redox center m including protonations of oxidized and reduced species, where the superscripted t denotes total species (see eq. 5.5). The equilibrium between deprotonated and protonated species is indicated by the dissociation constants K_m^o and K_m^r . Possible transitions are indicated by black arrows, E^0 and E^{0H} denote the standard redox potentials for the deprotonated and the protonated species, respectively (from: Srajer et al. [63], p. 35).

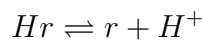
5.2.2. Redox Reactions and Protonation

In the course of a transition from an oxidised to a reduced state, protons (H^+) can be added to or removed from the oxidised or reduced centre, a reaction called protonation/deprotonation. As described in Srajer et al. [63] (p. 35) and depicted in figure 5.2, when looking at a single redox couple, there are four possible transitions. These redox reactions combined with protonation/deprotonation depend on the dissociation constants K_m^o and K_m^r , as well as the standard redox potentials E^0 (for the deprotonated species) and E^{0H} (for the protonated species).

To describe the relation between the overall applied potential and the ratio of concentrations of available redox species the Nernst equation is used [63]. For an unprotonated redox center m it is given by

$$E = E_m^0 + \frac{RT}{nF} \ln \frac{[o_m]}{[r_m]}, \quad (5.3)$$

where E indicates the electrode potential, E_m^0 the standard redox potential for center m , both measured relative to the standard hydrogen electrode (SHE). Furthermore T is the temperature, n the number of electrons involved in this process and R and F are the universal gas and Faraday constant, while expressions in squared brackets denote concentrations. A chemical reaction at equilibrium, i.e. for reduced species, of the form



is called dissociation.

The ratio of concentrations between deprotonated and protonated species is given by the acid

dissociation constants of the redox center m

$$K_m^r = \frac{[H][r_m]}{[Hr_m]}, \quad K_m^o = \frac{[H][o_m]}{[Ho_m]}, \quad (5.4)$$

Since, with current methods, it is not always possible to distinguish between protonated and deprotonated species experimentally, Srajer et al. adopted the idea of total reduced and oxidised species, developed by Walz [67]:

$$\begin{aligned} [r_m^t] &= [r_m] + [Hr_m] = [r_m](1 + 10^{pK_m^r - pH}) \\ [o_m^t] &= [o_m] + [Ho_m] = [o_m](1 + 10^{pK_m^o - pH}), \end{aligned} \quad (5.5)$$

where $pK_m^s = -\log K_m^s$ and pH refers to the pH-value in the solution. Inserting 5.5 into the Nernst equation (eq. 5.3) leads to

$$E = E_m^0 + \frac{RT}{nF} \ln \frac{[o_m^t]}{[r_m^t]} + \frac{RT}{F} \ln \frac{1 + 10^{pK_m^r - pH}}{1 + 10^{pK_m^o - pH}}. \quad (5.6)$$

Hence, the electrode potential is a function of the total concentrations of oxidised and reduced species as well as of the pK and pH values. The first and third term of 5.6 are usually combined to the midpoint potential E^M

$$E_m^M = E_m^0 + \frac{RT}{F} \ln \frac{1 + 10^{pK_m^r - pH}}{1 + 10^{pK_m^o - pH}}, \quad (5.7)$$

for redox center m [42, 67].

Considering a reaction between two such centers, a second order reaction equation can be written in terms of total species

$$[r_l^t] + [o_m^t] \rightleftharpoons [o_l^t] + [r_m^t]. \quad (5.8)$$

The equilibrium constant for this reaction is defined as

$$K = \frac{k_{lm}}{k_{ml}} = \frac{[o_l^t][r_m^t]}{[r_l^t][o_m^t]}, \quad (5.9)$$

and thus given by the ratio of forward and backward reaction rates at equilibrium. Using the modified Nernst equation (eq. 5.6) for two redox centers l and m , the equilibrium constant follows as

$$K = \frac{k_{lm}}{k_{ml}} = e^{\frac{nF(E_m^0 - E_l^0)}{RT}} \frac{(1 + 10^{pK_l^o - pH})(1 + 10^{pK_m^r - pH})}{(1 + 10^{pK_l^r - pH})(1 + 10^{pK_m^o - pH})}. \quad (5.10)$$

As discussed by Srajer et al. [63], “the flow of electrons J_{lm} ” (p. 36f) between center l and m in the course of a change from the conformational state i to j can be computed, provided

5. Kinetic Model of CcO

that the forward reaction rate constant k_{lm} has been determined (e.g. through measurement),

$$J_{lm} = \Gamma A \left(k_{lm} P_i - k_{lm} e^{\frac{nF(E_l^0 - E_m^0)}{RT}} \frac{(1 + 10^{pK_l^r - pH}) (1 + 10^{pK_m^o - pH})}{(1 + 10^{pK_l^o - pH}) (1 + 10^{pK_m^r - pH})} P_j \right), \quad (5.11)$$

with Γ denoting the surface concentration of the enzyme and A the surface area.

The quantity J_{lm} can also be interpreted as the net probability change dP_j of conformation j per time span dt , resulting from the “probability flow” between state i with probability P_i and state j with P_j , which is induced through the ET between center l and m .

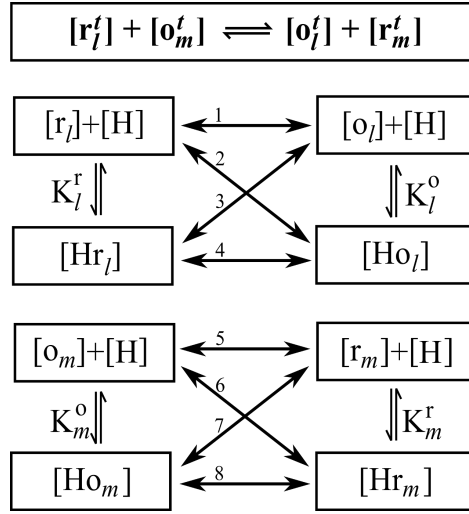


Figure 5.3.: Reaction scheme for a second-order reaction of total species (see eq. 5.5): Four different transitions are possible for each redox center separately (center l : 1-4, center m : 5-8), yielding a total of 16 unique combinations (from: Srajer et al. [63], p. 37).

5.2.3. Electrode Reactions

Second-order chemical reactions are useful to describe ET between the four centers of CcO. In the model system described in Srajer et al. [63], however, only the first redox center, Cu_A , is directly connected to the electrode. To describe the ET between the electrode and Cu_A , the authors use the Butler-Volmer equation [63, 3]. According to this empirically derived equation, ET rates between an electrode and a redox compound depend exponentially on the driving potential $E(t)$ (Léger and Bertrand [42], p. 2385):

$$\begin{aligned} k_{\text{ox}} &= k_e \exp[(1 - \alpha)f(E(t) - E^0)] \\ k_{\text{red}} &= k_e \exp[-\alpha f(E(t) - E^0)], \end{aligned} \quad (5.12)$$

with $f = F/RT$ and the transfer (symmetry) coefficient α .

Substituting E^0 with the midpoint potential (eq. 5.7) for the first center to account for protonation of the reduced and oxidised species, Srajer et al. [63] derive an expression for the current i^e resulting from the electron transfer between the electrode and Cu_A

$$\begin{aligned}
 k_{\text{ox}} &= k_e \exp \left[(1 - \alpha) f \left(E(t) - E_1^0 - \frac{1}{f} \ln \frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right) \right] \\
 &= k_e \exp \left[-(1 - \alpha) f \left((E_1^0 - E(t)) + \frac{1}{f} \ln \frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right) \right] \\
 &= k_e \left[e^{\frac{-(1-\alpha)F(E_1^0 - E(t))}{RT}} \left(\frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right)^{-(1-\alpha)} \right], \tag{5.13}
 \end{aligned}$$

$$\begin{aligned}
 k_{\text{red}} &= k_e \exp \left[-\alpha f \left(E(t) - E_1^0 - \frac{1}{f} \ln \frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right) \right] \\
 &= k_e \exp \left[\alpha f \left((E_1^0 - E(t)) + \frac{1}{f} \ln \frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right) \right] \\
 &= k_e \left[e^{\frac{\alpha F(E_1^0 - E(t))}{RT}} \left(\frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right)^{\alpha} \right]. \tag{5.14}
 \end{aligned}$$

The resulting current from this ET process (also called electron uptake) can then be calculated by summing up all the conformational state changes $P_{\gamma_j} \leftrightarrow P_{\delta_j}$ induced through an electron transfer between electrode and Cu_A , indicated by the row vectors γ and δ , where $\gamma = (1, 3, 4, 5, 9, 10, 11, 15)$ and $\delta = (2, 6, 7, 8, 12, 13, 14, 16)$ (see fig. 5.4):

$$\begin{aligned}
 i^e &= -F A \Gamma \sum_{j=1}^8 (k_{\text{red}} P_{\gamma_j} - k_{\text{ox}} P_{\delta_j}) \\
 &= \Omega \sum_{j=1}^8 \left(P_{\gamma_j} e^{\frac{F\alpha(E_1^0 - E(t))}{RT}} \left(\frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right)^{\alpha} - P_{\delta_j} e^{\frac{-F(1-\alpha)(E_1^0 - E(t))}{RT}} \left(\frac{1 + 10^{pK_1^r - pH}}{1 + 10^{pK_1^o - pH}} \right)^{-(1-\alpha)} \right), \tag{5.15}
 \end{aligned}$$

with $\Omega = -FAk_e\Gamma$, A the surface of the electrode-enzyme interface, k_e the standard rate constant of the electron transfer, Γ the enzyme surface density and α the charge transfer coefficient.

Essentially all of the reduced and oxidized species are prone to protonation, depending on

5. Kinetic Model of CcO

their specific pK^r and pK^o values (eq. 5.4). The total number of possible redox and protonation states therefore adds up to 256 different states of the enzyme, which are not accounted for in fig. 5.4.

The equilibrium constants 5.4 for a dissociation/association reaction determine the number of protons taken up or released in the course of a second order reaction like those depicted in scheme 5.3. Assuming equilibrium conditions, the author of this thesis contributed to this model specifically regarding the derivation of proton uptake/release probabilities from the dissociation constants for each of the four centres. The transition rates of protons can then be calculated by multiplication of the electron flow 5.11 with these probabilities (see the electronic supplementary information of Srajer et al. [63] for all the details). From that, the concentration of protons in the aqueous phase, $[H]$, can be computed. According to Srajer et al. [63], the reduction of two protons to one hydrogen molecule H_2 at the electrode leads to an observable current. These protons are then called *pumped protons* (p. 38) since they reach the electrode “[...] in the course of enzyme turnover” [63]. The resulting currents predicted by the model for different pK^s values can then be compared to the measured currents.

Regarding enzyme turnover under aerobic conditions, Srajer et al. [63] consider oxygen as a fifth redox couple, capable of exchanging electrons and protons in yet another second-order reaction with reduced Cu_B . The reduction of one O_2 molecule is then divided into four steps (transfer of one electron per step), where each one takes up $1/2 O + H$ to yield $1/2 H_2O$. As proposed by Srajer et al. [63], a full cycle of the CcO consists of these four steps and takes four protons and one oxygen molecule to produce two H_2O molecules. Schematics for all possible redox states and the transitions between them under anaerobic and aerobic conditions are depicted in fig. 5.4, respectively.

For the remaining part of this thesis, we want to focus on anaerobic conditions: the electrochemical electron uptake by Cu_A coupled to a series of second order electron exchange reactions to heme a , heme a_3 and Cu_B . The scheme depicted in fig. 5.4 can be seen as a graphical representation of the master equation 5.1. Regarding states that are connected through ET between the four centers, we interpret the flow J_{lm} 5.11 as the net probability change dP_j of conformation j per time interval dt , resulting from the “probability flow” between state i with probability P_i and state j with P_j , which is induced through ET between center l and m . State changes caused by electron transfers between the electrode and Cu_A (uptakes), on the other hand, are governed by the Butler-Volmer equation 5.15.

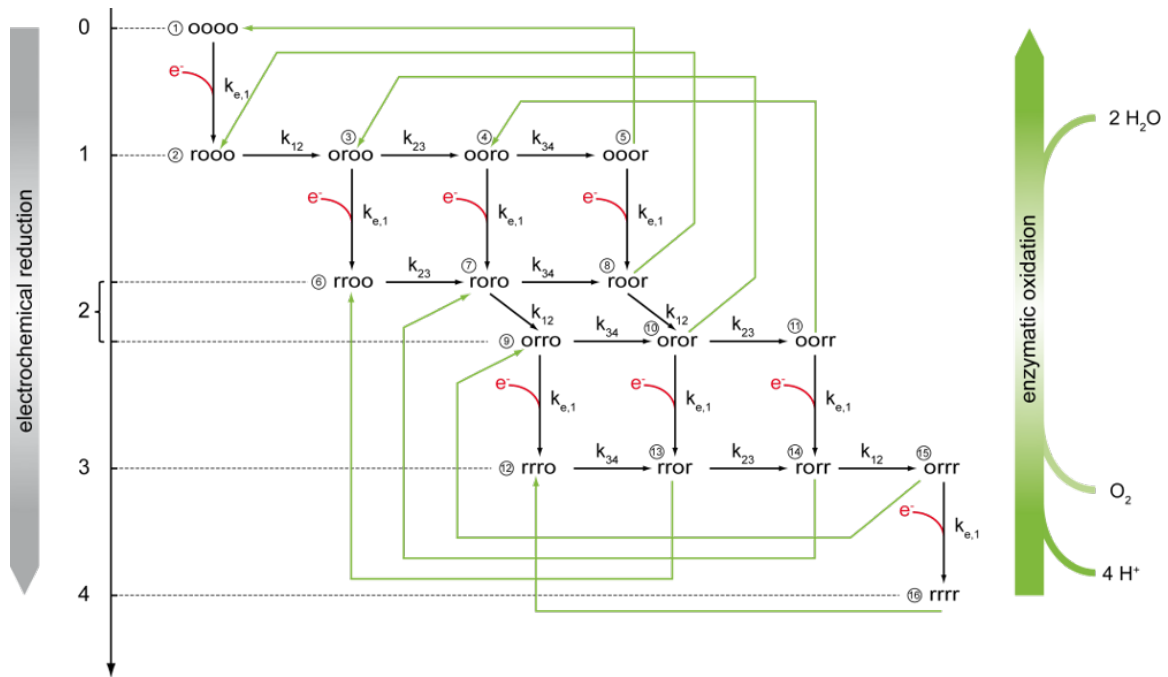


Figure 5.4.: Schematic view of the sequential electron transfer mechanism in CcO under anaerobic and aerobic conditions. The four centers of the enzyme are displayed in the sequence Cu_A , heme a , heme a_3 and Cu_B and can take on either a reduced (r) or an oxidised (o) state. Each of the reduced or oxidized centers can be protonated or deprotonated, yielding a total of 16 different redox states and 256 protonation states for the enzyme. Black vertical arrows with red branches depict electron uptake from the electrode, black horizontal arrows represent second-order redox reactions between the centers. Green arrows represent enzymatic oxidation under aerobic conditions: Different redox states of the enzyme are thus revisited in the course of these oxidation steps, with reduced Cu_B as starting point (from: Srajer et al. [63], p. 39).

5. Kinetic Model of CcO

5.2.4. Absorbance Changes Associated with Changes in Redox States

A central aim of this thesis is to obtain reasonable estimations for as many of the model's free parameters (listed in tab. 7.1) as possible. In collaboration with AIT and based on the work of Schach [56] (p. 59ff), the kinetic model was used to obtain estimations on some of the parameters, through comparing its output to data obtained by an experimental method called *time-resolved surface-enhanced IR-absorption spectroscopy* (tr-SEIRAS, see section 5.3.3) with the help of Bayesian inference. In general, Fourier transform infrared spectroscopy (see section 5.3.2) is a technique that can be used to visualize conformational changes of secondary structures within the CcO molecule. These conformational changes can then be assigned to redox centers and hence, changes in IR absorbance spectra can be predicted by our model, as will be discussed in the following.

Transmittance and Absorbance

If we direct infrared radiation towards a sample, some frequencies will be absorbed and therefore be attenuated in the emitted spectrum of the sample. We use a similar notation convention as in Schach [56] (p. 41) and define $I_0(\tilde{\nu})$ as the wavenumber-dependent initial radiation intensity and $I(\tilde{\nu})$ as the radiation intensity emitted by the sample. The transmittance $T(\tilde{\nu})$ is then defined as

$$T(\tilde{\nu}) = \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})}.$$

Lambert postulated the following differential equation, which states that the decrease in intensity is proportional to the intensity of the incident radiation and the thickness of the medium:

$$dI = -\alpha I dx.$$

We can solve this by separating the variables

$$\int_{I_0}^I \frac{dI'}{I'} = \int_0^d -\alpha dx.$$

It follows that

$$\ln\left(\frac{I}{I_0}\right) = -\alpha d.$$

Following Beer, who introduced the molar concentration c (in mol/l) to this equation, we get

$$\ln\left(\frac{I}{I_0}\right) = -\alpha' cd$$

where α' is the Napierian absorption coefficient (defined via the natural logarithm). We finally exponentiate both sides to obtain Beer-Lambert's equation

$$T(\tilde{\nu}) = \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})} = \exp(-\alpha' cd) = 10^{-\epsilon(\tilde{\nu})cd}. \quad (5.16)$$

In the last step we applied a base change with the decadic extinction coefficient $\epsilon(\tilde{\nu}) = \alpha'(\tilde{\nu}) \log(e)$. This exponent is also known as absorbance

$$A(\tilde{\nu}) = \epsilon(\tilde{\nu})cd.$$

In case of a mixture of substances with concentrations c_i , one can write Beer-Lambert's law for each substance as

$$\log_{10} \frac{1}{T(\tilde{\nu})} = A_i(\tilde{\nu}) = \epsilon_i(\tilde{\nu})c_i d.$$

From which one derives the total absorbance

$$A(\tilde{\nu}) = \sum_{i=1}^N \epsilon_i(\tilde{\nu})c_i d. \quad (5.17)$$

Calculation of Absorbance Changes

As described before, the various redox states (oxidized or reduced) of the centers arising from electron transfers are considered as different conformational states of the complex, which are numbered sequentially (sec. 5.2.1). The *overall* probabilities for the i -th center to take on either a reduced or an oxidized state, can be derived as follows (Schach [56], p. 37):

$$p_{i,\text{red}} = \sum_k P_{k|i,\text{red}}, \quad p_{i,\text{ox}} = 1 - p_{i,\text{red}}, \quad (5.18)$$

where $P_{k|i,\text{red}}$ designates the probability of the k -th conformation with a reduced i -th center. The relation for $p_{i,\text{ox}}$ follows from eq. 5.2. We interpret these probabilities as concentrations c_i of the respective species. Considering a band associated with the i -th center, Schach [56] (p. 59f) relates the contributions of the redox states of this center to the absorbance A_i at wavenumber $\tilde{\nu}$ and instant t with the help of Beer-Lambert's law (eq. 5.17) and the overall probabilities defined in eq. 5.18:

$$\begin{aligned} A_i(\tilde{\nu}, t) &= k_{i,\text{ox}}(\tilde{\nu})p_{i,\text{ox}}(t) + k_{i,\text{red}}(\tilde{\nu})p_{i,\text{red}}(t) \\ &= k_{i,\text{ox}}(\tilde{\nu}) + \Delta A_i(\tilde{\nu})p_{i,\text{red}}(t) \end{aligned} \quad (5.19)$$

with

$$\Delta A_i(\tilde{\nu}) = k_{i,\text{red}}(\tilde{\nu}) - k_{i,\text{ox}}(\tilde{\nu}).$$

According to Schach [56], the proportionality constants $k_{i,\text{red}}(\tilde{\nu})$ and $k_{i,\text{ox}}(\tilde{\nu})$ are “[...] products of the effective optical path length, the concentration of protein and the respective molar

5. Kinetic Model of CcO

extinction coefficient” (p. 60). Choosing the absorbance $A_i(\tilde{\nu}, 0)$ at $t = 0$ (the start of the periodic potential pulse) as reference, it follows from eq. 5.19 that

$$A_i(\tilde{\nu}, t) - A_i(\tilde{\nu}, 0) = \Delta A_i(\tilde{\nu})[p_{i,\text{red}}(t) - p_{i,\text{red}}(0)].$$

Schach [56] then derives for the change of absorbance in time

$$\Delta A_i(\tilde{\nu}, t) = \Delta A_i(\tilde{\nu})[p_{i,\text{red}}(t) - p_{i,\text{red}}(0)] + A_{i0}(\tilde{\nu}). \quad (5.20)$$

The term $A_{i0}(\tilde{\nu})$ is added in eq. 5.20 to account for a possible deviation of $A_i(\tilde{\nu}, 0)$ due to experimental scatter. In order to reduce the effect of experimental scatter for A_i at individual wavenumbers, $\Delta A_i(\tilde{\nu})$ is integrated over $\tilde{\nu}$, thus yielding the *band area change* $B_{i,\hat{\nu}}(t)$ (the constants $\hat{\nu}$, $\tilde{\nu}_{\min}$ and $\tilde{\nu}_{\max}$ refer to the peak wavenumber and to the boundaries of the band, respectively):

$$B_{i,\hat{\nu}}(t) = \int_{\tilde{\nu}_{\min}}^{\tilde{\nu}_{\max}} \Delta A_i(\tilde{\nu}, t) d\tilde{\nu}. \quad (5.21)$$

It follows that

$$B_{i,\hat{\nu}}(t) = [p_{i,\text{red}}(t) - p_{i,\text{red}}(0)] \cdot \left(\int_{\tilde{\nu}_{\min}}^{\tilde{\nu}_{\max}} \Delta A_i(\tilde{\nu}) d\tilde{\nu} \right) + \int_{\tilde{\nu}_{\min}}^{\tilde{\nu}_{\max}} A_{i0}(\tilde{\nu}) d\tilde{\nu} \quad (5.22)$$

and with the abbreviations

$$\overline{\Delta A_i}(\hat{\nu}) := \int_{\tilde{\nu}_{\min}}^{\tilde{\nu}_{\max}} \Delta A_i(\tilde{\nu}) d\tilde{\nu} \quad (5.23)$$

$$\overline{A_{i0}}(\hat{\nu}) := \int_{\tilde{\nu}_{\min}}^{\tilde{\nu}_{\max}} A_{i0}(\tilde{\nu}) d\tilde{\nu} \quad (5.24)$$

we can write $B_{i,\hat{\nu}}(t)$ as

$$B_{i,\hat{\nu}}(t) = [p_{i,\text{red}}(t) - p_{i,\text{red}}(0)] \overline{\Delta A_i}(\hat{\nu}) + \overline{A_{i0}}(\hat{\nu}). \quad (5.25)$$

5.2.5. Caveats of the Kinetic Model

For every model that seeks to emulate a real biophysical system, it is vital to map out the scope of its application. Due to the approximative character of most practical models there is some need of critical assessment of possible caveats of a model as well as a discussion of probable oversimplifications with respect to the real system.

As noted by Léger and Bertrand [42] (p. 2385), a “rigorous theoretical treatment” such as the Marcus theory does not yet exist for multicenter proteins in case of nonequilibrium conditions. We therefore make use of chemical kinetics, coupled to electrochemical kinetics in form of the simple Butler-Volmer approach, which describes ET at an interface in terms of the activation enthalpy $\Delta H^\ddagger$ of the ET. The idea behind this is the notion that a chemical reaction passes a transition state before yielding the reaction products. The energy it takes to get beyond the transition state is called activation energy E_A and can subsequently be

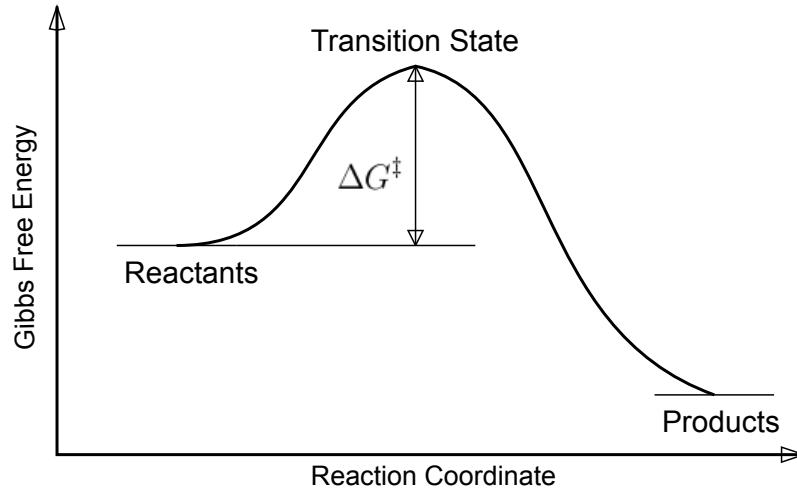


Figure 5.5.: Transition State Theory

identified with the activation enthalpy $\Delta H^\ddagger$. To calculate the change of Gibbs free energy $\Delta G_{\mathcal{X} \rightarrow \mathcal{Y}^\ddagger}$ of such a reaction at non-standard conditions, we make use of the ansatz [10]

$$\Delta G_{\mathcal{X} \rightarrow \mathcal{Y}^\ddagger} = \Delta G^\ddagger + RT \ln Q,$$

with $\Delta G^\ddagger$ the Gibbs energy change to the activated (=transition) state at standard conditions and Q the ratio of concentrations of transition state ($\mathcal{Y}^\ddagger$) and reactants ($\mathcal{X}$). If two reactants A, B form a transition state $AB^\ddagger$, then $K^\ddagger$ represents the fixed ratio of concentrations

$$K^\ddagger = \frac{[AB]^\ddagger}{[A][B]}$$

at (quasi-)equilibrium [17]. Since we assume a (quasi-)equilibrium between $A + B$ and $AB^\ddagger$, it follows that $\Delta G_{A+B \rightleftharpoons AB^\ddagger} = 0$ and therefore

$$\Delta G^\ddagger = -RT \ln K^\ddagger.$$

Knowing that, on the other hand,

$$\Delta G^\ddagger = \Delta H^\ddagger + T\Delta S^\ddagger$$

with the entropy of activation $\Delta S^\ddagger$, we get

$$K^\ddagger = \exp \left[-\frac{(\Delta H^\ddagger + T\Delta S^\ddagger)}{RT} \right]. \quad (5.26)$$

5. Kinetic Model of CcO

The right-hand side of eq. 5.26 is dimensionless and so we can introduce the reaction rate k and a constant C of dimension time (in seconds) with $K^\ddagger = k C$ to get

$$k = A \exp \left[-\frac{\Delta G^\ddagger}{RT} \right]$$

with $A = C^{-1}$.

Consider now a one-step, one electron redox reaction at an electrode with potential E : A small potential change $\Delta E = E - E^0$ with respect to an equilibrium potential E^0 of a redox center changes the energy of an electron present on the electrode by $-F\Delta E$ and hence the cathodic Gibbs energy for a redox reaction $O + e^- \rightarrow R$ changes by (Bard and Faulkner [3], p. 95)

$$\Delta G_c^\ddagger = \Delta G_{0c}^\ddagger + \alpha F(E - E^0),$$

where $\Delta G_{0c}^\ddagger$ is the cathodic activation energy at E^0 , $\alpha \in (0, 1)$ is the transfer coefficient and F is the Faraday constant. For the forward reaction rate this yields (Bard and Faulkner [3], p. 95)

$$k = A \exp(-\Delta G_{0c}^\ddagger/RT) \exp[-\alpha f(E - E^0)]. \quad (5.27)$$

When using these concepts (see section 5.2) to describe the kinetics of our electrochemical system, it is necessary to assume that for the applied potential E , the redox centers are “[...] close to equilibrium with respect to E ” (Schach [56], p. 38). This is the case for sufficiently low *scan rates* $\nu = \frac{dE}{dt}$ (see fig. 5.7). According to Léger and Bertrand, there is an upper boundary to this so-called *Nernstian regime* (Léger and Bertrand [42], p. 2382), where deviations from the typical Nernstian peak shapes begin to occur when “[...] the rate of electron transfer is not fast enough to maintain equilibrium between the redox site and the restless electrode potential” (Léger and Bertrand [42], p. 2385).

As we have seen, within the context of this Butler-Volmer approach, the electron transfer rate increases exponentially with the applied potential between a single redox center and an electrode (Léger and Bertrand [42], p. 2385; see also eq. 5.15):

$$\begin{aligned} k_{\text{ox}} &= k_0 \exp[(1 - \alpha)f(E - E^0)] \\ k_{\text{red}} &= k_0 \exp[-\alpha f(E - E^0)]. \end{aligned}$$

The constant k_0 hereby is the rate at zero driving force and $f = F/RT$. Léger and Bertrand [42] follow Laviron [41] and state that for scan rates smaller than k_0/f the Nernstian limit is restored. In our model, the values of k_0 were assumed to lie between $200 < k_0 < 700 \text{ s}^{-1}$ (see table 7.1). It follows that according to the criterion above, the maximum scan rate that can be applied to still remain in the Nernstian regime would be around (given $k_0 = 700 \text{ s}^{-1}$)

$$\nu_{\text{max}} = \frac{k_0}{f} = \frac{700}{38.917} = 17.987 \frac{\text{V}}{\text{s}}, \quad (5.28)$$

with

$$f = \frac{F}{RT} = \frac{96485.399}{8.315 \cdot 298.15} = 38.917 \text{ V}^{-1}.$$

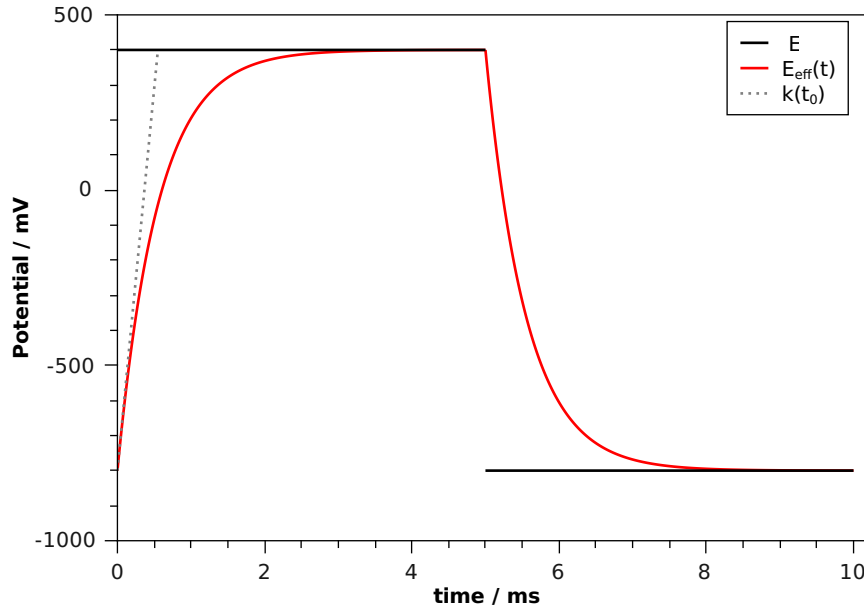


Figure 5.6.: Applied and effective Potentials

Remarks on the Simulation of Absorbance Measurements

When simulating cyclic voltammetry measurements, we model linear potential sweeps with ascending and descending slopes of 50 mV/s (see Fig 5.7) and therefore remain well below the boundary calculated in eq. 5.28.

For the 100 Hz square wave potential pulse applied to the electrochemical system in case of tr-SEIRAS measurements and the effective potential $E_{\text{eff}}(t)$ directly influencing the first center of the enzyme (compare section 5.3.3), however, our model is far beyond this limit.

Forming the derivative $\frac{dE}{dt}$ of the first leg of the effective potential function 5.35 with $\tau = 0.55$ ms

$$\frac{dE}{dt} = k(t) = -1200 \exp\left[-\frac{t}{0.55}\right] \cdot (-1/0.55),$$

we would get for $t = 0$ the theoretical scan rate (illustrated as grey dotted line in fig. 5.6)

$$\nu_{t_0} = -1200 \cdot (-1/0.55) = 2181.818 \frac{V}{s}.$$

For the first half of the potential pulse, for example, we can calculate the time it takes for the scan rate to slow down enough to be considered Nernstian again:

$$\ln(17.987 \cdot 0.55/1200) = -\frac{t}{0.55}$$

$$t_{\text{Nernst}} = -\ln(17.987 \cdot 0.55/1200) \cdot 0.55 = 2.639 \text{ ms}.$$

5. Kinetic Model of CcO

Therefore, with frequencies as high as 100 Hz, the conditions of the Nernstian regime are not obtained at less than 2.639 ms. Between 2.639 and 5 ms electrochemical kinetics and Marcus theory “converge”, afterwards the second pulse of the cycle steeply increases the scan rate again. However, it is important to point out that the potential pulse frequency of 100 Hz was chosen, because of the quality of the spectra optimized under these conditions.

Other Caveats

Further simplifications of our kinetic model are as follows:

- Cooperative effects between different redox centers were neglected
- Protonation is considered to be at equilibrium
- In our model system (cf. fig. 5.1), electron transfer between the Au electrode and Cu_A occurs via the pathway of a 2.3 nm wide “spacer”—i.e. the nitrilotriacetic acid (NTA)-terminated linker molecules binding via Nickel (Ni²⁺) complexation to the polyhistidine-tag (“His-tag”) that was engineered to subunit II of CcO (see Schach [56], pp. 22, 68). Direct ET in this context is usually described as quantum mechanical tunneling of an electron through the potential energy barrier between the states on the electrode and the states occupied by the reactant (see Bard and Faulkner [3], p. 130). The probability of a successful reaction therefore decreases exponentially with the distance between the reaction pair and this distance is usually held below 2 nm. Given our assumed range for k_0 (see tab. 7.1), the Ni complex itself must be assumed as an intermediate redox center (see Friedrich et. al. [21], p. 3700).
- Electrostatic effects were not considered.

Considering the listed points, we would like to point out that the main purpose of the present model calculation is to draw attention to the influence of different protonation equilibria on the energetics between redox centers, which frequently are considered without any or only a single pK value.

5.3. Measurement Methods

5.3.1. Cyclic Voltammetry

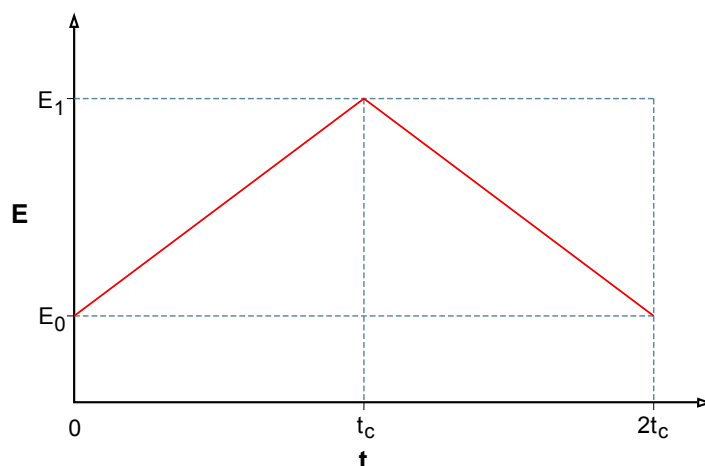
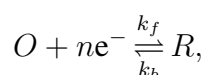


Figure 5.7.: Applied potential sweep in cyclic voltammetry with reversal potentials E_0 , E_1 and scan rate $v = \frac{E_1 - E_0}{t_c}$. This Illustration was redrawn and adapted from Schach [56], p. 32, Figure 5.9.

Cyclic voltammetry is a method used to “[...] gain information about thermodynamics of redox processes, kinetics of heterogenous electron transfer reactions, adsorption processes as well as coupled chemical reactions” (cf. Schach [56], p. 32).

A time-dependent slowly changing potential is applied between the working electrode with CcO attached to it and a reference electrode. The applied potential changes linearly during the first half cycle, starting from a initial potential E_0 , until a reversal point E_1 is reached (see fig. 5.7). In the second half cycle the linear potential change commences in the opposite direction until reaching E_0 again. These cycles are repeated and the resulting currents that pass through a counter electrode are recorded. Two types of currents are distinguished: *Faradaic* and *non-faradaic*.

Non-faradaic currents are caused mainly by the “[...] reorganisation of ions near the electrode interface” (Schach [56], p. 33). More important for the investigation of redox reactions of the enzyme are faradaic currents, stemming from charge transfer processes during reduction/oxidation of species at the electrode, i.e. reactions of the form



where k_f and k_b are forward and backward rate constants and n is the number of electrons involved [56].

5. Kinetic Model of CcO

Figure 5.8 schematically shows a typical cyclic voltammogram, plotting applied potential versus measured current at the counter electrode. For the above mentioned experimental setup for CcO measurements under anaerobic conditions the resulting CVs feature a similar shape (see fig. 5.9).

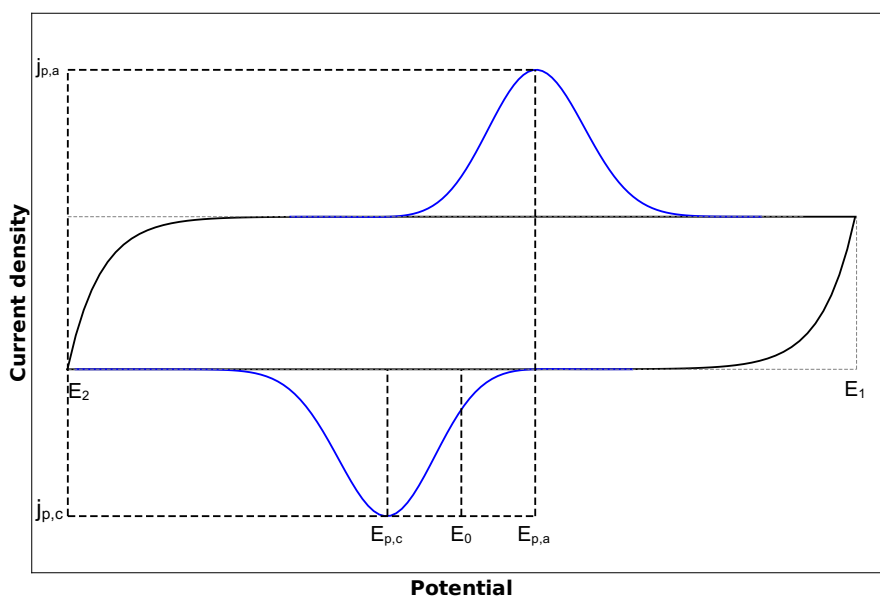


Figure 5.8.: Cyclic Voltammogram (in this case for a diffusion controlled redox process), schematically; black: non-faradaic currents; blue: faradaic currents; $j_{p,c}$, $E_{p,c}$: anodic peak current and potential; $j_{p,a}$, $E_{p,a}$: cathodic peak current and potential; E_0 : equilibrium potential; E_1 , E_2 : reversal potentials. This Illustration was redrawn and adapted from Schach [56], p. 33, Figure 5.10.

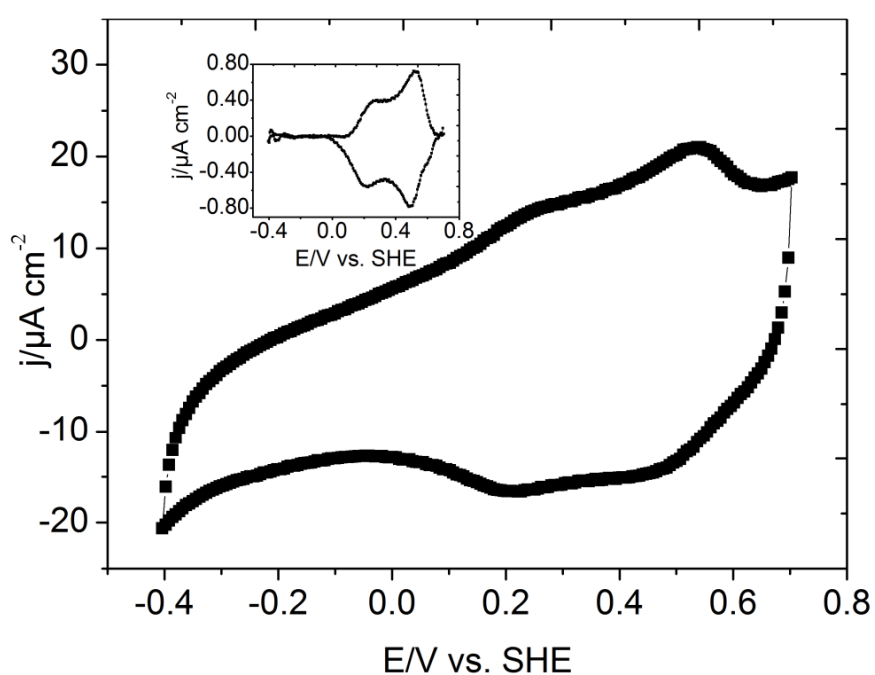


Figure 5.9.: Measured CV of CcO under anaerobic conditions in the non-activated state. The inset displays the baseline corrected signal, normalized to the scan rate (50 mVs^{-1}), thereby excluding currents that are caused by charging effects of the ptBLM (from: Srajer et al. [63], p. 42).

5.3.2. Fourier-Transform Infrared Spectroscopy

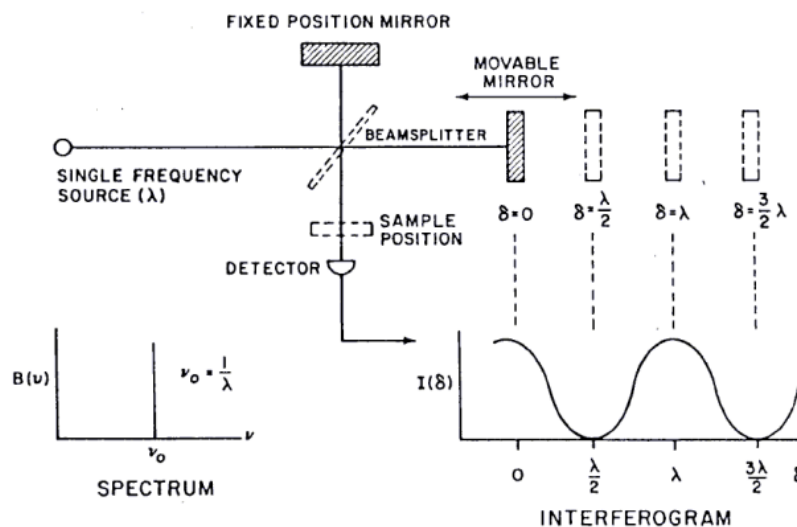


Figure 5.10.: Fourier-Transform Interferometer (from: Perkins [53], p. A5).

Fourier transform spectroscopy is a method to record an IR absorption spectrum, using a light source containing a full spectrum of wavelengths (as opposed to a monochromatic source). A beam is directed into a Michelson interferometer, containing a sample with unknown absorption spectrum. As shown in fig. 5.10, this device basically consists of a beam splitter and a stationary as well as a movable mirror. Due to interference with itself, the beam gets periodically annihilated and amplified, depending on the optical path difference δ , twice the difference between the linear distance of the stationary mirror to the beam splitter and the distance of the movable mirror to the beam splitter.

Generally, we can write for two separate, one-dimensional, monochromatic plane waves moving in the positive x -direction with amplitude A_i and phases $\varphi_1(x) = kx$, $\varphi_2(x) = k(x + \delta)$, total path difference δ and $\Delta\varphi = \varphi_1(x) - \varphi_2(x)$ (see for example Kauppinen and Partanen [35], p. 77)

$$W_1(x, t) = A_1 \exp[i(\varphi_1(x) - \omega t)]$$

$$W_2(x, t) = A_2 \exp[i(\varphi_2(x) - \omega t)].$$

And for the recombined wave follows

$$W(x, t) = A_1 \exp[i(\varphi_1(x) - \omega t)] + A_2 \exp[i(\varphi_2(x) - \omega t)].$$

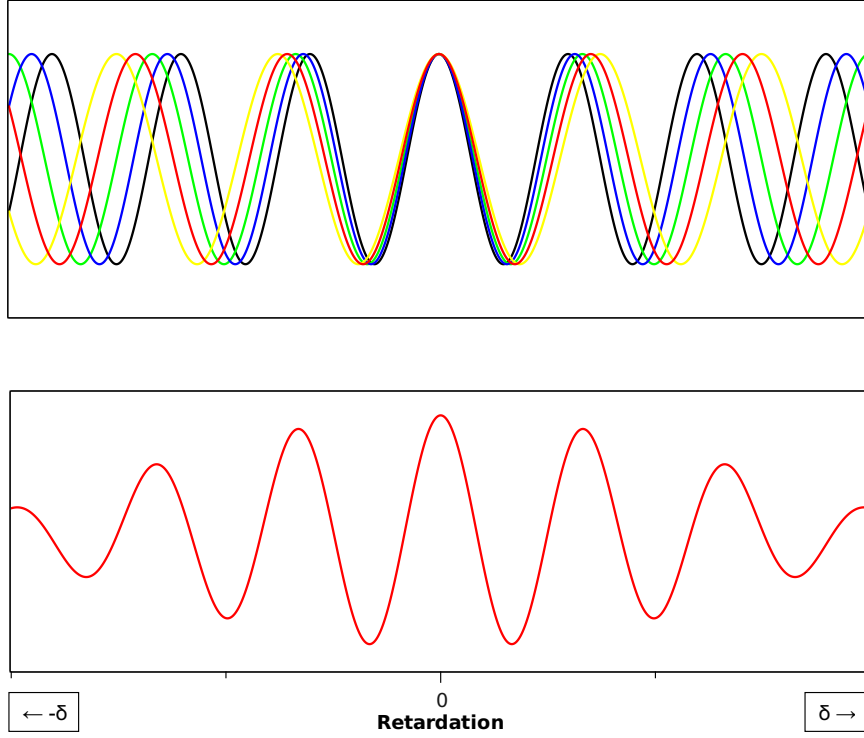


Figure 5.11.: Polychromatic interference: For each wavelength there is a separate cosine interference pattern 5.30. The sum over all wavelengths yields the interferogram displayed in the lower half. This Illustration was redrawn and adapted from Arnolds [2].

The intensity of the recombined beam is in general a function of x and t and proportional to the squared amplitude of W [35],

$$I(x, t) \propto |W(x, t)|^2 := A^2. \quad (5.29)$$

In our case, however, it only depends on the total path difference δ

$$I(\delta) \propto W(x, t)W^*(x, t) = A^2 = A_1^2 + A_2^2 + 2A_1A_2 \cos(\Delta\varphi).$$

Identifying the squared amplitudes with the intensities of the two beams $I_i = A_i^2$ and inserting $k = \frac{2\pi}{\lambda}$ we get [2]

$$I(\delta) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\left(\frac{2\pi}{\lambda}\delta\right). \quad (5.30)$$

Since our source is polychromatic, it emits a range of different frequencies, all of which separately show a cosine interference pattern. These components add up to a single total interference pattern, which is—at least in theory—symmetric around the point of zero path

5. Kinetic Model of CcO

difference. To get an expression for the intensity measured by a detector at a fixed point as function of the optical path difference δ , we integrate over a range of wavenumbers (denoted by $\mathcal{B}$) [2]:

$$I(\delta) = \int_{\mathcal{B}} I(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta) d\tilde{\nu}.$$

By defining an indicator function

$$\mathcal{C}(\tilde{\nu}) = \begin{cases} 1 & \tilde{\nu} \in \mathcal{B} \\ 0 & \tilde{\nu} \notin \mathcal{B} \end{cases}$$

one can write [2]

$$I(\delta) = \int_{-\infty}^{\infty} I(\tilde{\nu}) \mathcal{C}(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta) d\tilde{\nu}. \quad (5.31)$$

Comparing 5.31 with the Fourier Transform of $I_{\mathcal{B}}(\tilde{\nu}) := I(\tilde{\nu})\mathcal{C}(\tilde{\nu})$,

$$\begin{aligned} \mathcal{F}(I_{\mathcal{B}}(\tilde{\nu})) &= \int_{-\infty}^{\infty} I(\tilde{\nu}) \mathcal{C}(\tilde{\nu}) e^{-2\pi i \tilde{\nu} \delta} d\tilde{\nu} \\ &= \int_{-\infty}^{\infty} I(\tilde{\nu}) \mathcal{C}(\tilde{\nu}) \cos(2\pi\tilde{\nu}\delta) d\tilde{\nu} - i \int_{-\infty}^{\infty} I(\tilde{\nu}) \mathcal{C}(\tilde{\nu}) \sin(2\pi\tilde{\nu}\delta) d\tilde{\nu}, \end{aligned}$$

one can interpret $I(\delta)$ as the Fourier cosine transform of $I_{\mathcal{B}}(\tilde{\nu})$. The cosine is an even function ($\cos(x) = \cos(-x)$), while the sine is odd ($-\sin(x) = \sin(-x)$). A Fourier cosine transform of a function $I_{\mathcal{B}}(\tilde{\nu})$ is therefore an even function in δ , hence $I(-\delta) = I(\delta)$. Performing the inverse transformation² yields

$$I_{\mathcal{B}}(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta) \cos(2\pi\delta\tilde{\nu}) d\delta. \quad (5.32)$$

The product of $g_e(x)$, an even function, and $g_o(x)$, an odd function, is odd. Integration of an odd function yields an even function and the definite integral of an odd function over an interval of the form $(-a, a)$ is zero:

$$\int_{-a}^a g_o(x) dx = G_o(a) - G_o(-a) = G_o(a) - G_o(a) = 0.$$

The inverse Fourier cosine transform 5.32 is therefore equivalent to the inverse Fourier transform of $I(\delta)$,

$$I_{\mathcal{B}}(\tilde{\nu}) = \int_{-\infty}^{\infty} I(\delta) e^{2\pi i \delta \tilde{\nu}} d\delta, \quad (5.33)$$

because the definite integral of the product of $\sin(2\pi\delta\tilde{\nu})$ and the even function $I(\delta)$ over $(-\infty, \infty)$ vanishes. Thus, having measured intensity changes for various optical path differences (an interferogram), one can deduce a spectrum $I_{\mathcal{B}}(\tilde{\nu})$ [2].

A common practice is to record the intensity of a reference beam $I_R(\tilde{\nu})$ (after passing the

²The inverse Fourier cosine transform of an even function is also a Fourier cosine transform [2]

empty interferometer). The resulting reference spectrum contains information about the spectral composition and energy distribution of the light source and the optical properties of the interferometer. Ultimately, after recording the spectrum of a sample $I_S(\tilde{\nu})$ we can calculate the (time-dependent) absorbance of a sample (see Schach [56], p. 52):

$$A(\tilde{\nu}, t) = -\log \frac{I_S(\tilde{\nu}, t)}{I_R(\tilde{\nu})}. \quad (5.34)$$

Attenuated Total Reflectance

Attenuated total reflectance (ATR) is a technique used in FTIR using total inner reflection of a crystal which is diaphanous for wavelengths of the infrared region in order to measure absorbance spectra of a sample. This method takes advantage of a property of total reflection called evanescent wave which is capable of extending the beam into the interfacial area between the crystal and the sample and even into the sample itself, up to a certain penetration depth. The intensity of the electric field decreases exponentially with the penetrating distance. For total reflectance to occur the incident angle θ must be greater than a particular critical angle $\theta > \theta_c$.

The penetrating depth at which the amplitude of the electric field has decreased to the $1/e$ -th part of the field amplitude inside the crystal can be calculated as follows

$$d_p = \frac{\lambda_{\text{vac}}}{2\pi \sqrt{(n_1^2 \sin^2(\theta)) - n_2^2}},$$

with the vacuum wavelength λ_{vac} and the refractive indices n_1 and n_2 of the crystal and the sample, respectively (see Schach [56], p. 53). One of the main advantages of this method is a far easier sample preparation process.

Surface Enhancement Effect

If a sample prepared for absorbance measurements contains only trace amounts of the substance of interest, a technique that utilizes the surface enhancement effect of infrared absorption (SEIRA) of rough metallic surfaces can be employed. It makes use of effects like the amplification of the electromagnetic field near those surfaces (e.g.: Au, Fe, Ag) and thus the “[...] enhancement of optical processes by a factor of 10^2 to 10^{12} ” (Schach [56], p. 53). The phenomenon is generally interpreted in twofold ways: Enhancement of the electromagnetic field as a result of the excitation of so-called localized surface plasmon oscillations (from a classical point of view: collective charge density oscillations of the “free” electron gas of the rough metal surface) and an augmentation of the transition dipole moment of adsorbed molecules (see Schach [56], p. 53f).

5. Kinetic Model of CcO

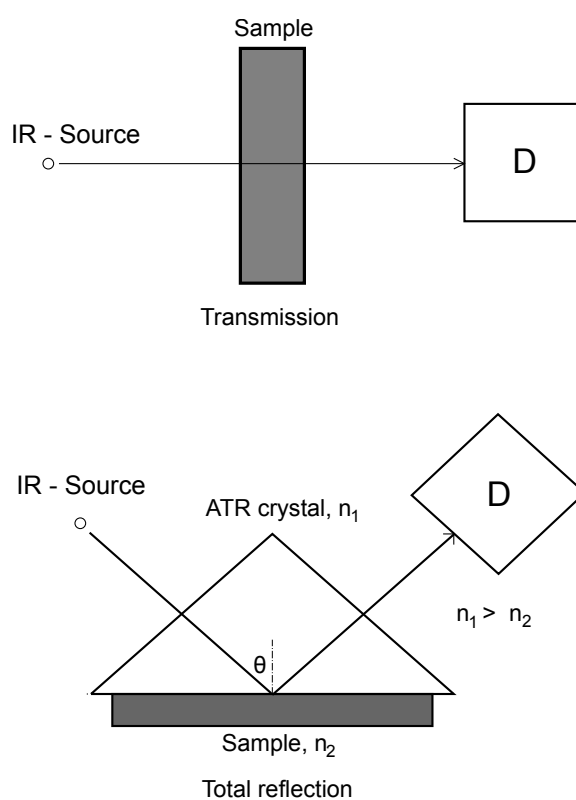


Figure 5.12.: Sampling techniques: Transmission and attenuated total reflectance (ATR), n_1 and n_2 denote refractive indices, D denotes the detector. This illustration was redrawn and adapted from Fringeli [22]^a (p. 277).

^aAs cited in Baurecht [5], p. 54

Time-Resolved Surface Enhanced Infrared Absorption Spectroscopy (tr-SEIRAS)

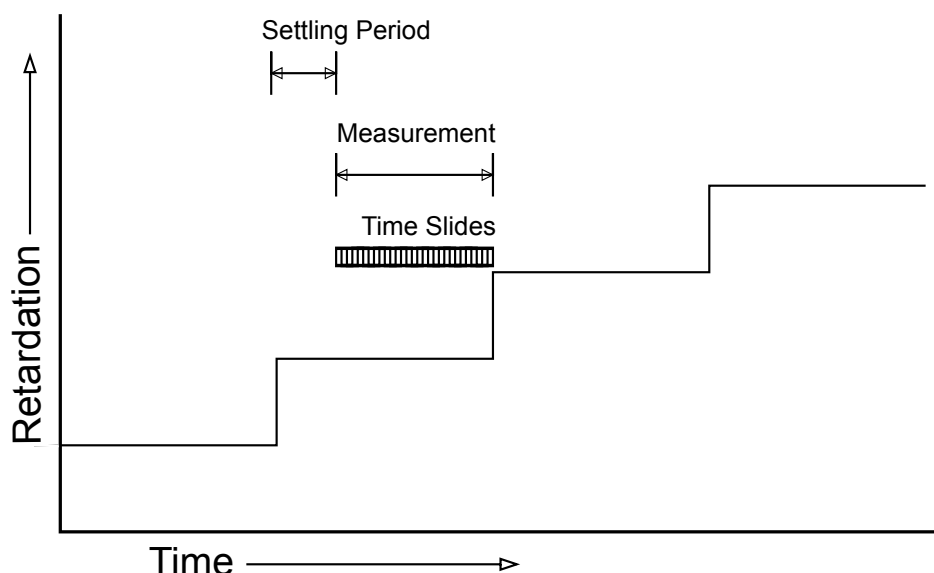


Figure 5.13.: Step Scan Technique. This illustration was redrawn and adapted from Nakano [48], p. 1356

Time-resolved spectroscopy is a method to study the temporal evolution of a system's response to external stimulation. In our case we want to focus on a technique called *step scan mode* in order to get time-dependent IR absorption spectra: The movable mirror of the interferometer is brought to a particular position x_i and after a short settling period the absorbance intensity is recorded. Since the absorbance of our sample can change over time, we divide the measuring period (at mirror position x_i) into discrete intervals and record the absorbance for each one of this so-called *time slides*. Subsequently, the mirror is moved to the next position x_{i+1} and another set of data points is recorded (fig. 5.13 illustrates the process). After all mirror positions $x_0, \dots, x_N$ have been probed in that manner, data points captured at equal points in time (time slides) are recombined to yield one interferogram for each time slide. Applying Fourier transformations, one can then obtain the respective IR absorbance spectra (see Schach [56], p. 56ff).

5.3.3. Electrochemical Excitation of CcO and tr-ATR-SEIRAS Measurements

In section 5.2.4, in a fashion similar to the approach described in Schach [56] (p. 59ff), we explained how absorbance changes can be identified with changes of the overall redox states of our model system and hence with conformational changes of CcO.

In collaboration with AIT, this model was then implemented in MATLAB2010Rb within the

5. Kinetic Model of CcO

context of this thesis and subsequently used to acquire estimations on some of the free parameters of the model, by comparing its output to data obtained by the tr-SEIRAS measurements with the help of Bayesian inference. Ultimately, great effort was put into the objective of obtaining an optimal set of parameters, preferably yielding the best possible agreement between the band area change for the first three centres as predicted by the model (compare eq. 5.25) and the measured data.

The details of the Bayesian parameter estimation model will be provided in section 6.3.1. We now want to give a brief overview of the tr-ATR-SEIRAS measurement process as described in Schwaighofer et al. [59] to explain how the experimental data of the band area changes was obtained.

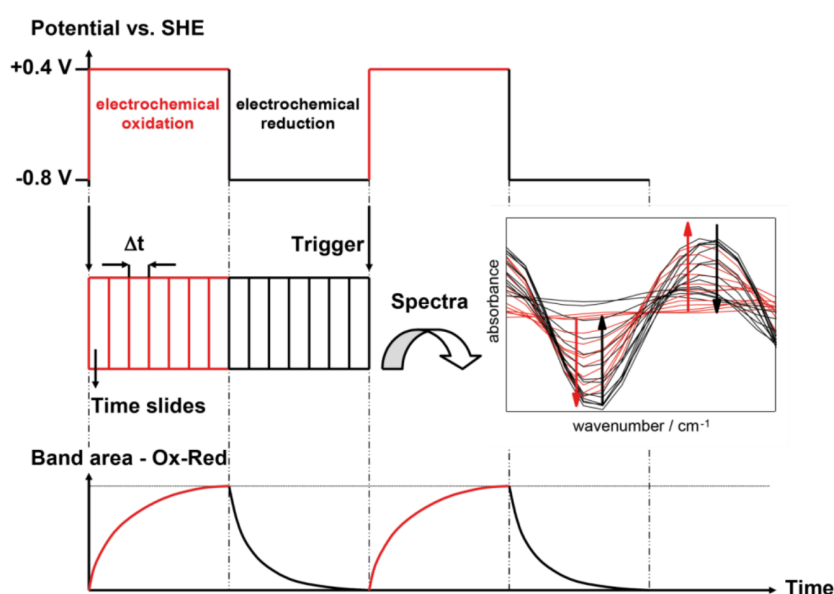


Figure 5.14.: Principle of electrochemically modulated tr-ATR-SEIRAS. Top: applied square wave potential pulse vs Standard Hydrogen Electrode (SHE); middle: recording of time-resolved FTIR spectra triggered by rapid potential changes at the beginning of each period; bottom: band area vs. time; from: Schwaighofer et al. [59], Supplementary Information p. 2.

Experimental Setup

As described by Schwaighofer et al. [59], an electrochemical cell (depicted in Fig 5.1) was installed on a trapezoid single-reflection silicon ATR crystal. The CcO molecule was immobilized on the crystal (for details see Schwaighofer et al. [58]) with the first center Cu_A facing the electrode and reconstituted into a membrane system. In preparation for the IR measurements, up to 20 cyclic voltammograms were recorded under aerobic conditions in order to change the conformational state of the enzyme from inactivated to activated. Thereafter, the IR measurements were performed in an anaerobic environment.

An infrared beam from a Fourier-transform IR spectrometer was then coupled into the crystal at an incidence angle of $\Theta = 60^\circ$, utilizing a custom-made setup. The intensity of the totally reflected IR beam was measured and Fourier-transform IR spectra were recorded. A square wave potential was applied to the membrane system (containing CcO) located on the gold film electrode to induce a periodical change from the fully reduced (at -800 mV) to the fully oxidized conformational state (at 400 mV) of the enzyme. As a result of charging effects of the gold/water interphase, the effective potential $E_{\text{eff}}(t)$ directly affecting the first center of the enzyme deviates from the applied potential:

According to Schach, “[...] each change in applied potential E_p gives rise to a capacitive current which causes the effective potential relevant to the redox centers to deviate from E_p ” (Schach [56], p. 59), due to the charging of the interface at the electrode surface. The potential effective for the redox centers, $E_{\text{eff}}(t)$, is therefore modeled by the consecutive charging and “discharging” of an RC-circuit with relaxation time τ , depending on the applied rectangular potential pulse, hence

$$E_{\text{eff}}(t) = \begin{cases} E_{p,1} - \Delta E \cdot \exp\left[-\frac{t}{\tau}\right] & \text{for } 0 \leq t \leq \frac{T}{2} \\ E_{p,0} + \Delta E_{T/2} \cdot \exp\left[-\frac{t-(T/2)}{\tau}\right] & \text{for } \frac{T}{2} < t \leq T \end{cases} \quad (5.35)$$

with the vertex potentials $E_{p,0} = -800$ mV, $E_{p,1} = 400$ mV and

$$\begin{aligned} \Delta E &= E_{p,1} - E_{p,0}, \\ \Delta E_{T/2} &= E_{p,1} - \Delta E \cdot \exp\left[-\frac{T}{2\tau}\right] - E_{p,0}. \end{aligned}$$

Note that in general there is a certain transient effect (depending on the frequency and amplitude of the periodic pulse and τ) until the effective potential turns periodic. For the Bayesian parameter estimation that will later be applied to the model, a routine was included to detect the completion of this transient period (see sec. 6.1).

FTIR measurements were triggered by the steep potential change at the beginning of each period, effectively recording a sequence of spectra which were then used to reconstruct IR absorbance changes as a function of time.

The spectra were recorded at a square wave potential pulse frequency of 100 Hz. The spectrum of the initial time slide of a period, representing the fully reduced state of CcO, hereby serves as a reference and is subtracted from the spectra of the subsequent time slides. As a result, these difference spectra were used by Schwaighofer et al. [59] to calculate absorbances. This process is illustrated in the middle part of fig. 5.14.

The idea behind this method is to gain insight into the kinetics of the redox centers of the probe by plotting absorbance vs. time for each redox center (lower part of fig. 5.14). Thereby, for practical reasons, it was chosen to map the change of absorbances over time of specific bands (each characterized by a particular peak wavenumber $\hat{\nu}$) associated with specific redox centers. This is done by plotting the change of the area between the wavenumber axis and the band curve (called *band area*) in reference to the band area of the spectrum of the first

5. Kinetic Model of CcO

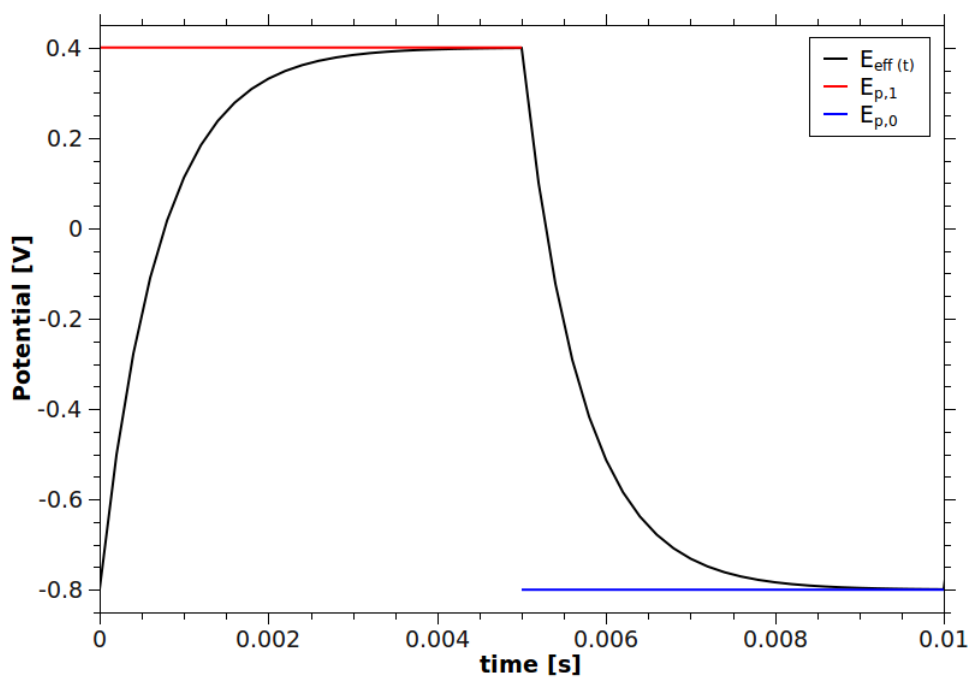


Figure 5.15.: Applied and effective Potentials

time slide. For the first three centers of the enzyme, fig. 7.1 depicts the change of band area vs. time, stemming from absorbance spectra of three particularly well suited wavenumbers associated with the respective centers.

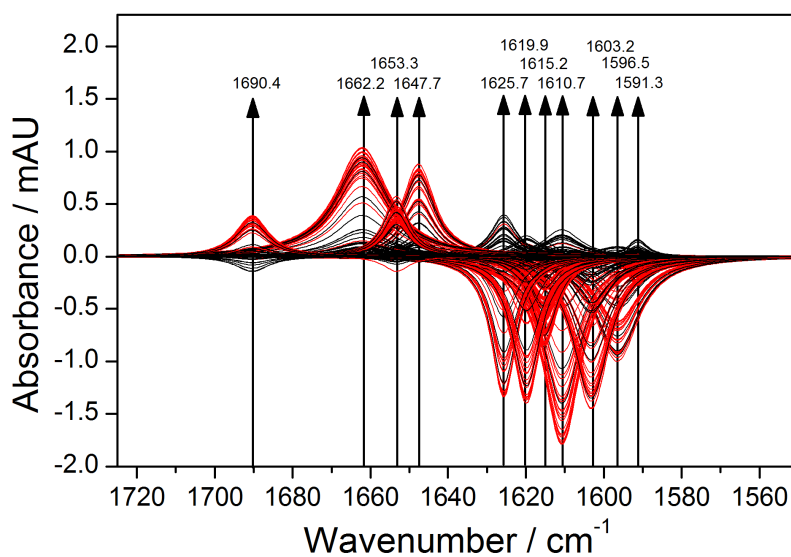


Figure 5.16.: Absorbance spectra obtained from tr-ATR-SEIRAS measurements^a. Red curves represent spectra obtained during the first half-periods of the potential pulse, black curves depict spectra obtained during the second half-periods. From: Schwaighofer et al. [59], p. 2710.

^aNote: Absorbance is a dimensionless quantity. It is, however, sometimes stated in terms of “absorbance units” (AU).

5. *Kinetic Model of CcO*

6. Implementation

The kinetic model described in the previous sections was implemented using the popular numerical computations engine MATLAB R2010b. The master equation 5.1, which is at the core of our model, can be solved using the standard variable step-size ODE solver `ode15s`. The solution vector yields the time evolution of the probabilities of each one of the 16 conformational states of the protein. As described in section 5.2.4, one can then calculate the absorbance changes for each one of the protein's centers in order to predict the band area change curves derived from tr-ATR-SEIRAS measurements. The solution of the master equation and thus everything derived from it, however, depends on a multitude of unknown parameters (tab. 6.1 and tab. 7.3). In section 6.3 we will discuss in detail how we attempted to gain sensible predictions for some of those parameters via Bayesian estimation techniques (see chapter 3).

We will start with discussing the main parts of the MATLAB source code, effectively implementing the kinetic model.

6.1. Function Declarations

At first, the electron uptake and exchange between the centers is defined in the form of [anonymous functions](#) [43] with the function handles `j_up` and `j_exHex`, equivalent to eq. 5.15 and eq. 5.11

Listing 6.1 Excerpt from `IRsumosqNO.m` (cf. p. 93)

```
%% Functions
%Midpointpotential
EM=@(pH) Eo_1+(R*temp/F).*log((1+10.^(pKr1-pH))./(1+10.^(pKo1-pH)));
%Electron uptake from the Electrode
j_up=@(P1,P2,kke,phi_eff,ph1)...
    kke.*(P1.*exp(F.*(EM(ph1)-phi_eff)./(2*R*temp)))...
    -P2.*exp(-F.*(EM(ph1)-phi_eff)./(2*R*temp)));
%Electron Exchange with Protonation
j_exHex=@(P1,P2,kk,E01,E02,pKr1,pKo1,pKr2,pKo2,ph1,ph2)...
    kk.*(P1-P2.*exp(F.*(E01-E02)./(R*temp)))...
    .*(1+10.^(pKr1-ph1)).*(1+10.^(pKo2-ph2))...
    ./((1+10.^(pKo1-ph1)).*(1+10.^(pKr2-ph2)));
```

As we can see, these functions depend on a large number of parameters. These are the pK^s values for the oxidised and reduced states of all four centers pK_{ri} , pK_{oi} , the electron uptake and transfer rates kke , kk as well as the standard redox potentials E_{0i} . All unknown parameters of the kinetic model as well as the parameters needed for absorbance change

6. Implementation

calculations are listed in the tables 6.1 and 7.1.

The function $USF(t)$ represents the consecutive charging and “discharging” of the electrode surface as described in eq. 5.35.

Listing 6.2 Excerpt from IRsumosqNO.m (cf. p. 93)

```
%VOLTAGE MINIMUM [V]
U1=-0.8
%VOLTAGE MAXIMUM [V]
U2=0.4
%NUMBER OF CYCLES
D1=8;
%FREQUENCY [Hz]
freq=100;
T=1/freq;
%POTENTIAL DIFFERENCE
dU=U2-U1;
```

Listing 6.3 Excerpt from IRsumosqNO.m (cf. p. 93)

```
dUa=U2-dU*exp(-(T/2)/tau)-U1;

USF=@(t) (U2-dU*exp(-(t)/tau)).*(t<=(T/2))+...
          (U1+dUa*exp(-(t-T/2)/tau)).*(t>(T/2) & t<=T);

for q=1:D1

    dUn=U2 - USF(q*T);
    dUn2=(U2-dUn*exp(-(T/2)/tau))-U1;
    USF=@(t) USF(t) + (U2-dUn*exp(-(t-q*T)/tau)).*(t>(q*T) & ...
        t<=((2*q+1)*T/2))+...
        (U1+dUn2*exp(-(t-(2*q+1)*T/2)/tau)).*(t>(2*q+1)*T/2 & ...
        t<=(q+1)*T);
    dUn2=USF((2*q+1)*(T/2))-U1;

end
```

Since there is in general—depending on τ —a certain transient interval until the effective potential reaches periodicity (See fig. 6.1), a routine was implemented to calculate the number of cycles necessary to overcome this transient period. This is done through checking the ratio of the absolute value of the potential minimum of the current cycle to that of the previous cycle. Once this ratio is greater than 0.999, the effective potential change is considered to be periodic.

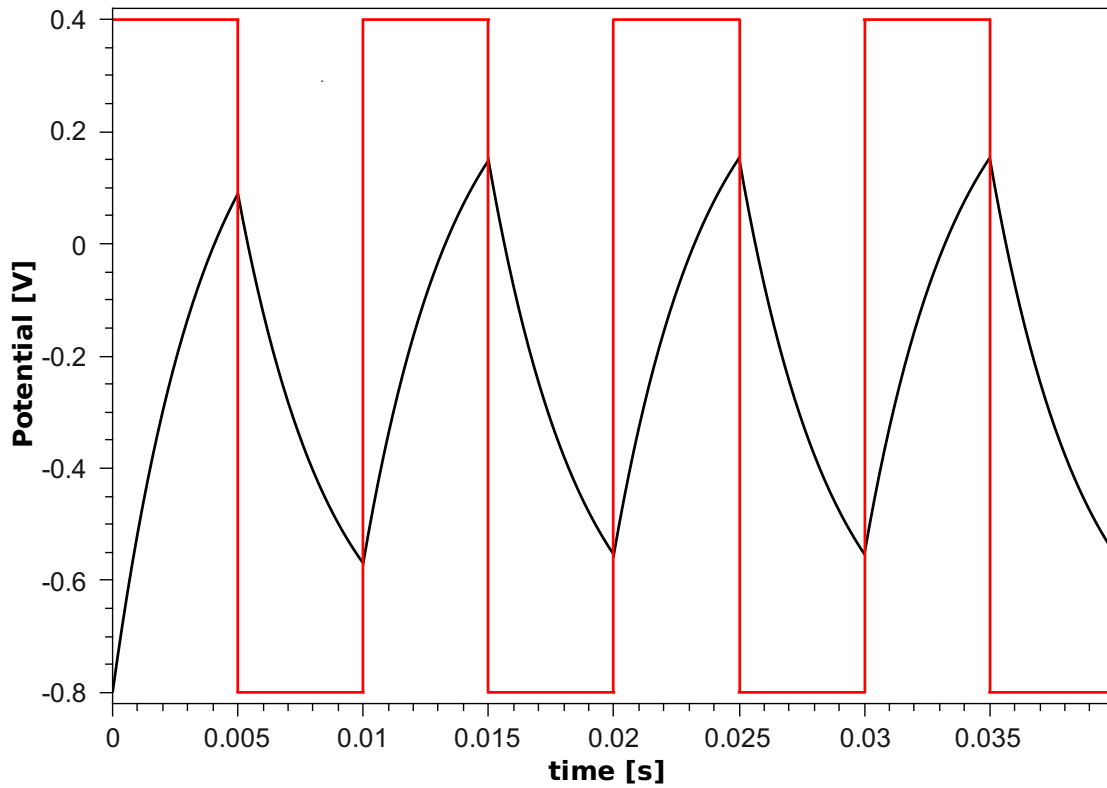


Figure 6.1.: Transient interval of the effective potential for a theoretical $\tau = 3.5$ ms

Listing 6.4 Excerpt from IRsumosqNO.m (cf. p. 93)

```
%calculate minimum number of cycles needed
Min(1)=-0.8;
D1_stat=0;
for ug=1:D1

Min(ug+1)=Ua(2*ug);
if abs(Min(ug+1))/abs(Min(ug))> 0.999
    A=ug+1;
else A=0;
end
D1_stat=[D1_stat; A];
end
D1_stat=D1_stat(D1_stat~=0);
fprintf('\nD1_stat(2): %f \n', D1_stat(2));
```

Note: For the range of τ values considered in the MCMC Simulation described later on, eight cycles proved to be sufficient to get a periodic potential. The number of cycles was therefore set to $D1=8$ in every iteration.

6. Implementation

6.2. Solving the Master Equation

We implement the Master equation 5.1 with the help of the functions defined in section 6.1. A graphical representation of all possible states and the connections between them can be seen in fig. 5.4.

Listing 6.5 Excerpt from IRsumosqNO.m (cf. p. 93)

```
%pH
pH=@(P) pH_N;

%Transitions between the 16 connected states

j_12=@(P,t) j_up(P fo(P),P(1),k(6),USF(t),pH(P));
j_23=@(P) ...
    j_exHex(P(1),P(2),tr12,k(1),k(2),k(13),k(14),k(15),k(16),pH(P),pH_N);
j_34=@(P) ...
    j_exHex(P(2),P(3),tr23,k(2),k(3),k(15),k(16),k(17),k(18),pH_N,pH_N);
j_45=@(P) ...
    j_exHex(P(3),P(4),tr34,k(3),k(4),k(17),k(18),k(19),k(20),pH_N,pH_N);
j_36=@(P,t) j_up(P(2),P(5),k(6),USF(t),pH(P));
j_67=@(P) ...
    j_exHex(P(5),P(6),tr23,k(2),k(3),k(15),k(16),k(17),k(18),pH_N,pH_N);
j_47=@(P,t) j_up(P(3),P(6),k(6),USF(t),pH(P));
j_78=@(P) ...
    j_exHex(P(6),P(7),tr34,k(3),k(4),k(17),k(18),k(19),k(20),pH_N,pH_N);
j_58=@(P,t) j_up(P(4),P(7),k(6),USF(t),pH(P));
j_79=@(P) ...
    j_exHex(P(6),P(8),tr12,k(1),k(2),k(13),k(14),k(15),k(16),pH(P),pH_N);
j_810=@(P) ...
    j_exHex(P(7),P(9),tr12,k(1),k(2),k(13),k(14),k(15),k(16),pH(P),pH_N);
j_910=@(P) ...
    j_exHex(P(8),P(9),tr34,k(3),k(4),k(17),k(18),k(19),k(20),pH_N,pH_N);
j_1011=@(P) ...
    j_exHex(P(9),P(10),tr23,k(2),k(3),k(15),k(16),k(17),k(18),pH_N,pH_N);
j_912=@(P,t) j_up(P(8),P(11),k(6),USF(t),pH(P));
j_1213=@(P) ...
    j_exHex(P(11),P(12),tr34,k(3),k(4),k(17),k(18),k(19),k(20),pH_N,pH_N);
j_1013=@(P,t) j_up(P(9),P(12),k(6),USF(t),pH(P));
j_1314=@(P) ...
    j_exHex(P(12),P(13),tr23,k(2),k(3),k(15),k(16),k(17),k(18),pH_N,pH_N);
j_1114=@(P,t) j_up(P(10),P(13),k(6),USF(t),pH(P));
j_1415=@(P) ...
    j_exHex(P(13),P(14),tr12,k(1),k(2),k(13),k(14),k(15),k(16),pH(P),pH_N);
j_1516=@(P,t) j_up(P(14),P(15),k(6),USF(t),pH(P));
```

Since at $t = 0$ the electrode potential is at -0.8 V, the enzyme is thought to be in a fully reduced state. We therefore set the vector of the initial occupation probabilities $P(t = 0)$ for all possible redox states to

6.2. Solving the Master Equation

Listing 6.6 Excerpt from IRsumosqNO.m (cf. p. 93)

```
P0 = [0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;1];
```

With the transition functions defined above, the master equation can finally be written as

Listing 6.7 Excerpt from IRsumosqNO.m (cf. p. 93)

```
Pdot =@(t,P) [
    j_12(P,t)-j_23(P);...
    j_23(P)-j_34(P)-j_36(P,t);...
    j_34(P)-j_45(P)-j_47(P,t);...
    j_45(P)-j_58(P,t);...
    j_36(P,t)-j_67(P);...
    j_67(P)+j_47(P,t)-j_78(P)-j_79(P);...
    j_78(P)+j_58(P,t)-j_810(P);...
    j_79(P)-j_912(P,t)-j_910(P);...
    j_910(P)+j_810(P)-j_1011(P)-j_1013(P,t);...
    j_1011(P)-j_1114(P,t);...
    j_912(P,t)-j_1213(P);...
    j_1213(P)+j_1013(P,t)-j_1314(P);...
    j_1314(P)+j_1114(P,t)-j_1415(P);...
    j_1415(P)-j_1516(P,t);...
    j_1516(P,t)];
```

We solve this system of differential equations using the solver `ode15s`. It is part of the standard MATLAB ODE suite and proved to be the most efficient one for this particular problem. According to Shampine and Reichelt [61], `ode15s` “[...] is a quasi-constant step size implementation in terms of backward differences” (p. 4) of Klopfenstein-Shampine Numerical Differentiation Formulas (NDFs).

Solving this system of ODEs proved to be computationally expensive for lower tolerances than the default of (`'AbsTol', 1e-6, 'RelTol', 1e-3`).

MATLAB provides error control properties in the form of *relative* and *absolute* tolerances. According to the MATLAB documentation, section “ODESET”, a “local error” e in the i -th component is estimated at each step. A constraint for this “local error” is then defined in terms of the relative `RelTol` and absolute `AbsTol` tolerances [44]:

$$|e(i)| \leq \max(\text{RelTol} \cdot |y(i)|, \text{AbsTol}(i)). \quad (6.1)$$

To illustrate this concept of “local error” we briefly want to discuss some of the basics and terminology concerning the numerical solution of a first-order differential equation

$$y'(t) = f(t, y(t)), \quad (6.2)$$

where we call $y(t)$ the analytical or global solution. For a distinct point in time t_n and a short time interval Δt , the fundamental theorem of calculus states that the solution $y(t_n + \Delta t)$ can be written as

$$y(t_n + \Delta t) = y(t_n) + \int_{t_n}^{t_n + \Delta t} f(t, y(t)) dt. \quad (6.3)$$

6. Implementation

The main idea is now to approximate the definite integral numerically through approximating the function $f(t_n, y(t_n))$ with help of a polynomial $P(t)$ (see Shampine et al. [60], p. 41ff). One therefore expands $f(t_n, y(t_n))$ ¹ into a Taylor series around t_n (see, e.g., Gupta et al. [26], p. 8):

$$y(t_n + \Delta t) = y(t_n) + \int_{t_n}^{t_n + \Delta t} dt \sum_{k=0}^{\infty} \frac{(t - t_n)^k}{k!} \left(\frac{d^k}{dt^k} f(t, y(t)) \right)_{t=t_n}. \quad (6.4)$$

Evaluating the integral yields

$$y(t_n + \Delta t) = y(t_n) + \sum_{k=1}^{\infty} \frac{(\Delta t)^k}{k!} \left(\frac{d^{k-1}}{dt^{k-1}} f(t, y(t)) \right)_{t=t_n}. \quad (6.5)$$

Hence, for a given (exact) initial condition $y(t = 0) = y_{t_0}$, we get a method for calculating an approximate solution y_1 at $t_0 + \Delta t$, for example if we only use the linear term of the Taylor expansion

$$y_1 = y_{t_0} + \Delta t f(t_0, y_{t_0}). \quad (6.6)$$

Having obtained a numerical approximation y_n at t_n , the **local solution** $u(t)$ of eq. 6.2 is then defined as the initial value problem (see Shampine et al. [60], p. 40)

$$u'(t) = f(t, u(t)) \quad \text{with} \quad u(t_n) = y_n, \quad (6.7)$$

hence

$$u(t) = y_n + \int_{t_n}^t f(x, u(x)) dx, \quad (6.8)$$

where the integral is evaluated analytically. By using the Taylor expansion shown in eq. 6.5 up to a specific term to approximate the integral, we therefore obtain a numerical scheme that approximates the local solution 6.8 (and the global solution $y(t_{n+1})$) at t_{n+1} , e.g.

$$y_{n+1} = y_n + \Delta t f(t_n, y_n) \approx u(t_{n+1}), \quad (6.9)$$

with $t_{n+1} = t_n + \Delta t$.

The procedure shown in eq. 6.9 is generally known as (forward) Euler's method and we can see that it deviates from the Taylor expansion 6.5 after the first term $\Delta t f(t_n, y_n)$. In other words, the deviation (or error) is of $\mathcal{O}(\Delta t^2)$. We generally call a numerical scheme that approximates the solution of an ODE to be of p -th order if it is identical to the Taylor expansion 6.5 up until (including) the p -th term, where $p = 1, 2, \dots$ (Gupta et al. [26], p. 8). The **local error** at t_{n+1} , on the other hand, is defined as the deviation (Shampine et al. [60], p. 46)

$$\text{LE}_{n+1} = u(t_{n+1}) - y_{n+1}, \quad (6.10)$$

¹One could also directly expand the solution $y(t)$ into a Taylor series around t_n and evaluate the resulting expression at t_{n+1} to obtain the same result.

6.2. Solving the Master Equation

and thus of $\mathcal{O}((\Delta t)^{p+1})$, while the **global error** is defined as the deviation from the analytical solution evaluated at t_{n+1} ,

$$\text{GE}_{n+1} = y(t_{n+1}) - y_{n+1}. \quad (6.11)$$

Finally, the **local truncation error**, is defined as the deviation between the true difference between two consecutive solution points $y(t_{n+1}) - y(t_n)$ and the numerical discretization scheme that approximates the definite integral in eq. 6.3, evaluated at $(t_n, y(t_n))$, hence, in case of a one-step procedure like Euler's method (see Gupta et al. [26], pp. 7–8),

$$\text{LTE}_{n+1} = y(t_{n+1}) - y(t_n) - \Delta t f(t_n, y(t_n)). \quad (6.12)$$

To give a basic idea of how the quantities `AbsTol` and `RelTol` are to be understood in this context of error estimation, we follow Shampine et al. [60] (p. 46):

The standard ODE solver `ode45` estimates the local error at t_{n+1} with help of the difference of solutions computed by two numerical discretization schemes of order p and $p + 1$, respectively:

$$\begin{aligned} \text{EST}_{n+1} &= y_{n+1}^* - y_{n+1} \\ &= [u(t_{n+1}) - y_{n+1}] - [u(t_{n+1}) - y_{n+1}^*] \\ &= \text{LE}_{n+1} + \mathcal{O}(\Delta t^{p+2}), \end{aligned}$$

with y_{n+1}^* the solution computed by the numerical method of order $p + 1$ and y_{n+1} the solution computed by the numerical method of order p . In case of `ode45`, for example, explicit Runge-Kutta formulas of orders 4 and 5—known as the Dormand and Prince or “DOPRI5” [60] (p. 53) method—are used. According to Shampine et al. [60] (p. 46), the computation is then advanced with the higher-order solution y_{n+1}^* , if the following condition for the local error estimate is met (cf. Shampine et al. [60]: exercise 2.4, pp. 56–57):

$$\text{EST}_{n+1} \leq \text{RelTol} \times |y_{n+1}^*| + \text{AbsTol}. \quad (6.13)$$

If qualification 6.13 is not met, the ODE solver reduces the step size according to specific conditions and tries again. Reducing the step size will be perpetuated, if necessary, until a minimal step length is reached at which point the solver simply aborts the computation. In this situation, higher error tolerances could be applied or, as an alternative, a different solving algorithm could be tried altogether.

As we have already mentioned before, the ODE solver that we used in this thesis, `ode15s`, mainly uses numerical differentiation formulas of order k (NDFk) that are based on backward differentiation formulas of order k (BDFk), which themselves are implicit formulas that are often used to solve stiff problems. Optionally, the “pure” BDFk can be used. The solver generally can make use of NDFk and BDFk methods between the orders of 1 and 5. Using this nomenclature, the BDF1, for example, is the backwards Euler method (see Shampine et al. [60], p. 60),

$$y_{n+1} = y_n + \Delta t f(t_{n+1}, y_{n+1}). \quad (6.14)$$

6. Implementation

Regarding the BDFk and the NDFk, the local *truncation* error, rather than the local error is estimated by `ode15s` and compared to the relative and absolute tolerances. To give a simple example, in case of the BDF1, one can expand $y(t)$ into a Taylor series around t_{n+1} [60] (p. 61–62),

$$y(t) = y(t_{n+1}) + (t - t_{n+1})y'(t_{n+1}) + \frac{(t - t_{n+1})^2}{2}y''(t_{n+1}) + \dots, \quad (6.15)$$

and then insert $t = t_n$ to get

$$y(t_n) = y(t_{n+1}) - \Delta ty'(t_{n+1}) + \frac{(\Delta t)^2}{2}y''(t_{n+1}) - \dots$$

Comparison with the local truncation error of the BDF1 (eq. 6.14)

$$\text{LTE}_{n+1} = y(t_{n+1}) - y(t_n) - \Delta ty'(t_{n+1}) \quad (6.16)$$

yields [60]

$$\text{LTE}_{n+1} = -\frac{\Delta t^2}{2}y''(t_{n+1}) + \dots = -\frac{\Delta t^2}{2}y''(t_n) + \dots \quad (6.17)$$

The BDF1 uses linear interpolation between t_n and t_{n+1} to approximate the solution $y(t)$ at t_{n+1} . Similarly, according to Shampine et al. [60] (p. 60), the higher order BDFs make use of a polynomial $P(t)$ that interpolates $y_{n+1}, y_n, y_{n-1}, \dots, y_{n+1-k}$, where $P(t)$ is required to satisfy the ODE at t_{n+1}

$$P'(t_{n+1}) = f(t_{n+1}, P(t_{n+1})) = f(t_{n+1}, y_{n+1}).$$

According to Shampine et al. [60] (p. 73), the leading term of the local truncation error in case of the BDF1 (eq. 6.17), for example, is therefore estimated by `ode15s` with help of a quadratic polynomial $Q(t)$, which interpolates y_{n+1}, y_n, y_{n-1}

$$\text{EST} = -\frac{(\Delta t)^2}{2}Q''(t_n) \approx -\frac{\Delta t^2}{2}y''(t_n). \quad (6.18)$$

Approximative solution points that have been obtained in the past also play a role in the iterative solution of the algebraic equations that arise in the course of the evaluation of implicit schemes like eq. 6.14 (so-called “predictor-corrector” methods). Furthermore, adjusting of the order of the BDFk (NDFk) to the solution, based on estimates of the local truncation errors of formulas of different orders, enables the use of larger step sizes. For more details on this subject, we want to refer to Shampine et al. [60] (pp. 57–81) and Shampine and Reichelt [61].

Anticipating the tens of thousands of iterations that are usually conducted in MCMC simulations in the context of Bayesian parameter estimation (solving this system for every set of newly proposed parameters) we therefore set the following solver options:

6.2. Solving the Master Equation

Listing 6.8 Excerpt from IRsumosqNO.m (cf. p. 93)

```
options=odeset('AbsTol',1e-5,'RelTol', 1e-3);
[t,P] = ode15s(Pdot,zeit,P0,options);
```

In various test runs with different parameter vectors, this setting showed satisfying results (smooth solution curves) while keeping the computation time relatively low (see section 6.3.1). Once a fifteen-dimensional solution vector of the above system is obtained, we can calculate the occupation probability over time for the first (fully oxidized) state P_{fo} :

Listing 6.9 Excerpt from IRsumosqNO.m (cf. p. 93)

```
P_fo=1-sum(P(:,1:15),2);
P=[P_fo P(:,1:15)];
```

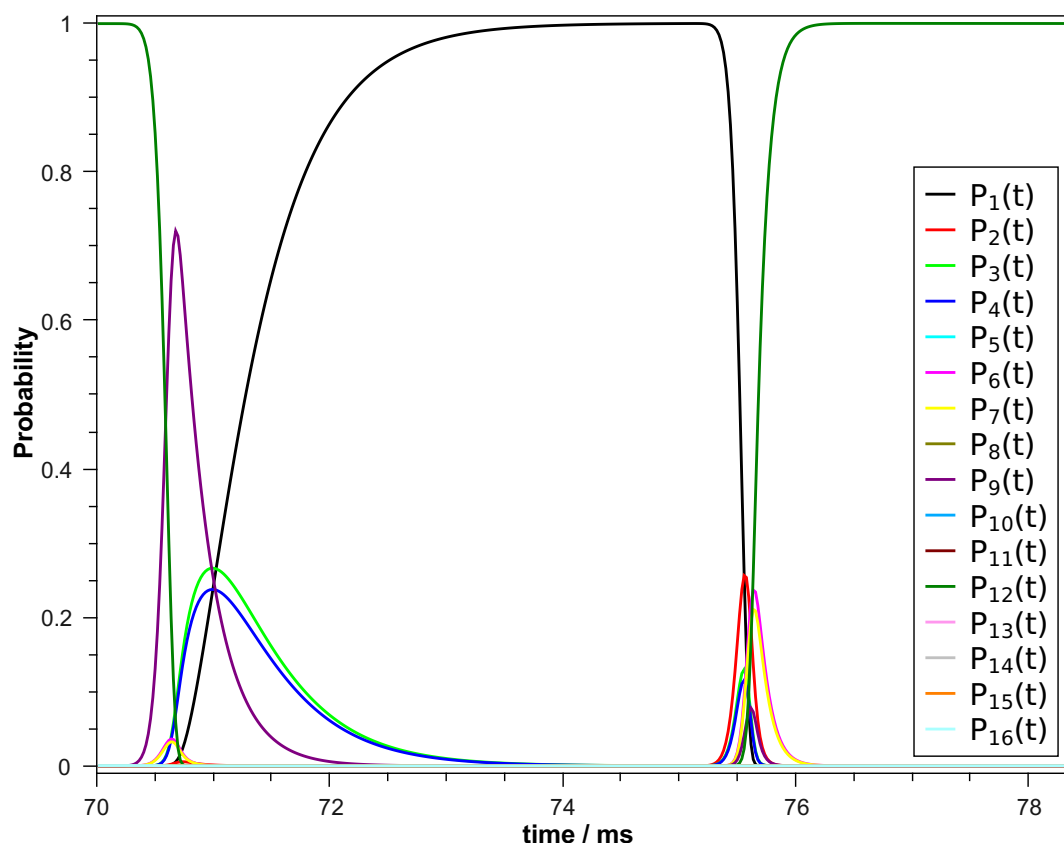


Figure 6.2.: Solutions to the Master Equation for the 8th cycle of the applied potential; $P_i(t)$ refers to the time-dependent occupation probabilities of the 16 conformational states displayed in fig. 5.4

In figure 6.2 a solution of 5.1, obtained with the settings mentioned above (parameter values used: see Table 7.3, column MLE) is displayed. The electrochemical system was simulated for 8 cycles under periodic potential change as described in section 6.1; figure 6.2 displays

6. Implementation

the solution for the eight cycle. Finally, the overall probabilities for the reduced state for each *center* can be calculated (See eq. 5.18):

Listing 6.10 Excerpt from IRsumosqNO.m (cf. p. 93)

```
p1r=P(:,2)+P(:,6)+P(:,7)+P(:,8)+P(:,12)+P(:,13)+P(:,14)+P(:,16);
p2r=P(:,3)+P(:,6)+P(:,9)+P(:,10)+P(:,12)+P(:,13)+P(:,15)+P(:,16);
p3r=P(:,4)+P(:,7)+P(:,9)+P(:,11)+P(:,12)+P(:,14)+P(:,15)+P(:,16);
```

Following eq. 5.20, the band area change for the first three centres can then be computed, given the constants ΔA_i and A_{i0} .

6.3. Bayesian Inference and MCMC Simulation

The above method of solving the master equation for the occupation probabilities of the various redox states of CcO was used in a broader context of a Bayesian parameter estimation. There are 23 unknown parameters of the electrochemical model of CcO in combination with the extension to simulate absorbance changes (as described in section 5.2.4). The author of this thesis subscribes to the opinion that it is in principle questionable to conduct a parameter fit/estimation for a model with such a large number of unknown parameters, especially given the relatively small number of data points available from the absorbance change measurements. Nevertheless, a Bayesian parameter estimation was conducted with the help of MCMC simulation techniques in order to obtain a set of parameters in best possible agreement with the data provided by Schwaighofer et al. [59].

6.3.1. Bayesian Parameter Estimation

Out of the 23 unknown parameters of the model, 14 were chosen as parameters of interest for a Bayesian parameter estimation routine. Table 6.1 lists all fixed parameters in the sense that they are not directly altered by means of a MCMC Simulation procedure.

Parameter Space

For our 14 parameters of interest listed in tab. 7.1, we restrict the parameter space to the following subset:

$$\bar{\Theta} = \{\theta \in \Theta | (\theta_{\min} \leq \theta \leq \theta_{\max}) \wedge (\theta_{pK_i^r} > \theta_{pK_i^o})\}. \quad (6.19)$$

This restriction arises due to

1. Choice of parameter boundaries $\theta_{\min}$, $\theta_{\max}$ to explore physically plausible regions of Θ . For example acid association and dissociation constants pK_i^r , pK_i^o are generally restricted to values between 0 and 14. The boundaries for all 14 parameters of interest listed in tab. 7.1 were determined through test runs in accordance with Dr. Naumann of AIT. Due to the sheer number of unknown parameters, the parameter space has

6.3. Bayesian Inference and MCMC Simulation

been significantly restricted for some parameters (e.g. τ). This was done in the hope of circumventing inefficiencies in the MCMC sampling process and to ensure faster convergence according to the criteria described in section 4.3. A drawback of this strategy, however, is the fact that we may be leaving significant portions of plausible regions of Θ unexplored.

2. $pK_i^r > pK_i^o$ due to electrostatic reasons.

Since a reduced center is more negative than an oxidised one, it follows that $[H^+r_i] > [H^+o_i]$ and $[H^+][r_i] < [H^+][o_i]$. The acid dissociation constants are defined as

$$pK_i^r = -\log_{10} \frac{[H^+][r_i]}{[H^+r_i]},$$

$$pK_i^o = -\log_{10} \frac{[H^+][o_i]}{[H^+o_i]}.$$

Hence, $\frac{[H^+][r_i]}{[H^+r_i]} < \frac{[H^+][o_i]}{[H^+o_i]}$ and $pK_i^r > pK_i^o$.

$k_{12} \text{ (s}^{-1}\text{)}$	23411
$k_{23} \text{ (s}^{-1}\text{)}$	5759469
$k_{34} \text{ (s}^{-1}\text{)}$	20300649
$\overline{\Delta A_{\text{Cu}_A}} \text{ (cm}^{-1}\text{)}$	adjusted
$\overline{\Delta A_{\text{Cu}_A,0}} \text{ (cm}^{-1}\text{)}$	-0.0014
$\overline{\Delta A_a} \text{ (cm}^{-1}\text{)}$	adjusted
$\overline{\Delta A_{a,0}} \text{ (cm}^{-1}\text{)}$	0.002
$\overline{\Delta A_{a_3}} \text{ (cm}^{-1}\text{)}$	adjusted
$\overline{\Delta A_{a_3,0}} \text{ (cm}^{-1}\text{)}$	0.0025

Table 6.1.: Fixed parameters; For a description of the adjustment procedure of $\overline{\Delta A_{\text{Cu}_A}}$, $\overline{\Delta A_a}$ and $\overline{\Delta A_{a_3}}$ see eq. 6.27

Likelihood Function

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{m}) = \frac{1}{(2\pi\tilde{\sigma})^{N/2}} \prod_{i=1}^N \exp\left(-\frac{1}{2\tilde{\sigma}^2}(m_i - s_i(\boldsymbol{\theta}))^2\right) \quad (6.20)$$

where $\mathbf{m}$ is the vector of measured data points, $\mathbf{s}(\boldsymbol{\theta})$ the vector of computer-generated points $\tilde{\sigma} = 9.6317 \times 10^{-6}$ and $N = 147$ the total number of fitted data points. The parameter $\tilde{\sigma}$ was derived from the mean square error of previous trial runs (see 6.26).

Prior Distribution

$$\pi(\boldsymbol{\theta}) = \begin{cases} 1 & \text{for } \boldsymbol{\theta} \in \bar{\Theta} \subset \Theta \\ 0 & \text{otherwise} \end{cases} \quad (6.21)$$

6. Implementation

Posterior Distribution

It follows from 3.11 and 6.21, that

$$\pi(\boldsymbol{\theta}|\mathbf{m}) = \frac{\mathcal{L}(\boldsymbol{\theta}; \mathbf{m})}{\int_{\Theta} \mathcal{L}(\boldsymbol{\theta}; \mathbf{m}) d\boldsymbol{\theta}} = \frac{\exp\left(\sum_{i=1}^N \left(-\frac{1}{2\hat{\sigma}^2} (m_i - s_i(\boldsymbol{\theta}))^2\right)\right)}{\int_{\Theta} \exp\left(\sum_{i=1}^N \left(-\frac{1}{2\hat{\sigma}^2} (m_i - s_i(\boldsymbol{\theta}))^2\right)\right) d\boldsymbol{\theta}} \quad (6.22)$$

and we define χ^2 , the sum of squared deviations between measured and computer-generated data points as

$$\chi^2 = \sum_{i=1}^N (m_i - s_i(\boldsymbol{\theta}))^2. \quad (6.23)$$

A vector of parameters $\boldsymbol{\theta}$ that minimizes χ^2 also maximizes our likelihood and posterior functions.

The posterior distribution function depends on the solution $\mathbf{s}(\boldsymbol{\theta})$ of our master equation 5.1, which we can only solve numerically. In the simplest case a posterior will have a single maximum which yields the most probable set of parameters. Here, however, it depends non-linearly on our 14 parameters of interest and will generally have multiple peaks and valleys of higher and lower probability. The main goal is to obtain a plausible set of parameters leading to computer simulated absorbance (band area) graphs which compare well to the measurements. In chapter 7 we will discuss estimates consisting of probable sets of parameters gained from our Bayesian model like the maximum likelihood (MLE) and maximum a-posteriori (MAP) estimate. Those are essentially our best bets for the “real” values of the parameters of interest. We obtain those estimates through use of MCMC simulation, in particular the Metropolis algorithm (section 4.1) with a Gaussian proposal density. After a sufficiently long run-time, we use the convergence criteria defined in section 4.3 to determine whether a sample approximately distributed according to this posterior density has been generated.

We will now discuss the details of this MCMC simulation procedure.

Gaussian Proposal Density

$$q(\boldsymbol{\theta}', \boldsymbol{\theta}) = \frac{1}{(2\pi)^{14/2} |S|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}' - \boldsymbol{\theta})^T S^{-1}(\boldsymbol{\theta}' - \boldsymbol{\theta})\right), \quad (6.24)$$

where the covariance matrix S is defined as

$$S = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{14}^2 \end{pmatrix}. \quad (6.25)$$

We implement the MCMC Simulation procedure with help of a MATLAB package written by Marko Laine [39] which we will slightly modify to fit our purposes. At its core, the

6.3. Bayesian Inference and MCMC Simulation

algorithm is realized as described in the following:

Listing 6.11 Excerpt from `mcmcrunPKNO.m` (cf. p. 93)

```
% sample new candidate from Gaussian proposal; increase stepsize for ...  
    about 15% of all draws  
dice=rand(1,1);  
fprintf('\n dice:%f\n', dice);  
u = randn(1,npar);  
if dice > 0.85  
  
    newpar=oldpar+u*2.5*R;  
else  
  
    newpar=oldpar+u*R;  
end
```

Here we introduce a change to the original code: In an effort to increase efficiency of the Gaussian random walk and decrease the chance of getting stuck in local minima, we increase the step size to $2.5 \times \sigma_i$ for 15 % of all draws (with $\mathbf{R}$ being a diagonal matrix with the step sizes σ_i in the main diagonal).

6. Implementation

We continue to implement our chosen prior—the indicator function of $\bar{\Theta}$ —and proceed to the actual Metropolis algorithm:

Listing 6.12 Excerpt from `mcmcrunPKNO.m` (cf. p. 93)

```
% reject points outside boundaries and if pKr < pKo
if any(newpar<low(parind)) || any(newpar>upp(parind))...
    || (newpar(1)<newpar(2)) || (newpar(3)<newpar(4))...
    || (newpar(5)<newpar(6)) || (newpar(7)<newpar(8))
    accept = 0;
    newprior = 0;
    tst      = 0;
    ssq1      = Inf;
    ssq2      = ssq;
    outbound = 1;

else
    outbound = 0;

    % calculate sum-of-squares
    ssq2 = ssq;          % old ssq
    ssq1 = ...
        ssqeval(ssqfun,ssqstyle,newpar,parind,value,local,data,modelfun);

    %tst= ratio of posteriors: new parameters / old parameters
    tst = exp(-0.5*( sum((ssq1-ssq2)./sigma2) ) );

    % negative log-likelihood
    llh= sum((ssq1-ssq2)./sigma2);

    %Metropolis algorithm
    if tst <= 0                %ssq2/ssq1
        accept = 0;
    elseif tst >= 1            %ssq2/ssq1
        accept = 1;
    elseif tst > rand(1,1)     %ssq2/ssq1
        accept = 1;
    else
        accept = 0;
    end
```

The two code fragments listed above are then executed for every iteration of the sampling process, realized in the form of a `for`-loop. They constitute the essential part of the MCMC simulation script `mcmcrunPKNO.m` used for this thesis.

6.3. Bayesian Inference and MCMC Simulation

Generally, there are three scripts used to perform MCMC simulation² and to generate a sample of parameter vectors distributed according to the 14-dimensional posterior density:

1. `startIRNO.m`, the executable script which calls the functions `IRsumosqNO.m` and `mcmcrunPKNO.m`.
2. `IRsumosqNO.m`, responsible for solving the master equation and calculation of the sum of squared deviations to the measured points for every new set of parameters generated by 3.
3. `mcmcrunPKNO.m`, the function responsible for the actual MCMC simulation procedure.

The main script `startIRNO.m` starts eight parallel MCMC sampling chains. As starting values for these eight chains, eight previously obtained sets of parameters leading to the smallest found sum of squared deviations to the measured points are used. One of these modes is also used to determine the free parameter $\tilde{\sigma}$ of the likelihood function:

$$\tilde{\sigma}^2 = \frac{\text{sum of squared deviations}}{\text{number of data points}}. \quad (6.26)$$

Listing 6.13 Excerpt from `startIRNO.m` (cf. p. 93)

```
%calculate sum of squares for set of parameters 'mode2'
ss0 = feval(@IRsumosqNO,mode2(1:14))

%(sum-of-squares)/(number of data points) = sigma^2
mse=ss0/147;
```

Furthermore, we use a routine to generate starting points in the *vicinity* of our previously found modes in an effort to prevent the chains from being stuck in possibly very pronounced local minima right at the beginning of the sampling procedure. To realize this, we take the parameter values of every mode and add 4 % Gaussian deviation, while checking if the newly generated starting values still lie within the boundaries.

²As we have mentioned before, the MCMC simulation relies on a MATLAB toolbox written by Marko Laine [39], whereas `mcmcrunPKNO.m` is a slightly modified version of the original function `mcmcrun.m` (see Listing 6.11).

6. Implementation

Listing 6.14 Excerpt from startIRNO.m (cf. p. 93)

```
%boundaries
bond= [3 11;1 5; 9 11; 2 6;3 9; 0 2; 0 4;...
       0 2;-0.25,-0.1;-0.35,-0.15;-0.2,-0.05;...
       -0.45,-0.25;200,700;0.00045,0.0006];

%generate new starting values within boundaries
while 1
for rn=1:14
    nmode1(rn)=mode1(rn)+0.04*mode1(rn).*randn(1,1);
    nmode2(rn)=mode2(rn)+0.04*mode2(rn).*randn(1,1);
    nmode3(rn)=mode3(rn)+0.04*mode3(rn).*randn(1,1);
    nmode4(rn)=mode4(rn)+0.04*mode4(rn).*randn(1,1);
    nmode5(rn)=mode5(rn)+0.04*mode5(rn).*randn(1,1);
    nmode6(rn)=mode6(rn)+0.04*mode6(rn).*randn(1,1);
    nmode7(rn)=mode7(rn)+0.04*mode7(rn).*randn(1,1);
    nmode8(rn)=mode8(rn)+0.04*mode8(rn).*randn(1,1);
    if (bond(rn,1)<nmode1(rn) && nmode1(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode2(rn) && nmode2(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode3(rn) && nmode3(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode4(rn) && nmode4(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode5(rn) && nmode5(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode6(rn) && nmode6(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode7(rn) && nmode7(rn)<bond(rn,2) && ...
        bond(rn,1)<nmode8(rn) && nmode8(rn)<bond(rn,2))

        isgood(rn)=1;
    else isgood(rn)=0;
        fprintf('outside bound');
    end
end

fprintf('%f\n',isgood)
    if (isgood==[1,1,1,1,1,1,1,1,1,1,1,1,1,1])
        break
    end
end

end
```

6.3. Bayesian Inference and MCMC Simulation

After specifying the sum-of-squares function (`ssqfun`), the maximum number of iterations (`nsimu`) and the algorithm we want to use (`'mh'` stands for “Metropolis-Hastings”), we can finally run eight chains in a parallel `for`-loop:

Listing 6.15 Excerpt from `startIRNO.m` (cf. p. 93)

```
model.ssqfun = @IRsumosqNO;
model.sigma2 = mse;

options.nsimu = 40000;
options.waitbar=0;
options.method = 'mh';
options.debug=1;
options.updatesigma = 0;
options.verbosity= 5;

%open pool with 8 workers, run 8 chains
matlabpool open 8
parfor rns=1:8
    abc(rns,:)=(rns+13*(rns-1)):(rns+13*rns);

    parpointer={
        params{abc(rns,:)}
    };
    mcmcrunPKNO(model,data,parpointer,options);

end

matlabpool close;
```

Note: `params` is a structure of the following form (See Table 7.1):

Listing 6.16 Excerpt from `startIRNO.m` (cf. p. 93)

```
params={
{'pKr1' , initial, lower bound, upper bound, optional prior ...
    parameter, ...
    step size}
{'pK01' , initial, lower bound, upper bound, optional prior ...
    parameter, ...
    step size}
{'pKr2' , initial, lower bound, upper bound, optional prior ...
    parameter, ...
    step size}
...
}
```

Each one of those eight MATLAB “workers” runs `mcmcrunPKNO`, which in turn calls `IRsumosqNO` for every newly proposed set of parameters. The master equation is then solved for a particular parameter vector in the fashion discussed in section 6.2 and the function `IRsumosqNO` returns the sum of the squared deviations of the computer generated band area change graphs with respect to the measurements. Apart from the 14 parameters that are al-

6. Implementation

tered at each iteration by the Metropolis algorithm, there are the two absorbance constants $\overline{\Delta A_i}$, $\overline{A_{i0}}$ for every center (see 5.2.4) listed in Table 6.1. In order to reduce the number of parameters and thus dimensions and complexity, $\overline{A_{i0}}$ remain fixed and $\overline{\Delta A_i}$ are adapted dynamically depending on to solution of the master equation for a particular set of parameters. We adjust the height of the computer generated band area change curve $\mathbf{s}(\boldsymbol{\theta})$ to the height of the measured curve $\mathbf{m}$ by setting $\overline{\Delta A}$ proportional to the ratio of the minima of both curves and make the ansatz:

$$\overline{\Delta A_i} = a_i \cdot \frac{\min(\mathbf{m}^{A_i})}{\min(\mathbf{s}(\boldsymbol{\theta})^{A_i})} + b_i. \quad (6.27)$$

The constants a_i , b_i are chosen arbitrarily and were determined during test runs, solving the master equation for different parameter combinations within the predefined boundaries:

$$\begin{aligned} \overline{\Delta A_1} &= 0.86 \cdot \frac{\min(\mathbf{m}^{\text{CuA}})}{\min(\mathbf{s}(\boldsymbol{\theta})^{\text{CuA}})} \\ \overline{\Delta A_2} &= 0.999 \cdot \frac{\min(\mathbf{m}^{\text{hemea}})}{\min(\mathbf{s}(\boldsymbol{\theta})^{\text{hemea}})} \\ \overline{\Delta A_3} &= \frac{\min(\mathbf{m}^{\text{hemea3}})}{\min(\mathbf{s}(\boldsymbol{\theta})^{\text{hemea3}})} + 0.0016. \end{aligned}$$

Listing 6.17 Excerpt from IRsumosqNO.m (cf. p. 93)

```
% find minima of absorbtion change curves (measurements)
Ma3=min(hemea3);
Mcua=min(cua);
Ma=min(hemea);

% Adapt solution curves p1r, p2r, p3r to measurements
% by changing absorbance constants:

% Delta A_1 (CuA):
k(7)=0.86*(Mcua/min(p1r((D1-1)*50+1:D1*50)-p1r((D1-1)*50+1)));
% Delta A_2 (heme a):
k(9)=0.999*(Ma/min(p2r((D1-1)*50+1:D1*50)-p2r((D1-1)*50+1)));
% Delta A_3 (heme a3):
k(11)=Ma3/min(p3r((D1-1)*50+1:D1*50)-p3r((D1-1)*50+1))+0.0016;
```

Total Running Time of the Simulation Procedure

As mentioned in section 6.2, it is necessary to solve the system of ODEs 5.1 for each newly drawn vector of parameters. In order to obtain smooth solution curves, tolerance settings near to or lower than the default tolerance of ('AbsTol', 1e-6, 'RelTol', 1e-3) are advisable.

In the course of several simulation test runs different combinations of AbsTol and RelTol were tried in order to obtain a near optimal setting requiring minimal total computation time

6.3. Bayesian Inference and MCMC Simulation

while yielding reasonably “good” solution curves. In other words, we tried to find tolerance settings which keep the *per iteration* computation time of `ode15s` as low as possible. At the same time, however, we require the average point wise deviation of a thereby obtained solution from more accurate numerical solutions achieved with the help of very low tolerance settings to be still negligibly small compared to the measurement error of each measured point.

Regarding the proposal density of the MCMC procedure, different “jump sizes” for each parameter of interest were tried out (the square roots of the diagonal entries of the covariance matrix 6.25) in addition to that, and a number of test runs were carried out, often over a period of several weeks, to determine which combination of σ_i and tolerance settings would result in a simulation procedure that is capable of exploring the parameter space within our boundaries in a feasible amount of time. If we conduct such a test run such with the settings `( 'AbsTol', 1e-7, 'RelTol', 1e-3 )`, for example, using a single Metropolis-Hastings sampling chain, we measure an average pure solving time of 4.384 minutes for each newly drawn trial parameter vector when looking at ten consecutive iterations. Hence, based on this number, we can extrapolate to around 30 days of total computation time on a Intel® Core™ i7-720QM 1.6 GHz CPU for only 10000 iterations and consider higher tolerance settings for the next test run.

In the end, the setting `( 'AbsTol', 1e-5, 'RelTol', 1e-3 )` was chosen, because the solutions obtained using these tolerances yielded results close enough to lower tolerance settings for several sets of test parameters while significantly reducing computation time: A total of around 320000 simulation points were obtained by eight parallel running Metropolis-Hastings sampling procedures on a Intel® Core™ i7 3.4 GHz quad-core CPU within a period of eight days.

6. *Implementation*

7. Results

Convergence

Using the between-within variance method as described in 4.3.2, scale reduction factors were monitored for 8 chains at around 40,000 Iterations each. The final scale reduction factors that were obtained for each parameter are listed in tab. 7.3. The time evolution of this diagnostic quantity is depicted in fig. 7.2.

Histograms and Marginal Posterior Densities

To approximate the marginal posterior densities for each parameter, the second halves of each of the eight chains were collected and treated as one sample of this density (around 160000 sample points in total). The resulting histograms are displayed at the end of this chapter.

The histograms were combined with box and whisker plots that show the 25 % and 75 % quantiles (box) as well the median (bold black line). The ends of the whisker represent the sample minimum and maximum.

The marginal maximum a posteriori estimate (MMAP) yields the most probable value of each parameter, separately obtained from the respective simulated marginal posterior densities, and is listed in tab. 7.3 :

$$\theta_k^{\text{MMAP}} = \arg \max_{\Theta} \pi(\theta_k | \mathbf{m}). \quad (7.1)$$

As credible intervals for our parameters, the 95 % highest posterior density regions (see section 3.3.3) were used. These HPD regions were calculated with help of the function “*emp.hpd()*” of the R package “*TeachingDemos*” [62]. The 95 % credible intervals are indicated by the green lines (lower and upper boundaries) in the respective box and whisker plots, as well as listed in the results table 7.3.

For parameters with marginal densities that resemble uniform distributions, however, highest marginal posterior density regions were not calculated. These consist of the pK^s values (except pK_4^r) and k_e .

7.1. Conclusion

Our prior density (eq. 6.21), defined on the restricted parameter space $\bar{\Theta}$ (eq. 6.19), incorporates pre-determined boundaries for each parameter of interest (tab. 7.1) which are based on previously gained experience: Cyclic voltammetry measurements [63], for example, helped

7. Results

to map out plausible ranges for the pK^s constants. Various test runs of the MCMC simulation also aided setting up boundaries as listed in tab. 7.1. They were obtained by excluding certain parameter combinations for physical reasons (see eq. 6.19) and recording pK^s intervals yielding better consistency between simulated and measured absorbance change curves. In the end, however, further refinement of those initial estimations that were incorporated into the prior function could not be accomplished with help of the Bayes Theorem 3.11. The resulting histograms, representing marginal posterior densities for those parameters, almost resemble uniform distributions and there is an absence of clearly distinguishable peaks. Hence, for a single pK^s parameter, the marginal posterior distribution apparently does not deviate much from the uniformly distributed prior and no new information was gained in the light of tr-ATR-SEIRAS measurements¹. Possible explanations for this result could be:

1. The prior distribution already contained most of the information that could be obtained for the pK^s parameters of our kinetic model in the light of the tr-ATR-SEIRAS measurements before we even fully considered it. Given the assumptions made on the probability model (e.g. the form of the likelihood function), the knowledge incorporated into the prior function and the new data from the measurements, the marginal posterior distribution does not provide a sharper measure of uncertainty² for those parameters compared to what was expected a priori.

For more definite inferences on the acid association/dissociation constants, the probability model needs to be reconsidered.

2. The histograms still do not represent the true marginal posterior densities for those parameters. A much longer runtime of the MCMC procedure is probably needed (more iterations).

For those reasons, interval estimators such as the highest posterior density intervals were not calculated.

The reduction potentials E_m^0 were expected to lie in the negative range around -200 mV, based on cyclic voltammetry [21] and computer modeling studies [57, 63]. Accordingly, interval boundaries in the vicinity of -200 mV were defined for all four E_m^0 , using crude estimations and plausibility considerations in consultation with Dr. Naumann of AIT. A uniform prior density was defined inside the predetermined boundaries. Using Bayes' theorem 3.11, the uncertainty reflected in the prior density could be narrowed down in the light of the new absorbance data, as can be seen in the histograms of E_1^0 (fig. 7.3), E_2^0 (fig. 7.4), E_3^0 (fig. 7.5) and E_4^0 (fig. 7.6).

Estimates for the time constant τ turned out to be lower than what was expected from previous efforts using IR absorbance data and yielding $\tau \sim 0.8$ ms (see Schach [56], p. 88). Values above $\tau = 0.7$ ms showed increasingly poor compliance with measured band area change curves and a trial run of $\tau > 0.7$ ms with different combinations of the other parameters within their boundaries did not seem to affect this result significantly. The marginal

¹With the exception of maybe pK_4^s . The resulting histogram 7.8 seems to indicate a higher probability in the upper half of the examined interval

²In the sense as it is described by Bernardo [7]

posterior density of τ (fig. 7.7) yielded a reduction of uncertainty compared to the uninformative prior with a clear peak at around 0.53 ms.

In summary, however, the maximum likelihood estimator (MLE) denotes our “best fit” in the sense of a total (including Cu_A , heme a and heme a_3) minimal sum of squared deviations of each computer-generated data point compared to the respective measurement. On the other hand, the histograms and 95 % credible regions in the Bayesian sense corner the most probable intervals for the parameters according to our computer model of CcO, at least within our pre-defined boundaries. The estimates of the reduction potentials E_m^0 are deemed to be the most important results of this study [59], as they seem to confirm a shift to negative potential values, in accordance with previous cyclic voltammetry measurements [21] and modeling studies [57, 63]. The set of curves for Cu_A , heme a and heme a_3 generated through MLE parameter values in comparison to the measured band area change graphs are depicted in fig. 7.1.

7.1.1. Fit Results: Tables and Histograms

θ	$[\theta_{\min}, \theta_{\max}]$	σ_i
$\text{p}K_1^r$	[3,11]	0.44
$\text{p}K_1^o$	[1,5]	0.2
$\text{p}K_2^r$	[9,11]	0.11
$\text{p}K_2^o$	[2,6]	0.2
$\text{p}K_3^r$	[3,9]	0.4
$\text{p}K_3^o$	[0,2]	0.13
$\text{p}K_4^r$	[0,4]	0.22
$\text{p}K_4^o$	[0,2]	0.15
E_1^0 (V)	[-0.25,-0.1]	$8.33 \cdot 10^{-3}$
E_2^0 (V)	[-0.35,-0.15]	$12 \cdot 10^{-3}$
E_3^0 (V)	[-0.2,-0.05]	$9 \cdot 10^{-3}$
E_4^0 (V)	[-0.45,-0.25]	$11.11 \cdot 10^{-3}$
k_e (s^{-1})	[200,700]	20
τ (ms)	[0.45,0.6]	$6.8 \cdot 10^{-6}$

Table 7.1.: Parameter boundaries and random walk step sizes

7. Results

θ	chain 1	chain 2	chain 3	chain 4	chain 5	chain 6	chain 7	chain 8
pK_1^r	5.337	6.9	6.497	7.768	4.566	6.728	7.001	6.476
pK_1^o	4.466	1.146	3.344	1.52	2.632	3.358	2.748	2.569
pK_2^r	9.802	9.331	9.486	9.236	10.01	10.81	10.29	10.34
pK_2^o	5.752	2.752	3.444	2.416	3.017	4.305	4.167	5.148
pK_3^r	5.814	6.811	4.621	3.802	6.261	6.1	6.66	6.414
pK_3^o	1.88	0.6029	0.8352	0.2928	0.53	1.68	0.655	1.366
pK_4^r	2.418	1.788	2.185	2.996	1.661	2.573	2.198	1.896
pK_4^o	1.14	1.802	0.2477	1.216	0.2514	0.7154	1.082	1.531
E_1^0 (V)	-0.1722	-0.1711	-0.1673	-0.1911	-0.1758	-0.1642	-0.1623	-0.1786
E_2^0 (V)	-0.2148	-0.1866	-0.2487	-0.219	-0.2202	-0.2647	-0.2503	-0.2479
E_3^0 (V)	-0.1309	-0.1177	-0.1314	-0.1297	-0.1312	-0.1248	-0.1155	-0.1311
E_4^0 (V)	-0.3875	-0.4043	-0.4077	-0.4279	-0.4054	-0.4268	-0.3197	-0.3959
k_e (s ⁻¹)	456.8	414.1	696.3	232.1	630.2	432	337.2	389.6
τ (ms)	0.514	0.534	0.589	0.468	0.588	0.532	0.52	0.548

Table 7.2.: Starting values

7.1. Conclusion

θ	MMAP	MLE	HPD _{min}	HPD _{max}	median	$\hat{R}$
pK_1^r	4.95	8.092	-	-	6.235	1.053502
pK_1^o	1.175	2.479	-	-	2.662	1.060104
pK_2^r	10.375	9.288	-	-	10.098	1.032156
pK_2^o	5.375	5.382	-	-	4.143	1.011017
pK_3^r	8.45	4.514	-	-	5.845	1.069696
pK_3^o	1.925	1.837	-	-	1.072	1.054094
pK_4^r	3.075	1.487	0.956	3.999	2.587	1.023582
pK_4^o	0.925	0.776	-	-	0.939	1.050650
E_1^0 (V)	-0.169	-0.19208	-0.2463	-0.144	-0.1763	1.041318
E_2^0 (V)	-0.2275	-0.2023	-0.3174	-0.1725	-0.2463	1.042709
E_3^0 (V)	-0.131	-0.1277	-0.19	-0.0959	-0.1362	1.009292
E_4^0 (V)	-0.4275	-0.4437	-0.4498	-0.2897	-0.3743	1.020213
k_e (s ⁻¹)	475	536.36	-	-	448.7	1.031495
τ (ms)	0.525	0.5504	0.465	0.588	0.530	1.072409
$\overline{\Delta A_1^*}$ (cm ⁻¹)	0.02219	0.02219	-	-	0.02219	-
$\overline{\Delta A_2^*}$ (cm ⁻¹)	0.03657	0.03285	-	-	0.03286	-
$\overline{\Delta A_3^*}$ (cm ⁻¹)	0.01983	0.01849	-	-	0.01849	-
χ^2 (cm ⁻²)	0.0034042	0.0014071	-	-	0.0014647	-
$\sqrt{\chi^2}/N$ (cm ⁻¹)	3.969 ·10 ⁻⁴	2.552 ·10 ⁻⁴	-	-	2.604 ·10 ⁻⁴	-

Table 7.3.: Results. Note: Absorbance parameters (eq. 5.23) marked with an asterisk are not part of the MCMC Simulation procedure but indirectly depend on the solution of the master equation at each iteration. See eq. 6.27 for details. $\sqrt{\chi^2}/N$ is the average deviation per data point 6.23

7. Results

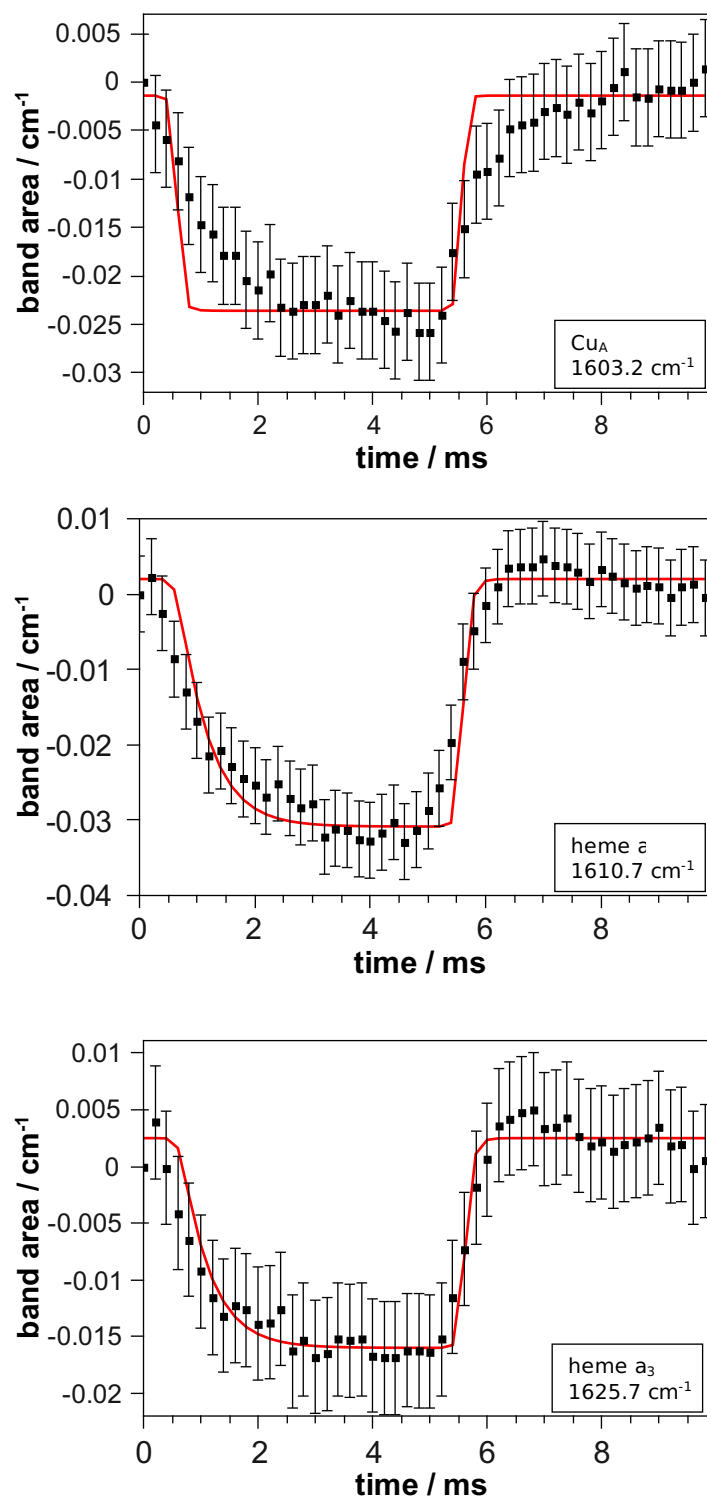


Figure 7.1.: Measured absorbance change plots (black squares) vs. computer-generated plots according to our model with the parameter values listed in the MLE column of tab. 7.3 (red); see Schwaighofer et al. [59], p. 5.

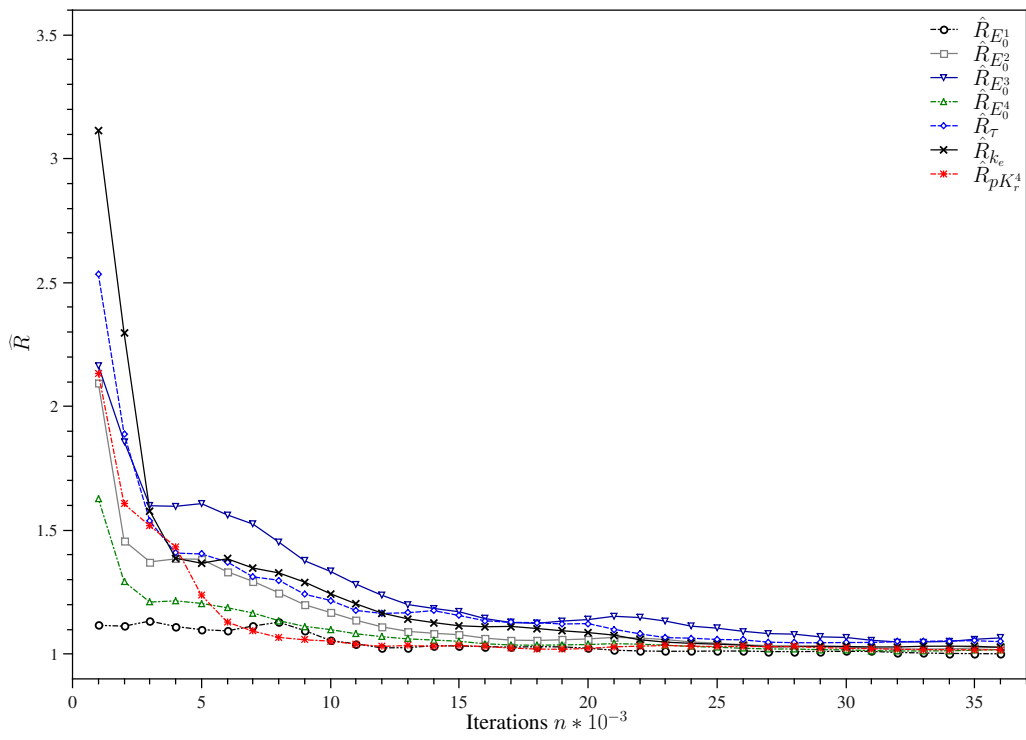


Figure 7.2.: Time evolution of the scale reduction factor $\hat{R}$ for the seven parameters of interest.

7. Results

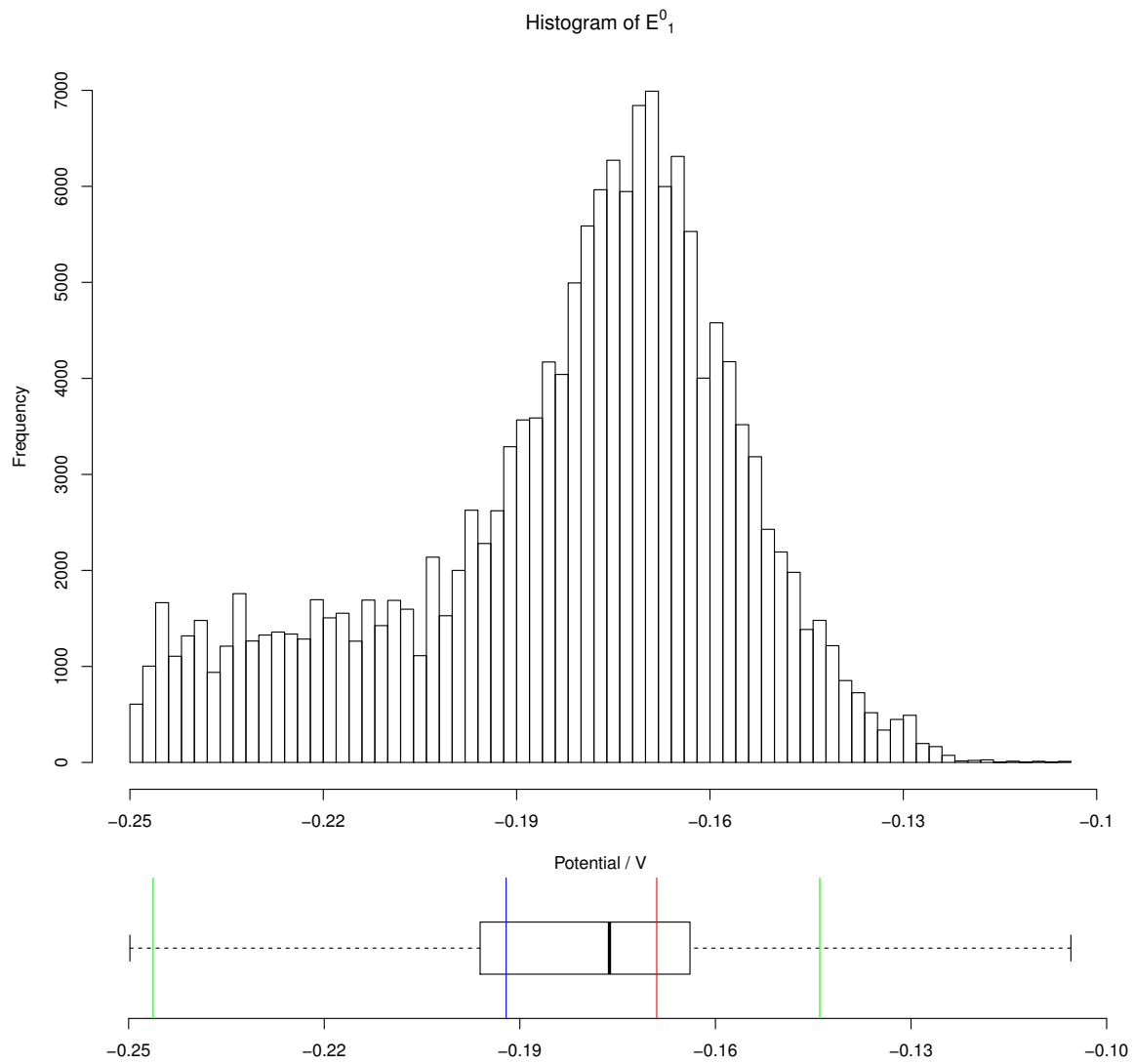


Figure 7.3.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

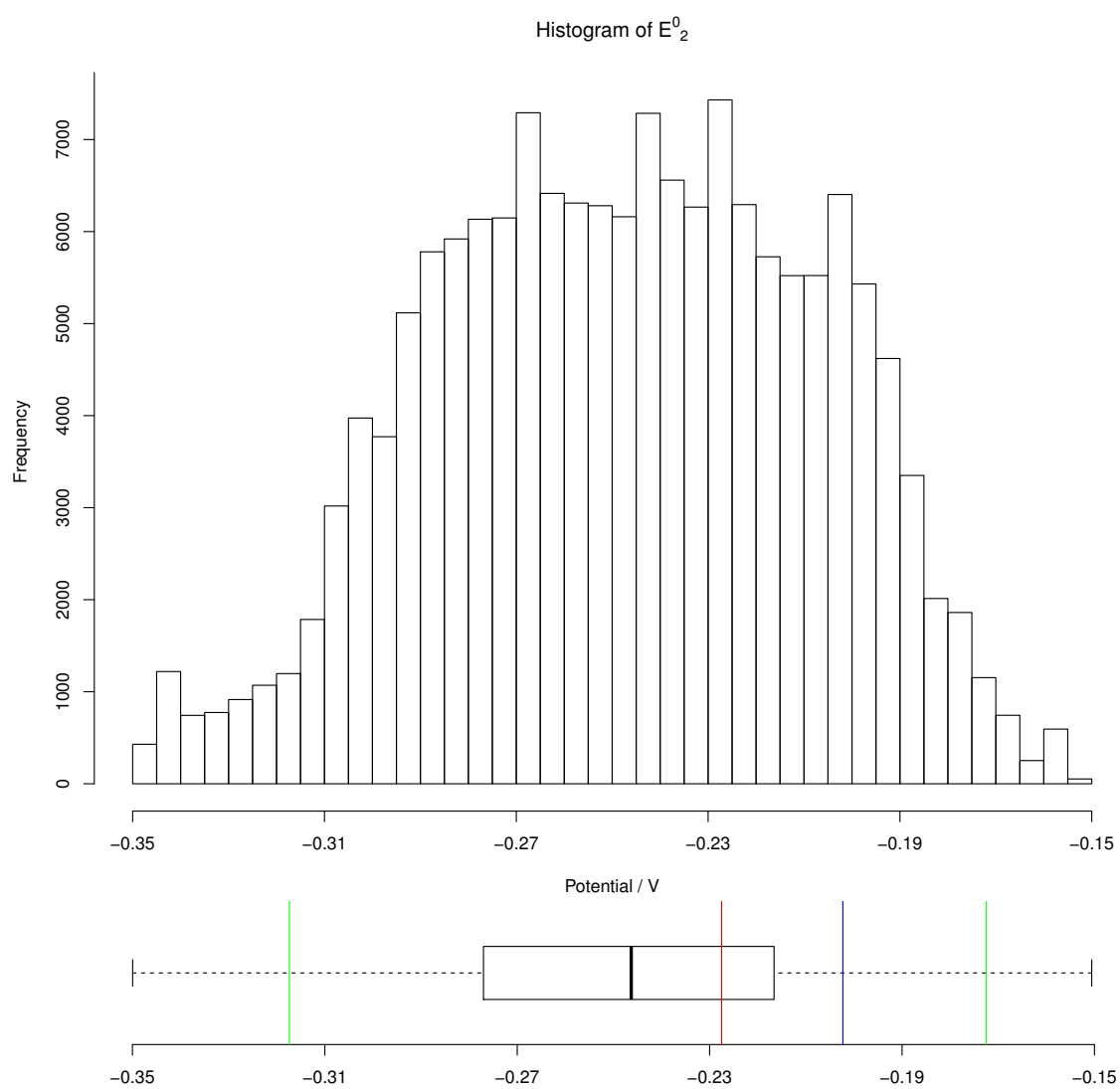


Figure 7.4.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

7. Results

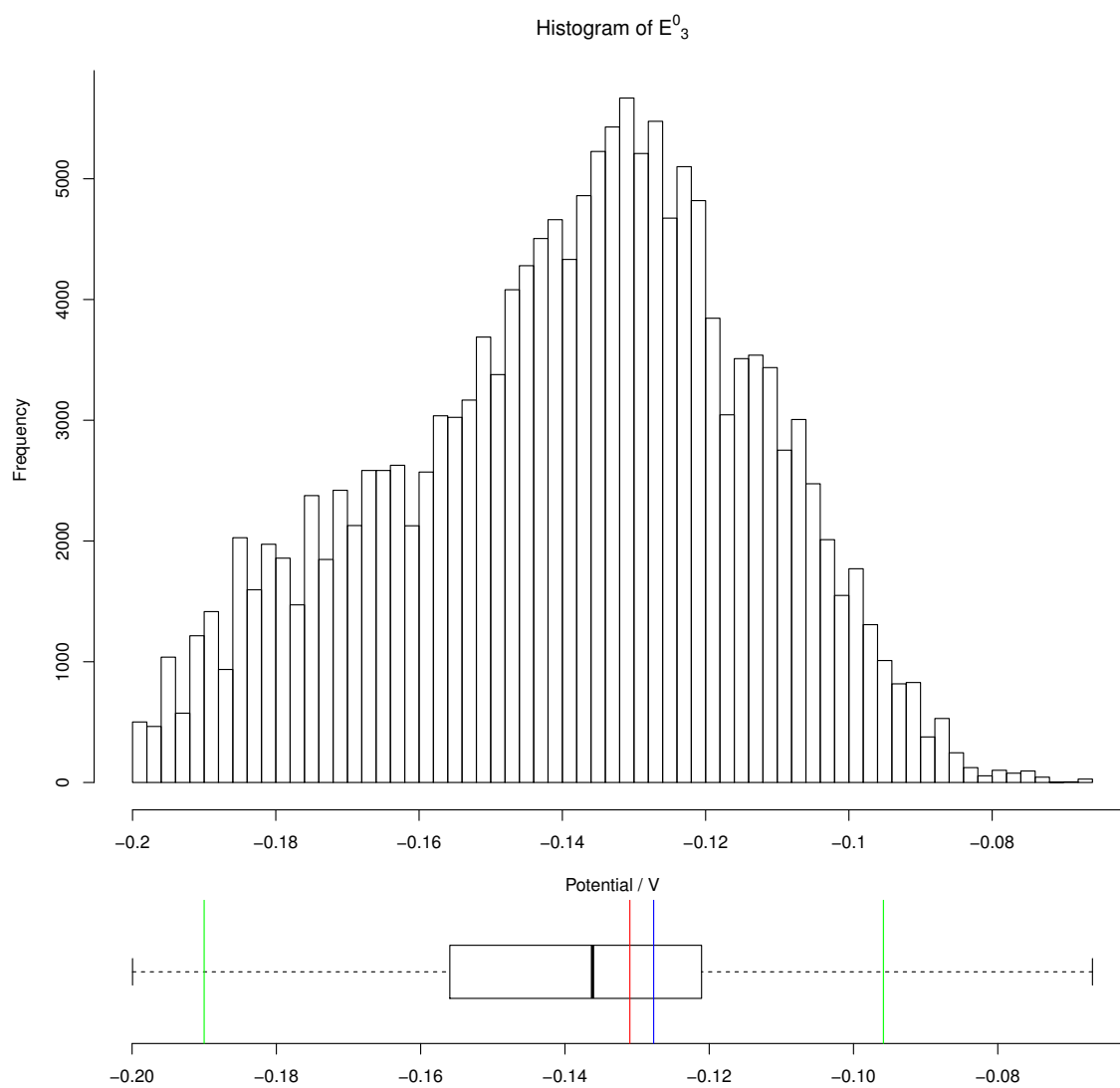


Figure 7.5.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

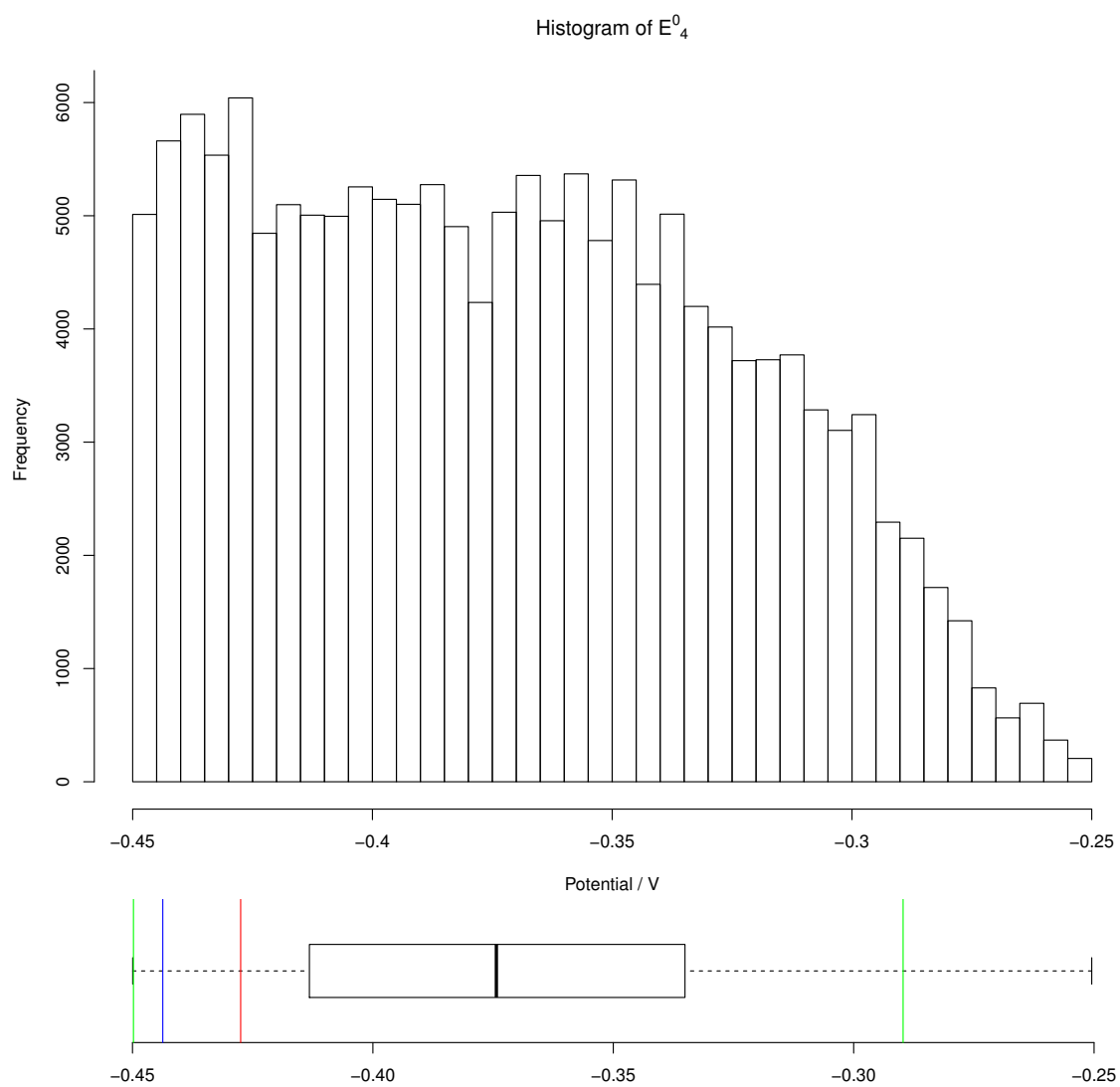


Figure 7.6.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

7. Results

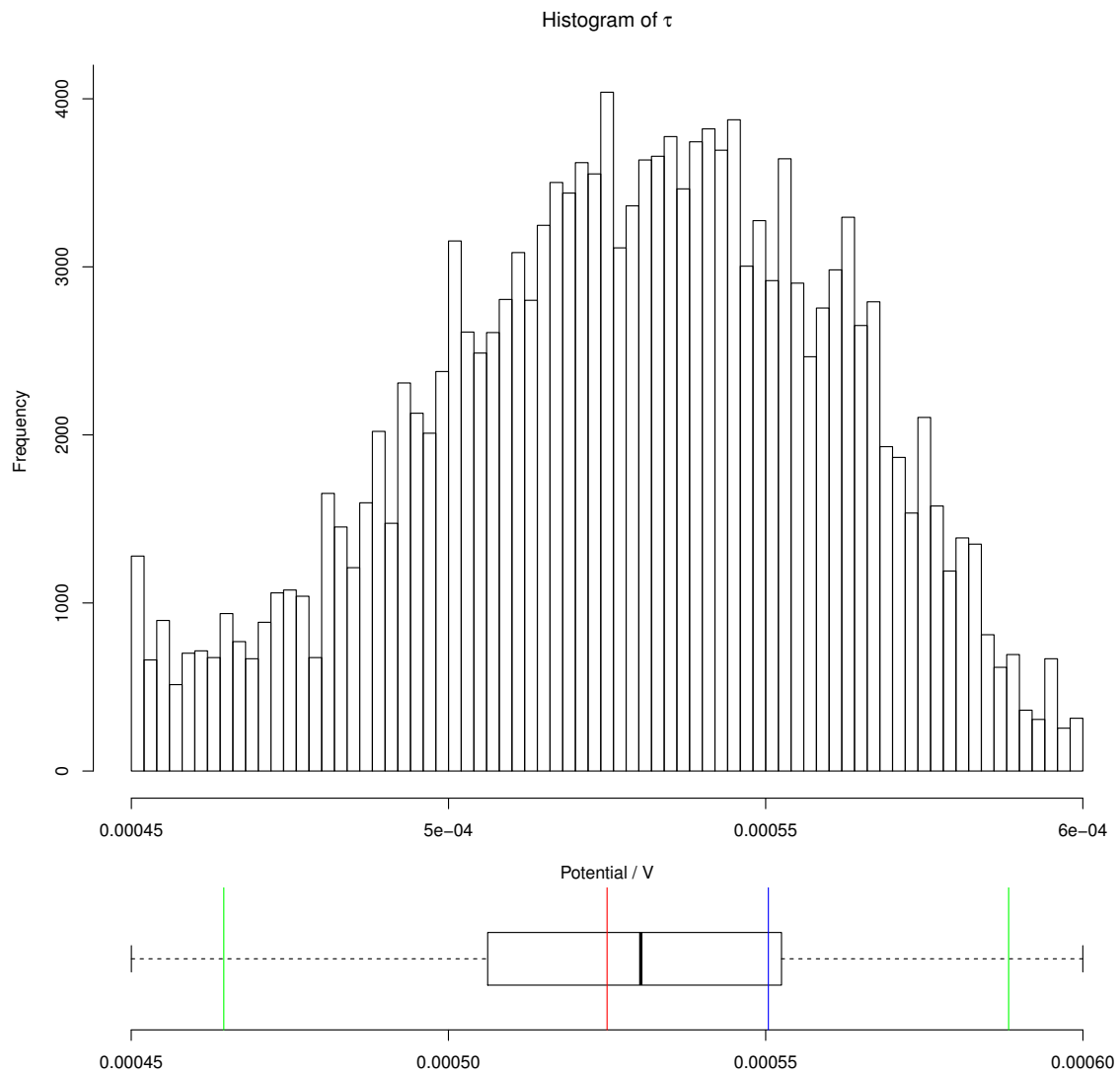


Figure 7.7.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

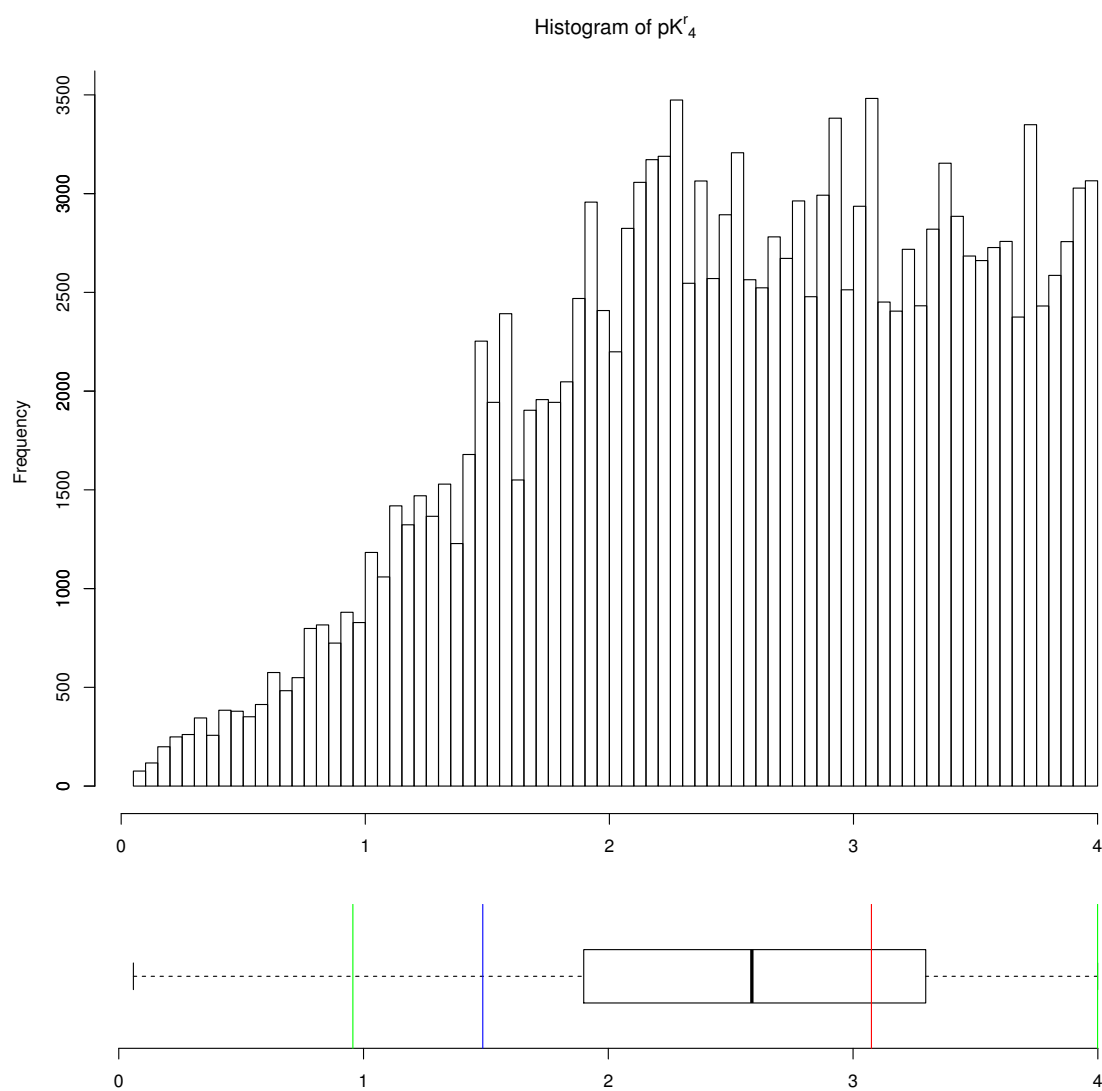


Figure 7.8.: Histogram and box and whisker plot. Red line: MMAP, blue line: MLE, green lines: HPD

7. Results

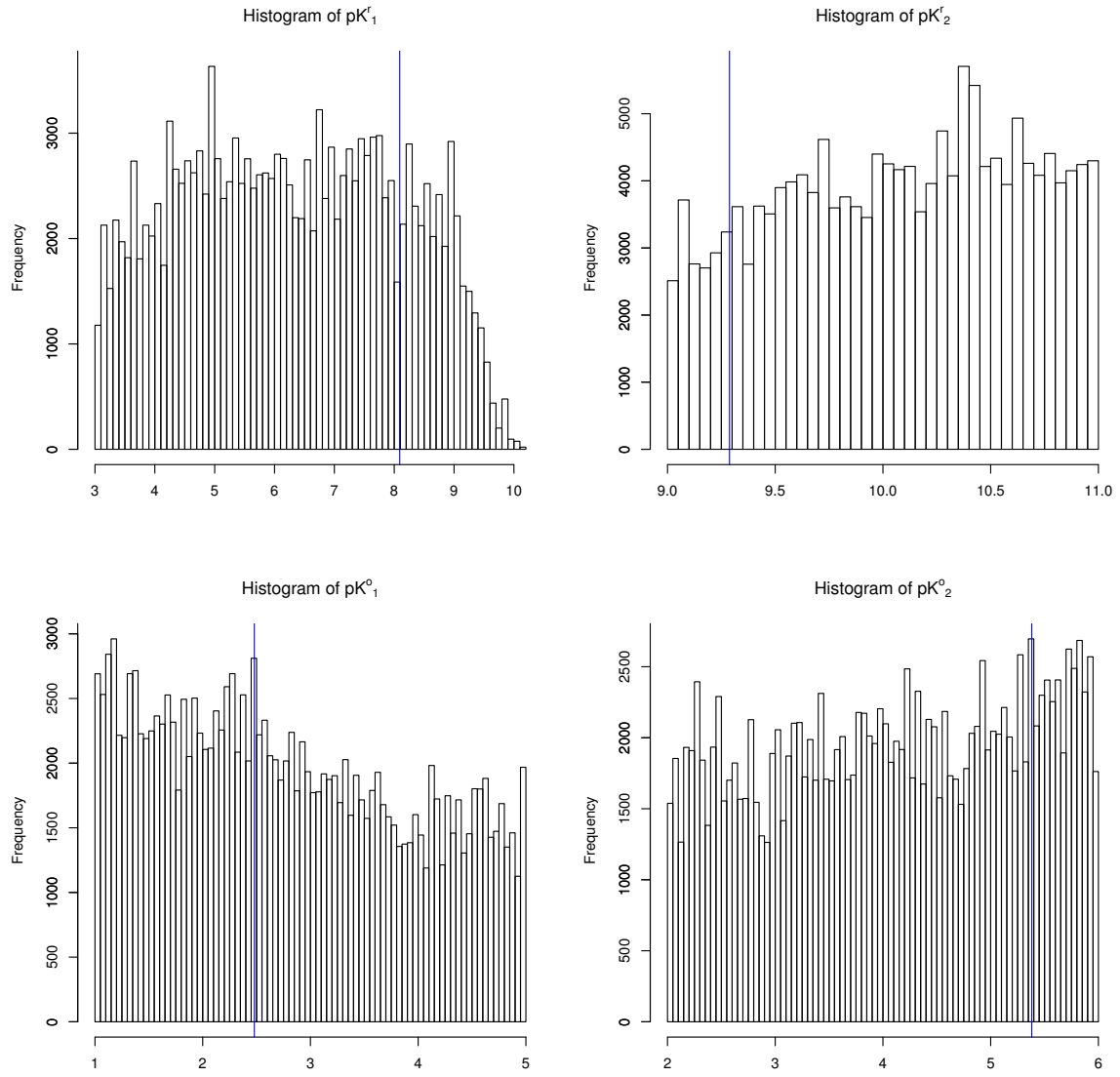


Figure 7.9.: Histograms of nuisance parameters. Blue line: MLE

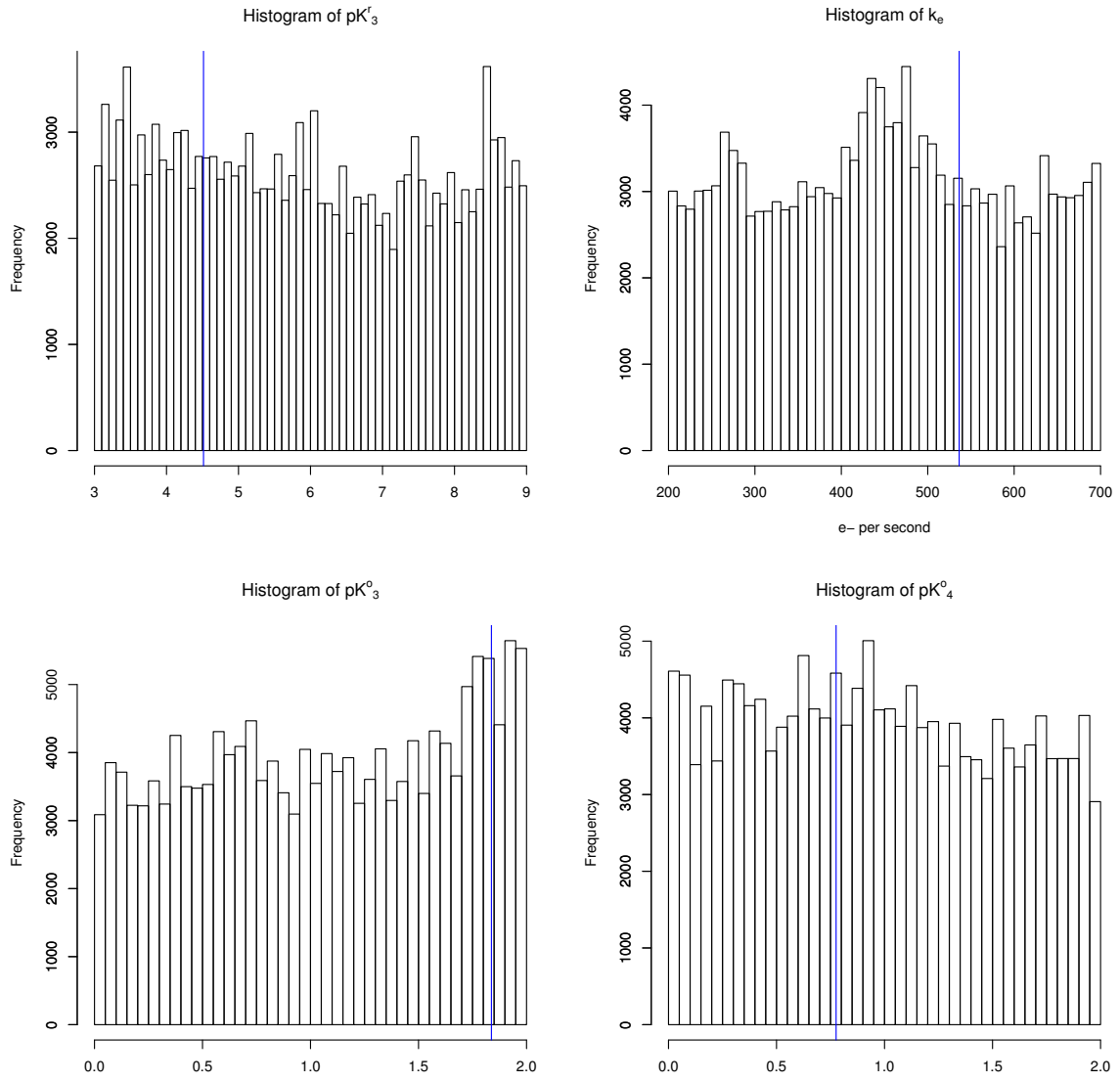


Figure 7.10.: Histograms of nuisance parameters. Blue line: MLE

7. Results

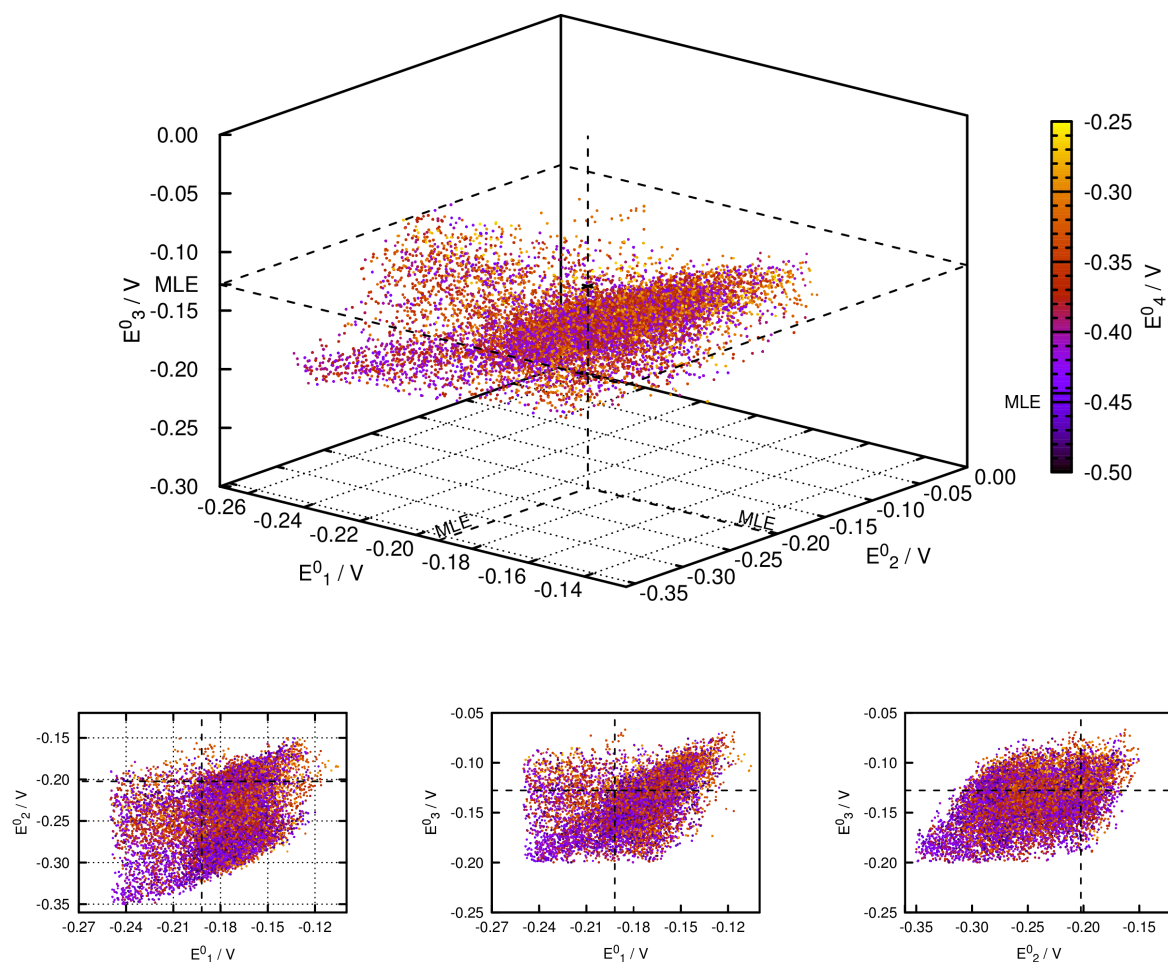


Figure 7.11.: Visual representation of the second half of the sample generated by the MCMC procedure showing only the sample draws of the four standard redox potentials E_i^0 . The sample draws of the redox potentials of the first three centers of CcO are plotted on the three (projected) spatial axes, while the sample draws of the standard redox potential of the fourth center are represented with help of the scaled color gradient displayed on the right-hand side. The respective values of the maximum likelihood estimate (see tab. 7.3) of the four redox potentials are indicated on each of the four axes (MLE).

Bibliography

- [1] ARISTOTLE. *Rhetoric*. I. 2, 1357a 34. Available from: <http://www.perseus.tufts.edu/hopper/text?doc=urn:cts:greekLit:tlg0086.tlg038.perseus-eng1:1357a> [accessed January 6th, 2014]. (Cited on p. 10).
- [2] ARNOLDS, R. Prinzip der FTIR-Spektroskopie [online]. Available from: <http://www.ir-spektroskopie.de/spec/ftir-prinzip/index.html> [accessed January 6th, 2014]. (Cited on pp. 69, 70).
- [3] BARD, A. J., AND FAULKNER, L. R. *Electrochemical Methods: Fundamentals and Applications*, 2nd ed. John Wiley & Sons, Inc., New York, 2001. (Cited on pp. 54, 62, 64).
- [4] BARKER, A. Monte Carlo Calculations of the Radial Distribution Functions for a Proton-Electron Plasma. *Australian Journal of Physics* **18**, 2 (1965), pp. 119–134. DOI: [10.1071/ph650119](https://doi.org/10.1071/ph650119). (Cited on p. 30).
- [5] BAURECHT, D. Biophysik I: Aufbau und Strukturanalyse von Biomolekülen. Lecture Notes [online]. November 2007, Institut für Biophysikalische Chemie, University of Vienna. Available from: http://www2.bpc.univie.ac.at/biophysik/pub/VO_Biophysik_1_DB.pdf [accessed December 20th, 2014]. (Cited on p. 72).
- [6] BAYES, T., AND PRICE, R. An Essay towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London* **53** (1763), pp. 370–418. DOI: [10.1098/rstl.1763.0053](https://doi.org/10.1098/rstl.1763.0053). (Cited on p. 19).
- [7] BERNARDO, J. M. Bayesian Statistics. In *Probability and Statistics*, R. Viertl, Ed., vol. 2, pp. 345–407, In *Encyclopedia of Life Support Systems (EOLSS)*. Developed under the Auspices of the UNESCO, EOLSS Publishers (www.eolss.net), Oxford, UK., 2003. Available from: <http://www.uv.es/~bernardo/BayesStat2.pdf> [accessed January 6th, 2014]. (Cited on pp. 1, 11, 14, 20, 100).
- [8] BOX, G. E. P., AND TIAO, G. C. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA, USA, 1973. (Cited on p. 23).
- [9] BROOKS, S. P., AND GELMAN, A. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* **7**, 4 (1998), pp. 434–455. DOI: [10.2307/1390675](https://doi.org/10.2307/1390675). (Cited on p. 47).

Bibliography

- [10] BUI, M., KAN, C. F. K., LE, S., AND TAN, E. UCDAVIS ChemWiki. The Arrhenius Law: Activation Energies [online]. August 2013, University of California. Available from: http://chemwiki.ucdavis.edu/Physical_Chemistry/Kinetics/Reaction_Rates/Temperature_Dependence_of_Reaction_Rates/Arrhenius_Equation/Activation_Energy [accessed January 6th, 2014]. (Cited on p. 61).
- [11] CALVETTI, D., AND SOMERSALO, E. *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, S. S. Antman, J. E. Marsden, and L. Sirovich, Eds., vol. 2 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer Science+Business Media, LLC, New York, USA, 2007. Available from: <http://books.google.com/books?vid=ISBN9780387733944> [accessed January 6th, 2014]. (Cited on pp. 7, 11, 13, 32).
- [12] CHEN, M.-H., AND SHAO, Q.-M. Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics* **8**, 1 (1999), pp. 69–92. DOI: [10.1080/10618600.1999.10474802](https://doi.org/10.1080/10618600.1999.10474802). (Cited on pp. 24, 26).
- [13] CHIB, S., AND GREENBERG, E. Understanding the Metropolis-Hastings Algorithm. *American Statistician* **49**, 4 (1995), pp. 327–335. DOI: [10.2307/2684568](https://doi.org/10.2307/2684568). (Cited on p. 31).
- [14] COUSINS, R. D. Why isn't every physicist a Bayesian? *American Journal of Physics* **63**, 5 (1995), pp. 398–410. DOI: [10.1119/1.17901](https://doi.org/10.1119/1.17901). (Cited on pp. 14, 15, 16, 17, 18).
- [15] COWLES, M. K., AND CARLIN, B. P. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association* **91**, 434 (1996), pp. 883–904. DOI: [10.1080/01621459.1996.10476956](https://doi.org/10.1080/01621459.1996.10476956). (Cited on pp. 37, 38).
- [16] D'AGOSTINI, G. Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics* **66**, 9 (2003), pp. 1383–1419. DOI: [10.1088/0034-4885/66/9/201](https://doi.org/10.1088/0034-4885/66/9/201). (Cited on p. 1).
- [17] DAO, P., AND LATIF, B. UCDAVIS ChemWiki. Transition State Theory 2 [online]. August 2013, University of California. Available from: http://chemwiki.ucdavis.edu/Physical_Chemistry/Kinetics/Complex_Reactions/Transition_State_Theory_2 [accessed January 6th, 2014]. (Cited on p. 61).
- [18] DE FINETTI, B. *Theory of Probability. A critical introductory treatment.*, 1st ed., vol. 1. Wiley, 1974. (Cited on p. 1).
- [19] DE FINETTI, B. *Theory of Probability. A critical introductory treatment.*, 1st ed., vol. 2. Wiley, 1975. (Cited on p. 1).
- [20] DOUCET, A., GODSILL, S. J., AND ROBERT, C. P. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing* **12**, 1 (2002), pp. 77–84. DOI: [10.1023/A:1013172322619](https://doi.org/10.1023/A:1013172322619). (Cited on p. 22).

- [21] FRIEDRICH, M. G., ROBERTSON, J. W. F., WALZ, D., KNOLL, W., AND NAUMANN, R. L. C. Electronic Wiring of a Multi-Redox Site Membrane Protein in a Biomimetic Surface Architecture. *Biophysical Journal* **94**, 9 (2008), pp. 3698–3705. DOI:[10.1529/biophysj.107.117846](https://doi.org/10.1529/biophysj.107.117846). (Cited on pp. [64](#), [100](#), [101](#)).
- [22] FRINGELI, U. P. In Situ Infrared Attenuated Total Reflection Membrane Spectroscopy. In *Internal Reflection Spectroscopy: Theory and Applications*, F. M. Mirabella, Jr., Ed., vol. 15 of *Practical Spectroscopy Series*. Marcel Dekker, Inc, New York, 1992, ch. 10, pp. 255–324. Available from: <http://books.google.at/books?vid=ISBN0824787307>. (Cited on p. [72](#)).
- [23] GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. *Bayesian Data Analysis*, 2nd ed. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton, Fla., USA, 2004. (Cited on pp. [21](#), [22](#), [37](#), [39](#), [46](#), [47](#)).
- [24] GELMAN, A., AND RUBIN, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**, 4 (1992), pp. 457–472. DOI:[10.1214/ss/1177011136](https://doi.org/10.1214/ss/1177011136). (Cited on pp. [37](#), [38](#), [40](#), [44](#), [45](#), [46](#)).
- [25] GREEN, P. J., AND MIRA, A. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika* **88**, 4 (2001), pp. 1035–1053. DOI:[10.1093/biomet/88.4.1035](https://doi.org/10.1093/biomet/88.4.1035). (Cited on p. [34](#)).
- [26] GUPTA, G. K., SACKS-DAVIS, R., AND TISCHER, P. E. A Review of Recent Developments in Solving ODEs. *Computing Surveys* **17**, 1 (1985), pp. 5–47. DOI:[10.1145/4078.4079](https://doi.org/10.1145/4078.4079). (Cited on pp. [84](#), [85](#)).
- [27] GUT, A. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer Science+Business Media, Inc., New York, NY, USA, 2005. (Cited on pp. [3](#), [4](#)).
- [28] HAARIO, H., LAINE, M., MIRA, A., AND SAKSMAN, E. DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16**, 4 (2006), pp. 339–354. DOI:[10.1007/s11222-006-9438-0](https://doi.org/10.1007/s11222-006-9438-0). (Cited on pp. [34](#), [35](#), [36](#)).
- [29] HÁJEK, A. Interpretations of Probability. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Winter 2012 ed. Available from: <http://plato.stanford.edu/archives/win2012/entries/probability-interpret> [accessed December 9th, 2013]. (Cited on p. [10](#)).
- [30] HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 1 (1970), pp. 97–109. DOI:[10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97). (Cited on pp. [27](#), [30](#), [31](#), [32](#)).
- [31] HOFF, P. D. Statistics 502 Lecture Notes [online]. December 2009. Available from: <http://www.stat.washington.edu/hoff/courses/stat421-502/LectureNotes/notes.pdf> [accessed January 6th, 2014]. (Cited on pp. [40](#), [41](#), [43](#)).

Bibliography

- [32] HYNDMAN, R. J. Computing and Graphing Highest Density Regions. *The American Statistician* **50**, 2 (1996), pp. 120–126. DOI:[10.2307/2684423](https://doi.org/10.2307/2684423). (Cited on pp. [23](#), [24](#), [25](#)).
- [33] JEFFREYS, H. *Theory of Probability*, 3rd ed. Clarendon Press, Oxford, UK, 1961. (Cited on p. [1](#)).
- [34] KARLIN, S., AND TAYLOR, H. M. *A First Course in Stochastic Processes*, 2nd ed. Academic Press, San Diego, California, USA, 1975. Available from: <http://books.google.com/books?vid=ISBN0080570410> [accessed March 3rd, 2014]. (Cited on pp. [27](#), [28](#)).
- [35] KAUPPINEN, J., AND PARTANEN, J. *Fourier Transforms in Spectroscopy*, 1st ed. WILEY-VCH, Berlin, Germany, 2001. Available from: <http://books.google.at/books?vid=ISBN3527635017> [accessed December 9th, 2014]. (Cited on pp. [68](#), [69](#)).
- [36] KEYNES, J. M. *A Treatise on Probability*, EBook ed. No. 32625. Project Gutenberg, May 31, 2010. Available from: <http://www.gutenberg.org/files/32625/32625-pdf.pdf> [accessed December 9th, 2013]. (Cited on pp. [9](#), [10](#)).
- [37] KOLMOGOROV, A. N. Grundbegriffe der Wahrscheinlichkeitsrechnung. In *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2, 3. Springer, Berlin, 1933. (Cited on p. [3](#)).
- [38] KOLMOGOROV, A. N. *Foundations of the Theory of Probability*, 2nd ed. Chelsea Publishing Company, New York, USA, 1956. Translation edited by Nathan Morrison. (Cited on p. [3](#)).
- [39] LAINE, M. MCMC toolbox for Matlab [online]. August 2013, University of Helsinki. Available from: <http://helios.fmi.fi/~lainema/mcmc/> [accessed January 6th, 2014]. (Cited on pp. [90](#), [93](#)).
- [40] LAPLACE, P.-S. *A Philosophical Essay on Probabilities*, 1st ed. Translated from the sixth French edition by Truscott, F. W. and Emroy, F. L. John Wiley & Sons, New York, 1902. Available from: <https://archive.org/details/philosophicaless00lapliala> [accessed January 6th, 2014]. (Cited on p. [9](#)).
- [41] LAVIRON, E. General expression of the linear potential sweep voltammogram in the case of diffusionless electrochemical systems. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* **101**, 1 (July 1979), pp. 19–28. DOI:[10.1016/S0022-0728\(79\)80075-3](https://doi.org/10.1016/S0022-0728(79)80075-3). (Cited on p. [62](#)).
- [42] LÉGER, C., AND BERTRAND, P. Direct Electrochemistry of Redox Enzymes as a Tool for Mechanistic Studies. *Chemical Reviews* **108**, 7 (2008), pp. 2379–2438. DOI: [10.1021/cr0680742](https://doi.org/10.1021/cr0680742). (Cited on pp. [53](#), [54](#), [60](#), [62](#)).

- [43] MATLAB DOCUMENTATION CENTER. Anonymous Functions [online]. MathWorks. Available from: http://www.mathworks.de/de/help/matlab/matlab_prog/anonymous-functions.html [accessed January 6th, 2014]. (Cited on p. 79).
- [44] MATLAB DOCUMENTATION CENTER. Odeset [online]. MathWorks. Available from: <http://www.mathworks.de/de/help/matlab/ref/odeset.html> [accessed January 6th, 2014]. (Cited on p. 83).
- [45] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., AND TELLER, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 6 (1953), pp. 1087–1092. DOI:[10.1063/1.1699114](https://doi.org/10.1063/1.1699114). (Cited on pp. 27, 30).
- [46] MIRA, A. On Metropolis-Hastings algorithms with delayed rejection. *Metron* **LIX**, 3-4 (2001), pp. 231–241. Available from: http://www.metronjournal.it/volumi/sommari_en.lasso?volume=LIX&numero=3-4 [accessed January 6th, 2014]. (Cited on pp. 34, 36).
- [47] MIRA, A. Ordering and Improving the Performance of Monte Carlo Markov Chains. *Statistical Science* **16**, 4 (2001), pp. 340–350. DOI:[10.1214/ss/1015346319](https://doi.org/10.1214/ss/1015346319). (Cited on p. 34).
- [48] NAKANO, T., YOKOYAMA, T., AND TORIUMI, H. One- and Two-Dimensional Infrared Time-Resolved Spectroscopy Using a Step-Scan FT-IR Spectrometer: Application to the Study of Liquid Crystal Reorientation Dynamics. *Applied Spectroscopy* **47**, 9 (1993), pp. 1354–1366. DOI:[10.1366/0003702934067702](https://doi.org/10.1366/0003702934067702). (Cited on p. 73).
- [49] NEYMAN, J. *A Selection of Early Statistical Papers of J. Neyman*. University of California Press, Berkeley, CA, 1967. (Cited on p. 14).
- [50] NORRIS, J. R. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK, 1998. (First Paperback Edition 1998, Reprinted 1999). Available from: <http://books.google.at/books?vid=ISBN9780521633963> [accessed November 8th, 2014]. (Cited on p. 27).
- [51] PAWITAN, Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, Oxford, UK, 2013. Available from: <http://books.google.com/books?vid=ISBN9780199671229> [accessed November 8th, 2014]. (Cited on pp. 6, 7, 16).
- [52] PENROSE, O. *Foundations of Statistical Mechanics: A Deductive Treatment*. Dover Publications, Inc., New York, USA, 2005. Reprint of the work originally published by Pergamon Press, Oxford, 1970. Available from: <http://books.google.com/books?vid=ISBN0486438708> [accessed January 9th, 2015]. (Cited on p. 29).

Bibliography

- [53] PERKINS, W. D. Fourier Transform-Infrared Spectroscopy: Part I. Instrumentation. *Journal of Chemical Education* **63**, 1 (1986), p. A5. DOI: [10.1021/ed063pA5](https://doi.org/10.1021/ed063pA5). (Cited on p. 68).
- [54] PESKUN, P. H. Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**, 3 (1973), pp. 607–612. DOI: [10.1093/biomet/60.3.607](https://doi.org/10.1093/biomet/60.3.607). (Cited on p. 34).
- [55] ROBERT, C. P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer Science+Business Media, LLC, New York, USA, 2007. (Cited on pp. 6, 8, 13, 19, 20, 21, 23).
- [56] SCHACH, D. *Direct Electron Transfer to Cytochrome c Oxidase investigated by Electrochemistry and Time-Resolved Surface-Enhanced Infrared Absorption Spectroscopy*. PhD thesis, University of Mainz, 2011. Available from: <http://ubm.opus.hbz-nrw.de/volltexte/2011/2826/> [accessed January 6th, 2014]. (Cited on pp. 58, 59, 60, 62, 64, 65, 66, 71, 73, 75, 100).
- [57] SCHACH, D., NOWAK, C., GENNIS, R. B., FERGUSON-MILLER, S., KNOLL, W., WALZ, D., AND NAUMANN, R. L. C. Modeling direct electron transfer to a multi-redox center protein: Cytochrome c oxidase. *Journal of Electroanalytical Chemistry* **649**, 1–2 (2010), pp. 268–276. DOI: [10.1016/j.jelechem.2010.07.009](https://doi.org/10.1016/j.jelechem.2010.07.009). (Cited on pp. 1, 49, 51, 100, 101).
- [58] SCHWAIGHOFER, A., BAURECHT, D., NAUMANN, R. L. C., AND NOWAK, C. Protein Dynamics of Cytochrome C Oxidase: Analysis of Vibrational Components by Phase Sensitive Detection Applied to Modulated Excitation-Seiras. *Biophysical Journal* **104**, 2, Supplement 1 (Jan. 2013), p. 224a. DOI: [10.1016/j.bpj.2012.11.1265](https://doi.org/10.1016/j.bpj.2012.11.1265). (Cited on p. 74).
- [59] SCHWAIGHOFER, A., STEININGER, C., HILDENBRANDT, D. M., SRAJER, J., NOWAK, C., KNOLL, W., AND NAUMANN, R. L. C. Time-Resolved Surface-Enhanced IR-Absorption Spectroscopy of Direct Electron Transfer to Cytochrome c Oxidase from *R. sphaeroides*. *Biophysical Journal* **105**, 12 (Dec. 2013), pp. 2706–2713. DOI: [10.1016/j.bpj.2013.10.037](https://doi.org/10.1016/j.bpj.2013.10.037). (Cited on pp. 2, 49, 74, 75, 77, 88, 101, 104, ix, xi).
- [60] SHAMPINE, L. F., GLADWELL, I., AND THOMPSON, S. *Solving ODEs with MATLAB*. Cambridge University Press, Cambridge, UK, 2003. (Cited on pp. 84, 85, 86).
- [61] SHAMPINE, L. F., AND REICHEL, M. W. The MATLAB ODE Suite. *SIAM Journal on Scientific Computing* **18**, 1 (1997), pp. 1–22. DOI: [10.1137/S1064827594276424](https://doi.org/10.1137/S1064827594276424). (Cited on pp. 83, 86).
- [62] SNOW, G. R package "TeachingDemos" [online]. January 2013. Available from: <http://cran.r-project.org/web/packages/TeachingDemos/index.html> [accessed January 6th, 2014]. (Cited on p. 99).

- [63] SRAJER, J., SCHWAIGHOFER, A., HILDENBRANDT, D. M., KIBROM, A., AND NAUMANN, R. L. C. A kinetic model of proton transport in a multi-redox centre protein: cytochrome *c* oxidase. *Progress in Reaction Kinetics and Mechanism* **38**, 1 (2013), pp. 32–47. DOI:[10.3184/146867812X13558465325118](https://doi.org/10.3184/146867812X13558465325118). (Cited on pp. [1](#), [2](#), [49](#), [50](#), [51](#), [52](#), [53](#), [54](#), [55](#), [56](#), [57](#), [67](#), [99](#), [100](#), [101](#), [ix](#), [xi](#)).
- [64] TARANTOLA, A. Popper, Bayes and the Inverse problem. *Nature Physics* **2**, 8 (2006), pp. 492–494. DOI:[10.1038/nphys375](https://doi.org/10.1038/nphys375). (Cited on p. [13](#)).
- [65] TIERNEY, L., AND MIRA, A. Some adaptive Monte Carlo Methods for Bayesian Inference. *Statistics in Medicine* **18**, 17–18 (1999), pp. 2507–2515. DOI:[http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18<2507::AID-SIM272>3.0.CO;2-J/references](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2507::AID-SIM272>3.0.CO;2-J/references). (Cited on pp. [34](#), [36](#)).
- [66] VENN, J. *The logic of chance: an essay on the foundations and province of the theory of probability, with especial reference to its logical bearings and its application to moral and social science, and to statistics.*, 3rd ed. Macmillan and Company, London, 1888. DOI: [10.5962/bhl.title.19044](https://doi.org/10.5962/bhl.title.19044). (Cited on p. [10](#)).
- [67] WALZ, D. Thermodynamics of oxidation-reduction reactions and its application to bioenergetics. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **505**, 3–4 (1979), pp. 279–353. DOI:[10.1016/0304-4173\(79\)90007-7](https://doi.org/10.1016/0304-4173(79)90007-7). (Cited on pp. [1](#), [49](#), [53](#)).

Bibliography

A. Abstract

Within the scope of this thesis, we adopted a kinetic model of cytochrome *c* oxidase (CcO) developed by Srajer et al. [63] which proposes a mechanism for proton transport coupled to electron transfer reactions of CcO from *R. sphaeroides*. This model uses a modified Nernst equation to account for protonation/deprotonation in the course of a second-order redox reaction between two of the four redox centers, respectively. Overall, there are sixteen possible redox states or conformations of the enzyme, according to the model.

Srajer et al. have used a master equation approach to describe the time evolution of the probability of occupation of each of the 16 possible redox states. The solution of this master equation depends on a number of unknown model parameters such as the standard redox potential of each of the four centers of the molecule and fixed acid dissociation constants to account for protonation equilibria of each redox site.

The author of this thesis contributed to the implementation of the model on the computer via MATLAB, particularly with regard to the numerical solution of the master equation. Subsequently, in the context of this thesis, the implementation of this model was extended in order to simulate absorbance changes visualized by means of time-resolved attenuated total reflection surface-enhanced IR-absorption spectroscopy (tr-ATR-SEIRAS) which can be associated with changes in redox states of the enzyme. Possible caveats of the kinetic model with regard to the simulation of absorbance data in particular as well as some of the simplifications used in this model in general were pointed out and discussed. Ultimately, with the help of IR absorbance data, a Bayesian parameter estimation based on a posterior probability distribution was conducted and published in Schwaighofer et al. [59].

The central aim of this thesis consists in the application of Bayesian inference in order to gain physically meaningful point estimates as well as credible regions for as many of the models' free parameters as possible. In particular, an evaluation of and improvement upon initial estimations for some of the fourteen free parameters, deduced from previous experimental evidence such as fast-scan cyclic voltammetry measurements, was intended.

For this purpose a prior probability density and a likelihood function, both obtained from experimental evidence and plausibility arguments, were defined. In order to approximately generate a representative sample from the resulting Bayesian posterior density, a Markov chain Monte Carlo simulation procedure running several parallel chains from different starting points was implemented. A maximum likelihood point estimate as well as highest posterior density (HPD) intervals for each parameter of interest were subsequently calculated. The main results of this thesis are the empirical maximum likelihood point estimate on the one hand as well as the empirical highest marginal posterior density regions on the other hand, both obtained from the sample generated by the MCMC simulation procedure. The probability for a particular parameter to lie within such a credible region (e.g. a 95 % HPD interval)

A. Abstract

should be understood as conditional probability, given the assumptions of our posterior and likelihood model as well as the measured tr-ATR-SEIRAS absorbance data. Results are presented in numerical as well as in graphical form, in particular as graphical comparison of the output of our kinetic model, generated with help of these updated, fitted parameters to IR absorbance measurement data of three of the four redox centers of CcO. In addition to that, histograms of all parameters of interest, based on the MCMC simulation data, graphically represent empirical estimates of marginal posterior densities of each parameter.

B. Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde ein von Srajer et al. [63] entwickeltes, reaktionskinetisches Modell der Cytochrom-*c*-Oxidase (CcO) übernommen, welches einen möglichen Mechanismus für den an den Elektronentransfer gekoppelten Protonentransport dieses Enzyms vorschlägt. Dieses Modell verwendet eine modifizierte Form der Nernst-Gleichung um Protonierungen und Deprotonierungen im Zuge einer Redoxreaktion zweiter Ordnung zwischen jeweils zwei der vier Zentren der CcO zu beschreiben. Insgesamt ergeben sich sechzehn Redox-Zustände oder Konformationen für das Enzym. Srajer et al. beschreiben die Zeitentwicklung der Besetzungswahrscheinlichkeiten dieser 16 Zustände mithilfe einer Mastergleichung. Die Lösung dieser Mastergleichung hängt allerdings von einer Reihe von experimentell schwer zu bestimmenden Parametern ab, unter anderem von den Standard-Redox-Potentialen der vier Zentren oder von den als fest angenommenen Dissoziationskonstanten, die zur Beschreibung der Protonierungs-Gleichgewichte verwendet werden.

Der Autor dieser Arbeit hat vornehmlich zur Implementierung dieses Modells am Computer beigetragen, insbesondere auch zur numerischen Lösung der Mastergleichung mittels MATLAB. Darauf aufbauend wurde im Rahmen dieser Arbeit die bestehende Implementierung erweitert um mittels *time-resolved attenuated total reflection surface-enhanced IR-absorption spectroscopy (tr-ATR-SEIRAS)* visualisierbare Absorbanzänderungen, denen Änderungen der Redoxzustände der CcO zugeordnet werden können, am Computer zu simulieren. Mögliche Kritikpunkte an der Anwendung unseres Modells zur Simulation dieser speziellen Messergebnisse sowie Vorbehalte allgemeinerer Art wurden aufgeführt und diskutiert. Mithilfe dieses Computermodells wurde anschließend eine Parameter-Schätzung basierend auf einer A-posteriori-Wahrscheinlichkeitsdichte im Sinne der Bayesschen Statistik durchgeführt und in Zusammenarbeit mit Schwaighofer et al. [59] veröffentlicht.

Das Hauptziel dieser Arbeit besteht in der praktischen Anwendung Bayesscher Inferenzstatistik, um für möglichst viele der freien Parameter unseres Modells mit den Messergebnissen vereinbare, physikalisch plausible Punktschätzungen und Konfidenzintervalle (im Bayesschen Sinn) zu erlangen. Insbesondere wurde mit dieser Methode eine Überprüfung und genauere Bestimmung der aus vorangehenden *fast-scan* Cyklovoltammetriemessungen gewonnenen Schätzwerte mancher der insgesamt vierzehn freien Parameter des Modells im Lichte der durch die IR-Absorbanz-Messungen neu erlangten Daten angestrebt. Zu diesem Zweck wurde eine A-priori-Wahrscheinlichkeitsdichte sowie eine Likelihood-Funktion definiert, basierend auf vorangehenden Messungen und Plausibilitätsüberlegungen. Um aus der resultierenden A-posteriori-Wahrscheinlichkeitsdichte eine (annähernd) repräsentative Stichprobe zu ziehen, wurde eine Markov-Chain-Monte-Carlo Simulations-Prozedur implementiert, realisiert in Form von mehreren parallel laufenden Markov-Ketten, ausgehend von verschiedenen, speziell gewählten Punkten des Parameterraums. Mithilfe dieser Stichprobe

B. Zusammenfassung

wurden ein empirischer Maximum-Likelihood Punkt-Schätzwert und eindimensionale, empirische *highest posterior density regions* der marginalen A-posteriori-Wahrscheinlichkeitsdichten für einige der Parameter berechnet.

Als Hauptresultat dieser Arbeit gilt zum einen der gewonnene empirische Maximum-Likelihood Schätzwert aller Parameter. Zum anderen sind es die für einige Parameter einzeln bestimmten, empirischen *highest marginal posterior density* Intervalle. Die angegebenen Realisierungswahrscheinlichkeiten für einen Parameter innerhalb eines solchen Intervalls (zbsp. 95 %) sind als bedingte Wahrscheinlichkeiten zu verstehen, abhängig von den Annahmen unseres Wahrscheinlichkeitsmodells und den gemessenen tr-ATR-SEIRAS Daten. Die Arbeit schließt mit der Darstellung aller Ergebnisse in numerischer und grafischer Form. Insbesondere wurde eine grafische Gegenüberstellung der tr-ATR-SEIRAS Absorbanzänderungs-Messkurven mit den vom Modell theoretisch mithilfe der Maximum-Likelihood Schätzwerte errechneten Kurven für drei der vier Redoxzentren der CcO angefertigt. Zusätzlich dazu stellen Histogramme aller Parameter eine Annäherung an die tatsächlichen marginalen A-posteriori-Wahrscheinlichkeitsdichten dieser dar.

C. Acknowledgements

I hereby want to thank my advisor Dr. Renate Naumann of Austrian Institute of Technology (AIT), firstly for providing me with the opportunity to work on a challenging project that allowed me to broaden my knowledge on a variety of subjects such as Bayesian statistics, stochastic simulation and the computational modeling of the kinetics of real life systems such as the Cytochrome *c* Oxidase. Secondly, for sharing her profound knowledge of enzyme kinetics and electro chemistry and for providing a multitude of useful suggestions that helped me complete this thesis.

I also want to express my sincere gratitude to my advisor ao. Univ.-Prof. Dr. Martin Neumann of University of Vienna, for sharing his expertise in stochastic simulation in general and Monte Carlo methods in particular and for readily discussing the various practical questions concerning MCMC simulation, that arose during the course of thesis—as well as for his patience and constructive critique. Special thanks go to Dr. Johannes Srajer of AIT for sharing his wisdom in MATLAB and his general helpfulness, to Dr. Asmorom Kibrom and Mag. Ciril Rozman for always being up for an interesting conversation and to the people that used to frequent the Copyteria.

I want to thank my brother Florian and all of my friends for providing proper distraction during periods of exhaustion.

I am immensely grateful to my parents for their perpetual moral and financial support during the course of my studies.

C. Acknowledgements

D. Curriculum Vitae

Persönliche Informationen

Name: David Markus Hildenbrandt
Geburtsdatum: 06.12.1984
Staatsbürgerschaft: Österreich

Schulische Ausbildung

1991-1995 Volksschule Gunskirchen
1995-2004 Bundesgymnasium Dr. Schauerstrasse, Wels

Zivildienst

SS 2004 - SS 2005 Lebenshilfe Wels

Universitäre Ausbildung

WS 2004 Beginn des Diplomstudiums der Physik, Universität Wien
SS 2004 - SS 2005 Unterbrechung des Studiums (Beurlaubung): Zivildienst
SS 2005 Wiederaufnahme des Physikstudiums, Universität Wien
WS 2012 Beginn des Bachelorstudiums der Physik, Universität Wien
Mai 2013 Abschluss des Bachelor of Science (Physik), Universität Wien
seit SS 2013 Masterstudium Physik, Universität Wien

Bisherige Arbeitsverhältnisse

01.08.2007 - 31.08.2007 Ferialpraktikum E+E Elektronik Ges.m.b.H
01.08.2009 - 31.08.2009 Ferialpraktikum Fronius International GmbH
01.08.2011 - 01.01.2012 geringfügige Beschäftigung bei aforms2web
15.02. 2012 - 15.01.2013 geringfügige Beschäftigung bei Austrian Institute of Technology (AIT), Abteilung Biosensors Technology

Publikationen

1. Srajer, J., Schwaighofer, A., Hildenbrandt, D. M., Kibrom, A., and Naumann, R. L. C. A kinetic model of proton transport in a multi-redox centre protein: cytochrome c oxidase. *Progress in Reaction Kinetics and Mechanism* **38**, 1 (2013), pp. 32–47.
2. Schwaighofer, A., Steininger, C., Hildenbrandt, D. M., Srajer, J., Nowak, C., Knoll, W., and Naumann, R. L. C. Time-Resolved Surface-Enhanced IR-Absorption Spectroscopy of Direct Electron Transfer to Cytochrome c Oxidase from *R. sphaeroides*. *Biophysical Journal* **105**, 12 (Dec. 2013), pp. 2706–2713.

Kenntnisse

Sprachen

Deutsch (Muttersprache), Englisch fließend in Wort und Schrift, Grundkenntnisse in Spanisch

Programmiersprachen

MATLAB, R, Mathematica, C, L^AT_EX

Betriebssysteme

GNU/Linux, MS Windows, Mac OS X

Software

Origin, MS Office, Openoffice, LabView, gnuplot