Polarised neutron diffraction measurements of $PrBa_2Cu_3O_{6+x}$ and the Bayesian statistical analysis of such data

Anders Johannes Markvardsen

St. Edmund Hall

A thesis submitted for the degree of

Doctor of Philosophy at the

University of Oxford

Clarendon Laboratory

September 2000

To

My Parents

Polarised neutron diffraction measurements of $PrBa_2Cu_3O_{6+x}$ and the Bayesian statistical analysis of such data

Anders J. Markvardsen, *St. Edmund Hall, University of Oxford*

Thesis submitted for the degree of Doctor of Philosophy, September 2000

Abstract

The physics of the series $Pr_yY_{1-y}Ba_2Cu_3O_{6+x}$, and ability of Pr to suppress superconductivity, has been a subject of frequent discussions in the literature for more than a decade. This thesis describes a polarised neutron diffraction (PND) experiment performed on $PrBa_2Cu_3O_{6.24}$ designed to find out something about the electron structure. This experiment pushed the limits of what can be done using the PND technique. The problem is one of a limited number of measured Fourier components that need to be inverted to form a real space image. To accomplish this inversion the maximum entropy technique has been employed. In some cases, the maximum entropy technique has the ability to increase the resolution of 'inverted' data immensely, but this ability is found to depend critically on the choice of constants used in the method. To investigate this a Bayesian robustness analysis of the maximum entropy method is carried out, resulting in an improvement of the maximum entropy technique for analysing PND data. Some results for nickel in the literature have been re-analysed and a comparison is made with different maximum entropy algorithms.

Equipped with an improved data analysis technique and carefully measured PND data for $PrBa_2Cu_3O_{6.24}$ a number of new interesting features are observed, putting constraints on existing theoretical models of $Pr_yY_{1-y}Ba_2Cu_3O_{6+x}$ and leaving room for more questions to be answered.

Acknowledgements

I would like to thank Andrew T. Boothroyd and Brian Buck, my supervisors, for their stimulating relationships, many ideas and suggestions and most importantly their never-failing support during this thesis. Also, many thanks to their great patience towards my sometimes optimistic proposed deadlines.

Outside the laboratory, the list of people I should like to acknowledge is too long to mention but a big thanks must go to them all for reminding me that there's more to life than physics.

Of course, I am very grateful to my parents for their patience and love. Without them this work would never have come into existence (literally). Likewise, I would like to thank my brothers for their support.

I wish to thank the following: Harriott Nowell for carefully (in addition to my supervisors) reading through my thesis and for making everyday life that bit more enjoyable doing the final stages of this work; and Bill David, my Post Doc employer at Rutherford Appleton Laboratory during the past year, for providing me with flexible working hours to finish this thesis.

Finally, many thanks to Geoff J. Daniell and Neil F. Johnson for helpful discussions.

Contents

Introduction

One of the most engaging problems that has emerged from the study of cuprate high T_C superconductors is the suppression of superconductivity that occurs when Pr is substituted for other lanthanide ions in certain compounds. The effect is best known when Pr is doped into $YBa_2Cu_3O_{6+x}$, where the superconducting transition temperature, T_C, decreases as the amount of Pr increases until superconductivity is entirely surpressed. This is in contrast with the substitution of most other rare earths into $YBa_2Cu_3O_{6+x}$, which has little effect on T_C.[1]

The anomalous electrical and magnetic properties of $(Pr_yY_{1-y})Ba_2Cu_3O_{6+x}$ have been studied by many research groups for over a decade and the problem has recently been highlighted by reports of superconductivity in a single crystal of $PrBa_2Cu_3O_{6+x}$ grown under special conditions (for a recent review paper see Ref. [1]), a result which may provide help in the understanding of why samples prepared by standard methods do not superconduct.

Many theories have been put forward to explain the reason for Pr having an adverse effect on superconductivity. The majority of these focus on an interaction between Pr and the superconducting cuprate planes. It is the test of such models which is the aim of the polarised neutron diffraction experiment on $PrBa_2Cu_3O_{6.24}$ described in this thesis.

Polarised neutron diffraction is a technique for revealing the magnetization density distribution induced by a magnetic field in a single crystal. Unfortunately the small cross-section for scattering neutrons by magnetic moments in the sample

[1] Don't give up there are only 6 chapters and about 200 pages left to read:-)

results in data sets which are much smaller than those obtained from other types of diffraction techniques (e.g. standard X-ray diffraction). To assist the data analysis of the PND data the so-called maximum entropy technique has been employed.

Maximum entropy methods existed before their first application in data analysis. The maximum entropy method as defined in statistical or information theoretical textbooks is a method for selecting a probability distribution from a set of possible distributions, and in this context the maximum entropy method is abbreviated as MEM. Its first application in data analysis was in the spectral analysis of time series [2] and in the following years it evolved into forms which, in many cases, no longer justify a link to the original concept of a MEM. Therefore, in this thesis a different abbreviation, MEMx, is used to represent the way the maximum entropy method is used in image analysis today. The MEMx technique is critically analysed by a Bayesian robustness analysis and the result is an improved data analysis technique tuned to 'invert' PND data into magnetization density. The thesis is structured as follows:

Chapter 1 and 2. Two introductory chapters are included. The first deals with the PND experimental technique. This will be particular useful when reading Chap. 5 and Chap. 6. The second introductory chapter is an introduction to Bayesian statistics with an emphasis on the topics relevant for understanding Chap. 3 and Chap. 4. Further, it includes the probabilistic notation used in this thesis, such as, the concept of a 'prior model constant', a constant that is part of a prior probability expression. These two introductory chapters cover two research areas that are not normally seen together and many researchers will be familiar with one, but not both.

Chapter 3. Contains a summary of important early works on the maximum entropy method. It complements two papers from the literature which review the use of the maximum entropy method in data analysis (see

Sec. 3.4), as it includes early works that are not discussed in these review papers, but does not discuss the practical implication of the maximum entropy method. Moreover, the definition for MEMx is given in Sec. 3.3.3; the technique which will be the subject of a critical Bayesian analysis in the following chapter.

Chapter 4. A Bayesian robustness analysis of the MEMx technique is performed and the end result is a data analysis method tuned for the analysis of PND data. These results are summarized in Sec. 4.7.

Chapter 5. Perhaps the main chapter of the thesis. The Bayesian image analysis technique described in Chap. 4 is applied on PND data measured on $PrBa_2Cu_3O_{6.24}$. A number of new features are observed in the induced magnetization density, these are discussed in relation to previous experiment and theoretical works.

Chapter 6. MEMx algorithms are compared for the analysis of PND data measured on nickel. The works of Papoular et al. and Sakata et al. are reviewed in detail. These works are compared to each other and to the work of this thesis. The induced magnetization density of nickel is also discussed.

To briefly summarise what is novel in this thesis: *i*) The highly critical and detailed study of the maximum entropy method in Chap. 4 provides a new MEMx recipe tuned to the analysis of PND data, and is a more efficient and rigorous method for analysing such data. *ii*) Chap. 3 is a review chapter giving a discussion of early maximum entropy work, some of which has not been reviewed previously. *iii*) PND data of $PrBa_2Cu_3O_{6.24}$ collected at ILL, Grenoble are studied in Chap. 5. The induced magnetization density near the Pr site can be largely accounted for within a crystal field model, when comparing it to a theoretical calculation performed using such a model. However, a number of other features are

observed in various different places within the unit cell, where none of these can be explained within existing theoretical models, and these provide ideas for further investigations. *iv*) Previously collected PND data of nickel are re-analysed in Chap. 6. The work in Chap. 4 is compared with previous maximum entropy work done on nickel PND data, in order to illustrate possible differences, and to show that the data analysis method of Chap. 4 provides an improvement to the MEMx analysis of PND data. Given the accuracy of the nickel data, this chapter also reveals new features of the negative magnetization density distribution of nickel.

Chapter 1

Polarised neutron diffraction (PND)

A number of good textbooks exist on thermal neutron scattering for instance those of Squires and Lovesey [3],[4]. For a textbook which specifically deals with polarised neutrons see Ref. [5]. The neutron scattering theory relevant for understanding the polarised neutron diffraction experiments presented in this thesis will be discussed here. This discussion will be centered around coherent elastic scattering of polarised neutrons by a crystal.

1.1 Flipping ratios

When neutrons are scattered by condensed matter the scattering is caused by the strong nuclear force between neutron and atomic nuclei, and the interaction between the neutron's magnetic moment and any electrons with a moment in the target material. Thus, neutron scattering can provide useful information about the positions of the atoms and about their magnetic properties.

Scattering is described in physics by counting the number of particles (in this case neutrons) scattered into some solid angle $\Delta\Omega$ with energy between E and

$E+\Delta E$. Mathematically we write this as $I_0(\frac{d^2\sigma}{d\Omega dE})\Delta\Omega\Delta E$, where I_0 is the incident flux and the quantity of interest $\frac{d^2\sigma}{d\Omega dE}$ is the *partial differential cross-section* (or just referred to as cross-section in short hand notation). Since, in this thesis, we are only concerned with coherent elastic scattering it is sufficient to consider the cross-section $\frac{d\sigma}{d\Omega}$ [1], which does not discriminate the scattered neutron's energy.

Taking into account that thermal neutrons only interact weakly with condensed matter, we can describe the cross-section within the Born approximation. The coherent, elastic, nuclear cross-section for a target crystal of volume V in this approximation is calculated to be

$$\frac{d\sigma}{d\Omega} = N\frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta(\mathbf{q}-\mathbf{K})|F_N(\mathbf{K})|^2, \tag{1.1}$$

where $\mathbf{K}$ is a reciprocal lattice vector. $\mathbf{q}$ is the scattering vector equal to the difference between the wave vector of the incoming and outgoing plane wave: $\mathbf{q} = \mathbf{k} - \mathbf{k}'$. As seen from Eq. (1.1), the delta function means that the cross-section for elastic scattering is zero unless $\mathbf{q} = \mathbf{K}$, this is the condition for Bragg scattering. The cross-section in Eq. (1.1) assumes that the crystal is a rigid lattice. In a real crystal, the thermal motion of atoms around their equilibrium position changes the magnitude of the observed cross-section. To account for this, Eq. (1.1) must be multiplied by a Debye-Waller factor. Also, since Eq. (1.1) is calculated within the Born approximation, the neutrons are assumed not to make more than one collision before escaping the sample. In a real experiment multiple scattering does occur, and this phenomenon is referred to as extinction. Depending on the experimental conditions and the nature of the crystal the degree of extinction varies. Extinction cannot be avoided but reliable methods exist to correct the observed structure factors for this multiple scattering phenomenon.

The main interest here is the use of thermal neutrons to probe the magnetic structure of crystals. It is the neutron spin $\boldsymbol{\sigma}$ that allows the interaction between the incident neutrons and the magnetic electrons of the scattering material. The

[1] Strictly speaking $\left(\frac{d\sigma}{d\Omega}\right)_{coh}$ but for convenience we write it as $\frac{d\sigma}{d\Omega}$.

interaction potential is written as $V_M = -\gamma\mu_N \boldsymbol{\sigma} \cdot \mathbf{B}$, where

$$\mathbf{B} = -\frac{\mu_0\mu_B}{2\pi}\left(\nabla \times \left(\frac{\mathbf{S}\times\mathbf{R}}{R^3}\right) + \frac{1}{\hbar}\frac{\mathbf{p}\times\mathbf{R}}{R^3}\right) \tag{1.2}$$

is the total magnetic field due to an electron with momentum $\mathbf{p}$ and spin $\mathbf{S}$ [2]. From the expression in Eq. (1.2) we see that the first term results from the interaction of the neutron with the electron spin, and the second is due to the orbital angular momentum. Within the Born approximation the coherent elastic cross section for magnetic scattering from a crystal may be written as

$$\left(\frac{d\sigma}{d\Omega}\right)_{\sigma\to\sigma'} = \left(\frac{\gamma r_0}{2\mu_B}\right)^2 N\frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta(\mathbf{q}-\mathbf{K})\left|\langle\sigma|\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}(\mathbf{K})|\sigma'\rangle\right|^2 \quad , \tag{1.3}$$

where r_0 is the classical electron radius and

$$\mathbf{F}_{M_\perp} = \hat{\mathbf{q}}\times(\mathbf{F}_M\times\hat{\mathbf{q}}) = \mathbf{F}_M - \hat{\mathbf{q}}(\mathbf{F}_M\cdot\hat{\mathbf{q}}) \tag{1.4}$$

and $\mathbf{F}_M$ is a magnetic structure factor. $\mathbf{F}_M$ contains a spin and an orbital part corresponding to the terms in Eq. (1.2). From Eq. (1.4) we see the significance of the subscript $\perp$ in Eq. (1.3); $\mathbf{F}_{M_\perp}$ is the vector projection of $\mathbf{F}_M$ on to the plane perpendicular to the scattering vector $\mathbf{q}$, and it is only this component of $\mathbf{F}_M$ that is measured. As for the nuclear scattering, the magnetic cross section should include the Debye-Waller factor to account for the thermal motion of atoms in a real crystal.

In contrast to coherent nuclear scattering, the coherent magnetic scattering is sensitive to the spin direction of the incoming (and outgoing) neutrons. This is emphasized by explicitly including the Dirac bra σ and ket σ' in Eq. (1.3). Combining Eq. (1.1) and Eq. (1.3) we see that the total coherent elastic scattering from a crystal with magnetic order is[3]

$$\left(\frac{d\sigma}{d\Omega}\right)_{\sigma\to\sigma'} = N\frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta(\mathbf{q}-\mathbf{K})\left|\langle\sigma|F_N(\mathbf{K}) + \frac{\gamma r_0}{2\mu_B}\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}(\mathbf{K})|\sigma'\rangle\right|^2 \quad . \tag{1.5}$$

[2] $\mathbf{R}$ is the displacement from the electron.

[3] The relative sign of the nuclear and magnetic scattering is obtained by comparing page 181 and 134 in Squires [3].

To calculate the cross section, measured using the instrument described in the next section, we will consider in more detail cross sections with specific final and starting neutron spin states σ' and σ in Eq. (1.5). The neutron has spin $\frac{1}{2}$ and we denote its two spin states by u and d. They may be regarded as 'spin up' and 'spin down' relative to a specified axis which is taken to be the z axis. By going through some elementary quantum mechanic algebra it can be shown that

$$\begin{aligned}
\langle u|F_N + a\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}|u\rangle &= F_N + aF_{M_{\perp z}} \\
\langle d|F_N + a\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}|d\rangle &= F_N - aF_{M_{\perp z}} \\
\langle d|F_N + a\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}|u\rangle &= a(F_{M_{\perp x}} + iF_{M_{\perp y}}) \\
\langle u|F_N + a\boldsymbol{\sigma}\cdot\mathbf{F}_{M_\perp}|d\rangle &= a(F_{M_{\perp x}} - iF_{M_{\perp y}}) \quad , \qquad (1.6)
\end{aligned}$$

where $a = \frac{\gamma r_0}{2\mu_B}$. If the incident beam is fully polarised i.e. either with all the incoming neutrons in spin state u or d, and the detector is impartial to the different spin states of scattered neutrons, the cross sections which are measured are $\left(\frac{d\sigma}{d\Omega}\right)_u = \left(\frac{d\sigma}{d\Omega}\right)_{u\to u} + \left(\frac{d\sigma}{d\Omega}\right)_{u\to d}$ and $\left(\frac{d\sigma}{d\Omega}\right)_d = \left(\frac{d\sigma}{d\Omega}\right)_{d\to u} + \left(\frac{d\sigma}{d\Omega}\right)_{d\to d}$. Using the expression for the matrix elements in Eq. (1.6) we have

$$\left(\frac{d\sigma}{d\Omega}\right)_u = N\frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta(\mathbf{q}-\mathbf{K})\left[|(F_N + aF_{M_{\perp z}})|^2 + a^2|F_{M_{\perp x}} + iF_{M_{\perp y}}|^2\right] \quad (1.7)$$

and

$$\left(\frac{d\sigma}{d\Omega}\right)_d = N\frac{(2\pi)^3}{V}\sum_{\mathbf{K}}\delta(\mathbf{q}-\mathbf{K})\left[|(F_N - aF_{M_{\perp z}})|^2 + a^2|F_{M_{\perp x}} + iF_{M_{\perp y}}|^2\right]. \quad (1.8)$$

It will be seen from Eq. (1.7) and Eq. (1.8) that the u and d cross-sections differ. This difference enables us to use a polarised incident beam to distinguish the magnetic scattering from the scattering of the nuclei (nuclear scattering). The method is as follows. In a polarised neutron diffraction experiment the polarisation of the incident neutron beam is either with all the neutrons in spin state u or d. For each Bragg reflection the scattered intensity is then measured with both polarisations of the incident beam, and the ratio of the intensities

is recorded: $R(\mathbf{K}) = I_u/I_d$. It is the measurement of these so-called flipping ratios which characterize a polarised neutron diffraction experiment. Measuring flipping ratios increases the sensitivity to the magnetic scattered neutrons. This is especially true when the $\mathbf{F}_M$'s are small relative to the F_N's. A polarised neutron experiment is, however, time consuming compared to an unpolarised diffraction experiment[4] and is therefore not recommended, for example, to measure the magnetic ordering of an antiferromagnetic compound, but is ideal for measuring the magnetization density in a paramagnet[5] like $PrBa_2Cu_3O_{6.24}$.

If we assume that the incident neutron beam has perfect polarisation, the crystal is centrosymmetric[6] and that $\mathbf{F}_M(\mathbf{K})$ is everywhere parallel to the direction of the polarisation vector and perpendicular to the scattering vector $\mathbf{q}$ then the flipping ratio takes on the particular simple form

$$R = \frac{(F_N + aF_M)^2}{(F_N - aF_M)^2} \quad . \tag{1.9}$$

Under the conditions satisfied by Eq. (1.9), and together with knowledge of F_N a magnetic structure factor F_M can be calculated from each flipping ratio measured using Eq. (1.9).

A nuclear structure factor F_N is equal to the Fourier transform of the nuclear density in the crystal. Similarly a magnetic structure $\mathbf{F}_M$ is related to the magnetic vector density $\mathbf{M}(\mathbf{r})$ of a unit cell by

$$\mathbf{F}_M(\mathbf{K}) = \int_{cell} \mathbf{M}(\mathbf{r}) e^{i\mathbf{K}\cdot\mathbf{r}} d^3\mathbf{r} \quad . \tag{1.10}$$

It is knowledge about the magnetisation (vector) density that we aim to obtain through rigorous PND measurements, see Chap. 5, and by analysing such data

[4]Because of a factor of two caused by the selection of neutrons in a specific spin state from the incoming reactor neutron beam and a factor of ~ 2 because the material used for selecting neutrons in a specific spin state typically has a reflectivity of $\sim 50\%$.

[5]For temperatures larger than the Néel temperature (T_N) of the magnetically ordered Pr moments in $PrBa_2Cu_3O_{6.24}$, Pr is described as a paramagnetic ion in that compound. Hence, for $T > T_N$ and from the point of view of Pr, $PrBa_2Cu_3O_{6.24}$ may be referred to as a paramagnet.

[6]This condition ensures that $\mathbf{F}_M$ is real.

using the Bayesian data analysis technique developed in Chap. 4. For the majority of applications it is assumed that the vector field $\mathbf{M}(\mathbf{r})$ is pointing in the same direction everywhere and that it is this component which is measured[7]. $\mathbf{M}(\mathbf{r})$ is almost always treated as a scalar field $M(\mathbf{r})$ where $\mathbf{M}(\mathbf{r})$ is assumed to point in some known direction.

From Eq. (1.2) it is evident that we may write the magnetisation density as the sum of two terms: $\mathbf{M} = \mathbf{M}_S + \mathbf{M}_L$, where $\mathbf{M}_S$ is the term arising from the spin of the magnetic electrons and $\mathbf{M}_L$ from the orbital magnetic moments. The spin magnetisation density can be written as $\mathbf{M}_S = -2\mu_B \boldsymbol{\rho}_S$, where $\boldsymbol{\rho}_S$ is the electron spin density. Therefore the spin magnetisation density is proportional to the electron density of the magnetic electrons in the crystal. $\mathbf{M}_L$ cannot be directly associated with electron density. It may be written as $\nabla \times \mathbf{M}_L = \nabla\phi - \mathbf{j}$, where $\mathbf{j}$ is the current density due to the orbital moment, ϕ, the so-called conduction current density (see Squires).

1.2 The D3 instrument at ILL

The polarised neutron diffraction data which will be presented in Chap. 5 were collected on the D3 instrument at the Institut Laue-Langevin (ILL), Grenoble. A drawing of the instrument is shown in Fig. 1.1 and the picture, except for a few modifications, is a copy of a picture taken from the D3/ILL web-site www.ill.fr. In this figure the horizontal neutron beam appears from the left. The beam originates from a 58MW high flux nuclear reactor. The initial polarisation of the beam is achieved by the polarising monochromator indicated in the figure. A monochromator of either $Co_{0.92}Fe_{0.08}$ or Cu_2MnAl (Heusler) is used. Both these crystals are ferromagnets and have the property that their magnetic cross section is equal in magnitude to their nuclear cross section. Thus, if a $\mathbf{B}$ field is applied perpendicular to the scattering plane then only the non spin flip cross sections are

[7]For a further discussion of this see e.g. Squires page 186-8.

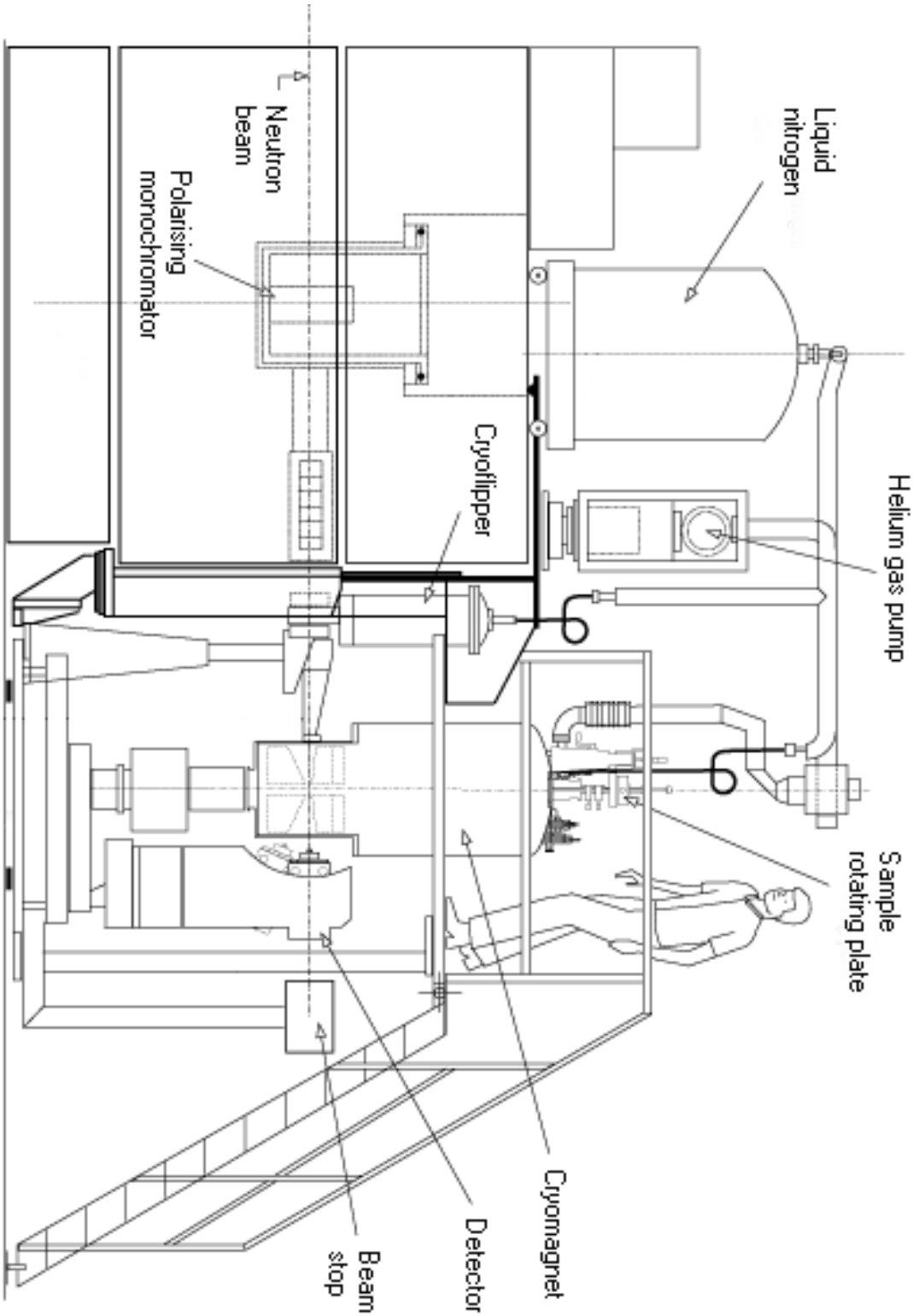

Figure 1.1: The layout of the D3 instrument at Ill, Grenoble.

non-zero (see Eq. (1.6)) and in addition one of these non spin flip cross sections will be zero where the nuclear and magnetic terms cancel out. This is how a polarised neutron beam is generated on D3.

The direction of polarisation of the incident beam is controlled by a cryoflipper. The cryoflipper consists of two magnetic field guides positioned closely either side of a thin sheet of superconducting niobium foil which lies transverse to the beam. In the flipper 'on' state the fields lie anti-parallel to each other and the neutrons will experience a rapid reversal of field across the narrow width of the superconducting foil. These conditions cause a change in neutron spin state relative to the field. In the flipper 'off' state the fields lie parallel and no reversal of field occurs across the superconducting foil. Hence, with the cryoflipper we can choose the neutron spin state to be either u or d before they hit the sample.

The scattered neutrons are detected using a single ^{3}He detector. To provide the low-temperature and high-field sample conditions a $4.6T$ Oxford instrument cryomagnet is used, providing stable sample conditions from $1.5K$ to $273K$ and from $1.5T$ to $4.6T$.

The instrument at ILL differs from the example described in the textbooks by Squires and Lovesey[8] as it includes an additional flipper plus an analyser between the sample and detector, this adds the possibility of selecting scattered neutrons in a specific spin state. The detected signal on D3 therefore includes neutrons in both spin states and for a general instrument geometry the two cross sections which can be measured on D3 are those in Eqs. (1.7-1.8).

[8] where the instrument at Oak Ridge National Laboratory is used as an example.

Chapter 2

Bayesian statistics

This chapter provides some background material on statistics and probability theory. There are many monographs on statistics, some of which are mentioned in Sec. 2.4, the last section of this chapter. In this brief general introduction to Bayesian statistics the emphasis will be on topics of statistics relavant to later parts of the thesis.

2.1 General introduction

Probability theory has a long and controversial history, and can be dated back to a gambler's dispute in 1654 discussed by two French mathematicians, Blaise Pascal and Pierre de Fermat. Founders of the Bayesian approach are Thomas Bayes who died in 1761 at the age of 59 and Pierre Simon Laplace (1749-1827)[1]. The further development of the Bayesian approach led to the branch of 'applied probability' which we now refer to as statistics; the subject of probability theory is therefore the foundation upon which statistics is built. An illustration of the different tasks of modern probability theory and statistics is found when tossing

[1]There are numerous historical accounts of probability theory, in which the authors do not alway agree upon the significance of individual scientist's contributions, see references in Sec. 2.4.

a coin once. The outcome of a coin experiment is either head, H, or tail, T. If we feel that the coin is unbiased so that there is no a priori reason to prefer heads or tails, it seems natural to assign

$$\Pr(H) = \Pr(T) = 1/2 \ .$$

For a given coin however, there is no mathematical way to determine whether the coin is "really" unbiased. But, if we believe the coin is unbiased then we can deduce the logical consequences of this assumption. Probability theory does not assess the validity of such a choice, but is used instead in the study of logical consequences that can be derived once an initial probability assignment is given. On the contrary, statistics is about assessing the hypothesis of the coin being unbiased or the (Bayesian) probability that it is unbiased. This is done by performing a number of experiments with the coin and studying the agreement between theory and empirical evidence. Statistics can be referred to as the art of making inferences about unknown parameters from incomplete data.

A number of textbooks treat probability on an axiomatic basis. Since this idea was first proposed[2] the ideas have been refined somewhat and mathematical probability is now part of a more general discipline known as measure theory. Although it is beyond the context of this thesis to outline this approach, it is still interesting that such an approach exists; it has the property that it defines probability without a reference to a specific interpretation of probability, and is concerned only that the probabilities are defined by a function satisfying some axioms (see e.g. [7] and [6]). As long as we are dealing with, in particular, the more mathematical aspect of probability theory, this procedure is fine. However, when applying statistics we may come across problems where it is necessary to attach an interpretation to the meaning of probabilities, and this choice can affect the final outcome of our statistical analysis. The two most well-known interpretations are

[2]By a Russian mathematician A. Kolmogorov in 1933 (see e.g. [6] and [7]).

- **Frequency interpretation of probability:** The probability of an event is defined as the limiting proportion of times the event occurs in an infinite sequence of independent repetitions of the experiment. In this interpretation, probabilities only make sense in connection with experiments which can (at least conceptually) be repeated.

- **Subjective (or Bayesian) interpretation of probability:** Subjective probability concerns the assessments of a given person("You") about things which are not known with certainty, in the presence of partial information. Think of $\Pr(A)$ as Your personal probability assessment of an 'event' A, an numerical measure of the strength of Your degree of belief that A will occur, in the light of available information. Bayesian probability can be associated with a wider class of events than those to which the frequency interpretation of probability pertains. In addition to the outcome of a repeatable experiment we can ask for the probability of the outcome of a non-repeatable experiment (e.g. that Denmark will beat England in football in the next European Championship) or propositions about nature (e.g. this surgical procedure results in increased life expectancy over the current procedure for disease X).

The Bayesian approach can be justified from various points of view. For instance from a set of reasonable requirements for how to reason in the presence of uncertainty while avoiding internal inconsistencies. It can be shown that for such requirements[3] to be satisfied we must perform inference according to the usual rules of probability. This gives credibility towards defining probability in the broader sense, hence as subjective probability.

[3]There exist various methods of axiomatizing these requirements, see e.g. Refs. [8],[9] and references therein.

2.1.1 The choice of notation used in this thesis

The notation used in this chapter and the rest of this thesis is a compromise between the notations found in textbooks on statistics (such as [6],[10],[11],[8],[9],[12],[13]) for students reading for a degree in statistics and the notation often employed by scientist in other fields.

Let the *probability* of an event A be denoted as $\Pr(A)$. Mathematically a *random variable(r.v.)* is defined as a function which maps the space of all possible events into the real numbers. For our purpose here we will think of a r.v. in the rather broad sense of any real-valued variable to which a probability distribution is associated. Let x be a r.v. and lets say x can take on the values $\mathcal{X} = \{1, 2, 3$ and $4\}$ only. Then the probability that x takes on the value $x = 1$ is written as $\Pr(x = 1)$, and to each of the possible outcomes of x there is associated a probability of that outcome. It is standard to introduce the "point" probability function $p(x)$ of a r.v. x. For the example above $p(x)$ is the function defined as $p(i) = \Pr(x = i)$, $i = 1, 2, 3, 4$. We will refer to $p(x)$ as the *probability density function* or just as the probability density (p.d.)[4]. We also say that a r.v. x is *distributed* as p, which is written in shorthand notation as $x \sim p$, and refer alternatively to $p(x)$ as the probability distribution of x. Therefore the abbreviation p.d. can be read either as 'probability density' or 'probability distribution'.

The difference between p and its 'big-brother' Pr is from a simplified point of view that Pr can handle more complicated arguments. Using Pr the probability of the event $1 \leq x \leq 3$ is written as $\Pr(1 \leq x \leq 3)$, on the other hand, using the

[4]A common alternative notation is to refer to $p(x)$ as the probability mass function when x is a discrete r.v. and probability density function when x is a continuous r.v., but $p(x)$ is chosen to be called a probability density (function) whether or not x is is discrete or continuous r.v.. This is also the notation used in Ref. [13].

p.d. to obtain the probability of this event we must perform the sum,

$$\Pr(1 \leq x \leq 3) = \sum_{i=1}^{3} p(i) \ .$$

For mathematicians this distinction between a probability[5] $\Pr(A)$ and a probability density $p(x)$ is important, and to make our notation compatible this notation is also adopted. For a one-dimensional continuous r.v. the probability and its probability density is related as

$$\Pr(a \leq x \leq b) = \int_a^b p(x)dx \ .$$

2.1.2 Parametric model

Lets us denote the unknown state of nature by θ and use capital Θ to denote the set of all possible states of nature. Typically, when experiments are performed to obtain information about θ, the experiments are designed so that the observations are distributed according to some probability distribution which has θ as an unknown parameter. In such situations θ will be called the *parameter* and Θ the *parameter space*. A *statistical parametric model* for observation x we write as $p(x|\theta)$, where θ is a parameter. For a given value of θ the data x is then predicted to be distributed as $p(x|\theta)$.

2.1.3 Bayes' theorem

For a one-dimensional continuous parameter space Θ Bayes's theorem reads

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{\int p(\theta)p(x|\theta)d\theta} \ . \tag{2.1}$$

Obviously if Θ is discrete the integral in the denominator becomes a sum. θ denotes our unknown state of nature and x an observation. Bayes' theorem then

[5]Mathematicians refer to Pr as a measure - a positive finite-additive set function, which also satisfies $\Pr(S) = 1$, where S is the set (sample space) of all conceivable outcomes.

describes how to update our prior belief, $p(\theta)$, in the light of data (observations) to give us $p(\theta|x)$. The terminology used for the different entries in Bayes' theorem is as follows

- $p(\theta)$, the p.d. of θ before the data are observed, is called the **prior** probability density or just the prior.
- $p(x|\theta)$, the p.d. of the data given a value for θ, this is called the **likelihood** (function). This is because we may think of $p(x|\theta)$ as a function of θ for a fixed observation x, which is how a likelihood function is defined.
- $p(\theta|x)$, the p.d. of θ given the data, is called the **posterior** probability density or just the posterior.

From elementary probability theory the *joint* distribution of θ and x is $p(\theta, x) = p(\theta)p(x|\theta)$. A *marginal* distribution is obtained by integrating out one of its arguments, e.g., the marginal p.d. of x is

$$p(x) = \int p(\theta)p(x|\theta)d\theta \ , \tag{2.2}$$

Hence, we see that the denominator in Bayes' theorem is the marginal p.d. $p(x)$, the unconditional probability density of x. For a fixed data set $p(x)$ is a constant and acts as a normalisation constant.

It is trivial to derive Bayes' theorem from modern axiomatized probability theory[6]. However, it appears as a major conceptual step in the history of statistics, being the first *inversion* of probabilities; since it aims at retrieving the parameters which characterize the object under study from 'effects' (observations), by obtaining the posterior $p(\theta|x)$ from $p(\theta)$ and $p(x|\theta)$. While this inversion theorem is quite natural from a probabilistic point of view, it contains another important element, the prior density $p(\theta)$. It is based on the idea of putting causes (parameters) and effects (observations) on the same conceptual level, since here both of

[6]Use $p(x|\theta)p(\theta) = p(\theta, x) = p(\theta|x)p(x)$ and Eq. (2.2).

them have probability distributions. Thus, from a statistical modeling viewpoint there is little difference between observations and parameters, since conditional manipulations allow for an interplay of their respective roles. Finally we observe that a *Bayesian parametric statistical model* is then made up of a parametric statistical model, $p(x|\theta)$ for the data, and a prior distribution on the parameters, $p(\theta)$.

2.1.4 Nuisance parameters

Suppose $\theta = (\gamma_1, \gamma_2)$ has two components. In some settings one component, γ_1 say, will be of primary interest. The second component is then often described as a *nuisance parameter*. Given data x, one obtains the joint posterior $p((\gamma_1, \gamma_2)|x)$ of (γ_1, γ_2) via Bayes' theorem. If γ_1 is the focus of interest, beliefs about γ_1 are described by the marginal posterior of γ_1

$$p(\gamma_1|x) = \int p((\gamma_1, \gamma_2)|x) d\gamma_2 \ .$$

Thus, in the Bayesian framework, one handles nuisance parameters simply by integrating them out of the posterior.

2.2 Prior Distributions

From an applied viewpoint what makes the main difference between Bayesian statistics and non-Bayesian statistics is that in a Bayesian analysis we assign a prior distribution.

2.2.1 Subjective determination of the prior

In the subjective Bayes world, the prior distribution is intended to capture the information available about the parameter θ before the data is observed. For a discrete parameter space one needs to determine the subjective probability assigned

to each event of the parameter space Θ, which is in principle straightforward. For a continuous parameter space, however, several approaches are possible. For instance a continuous prior can be achieved by the histogram approach: If Θ is an interval of the real line, one can discretize Θ and assess a subjective probability for each discrete sub-interval. This histogram (a step function) may be used as a continuous prior or a smoothed version of the histogram.

Alternatively, the prior can be matched to a particular functional form by restricting it to a particular family of distributions. Say, as an example, that θ is normally distributed (same as Gaussian distributed) with mean μ and standard deviation σ, i.e.

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(\theta-\mu)^2}{\sigma^2}) \quad . \tag{2.3}$$

Then by providing values of μ and σ a prior assignment is made. In Eq. (2.3) μ and σ are constants rather than parameters, as θ is. To make this absolutely explicit we may write $p(\theta|\mu,\sigma)$ instead of $p(\theta)$ in Eq. (2.3), i.e.

$$p(\theta|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(\theta-\mu)^2}{\sigma^2}) \quad . \tag{2.4}$$

μ and σ are referred to as *prior model constants* in this thesis, and in Eq. (2.3) and Eq. (2.4) we therefore have the following items in the prior distribution

- μ and σ are both **prior model constants**.
- θ is a **parameter**.

Hence, before a Bayesian analysis with a prior of the type in Eq. (2.4) can be started, values for both μ and σ must be provided. In later chapters it will be important to keep in mind the difference between what are prior model constants and parameters in the prior distribution, as e.g. in Chap. 4. In addition it is possible to change a prior model constant, for example σ, into a parameter. Obviously this must require a change in the prior p.d. in Eq. (2.4), such that $p(\theta|\mu,\sigma)$ becomes $p(\theta,\sigma|\mu)$. The prior then contains two parameters and by

the elementary probability relation for conditional probabilities this prior can be decomposed into

$$p(\theta, \sigma|\mu) = p(\sigma|\mu)p(\theta|\mu, \sigma) \ . \tag{2.5}$$

Of the two terms on the right hand side of Eq. (2.5) we already know $p(\theta|\mu, \sigma)$ from Eq. (2.4). Hence, to complete the transformation of σ from a prior model constant into a parameter we have to supply the p.d. $p(\sigma|\mu)$.

It is in general desirable that small changes in the choices of prior model constants do not largely affect the conclusions drawn from the posterior distribution. It is therefore important to assess the sensitivity of the conclusions to aspects of the prior, e.g. to analyse how changes in the choices of the prior model constants affect the posterior. Such a study may be difficult to perform depending on the dimensionality and analytical complexity of the posterior (e.g. characterizing how changes in a prior model constant a affect a posterior of the form $e^{-(x-a)^2}$ is clearly simpler than a posterior with the form $e^{-\sum_{i=1}^{I} x_i \sinh^{-1}(x_i/a)}$), this is called a **robustness** analysis. More precisely, for the purpose of this thesis, a robustness analysis is the functional investigation of the posterior distribution as a function of its prior model constant(s). In Chap. 4 the MaxEnt method is studied from the point of view of the mode of a posterior probability distribution, and a robustness analysis is performed. Such an analysis is novel in the context of MaxEnt and provides an insight into improving the overall performance of the "MaxEnt" technique when used in connection with polarised neutron diffraction data and possibly other types of diffraction data.

2.2.2 Non-informative priors

An interesting class of priors are the so-called non-informative priors. Such a prior can be appropriate in situations in which little is known about a parameter θ, and it is hard to justify one particular density assignment $p(\theta)$ from any other assignment $p'(\theta) \neq p(\theta)$. We may therefore take another point of view and try to

aim for a prior assignment which "favours" no particular value of the parameter over any other; this is what we refer to as a non-informative prior assignment. If the parameter space Θ is finite, with n elements, then the obvious non-informative prior places mass n^{-1} at each possible parameter value, $p(\theta) = \frac{1}{n}$ for all θ in Θ. If Θ is infinite, it turns out that there are several senses in which one might seek a non-informative prior. Hence, there may be distinct priors which each claim to be non-informative.

One characteristic of non-informative assignments for infinite parameter spaces is that the resulting prior is often what is called improper. This means that it cannot be normalized to unity, i.e. $\int p(\theta)d\theta$ integrates, or $\sum p(\theta)$ sums, to infinity. Such a prior cannot be classified as a probability distribution since formally for a distribution to be a probability density it must be normalizable. However, it still makes sense to apply such priors as long as the posterior distribution is normalizable in θ.

The two most common non-informative priors for a continuous parameter space are

Non-informative priors for location parameters

[7] Suppose the parameter space Θ is continuous. If the p.d. of x given θ is of the form $f(x-\theta)$, so that it depends only on the difference between x and θ, the p.d. is said to be a *location p.d.*, and θ is called a *location parameter*.

To derive a non-informative prior for θ, consider observing $y = x + c$ for some fixed c. If $\gamma \equiv \theta + c$, it is clear that y has p.d. $f(y-\gamma)$. The (x, θ) and (y, γ) problems are thus similar in structure. Further, if Θ is equal to the real line, i.e. $\theta \in \Theta = \mathbf{R}$, then both problems also have the same parameter space, and it may therefore seem natural to insist that a prior should be the same in each problem. This implies that the form of the function $p(\theta)$ stays the same under

[7] This sub-section is slightly technical and may be skipped.

the coordinate transformation $\theta \to \gamma = \theta + c$, hence, $p(\theta)d\theta = p(\gamma)d\gamma$, where $p(\theta)$ and $p(\gamma)$ have the same form (e.g. $p(\theta) = \theta^3$ implies $p(\gamma) = \gamma^3$ and vice versa). But by direct substitution $\theta \to \gamma = \theta + c$: $p(\theta)d\theta = p(\gamma - c)d\gamma$, which together with $p(\theta)d\theta = p(\gamma)d\gamma$ means that the function p must satisfy $p(\gamma) = p(\gamma - c)$ for all real values of c. Only if p is a constant function is this satisfied. Hence, for a location parameter a 'natural' non-informative prior is

$$p(\theta) = const. \quad \theta \in \mathbf{R} \ , \tag{2.6}$$

which is improper.

Non-informative prior for scale parameter

[8] The family of p.d. $\frac{1}{\sigma}p\left(\frac{x}{\sigma}\right)$, indexed by σ, where $\sigma > 0$, is called a *scale family* and a member of the family a *scale p.d.*. The parameter σ is called a *scale parameter*.

To derive a non-informative prior for σ we follow the same procedure which leads to the assignment of a location parameter in Eq. (2.6). Consider $y = cx$ where $c > 0$. If $\gamma = c\sigma$, it follows that y has the p.d. $\frac{1}{\gamma}p\left(\frac{y}{\gamma}\right)$. The (x, σ) and (y, γ) problems now have the same structure. It can therefore be argued that they should then have the same prior, i.e. $p(\sigma)d\sigma = p(\gamma)d\gamma$. Further the coordinate transformation $\sigma \to \gamma = c\sigma$ implies $p(\sigma)d\sigma = \frac{1}{c}p\left(\frac{\gamma}{c}\right)d\gamma$. This requires

$$p(\gamma) = \frac{1}{c}p\left(\frac{\gamma}{c}\right)$$

for arbitrary values of c. Only functions $p \propto \frac{1}{\gamma}$ satisfy this. Hence, for a scale parameter we may assign the following non-informative prior assignment

$$p(\sigma) = \frac{1}{\sigma} \quad , \quad \sigma > 0 \ , \tag{2.7}$$

which, like the prior in Eq. (2.6), also is improper.

[8] This sub-section is slightly technical and may be skipped.

As an example of a scale family we have the normal distribution $N(\theta, \sigma^2)$, indexed by the scale parameter σ - the standard deviation[9]. Likewise, $N(\theta, \sigma^2)$ can also be looked upon as a location p.d. with θ as the location parameter. Together $N(\theta, \sigma^2)$ may therefore be referred to as a *location-scale p.d.*.

For further information on non-information priors see, for example, [8],[12],[9] and references therein.

2.3 Chi-square random variable and PND data

2.3.1 Chi-square distributions

Chi-square distributions appear most often in the context of normally (or Gaussian) distributed r.v.. If x is a standard normal r.v., i.e. $x \sim N(0, 1)$, the p.d. of $u = x^2$ is called the chi-square distribution with 1 degree of freedom, and is denoted χ_1^2. It is useful to note that if $x \sim N(\theta, \sigma^2)$, then $\frac{x-\theta}{\sigma} \sim N(0, 1)$, and therefore $[\frac{x-\theta}{\sigma}]^2 \sim \chi_1^2$.

Further if $u_1, u_2, \ldots, u_n$ are independent chi-square r.v. with 1 degree of freedom, then the distribution of the *statistic*[10] $v = u_1 + u_2 + \cdots + u_n$ is called the chi-square density with n degrees of freedom and is denoted by χ_n^2, and

$$\chi_n^2(v) = \frac{v^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-v/2} \quad , \quad v \geq 0 \; . \tag{2.8}$$

$\Gamma(n)$ is the gamma function.

2.3.2 PND data and its likelihood

Take the example of a polarized neutron diffraction (PND) experiment. A description of the PND technique is given in Chap. 1. For the purpose here it is

[9]Statisticians often use the shorthand notation $N(\theta, \sigma^2)$ for the normal distribution (same as the Gaussian distribution).

[10]A r.v. formed from a set of other r.v. is often referred to as a statistic. The idea of a statistic is to find an efficient way of summarizing the information contained in a set of r.v..

enough to know that a PND experiment is designed such that information about a number of magnetic structure factors, where P is the number of data points, is obtained by counting the number of neutrons, in a given time-interval, entering into a detector at P distinct points in space - at P so-called Bragg positions. To each Bragg position we have an unknown parameter, a magnetic structure factor F. By counting the number of neutrons entering the detector at a fixed Bragg position and doing some magic black-box pre-data processing we obtain an observed magnetic structure factor value, denoted by F^{obs}, and associated with this measurement a standard deviation σ. To summarize this observation we write down the p.d. of F^{obs} given a value of the parameter F, the likelihood,

$$p(F^{obs}|F,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{(F^{obs}-F)^2}{\sigma^2}\right) \ , \tag{2.9}$$

or alternatively we may write $F^{obs} \sim N(F,\sigma^2)$. The likelihood in Eq. (2.9) is somewhat non-standard because the standard deviation σ is a quantity not known before F^{obs} is observed and should therefore, in principle, not appear on the right hand side of the vertical line in the argument of $p(F^{obs}|F,\sigma)$ in Eq. (2.9), but instead on the left hand side together with F^{obs}. However the parametric statistical model in Eq. (2.9) is the one always used for the analysis of PND data, and we will also use this expression for the likelihood, but it may be possible to improve on Eq. (2.9) in the future.

In a full PND experiment P observations are obtained. Index these observations as $\mathbf{F}^{obs} = (F_1^{obs}, F_2^{obs}, \ldots, F_P^{obs})$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_P)$, and parameters as $\mathbf{F} = (F_1, F_2, \ldots, F_P)$. The likelihood for P independent observed magnetic structure factor values follows directly from Eq. (2.9) and is

$$p(\mathbf{F}^{obs}|\mathbf{F}) \equiv p(\mathbf{F}^{obs}|\mathbf{F},\boldsymbol{\sigma}) = \frac{1}{(2\pi)^{P/2}\prod_{i=1}^{P}\sigma_i}\exp(-\frac{1}{2}Q) \ , \tag{2.10}$$

where the statistic (or r.v.) Q is

$$Q = \sum_{i=1}^{P}\left(\frac{F_i^{obs}-F_i}{\sigma_i}\right)^2 \ . \tag{2.11}$$

In the remainder of this thesis $\boldsymbol{\sigma}$ will be suppressed in the argument of the likelihood in Eq. (2.10) and will be assumed 'given'. The likelihood in Eq. (2.10) is an example of what is called a *multivariate normal distribution*, and in this simple case consists of a product of P independent normal distributions. Occasionally it is found to be advantageous to write Eq. (2.10) in the vector-matrix form

$$p(\mathbf{F}^{obs}|\mathbf{F}) = \frac{1}{(2\pi)^{P/2}\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-1/2(\mathbf{F}^{obs}-\mathbf{F})^T\boldsymbol{\Sigma}^{-1}(\mathbf{F}^{obs}-\mathbf{F})\right] \ , \qquad (2.12)$$

where $\boldsymbol{\Sigma}$ is called the covariance matrix, because its elements σ_{ij} are the variances and covariances of the density in Eq. (2.12),

$$\langle (F_i^{obs}-F_i)(F_j^{obs}-F_j)\rangle_{p(\mathbf{F}^{obs}|\mathbf{F})} = \sigma_{ij} \ . \qquad (2.13)$$

For the parametric statistical model in Eq. (2.10) and Eq. (2.12) the covariance matrix elements are

$$\begin{aligned} \sigma_{ij} &= 0 \qquad i \neq j \\ \sigma_{ij} &= \sigma_i^2 \qquad i = j \end{aligned}$$

This vector-matrix notation will be used in Chap. 4. Also, the multivariate normal distribution in Eq. (2.12) can be written in shorthand notation as $\mathbf{F}^{obs} \sim N(\mathbf{F}, \boldsymbol{\Sigma})$.

2.3.3 Chi-square constraint and PND data

From the likelihood function in Eq. (2.10) it is clear that $\frac{F_i^{obs}-F_i}{\sigma_i} \sim N(0,1)$ for all $i = 1, 2, \ldots, P$, thus, each of the terms in the sum of the statistic Q in Eq. (2.11) is chi-squared distributed with 1 degree of freedom and therefore $Q \sim \chi^2_P$. The statistic Q is one example of a statistic which can be used to indicate how well a given data set $F_1^{obs}, F_2^{obs}, \ldots, F_P^{obs}$ fits a given set of parameter values $F_1, F_2, \ldots, F_P$ (and a set of standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_P$). Q is called a *chi-square statistic*, for obvious reasons, and is probably the most common summary statistic used in experimental physics.

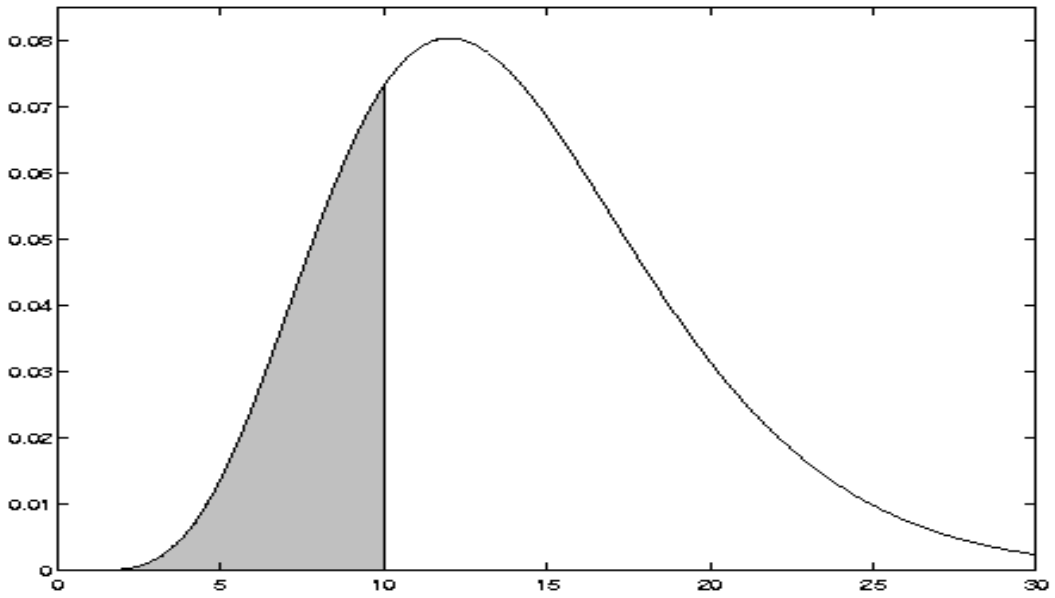

Figure 2.1: The chi-square distribution with 14 degrees of freedom, χ^2_{14}. For our example $Q = 10$. The area of the shaded portion of the distribution is the probability that we would have observed a smaller value of Q than $Q = 10$ if the model were correct.

Intuition tells us that the larger the value of Q, the worse the fit between data and parameters is. This picture fits in well with the following example. Let the number of data collected be $P = 14$ and the value of Q for this data set be $Q = 10$ relative to a particular model. Q is distributed as χ^2_{14} and this distribution is shown in Fig. 2.1. It depicts the probabilities of observing various values of Q if a given model was correct (**F** plus $\boldsymbol{\sigma}$ given). From this graph it is seen that values of 10-20, say, would not be unusual if the model were correct. Thus, the observation of a Q value between 10 and 20 would not give reason to seriously doubt the model. However, suppose a value of 28 is observed; since values this large or larger would rarely occur if the model was correct, such a value may cast doubt on the adequacy of the model. In general, if Q is observed to take on the value γ, for instance, an often used Frequentist measure of evidence

against the model is the p-value,

$$\text{p-value} = \Pr(Q > \gamma | \text{ model is correct}) \ .$$

In the case at hand, numerical evaluation of the p-value corresponding to $Q = 10$ gives p-value = 0.76. In other words, if the model were correct, Q values as large or larger than the Q=10 would take place 76% of the time if one repeated the experiment many times. Therefore from a Frequentist point of view there is little reason to doubt the model on the basis of the chi-square statistic Q.

In fitting a model to diffraction data the chi-square statistic Q is often used to put a constraint on the likelihood in Eq. (2.10), by typically requiring that we can only accept a likelihood with Q values less than or equal to the number of data points P, or the stronger constraint that Q must be equal to P. Such a constraint is often referred to in the scientific literature as a **chi-square constraint**.

The χ^2_P distribution has the mean value P and standard deviation $\sqrt{2P}$, and in addition if $P \to \infty$ then $\chi^2_P \to N(P, 2P)$. This is illustrated in Fig. 2.2 which shows the chi-square distribution with 10, 100 and 1000 degrees of freedom. As is seen from these figures, χ^2_P gradually becomes more sharply distributed around P and with a peak shape resampling that of a Gaussian distribution. However, even when P is large, for any *one* given data set there is no guarantee that this particular data set has a $Q \simeq P$ relative to the true model. Hence, the use of a systematic Q constraint in a Bayesian statistical analysis, for example $Q = P$, should be justified before use. For the case of the MaxEnt method this issue is discussed in detail in Chap. 4.

2.3.4 Posterior distribution and PND data

For a PND data set Bayes' theorem reads

$$p(\mathbf{F}|\mathbf{F}^{obs}) \propto p(\mathbf{F}) p(\mathbf{F}^{obs}|\mathbf{F}) \ , \tag{2.14}$$

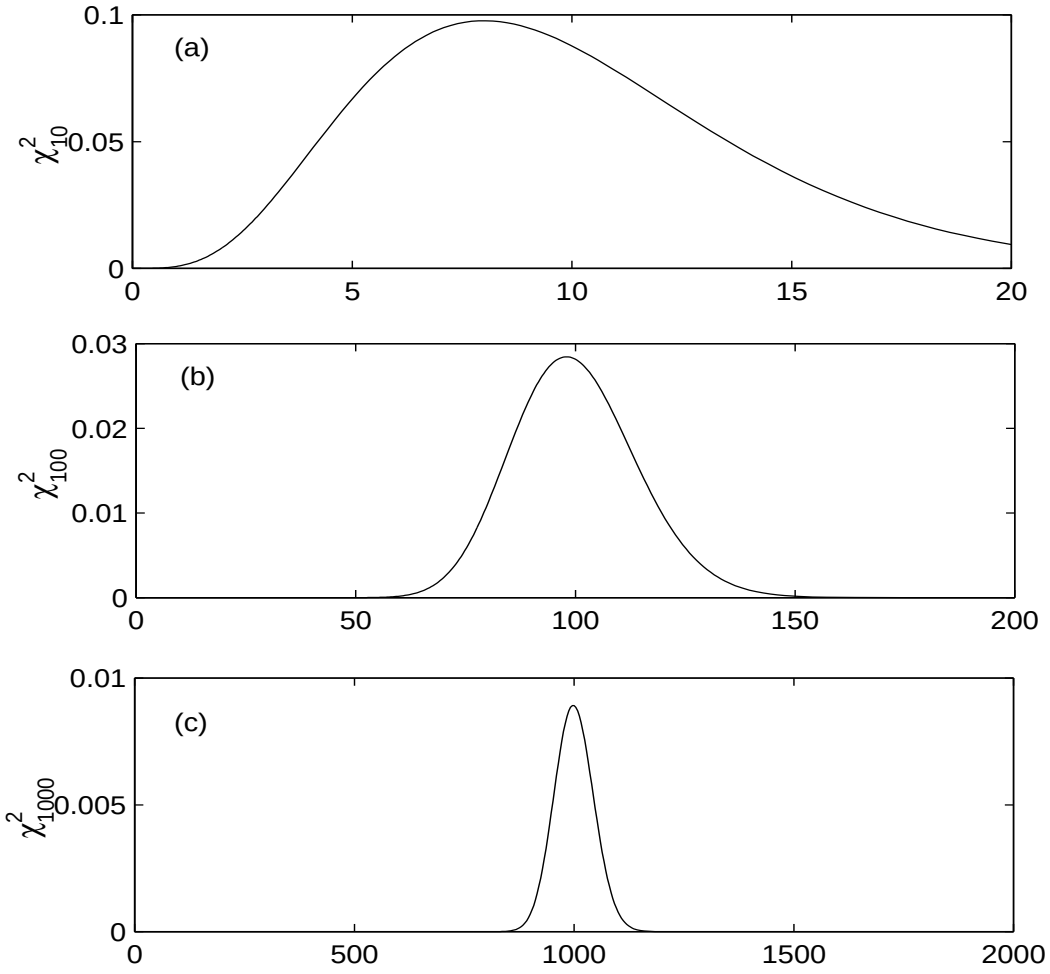

Figure 2.2: The chi-square distribution with 10, 100 and 1000 degrees of freedom. χ^2_{10} in Fig. (a), χ^2_{100} in Fig. (b) and χ^2_{1000} in Fig. (c).

where the marginal p.d. $p(F^{obs})$ is a constant (for a fixed data set). In Bayesian analysis obtaining an expression for the posterior distribution is, in principle, the final stage of the analysis. If the parameter space is one, two, or maybe even three dimensional we can simply present our posterior belief of the parameters, given the data, by plotting out the posterior in a graph. However, when the parameter space becomes of dimension 4 or higher, this task becomes increasingly difficult and other means for summarizing the posterior density must be used. To summarize a multi-dimensional posterior it is typical to use so-called 'point estimates' of the posterior, like the mode and the mean; the mode of a p.d. is just another name for the maximum of a p.d..

Intuitively the mode gives us the "most likely" value of the parameters given the data, and is sometimes called the maximum Bayesian likelihood estimator [8] or alternatively it may be referred to as the maximum *a posterior* estimator. In this thesis, the maximum a posterior estimator of a parameter will be denoted by the superscript 'max', like θ^{max} or $\mathbf{F}^{max}$.

2.4 References to textbooks in statistics

An introduction to Bayesian statistics for experimental physicists is Ref. [14]. As an example of a more advanced and detailed Bayesian textbook is Ref. [8]. A good, short and pure mathematical introduction to probability theory can be found in chapter 13 and 14 of Ref. [7]. Finally, Ref. [6] is an example of a well written modern textbook on statistics used at the statistical and mathematical departments of Oxford. Ref. [7] and Ref. [6] have the characteristic that they leave out any discussion about how to interpret probability, contrary to the first two where the Bayesian probability interpretation is preferred. The advantage of defining probability on an axiomatic basis is that it gives no opinion about how to interpret a probability and within the same textbook both Bayesian statistical methods and non-Bayesian methods are discussed. Other good references are the

thesis by Johnston [15] and the thesis by Macaulay [16]. Further, I would like to make references to the following textbooks and theses I have read in connection with this thesis [10],[17],[18],[11],[19],[20],[21]. In addition to my supervisors I would also like to thank Dr. Peter Clifford (stat. dep., Oxford) for interesting discussion on Bayesian statistic and Dr. Mario C. Borja (also stat. dep., Oxford).

Chapter 3

MEM in data analysis

3.1 Introduction

The aim of this chapter is to survey some important papers describing the maximum entropy method (MEM) in data analysis. The main emphasis will be on physical problems where the quantity of interest, say an object O, is linearly related to a set of measured data. In general we can write such a linear relation between data $d_1, d_2, \ldots, d_M$ and object as

$$d_m = \int_D A_m(x) O(x) dx \quad , \tag{3.1}$$

where $O(x)$ is zero outside the known region D, and the function $A_m(x)$ may be referred to as the measurement kernel of a physical problem. Later on in this thesis I will discuss the example where d_m represent magnetic structure data obtained by polarized neutron diffraction, O is the magnetization density in the unit cell of a crystal, and A is a complex exponential, so that the integral is a Fourier transform.

The object O in Eq. (3.1) is introduced as a one-dimensional continuous function, but it could just as well be a discrete function and/or of a higher dimension

than one. The discrete version of Eq. (3.1) may be written as

$$d_m = \sum_{i=1}^{I} A_{mi} O_i \ , \tag{3.2}$$

where A_{mi} is a matrix element of a $M \times I$ measurement kernel matrix $\mathbf{A}$.

The specific form of the measurement kernel $A_m(x)$ depends on the application. Problems of this type have arisen in a number of diverse fields; radio astronomy, crystallography, optics and so on. As mentioned above taking $A_m(x) = \exp(ik_m x)$ reduces Eq. (3.1) to a Fourier transform, representing the physical situation which applies for example, to polarised neutron diffraction. If we let $A_m(x) = A(x_m - x)$ then the data d_m measures the convolution of two functions $A(x)$ and $O(x)$, a physical situation which often arises in optics where $A(x)$ becomes a point spread function. The terminology used above is partly borrowed from optics, but we will also refer to $O(x)$ as an image (or picture) or any other quantity which is linearly related to its measured data. The process of obtaining an estimate of O is often referred to as a reconstruction process: if O is the true physical object then we would like to find the most likely reconstruction $\hat{O}$, of O, that the data can support.

This chapter will not exclusively be concerned with linear problems of the type in Eq. (3.1). The reason for this is that, historically, MEM was first applied in data analysis to the spectral analysis of time series. Here a measured time series $x_0, x_1, \ldots, x_n, \ldots$, and its individual data points x_n are not linearly related to its so-called power spectrum $S(\omega)$. Instead, the power spectrum is given in terms of expectation values $\langle x_{n+i} x_n \rangle = R(i)$, where $R(i)$ is the autocorrelation function and

$$R(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) e^{im\omega} d\omega \ , \tag{3.3}$$

and in Eq. (3.3) it is assumed that the time series $x_0, x_1, \ldots$ is measured at times $t = 0, 1, \ldots$.

Objects described by Eq. (3.1) are estimated using different methods and algorithms than objects described by Eq. (3.3), as discussed in detail later.

We will find that MEM consideration together with a few extra assumptions, leads to an expression for the entropy of a time series as $H \propto \int \ln S(\omega) d\omega$, whereas for problems of the type in Eq. (3.1) the entropy is often expressed as $H \propto \int O(x) \ln O(x) dx$. Today a relatively large number of papers are regularly published which are either developments or applications of a method called "MEM" in solving problems of the type in Eq. (3.1). The reason for putting MEM in quotation-marks here is that often these newer papers no longer use the MEM in terms of the textbook definition of the MEM (see next section), but instead are applications of Bayes law or 'regularizer' methods or other methods. For this reason, in Sec. 3.3.3, another definition of the maximum entropy method is introduced to encapsulate how the maximum entropy method is most frequently employed in data analysis today.

3.1.1 Textbook definition of MEM

The MEM has existed since before its first application in data analysis. The MEM as defined in statistical or information theoretical textbooks is a method, for deciding which probability distribution should be chosen among a set of possible distributions; i.e. MEM is a statistical tool.

To introduce MEM, consider the example of tossing a die. Define the discrete random variable (abbreviated as r.v.) x which can take the values one to six. Assume we do not know whether the die is fair, i.e. we do not know the probability of one coming up, two coming up and so on. Instead we are given only the information that the mean value of any number coming up when throwing the dice is equal to $\bar{x}$. Then among the infinite number of possible distributions which satisfy the normalization constraint $\sum_{i=1}^{6} p(x_i) = 1$ and mean value constraint $\sum_{i=1}^{6} x_i p(x_i) = \bar{x}$, the maximum entropy method[1] selects the maximum

[1]Or principle, depending on textbook; in the context of statistical mechanics it is often referred to as a principle rather than method. See e.g. Ref. [22], [23], [24], which all use [25] as

entropy distribution (MED), which is defined as the distribution that maximizes the functional $H[p] = -\sum_i p(x_i) \ln p(x_i)$ subject to the normalization and mean value constraints.

In general, we define the (information) entropy of a discrete r.v. x with range $\mathcal{X}$ as

$$H[p] = -\sum_{x \in \mathcal{X}} p(x) \ln p(x) \quad . \tag{3.4}$$

Hence, using Eq. (3.4) we associate a number $H[p]$ with any possible probability distribution $p(x)$. Given that we know a set of properties about $p(x)$ of the linear form

$$\sum_{x \in \mathcal{X}} g_k(x) p(x) = \langle g_k(x) \rangle_p = \omega_k \quad , \; k = 1, 2, \ldots, K \; , \tag{3.5}$$

then we introduce our first definition of MEM

Definition 1 (Discrete variable MEM) *The MEM for a discrete r.v. x is to choose, from all distributions which satisfy constraint equations of the type in Eq. (3.5), the distribution $p^*(x)$ that maximizes the entropy $H[p]$ in Eq. (3.4). $p^*(x)$ is referred to as the maximum entropy distribution (MED).*

The solution to the maximization problem in the above definition is well known. By forming the functional

$$J[p] = H[p] + \lambda_0 \sum_{x \in \mathcal{X}} p(x) + \sum_{k=1}^{K} \lambda_k \sum_{x \in \mathcal{X}} p(x) g_k(x) \tag{3.6}$$

and by setting $\partial J / \partial p(x)$ equal to zero we obtain the following form of the MED

$$p^*(x) = \exp(\lambda_0 - 1 + \sum_{k=1}^{K} \lambda_k g_k(x)) \; , \tag{3.7}$$

where the Lagrangian multipliers $\lambda_0, \lambda_1, \ldots, \lambda_K$ are chosen so that p^* satisfies the normalization condition and the K constraints in Eq. (3.5). Normalization is automatically satisfied if we write the MED as

$$p^*(x) = \frac{\exp(\sum_{k=1}^{K} \lambda_k g_k(x))}{\sum_{x \in \mathcal{X}} \exp(\sum_{k=1}^{K} \lambda_k g_k(x))} \; . \tag{3.8}$$

their main reference, when referring to MEM as the "maximum entropy principle".

There is some controversy with respect to the definition of the MEM for a continuous r.v.. Without going too much into this issue, for future reference, the MEM for a continuous r.v. will be defined in the following three ways. Firstly, as a simple extension of the definition of entropy in Eq. (3.4) we could define the entropy for a continuous r.v. x as [See e.g [26], [27]]

$$h[p] = -\int_{\mathcal{X}} p(x) \ln p(x) dx \ , \tag{3.9}$$

where $p(x)$ is equal to zero outside the continuous support set $\mathcal{X}$. Our second definition then follows

Definition 2 (Continuous variable MEM) *The definition of the MEM for a continuous r.v. x is the same as for a discrete variable MEM except that we maximize the continuous entropy in Eq. (3.9), subject to constraints $\int_{x\in\mathcal{X}} g_i(x)p(x) = \omega_i$, $i = 1, 2, \ldots, K$. As in definition 1 we denote MED by $p^*(x)$.*

A difference between the discrete entropy in Eq. (3.4) and the definition of continuous entropy in Eq. (3.9) is that in the discrete case the measure of entropy is invariant under change of variables ($x \to y = f(x)$), whereas the continuous entropy in Eq. (3.9) will, in general, change under coordinate transformation, for more details see [26]. Jaynes [28][29] argued in favour of using the following definition of entropy for a continuous random variable, which is invariant under coordinate transformations,

$$h[p; p_0] = -\int_{\mathcal{X}} p(x) \ln \frac{p(x)}{p_0(x)} dx \ , \tag{3.10}$$

where p_0 is a so-called natural 'invariant' non-informative prior p_0 for the problem. The modulus of the integral in Eq. (3.10) is known under other names, including the relative entropy, Kullback Leibler distance [27][30] and entropy distance[8] and is denoted in this context as $D(p||p_0) = -h[p; p_0]$. One may also come across names such as cross entropy, information divergence and information for discrimination [27]. Unfortunately, non-informative distributions, such as p_0,

are often highly non-trivial to assign, this makes the definition of entropy in Eq. (3.10) somewhat ambiguous [9],[12],[8]. However, it may still be useful, as articulated strongly in Jaynes' publications [28],[29],[31], and references therein. The definition of entropy in Eq. (3.10) leads us to

Definition 3 (Relative entropy MEM) *The same as for a continuous variable MEM except that we need to supply a non-informative distribution p_0, and maximize the entropy $h[p; p_0]$ in Eq. (3.10) instead of the entropy $h[p]$ in Eq. (3.9). Notice, if p_0 is the uniform distribution then definition 3 and 2 are equivalent.*

If we have no information of the form in Eq. (3.5) then the MED becomes $p^*(x) \propto p_0(x)$. Jaynes [29] used this observation to argue that the intuitive meaning of p_0 is a distribution describing "complete ignorance"; hence, a non-informative prior[2]. An alternative to Jaynes' interpretation, which we will come across a number of times later in this thesis, is to interpret the reference distribution p_0 not as being some non-informative distribution but as a density which summarizes any prior knowledge which we have before we are given the data. By this interpretation the MEM takes on the role of Bayes theorem; updating prior information, p_0, with additional data to obtain p^*. This may seem an appealing thing to do, but it is neither Bayesian statistics nor any kind of Frequentist approach. The notion that the MEM can update prior information p_0, with new data to obtain p^* could, perhaps, be classified as MEM-inference or MEM-statistics - as opposed to Bayesian statistics, which uses Bayes' theorem to update information as explained in Chap. 2. We define our last definition of MEM as

Definition 4 (Bayes law MEM) *Same as the relative entropy definition of MEM, but where p_0 can now summarize any available information for a given problem.*

[2]Non-informative priors are described in detail in, for example [12] or [8] and Chap. 2 in this thesis.

The credibility of the MED is a matter of opinion. The explanation for this is that the MED is ultimately connected with the choice of a distribution from a set of possible ones, and the MEM does not by itself offer any means of testing this choice. The significance of a MED, therefore depends on one's personal belief in the credibility of such a distribution; i.e. is it believed that MED is *the only* possible distribution one can accept, or is the MED interpreted as a sensible distribution to use and one which should be compared with other possible choices? For instance, if we chose $p(x) \neq p^*(x)$ then what is the quantitative 'error' associated with this choice as opposed to choosing $p(x) = p^*(x)$.

One limitation of the MEM is that there exist very few cases in practice where there is information available strictly of the form in Eq. (3.5). However, if we include noise corrupted data which are of the form in Eq. (3.5) with additional noise term, e_m, and neglect such noise terms, then we can apply MEM using these noise-free data as MEM-constraints.

Let us use the MEM in a couple of examples.

EXAMPLE 1: The throwing of a die with no constraints on the probability distribution other than normalization. In this case the sum over k in Eq. (3.7) vanishes, and the distribution that maximizes the entropy is the uniform distribution, $p^*(x_i) = 1/6$.

EXAMPLE 2: If we have the constraints $\langle x \rangle_p = \mu$ and $\langle (x-\mu)^2 \rangle_p = \sigma^2$. Applying these constraints to Eq. (3.8) and generalising to a continuous distribution we find that the MED is

$$p^*(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$$

which is the Gaussian (or normal) distribution.

3.1.2 Boltzmann and entropy

Boltzmann in 1877 was the first to write down a function of the form $\sum x \ln x$, and to associate it with Clausius's thermodynamic definition of entropy. In doing

that he introduced a statistical model for a dilute gas in a box, which may be regarded as essentially the same model to be used in later sections and chapters of this thesis. To illustrate this we will discuss some of Boltzmann's important work.

In 1872, Boltzmann wrote down the equation that bears his name to describe the evolution of the particle distribution of a diluted gas. The particle distribution f is defined such that $f(\mathbf{r},\mathbf{p},t)d\mathbf{r}d\mathbf{p}$ is the number of molecules which, at time t, have positions lying within a volume element $d\mathbf{r}$ about $\mathbf{r}$ and momenta lying within a momentum-space element $d\mathbf{p}$ about $\mathbf{p}$[3]. In addition he associated with f a quantity, $H_{Bol} = \int f(\mathbf{r},\mathbf{p},t)\ln f(\mathbf{r},\mathbf{p},t)d\mathbf{r}d\mathbf{p}$, which, apart from the sign and perhaps additive and multiplying constants, is identified as Clausius's definition of entropy. To understand this quantity, and his celebrated H-theorem and equation, he studied the following problem in 1877.

The physical system is a gas in a isolated box of unit volume containing N_0 molecules[4]. Before treating this problem Boltzmann considered an urn containing an equal number of white, black, red etc spheres. He noted that the probability of drawing N_0 spheres with replacement, such that n_w are white, n_b black, etc, is $\frac{N_0!}{n_w!n_b!n_r!\cdots}$ times as large as the probability of drawing N_0 spheres of only one colour. Without loss of much generality lets assume that all of the N_0 molecules are confined to a finite region of momentum-space. Together the momentum-region and box-region form the phase-space region (momentum-position region) in which each of the N_0 molecules is confined. Boltzmann now made the important step of dividing the one-molecule phase-space region into equal-size volume elements (cells), say I of these, and made the assumption that the probability

[3]In the usual sense, the volume elements $d\mathbf{r}$ and $d\mathbf{p}$ are not to be taken literally as mathematically infinitesimal quantities. They are finite volume elements which are large enough to contain a very large number of molecules and yet small enough so that compared to macroscopic dimensions they are essentially points.

[4]See Part I, Chapter I, §6 in Ref. [32].

of finding the molecule in any of the equal-sized cells is the same. We can now relate the problem of the gas in a box to that of the urn: Instead of n_w, the number of drawings of white spheres, we denote n_1 the number of molecules in cell 1 having momentum $\mathbf{p}_1 = \mathbf{v}_1 m$ and position $\mathbf{r}_1$ and so on. Furthermore, let x_i represent the position in phase-space of the i^{th} molecule, where the range of x_i is $\mathcal{X} = \{1, 2, \ldots I\}$. Each micro-state $(x_1, x_2, \ldots, x_{N_0})$ is equally probable and the probability of some macro-state $(n_1, n_2, \ldots, n_I)$ is equal to the number of possible micro-states which have the numbers $n_1, n_2, \ldots, n_I$, which is simply $N_0!/(n_1!n_2! \cdots n_I!)$ multiplied by the probability of one micro-state, which is $(1/I)^{N_0}$, i.e. $(n_1, n_2, \ldots, n_I)$ is given by the multinomial distribution

$$p(n_1, n_2, \ldots, n_I | N_0) = \frac{N_0!}{n_1!n_2! \cdots n_I!} \frac{1}{I^{N_0}} \ . \tag{3.11}$$

We define Boltzmann's statistical model as

Definition 5 (Boltzmann's stat-model) *Boltzmann's statistical model consists of converting a phase-space region into I discrete cells and assigning equal probability to a molecule being located in any of these cells. Mathematically, this means that N_0 molecules, represented by N_0 r.v. $x_1, x_2, \ldots, x_{N_0}$, each having the range $\mathcal{X} = \{1, 2, \ldots, I\}$ and each being uniformly distributed as $p(x_i) = \frac{1}{I}$, for all $i = 1, 2, \ldots N_0$. As a result, the probability of a macro-state $(n_1, n_2, \ldots, n_I)$ is given by Eq. (3.11).*

If the only thing we know is that the molecules are confined to a box region and some region of momentum space then obviously the most probable macro-state will be the one where all the n_i's are equal and[5] $n_i = N_0/I$. However the gas container may be isolated such that the energy E of the N_0 molecules is conserved. The energy is given by $E = \sum_{i=1}^{I} n_i \epsilon_i$, where ϵ_i is the energy of a molecule in the i^{th} cell and the interaction energy of molecules with each other is neglected. Boltzmann now argued that if N_0 is large enough then the macro-state

[5]Assuming N_0 is a multiple of I

that satisfies this energy constraint and maximizes the expression in Eq. (3.11) will be sharply distributed, meaning that any other macro-state will be far less probable. In fact what can be shown ([27] and references therein) is that the probability of the gas not being distributed in any finite neighbourhood of its most probable macro-state goes towards zero as N_0 goes to infinity (see App. A).

We aim to find the most probable macro-state which maximizes the probability distribution in Eq. (3.11) subject to constraints on the phase-space occupation numbers $n_1, n_2, \ldots, n_I$. Since N_0 is astronomically large for any macroscopic system of reasonable size[6], Boltzmann applied the Stirling approximation $N_0! \simeq N_0^{N_0} e^{-N_0}$ and the probability in Eq. (3.11) becomes

$$\begin{aligned} p(n_1, n_2, \ldots, n_I | N_0) &\simeq \frac{N_0^{n_1+n_2+\cdots+n_I} e^{-N_0}}{\prod n_i^{n_i} e^{-n_i}} \frac{1}{I^{N_0}} \\ &= \frac{1}{I^{N_0}} e^{-N_0 \sum_{i=1}^{I} \frac{n_i}{N_0} \ln \frac{n_i}{N_0}} \ . \end{aligned} \tag{3.12}$$

By introducing v the volume of a cell, we can write the discrete single-particle density as $f_i = n_i/v$, $i = 1, 2, \ldots, I$. We see from Eq. (3.12) that maximizing the probability in Eq. (3.11) is replaced by minimizing Boltzmann's $H_{bol} = \sum_{i=1}^{I} f_i \ln f_i$ or alternatively maximizing a macro-state entropy of the form

$$H_e(\frac{n_1}{N_0}, \frac{n_2}{N_0}, \ldots, \frac{n_I}{N_0}) = -\sum_{i=1}^{I} \frac{n_i}{N_0} \ln \frac{n_i}{N_0} \ , \tag{3.13}$$

Hence, Boltzmann interpreted his H_{bol}, and therefore also entropy, as a measure of likelihood[24]; the gas is more likely to be distributed as $\mathbf{f} = (f_1, \ldots, f_I)$ rather than $\mathbf{f}' = (f_1', \ldots, f_I')$ if $H_{bol}(\mathbf{f}) < H_{bol}(\mathbf{f}')$ or equivalently if the entropy H_e of $\mathbf{f}$ is higher than the entropy of $\mathbf{f}'$.

As explained in the text following Def. 1, Lagrangian multipliers are introduced to form the functional

$$J[\mathbf{n}/N_0] = H_e[\mathbf{n}/N_0] + \lambda_0 N_0 \sum_{i=1}^{I} n_i/N_0 - \beta N_0 \sum_{i=1}^{I} \epsilon_i n_i/N_0 \ \ , \tag{3.14}$$

[6]Under standard conditions there are about 3×10^{19} molecules/cm^3 in a gas [33].

where $\mathbf{n} = (n_1, n_2, \ldots n_I)$. By comparing Eq. (3.14) and Eq. (3.6) we see from Eq. (3.8) that the Boltzmann most probable macro-state is

$$n_i^{bol} = N_0 \frac{\exp(-\beta\epsilon_i)}{\sum_{i=1}^{I} \exp(-\beta\epsilon_i)} \quad . \tag{3.15}$$

Take the example of an equilibrium distribution for a dilute gas without the presence of an external force field, where $\epsilon_i = mv_i^2/2$. Recognizing $\beta = -1/(kT)$ it is clear that inserting this expression for ϵ_i into Eq. (3.15) and using an appropriate normalization constant, the famous Maxwell-Boltzmann distribution is obtained, see e.g. Ref. [24], Chap. 4 for more information and examples.

Notice that Boltzmann's reasoning can be cast into the form of Bayes's theorem. This is simply achieved by taking Boltzmann's statistical model in Def. 5 as the prior probability distribution and the likelihood function to be the delta-function $p(E|\mathbf{n}, N_0) = \delta(\sum_{i=1}^{I} n_i\epsilon_i - E)$. We obtain the following posterior probability distribution for Boltzmann's dilute gas

$$\begin{aligned} p(\mathbf{n}|E, N_0) &= \frac{p(\mathbf{n}|N_0)p(E|\mathbf{n}, N_0)}{p(E|N_0)} \\ &\propto e^{N_0 H_e} \delta(\sum_{i=1}^{I} n_i\epsilon_i - E)\delta(\sum_{i=1}^{I} n_i - N_0) \quad , \end{aligned} \tag{3.16}$$

where we have written the prior $p(\mathbf{n}|N_0)$ as in the Stirling approximation in Eq. (3.12), the last delta-function on the right hand side of Eq. (3.16) is included to emphasize that N_0 is a known fixed value.

The mode of the posterior in Eq. (3.16) is exactly equal to Boltzmann's most probable macro-state in Eq. (3.15).

3.1.3 Shannon and communication theory

In 1948 Shannon introduced the concept of entropy in communication theory [34],[26]. He showed that entropy naturally enters into the expression for channel capacity and quantified entropy as a measure of uncertainty and information in communication theory. Shannon, for instance, showed that the number of bits

on average required to describe a discrete r.v. x with probability function $p(x)$ is $H_2[p] = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)$. Hence, the more uncertain you are about a r.v. the more bits are needed to describe it, in this sense $H_2[p]$ is a measure of uncertainty. To identify entropy as a measure of information, in contrast to a measure of uncertainty, envision an experiment in which the r.v. x is witnessed by an observer, who subsequently transmits the observed value to you. Before you are provided with this information, the uncertainty is $H_2(x)$. Upon receipt of the information, the uncertainty is zero. Interpreting "information" as the removal of uncertainty, then $H_2(x)$ is a natural measure of information. Shannon also showed that it is possible to derive the expression for entropy from a set of reasonable requirements. Looking for something which measures the information content of a r.v. $x \sim p$ it is reasonable to require the following properties of it (following Ref. [34])

1. H should be a continuous function of $p(x)$, for each $x \in \mathcal{X}$.

2. If p is uniform on $\mathcal{X}$, $p(x) = 1/I$, then H should be a monotonic increasing function of I.

3. If a reasoning can be broken down into successive choices, the original H should be the weighted sum of the individual values of H. (For more details illustrating this assumption see Ref. [34], or any textbook on information theory).

It can then by shown that the only H satisfying the three above assumptions is of the form $H \propto -\sum_{x\in\mathcal{X}} p(x)\ln p(x)$, i.e. of the entropy form. It is worth mentioning that Shannon did not in any way consider the above axiomatic derivation of entropy as a necessary part of his theories, but given chiefly to lend a certain plausibility to some of his later definitions[7]. Later Jaynes [25] re-interpreted

[7]In Ref. [34] Shannon writes "This theorem, and the assumptions required for its proof, are in no way necessary for the present theory. It is chiefly to lend a certain plausibility to some

Shannon's definition of entropy to be *the* measure of uncertainty instead of *a* measure of uncertainty. Jaynes insisted on the interpretation of entropy as *the* measure of uncertainty and the unquestionable starting point for further inference. For example, Jaynes writes in [25]: "Now, however, we can take entropy as our starting concept, and the fact that a probability distribution maximizes the entropy subject to certain constraints becomes the essential fact which justifies use of that distribution for inference"[8]. From Jaynes point of view his maximum entropy principle (or MEM) was regarded as a very high principle; he saw statistical mechanics as an application of this principle [25], for example.

3.2 MEM in spectral analysis

[9]The first application of MEM was in spectral analysis of time series. A sequence of observations made at equidistant time steps is referred to as a discrete time series. Thus $x_0, x_1, \ldots$ is a time series composed of measurements at times $t = 0, 1, \ldots$ (after a suitable rescaling of the time axis if necessary). To model the uncertainty of a time series each of the measurements x_n is described as a realized value of a r.v. x_n (the same notation for a r.v. and its realized values is used, see Chap. 2), and associated with the time series as a whole is the set of r.v. $x_0, x_1, \ldots$ called a random sequence (or process). These r.v. might be correlated with each other (e.g. the measurement at time $t = n$ may depend on the value measured at time $t = n - 1$ etc.) and $p(x_0, x_1, \ldots)$ is the joint probability density for the

of our later definitions. The real justification of these definitions, however, will reside in their implications.".

[8]Also, in Ref. [28], Jaynes writes "It is by now amply demonstrated by many workers that the "information measure" introduced by Shannon has special properties of consistency and uniqueness which make it *the* correct measure of 'amount of uncertainty' in a probability distribution.".

[9]This section can be omitted without loss of continuity, particularly if you are not concerned with MEM applied to spectral analysis.

process as a whole. In a variety of applications, it is desirable to characterize what is called a zero-mean weakly stationary time series (WSS)[10] through its associated spectral density representation. This representation is useful in those situations where the time series is thought to contain hidden periodicities.

The spectral density (or power density) is formally given for a digital (or discrete-time) process by

$$S(\omega) = \sum_{m=-\infty}^{\infty} R(m)e^{im\omega} \ , \tag{3.17}$$

where $R(m)$ is the autocorrelation function and for a WSS process $R(m) \equiv R(i+m,i) = \langle x_{i+m}x_i\rangle_p$. In real applications a finite time series is measured $x_0, x_1, \ldots, x_N$. The aim is to determine a power spectrum $S(\omega)$ for this time series. From Eq. (3.17) it is seen that this requires the knowledge of an infinite number of autocorrelation coefficients $R(m)$. However, given only a finite time series it is not obvious how to obtain an infinite number of $R(m)$'s; we can only hope to get good estimates of a finite number of $R(m)$'s, say $R(0), R(1), \ldots, R(M)$. Hence, the following question arises: given a finite number of $R(m)$ how are we going to calculate $S(\omega)$? Ideas on how to tackle this problem are dealt with in some of the spectral analysis literature. Lets assume that we can obtain good estimates of the first $M+1$ values of $R(m)$ and the underlying stochastical process is the infinite WSS random sequence $x_0, x_1, \ldots$. To follow the original approach by Burg[2] lets further assume that the joint probability distribution of the underlying process is multivariate normal with zero mean values. For the finite set of r.v's $\mathbf{x} = x_0, x_1, \ldots, x_N$, this implies

$$p(x_0, x_1, \ldots, x_N) = \frac{1}{(2\pi)^{(N+1)/2}|\mathbf{R}_{(N+1)}|^{1/2}} \exp(-\frac{1}{2}\mathbf{x}\mathbf{R}^{-1}_{(N+1)}\mathbf{x}^T) \ , \tag{3.18}$$

where the elements of the $N+1$ by $N+1$ matrix $\mathbf{R}_{(N+1)}$ are the autocorrelation coefficients $R(i,j)$, and because the process is WSS $R(i,j) = R(i-j)$ then $\mathbf{R}_{(N+1)}$

[10] A stochastic process $x_0, x_1, \ldots$ is called weakly stationary if its mean is constant $\langle x_i\rangle = \mu_i = \mu =$ constant, and its autocorrelation depends only on $m = n_2 - n_1$: $R(i+m,i) = R(m)$. For non-stationary processes the notion of a spectrum is of limited interest.

is Toeplitz[11] with entries $R(0), R(1), \ldots, R(N)$ along the top row. In the spirit of information theory and channel capacity the entropy rate of a random sequence $x_0, x_1, \ldots$ is defined as

$$\bar{h}[p] = \lim_{N\to\infty} \frac{1}{N} h[p(x_0, x_1, \ldots, x_{N-1})] \ . \tag{3.19}$$

It can then be shown under the assumption that the underlying probability density for the process $x_0, x_1, \ldots$ is a multivariate normal, that the entropy rate for this process is

$$\bar{h}[p] = \ln \sqrt{2\pi e} \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln S(\omega) d\omega \ . \tag{3.20}$$

This is obtained by substituting p in Eq. (3.18) into the expression for the entropy in Eq. (3.9) and taking the limit $N \to \infty$ as prescribed in Eq. (3.19).

Hence, we see that Burg's assumption that the underlying probability density is multivariate normal has transformed the problem of finding the probability density p which maximizes the entropy into that of optimizing the entropy rate in Eq. (3.20) with respect to $S(\omega)$. This solution is fairly easy to find. The first $M+1$ values of R(m) are assumed to be known. The entropy rate $\bar{h}$ then depends on the unspecified values of $R(m)$, $|m| > M$ and it is maximum if $\partial\bar{h}/\partial R(m) = 0$ for all $|m| > M$ i.e.

$$\int_{-\pi}^{\pi} \frac{1}{S(\omega)} e^{-im\omega} d\omega = 0$$

for all $|m| > M$. This shows that the coefficients of the Fourier series expansion of $1/S(\omega)$ are zero for $|m| > M$. Hence,

$$1/S(\omega) = \sum_{k=-M}^{M} c_k e^{-ik\omega} \tag{3.21}$$

with $c_k = c_{-k}^*$ to guarantee that $1/S(\omega)$ is a real quantity. The $M+1$ independent coefficients c_k can be found by inserting the expression for the entropy rate in Eq. (3.21) into $M+1$ equations of the kind in Eq. (3.3) one for each of the

[11] Toeplitz; means that it is symmetrical with respect to its leading diagonal, and the elements of a Toeplitz matrix satisfy: $a_{ij} = a_{i-1,j-1}$ for $i, j = 2, 3, \ldots, N$.

good estimated values of $R(m)$. Further, by using the property that $S(\omega) \geq 0$ it is possible to reduce the task of finding $S(\omega)$ to a purely linear problem [2]. To summarize Burg's paper, under the assumption that the underlying process is modeled to have the p.d. in Eq. (3.18), then this implies that the entropy rate is related to $S(\omega)$ as in Eq. (3.20), and Burg chooses to represent the data using the power spectrum which maximizes this entropy rate subject to keep the first $M+1$ autocorrelation coefficients $R(0), R(1), \ldots, R(M)$ being fixed. This is Burg's estimate for $S(\omega)$.

For a collection of papers following up Burg's work see [35]. Alternative derivations of Burg's method are presented in Ref. [36] and Ref. [37]. These references claim that in deriving Burg's expression for $S(\omega)$ it is not necessary to assume that the underlying stochastical process is multivariate Gaussian with zero means. Instead it is assumed that all the autocorrelation coefficients $R(0), R(1), \ldots$ are known, in the sense that the sequence of constraint equation $R(m) = \langle x_i x_{i+m} \rangle_p$ is imposed which the underlying statistical probability distribution p must satisfy. The only distribution, according to MEM, which satisfies such constraints, is clearly the multivariate Gaussian distribution used by Burg. If the coefficients $R(0), R(1), \ldots$ had been assumed to be related to any other expectation values but $R(m) = \langle x_i x_{i+m} \rangle_p$, then according to MEM p would not have been multivariate Gaussian distributed. Hence, the claim made by the authors of Ref. [36] and Ref. [37] that they have managed to arrive at the conclusion that the underlying stochastical process is multivariate Gaussian, without any extra assumptions, is not true in my opinion. Because assuming the relations $R(m) = \langle x_i x_{i+m} \rangle_p$ and using MEM is the same as assigning p to be multivariate Gaussian distributed in the first place.

3.3 MEM for image reconstruction

3.3.1 The approach of Gordon & Herman

The suggestion to use MEM for image reconstruction first appeared in the open literature in 1970-71, [38],[39]. In these papers the authors aim to find best reconstructions of a picture from measured projections of the picture. As an example, their work is applicable to the problem of finding the three dimensional density in space of an object being viewed with an electron microscope. The focus of both papers is on reconstruction of discrete two-dimensional (2D) pictures from projections. We can describe a square discrete 2D picture by specifying $\sqrt{I} \times \sqrt{I}$ numbers; say, n_{ij}, $i, j = 1, 2, \ldots, \sqrt{I}$. Gordon & Herman refer in Ref. [39] to n_{ij} as the grey-level of the picture at the matrix position (i, j), and in the same paper so-called 'upper-bound-cell' quantized pictures are considered where the n_{ij} are restricted to the integer gray-level values

$$0 \leq n_{ij} \leq l - 1, \quad i, j = 1, 2, \ldots, \sqrt{I} \ , \tag{3.22}$$

where 'upper-bound-cell' refers to the quantization scheme in Eq. (3.22). A projection of a 2D picture is, for example, the set of row sums $r_1, r_2, \ldots, r_{\sqrt{I}}$ where

$$r_i = \sum_{j=1}^{\sqrt{I}} n_{ij} \quad , i = 1, 2, \ldots, \sqrt{I} \ . \tag{3.23}$$

Another example of a projection could be the set of column sums $c_j = \sum_{i=1}^{\sqrt{I}} n_{ij}$, $j = 1, 2, \ldots, \sqrt{I}$ and so on. For each picture, a projection can be considered to be a set of $\sqrt{I}$ linear equations in I unknown numbers $\{n_{ij}\} \equiv \{n_{ij} \ , i, j = 1, 2, \ldots, \sqrt{I}\}$. If we take m distinct projections, we have $m\sqrt{I}$ simultaneous linear equations in I unknowns. If $m < \sqrt{I}$ there will generally be more than one solution; i.e., more than one picture with exactly the same projection. To find the best reconstruction from projections Gordon & Herman present in Ref. [39] three Monte Carlo algorithms for quantizised pictures and in Ref. [38] Gordon, Bender & Herman

present two direct methods called Algebraic Reconstruction Techniques (ART) for non-quantizised images. Here I will focus my attention on the three Monte Carlo algorithms[12]. These algorithms distribute bits (equal to 1 gray-level unit each) randomly over the I cells. The first two algorithms keep distributing bits randomly until the projection constraints are satisfied, using similar techniques. Gordon & Herman associate with a picture the entropy-like function

$$H_{GH} = -\sum_{i,j=1}^{\sqrt{I}} n_{ij} \ln n_{ij} \quad . \tag{3.24}$$

They observed that if a number of independent reconstructions are averaged then the entropy of this averaged image asymptotically increases towards a certain value, where this certain value is the maximum value H_{GH} can take on in Eq. (3.24), call this value H_{GH}^{max}, where the picture-values $\{n_{ij}\}$ satisfy a number of projection constraints of the type in Eq. (3.23). This observation led Gordon & Herman to suggest the first example (to our knowledge) of a MEM (or a Maximum Entropy-like Method) in image reconstruction, which is an algorithm where bits are distributed randomly (Monte Carlo) until the projection constraints are satisfied *and* H_{GH} has reached its maximum value, H_{GH}^{max}.

Gordon & Herman did have some technical problems with their algorithm, which are all very well described in Ref. [39]. In appendix I of Ref. [39], Gordon & Herman give a lengthy discussion towards finding a mathematical justification of their method. Their search for such a justification leads them into an impressive variety of fields, including "information theory, integer programming, numerical integration, adaptive search processes, chemical reaction systems, statistical mechanics, thermodynamics and many-dimensional geometry". As the conclusion of all this they write "Although these studies have increased and sometimes corrected our intuition, we still have no clear proof of why our method works". I find this appendix interesting reading, but would argue that they could have

[12] Also notice that, these algorithms were studied before the ART algorithms, even though they appeared in the open literature later (see text in [38]).

made their argument simpler. Instead of going into the objections to and justifications for the so-called 'maximum entropy principle', they could have omitted any discussion about such a principle and noted the following. If one or more projections is given it follows from Eq. (3.23) that the total number of bits is known, $N_0 = \sum_{i,j=1}^{\sqrt{I}} n_{ij}$. The probability for any picture $\{n_{ij}\}$ is proportional to the number of possible ways of generating that picture by distributing randomly N_0 bits. This is just a multinomial coefficient, and,

$$p(\{n_{ij}\}) = \frac{N_0!}{\prod_{i,j=1}^{\sqrt{I}} n_{ij}!} \frac{1}{A} \ , \tag{3.25}$$

where A is the total number of ways of distributing N_0 bits among I cells. The extra upper-bound-cell constraints in Eq. (3.22) limits A because each of the n_{ij} must be smaller than[13] $l-1$. What we can do now is simply use $p(\{n_{ij}\})$ in Eq. (3.25) to calculate which picture $\{n_{ij}\}$ satisfies the given projection constraints and is most likely to be generated by a random distribution of N bits. We can draw a parallel from this situation to Boltzmann's reasoning in Sec. 3.1.2 and also to App. A. We may include the I upper-bound-cell constraints as data of the type

$$n_{kk'} = \sum_{i,j=1}^{\sqrt{I}} n_{ij} g_{kk'}(n_{ij}) \leq l-1 \quad , k, k' = 1, 2, \ldots, \sqrt{I} \ ,$$

and the function $g_{kk'}(n_{ij})$ has the special form $g_{kk'}(n_{ij}) = 1$ if $k = i$ and $k' = j$ and zero otherwise. Hence, the macro-state (or image) which maximizes the probability in Eq. (3.25) can be described by a Boltzmann's most probably macro-state $\{n_{ij}\}^{bol}$ and this macro-state will also, by definition, have the highest macro-state entropy $H_e^{max} = H_{GH}^{max}$ of all macro-states that satisfy the I upper-bound-cell constraints and available projection constraints. It is therefore obvious that as more and more pictures, randomly generated by Gordon & Herman's algorithm 1 and 2, are averaged, then if the number of bits is large the entropy of this averaged

[13] Obviously if $l \geq N_0$ then $A = I^{N_0}$ and Eq. (3.25) is a multinomial distribution.

picture must approach H_{GH}^{max} from below. This gives a simple 'mathematical justification' for why Gordon & Herman's method works.

3.3.2 The approach of Frieden

An alternative way of introducing MEM into image reconstruction was presented by Frieden 1972 [40]. Frieden was interested in optics, particularly in estimating a most likely object given a number of sample image values of an incoherent object scene $O(x)$, which is a spatial radiance distribution. One of the main tasks of Frieden's 1972 paper is to associate a probability distribution with a spatial radiance distribution $O(x)$ and then to use the MEM (as e.g. in Def. 1) to assign the 'best' reconstructed object as a MED. His justification for comparing $O(x)$ with a probability distribution is firstly that $O(x)$ is non-negative for all values of x and secondly based on a comparison to the work in Ref. [29] and Ref. [41].

Frieden considered the following model to justify associating $O(x)$ with a probability distribution and the use of MEM in image reconstruction. First the object splits into discrete cells that are indexed $i = 1, 2, \ldots I$. An object is then fully specified by the I object values $O_1, O_2, \ldots, O_I$. Denote the total radiant power as P_0 and now the assumption that P_0 originates from N_0 smaller amounts of equal size ΔO: $P_0 = N_0 \Delta O$. Each of the elements can be positioned in any of the I cells, and the number of elements in the i^{th} cell is $O_i/\Delta O$. Using this model as his starting point, Frieden goes through arguments similar to the ones in [29] and proposes a parallel between $O(x)$ and a probability distribution and the use of a MEM to estimate the 'most uncertain' $O(x)$ given the data. Associating $\frac{O_1}{P_0}, \frac{O_2}{P_0}, \ldots, \frac{O_I}{P_0}$ with a p.d. Frieden formulated the MEM as that of maximizing (see also [42][14])

$$H_{F1} = -\sum_{i=1}^{I} O_i \ln O_i \tag{3.26}$$

[14]Be aware that Frieden refers to capital O as o in Ref. [42], and vice versa.

subject to constraints of the form

$$d_m = \sum_{i=1}^{I} O_i S_{mi} \ , \ m = 1, 2, \ldots, M \quad . \tag{3.27}$$

S_{mi} is a point-spread function for the system[15]. The relation between the measured data and object in Eq. (3.27) assumes the data are noise-free. Let us denote the noise of each data point by the noise term e_m. Frieden treated not only the object O as a probability distribution but also the noise. Frieden added a constant to all noise values to ensure that all $E_m = e_m + B$ are positive. And finally Frieden suggest the following procedure for finding the most likely estimate for O by optimizing an object-error entropy of the form

$$H_{F2} = -\sum_{i=1}^{I} O_i \ln O_i - \rho \sum_{m=1}^{M} E_m \ln E_m \ , \tag{3.28}$$

where ρ is an additional parameter. Subject to the constraints

$$d_m = \sum_{i=1}^{I} O_i S(y_m, x_i) + E_m - B \quad , m = 1, 2, \ldots, M \quad , \tag{3.29}$$

and $P_0 = \sum_{i=1}^{I} O_i$. The additional parameters B in Eq. (3.29) and ρ in Eq. (3.28) are regarded as experimental inputs that supplement the data d_m and the known total radiant power P_0 [40].

For an application of this MEM[16] in optics to the restoration of pictures taken of one of the moons of Jupiter, Ganymede, see Ref. [43].

In summary, the Frieden approach is to interpret the image (here a spatial radiance distribution) as a probability function, and to select the one possible probability distribution according to the MEM, where the data have the role of constraint equations on the probability distribution. To my knowledge, Frieden was the first to draw a parallel between a picture and a probability distribution in image analysis, and today many papers on maximum entropy in image analysis also perceive an image in this way (see Chap. 6 for some examples).

[15]To compare more easily Eq. (3.26-3.27) with the definition of the MEM in Def. 1 notice that $H_{F1} = -\sum_i Q_i \ln Q_i = -P_0 \sum_i \frac{Q_i}{P_0} \ln \frac{Q_i}{P_0} - P_0 \ln P_0$, and the last term is a constant.

[16]In the form of Eq. (3.26) and Eq. (3.27) I believe.

3.3.3 The approach of D'Addario

In the brief conference paper entitled "Maximum a posterior probability and maximum entropy reconstruction" [44], D'Addario discusses the work of Frieden and Gordon & Herman in the context of Bayes theorem and Bayesian statistics. Frieden's model, used to justify the use of MEM, is introduced in the context of a prior probability function and as a possible statistical model for an image. The total integrated intensity (equivalent to Frieden's P_0) in the picture is assumed known and is divided into $N_0 = S/\Delta S$ units of size ΔS. If the N_0 units are distributed at random among I cells of a discrete picture, then we know from Eq. (3.11), for example, that the probability of obtaining the picture with n_1 units in cell 1, n_2 in cell 2, etc. is given by the multinomial distribution and in the Stirling approximation (see Eq. (3.12)) has logarithm proportional to

$$H_1 = -\sum_{i=1}^{I} \rho_i \log \rho_i + \text{const} \quad , \tag{3.30}$$

where $\rho_i = n_i \Delta S$. This provides D'Addario with a prior probability distribution. At this point D'Addario notices a difficulty with the prior distribution in the above equation, which is its dependence on the quantization chosen. A part of Chap. 4 addresses this issue for a statistical model of a picture which contains regions of both positive and negative intensity.

In applying Bayes theorem D'Addario maximizes the posterior probability distribution, whose logarithm is given by

$$f(\boldsymbol{\rho}) = H_1(\boldsymbol{\rho}) - \lambda \sum_{m=1}^{P} \frac{|d_m^{obs} - d_m(\boldsymbol{\rho})|^2}{\sigma_m^2} \quad , \tag{3.31}$$

where the notation in Eq. (3.2) has been used to denote d_i^{obs} as the i^{th} observed data point and d_i is related to $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots \rho_I)$ as $d_m = \sum_{i=1}^{I} A_{mi}\rho_i$. σ_m is the standard deviation of the m^{th} independent measured Gaussian data point, and λ is known if the quantization of the statistical model is specified.

D'Addario's aim is to find the map $\boldsymbol{\rho}$ which maximizes f in Eq. (3.31). He

notices that this is actually a 'maximum a posterior' probability method based on the statistical model of the object studied - rather than an application of MEM.

Finally, optimizing Eq. (3.31) is equivalent to maximizing H_1 subject to the constraint

$$\sum_{m=1}^{P} \frac{|d_m^{obs} - d_m^{cal}(\boldsymbol{\rho})|^2}{\sigma_m^2} = \text{const} \quad , \tag{3.32}$$

where λ becomes a Lagrangian multiplier.

To my knowledge D'Addario never optimized f in Eq. (3.31) for a real example. However, the scenario where an $\sum x \ln x$ quantity is optimized subject to a constraint on a chi-squared like function (as in Eq. (3.32)) is today the colloquial definition of the maximum entropy method in image analysis. To distinguish it from the MEM as defined in Sec. 3.1.1 it will be abbreviated as MEMx. Let us define MEMx as

Definition 6 (MEMx) *MEMx is the optimizion of a function of the form* $\sum_{i=1}^{I} x_i \ln x_i$, *where* x_i *is real and positive* ($x_i \in \mathbf{R}^+$), *or the form* $\sum_{i=1}^{I} x_i \sinh^{-1} x_i$, *where* x_i *is either real and positive or* x_i *is defined over the whole real line* ($x_i \in \mathbf{R}$). *In both cases, subject to a constraint on the* x_i*'s of the type* $Q(x_1, x_2, \ldots, x_I) = \sum_{i=1}^{P} (d_i^{obs} - d_i^{cal}(x_1, x_2, \ldots, x_I))^2/\sigma_i^2 =$ *constant.*

Often x_i in Def. 6 represents the amount of image material or probability associated with the i^{th} discrete cell of the image, or x_i may be the square of the amount of image material in the i^{th} cell and so on. $Q(x_1, x_2, \ldots, x_I)$ in Def. 6 is called a chi-squared like statistic, because if the data represent P independent Gaussian observations, then Q is distributed as the chi-square probability distributed with P degrees of freedom, see Chap. 2, Sec. 2.3. In Def. 6 the '$\sum_{i=1}^{I} x_i \sinh^{-1} x_i$' is inserted for later convenience, such as in Sec. 4.2. In fact '$\sum_{i=1}^{I} x_i \sinh^{-1} x_i$' is more precisely written as $\sum_{i=1}^{I} [x_i \sinh^{-1} x_i/A - \sqrt{x_i^2 + A^2}]$, where A is a constant to be specified, and this functional form has been applied where x_i is a real number or real and positive (the magnitude of a complex quantity, see Chap. 4, Sec. 4.3). A function of the form '$\sum_{i=1}^{I} x_i \ln x_i$' in Def. 6 may be written more generally

as $\sum_{i=1}^{I}[x_i \ln x_i/A + \text{const}]$ or $\sum_{i=1}^{I}[x_i \ln x_i/A - x_i + \text{const}]$ (see references in the two review papers cited in the next section), where A is again a constant to be specified, and x_i must be real and positive.

In Ref. [45] D'Addario gives a short review of Fourier reconstruction techniques in radio astronomy. In Ref. [46] Wernecke and D'Addario actually apply maximum entropy image reconstruction in a form very similar to Frieden's, with the only difference being the use of a $\sum_i \ln x_i$ 'entropy' rather than a $\sum_i x_i \ln x_i$ 'entropy'. Also, interesting enough in the work [46] this approach is referred to as non-Bayesian.

3.4 Summary

The broad definition in Def. 6 of MEMx is an attempt to encapsulate what is meant by maximum entropy in image analysis in papers on this subject following the work of Gordon & Herman, Frieden and D'Addario. For a review of MEMx see Ref. [47], which has a special emphasis on NMR, and Ref. [48], which has a special emphasis on Astronomy. Neither of these review papers mention the work by Gordon & Herman and D'Addario, but provide a detailed account of how MEMx has been implemented in practice and many more MEMx publications dated after 1975. In Ref. [49] a quantitative analysis of the $\sum_i x_i \ln x_i$ 'entropy' is given. In the same journal of [49] there follows a discussion of this paper and MEMx in general, presented as 16 individual comments by 16 authors. Reading these comments is interesting (and can also be rather entertaining) because it shows how well-known scientists disagree totally upon how to perceive the MEMx method and how to use it, and it illustrates some of the difficulties of this subject. In this thesis the MEMx method is considered from a pragmatic viewpoint and although the MEMx method is critically analysed from a Bayesian statistical approach in Chap. 4, the results are applicable irrespective of how the MEMx method is perceived.

Of particular interest to this thesis is to apply MEMx to the study of images containing both positive and negative image material (papers on this subject are compared in Sec. 4.3, Chap. 4). This is done by considering the positive and negative material in the image separately optimizing the sum of two functions of the form $\sum_{i=1}^{I} x_i \ln x_i + \sum_{i=1}^{I} y_i \ln y_i$, where x_i represents positive image material and y_i negative image material, instead of just one function of the type $\sum_{i=1}^{I} x_i \ln x_i$. The definition of MEMx in Def. 6 can be viewed as encapsulating both cases, i.e. optimization of functions of the form $\sum_{i=1}^{I} x_i \ln x_i$ and $\sum_{j=1}^{J} \sum_{i=1}^{I} x_{ij} \ln x_{ij}$, where all the x_{ij} are positive numbers (for $\sum_{j=1}^{J} \sum_{i=1}^{I} x_{ij} \sinh^{-1} x_{ij}$, where all the x_{ij} are any real numbers or magnitudes of complex numbers) and the different j's represent different types of material (for the case of $J = 2$, then $j = 1$ could represent positive image material and $j = 2$ could represent negative image material).

Chapter 4

Data analysis technique used to analyze PND data

The purpose of this chapter is to introduce the technique used to transform magnetic structure factors into magnetization density maps. This technique may be categorized under the group of techniques labelled Maximum Entropy methods, sometimes abbreviated as MaxEnt or MEED, or MEMx as done in Chap. 3. The MEMx method will be presented as an application of Bayes' theorem with no references to any maximum entropy concept or quantity. For that reason, it is argued that for the presentation of the MEMx method given here, the name MEMx can be misleading. However, this is the label used in the literature for techniques that are similar or the same, and the name also used here.

The chapter will provide a rigorous and critical analysis of the MEMx method when used to construct images in which the values of the image can take both positive and negative values (to be referred to as a pos/neg image), with special emphasis on Polarized Neutron Diffraction (PND) data. In this process, it is found that the MEMx method depends crucially on the choice of its prior model

constants[1], and the output of the MEMx method is critically analysed in terms of these constants. These prior model constants have been discussed previously, see e.g. [50],[51],[52],[53],[54],[55] and [56], where a variety of different names for these prior model constants have been used. In Sec. 4.1 the prior model constant $\bar{M}^s$ is defined and the following names and notations have been used for $\bar{M}^s$: In Ref. [51] scalar parameter (β); In Ref. [57] default value (*def*); In Ref. [54] default parameter (b); In Ref. [52] model (m); In Ref. [55] *a priori* estimator (Q_k) and so on. None of the references cited above go into a detailed study of prior model constants, but a number of different strategies for selecting them are suggested and different degrees of significance are attributed to these prior constants[2]. Perhaps the most detailed of these studies is found in Ref. [57], Chapter 5, where the following conclusion is drawn: "to achieve visually stunning MEMx reconstructions: Choose C_0 [3]sufficiently large to suppress all the features you don't like and *def* [4]sufficiently small to smooth the points between the surviving features.". One purpose of this chapter may therefore be said to be to quantify the process proposed in the above citation, such that "stunning" MEMx reconstructions can be achieved by an automated procedure which would require a minimum of 'subjective' selections of prior model constants from the user. This is done by studying in painstaking detail how the prior model constants effect the MEMx output. In the terminology of Bayesian statistics such an analysis is referred to as a robustness analysis. This analysis provides a robust Bayesian image analysis method for the analysis of PND data, and other types of data where a parallel can be drawn to this work.

[1]Prior model constants is the term used for the constants making up the prior probability expression, see Chap. 2, Sec. 2.2.1.

[2]As will be shown later in this chapter, the effect of the prior model constants on the MEMx output depends on the nature and quality of the data, which may partly account for the differences in opinions.

[3]C_0 is equivalent to Q in this thesis, see Sec. 4.2.

[4]*def* is equivalent to $\bar{M}^s$, see Sec. 4.1.

Brief summary of chapter

The chapter is organized as follows. Sec. 4.1 provides a discussion of prior distributions which can be related to the MEMx technique. It contains most of the necessary mathematics, and Eq. (4.12) gives the prior probability distribution which will be used thereafter. Sec. 4.2 defines the Bayesian image analysis technique and MEMx method. A number of papers on MEMx are compared in Sec. 4.3. In Sec. 4.4 the output of the MEMx method is studied as a function of its prior model constants. Based on the results in Sec. 4.4, possible options for selection of good values for these prior model constants are discussed in Sec. 4.5. Sec. 4.6 shows how to calculate uncertainty estimates for any feature in the reconstruction, and Sec. 4.7 provides a summary and conclusions.

4.1 Prior distribution

The MEMx method as explained in detail in Chap. 3 is a somewhat confusing concept because different people perceive it in different ways. However, for the MEMx method(s), which is the optimization of a posterior probability, we can talk about a MEMx prior, the prior function in such a posterior distribution. Such a prior may be deduced in several ways. The approach followed here is to think of a MEMx prior as originating from a statistical model for an image. In the context of maximum entropy in image reconstruction, this approach was first proposed by D'Addario [44], see also Chap. 3. The work of D'Addario's and others is concerned with a strictly positive image (see Chap. 3). In deducing a MEMx prior for a pos/neg image, e.g. a magnetisation density, a procedure closely related to that of Buck & Macaulay [50],[51],[16] will be followed. The derivation in this section differs from that of previous work in its rigorous detail of the derivation and it also provides comments on priors, closely related to the MEMx prior, which could be used in image reconstruction.

The aim is to obtain a prior which is based on a simple idea involving the

generation of an image assuming as little as possible about how the image may look. In subsection 4.1.2 a possibility for adding more detailed information about the image into the prior will be discussed.

Start with the following statistical model for a positive image. Let us assume that the image is built up of N_0 quanta of equal magnitude ϵ. If the image is discretized into I equal size bins then the probability of having the configuration with n_1 quanta in bin 1, n_2 quanta in bin 2, ..., and n_I quanta in bin I, is simply given by the multinomial distribution, i.e.

$$p(\mathbf{n}|N_0) = \frac{N_0!}{n_1! \cdots n_I!} \frac{1}{I^{N_0}} \quad , \tag{4.1}$$

where the notation $\mathbf{n} = (n_1, n_2, \ldots, n_I)$ denotes a configuration for the N_0 quanta in the I bins.

A statistical model slightly different from Eq. (4.1) is to let the total number of quanta $N = \sum n_i$ be Poisson distributed with some mean value $\bar{N}$, i.e. $p(N|\bar{N}) = \frac{\bar{N}^N}{N!} e^{-\bar{N}}$, rather than fixing the total number of quanta to the value N_0 as in Eq. (4.1). Adding the information that $p(N|\bar{N})$ is Poisson distributed to the prior in Eq. (4.1) results in the analytically more convenient expression

$$p(\mathbf{n}|\bar{N}) = p(N|\bar{N}) p(\mathbf{n}|N) = \prod_{i=1}^{I} \frac{(\bar{N}/I)^{n_i}}{n_i!} e^{-\bar{N}/I} \tag{4.2}$$

i.e. the product of I independent and identical Poisson distributions each with mean value $\bar{N}/I$.

Our purpose is to extend the statistical Poisson model in Eq. (4.2) to enable us to handle a pos/neg image. Imagine now that we have two sets of quanta all of equal magnitude but one of the two sets we associate with positive material and the other set with negative material. Say we have the Poisson mean value $\bar{N}^+$ for the total number of positive type quanta and $\bar{N}^-$ for the total number of negative type quanta. Then this statistical model simply results in the prior distribution

$$p(\mathbf{n}^+, \mathbf{n}^-|\bar{N}^+, \bar{N}^-) = \prod_{i=1}^{I} \frac{(\bar{N}^+/I)^{n_i^+}}{n_i^+!} e^{-\bar{N}^+/I} \prod_{i=1}^{I} \frac{(\bar{N}^-/I)^{n_i^-}}{n_i^-!} e^{-\bar{N}^-/I} \ , \tag{4.3}$$

where the vectors $\mathbf{n}^+ = (n_1^+, n_2^+, \ldots, n_I^+)$ and $\mathbf{n}^- = (n_1^-, n_2^-, \ldots, n_I^-)$ denote a possible configuration of the two types of quanta. Typically, and in particular in analysis of PND data, we are only interested in knowing the difference between the number of positive and negative type quanta in each of the I bins. Therefore, make the change of variable $\mathbf{n} = \mathbf{n}^+ + \mathbf{n}^-$ to obtain the sum image map and $\mathbf{q} = \mathbf{n}^+ - \mathbf{n}^-$ to obtain the difference image map. From the probability distribution in Eq. (4.3) a prior expression can be obtained solely in terms of the difference image parameters $\mathbf{q} = (q_1, q_2, \ldots, q_I)$, using the marginalising procedure shown in App. B. The marginal distribution for $\mathbf{q}$ becomes

$$p(\mathbf{q}|\bar{N}^+, \bar{N}^-) = \prod_{i=1}^{I} \left(\frac{\bar{N}^+}{\bar{N}^-}\right)^{q_i/2} \mathrm{I}_{q_i}\left(\frac{2\sqrt{\bar{N}^+\bar{N}^-}}{I}\right) e^{-(\bar{N}^+ + \bar{N}^-)/I} \ , \tag{4.4}$$

where I_{q_i} is the modified Bessel function of order q_i. This distribution was studied in 1937, see for example Ref. [58] or Ref. [59] and references therein, and in the context of MEMx by Buck & Macaulay [50],[51]. The modified Bessel distribution in Eq. (4.4) does not occur often in literature, and so a brief discussion of some of its properties will be presented here. To this end it is convenient to first change the prior model constants $\bar{N}^+$ and $\bar{N}^-$ in Eq. (4.4) to

$$\begin{aligned} \bar{N}^s &= \bar{N}^+ + \bar{N}^- \\ \bar{N}^d &= \bar{N}^+ - \bar{N}^- \ , \end{aligned} \tag{4.5}$$

where, since both $\bar{N}^+$ and $\bar{N}^-$ are greater than zero, $|\bar{N}^d| < \bar{N}^s$. Eq. (4.4) then reads

$$p(\mathbf{q}|\bar{N}^s, \bar{N}^d) = \prod_{i=1}^{I} \left(\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d}\right)^{q_i/2} \mathrm{I}_{q_i}\left(\frac{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}{I}\right) e^{-\bar{N}^s/I} \ . \tag{4.6}$$

In App. B it is shown that the expectation value of q_i is

$$\langle q_i \rangle = \bar{N}^d/I \ \ , \ i = 1, 2, \ldots, I \tag{4.7}$$

with variance

$$\langle (q_i - \langle q_i \rangle)^2 \rangle = \bar{N}^s/I \ \ , i = 1, 2, \ldots, I \ . \tag{4.8}$$

This means that the expected value of the sum of the q_i's is $\bar{N}^d$, i.e. $\langle \sum_i q_i \rangle = \bar{N}^d$, and using the results in App. B it follows that $\sum_i q_i$ has variance $\bar{N}^s$. Hence, in the same manner as we are accustomed to summarize a Gaussian distribution by its mean and standard deviation, we can summarize the modified Bessel distribution by its mean value $\bar{N}^d$ and standard deviation $\sqrt{\bar{N}^s}$.

The prior in Eq. (4.6) is a discrete prior, its parameters $q_1, q_2, \ldots, q_I$ are integers. To transform $p(\mathbf{q}|\bar{N}^s, \bar{N}^d)$ into a recognizeable MEMx prior we aim to expand $p(\mathbf{q}|\bar{N}^s, \bar{N}^d)$ to a continuous probability distribution. The most obvious such transition simply involves replacing the discrete q_i's in Eq. (4.6) by continuous ones, hence the transformation

$$\begin{aligned} q_i &\rightarrow x_i \quad , i = 1, 2, \ldots, I \\ p(\mathbf{q}|\bar{N}^s, \bar{N}^d) &\rightarrow p(\mathbf{x}|\bar{N}^s, \bar{N}^d) = \frac{1}{A} \prod_{i=1}^{I} \left(\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d} \right)^{x_i/2} \mathrm{I}_{|x_i|}\left(\frac{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}{I}\right) , \end{aligned} \tag{4.9}$$

where the x_i's are real numbers and A is an appropriate normalization constant. This continuous analogue to the discrete prior in Eq. (4.6) could be used right-away as a prior. It has been implemented and tested and for the tests it was applied to, it did not make a significant improvement to the image analysis as compared to using the MEMx prior in Eq. (4.12) to be introduced shortly. Notice also that, rather than using the continuous Bessel prior in Eq. (4.9) the discrete Bessel prior in Eq. (4.6) is an interesting alternative. However, to limit the discussion in this thesis, the prior in Eq. (4.9) will not be discussed further, and to move forward in deducing a MEMx prior consider asymptotic expansions of the modified Bessel function as a function of the order parameter[5]. As a first asymptotic approximation to the modified Bessel function use[6] $\mathrm{I}_x(z) \sim e^{\sqrt{x_i^2+z^2} - x_i \sinh^{-1}(x_i/z)}$.

[5]The order parameter of a modified Bessel function $\mathrm{I}_x(z)$ is x, and z is its argument.

[6]See the books on special functions in Refs. [60],[61],[62].

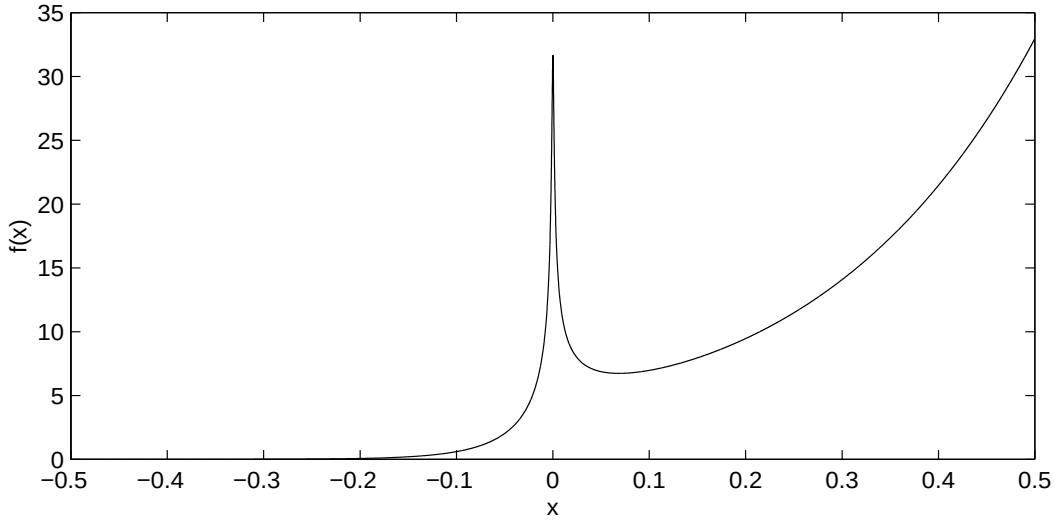

Figure 4.1: $f(x) = [x^2 + z^2]^{-1/4} e^{\sqrt{x^2+z^2} + x\left(\sinh^{-1}\left(\bar{N}^d/z\right) - \sinh^{-1}(x/z)\right)}$ with $\bar{N}^d = 100$ and $z = \sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}} = 0.001$. $f(x)$ is obtained by inserting $\mathrm{I}_x(z) = \frac{1}{\sqrt{2\pi}}[x^2 + z^2]^{-1/4} e^{\sqrt{x^2+z^2} - x\sinh^{-1}(x/z)}$ into Eq. (4.9) and not including the normalization constant A in that equation.

The prior in Eq. (4.9) becomes,

$$p(\mathbf{x}|\bar{N}^s, \bar{N}^d) = \frac{1}{B} \prod_{i=1}^{I} e^{\sqrt{x_i^2 + \frac{\bar{N}^{s\,2} - \bar{N}^{d\,2}}{I^2}} + x_i \left(\sinh^{-1}\left(\frac{\bar{N}^d}{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}\right) - \sinh^{-1}\left(\frac{x_i I}{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}\right) \right)}, \tag{4.10}$$

where B is a normalization constant and the relation $\left(\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d}\right)^{x/2} = \exp(x \sinh^{-1}\left(\frac{\bar{N}^d}{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}\right))$ has been used. Before continuing with Eq. (4.10) to deduce a MEMx prior, as a curiosity, consider briefly the alternative asymptotic approximation $\mathrm{I}_x(z) \sim \frac{1}{\sqrt{2\pi}}[x^2 + z^2]^{-1/4} e^{\sqrt{x^2+z^2} - x\sinh^{-1}(x/z)}$, the only difference to the first approximation being of the factor $\frac{1}{\sqrt{2\pi}}[x^2 + z^2]^{-1/4}$. Interestingly enough, it seems that this second approximation, though a more accurate representation of the modified Bessel function, when inserted into Eq. (4.9) leads to a prior that is physically not acceptable. For instance, the case with $\bar{N}^d = 100$, $\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}} = 0.001$ and $I = 1$ is drawn for this prior in Fig. 4.1, where the region close to the origin is emphasized. Clearly a continuous analogue of Eq. (4.6) must be a monoton-

ically increasing function for $x < \bar{N}^d$ and a monotonically decreasing function for $x > \bar{N}^d$. As seen in Fig. 4.1 this is not satisfied because of the sharp peak around $x = 0$. This feature can be attributed to the factor $[x^2 + z^2]^{-1/4}$ and to the fact that this prior is not strictly concave everywhere, in contrast to the priors in Eq. (4.9) and in Eq. (4.10).

In Fig. 4.2, the continuous priors in Eq. (4.9) and Eq. (4.10) are compared with their discrete analogue in Eq. (4.6), with $\bar{N}^d$ kept equal to zero. It might be surprising to notice that the $\sinh^{-1}$ prior in Eq. (4.10) seems in general to be a 'better' continuous analogue to the discrete prior than the modified Bessel prior in Eq. (4.9). Also shown in Fig. (4.2) is the Gaussian distribution with a mean value of zero and variance $\bar{N}^s$. As might have been expected, for $\bar{N}^s$ values of about 10 or greater the Gaussian distribution begins to become indistinguishable from the priors in Eq. (4.9-4.10), a property which will be used to explain later observations, in Sec. 4.4 for example.

The final task in obtaining an appropriate prior is to introduce the quantum magnitude ϵ into the $\sinh^{-1}$ prior in Eq. (4.10). Since we will be dealing almost exclusively with images that are naturally represented by a density, in particular a magnetization density, the transition from dimensionless variables (x_i's) to dimensional ones is written by means of the substitution $\mathbf{m} = \epsilon\mathbf{x}/v$, where the quantum ϵ is measured in the magnetic unit μ_B (Bohr magneton) and v is the volume of an image bin, so that the elements of the vector $\mathbf{m}$ have units of magnetization density; Bohr magneton per unit volume. The probability distribution for $\mathbf{m}$ is

$$p(\mathbf{m}|\epsilon, \bar{N}^s, \bar{N}^d) = \frac{1}{B(\bar{N}^s, \bar{N}^d)} \frac{v^I}{\epsilon^I} \times \prod_{i=1}^{I} e^{\frac{v}{\epsilon}\left(\sqrt{m_i^2 + \frac{\epsilon^2(\bar{N}^{s\,2} - \bar{N}^{d\,2})}{v^2 I^2}} + m_i(\sinh^{-1}(\frac{\bar{N}^d}{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}) - \sinh^{-1}(\frac{v m_i I/\epsilon}{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}))\right)}, \quad (4.11)$$

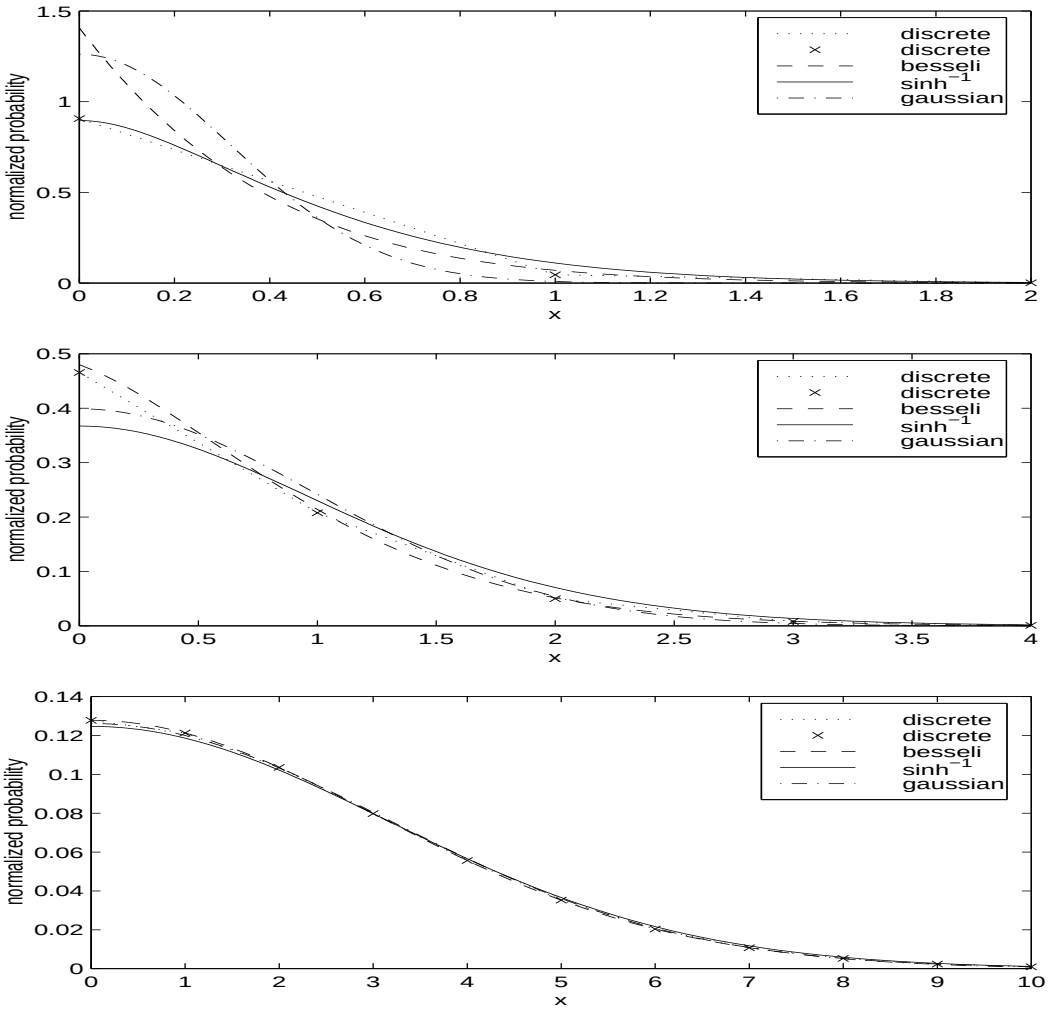

Figure 4.2: A number of the priors in the text are compared for three different values of $\bar{N}^s$ while $\bar{N}^d$ is kept equal to zero. In the top figure $\bar{N}^s = 0.1$, in the middle figure $\bar{N}^s = 1$ and in the bottom figure $\bar{N}^s = 10$. The discrete prior in Eq. (4.6) is plotted with a cross for each of its integer arguments and the crosses are connected with straight dotted lines. The continuous analogues in Eq. (4.9) and Eq. (4.10) are plotted as the dashed and solid graphs respectively. Also, the Gaussian distribution with mean zero and standard deviation $\bar{N}^s$ is plotted as the dash-dotted graph. It is seen that when $\bar{N}^s = 10$, any of the priors can be approximated well with a Gaussian distribution.

where $(v/\epsilon)^I$ is the determinant of the Jacobian for the coordinate transformation $\mathbf{m} = \epsilon\mathbf{x}/v$. In addition the dimensionless prior constants $\bar{N}^s$ and $\bar{N}^d$ can be replaced with the dimensional ones $\bar{M}^s = \epsilon\bar{N}^s$ and $\bar{M}^d = \epsilon\bar{N}^d$ so that Eq. (4.11) reads

$$\boxed{\begin{aligned} p(\mathbf{m}|\epsilon, \bar{M}^s, \bar{M}^d) &= \frac{1}{B(\bar{M}^s/\epsilon, \bar{M}^d/\epsilon)} \frac{v^I}{\epsilon^I} \\ &\times \prod_{i=1}^{I} e^{\frac{v}{\epsilon}\left(\sqrt{m_i^2 + \frac{\bar{M}^{s2} - \bar{M}^{d2}}{V^2}} + m_i(\sinh^{-1}(\frac{\bar{M}^d}{\sqrt{\bar{M}^{s2} - \bar{M}^{d2}}}) - \sinh^{-1}(\frac{Vm_i}{\sqrt{\bar{M}^{s2} - \bar{M}^{d2}}}))\right)} \end{aligned}}, \quad (4.12)$$

where $V = vI$ is the volume of the unit cell. The priors in Eq. (4.11) and Eq. (4.12) are of course identical, since fixing ϵ and $\bar{N}^s$ is equivalent to fixing ϵ and $\bar{M}^s$ (or $\bar{N}^s$ and $\bar{M}^s$), because these 3 quantities are constrained by the relation $\bar{M}^s = \epsilon\bar{N}^s$. The prior in Eq. (4.11) and Eq. (4.12) will mainly be discussed in terms of the prior model constants $\bar{M}^s$ and $\bar{M}^d$ but occasionally it will be convenient to discuss it in terms of the dimensionless constants $\bar{N}^s$ and $\bar{N}^d$. This prior, when written in the form of Eq. (4.12), is referred to as a MEMx prior in Refs. [50],[51] and Refs. [52],[53] (with $\bar{M}^d = 0$). To compare easier the prior expression in Eq. (4.12) with that of Buck & Macaulay [50],[51] we may write Eq. (4.12) as

$$\boxed{\begin{aligned} p(\mathbf{m}|\epsilon, \bar{M}^s, \bar{M}^d) &= \frac{1}{\tilde{B}(\bar{M}^s/\epsilon, \bar{M}^d/\epsilon)} \frac{v^I}{\epsilon^I} \\ &\times \prod_{i=1}^{I} e^{\frac{v}{\epsilon}\left(\sqrt{m_i^2 + \frac{\bar{M}^{s2} - \bar{M}^{d2}}{V^2}} - \frac{\bar{M}^s}{V} + m_i(\sinh^{-1}(\frac{\bar{M}^d}{\sqrt{\bar{M}^{s2} - \bar{M}^{d2}}}) - \sinh^{-1}(\frac{Vm_i}{\sqrt{\bar{M}^{s2} - \bar{M}^{d2}}}))\right)} \end{aligned}}, \quad (4.13)$$

where the term $e^{I\frac{v}{\epsilon}\frac{\bar{M}^s}{V}} = e^{\bar{N}^s}$ has been extracted from the normalization constant B, and $B = \tilde{B}e^{\bar{N}^s}$. Eq. (4.12) is the MEMx prior which will be of the main focus in this thesis. It will form part of a posterior distribution in Sec. 4.2 and the

resulting MEMx technique will be studied critically by a robustness analysis etc., and used in the analysis of real data in Chap. 5-6.

In a statistical sense the $\sinh^{-1}$ prior in Eq. (4.11-4.12) is a distribution that has been studied even less well than the modified Bessel distribution in Eq. (4.6). Moreover, it is not easy to describe its properties, because neither the normalization constant B, nor its mean values or variances can be written down in closed form. However, since the $\sinh^{-1}$ prior is a continuous analogue of the discrete Bessel prior and by comparing it with Eqs. (4.7,4.8), we may expect that, in particular when $\bar{N}^s/I \gg 1$, the approximate mean value for m_i is given by

$$\langle m_i \rangle \simeq \bar{M}^d/V \quad , \; i = 1, 2, \ldots, I \tag{4.14}$$

with variance

$$\langle (m_i - \langle m_i \rangle)^2 \rangle \simeq \epsilon \bar{M}^s/(v^2 I) \quad , i = 1, 2, \ldots, I \; . \tag{4.15}$$

Hence, the expected value of the total amount of material (i.e. $v \sum_{i=1}^{I} m_i$) is approximately $\bar{M}^d$ with variance $\epsilon \bar{M}^s$.

Observe also that the mode of the $\sinh^{-1}$ distribution in Eq. (4.12) is $m_i = \bar{M}^d/V$, $i = 1, 2, \ldots, I$. In the context of MEMx the mode of the prior is often referred to as the 'default model', see e.g. Ref. [51] and see Sec. 4.4.6. In the interpretation of MEMx where we are not searching for the probability of an image $p(\mathbf{m})$, but where the image itself is interpreted as a probability distribution, the mode of the prior is often referred to as the prior itself! See Chap. 3 for more details on this.

In this subsection the MEMx prior in Eq. (4.12) was deduced. Note that, in this process MEM as described in Sec. 3.1.1 was not applied and no reference was made to any maximum entropy principle. For this reason the prior in Eq. (4.12) would arguably be better referred to simply as a possible Bayesian image prior, rather than a MEMx prior. Also, referring to Eq. (4.12) as a possible Bayesian image prior puts it on equal footing with other possible priors, meaning that there

may be more appropriate priors to use than the prior in Eq. (4.12) depending on the system under consideration. This contradicts the notion sometimes found in MEMx literature that the MEMx prior is *the* prior which must always be used to describe a pos/neg image or pos image (see e.g. Refs. [63],[64] and references therein and Chap. 3).

4.1.1 An alternative continuous analogue

In this subsection an alternative MEMx prior[7] for a pos/neg image is deduced which is different from the MEMx prior given by Eq. (4.12).

Consider the discrete prior in Eq. (4.3) again. This time aim to find a continuous analogue to Eq. (4.3) directly without first marginalising out redundant parameters (to obtain Eq. (4.4)). Following the same procedure as was used to get from Eq. (4.6) to Eq. (4.10), firstly a continuous analogue to Eq. (4.3) is suggested,

$$\begin{aligned} n_i^+ &\rightarrow x_i^+ \quad , i = 1, 2, \ldots, I \\ n_i^- &\rightarrow x_i^- \quad , i = 1, 2, \ldots, I \\ p(\mathbf{n}^+, \mathbf{n}^- | \bar{N}^+, \bar{N}^-) &\rightarrow p(\mathbf{x}^+, \mathbf{x}^- | \bar{N}^+, \bar{N}^-) = \frac{1}{C} \prod_{i=1}^{I} \frac{e^{x_i^+ \ln(\bar{N}^+/I) + x_i^- \ln(\bar{N}^-/I)}}{\Gamma(x_i^+ + 1)\Gamma(x_i^- + 1)} \ , \end{aligned} \tag{4.16}$$

where $\Gamma(x)$ is the gamma function, and x_i^+, x_i^- are non-negative real numbers. Using the Stirling's asymptotic series for the gamma function, we can write $\Gamma(x+1) \sim x^x e^{-x}$ and rewrite Eq. (4.16) in the form

$$p(\mathbf{x}^+, \mathbf{x}^- | \bar{N}^+, \bar{N}^-) = \frac{1}{D} \prod_{i=1}^{I} e^{x_i^+ + x_i^- - x_i^+ \ln(\frac{x_i^+ I}{\bar{N}^+}) - x_i^- \ln(\frac{x_i^- I}{\bar{N}^-})} \ , \tag{4.17}$$

where D is an appropriate normalization constant. If, instead of substituting for the Gamma function using $\Gamma(x+1) \sim x^x e^{-x}$ we use the Stirling series approximation $\Gamma(x+1) \sim \sqrt{2\pi x} x^x e^{-x}$, the result is a continuous analogue not acceptable

[7] A MEMx prior may be defined as any prior which can be related to a prior in literature where it is used as part of a maximum entropy technique.

because the additional factor of $\sqrt{2\pi x}$ causes a $1/\sqrt{x}$ singularity at the origin of Eq. (4.16). As an example, when the Poisson mean values $\bar{N}^+$ and $\bar{N}^-$ are huge we expect the prior to have its maximum near $\bar{N}^+$ and $\bar{N}^-$, but because of the $1/\sqrt{x}$ singularity this will clearly not be the case.

Introducing the dimensional vector maps $\mathbf{m}^+ = \epsilon\mathbf{x}^+/v$ and $\mathbf{m}^- = \epsilon\mathbf{x}^-/v$ representing positive and negative magnetization density respectively, as well as $\bar{M}^+ = \epsilon\bar{N}^+$ and $\bar{M}^- = \epsilon\bar{N}^-$, then Eq. (4.17) can be written as

$$p(\mathbf{m}^+, \mathbf{m}^- | \bar{M}^+ \bar{M}^-) = \frac{1}{D}\frac{v^I}{\epsilon^I}\prod_{i=1}^{I} e^{\frac{v}{\epsilon}\left(m_i^+ + m_i^- - m_i^+ \ln(Vm_i^+/\bar{M}^+) - m_i^- \ln(Vm_i^-/\bar{M}^-)\right)} \,. \tag{4.18}$$

In Refs. [65],[66],[67] the logarithm of the prior in Eq. (4.18) is referred to as an 'entropy' for a pos/neg image, where this 'entropy' forms part of a maximum entropy technique[8], and Eq. (4.18) is said to be a MEMx prior. The MEMx priors in Eq. (4.18) and Eq. (4.12) are different, since it is not mathematically possible to connect these two priors by performing a marginalization[9]. But when these two priors are used as part of two MEMx techniques using data that depends only on the difference map $\mathbf{m} = \mathbf{m}^+ - \mathbf{m}^+$, it will be shown in Sec. 4.3 that the same results are obtained[10].

4.1.2 Adding more information to the statistical model

A possibility of adding more information to the statistical model in Eq. (4.4) is considered. Suppose that it is known, or at least a good guess can be made at the Poisson mean values of n_i^+ and n_i^- for *all* $i = 1, 2, \ldots, I$ bins. Denote these values as $\bar{\mathbf{n}}^+ = (\bar{n}_1^+, \bar{n}_2^+, \ldots, \bar{n}_I^+)$ and $\bar{\mathbf{n}}^- = (\bar{n}_1^-, \bar{n}_2^-, \ldots, \bar{n}_I^-)$ respectively. Then

[8] See also Sec. 4.2-4.3 and in Chap. 6 Papoular et al. [66] and Sakata et al. [67] works are discussed in more detail.

[9] The integrals necessary for performing the marginalisation of the components of $\mathbf{m}^+ + \mathbf{m}^-$ in Eq. (4.18) cannot be evaluated analytically.

[10] Provided also matching prior model constants are used, see Sec. 4.3, and for a practical demonstration see end of Sec. 6.3, Chap. 6.

for such a statistical model the resulting prior distribution reads[11]

$$p(\mathbf{n}^+\mathbf{n}^-|\bar{\mathbf{n}}^+\bar{\mathbf{n}}^-) = \prod_{i=1}^{I} \frac{(\bar{n}_i^+)^{n_i^+}}{n_i^+!} e^{-\bar{n}_i^+} \prod_{i=1}^{I} \frac{(\bar{n}_i^-)^{n_i^-}}{n_i^-!} e^{-\bar{n}_i^-} \quad . \tag{4.19}$$

To transfer the discrete prior in Eq. (4.19) to a continuous analogue repeat the steps from Eq. (4.6) to Eq. (4.10). Introducing the dimensional Poisson mean value maps $\bar{\mathbf{m}}^s = \epsilon(\bar{\mathbf{n}}^+ + \bar{\mathbf{n}}^-)/v$ and $\bar{\mathbf{m}}^d = \epsilon(\bar{\mathbf{n}}^+ - \bar{\mathbf{n}}^-)/v$ we obtain

$$p(\mathbf{m}|\epsilon, \bar{\mathbf{m}}^s, \bar{\mathbf{m}}^d) = \tag{4.20}$$
$$\frac{1}{\tilde{B}} \frac{v^I}{\epsilon^I} \prod_{i=1}^{I} e^{\frac{v}{\epsilon}\left(\sqrt{m_i^2+(\bar{m}_i^s)^2-(\bar{m}_i^d)^2} - \bar{m}_i^s + m_i(\sinh^{-1}(\frac{\bar{m}_i^d}{\sqrt{(\bar{m}_i^s)^2-(\bar{m}_i^d)^2}}) - \sinh^{-1}(\frac{m_i}{\sqrt{(\bar{m}_i^s)^2-(\bar{m}_i^d)^2}})) \right)} \quad .$$

which reduces to the distribution in Eq. (4.13) when $\bar{m}_i^s = \bar{M}^s/V$ and $\bar{m}_i^d = \bar{M}^d/V$, $i = 1, 2, \ldots, I$.

There may be physical situations where the use of the prior in Eq. (4.20) can be rigorously justified. However, it is definitely not applicable to the modelling of magnetization density images, simply because we do not have *a priori* such a large amount of information available before a PND experiment. In attempts where Eq. (4.20) is used it should be mentioned that, in general, very good data is needed if the results are not going to be dominated by the choice of the prior model maps $\bar{\mathbf{m}}^s$ and $\bar{\mathbf{m}}^d$. Also, if there is not a general agreement for the choice of the Prior model maps $\bar{\mathbf{m}}^s$ and $\bar{\mathbf{m}}^d$ then the final output can result in different conclusions about the physical system under consideration given the same measured data set.

4.2 Bayesian image analysis method and MEMx

A Bayesian image analysis method is typically achieved by firstly setting up a posterior probability distribution. For a fixed data set the posterior is proportional to the product of the prior and likelihood function, see Chap. 2. The prior

[11] Eq. (4.19) reduces to Eq. (4.3) when $\bar{n}_i^+ = \bar{N}^+/I$ and $\bar{n}_i^- = \bar{N}^-/I$ for all $i = 1, 2, \ldots I$.

distribution is taken to be the $\sinh^{-1}$ prior in Eq. (4.12) suitable for the image analysis problem of polarised neutron diffraction and similar physical systems involving the reconstruction of pos/neg images. The likelihood function is taken to be the product of P univariate Gaussian distributions, one for each data point, and the likelihood function reads

$$p(\mathbf{F}^{obs}|\mathbf{m}) = \frac{1}{(2\pi)^{P/2}\prod_{i=1}^{P}\sigma_i}\exp(-\frac{1}{2}Q(\mathbf{m})) \quad , \tag{4.21}$$

where $Q(\mathbf{m})$ is

$$Q(\mathbf{m}) = \sum_{i=1}^{P}\frac{(F_i^{obs}-F_i)^2}{\sigma_i^2} \quad . \tag{4.22}$$

F_i^{obs} is the i^{th} observed structure factor and $\mathbf{F}^{obs}$ is shorthand for $\mathbf{F}^{obs} = (F_1^{obs}, F_2^{obs}, \ldots, F_P^{obs})$. F_i is referred to as the i^{th} 'calculated' structure factor (as opposed to the i^{th} 'observed' structure factor) and is related to $\mathbf{m}$ through the linear transformation $\mathbf{F} = \mathbf{Am}$, where the $P \times I$ image to data Fourier[12] transformation matrix $\mathbf{A}$ has the elements

$$A_{ij} = v\exp(2\pi i\mathbf{K}_i \cdot \mathbf{r}_j) \quad . \tag{4.23}$$

The posterior which will studied in this chapter and used in Chap. 5-6 has, for a given fixed data set, the following relation

$$\boxed{p(\mathbf{m}|\mathbf{F}^{obs}, \epsilon, \bar{M}^s, \bar{M}^d) \propto p(\mathbf{m}|\epsilon, \bar{M}^s, \bar{M}^d)p(\mathbf{F}^{obs}|\mathbf{m})} \quad , \tag{4.24}$$

where the first term is the prior in Eq. (4.12) and the second term the likelihood in Eq. (4.21). The posterior in Eq. (4.24) contains all the information needed,

[12]The Fourier transform can be written as

$$F_i = \int_{cell} m(\mathbf{r})\exp(2\pi i\mathbf{K}_i \cdot \mathbf{r})d\mathbf{r}$$

where $\mathbf{K}_i$ is a reciprocal lattice vector and the integral is calculated as

$$F_i = v\sum_{j=1}^{I} m_i\exp(2\pi i\mathbf{K}_i \cdot \mathbf{r}_j) \quad ,$$

i.e. in the usual discrete Fourier series approximation of the Fourier transform.

but since the image $\mathbf{m}$ may easily consist of 10000 or more image bins it is not possible to comprehend all this information in one go. The reason for using a 'point estimate' of multi-dimensional probability distributions, as discussed in Chap. 2, is to choose from Eq. (4.24) the information which best characterizes this probability distribution in a form which we can comprehend (i.e. is practically useful). If the posterior in Eq. (4.24) is chosen to be summarized by the image ($\mathbf{m}$) which maximizes it (hence its mode) then the result is a Bayesian image analysis technique can be compared to MEMx techniques in literature, and define the following Bayesian image analysis method

Definition 7 (Bayesian image analysis method) *Consist of summarising the posterior in Eq. (4.24) by its mode.*

In Sec. 4.6 it will be shown how information from the posterior, other then its mode[13], can be taken advantage of in the data analysis process. However, before Sec. 4.6 the focus will be on the analysis of the mode of the posterior in Eq. (4.24).

Def. 7 is in agreement with the broad definition of MEMx in Def. 6, Chap. 3, which attempts to cover how MEMx is used and defined in the literature. To see this more clearly we write the $\sinh^{-1}$ prior in Eq. (4.12) in the form $p(\mathbf{m}|\epsilon, \bar{M}^s, \bar{M}^d) \propto \exp(\frac{v}{\epsilon}S(\mathbf{m}))$ where

$$\begin{aligned} S(\mathbf{m}) &= \sum_{i=1}^{I} \big\{ [m_i^2 + (\bar{M}^{s\,2} - \bar{M}^{d\,2})/V^2]^{1/2} + m_i \big(\sinh^{-1}(\bar{M}^d/[\bar{M}^{s\,2} - \bar{M}^{d\,2}]^{1/2}) \\ &\quad - \sinh^{-1}(Vm_i/[\bar{M}^{s\,2} - \bar{M}^{d\,2}]^{1/2}) \big) \big\} \quad , \end{aligned} \tag{4.25}$$

Maximizing the posterior in Eq. (4.24) is the same as maximizing the logarithm of the posterior and the logarithm of the posterior is, using the notation of Eq. (4.25) (and Eq. (4.22)), proportional to

$$\log \mathrm{P}(\mathbf{m}) = \frac{v}{\epsilon} S(\mathbf{m}) - \frac{1}{2} Q(\mathbf{m}) \quad . \tag{4.26}$$

[13]Information about the sharpness of the posterior distribution around its mode will be used to calculate error estimation, see Sec. 4.6.

Considering the likelihood function in Eq. (4.21) by itself, then the statistic Q in Eq. (4.22) is χ^2 distributed with P degrees of freedom, provided that $\mathbf{m}$ is the true image (see Chap. 2, Sec. 2.3). For this reason, a constraint on Q, i.e. Q = some constant, is often referred to as a chi-squared constraint, (see Sec. 2.3.3). Furthermore, $\frac{v}{\epsilon}$ in Eq. (4.26) may be intepreted as a Lagrangian multiplier and optimizing Eq. (4.26) can be perceived as maximizing $S(\mathbf{m})$, an 'entropy', subject to a chi-squared constraint, Q = come constant. This is how the MEMx technique is most frequently thought of and how it is defined in Def. 6, Chap. 3.

In addition, the maximizing problem of Def. 7 can be perceived as a regularizer method, where $S(\mathbf{m})$ is the regularizer and the degree of regularizing or smoothing is determined by the constraint equation involving $Q(\mathbf{m})$, see Refs. [63],[64] for a detailed explanation of this. If the posterior in Eq. (4.24) had been chosen to be summarized not by its mode, but by its mean for example, then the comparison to MEMx and the broad definition of MEMx in Def. 6, Chap. 3 could not have been made.

The number of prior model constants of the MEMx method of Def. 7 is three: ϵ, $\bar{M}^s$ and $\bar{M}^d$. A value for each of these three constants needs to be specified before the MEMx method can be used. Rather then specifying a ϵ value as explained above and seen from the expression in Eq. (4.26), ϵ can alternatively be interpreted as a Lagrangian multiplier which is determined from a specified constraint value on Q. This is preferable because one generally has a stronger intuition about a quantity such as Q compared to quantity such as ϵ, also using Q (to fix a value for ϵ) is the familiar approach in MEMx literature. Sec. 4.4-4.5 provides a rigorous study of the MEMx method as a function of its prior model constants Q, $\bar{M}^s$ and $\bar{M}^d$, which is called an robustness analysis in Bayesian statistics (see Sec. 2.2.1, Chap. 2).

4.3 Comparing MEMx publications

Although the MEMx method is supposedly just one technique in image analysis, it is interesting to note that the relatively large number of papers with "MEMx" in the title can be separated into many smaller sets which do not cite one another. Only MEMx papers used to reconstruct pos/neg images will be briefly reviewed here, see Chap. 3 for a general discussion of early maximum entropy related works.

Buck & Macaulay, in Refs. [50],[51], used the prior in Eq. (4.12) (or Eq. (4.20)) to analyse solar magnetic cycles from sunspot data and to construct neutron charge density maps from elastic electron scattering cross-sections. It must be noted that Buck & Macaulay use a notation, explained in App. C, adapted from papers of Skilling (see App. C), which can be confusing. This notation introduces a so-called metric as part of the prior expression which may lead to the illusion that this metric is a real physical quantity and that the optimization problem that is treated by Buck & Macaulay is different from the one in Def. 7; This is not the case (see App. C for a discussion of this). From a Bayesian statistical point of view Buck & Macaulay [50],[51] were the first to marginalise out redundant parameters to obtain the modified Bessel function expression in Eq. (4.4) in the context of a MEMx analysis.

Consider the following selection of papers: Laue et al. [65],[68]; David [55],[69]; Papoular et al. [66],[70]; and Sakata et al. [67]. Laue et al. studied NMR data and were the first to use a MEMx method for reconstructing pos/neg images and their work will be discussed further at the end of this section. David was interested in solving crystal structures from powder diffraction data and this work involves obtaining Patterson maps, which for neutron powder diffraction data is a pos/neg image. The work of Papoular et al. and Sakata et al. will be discussed in more detail in Chap. 6. Common to these approaches is the use of what may be called a 'double ln entropy'. It is therefore of interest to study briefly whether

the output from the MEMx method differs as a result of using a double ln entropy rather than a $\sinh^{-1}$ entropy. Take the double ln entropy to be the exponential of the double ln prior in Eq. (4.18)

$$S(\mathbf{m}^+, \mathbf{m}^-) = \sum_{i=1}^{I} [m_i^+ + m_i^- - m_i^+ \ln(\frac{V m_i^+}{\bar{M}^+}) - m_i^- \ln(\frac{V m_i^-}{\bar{M}^-})] \ . \tag{4.27}$$

If the data only depends on the difference map $\mathbf{m} = \mathbf{m}^+ - \mathbf{m}^-$ (as is the case for data measured to obtain the magnetization density of a unit cell) then the resulting MEMx optimization problem may be defined as the maximization of

$$\log\mathrm{P}(\mathbf{m}^+, \mathbf{m}^-) = \frac{v}{\epsilon} S(\mathbf{m}^+, \mathbf{m}^-) - \frac{1}{2} Q(\mathbf{m}^+ - \mathbf{m}^-) \tag{4.28}$$

subject to a constraint on Q, where Q is a chi-squared like function such as Q in Eq. (4.22). Eq. (4.28) is the equivalent of Eq. (4.26) in the previous section except for the use of the double ln prior instead of the $\sinh^{-1}$ prior. It will now be shown that optimizations of Eq. (4.28) and Eq. (4.26) yield exactly the same output provided equivalent prior model constants are used[14].

For notational convenience (and without loss of generality) assume that we have just one bin ($I = 1$) with volume $v = 1$. Changing the variables to sum density $m^s = m^+ + m^-$ and difference density $m = m^+ - m^-$ and $\log\mathrm{P}(m^+, m^-)$ in Eq. (4.28) reads

$$\begin{aligned} \log\mathrm{P}(m^+, m^-) \ &= \ m^s - (1/2) m^s \ln[((m^s)^2 - m^2)/(4\bar{M}^+\bar{M}^-)] - Q(m)/2 \\ &\quad -(1/2) m \ln[(m^s + m)\bar{M}^-/((m^s - m)\bar{M}^+)] \ . \end{aligned} \tag{4.29}$$

To optimize $\log\mathrm{P}(m^+, m^-)$ take the derivative of this expression first with respect to m^s. This first derivative is equal to zero when $(m^s)^2 = 4\bar{M}^+\bar{M}^- + m^2$. Inserting this value for m^s into the expression in Eq. (4.29) and using the relations between the prior model constants $\bar{M}^s = \bar{M}^+ + \bar{M}^-$ and $\bar{M}^d = \bar{M}^+ - \bar{M}^-$ (see Eqs. (4.5))

[14] A similar observation has been noted previously, for instance see Refs. [56],[52] and [14], page 152.

and $\sinh^{-1} x = \ln(x+\sqrt{x^2+1})$ it can be deduced that $\log P(m^+, m^-)$ in Eq. (4.29) reduces to exactly $\log P(m)$ in Eq. (4.26) with I = 1 (and $v = 1$). Hence, it has been shown that the MEMx optimization of Eq. (4.28) and Eq. (4.26) result in exactly the same solution. This shows that it is possible to relate publications on MEMx which use the double ln prior and publications which use the $\sinh^{-1}$ prior directly, as will also be illustrated in Chap. 6, where Papoular et al. [66],[70] and Sakata et al. [67] are discussed in more detail.

In the work of Laue et al. [65],[68] the MEMx method was used to obtain NMR frequency spectra containing both positive and negative peaks or even complex NMR spectra[15]. Following the work of Laue et al. and others, and based on the particular physics of a NMR experiment, it was argued by Daniell & Hore in Ref. [54] that for the purpose of analysing NMR data, a $\sinh^{-1}$ entropy is advantageous to use, where this $\sinh^{-1}$ entropy is a function of magnitudes of complex numbers[16]. For a quantitative comparison of the work of Daniell & Hore and Laue et al. see Ref. [56], where it is concluded that for the analysis of NMR data the approach of Daniell & Hore seems to give the best results.

A recent paper on MEMx is Maisinger et al. [52],[53], who study cosmic microwave background radiation data. In Maisinger et al. [52],[53] the $\sinh^{-1}$ prior expression is used (use in Ref. [52] in Eq. (8) of this work that $\ln(x + \sqrt{x^2+1}) = \sinh^{-1} x$).

Practical applications of MEMx applied to the study of PND data is found in Refs. [74], [75], [76], [77], [78], [79], [80], [81] and [82]. Refs. [80], [81] will be discussed in more detail in Chap. 6, and Ref. [82] in Chap. 5.

[15] A large class of techniques known as two-dimensional NMR return frequency spectra, which, in general, may need to be described as 2D-complex spectra. An example of such a NMR technique is COSY, see e.g. Ref. [71].

[16] In Refs. [54],[72],[73] the entropy used is $S_{1/2} = -\sum[|M_i| \sinh^{-1} |M_i|/b - \sqrt{b^2 + |M_i|^2}]$, where $|M_i|$ is the modulus of a complex magnetization. $S_{1/2}$ has exactly the same form as $S(\mathbf{m})$ (with $\bar{M}^d = 0$) in Eq. (4.25). For more details see the cited papers.

4.4 Robustness analysis of mode of posterior

This section contains a rigorous analysis of the Bayesian image analysis method as defined in Def. 7 as a function of the methods prior model constants for a fixed data set. In Bayesian statistics this is called a robustness analysis. Such an analysis has never been carried out previously in the context of the posterior in Eq. (4.24) and the MEMx technique.

The analysis aims to reconstruct a known model from an incomplete data set generated by that model. The model which will be used for this purpose is described in Sec. 4.4.2. Since the data is generated from a known source, different choices of the prior model constant, and the resemblance of the resulting mode of the posterior to the actual model can be compared. Such analysis may be referred to as a simulation. Before performing any simulations a number of possible numerical algorithms for carrying out MEMx optimization are discussed first.

4.4.1 Numerical consideration

The numerical problem of finding the mode of a posterior in Eq. (4.24) for given prior model constants is in general a fairly straightforward one. The main reason for this is that the posterior is strictly concave and therefore has only one maximum, the mode of the posterior, and no additional local extrema. Also, the gradient and Hessian matrix can be calculated with relative ease. On the other hand, since the number of variables of the posterior function, I (equal to the number of bins), may be of the order of a million or more, the discussion will be limited to algorithms which require storage of the order I rather than I^2. Further, finding the mode of the posterior is the same as finding the maximum of the logarithm of the posterior, and the later is computationally favourable.

An obvious first candidate to try is the conjugate gradient method. The reason for this is that the logarithm of the posterior is expected not to deviate too

much from a quadratic form, since the likelihood has a Gaussian distribution (see Eq. (4.21)), for a broad range of prior model constants. The closer the posterior is to being a multivariate Gaussian the more efficient the conjugate gradient method will be. Described shortly, this search algorithm finds the maximum of a function along a set of conjugate search directions, one search direction at a time. The so-called Polak-Risiere variant of the conjugate gradient method has been implemented following Press et. al. [83]. The method has worked satisfactorily for the cases to which it was applied. In general this algorithm was used together with the algorithm described below in order to make absolutely sure that the correct answers are obtained.

The first algorithm that was used was a copy of a program used previously by Buck & Macaulay to study sunspots and nuclear charge densities in Refs. [50],[51], it is described in some detail in appendix A of Macaulay's thesis [16]. The program follows the method of Skilling & Bryan described in Ref. [84], a specialized algorithm to deal with the optimization problem related to that of Eq. (4.26) (or Eq. (4.28)). The method resembles a steepest descent method but is somewhat more involved. First, a number of search directions are selected. Typically three search directions are used, where one direction is the gradient of the posterior. Then, a second order Taylor expansion of the logarithm of the posterior is made along these search directions, and the image is found that optimizes the Taylor approximated posterior within the three-dimensional search space. Finally, reset the old image with the newly found image. This iterative process is repeated until the maximum value for the posterior is reached. In addition Skilling & Bryan use a metric in their algorithm to adjust the search directions and the step-size in each iterative step. They find that choosing the metric to be equal to the inverse of the Laplacian of the logarithm of the prior is a particularly useful choice.

Recently the Newton-Raphson method was used in Refs. [52],[53] to calculate the MEMx Problem. The Newton-Raphson iterative process updates the image

using

$$m_i^{n+1} = m_i^n - \gamma \left(\frac{\partial f / \partial m_i|_{m_i^n}}{\partial^2 f / \partial m_i^2|_{m_i^n}} \right) , \tag{4.30}$$

where f is the logarithm of the posterior in Eq. (4.24) and an appropriate value for γ is chosen for each iteration to maximize speed and stability.

Freely available MEMx programs exist on the web. For instance, a MEMx program which is an extended version of the program in Ref. [85] can be downloaded from the web-side "http://www.uni-tuebingen.de/uni/pki". For a full description of the web-program see Ref. [86]. This program differs from the ones described above by not keeping the MEMx prior model constants fixed in the iteration search algorithm it is using (except for the equivalent quantity to Q in Eq. (4.22)). This is a consequence of an approximation that is applied which is referred to as the zeroth-order single-pixel approximation, for more details see Ref. [87] and Sec. 6.2 in this thesis. Thus, this type of algorithm is not appropriate for the analysis carried out in this chapter. Another example, of a freely available MEMx software is described in Ref. [88], which is available from hoch@rowland.org.

The code which was obtained from Buck & Macaulay together with an alternative code that was written by myself are very solid and complement each other very well. The program includes a simple interface and will be made freely available to anyone interested.

My general experience is that the overall speed depends heavily on the choice of the prior model constants used relative to the data set, and that each algorithm has preferred regions of prior models constants. However this subject has not been studied exhaustively.

4.4.2 Thesis model used for simulations

The model chosen here for studying the robustness of the posterior distribution in Eq. (4.24) has a degree of randomness to it, and could reflect a real physical

system. It is a 2D model rather than a 3D model simply because 2D images are easier to visualize, however it does originate from a 3D charge density model of one of the crystal structures that was studied experimentally.

The unit cell is chosen to be rectangular with point group mm[17] and with lattice parameters $a = 3.9\mathring{A}$ and $b = 11.8\mathring{A}$. The image density within the 2D-unit cell consists of a collection of densities centered around a set of crystallographic sites specified as fractional coordinates in table 4.1. These local densities are generated by the same function for all the sites, and are given by the values in a plane though the center of the Pr $4f$ charge density taken from Ref. [89]. Taking an atomic unit to be AU$= 0.528\mathring{A}$, this density in Cartesian coordinates is

$$\begin{aligned}\rho(x,y,z) &= \frac{1}{4\pi(\mathrm{AU})^3}\,(r/\mathrm{AU})^6\,[902.183\exp(-10.271\,r/\mathrm{AU}) + \\ &\quad 129.67688\exp(-5.828\,r/\mathrm{AU}) + 20.006113\exp(-3.885\,r/\mathrm{AU}) + \\ &\quad 0.52792993\exp(-2.125\,r/\mathrm{AU})]\ , \end{aligned} \tag{4.31}$$

where $r = \sqrt{x^2+y^2+z^2}$ is in units of $\mathring{A}$, and the density ρ satisfies $\int_{\mathbf{R}^3}\rho\,dxdydz = 1$. The magnetization at any point in the 2D-unit cell is then given by the sum over the atomic sites in table 4.1

$$m^{model}(x,y) = \sum_{i=1}^{9} A_i \sum_j \rho(x - x_{ij}, y - y_{ij}, z = 0)\ , \tag{4.32}$$

where (x_{ij}, y_{ij}) is the position of the i^{th} atom at the j^{th} symmetry site (e.g. atom $i = 2$ has the symmetry equivalent positions $(0, 0.45)$, $(0, 0.55)$, $(1, 0.45)$ and $(1, 0.55)$). A_i is a relative intensity coefficient in units of $\mu_B\mathring{A}$ and the density $m(x,y)$ is then in units of $\mu_B\mathring{A}^{-2}$. This choice of units is made because the main focus here is on the analysis of polarized neutron diffraction data.

The model in Eq. (4.32) is transformed into a discrete 64×128 grid which will be referred to as the thesis model; it is then defined by specifying a magnetization density for each of the 64×128 bins: $m^{model}(0,0)$, $m^{model}(\frac{1}{64}a, 0)$, ..., $m^{model}(\frac{63}{64}a, 0)$, $m^{model}(0, \frac{1}{128}b)$, ..., $m^{model}(\frac{63}{64}a, \frac{127}{128}b)$, as shown in Fig. 4.3.

[17]The mm point group has two mirror lines through the center of the cell.

Atom name	Position		Intensity coefficient	
i	x (units of a)	y (units of b)	A [$\mu_B \mathring{A}$]	No. of atoms
1	0.5	0.5	1	1
2	0	0.45	-0.08	2
3	0.5	0.37	-0.05	2
4	0	0.36	0.02	2
5	0.5	0.18	0.18	2
6	0	0.15	0.20	2
7	0.5	0.09	-0.04	2
8	0.5	0	0.04	1
9	0	0	0.06	1

Table 4.1: The fractional coordinates together with the intensity coefficients of the thesis model. The thesis model is shown in Fig. 4.3. The last column shows the number of atoms of each type in the unit cell.

From the thesis model a set of structure factors can be generated. An incomplete data set is selected including all structure factors where the absolute values of the Miller indices $|h|$ and $|k|$ are smaller than, or equal to, 2 and 4 respectively. Hence, there are a total of 45 structure factors. However, because the unit cell has point group symmetry mm some of the 45 structure factors are symmetry related, so the total number of unique structure factors is 15, these are the structure factors listed as F_{hk}^{model} in table 4.2. Each of the 45 structure factors is associated with a Gaussian standard deviation $\sigma = 0.01\mu_B$. Mock data is generated in accordance with the $\sigma = 0.01\mu_B$ standard deviations using a numerical Gaussian sampling routine. The resulting data set is shown in columns 4 and 5 in table 4.2 and has $Q = 18.2$ relative to the model. The standard deviations in column 5 appear to be smaller than $\sigma = 0.01\mu_B$, this is because the standard deviations listed refer to 'Gaussian averaged' standard deviations. For instance, the standard deviation $\sigma_{12} = 0.005\mu_B$ in table 4.2 is the average

Miller indices						
h	k	F_{hk}^{model} $[\mu_B]$	F_{hk}^{obs} $[\mu_B]$	σ_{hk} $[\mu_B]$	F_{hk}^{obs} $[\mu_B]$	σ_{hk} $[\mu_B]$
0	0	2.012013	2.021786	0.010000	1.490494	1.0000
0	1	-0.488527	-0.496344	0.007071	-1.774944	0.7071
0	2	0.711790	0.723847	0.007071	1.134492	0.7071
0	3	-1.792406	-1.796752	0.007071	-2.907177	0.7071
0	4	0.806576	0.809126	0.007071	1.124114	0.7071
1	0	-1.004830	-1.009398	0.007071	-0.580080	0.7071
2	0	1.251381	1.270658	0.007071	1.185319	0.7071
1	1	1.382979	1.383121	0.005000	1.384047	0.5000
1	2	-1.065333	-1.057195	0.005000	-1.699958	0.5000
1	3	1.202411	1.198062	0.005000	1.269323	0.5000
1	4	-1.366915	-1.370709	0.005000	-1.460932	0.5000
2	1	-0.304160	-0.305367	0.005000	0.856251	0.5000
2	2	0.444078	0.447442	0.005000	0.229245	0.5000
2	3	-1.121853	-1.119203	0.005000	-1.091098	0.5000
2	4	0.506383	0.506306	0.005000	0.046348	0.5000

Table 4.2: The table lists a set of 15 unique model structure factors in column 3 and two mock data sets in columns 4-5 and 6-7 respectively. See text for more information.

standard deviation of the four symmetry equivalent structure factors with Miller indices $(h,k) = (1,2),(-1,2),(-1,-2),\ (1,-2)$. Since we assign the Gaussian standard deviation $\sigma = 0.01\mu_B$ to each of these structure factors, the average standard deviation is $\sigma_{12} = 0.01/\sqrt{4} = 0.005\mu_B$.

The mock data set in columns 6 and 7 is generated in the same way as above, but with the higher standard deviation, $\sigma = 0.5\mu_B$, on each of the 45 structure factors. For this mock data set $Q = 15.1$.

4.4.3 Varying $\bar{M}^s$

The first robustness analysis of the MEMx method of Def. 7 concerns the sensitivity of the mode of the posterior in Eq. (4.24) to changes of the prior model constant, $\bar{M}^s$. To study this, the series of figures from Fig. 4.11 to Fig. 4.13 show reconstructions of the data in columns 4 and 5 of table 4.2 for fixed $Q = 15$ and varying values of $\bar{M}^s$, starting from $\bar{M}^s = 100\mu_B$ moving down to $\bar{M}^s = 10^{-5}\mu_B$. In this section $\bar{M}^d$ is kept equal to zero, the behaviour of the mode of the posterior as a function of $\bar{M}^d$ will be studied in Sec. 4.4.6.

In Figs. 4.11-4.13 the mode is visualized as follows. The graphs to the left-hand side of these figures show the mode as magnetization density surface plots and represent $\mathbf{m}^{max}$. The following notation is introduced to explain the graphs to the right-hand side of these figures

- $\mathbf{F}^{max}$ are the structure factors obtained by taking the Fourier transform of $\mathbf{m}^{max}$, i.e. $\mathbf{F}^{max} = \mathbf{A}\mathbf{m}^{max}$, where $\mathbf{A}$ is defined in Eq. (4.23). Thus, $\mathbf{F}^{max}$ is just another representation of the mode but in 'data space'.

- $\mathbf{F}^{model}$ is the vector representing the structure factors which is the Fourier transform of the model in Eq. (4.32).

- $\mathbf{F}^{obs}$ represents the observed structure factors, here the mock data in columns 4 and 5 in table 4.2.

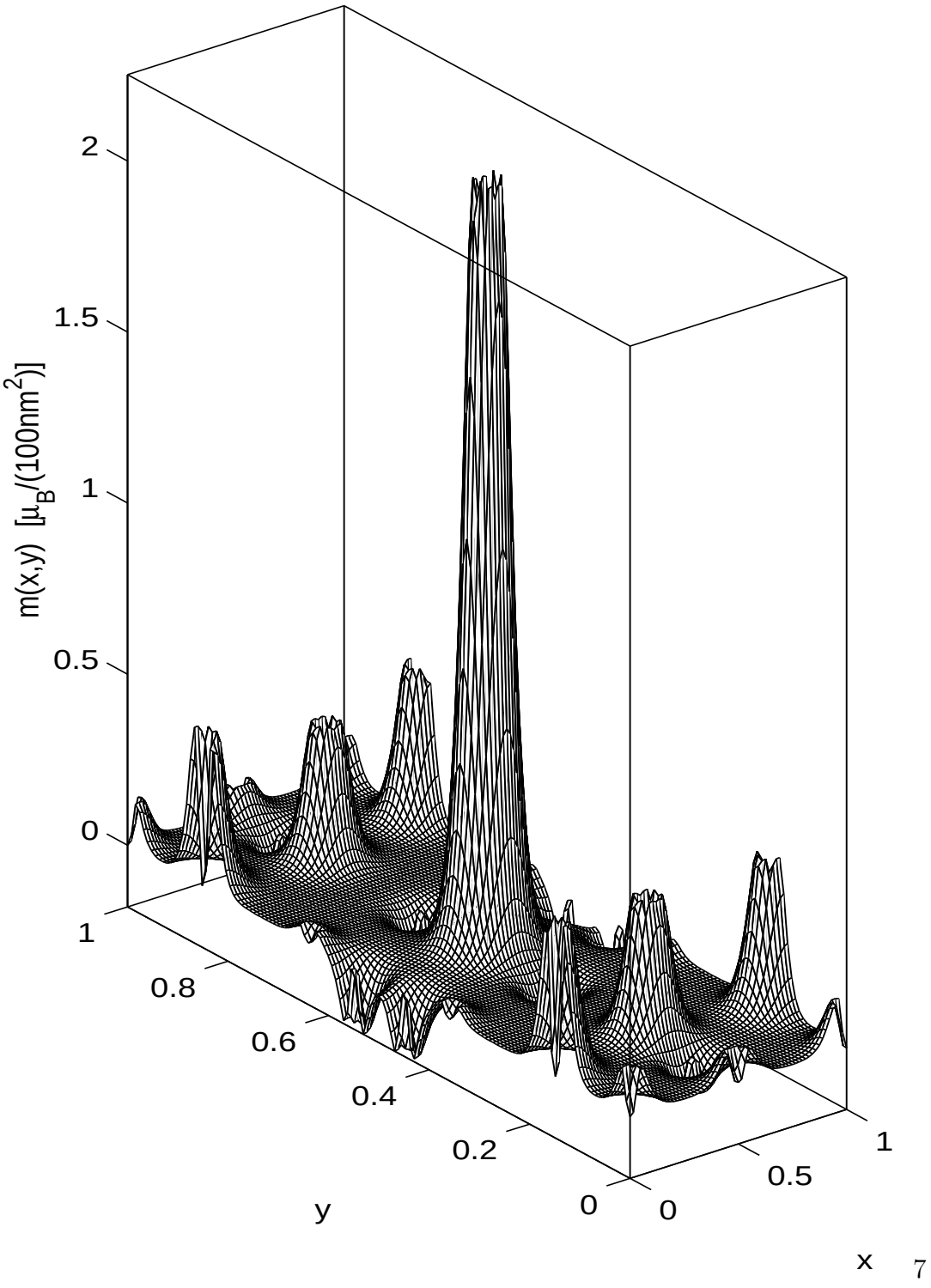

Figure 4.3: The figure shows the thesis model described below Eq. (4.32).

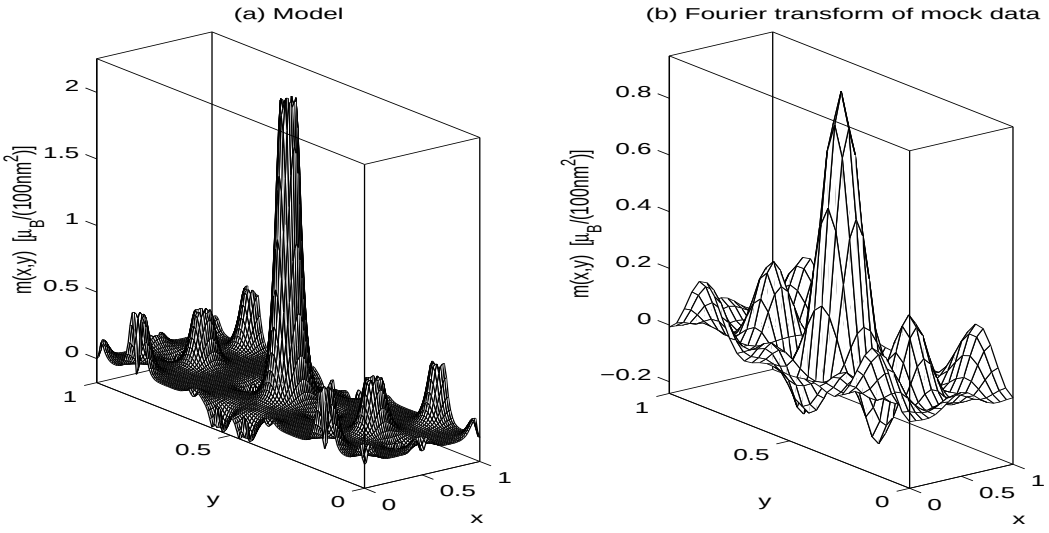

Figure 4.4: (a) shows the same magnetization density image as in Fig. 4.3, and is included to make it easier to compare the thesis model with reconstructions in Fig. 4.11-4.30. (b) shows the Fourier transform of the mock data mapped onto a 16×32 grid.

In the figures to the right-hand side the mode $\mathbf{F}^{max}$ is plotted, as circles, as a function of the scattering length $q = 2\pi\sqrt{(h/a)^2 + (k/b)^2}$. To compare the calculated structure factors $\mathbf{F}^{max}$ with the observed and model structure factors, $\mathbf{F}^{obs}$ are plotted as the big crosses and $\mathbf{F}^{model}$ as the small crosses.

To aid the comparison of a reconstruction with the thesis model in 'magnetization density space', the thesis model is plotted in Fig. 4.4(a) to the same scale as the magnetization density pictures in Figs. 4.11-4.13.

From Figs. 4.11-4.13 first observe the following. It is clear from the structure factor pictures that we have lost most of the fine structure contained in the structure factors with $q > 3.86\mathring{A}^{-1}$ because only data with scattering vector up to $q = 2\pi\sqrt{(\frac{2}{a})^2 + (\frac{4}{b})^2} \simeq 3.86\mathring{A}^{-1}$ are included in table 4.2. It is not possible to apply our Bayesian analysis in Def. 7 in order to recover such fine details. It is also clear from the magnetization density plots (when compared to the thesis model) that fine details, on a scale smaller than 1/4 along the x-direction and 1/8

along the y-direction, are not reconstructed. This is for the simple reason that such information is not present in the input data and is certainly not present in the prior.

Comparing the reconstruction in the top left-hand figure in Fig. 4.11, where the prior constant $\bar{M}^s$ is equal to $100\mu_B$, with the inverse-Fourier density map in Fig. 4.4(b), these two maps look essentially identical. Hence when $\bar{M}^s = 100\mu_B$ is chosen as the prior constant, it is sufficient to take simply the Fourier transform of the data, rather than spending time calculating the mode of the posterior in Eq. (4.24). This is because when $\bar{M}^s$ and, in particular, $\bar{N}^s$ is large, we know from consideration of Fig. 4.2 that it becomes valid to approximate the $\sinh^{-1}$ prior with a Gaussian distribution, i.e.

$$p(\mathbf{m}|\epsilon, \bar{M}^s) = \frac{1}{[2\pi\epsilon\bar{M}^s/(v^2 I)]^{I/2}} \exp\left(-\frac{v^2 I}{2\epsilon\bar{M}^s}\sum_{i=1}^{I} m_i^2\right) . \tag{4.33}$$

Using Eq. (4.33) in the posterior instead of the $\sinh^{-1}$ prior in Eq. (4.12) an analytical solution for $\mathbf{m}^{max}$ is found by changing to data space coordinates using $\mathbf{F}^{cal} = \mathbf{Am}$ and the relation $v^2 I \sum_{i=1}^{I} m_i^2 = \sum_{i=1}^{I} (F_i^{cal})^2$. The mode of the posterior with a Gaussian prior is[18] (see also App. D)

$$F_i^{max} = \begin{cases} \frac{F_i^{obs}}{1+\frac{\sigma_i^2}{\epsilon\bar{M}^s}} & \text{for Miller indices inside } \mathbf{F}^{obs} \\ 0 & \text{otherwise} \end{cases} . \tag{4.34}$$

Here two things are observed. By using a Gaussian prior it is not possible to generate a reconstruction which has non-zero calculated structure factors outside the region of structure factors in $\mathbf{F}^{obs}$ and, in addition, when $\bar{M}^s = 100\mu_B$, $\frac{\sigma^2}{\epsilon\bar{M}^s} = \frac{\sigma^2\bar{N}^s}{(\bar{M}^s)^2} = \frac{0.01^2 * 959400}{100^2} \simeq 0.0096 << 1$ (see Eq. (4.34) and $\bar{N}^s$ is quoted in the top left-hand figure in Fig. 4.11), which explains the reconstruction observed with $\bar{M}^s = 100\mu_B$.

As $\bar{M}^s$ is reduced from $100\mu_B$ to $1\mu_B$, the resemblance of the reconstructed magnetization density maps to the thesis model is improved and the Bayesian

[18] Unit cell is centrosymmetric and $F_i = F_i^*$

image analysis method also returns calculated structure factor values which are non-zero outside the region of structure factors in $\mathbf{F}^{obs}$. Although this is a subjective judgment, the best overall reconstructions are found in the interval $\bar{M}^s \in [0.1; 2]\mu_B$. Furthermore, observe that, as $\bar{M}^s$ is lowered, the areas with non-zero density are truncated to gradually smaller areas and hence more spiky maps appear. This behaviour is also seen for example, by following the calculated structure factors in the right-hand side figures in Fig. 4.13, as $\bar{M}^s$ decreases.

The main purpose of these figures is to obtain an idea of the sensitivity of the mode of the posterior to changes in $\bar{M}^s$. As is observed, the prior constant $\bar{M}^s$ does indeed play an important role in determining the quality of the reconstruction. It shows that a prior model constant like $\bar{M}^s$ cannot be ignored, it must be *clearly specified* to ensure that the prior information in the MEMx application is fully specified.

Although some details may be different the overall behaviour which is observed in Figs. 4.11-4.13 will apply in general. The exception is when $I = P$ and possibly when I is very close to P[19] but the $I = P$ case is unlikely to be of practical interest. Possible options for selecting good $\bar{M}^s$ values will be discussed in Sec. 4.5.2.

4.4.4 Varying Q with smaller errorbars

As for the experiment (i.e. picture series) discussed in the previous subsection the data in columns 4 and 5 in table 4.2 with the smaller errorbars (smaller σ's) is used and Fig. 4.14 and Fig. 4.15 show reconstructions with Q fixed and $\bar{M}^s$ varying. In Fig. 4.14 $Q = 1500$ and in Fig. 4.15 $Q = 0.15$. The observed behaviour is exactly the same as that described in Sec. 4.4.3 and indicates that the reconstructed magnetization densities are independent of the choice of Q.

[19]The exception to this rule is when $I = P$, or I slightly bigger than P, then the likelihood part of the posterior is expected to completely dominate the posterior and therefore be independent of any prior model constants.

This is supported by the series of pictures in Figs. 4.16-4.17, where $\bar{M}^s = 2\mu_B$ throughout, and Q is changed from 0.0015 to 1500, an increase of 6 orders of magnitude. Only as Q approaches 1500 do small changes in the reconstructed magnetization density images appear. Therefore, from Figs. 4.14-4.17 it is fair to conclude that the overall structure of the reconstruction for this range of Q values is independent of the value of Q when using the thesis data in table 4.2 with the smaller errorbars.

4.4.5 Varying Q with larger errorbars

The mock data used in the previous section was the data in columns 4 and 5 in table 4.2. This data set has $\frac{\sigma}{F^{obs}}$ ratios of the order of 0.01. The conclusion of the previous section remains the same for any data set which is measured with an accuracy equal to or better than this. For data of poorer quality, and higher $\frac{\sigma}{F^{obs}}$ ratios, the mode of the posterior will start to be less robust to changes in Q. This will happen for gradually smaller values of Q as the quality of the data is reduced.

That behaviour can be understood from the form of the posterior. If the posterior is written in structure factor coordinates, achieved by substituting the m_i's with F_i's using $\mathbf{F} = \mathbf{Am}$, then it should be clear that the likelihood part of the posterior depends only on the structure factors whose Miller indices match those of the observed structure factors, call these F_i's the 'observed' F_i's, whereas, the prior part of the posterior depends on all the F_i's (not only the 'observed' F_i). From the point of view of the prior the effect of a Q constraint is to impose limits on the 'observed' F_i's. If the Q constraint is $Q = 0$ the prior must 'select' maps such that $F_i = F_i^{obs}$ for all the 'observed' F_i, and the remaining F_i values (which may be called the 'unobserved' F_i's) are chosen according to the prior and prior model constant $\bar{M}^s$ while keeping $F_i = F_i^{obs}$ for all i's belonging to $\mathbf{F}^{obs}$. Consequently, if the $\frac{\sigma_i}{F_i^{obs}}$'s ratios are small, say smaller than 0.01, then

changing the Q constraint from $Q = 0$ to $Q = P$ implies that the 'observed' F_i's on an average must be $F_i^{obs} \pm \sigma_i$, which means the 'constraints' on these F_i's stay effectively unchanged as compared to the constraints on these values when $Q = 0$. As an example, say the i^{th} structure factor has $F_i^{obs} = 1$ with $\sigma = 0.01$. Then $Q = 0$ implies $F_i = F_i^{obs} = 1$ and using $Q = P$ implies $F_i \sim 1 \pm 0.01$. Thus from the point of view of the prior F_i is constrained to approximately the same value. Generalizing this argument for all $i = 1, 2, \ldots, P$ data points gives that the reconstruction with $Q = 0$ and $Q = P$ will be basically the same (keeping $\bar{M}^s$ fixed). From the argument above it is expected that the mode of the posterior will not change significantly as long as the data and value of Q implies that the 'observed' F_i's stay close, on an absolute scale, to the observed structure factors. More specifically, if the 'observed' F_i's on an average are allowed to vary more than say 10% from the F_i^{obs}'s, then a notable difference can be observed in the reconstructed image. Using this approximate thumbrule, we have, for Q values smaller than about

$$Q = \sum_{i=1}^{P} \frac{(F_i^{obs} - 0.9F_i)^2}{\sigma_i^2} = \sum_{i=1}^{P} \frac{(0.1F_i^{obs})^2}{\sigma_i^2} \quad , \tag{4.35}$$

the mode of the posterior will change little as Q is reduced below that value. Making the assumption that the $\frac{\sigma_i}{F_i^{obs}}$ ratios are of about the same magnitude for all measured data points and let $\frac{\sigma}{F^{obs}}$ represent a typical such value then Q in Eq. (4.35) simplifies to

$$Q = P(0.1F^{obs}/\sigma)^2 \quad . \tag{4.36}$$

For such data conditions the level of robustness of the mode of the posterior to variations in Q can be estimated according to Eq. (4.36). For Q values such that $\frac{\sigma\sqrt{Q}}{F^{obs}\sqrt{P}} < 0.1$ the mode of posterior is not sensitive to changes in Q, i.e. it is robust to changes in Q (as was the case in the previous subsection). But, for values such that $\frac{\sigma\sqrt{Q}}{F^{obs}\sqrt{P}} \geq 0.1$ the reconstructed maps can depend strongly on the choice of Q, this is illustrated below.

Figs. 4.18-4.20 represent an experiment similar to that explained in the previous sub-section, but the mock data in columns 6 and 7 is used instead of the mock data in columns 4 and 5 in table 4.2. This data has larger errorbars (larger standard deviations). In the first four frames of Figs. 4.18-4.20 $\bar{M}^s = 0.01\mu_B$ and Q is varied. These data have σ_i/F_i^{obs} ratios of the order one. Therefore, according to Eq. (4.36), the mode of the posterior will depend significantly on Q, for Q values bigger than $Q \simeq P(0.1)^2 = 0.15$. The reconstructed image changes only a little as Q is increased from $Q = 0.1$ to $Q = 1$, but already for $Q = 15$ a big difference is seen. As explained in the text above, this is because the larger $\frac{\sigma}{F^{obs}}$ means that the prior at $Q = 15$ gets a lot more flexibility, on an absolute scale, to 'select' values for the 'observed' structure factors. Frames 5 to 8 in Figs. 4.18-4.20 are the same as frames 1 to 4 but with the more optimal $\bar{M}^s$ value of $0.5\mu_B$, and the same pattern is observed. Note also that, as in Sec. 4.4.3, the $\bar{M}^s = 0.01\mu_B$ reconstructions appear artificially spiky as compared to the $\bar{M}^s = 0.5\mu_B$ reconstructions. The last frame in the picture series Figs. 4.18-4.20 shows the mode with the high $\bar{M}^s$ value of $100\mu_B$ and small value Q of 0.01. As expected from the discussion in Sec. 4.4.3, the resulting reconstruction is basically the same as that obtained by taking the Fourier transform of the mock data.

Based on the observation in Sec. 4.4.4-4.4.5 the prior model constants Q will be discussed further in Sec. 4.5.1.

4.4.6 Without the zero-q structure factor and $\bar{M}^d$

The last prior model constant of the MEMx method to be studied is $\bar{M}^d$. It will be argued that the best choice is to select $\bar{M}^d$ to zero for the type of data and systems considered in this thesis. $\bar{M}^d$ differs from Q and $\bar{M}^s$ in that, in theory, it is possible to relate it to a physical measurable quantity (at least for the case of reconstructing a magnetization density). This will be discussed below.

In a PND experiment it is not possible to measure the zero-q structure factor

which is equal to the total magnetic moment in the unit cell. However, this value can be obtained from a magnetic susceptibility measurement, or alternatively the physics of the system, or a conservation law, may imply that the total moment is known to be exactly equal to a certain value, in which case the probability distribution for the total magnetic moment is a delta function. However, for the data in columns 4 and 5 of table 4.2 such a data point is included and the zero-q structure factor F_{00}^{obs} is Gaussian distributed with mean value $F_{00}^{obs} = 2.021786\mu_B$ and standard deviation $\sigma_{00} = 0.01\mu_B$.

Information about the total magnetic moment in the cell is also contained in the prior model constant $\bar{M}^d$, in the form of the highly non-standard $\sinh^{-1}$ distribution in Eq. (4.12). This is reflected by the fact that, if no data were available, the MEMx method would output an image with a moment equal to $\bar{M}^d$, as discussed in Sec. 4.1. For this reason, it is sometimes found in MEMx literature that a quantity such as $\bar{M}^d$ is referred to as a default model. In which case, the default model relates to the idea of putting $\bar{M}^d$ equal to the magnetic moment which was expected before the data was measured. Although this seems sensible at first sight it actually turns out not to be such a good idea, see discussion of the picture series Figs. 4.25-4.26 below.

In summary, it is seen that the information in $\bar{M}^d$ and information about the value of the total moment in the likelihood (represented by F_{00}^{obs} and σ_{00}) are expected to overlap somewhat. It would therefore be interesting to study them together. At this point one may ask the following questions: *i*) How is information on $\bar{M}^d$ linked with information on F_{00}? *ii*) For the case where no information is available about the total moment, what value should be chosen for $\bar{M}^d$? *iii*) If information is available about the total moment should this information be included in $\bar{M}^d$ or in the likelihood or both?

In the picture series Figs. 4.21-4.22 the experiment illustrated in Sec. 4.4.3 is repeated, still with $\bar{M}^d = 0$, but the zero-q structure factor in column 4 of table 4.2 is excluded. Hence, for these reconstructions the data contains no

information about the total moment. What then is the effect of keeping $\bar{M}^d = 0$? We may have expected that the reconstructed images would have had a total calculated moment of zero; this is clearly not the case as seen from Figs. 4.21-4.22, and can be understood from the form of the prior and from the following example. The prior can be written as a product of I identical terms (see Eq. (4.12)). The purpose of each of these terms is to set the magnetic moment in each of the I bins to zero when $\bar{M}^d = 0$. Therefore the prior function works by attempting to make the moment as small as possible in each bin. As an example consider data which contains evidence of a positive moment around the center of the unit cell and nowhere else, let this data be without a zero-q structure factor and put $\bar{M}^d = 0$. However, the mode of the posterior will have a positive peak at the center of the unit cell and be zero everywhere else. Thus, the reconstruction has a positive total moment. This shows that as soon as any data are available, even data containing no information about the total moment, then there is a non-straightforward relation between $\bar{M}^d$ and the calculated total moment of the mode. Only when we have no data do we know that $\bar{M}^d$ is equal to the total calculated moment of the mode.

It is observed from the figures in Figs. 4.21-4.22 that, as $\bar{M}^s$ is reduced, the total calculated moment goes up until a saturation point, and thereafter it remains almost constant but may eventually start to decrease slowly as $\bar{M}^s$ becomes very small. This behaviour is illustrated in fig. 4.5 which shows $\bar{M}^s$ versus the calculated total moment. Here again we see that despite the lack of a zero-q structure factor in the data and the fact that $\bar{M}^d = 0$, the total calculated moment is not implicitly zero, even when the likelihood contains no information about the total moment.

Otherwise, except for the total magnetic moment, the reconstructions for various values of $\bar{M}^s$ look the same with or without the zero-q structure factor, as long as $\bar{M}^s \geq 0.1\mu_B$. Also note that adding a constant background of $2.012013\mu_B/(3.9\mathring{A} * 11.8\mathring{A}) \simeq 0.044\mu_B\mathring{A}^{-2}$ is the same as adding $2.012013\mu_B$ to

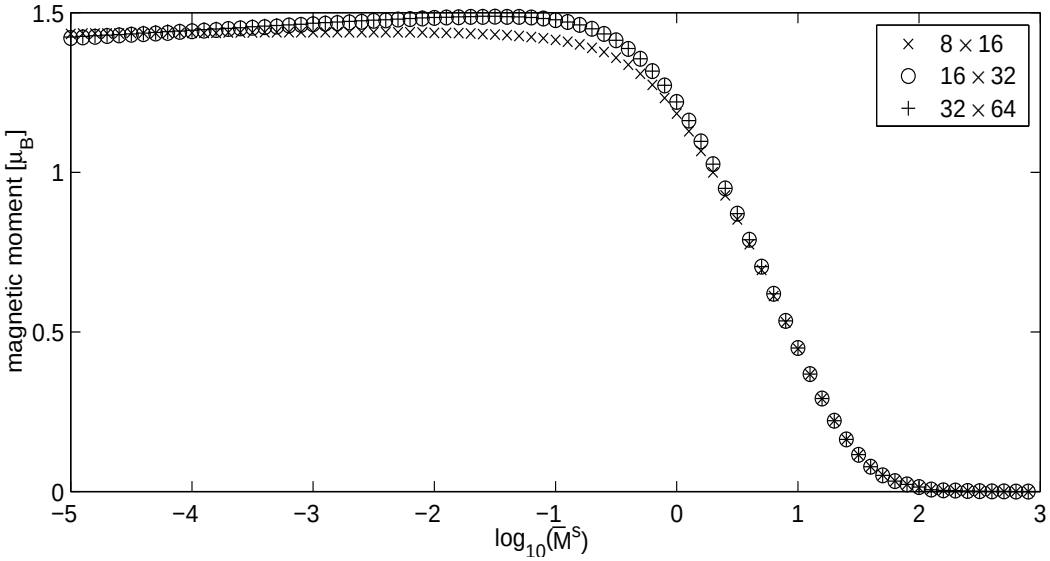

Figure 4.5: A plot of the total calculated moment as a function of $\bar{M}^s$ with $\bar{M}^d = 0$, $Q = 15$ and the mock data in the 4th and 5th column in table 4.2 without the zero-q structure factor. This is shown for the unit cell divided into 8×16 , 16×32 and 32×64 discrete bins.

the unit cell. If $\bar{M}^s$ is below about $\bar{M}^s = 0.1\mu_B$, the resulting magnetization density maps start to decline in quality relative to their counterparts in Sec. 4.4.3. This is seen, for instance, by comparing the bottom magnetization density map in Fig. 4.22 with the bottom left-hand figure in Fig. 4.13.

To generate the plots in Fig. 4.23-4.24 the zero-q structure factor is once again included in the data, and in addition $\bar{M}^d$ is fixed to be equal to the total moment of the thesis model, i.e. $\bar{M}^d = 2.012013\mu_B$, instead of $\bar{M}^d = 0$. Information about the total moment is then available in the form of the Gaussian data point $F_{00} = 2.021786\mu_B$ with $\sigma_{00} = 0.01\mu_B$ and in the prior as $\bar{M}^d = 2.012013\mu_B$. In Fig. 4.23-4.24 the prior constant $\bar{M}^s$ is reduced from $\bar{M}^s = 10\mu_B$ to $\bar{M}^s = 2.01202\mu_B$ (recall that $\bar{M}^s$ must be bigger than the absolute value of $\bar{M}^d$, see below Eq. (4.5)). Probably the most interesting observation here is that contrary to the equivalent reconstructions with $\bar{M}^d = 0$, the total calculated moment for these magnetization density maps is slightly larger rather than slightly smaller. For instance, with $\bar{M}^s = 5\mu_B$ and $\bar{M}^d = 0\mu_B$ the calculated total moment of $2.009\mu_B$ (see title of top figures in Fig. 4.12) is obtained whereas, with $\bar{M}^s = 5\mu_B$ and $\bar{M}^d = 2.012013\mu_B$ a moment of $2.032\mu_B$ is obtained (see title of middle figures Fig. 4.23).

Finally, in the picture series in Figs. 4.25-4.26 the same experiment as above is repeated, but this time with the zero-q data point excluded from the data and a value of $\bar{M}^d$ equal to the total moment of the thesis model is used. It is observed that the total calculated moments are larger than $\bar{M}^d = 2.012013\mu_B$, starting from $v\sum_i m_i^{max} = 2.598\mu_B$, when $\bar{M}^s = 10\mu_B$, increasing to $v\sum_i m_i^{max} = 4.347\mu_B$, when $\bar{M}^s = 2.0121\mu_B$. Comparison with the calculated moments in Figs. 4.11-4.13 shows that the total calculated moment when $\bar{M}^d = 2.0121013\mu_B$ is approximately equal to $2.0121013\mu_B$ plus the total moment calculated when $\bar{M}^d = 0$, keeping $\bar{M}^s$ and Q the same.

There is clear indication from the pictures presented in this subsection that if information about the total moment is available either from an experiment

or other source, then this information should be included as a constraint in the likelihood rather than as a non-zero value for $\bar{M}^d$. There are a number of reasons for this. One is a pragmatic reason: It has been found that choosing a non-zero value for $\bar{M}^d$ does not make a significant improvement to the reconstruction. Hence, for simplicity, keep $\bar{M}^d = 0$. If a non-zero value for $\bar{M}^d$ was used, then what value should be chosen? An intuitive choice would be to take $\bar{M}^d$ to be equal to the known total moment. But as we saw in Fig. 4.25-4.26 this results in a reconstruction that is worse than the reconstruction with $\bar{M}^d = 0$. Hence, because no evidence has been found to suggest otherwise, the use of $\bar{M}^d = 0$ appears to be the best choice and $\bar{M}^d = 0$ will therefore be used for the rest of this thesis.

4.4.7 Different grids

Our purpose in this subsection is to show that observations made in the previous subsections are not affected by the use of different grids. This is what is expected from the mode of the posterior in Eq. (4.24). Finding the mode of Eq. (4.24) is the same as optimizing the expression in Eq. (4.26). It is the term $\frac{v}{\epsilon}S(\mathbf{m})$ of Eq. (4.26) which could depend on I. However, for fixed prior model constants ϵ, $\bar{M}^s$ and $\bar{M}^d$ we have $\frac{v}{\epsilon}S(\mathbf{m}) \propto \frac{1}{I}S(\mathbf{m})$ is approximately independent of I for $\mathbf{m}$'s close to $\mathbf{m}^{max}$. This can be seen from the form of $S(\mathbf{m})$ in Eq. (4.25) and prefactor $\frac{1}{I}$. Hence, the mode of the posterior in Eq. (4.24) will basically be independent of the choice of grid, as observed in Fig. 4.27-4.30.

4.5 Best choices for the prior model constants $\bar{M}^s$ and Q

As has been seen in previous sections, the mode of the posterior is in general critically affected by the choice of $\bar{M}^s$ Q and, depending on the 'quality' of the

data, may be sensitive to the choice of Q. In this section, possible options will be investigated for assigning values to $\bar{M}^s$ and Q based on the results in the previous section with the aim of selecting prior model constants such that the balance between prior and likelihood result in the best possible output. In all that follows it is assumed that $\bar{M}^d = 0$ for reasons explained in Sec. 4.4.6.

4.5.1 Searching for an estimate of Q

According to Sec. 4.4, for the case where the structure factors are measured within 1% accuracy (i.e. $\sigma/F < 0.01$), the MEMx reconstruction is basically independent of the choice of Q, for any reasonable value of Q, which is say $Q \in [0; 100P]$. However, when some or all of the measured data are of poor quality then the MEMx reconstruction will start to depend on Q, as was seen in Sec. 4.4.5. Therefore the problem of selecting a value for Q in such a case needs to be addressed.

The fact that the mode of the posterior is independent of Q for good measured data may come as a surprise considering how a statistical quantity such as Q is used in other fields. For instance, in crystal structure determination from powder or single crystal diffraction data, a Q-like statistic is often monitored as part of a crystal structure refinement process. Here the number of parameters to be refined is much smaller than the number of data points (i.e. structure factor values) and a crystal structure is said to be solved when a set of values for the parameters (lattice parameters, atomic positions etc.) result in a Q value which is small enough, say, for example when $Q < P$ (where P = number of data points). However for the MEMx problem there are typically many more parameters than data points. This implies that for any choice of Q there will in general be an infinite number of parameter solutions which satisfy that Q constraint, where some of these can be classified as 'good' and some as 'bad' solutions. A purpose of applying the MEMx method is to try to select the 'best' solution among all

the solutions which satisfy a given Q value. Hence, a 'good' solution may be obtained from a large range of Q values using the MEMx, and it may happen as was the seen for the data discussed in Sec. 4.4.4 that this 'good' solution remains basically unchanged throughout such a Q interval.

The problem of selecting a Q value for a MEMx reconstruction has been discussed in a number of publications. For those papers which actually quote the value of Q used, the most commonly used value by far is $Q = P$. Hence, for those papers where a value of Q is not quoted it is likely that $Q = P$ was used. The argument for using the $Q = P$ comes from the fact that given just the likelihood probability distribution, the statistic Q is χ^2_P distributed with P degrees of freedom, see Chap. 2. Hence, for any moderate size P, χ^2_P peaks at $Q = P$, so a $Q = P$ constraint is employed.

To my knowledge, the best known attempt for coming up with an alternative to the $Q = P$ option is by Gull & Skilling [90],[91],[92]. These authors go so far as to use different MEMx names depending on how Q is selected. They refer to "Historic" maximum entropy when $Q = P$ is used. When following their alternative way of selecting a value for Q the name "Classical" maximum entropy is used [91],[92]. Gull & Skilling use a Bayesian probability analysis to come up with an alternative to $Q = P$. Although this approach is potentially sound, findings in the following subsections show that it is not to be recommended. Furthermore, various other alternatives for selecting a value for Q have been tested but none of these have been satisfactory. The reason for this is partly related to a property of the prior which is not shared by the mode of the posterior as will be demonstrated. Consequently, at the end of the discussion in the following subsections the use of $Q = P$ method for selecting a value for Q will be recommended, simply because this method appears to be the best choice when compared to possible alternatives.

The discussion in the remainder of this section (until Sec. 4.5.2) can be omitted without loss of continuity.

Obtaining $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$

Gull & Skilling in Ref. [91],[92] have worked on finding an alternative to the most commonly used constraint equation involving Q, which is $Q = P$, using Bayesian statistical reasoning. The idea is a fairly standard approach in Bayesian statistical applications, that of using the observed data itself to select a value for Q, which can potentially be a good idea for obtaining a good and objective estimate for Q. The aim is to calculate the probability distribution for prior model constants *given* the observed data, this function can be written symbolically as the probability p(prior model constants|data). Gull & Skilling [91],[92] explore this density where the prior model constant is ϵ, or equivalently Q. The probability they explore is therefore $p(\epsilon|\text{data})$. This will be extended to the discussion of $p(\bar{M}^s, \epsilon|\text{data})$ and to investigating whether this probability can provide us with any useful information.

As previously, the data is denoted by $\mathbf{F}^{obs}$ and the aim is to calculate $p(\bar{M}^s, \epsilon|\mathbf{F}^{obs})$. Using Bayes' theorem we get

$$p(\bar{M}^s, \epsilon|\mathbf{F}^{obs}) = \frac{p(\bar{M}^s, \epsilon)p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)}{p(\mathbf{F}^{obs})} \quad . \tag{4.37}$$

$p(\mathbf{F}^{obs})$ is a constant for a given fixed data set. If we don't have any idea *a priori* about what the prior model constants ϵ and $\bar{M}^s$ might be, then this is reflected by assuming $p(\bar{M}^s, \epsilon)$ to be uniformly distributed or at least approximately uniformly distributed, so that in Eq. (4.37) the dominant term is $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ [20]. With $p(\bar{M}^s, \epsilon)$ constant Eq. (4.37) reads

$$p(\bar{M}^s, \epsilon|\mathbf{F}^{obs}) \propto p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon) \quad . \tag{4.38}$$

By calculating $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ a relative probability distribution for $p(\bar{M}^s, \epsilon|\mathbf{F}^{obs})$ is obtained. Consider the joint probability distribution $p(\mathbf{F}^{obs}, \mathbf{m}|\bar{M}^s, \epsilon)$ which

[20] Other options for assigning a distribution to $p(\bar{M}^s, \epsilon)$ could be a Jeffreys (or scale) prior or other non-informative priors, see Chap. 2. It is always possible to go back and check if any such choice for $p(\bar{M}^s, \epsilon)$ changes the conclusions drawn by using Eq. (4.38).

can be written as

$$p(\mathbf{F}^{obs}, \mathbf{m}|\bar{M}^s, \epsilon) = p(\mathbf{m}|\bar{M}^s, \epsilon)p(\mathbf{F}^{obs}|\mathbf{m}, \bar{M}^s, \epsilon) \quad . \tag{4.39}$$

Clearly the last factor does not depend on the prior model constants $\bar{M}^s$ and ϵ, and the equation that will be used for calculating $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ is the following marginalised distribution of Eq. (4.39)

$$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon) = \int p(\mathbf{m}|\bar{M}^s, \epsilon)p(\mathbf{F}^{obs}|\mathbf{m})d\mathbf{m} \quad , \tag{4.40}$$

where $p(\mathbf{F}^{obs}|\mathbf{m})$ is a multivariate Gaussian likelihood and $p(\mathbf{m}|\bar{M}^s, \epsilon)$ is the prior distribution in Eq. (4.12) (with $\bar{M}^d = 0$).

As it stands, the integral in Eq. (4.40) is too complicated to be evaluated even by modern computer standards, and, of course, there is no hope of solving it analytically for the general case. Still, attempts have been made to handle an integral problem of the type in Eq. (4.40) with the purpose of obtaining a replacement for the $Q = P$ constraint. Attempts have been made by Gull & Skilling and by Johnston [15], Chap. 3, Sec. 3.5.3. Both involve the following approximation of the integral in Eq. (4.40). We know how to calculate $\mathbf{m}^{max}$ for which the integrand in Eq. (4.40) assumes its maximum value. This map is just the mode of the posterior and can be calculated with relative ease. Next, make a second order Taylor expansion of the logarithm of the integrand in Eq. (4.40) around $\mathbf{m}^{max}$ and assume that this quadratic approximation is not only valid in a region close to $\mathbf{m}^{max}$ but for all maps $\mathbf{m}$. This approximation transforms the integral in Eq. (4.40) into an analytically solvable integral and the result is a function of $\mathbf{m}^{max}$.

Inserting a Gaussian likelihood and the prior in Eq. (4.12) into the integral in Eq. (4.40) and using the approximation explained above, the integral is evaluated

to

$$
\begin{aligned}
p(\mathbf{F}^{obs}|\bar{M}^s,\epsilon) &= p(\mathbf{m}^{max}|\bar{M}^s,\epsilon)p(\mathbf{F}^{obs}|\mathbf{m}^{max})|\mathbf{\Sigma}|^{1/2}\sqrt{2\pi}^I \\
&= \frac{1}{B(\bar{M}^s/\epsilon)}\frac{v^I}{\epsilon^I}\frac{\sqrt{2\pi}^{I-P}|\mathbf{\Sigma}|^{1/2}}{|\mathbf{\Sigma}_0|^{1/2}}e^{-\frac{1}{2}\sum_{i=1}^{P}\frac{(F_i^{obs}-F_i^{max})^2}{\sigma_i^2}} \\
&\times \; e^{\sum_{i=1}^{I}\frac{v}{\epsilon}\left[\sqrt{(m_i^{max})^2+(\bar{M}^s/V)^2}-m_i^{max}\sinh^{-1}(\frac{m_i^{max}V}{\bar{M}^s})\right]} ,
\end{aligned}
\tag{4.41}
$$

where

$$
\mathbf{\Sigma}^{-1} = \mathbf{A}^T\mathbf{\Sigma}_0^{-1}\mathbf{A} + \mathbf{D} , \tag{4.42}
$$

and $\mathbf{A}$ is the image to data transformation matrix in Eq. (4.23), $\mathbf{\Sigma}_0$ a diagonal matrix with diagonal elements σ_i^2 and $\mathbf{D}$ a diagonal matrix with $D_{ii} = \frac{v}{\epsilon}\frac{1}{\sqrt{(m_i^{max})^2+(\bar{M}^s/V)^2}}$.

Further manipulations of Eq. (4.41) are made in search of a replacement for the $Q = P$ constraint equation in the work of Gull & Skilling and Johnston. However, there is no need for any further manipulations of Eq. (4.41), because it can just be plotted! If $\bar{M}^s$ is kept fixed this is just a one dimensional graph.

The expression in Eq. (4.41) may seem a bit tricky at first to evaluate, because it involves the calculation of a $I \times I$ matrix $\mathbf{\Sigma}$, where I is the number of bins and this number may be as high as e.g. 64*64*64 or more for a three dimensional grid. However, due to the structure of $\mathbf{\Sigma}$, we can reduce to the problem of calculating the determinant to that of a $P \times P$ eigenvalue problem, where P is the number of data points. Since a PND data set rarely contains more than $P = 100$ data points and finding the eigenvalues and eigenvectors of a 100×100 matrix takes little time, then the same applies to $p(\mathbf{F}^{obs}|\bar{M}^s,\epsilon)$ in Eq. (4.41) which contain $\mathbf{\Sigma}$. An explanation of how to do this is shown in App. E.

Clearly the accuracy of the quadratic approximation which led to Eq. (4.41) will depend on how close the integrand in Eq. (4.40) already is to a Gaussian form, this will depend on the choice of image grid I, the prior model constants and the data. As an example, the situation with $I = P$ is expected to give a better quadratic approximation than when $I > P$ in general. The reason being

that when $I > P$ there will be integration variables of the integral in Eq. (4.40) for which the Gaussian part of the integrand cannot dominate over the prior part of the integrand. We would expect there to be $I - P$ of these. Further, the discussion below will contain examples which will indicate the validity of Eq. (4.41).

$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ for large $\bar{M}^s$ values

For large values of $\bar{M}^s$ (and moderate size ϵ's) the prior will get gradually closer to a Gaussian distribution (see Sec. 4.4.3 and Fig. 4.2). Hence, as $\bar{M}^s$ increases the expression for the quadratic approximation in Eq. (4.41) will get closer to the exact expression in Eq. (4.40). With a Gaussian prior Eq. (4.40) can be evaluated analytically. It would therefore be interesting to consider the case where the $\sinh^{-1}$ prior is substituted with its corresponding Gaussian distribution (for large $\bar{M}^s$ only) in Eq. (4.33) in order to get a better understanding of how $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ behaves for large values of $\bar{M}^s$.

With the Gaussian prior in Eq. (4.33), $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ is calculated to be

$$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon) = \frac{1}{(\sqrt{2\pi})^P \prod_{i=1}^{P} \sqrt{\sigma_i^2 + \epsilon \bar{M}^s}} \exp\left(-\frac{1}{2} \sum_{i=1}^{P} \frac{(F_i^{obs})^2}{\sigma_i^2 + \epsilon \bar{M}^s} \right) \quad . \tag{4.43}$$

If all the standard deviations are the same and equal to σ, then the probability in Eq. (4.43) has its maximum when

$$\bar{M}^s \epsilon = \frac{1}{P} \sum_{i=1}^{P} (F_i^{obs})^2 - \sigma^2 \quad . \tag{4.44}$$

If there is symmetry in the unit cell, then it has its maximum when

$$\bar{M}^s \epsilon = \frac{1}{P_{sym}} \sum_{i=1}^{P} (F_i^{obs})^2 - \sigma^2 \quad , \tag{4.45}$$

where P_{sym} is the number of uniquely observed structure factors and σ is here referring to the standard deviation for each of the P structure factors[21]. Eq. (4.43) to Eq. (4.45) are derived in App. D.

[21] Say, F_{11}^{obs} has the three equivalent structure factors F_{1-1}^{obs}, F_{-11}^{obs} and F_{-1-1}^{obs} and each has

With a Gaussian prior the mode of the posterior can also be evaluated analytically. Using the solution for the mode in Eq. (4.34) the mode is related to Q by[22]

$$Q = \sum_{i=1}^{P} \frac{(F_i^{obs} - \frac{F_i^{obs}}{1+\sigma_i^2/(\epsilon \bar{M}^s)})^2}{\sigma_i^2} \quad . \tag{4.46}$$

Assuming again that $\sigma = \sigma_i$, $i = 1, 2, \ldots, P$, Eq. (4.46) simplifies to

$$\bar{M}^s \epsilon = \frac{\sigma}{\sqrt{Q}} \sqrt{\sum_{i=1}^{P} (F_i^{obs})^2 - \sigma^2} \quad , \tag{4.47}$$

which when $\bar{M}^s \epsilon$ in Eq. (4.47) is inserted into Eq. (4.45) tells us that the Q value which maximizes the probability distribution in Eq. (4.43) is

$$Q = \frac{P_{sym}^2 \sigma^2}{\sum_{i=1}^{P} (F_i^{obs})^2} \quad , \tag{4.48}$$

independent of $\bar{M}^s$.

$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ for a couple of simple cases

Numerically Eq. (4.40) can be calculated for a couple of simple cases without invoking Eq. (4.41). The two examples that will be considered are *i*) the special case with one bin and the zero-q data point *ii*) two bins and the zero-q data point.

In Fig. 4.6 $p(F^{obs}|\bar{M}^s, \epsilon)$ is plotted both exactly as in Eq. (4.40) and in the quadratic approximation as in Eq. (4.41) for the one bin one data point example. In the upper two frames the zero-q data point is $F^{obs} = 1\mu_B$ with $\sigma = 0.01\mu_B$ and for the lower two frames $F^{obs} = 1\mu_B$ with $\sigma = 0.5\mu_B$. Both with $\sigma = 0.01\mu_B$ and $\sigma = 0.5\mu_B$ there is a very good agreement between the quadratic approximation

the standard deviation σ. Then an equivalently way of representing this information is by the unique structure factor F_{11}^{obs} with standard deviation $\sigma/\sqrt{4}$ and a specific space group for the crystal. See also end of Sec. 4.4.2.

[22]This result is independent of symmetry as long as the standard deviations correspond to the P structure factors and not the P_{sym} unique structure factors, see previous footnote.

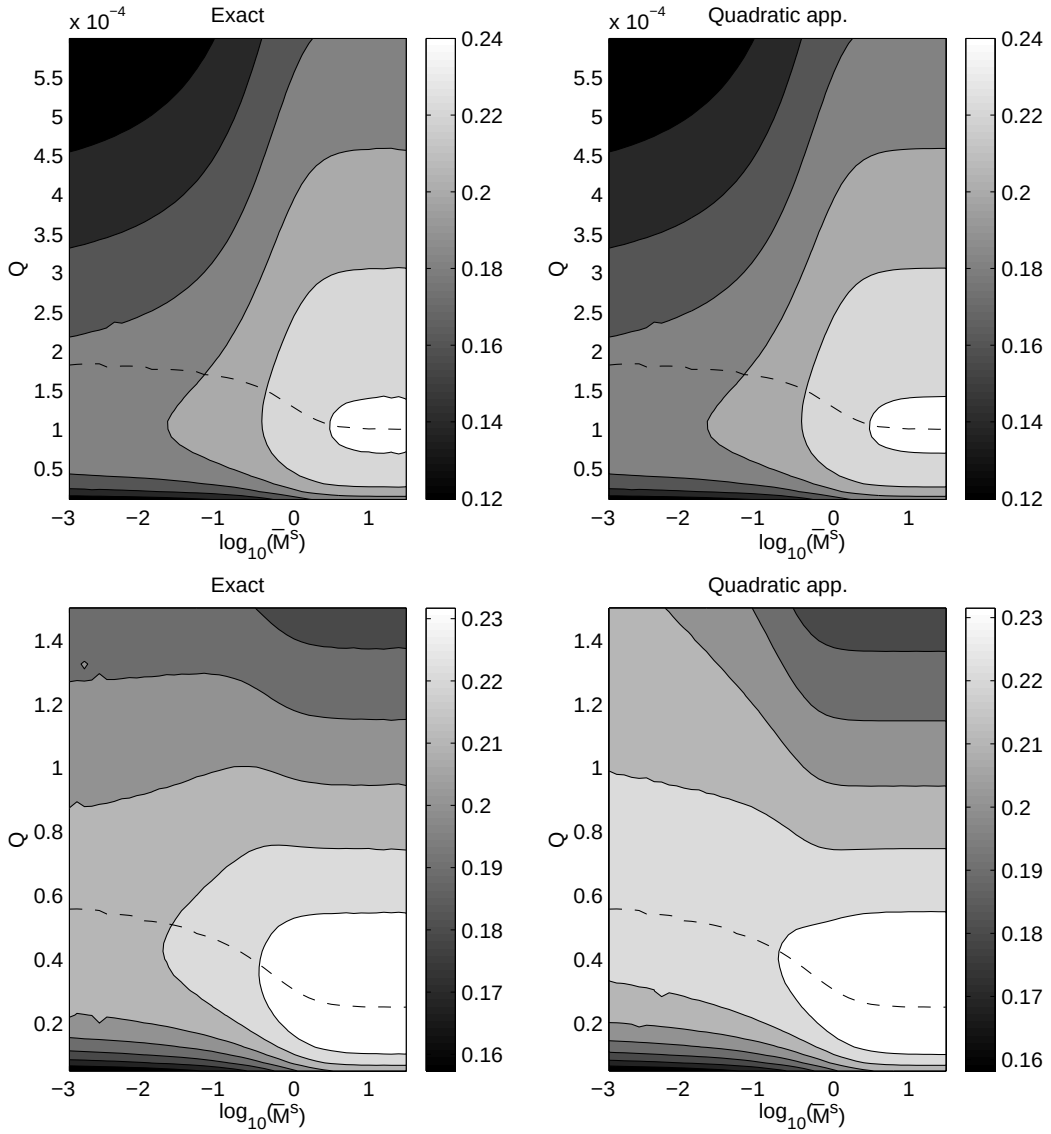

Figure 4.6: $p(F^{obs}|\bar{M}^s, \epsilon)$ is plotted for the one bin one data point example. In the upper two frames $F^{obs} = 1\mu_B$ with $\sigma = 0.01\mu_B$ and for the lower two frames $F^{obs} = 1\mu_B$ with $\sigma = 0.5\mu_B$. The volume of the bin and unit cell is $V = 1\mathring{A}^2$ (the notation from Sec. 4.4.2 is used).

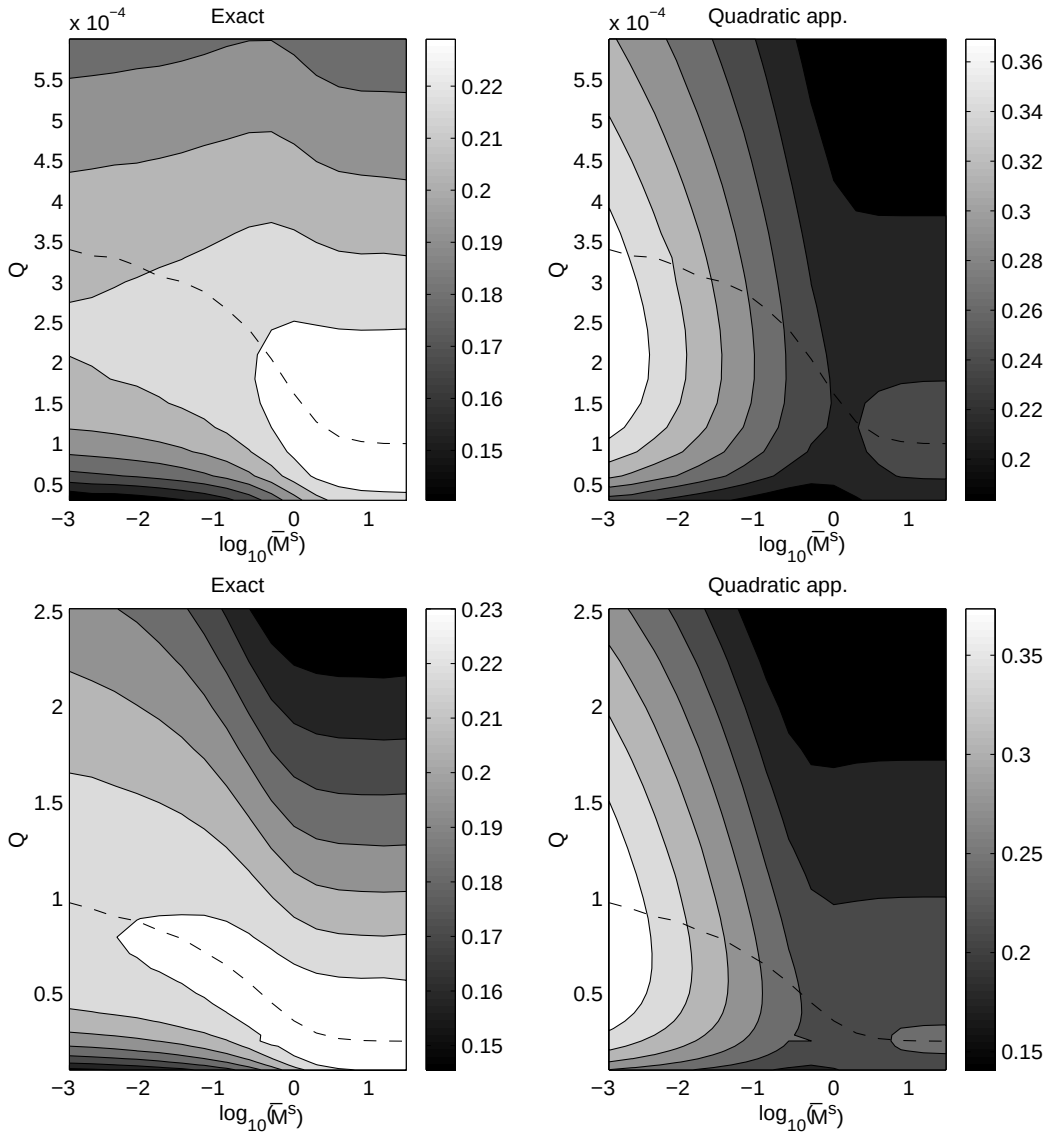

Figure 4.7: $p(F^{obs}|\bar{M}^s, \epsilon)$ is plotted for the two bin one data point example. In the upper two frames $F^{obs} = 1\mu_B$ with $\sigma = 0.01\mu_B$ and for the lower two frames $F^{obs} = 1\mu_B$ with $\sigma = 0.5\mu_B$. The volume of the unit cell is $V = 1\mathring{A}^2$.

and the exactly calculated $p(F^{obs}|\bar{M}^s, \epsilon)$. In all four plots a dotted line is drawn. This line represents the values of ϵ and $\bar{M}^s$ for which the variance of the total moment for the prior distribution is equal to the right-hand side of Eq. (4.45), i.e. the values of ϵ and $\bar{M}^s$ for which

$$f(\epsilon, \bar{M}^s) \equiv \mathrm{Var}[v \sum_{i=1}^{I} m_i]_{prior} = \frac{1}{P_{sym}} \sum_{i=1}^{P} (F_i^{obs})^2 - \sigma^2 \quad . \tag{4.49}$$

It can be seen from the plots in Fig. 4.6 that the dotted line coincides approximately with the maximum of $p(F^{obs}|\bar{M}^s, \epsilon)$ for any fixed value of $\bar{M}^s$. $f(\epsilon, \bar{M}^s)$ will be discussed in more detail below and in the next subsection. A common characteristic of $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ in general and in Fig. 4.6 is that $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ appears to peak sharply as a function of Q for constant $\bar{M}^s$ but not as a function of $\bar{M}^s$ for constant Q. This means that the probability distribution $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ has a clear preference for a Q value but not for a $\bar{M}^s$ value. Hence $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ could potentially be used to select a Q value, thereby giving an alternative to the $Q = P$ criterion. Unfortunately, this does not work, as will be discussed below and in the next subsection.

Fig. 4.7 shows the same plots as in Fig. 4.6 but for the two bins and zero-q data point example. The quadratic approximation is way off the numerically exactly calculated $p(F^{obs}|\bar{M}^s, \epsilon)$ for this example. In particular as a function of $\bar{M}^s$ for constant Q. But, for any fixed value of $\bar{M}^s$, the correct Q dependence seem to be reproduced. Hence, the plots in Fig. 4.7 show that Eq. (4.41) may still be used for the general case to calculate the Q dependence of $p(F^{obs}|\bar{M}^s, \epsilon)$ for fixed $\bar{M}^s$. However, the Q dependence of $p(F^{obs}|\bar{M}^s, \epsilon)$ depends on the grid used, i.e. on I (here whether $I = 1$ or $I = 2$), whereas we know from Sec. 4.4.7 that the mode is basically independent of I. For the simplified case of a system consisting of any number of bins and the zero-q data point the mode is exactly the same for any I. However, by comparing the plots in Fig. 4.6 and Fig. 4.7 it is seen that for values of $\bar{M}^s < 1\mu_B$ the Q dependence of $p(F^{obs}|\bar{M}^s, \epsilon)$ is not the same for the two examples. For higher values of $\bar{M}^s$ ($\bar{M}^s > 10\mu_B$) the prior will

be approximately Gaussian distributed and $p(F^{obs}|\bar{M}^s, \epsilon)$ will peak according to Eq. (4.48). This means that for the $\sigma = 0.01\mu_B$ plots $p(F^{obs}|\bar{M}^s, \epsilon)$ peaks when $Q = 0.0001$ and for the $\sigma = 0.5\mu_B$ plots it peaks when $Q = 0.25$, in accordance with Fig. 4.6-4.7.

Discussion of $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$

In Eq. (4.49) the variance of the total moment for the prior distribution was denoted as $f(\epsilon, \bar{M}^s)$. From all the plots in Fig. 4.6-4.7 it is seen that the Q dependence can be explained from Eq. (4.49), i.e. for a given $\bar{M}^s$, $p(F^{obs}|\bar{M}^s, \epsilon)$ peaks close to the ϵ value for which Eq. (4.49) is satisfied. Since f depends on I, except in the limit where the prior is close to being Gaussian distributed (where $f(\epsilon, \bar{M}^s) \simeq \epsilon\bar{M}^s$), the I dependence of $p(F^{obs}|\bar{M}^s, \epsilon)$ can be traced to the I dependence of f. This is a property not shared by the mode. The mode looks the same for any choice of grid as long as the prior model constants Q and $\bar{M}^s$ are kept constant. So, it must be required from $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ that it is likewise independent of the choice of grid, otherwise it would be possible to have any value of Q to maximize $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$, simply by choosing a particular grid, which is clearly not acceptable. As an example take the mock data in columns 4 and 5 in table 4.2 with $\bar{M}^s = 1$ and make a reconstruction on the following grids: 8×16, 16×32, 32×64, 64×128, 128×256 and 256×512 . For all six grids the mode looks identical however the Q values for which $p(\mathbf{F}^{obs}|\bar{M}^s = 1, \epsilon)$ peaks are: $Q \simeq 0.001$, $Q \simeq 0.008$, $Q \simeq 0.085$, $Q \simeq 1.15$, $Q \simeq 17$ and $Q \simeq 260$ respectively. These values of Q would keep increasing as the grid becomes even finer. The use of the Q value which maximizes $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ is therefore not recommended, and as stated in the beginning of this section the best option is to set $Q = P$.

4.5.2 Estimating $\bar{M}^s$

Based on the detailed analysis in Sec. 4.4.3, and taking into account the nature of the statistical model used in Sec. 4.1 to deduce the $\sinh^{-1}$ prior, the options for selecting a value for the prior model constant $\bar{M}^s$ will be discussed.

The determination of a prior constant like $\bar{M}^s$ can be difficult because it cannot be related directly to any measurable quantity. Even when the true answer is known, as it was when we tried to reconstruct the thesis model in Sec. 4.4, problems still arise, generally because there are different opinions about which of the reconstructions resembles the thesis model best. Therefore, it is the aim of this section not to select *one* value for $\bar{M}^s$ but to provide guidance on how to choose a value for $\bar{M}^s$ by outlining how different choices for $\bar{M}^s$ affect the resulting reconstructed images. It is then up to the individual person to choose a specific value for $\bar{M}^s$ according to this information. Most importantly, however, one should always remember to quote the value used for a prior model constant such as $\bar{M}^s$ in every MEMx application, and the value of any other prior model constants which likewise critically effect the output of a MEMx analysis.

One figure of merit in particular can be easily plotted and which summarizes the behavior of the mode of the posterior in Eq. (4.24) as a function of $\bar{M}^s$. This figure of merit is found to be[23]

$$\mathrm{FOM}_1 = v \sum_{i=1}^{I} |m_i^{max}| \tag{4.50}$$

i.e. the sum of the magnitude of the total negative magnetic moment and the total positive magnetic moment contributing to the unit cell ($\mathbf{m}^{max} = (m_1^{max}$, $m_2^{max}, \ldots, m_I^{max}$) is the mode of the posterior). In Eq. (4.50) the notation FOM_1 stands for Figure Of Merit and the subscript "1" indicates that it is one of many

[23]A number of other figure of merits, such as $v[\sum_{i=1}^{I}(m_i^{max})^2]^{1/2}$ etc., have been tried out as possible quantities which could potentially be used to characterize how the mode of the posterior depends on $\bar{M}^s$. The conclusion was that the best figure of merit is the one in Eq. (4.50).

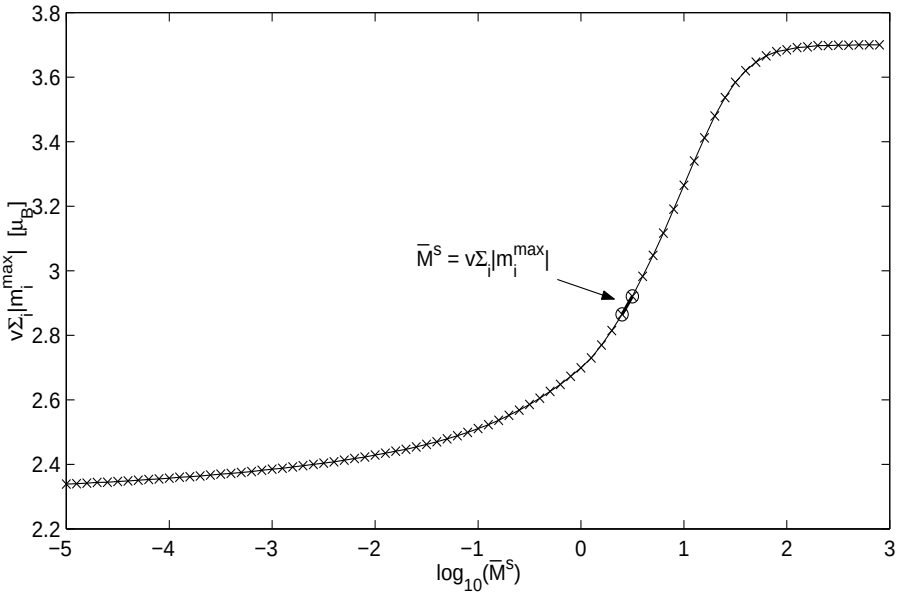

Figure 4.8: The solid curve shows the figure of merit $\text{FOM}_1 = v\sum_{i=1}^{I}|m_i^{max}|$ as a function of $\bar{M}^s$. $Q = 15$ and the mock data in columns 4 and 5 of table 4.2 is used.

possible figures of merit which can be used for the image $\mathbf{m}^{max}$. The use of the particular figure of merit in Eq. (4.50) is also supported by the statistical model used to deduce the $\sinh^{-1}$ prior in Sec. 4.1. In the statistical model of Sec. 4.1, $\bar{M}^s$ is interpreted as the sum of the magnitude of the Poisson mean value of the total negative moment and the Poisson mean value of the total positive moment, see Eq. (4.5), this is basically what is calculated by the figure of merit FOM_1 in Eq. (4.50). From this comparison it is expected that for $\bar{M}^s$ values around the point where the selected value for the prior model constant $\bar{M}^s$ coincides with FOM_1, the reconstruction will have a specific characteristic which can be used as a reference point to map out the ways in which different choices of $\bar{M}^s$ will affect the reconstruction. A FOM_1 versus $\bar{M}^s$ curve will, in general, have a particular shape, and as an example, in Fig. 4.8, FOM_1 is plotted as a function of $\bar{M}^s$ for the thesis data in the 4th and 5th columns of table 4.2. The arrow in Fig. 4.8 shows the value for which $\bar{M}^s = \text{FOM}_1$, and this point will be used as a reference

point for $\bar{M}^s$ and denoted $\bar{M}^s{}_{ref}$. When $\bar{M}^s$ greater than $\bar{M}^s{}_{ref}$ and is increased we move into the region of $\bar{M}^s$ values for which the reconstructions are basically identical to the image obtained by taking the 'inverse' Fourier transform of the data. For instance, in Fig. 4.8 for $\bar{M}^s > 100\mu_B$ we have $\text{FOM}_1 \simeq 3.7\mu_B$ which is close to $v\sum_{i=1}^{I}|m_i^{inv}| \simeq 3.694\mu_B$, where $\mathbf{m}^{inv} = (m_1^{inv}, m_2^{inv}, \ldots, m_I^{inv})$ is the 'inverse' Fourier transform of the data used in Fig. 4.8 (see Fig. 4.4(b) for a plot of $\mathbf{m}^{inv}$).

In Fig. 4.8 $\bar{M}^s$ is equal to FOM_1 when $\bar{M}^s{}_{ref} \simeq 2.85\mu_B$. This is what could be expected since $v\sum_{i=1}^{I}|m_i^{model}| = 2.79\mu_B$, where $\mathbf{m}^{model}$ is the thesis model in Fig. (4.3). As we move away from the 'inverse Fourier' area and down the slope of the solid curve in Fig. 4.8 towards smaller values of $\bar{M}^s$, the areas in the reconstructed image with non-zero magnetization density tend to become gradually smaller[24]. This starts to happen when $\bar{M}^s \approx 10 - 50\mu_B$ in Fig. 4.8. As $\bar{M}^s$ decreases the reconstruction begins to generate calculated structure factors which are non-zero for structure factors lying outside the set of measured structure factors $\mathbf{F}^{obs}$. The reference point is where we have a "natural" transition between the 'observed' calculated structure factors and the 'unobserved' calculated structure factors. Compare for instance the calculated structure factors, the circles, in the top right-hand plot in Fig. 4.11 with the circles plotted in the middle right-hand figure in Fig. 4.12. Also, from the figures in Fig. 4.11-4.13 it is seen that when $\bar{M}^s = \bar{M}^s{}_{ref}$ the reconstruction matches the overall features of the thesis model well. $\bar{M}^s = \bar{M}^s{}_{ref}$ is in general a good choice for $\bar{M}^s$. However, based on the rigously performed simulations in Sec. 4.4 and other simulations not shown here the conclusion is that even better choices for $\bar{M}^s$ lay in the interval

$$\bar{M}^s \in [\bar{M}^s{}_{ref}/8; \bar{M}^s{}_{ref}/2] \quad , \tag{4.51}$$

where the best of these are typically close to $\bar{M}^s = \bar{M}^s{}_{ref}/4$. Eq. (4.51) is based on a large number of simulations and the usefulness of FOM_1 in summarizing

[24]See footnote 19.

the effect of $\bar{M}^s$ on the posterior distribution is supported by the statistical model used to deduce the $\sinh^{-1}$ prior. In choosing $\bar{M}^s$ there is no reason to be too accurate, provided that $\bar{M}^s \sim \bar{M}^s{}_{ref}/4$ there will be no differences in the conclusions drawn from the reconstruction. To get a significant change in the reconstruction a value outside the range in Eq. (4.51) must be used. As an example for determining a $\bar{M}^s$ value in practice, take the thesis data in columns 4 and 5 of table 4.2. From Fig. 4.8, $\bar{M}^s{}_{ref} \simeq 2.85\mu_B$, which is close to $2.5\mu_B$ and divided by 5 (instead of 4) is equal to $0.5\mu_B$. Thus, in choosing a $\bar{M}^s$ value it is not necessary to be that accurate, as long as $\bar{M}^s$ is chosen within the interval in Eq. (4.51) and a good candidate for the thesis data is to chose $\bar{M}^s = 0.5\mu_B$.

One effect of choosing a value for $\bar{M}^s$ which is too small is that the reconstruction will start to appear artificially spiky relative to the real physical situation. This is due to a property of the $\sinh^{-1}$ prior discussed in Sec. 4.4. It is therefore possible (for most data) to check if the value of $\bar{M}^s$ is chosen too small by comparing the calculated and observed structure factors. If there is not a "natural" transition from the calculated structure factors which belong to $\mathbf{F}^{obs}$ to the calculated structure factors outside $\mathbf{F}^{obs}$, then this is evidence that the reconstructed image is artificially spiky. For examples see the bottom figure in Fig. 4.13 or Fig. 4.22.

4.6 Errorbars

So far we have been concentrating on the study of the mode of the posterior, i.e. the most probable image, as the values of the prior model constants are varied. However, what about other images close to the mode? We would like to know more about the probability distribution for images around the mode and formulate a measure of uncertainty for the integrated density of any specified region of the image. In theory all this is provided by the posterior, since it not only returns the value for the probability of the most probable image, i.e.

the mode, but the probability of any image. However, for many multivariate problems and the posterior in Eq. (4.24) this information is inaccessible. This is a reason why we have up until now, just been summarizing the posterior by its mode, hoping that the mode is a good representative of the posterior, as is true for the posterior in Eq. (4.24). The purpose here is to show how more information from the posterior, apart from its mode (its maximum), can be included in an accessible form, that will explain the behavior of the posterior for images about its maximum. To do this the posterior will be approximated with a multivariate Gaussian distribution around its mode. The Gaussian distribution is probably the best known continuous distribution and is fully defined by its mean value (which is equal to the mode in this case) and covariance matrix. Hence in this section the aim is to calculate such a covariance matrix and study it in a similar manner to the way the mode of the posterior was studied in Sec. 4.4-4.5.

In order to do this the second derivative matrix (Hessian matrix) is calculated at the mode and denoted by $-\mathbf{\Sigma}^{-1}$, as done in Eq. (4.42). The posterior is then simplified by approximating it with the multivariate Gaussian distribution[25]

$$\begin{aligned} p(\mathbf{m}|\mathbf{F}^{obs}, \bar{M}^s, \epsilon) &\simeq N(\mathbf{m}^{max}, \mathbf{\Sigma}) \qquad (4.52) \\ &= \frac{1}{\sqrt{2\pi}^I |\mathbf{\Sigma}|^{1/2}} \exp\left(-1/2(\mathbf{m}-\mathbf{m}^{max})^T \mathbf{\Sigma}^{-1}(\mathbf{m}-\mathbf{m}^{max})\right) , \end{aligned}$$

where $N(\mathbf{m}^{max}, \mathbf{\Sigma})$ is the notation commonly used for a Gaussian (or normal) distribution (see Chap. 2), which has mean $\mathbf{m}^{max}$ and covariance matrix $\mathbf{\Sigma}$.

The robustness of $\mathbf{m}^{max}$ as a function of its prior model constants was studied in Sec. 4.4-4.5. To study the behaviour of the correlation matrix $\mathbf{\Sigma}$ as a function of the prior model constants, $\bar{M}^s$ and ϵ ($\bar{M}^d = 0$, see Sec. 4.4.6), consider first the system consisting of one data point, the zero-q data point, and two bins. For this system the mode will be uniform, i.e. $m^{max} \equiv m_1^{max} = m_2^{max}$ and from Eq. (4.42)

[25] $\bar{M}^d = 0$ see Sec. 4.4.6.

we have

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} v^2/\sigma^2 + v/(\epsilon\gamma) & v^2/\sigma^2 \\ v^2/\sigma^2 & v^2/\sigma^2 + v/(\epsilon\gamma) \end{bmatrix} \quad , \tag{4.53}$$

where $\gamma = \sqrt{(m^{max})^2 + (\bar{M}^s/V)^2}$, and m^{max} is the solution to

$$-\frac{1}{\epsilon}\sinh^{-1}\left(\frac{m^{max}V}{\bar{M}^s}\right) + \frac{F^{obs} - Vm^{max}}{\sigma^2} = 0 \quad . \tag{4.54}$$

The inverse matrix of $\mathbf{\Sigma}^{-1}$ in Eq. (4.53) is

$$\mathbf{\Sigma} = \frac{\sigma^2 + v\epsilon\gamma}{\sigma^2 + 2v\epsilon\gamma} \begin{bmatrix} (\epsilon\gamma)/v & -(\epsilon\gamma)^2/(\sigma^2 + v\epsilon\gamma) \\ -(\epsilon\gamma)^2/(\sigma^2 + v\epsilon\gamma) & (\epsilon\gamma)/v \end{bmatrix} \quad . \tag{4.55}$$

For the special case considered, a Q constraint will fix a value for m^{max}. In fact a Q constraint implies $Vm^{max} = F^{obs} \pm \sqrt{Q}\sigma$. If F^{obs} is positive, then because ϵ must be positive in Eq. (4.54), the acceptable solution is $Vm^{max} = F^{obs} - \sqrt{Q}\sigma$, and there is no solution when $\sqrt{Q}\sigma > F^{obs}$ [26]. Insert $Vm^{max} = F^{obs} - \sqrt{Q}\sigma$ into Eq. (4.54) and solve for ϵ to give,

$$\epsilon = \frac{\sigma \sinh^{-1}\left([F^{obs} - \sqrt{Q}\sigma]/\bar{M}^s\right)}{\sqrt{Q}} \quad . \tag{4.56}$$

Insertion of this into one of the diagonal elements in Eq. (4.55) gives

$$\Sigma_{11} = \frac{\sigma\gamma}{\sqrt{Q}v}\sinh^{-1}\left(\frac{F^{obs} - \sqrt{Q}\sigma}{\bar{M}^s}\right)\left[\frac{\sigma^2 + v\epsilon\gamma}{\sigma^2 + 2v\epsilon\gamma}\right] \quad , \tag{4.57}$$

where the last term in the square bracket is a number between 0.5 and 1.0 and can be considered a constant. What is seen from Eq. (4.57) and Eq. (4.55) is an explicit dependence of the variances on the prior constants ϵ and $\bar{M}^s$ in Eq. (4.55) and Q and $\bar{M}^s$ in Eq. (4.57). Taking Eq. (4.57) for $\bar{M}^s/V \gg m^{max}$ (and $F^{obs} \gg \sqrt{Q}\sigma$), and because $\sinh^{-1}(x) \simeq x$ for small x, we have $\Sigma_{11} \simeq \frac{\sigma m^{max}}{\sqrt{Q}v}$ and Σ_{11} is independent of $\bar{M}^s$ but proportional to $1/\sqrt{Q}$. Since $\sinh^{-1}(x) \simeq \log(2x)$ for large values of x, when $\bar{M}^s/V \ll m^{max}$, we have $\Sigma_{11} \simeq \frac{\sigma m^{max}}{\sqrt{Q}v}\log(2\frac{m^{max}V}{\bar{M}^s})$, and

[26] When F^{obs} is negative the acceptable solution is $Vm^{max} = F^{obs} + \sqrt{Q}\sigma$ and there is no solution when $\sqrt{Q}\sigma > |F^{obs}|$.

Σ_{11} depends on $\bar{M}^s$ as $-\log(\bar{M}^s)$ and again is proportional to $1/\sqrt{Q}$. Thus, for the limits considered above, Σ_{11}, is found to be particularly sensitive to changes in Q and proportional to $1/\sqrt{Q}$.

A multivariate Gaussian, $N(\mathbf{m}^{max}, \mathbf{\Sigma})$ has the following convenient property: For any $r \times p$ matrix $\mathbf{C}$ it can be shown that

$$\mathbf{Cm} \sim N(\mathbf{Cm}^{max}, \mathbf{C\Sigma C}^T) \quad . \tag{4.58}$$

In Eq. (4.58) the practical shorthand notation $\sim$ stands for 'distributed as' (see Chap. 2) so Eq. (4.58) says that $\mathbf{Cm}$ is distributed as a multivariate Gaussian with mean values $\mathbf{Cm}^{max}$ and covariance matrix $\mathbf{C\Sigma C}^T$. Consider the following examples. Using the above described two-bin one-data point system as an example we want to determine the distribution of the magnetization density of bin 1. Take $\mathbf{C}$ to be the 1×2 matrix $\mathbf{C} = [1\ 0]$ then $\mathbf{C\Sigma C}^T = \Sigma_{11}$ and therefore the answer to the above question is $\mathbf{Cm} = m_1 \sim N(m^{max}, \Sigma_{11})$, where Σ_{11} (Eq. (4.57)) is the variance of this probability distribution. Likewise, by putting $\mathbf{C} = v[1\ 1]$, a variance estimate of the total magnetic moment, $v(m_1 + m_2)$, of the two-bin one-data point system is obtained

$$\mathrm{Var}[v(m_1 + m_2)] = \mathbf{C\Sigma C}^T = \frac{\sigma^2}{\sigma^2/(\epsilon\gamma V) + 1} \quad , \tag{4.59}$$

where $\mathbf{\Sigma}$ is given by Eq. (4.55). Using Eq. (4.58) it is possible to calculate the variance of the (marginalised) probability distribution of any linear combination of the m_i's. This is the measure of uncertainty which will be used to determine whether a feature in a reconstruction is present or absent.

In general, the covariance matrix elements which are the easiest to relate to are the diagonal elements. Furthermore, from the form of $\mathbf{\Sigma}$ in Eq. (4.42) it is expected that the diagonal elements will be approximately given by[27]

$$\Sigma_{ii} \approx \frac{\epsilon}{v}\sqrt{(m_i^{max})^2 + (\bar{M}^s/V)^2}\ ,\ i = 1, 2, \ldots, I \quad , \tag{4.60}$$

[27]For the diagonal elements of Eq. (4.42) it is expected that $D_i i$ will be the dominating term for most practical cases.

when $I \gg P$. In Fig. 4.9 an example of a full correlation matrix is shown, using the mock data set in columns 4 and 5 of table 4.2 and with $\bar{M}^s = 2\mu_B$ and $Q = 15$. The diagonal elements dominate the covariance matrix, and the top right picture in Fig. 4.10 shows these diagonal elements together with (on the left) a plot of the approximation in Eq. (4.60). The excellent agreement between these two pictures verifies the approximation. In the bottom two frames in Fig. 4.10 the same two pictures are shown but with $Q = 0.0015$ instead of $Q = 15$. It is seen that the values of the diagonal elements are distributed in the same way as in the top two pictures but are greater by one order of magnitude. In comparison we know from Figs. 4.16-4.17 that the reconstructions with $\bar{M}^s = 2\mu_B$ and $Q = 0.0015$ and $Q = 15$ look identical. Hence, we have managed to keep the reconstruction $\mathbf{m}^{max}$ constant and increase the value of the square-root of the variance (i.e. standard deviation) in each bin by a factor of ten [28], this shows that the diagonal elements of the covariance matrix are less robust to changes in Q than the mode of the posterior for this example.

Since it is argued in Sec. 4.4 that Q should always be chosen to be equal to P, Q may be perceived as a fixed quantity and therefore the fact that the diagonal elements of the covariance matrix are less robust to variations in Q compared to the mode of the posterior might seem unimportant. Let us study the usefulness of the measure of uncertainty calculated using Eq. (4.58) and $Q = P$ with some examples and test whether acceptable results are returned.

An important property of $\mathbf{\Sigma}$ is that given a fixed area of an image, the variance calculated for that area, using Eq. (4.58), will be basically unaffected by the choice of image grid used for that reconstruction. A property shared by the mode of the posterior, and this is the reason why $\mathbf{\Sigma}$ must have the same property. To illustrate this consider the covariance matrix to be fully represented by the diagonal elements in Eq. (4.60) and consider the top diagonal element Σ_{ii}, which

[28] In accordence with the discussion below Eq. (4.57), where approximately $\Sigma_{ii}^{1/2} \propto Q^{-1/4}$.

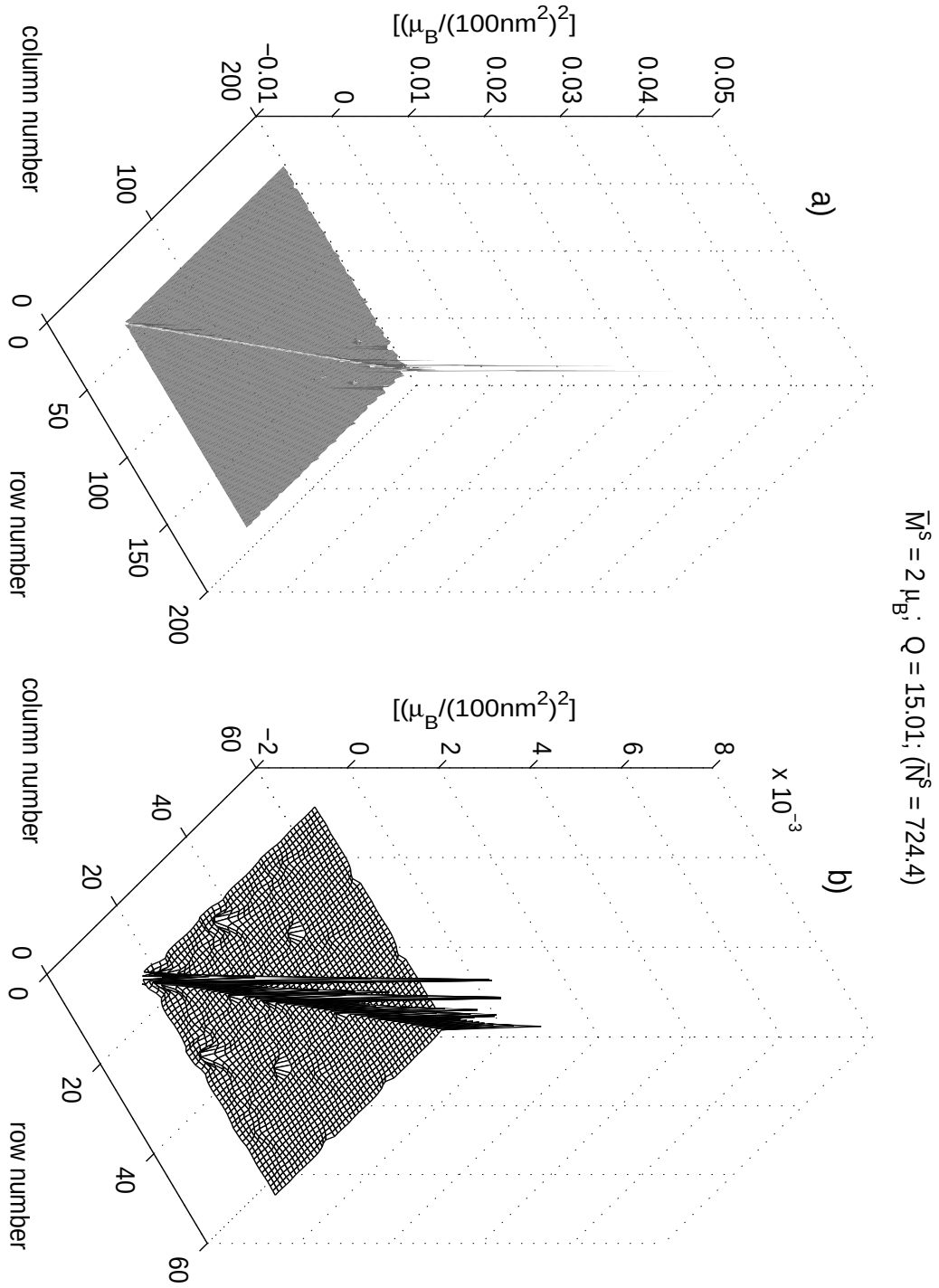

Figure 4.9: a) The correlation matrix is plotted as 153×153 matrix. $153 = 9 * 17$ is the number of unique bins for a 16×32 grid with mm symmetry. b) shows the first 50×50 bins on a finer scale.

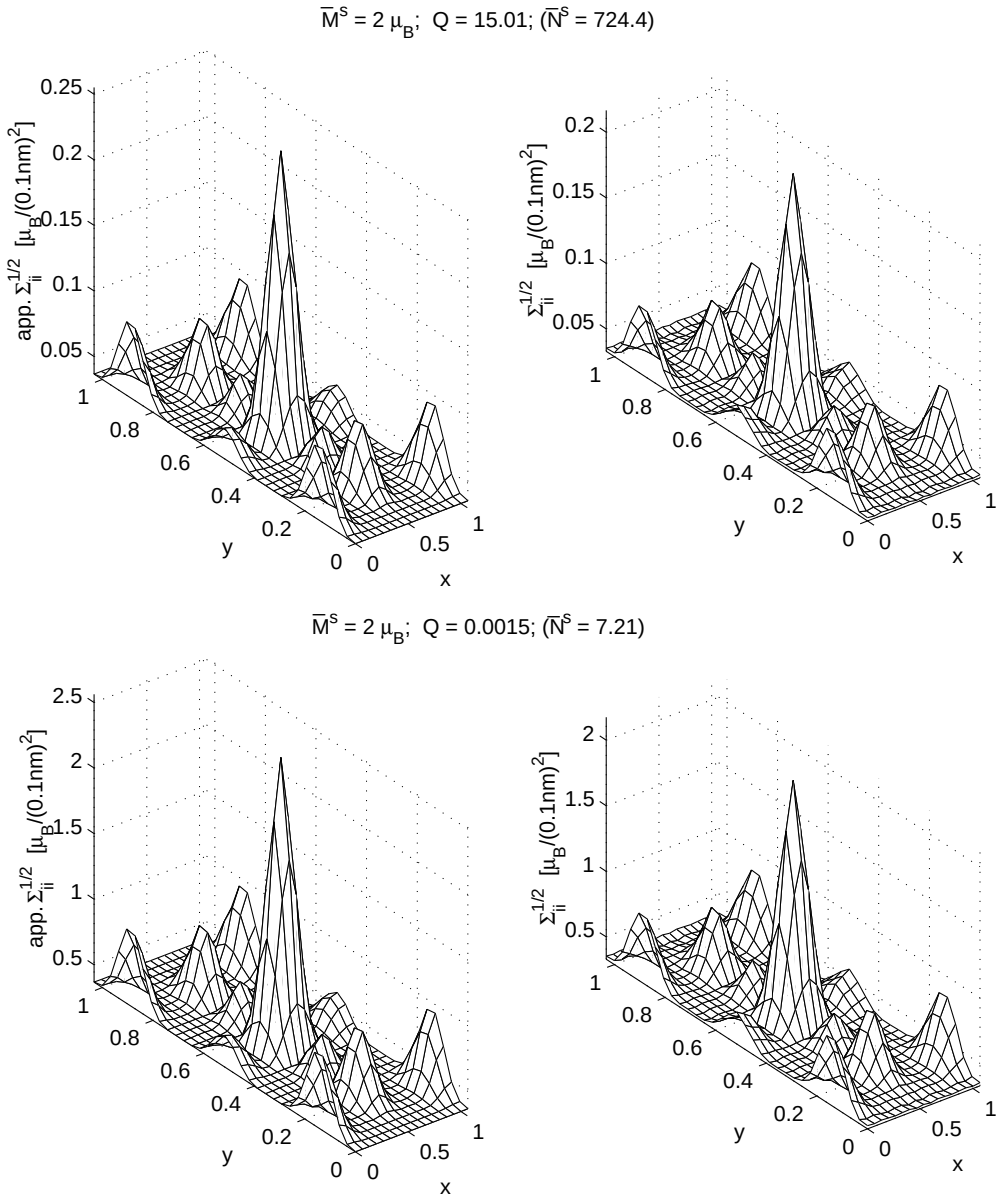

Figure 4.10: The right-hand figures show the square-root of the diagonal elements of the covariance matrix and the left-hand figures the approximation of these diagonal elements in Eq. (4.60). In the upper two frames $Q = 15$ and in the lower two frames $Q = 0.0015$.

is the variance of the (marginalised) probability distribution of the density in bin 1, $\mathrm{Var}[m_1] = \Sigma_{11}$. Say, the magnetization density is displayed on a 2D grid, and scale this grid up by a factor of 2 along both image axes. Then bin 1 in the old grid is split into 4 new bins, and let the values of each of these bins be m'_1, m'_2, m'_3 and m'_4. Because the mode stays almost the same (see Sec. 4.4.7) then we know that the densities satisfy $m'_1 = m'_2 = m'_3 = m'_4 = m_1$. From Eq. (4.60) we see that this implies $\Sigma'_{11} = \Sigma'_{22} = \Sigma'_{33} = \Sigma'_{44} = 4\Sigma_{11}$ and finally using Eq. (4.58) we get $\mathrm{Var}[(m'_1 + m'_2 + m'_3 + m'_4)/4] = \Sigma_{11}$. Thus, the variance of the average value of m'_1, m'_2, m'_3 and m'_4 for the new grid is the same as the variance of m_1 in the old grid. This supports the claim that for any fixed sized area of the image the variance calculated for that area will, for all practical purposes, be unaffected by the choice of grid.

To further examine $\mathbf{\Sigma}$ consider the following example. Project the thesis model 64×128 (Fig. 4.3) onto a 4×8 grid by grouping the bins together, such that the first 16×16 bins become bin number one in the new grid, and so on. The number of projected bins is $4 \cdot 8 = 32$ which is about equal to the number of data points and the calculated standard deviations of the projected bin values may be compared with the standard deviations of the data. The mock data in columns 4 and 5 in table 4.2 is used, and reconstructions are performed with $\bar{M}^s = 0.25\mu_B$, $\bar{M}^s = 0.5\mu_B$ and $\bar{M}^s = 1.0\mu_B$ and keeping $Q = P$. The reconstructions are performed on the 64×128 grid, the same as the thesis model grid (to allow easier comparison between reconstruction and model), and projected mean values and standard deviations for the three reconstructions are shown in table 4.4, table 4.5 and table 4.6 respectively. The tables show the first 3×5 bin values of the projected 4×8 reconstructions. To compare these reconstructed values with the thesis model, table 4.3 shows the mean value of the same 3×5 projected bins.

According to Eq. (4.51) the three $\bar{M}^s$ values used for tables 4.4-4.6 cover approximately the range of recommended $\bar{M}^s$ values[29]. Tables 4.4-4.6 indicate

[29] $\bar{M}^s{}_{ref} \simeq 2.85\mu_B$ for this data set, which means the interval covered by Eq. (4.51) is

that the standard deviations calculated using Eq. (4.58) do not change much as $\bar{M}^s$ is varied. However, even if such changes do occur they are perfectly acceptable since the mode of the posterior has a similar behavior. As an example, the mean density of the first bin with center at $(\frac{15}{128}, \frac{15}{256})$ varies from $3.89 \pm 1.11\ 10^{-2}\mu_B\mathring{A}^{-2}$ to $4.00 \pm 1.00\ 10^{-2}\mu_B\mathring{A}^{-2}$ over the full range of recommended $\bar{M}^s$ values and for the next bin to the right in tables 4.4-4.6 the variation goes from $7.17 \pm 1.42\mu_B\mathring{A}^{-2}$ to $7.10 \pm 1.24\mu_B\mathring{A}^{-2}$, which is fully acceptable.

Looking at the calculated errorbars in table 4.5 for instance, it is seen that apart from the bin features centered at $(\frac{15}{128}, \frac{79}{256})$, $(\frac{47}{128}, \frac{15}{256})$ and $(\frac{79}{128}, \frac{15}{256})$ the remaining mean value in the table are significant, meaning that they have larger mean values than standard deviations. Comparing with the mean values of the thesis model in table 4.3 it is clear that all the significant features in table 4.5 are also present in the thesis model. Hence, the error estimates calculated using $\mathbf{\Sigma}$ can be used to get a quick overview of which features may be present and absent in the reconstruction.

Tables 4.4-4.6 show also, in brackets for each table entry, the value of an approximate standard deviation obtained by using $\mathbf{\Sigma}$ with the approximate diagonal elements in Eq. (4.60) and with all its remaining elements put equal to zero (i.e. its off-diagonal elements = 0). In general, there is a close agreement between the standard deviation in the bracket and the standard deviation calculated using the full covariance matrix (listed just above the value in the bracket). Such an agreement is expected to become less good as the feature integrated over covers a larger fraction of the image. This should be clear because calculating a standard deviation for a gradually larger region using Eq. (4.58) involves more and more of the off-diagonal elements of $\mathbf{\Sigma}$ and although these may indivially contribute little the approximation of putting them all equal to zero clearly becomes less good for gradually larger areas. For small areas (one bin areas) there is almost

$\bar{M}^s \in [0.356; 1.425]$ and this range is close enough to the range of $\bar{M}^s$ covered in tables 4.4-4.6.

Projected thesis model

	$y = \frac{15}{256}$	$y = \frac{47}{256}$	$y = \frac{79}{256}$	$y = \frac{111}{256}$	$y = \frac{143}{256}$
$x = \frac{15}{128}$	2.72	8.34	0.60	-3.31	-3.51
$x = \frac{47}{128}$	-0.46	7.24	-1.01	17.69	22.24
$x = \frac{79}{128}$	-0.54	8.20	-1.15	20.39	24.94

Table 4.3: Shows 3×5 projeced bin mean values. The bin values are in units of $10^{-2}\mu_B \mathring{A}^{-2}$.

an exact agreement, see Fig. 4.10, and for the integrated 16×16 projected bins in tables 4.4-4.6 the agreement is still very good. Hence, for such smaller regions the standard deviations may be calculated using the approximate diagonal elements in Eq. (4.58) and putting all the remaining elements of $\mathbf{\Sigma}$ equal to zero. This can offer a very significant reduction in computational cost and computer storage. Calculating the approximate standard deviations costs almost nothing extra once the mode of the posterior has been found (this is apparent by looking at the expression of the diagonal elements in Eq. (4.60)). Whereas calculating the full covariance matrix can be non-trivial depending on the problem. For instance, for crystals with low symmetry the full storage of $\mathbf{\Sigma}$ requires of the order I^2 doubles (for most computers one double = 8 bytes)[30]. The physical systems treated in this thesis are all of a size and crystal symmetry which enable the full covariance matrix to be calculated, and unless otherwise stated, standard deviations will be obtained this way. App. F explains how $\mathbf{\Sigma}$ can be calculated.

A more obvious way of amalgamating the bins than projecting them onto a 4×8 grid would be to sum up the magnetic moments around the features of interest in the image. As an example, in table 4.7, the magnetic moment around five features of the thesis model is calculated by summing the moment in the bins surrounding these 5 features. The location of the center of these features are

[30]However, it is possible to partially get around this problem by never storing the full covariance matrix at once.

Projected $\bar{M}^s = 0.25\mu_B$ reconstruction

	$y = \frac{15}{256}$	$y = \frac{47}{256}$	$y = \frac{79}{256}$	$y = \frac{111}{256}$	$y = \frac{143}{256}$
$x = \frac{15}{128}$	3.89 ± 1.11 (1.26)	7.17 ± 1.47 (1.69)	0.31 ± 0.49 (0.53)	-2.79 ± 0.95 (1.10)	-3.23 ± 0.99 (1.17)
$x = \frac{47}{128}$	0.21 ± 0.48 (0.52)	6.32 ± 1.40 (1.59)	-1.36 ± 0.81 (0.90)	16.94 ± 2.34 (2.64)	22.64 ± 2.56 (3.02)
$x = \frac{79}{128}$	0.25 ± 0.48 (0.53)	7.53 ± 1.48 (1.73)	-1.56 ± 0.85 (0.95)	19.93 ± 2.48 (2.86)	26.73 ± 2.69 (3.28)

Table 4.4: The data in columns 4 and 5 in table 4.2 is used. The numbers in the brackets show the value of an approximate standard deviation and is discussed in the text. The units for the 3×5 mean values and standard deviations are $10^{-2}\mu_B \mathring{A}^{-2}$.

Projected $\bar{M}^s = 0.5\mu_B$ reconstruction

	$y = \frac{15}{256}$	$y = \frac{47}{256}$	$y = \frac{79}{256}$	$y = \frac{111}{256}$	$y = \frac{143}{256}$
$x = \frac{15}{128}$	3.93 ± 1.04 (1.18)	7.13 ± 1.34 (1.54)	0.36 ± 0.57 (0.62)	-2.87 ± 0.91 (1.04)	-3.30 ± 0.94 (1.10)
$x = \frac{47}{128}$	0.21 ± 0.56 (0.63)	6.32 ± 1.29 (1.45)	-1.51 ± 0.85 (0.93)	17.31 ± 2.15 (2.43)	22.66 ± 2.33 (2.74)
$x = \frac{79}{128}$	0.26 ± 0.56 (0.63)	7.52 ± 1.35 (1.58)	-1.72 ± 0.87 (0.98)	20.35 ± 2.27 (2.62)	26.73 ± 2.44 (2.97)

Table 4.5: See caption table 4.4.

Projected $\bar{M}^s = 1.0\mu_B$ reconstruction

	$y = \frac{15}{256}$	$y = \frac{47}{256}$	$y = \frac{79}{256}$	$y = \frac{111}{256}$	$y = \frac{143}{256}$
$x = \frac{15}{128}$	4.00 ± 1.00 (1.13)	7.10 ± 1.24 (1.42)	0.39 ± 0.69 (0.75)	-2.98 ± 0.91 (1.03)	-3.41 ± 0.92 (1.07)
$x = \frac{47}{128}$	0.19 ± 0.68 (0.76)	6.34 ± 1.20 (1.35)	-1.70 ± 0.90 (0.99)	17.77 ± 1.97 (2.22)	22.76 ± 2.11 (2.46)
$x = \frac{79}{128}$	0.26 ± 0.68 (0.77)	7.49 ± 1.25 (1.44)	-1.92 ± 0.92 (1.03)	20.85 ± 2.07 (2.39)	26.79 ± 2.20 (2.66)

Table 4.6: See caption table 4.4.

description	location	model	data : small s.d.		data : large s.d.	
center peak	(0.5, 0.5)	1.2717	1.2656	0.00993	1.1442	0.26654
edge negative	(0, 0.5)	-0.17977	-0.16258	0.00986	0.02287	0.19806
edge positive	(0, 0.15)	0.25599	0.25899	0.02547	0.10581	0.22358
middle positive	(0.5, 0.18)	0.21781	0.19146	0.02189	0.03201	0.12954
middle negative	(0.5, 0.37)	-0.05620	-0.05665	0.01304	-0.06414	0.1522

Table 4.7: The location of the center of five thesis model features are listed in the second column. The corresponding integrated moment in column 3 in units of μ_B. In columns 4 and 5 are the integrated moments and calculated standard deviations for the reconstruction with $\bar{M}^s = 0.5\mu_B$, $Q = 15$, grid $= 64 \times 128$ using the mock data with the smaller errorbars in table 4.2. In columns 6 and 7 the same but using the mock data with the larger errorbars in table 4.2.

listed in column 2 of table 4.7 and corresponding magnetic moments in column 3. The thesis model is reconstructed on a 64×128 grid using $\bar{M}^s = 0.5\mu_B$ ($Q = P$), using both the data with the smaller and larger errorbars in table 4.2. Again, when comparing thesis and reconstructed mean values it is my opinion that very reasonable standard deviations are obtained from $\mathbf{\Sigma}$.

4.7 Summary and conclusions

Def. 7 defines the Bayesian image analysis technique and MEMx method of this chapter. Once the three prior model constants are specified the optimization problem is well defined and using one of the computer algorithms described in Sec. 4.4.1 the mode of the posterior can be calculated. Furthermore, by calculating a covariance matrix and using Eq. (4.58), as discussed in Sec. 4.6, a reliable error estimate for any feature of the mode of the posterior can be calculated.

From the detailed analysis in Sec. 4.4-4.5 it is suggested that the three prior model constants of Def. 7 are selected as follows

- $\bar{M}^d$: select $\bar{M}^d = 0$ (see Sec. 4.4.6)
- Q : select $Q = P$ where P is the number of data points (see Sec. 4.5.1)
- $\bar{M}^s$: select value in the interval $\bar{M}^s \in [\bar{M}^s{}_{ref}/8; \bar{M}^s{}_{ref}/2]$ and preferably close to $\bar{M}^s{}_{ref}/4$ (see Sec. 4.5.2)

The above recipe for the use of the MEMx method is recommended for the analysis of polarised neutron diffraction data and other data with equivalent information content.

The main new feature in the above list is the suggestion of how to select an $\bar{M}^s$ value. The selections of $\bar{M}^d$ and Q, although used in the literature, have not previously been discussed in terms of a Bayesian robustness analysis. Most important, the analysis shows that the prior model constants critically affect

the output of the MEMx technique and demonstrate the importance of selecting sensible values.

Applications of the MEMx method will follow in the next two chapters. In addition, Chap. 6 will include a comparison between the approach to the MEMx method outlined here and other approaches to MEMx for a practical example, and it will be illustrate how knowledge gained from this chapter can be used to apply the MEMx method more convincingly to the analysis of polarised neutron diffraction data.

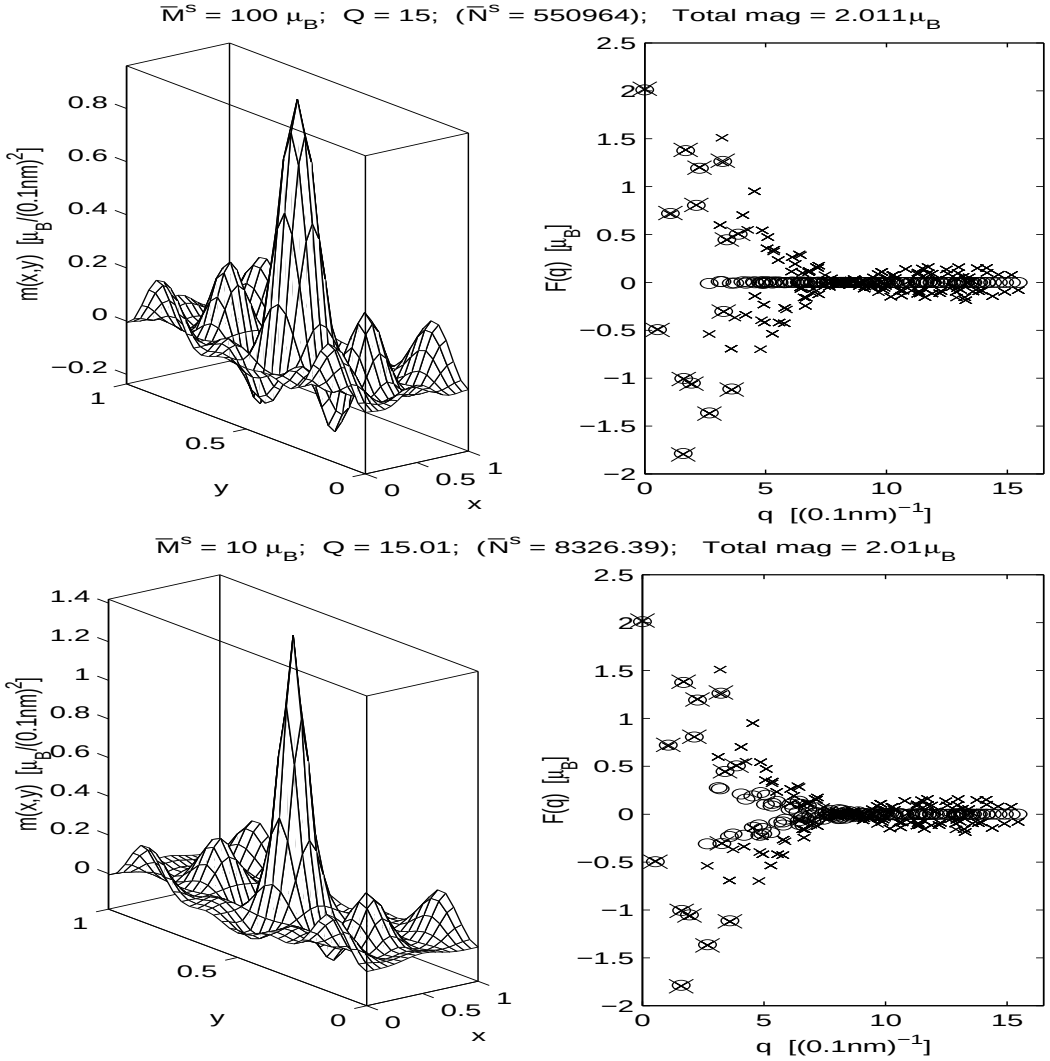

Figure 4.11: The mode of the posterior in Eq. (4.24) with the mock data in columns 4 and 5 of table 4.2, $\bar{M^d} = 0$ and $Q = 15$ over a range of $\bar{M^s}$ values are shown as magnetization density maps in the left-hand figures and as calculated structure factors in the right-hand figures. In the right-hand figures the calculated structure factors are shown as circles. These data points are compared with the structure factor obtained from the model and shown as crosses. The larger crosses show the mock data. q is the magnitude of the scattering vector, $q = 2\pi\sqrt{(h/a)^2 + (k/b)^2}$, where h and k are the miller indices.

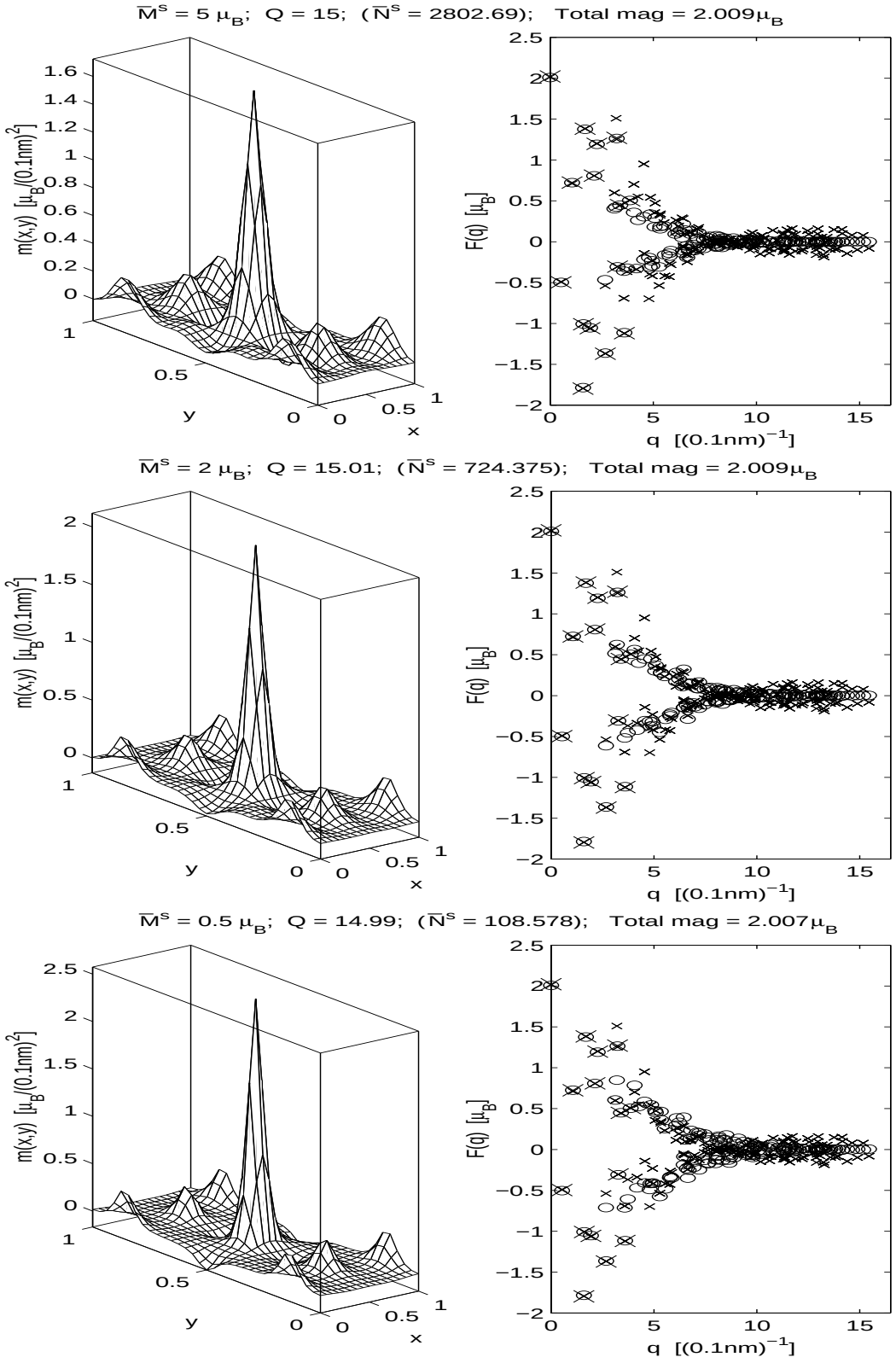

Figure 4.12: See caption Fig. 4.11.

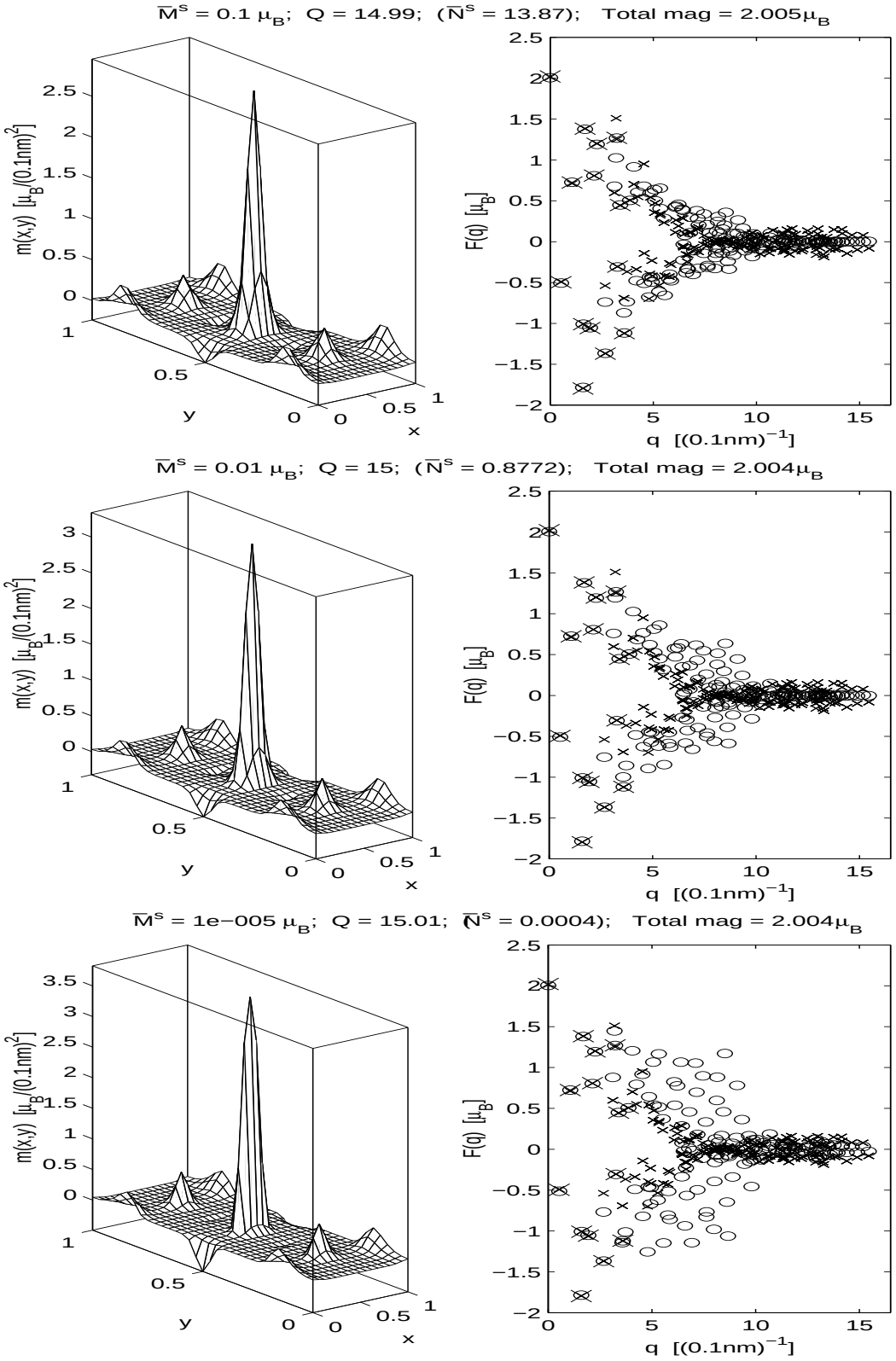

Figure 4.13: See caption Fig. 4.11.

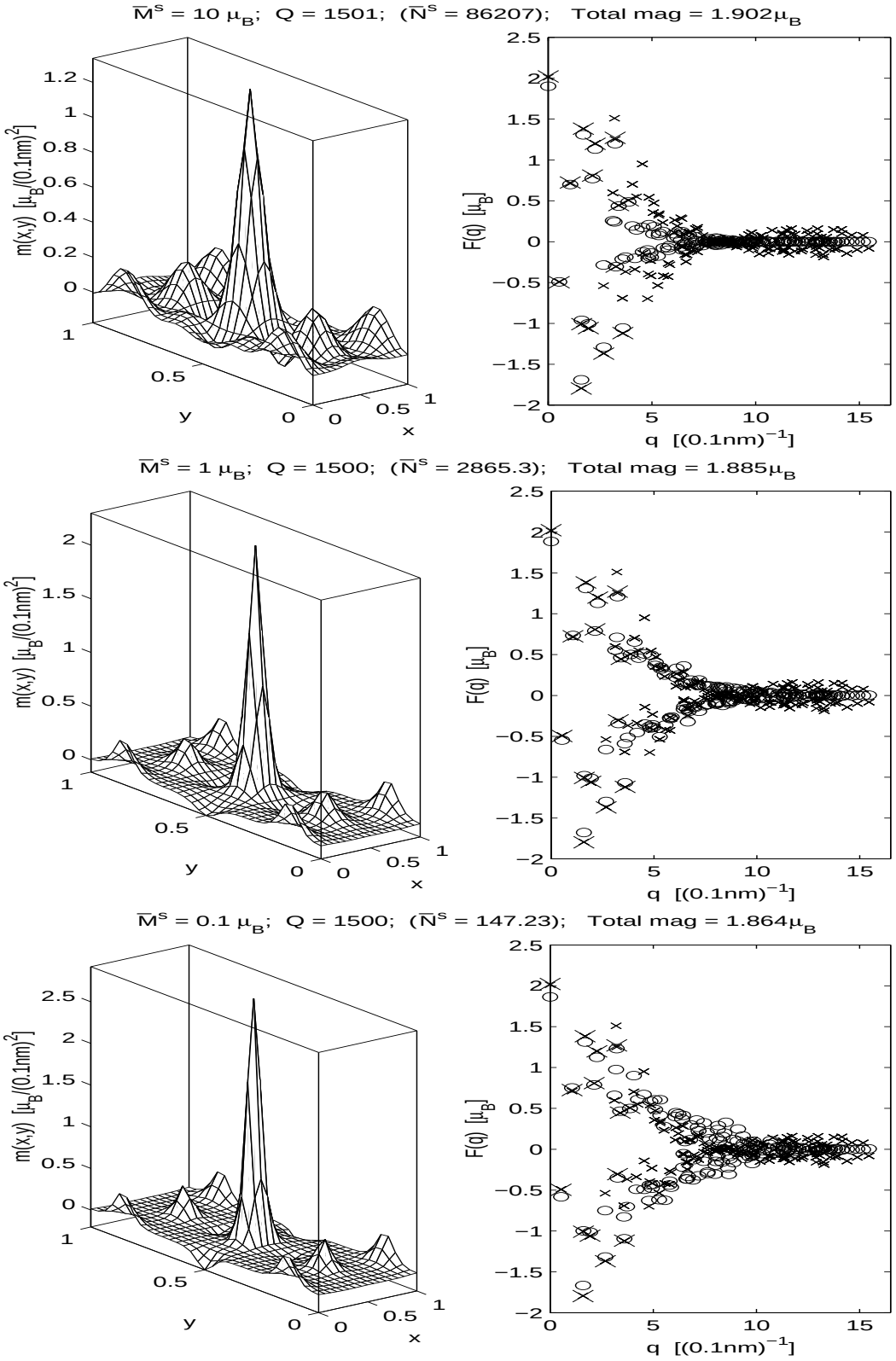

Figure 4.14: Same as in Fig. 4.11, but with $Q = 1500$.

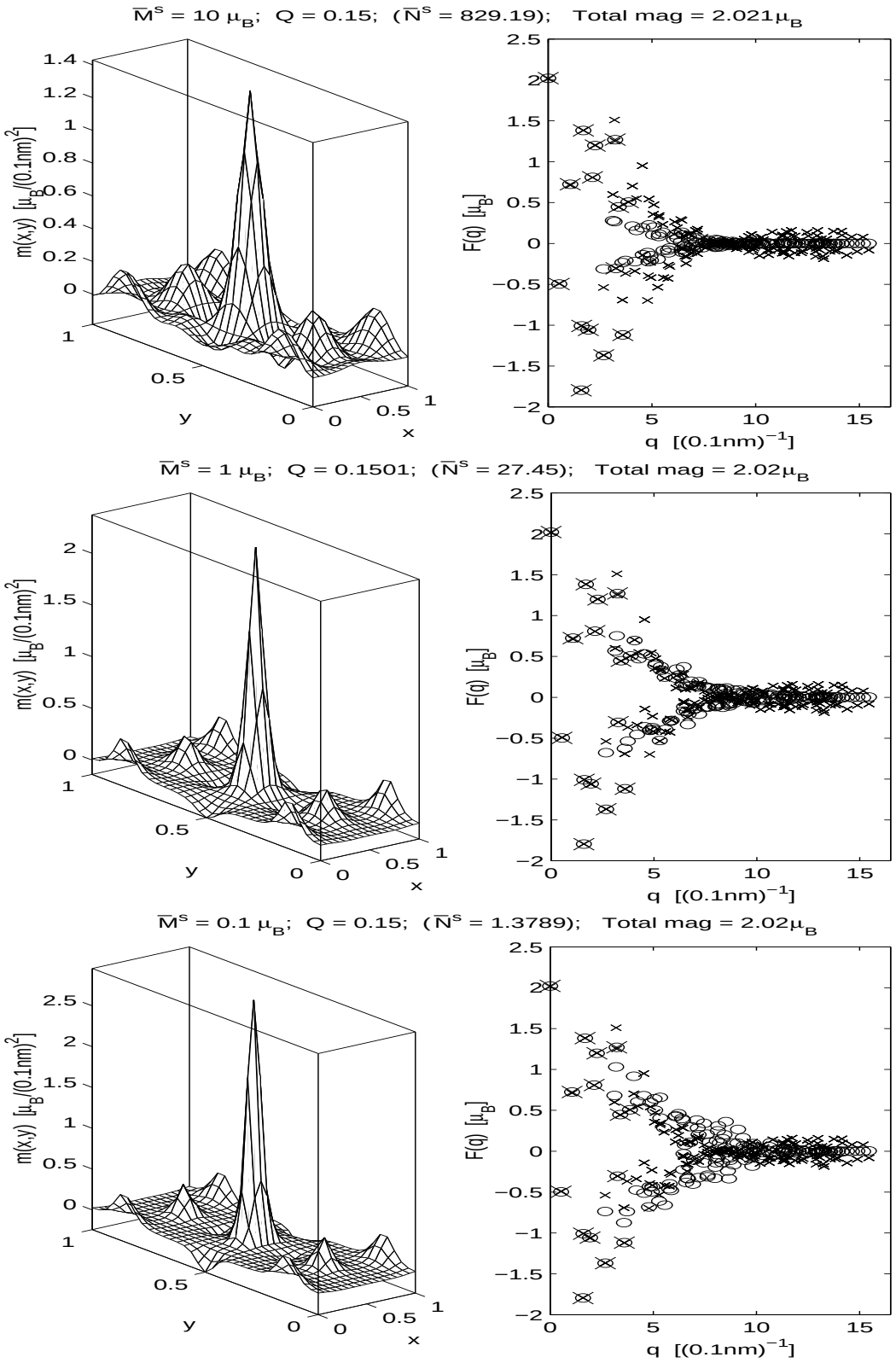

Figure 4.15: Same as in Fig. 4.11, but with $Q = 0.1$.

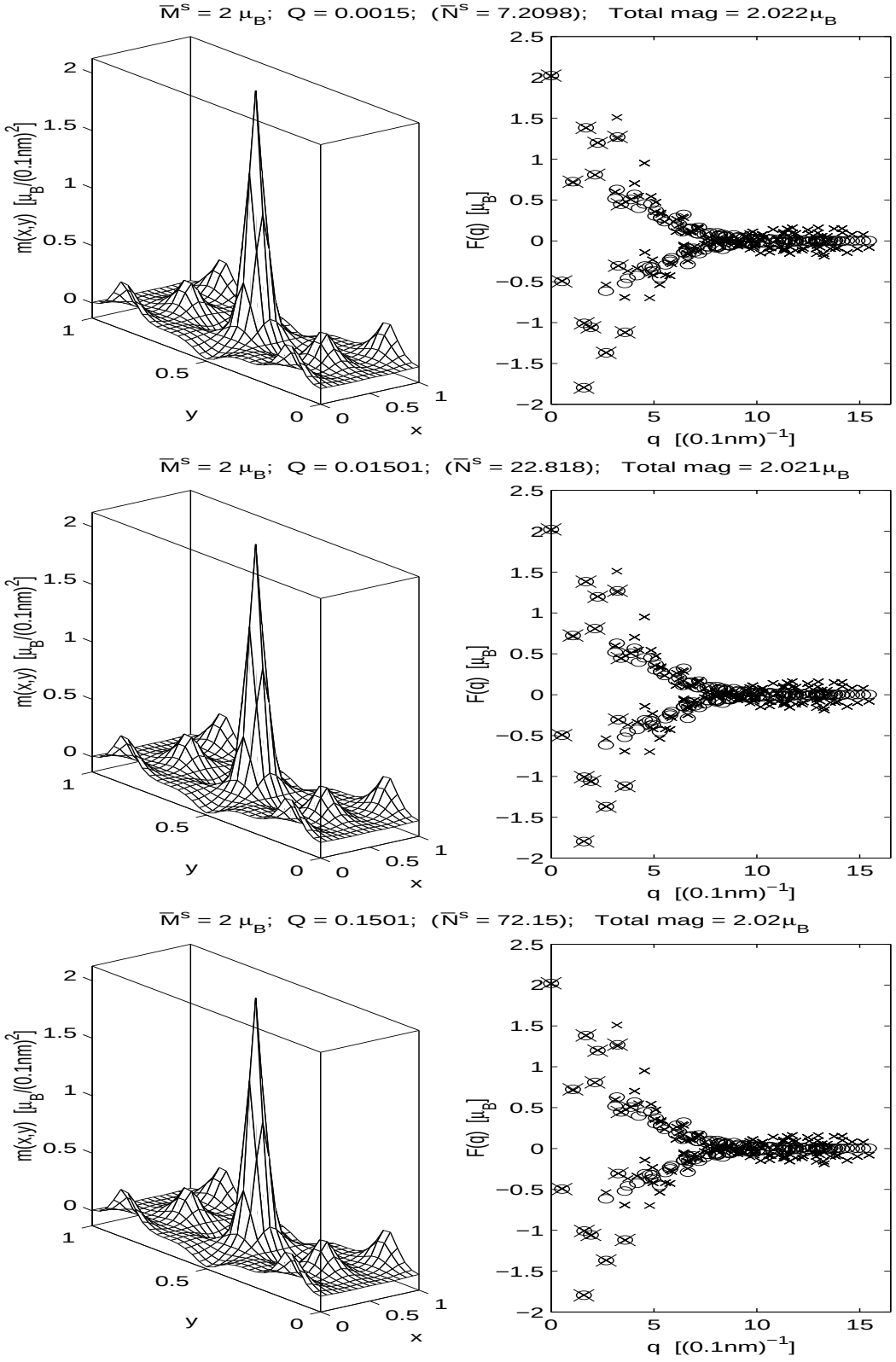

Figure 4.16: Same as in Fig. 4.11, but where $\bar{M}^s$ is kept constant instead of Q and $\bar{M}^s = 2\mu_B$.

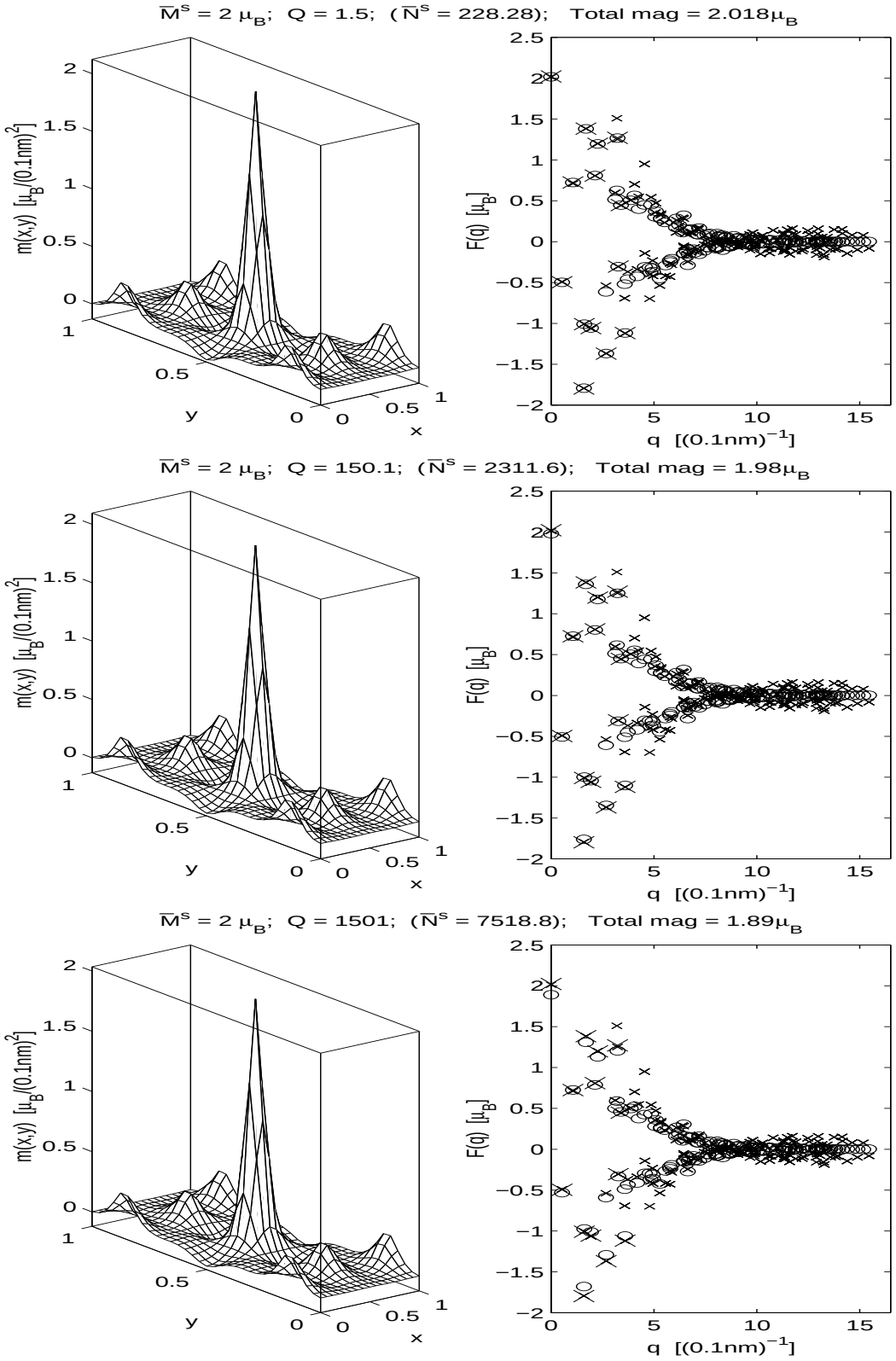

Figure 4.17: See caption Fig. 4.16.

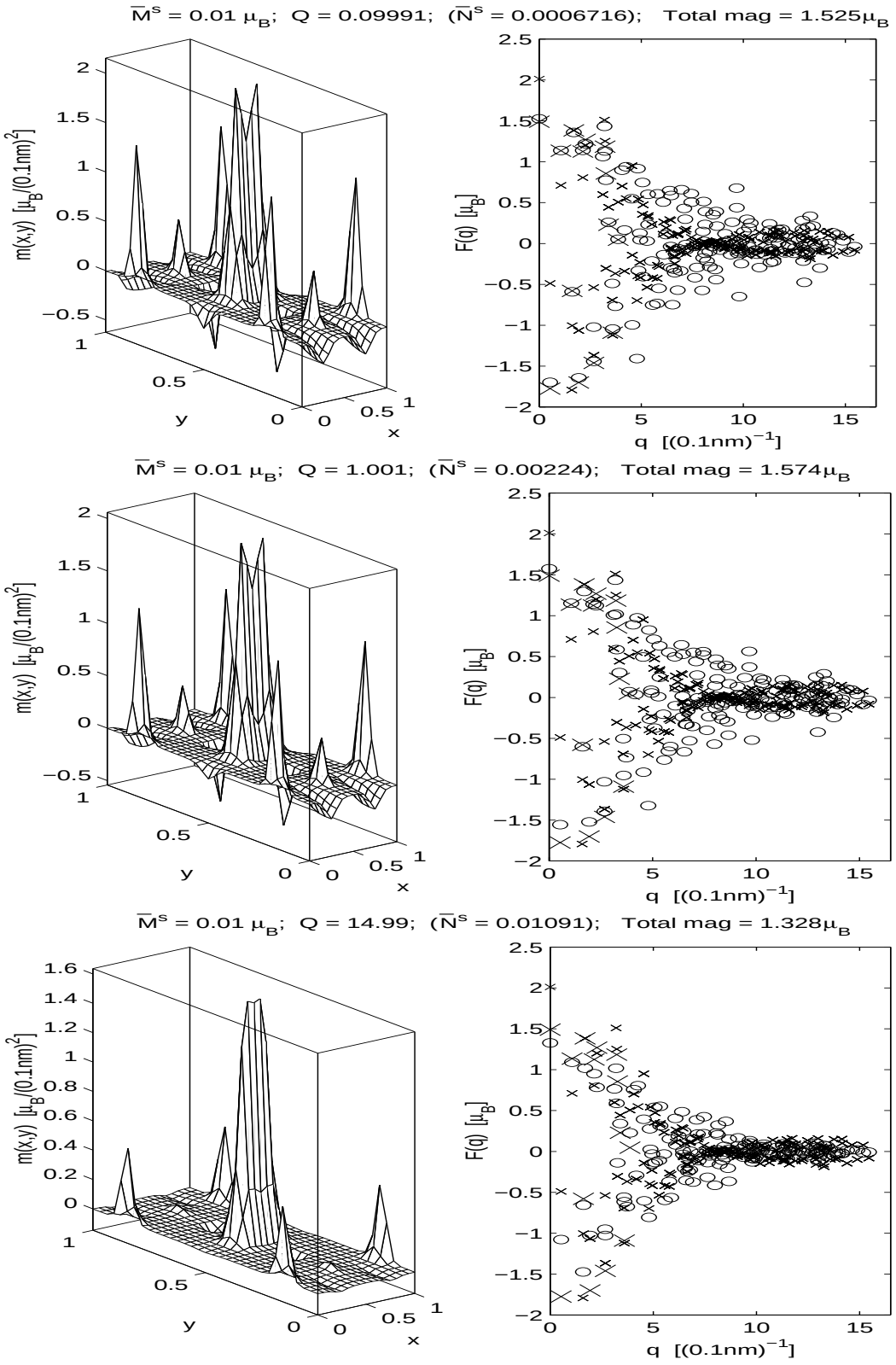

Figure 4.18: Same as in Fig. 4.11, but with the thesis data in columns 6 and 7 of table 4.2.

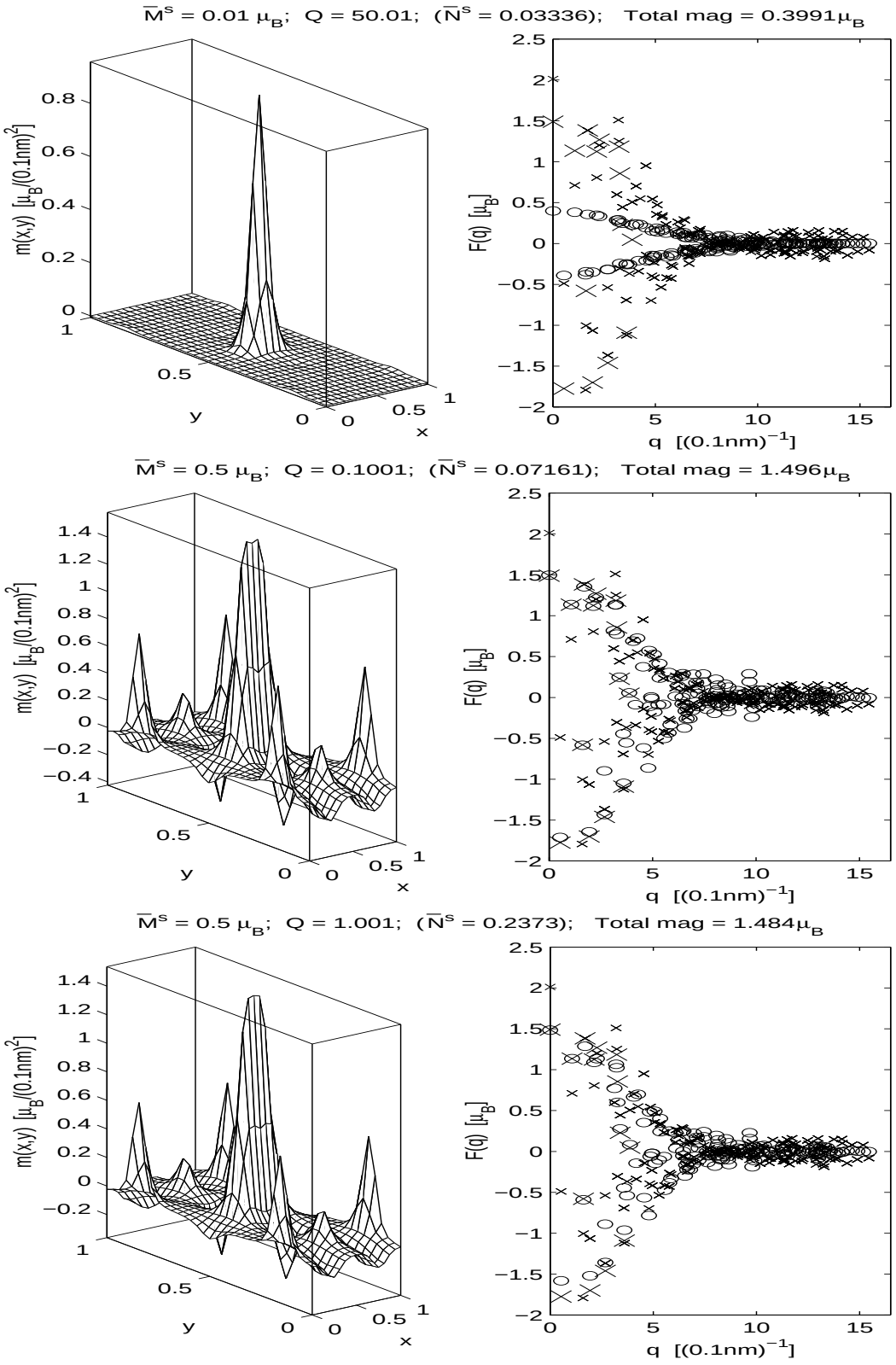

Figure 4.19: See caption Fig. 4.18.

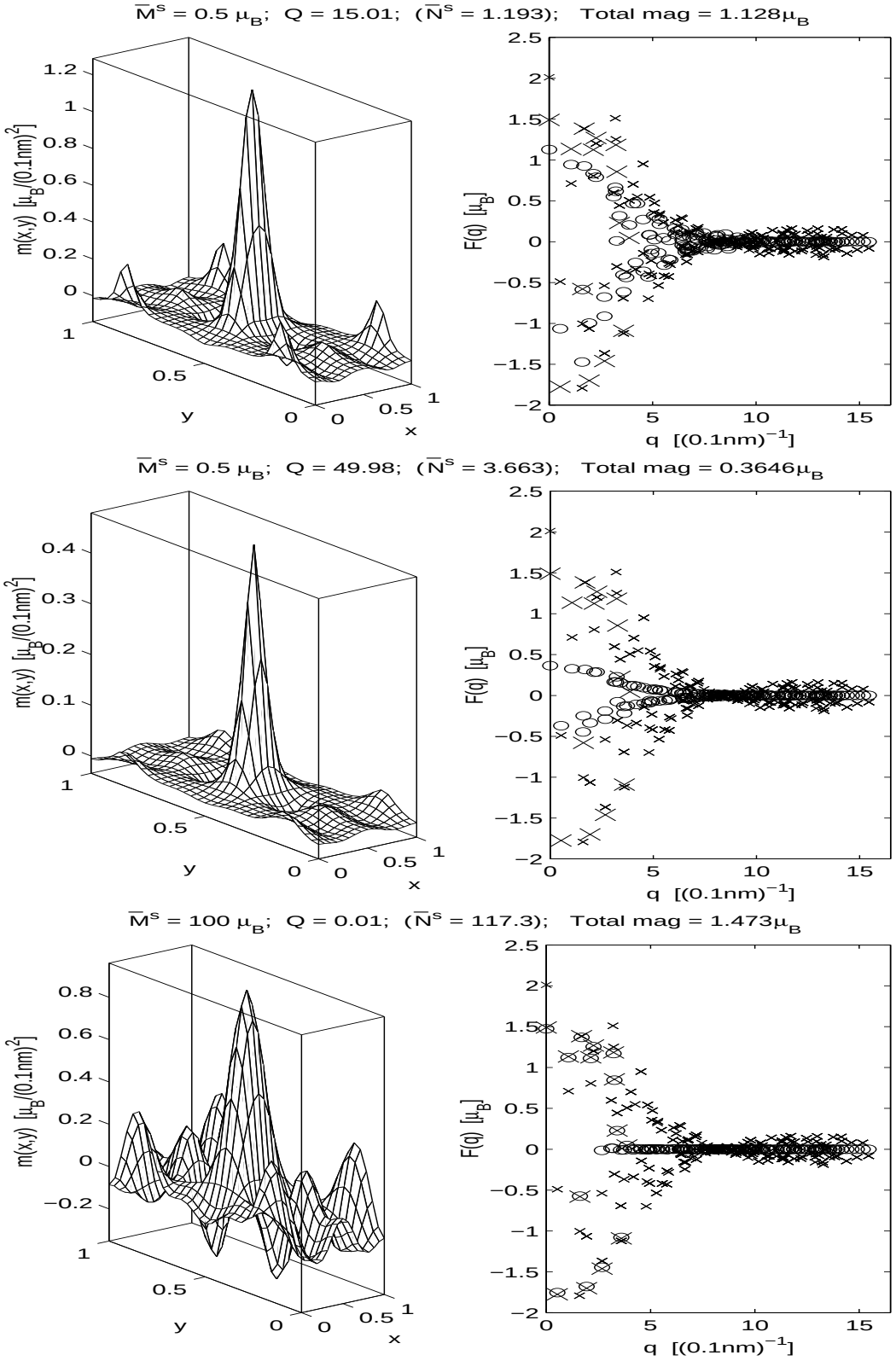

Figure 4.20: See caption Fig. 4.18.

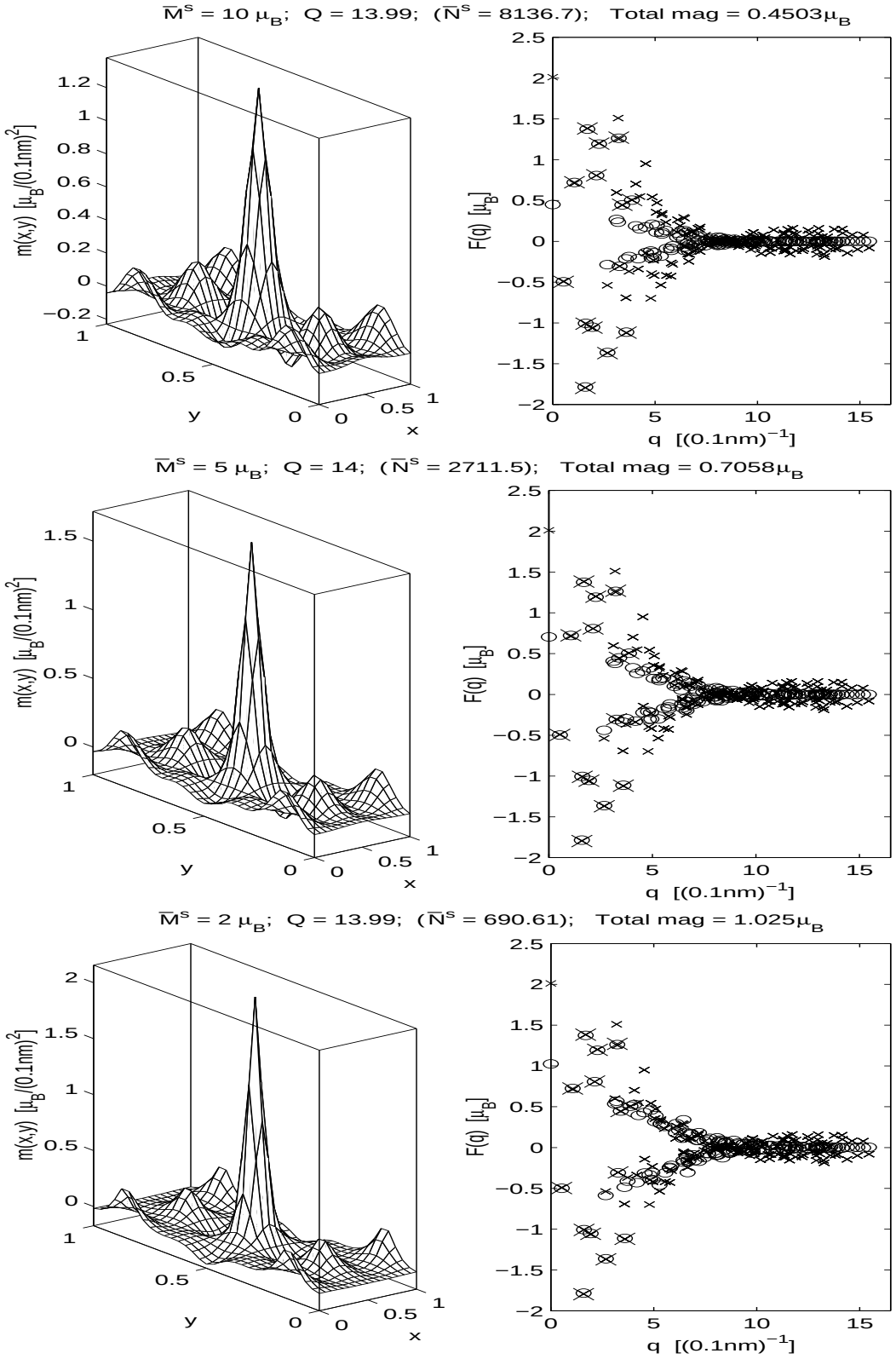

Figure 4.21: Same as in Fig. 4.11, but without the zero-q data point in table 4.2.

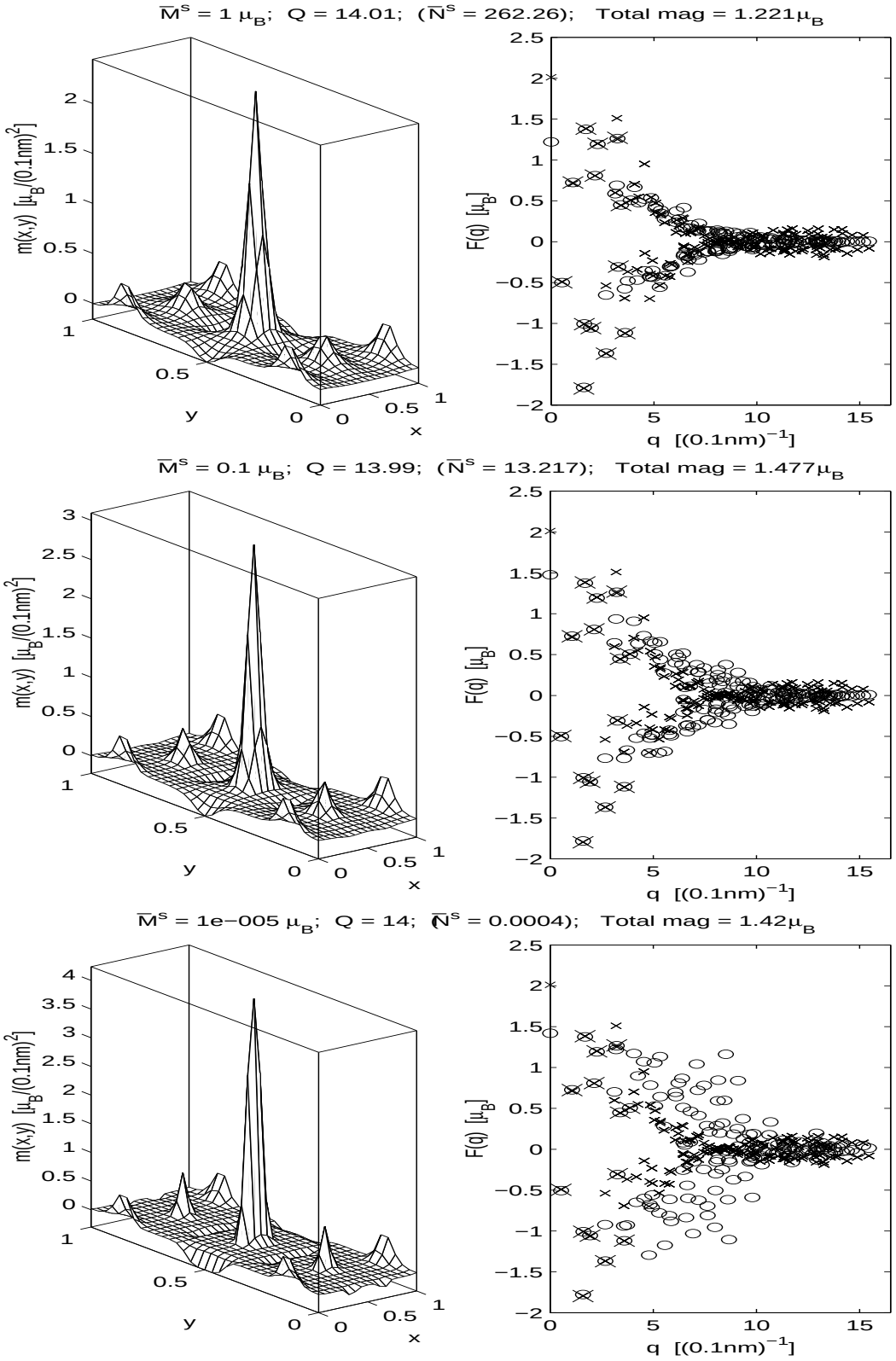

Figure 4.22: See caption Fig. 4.21.

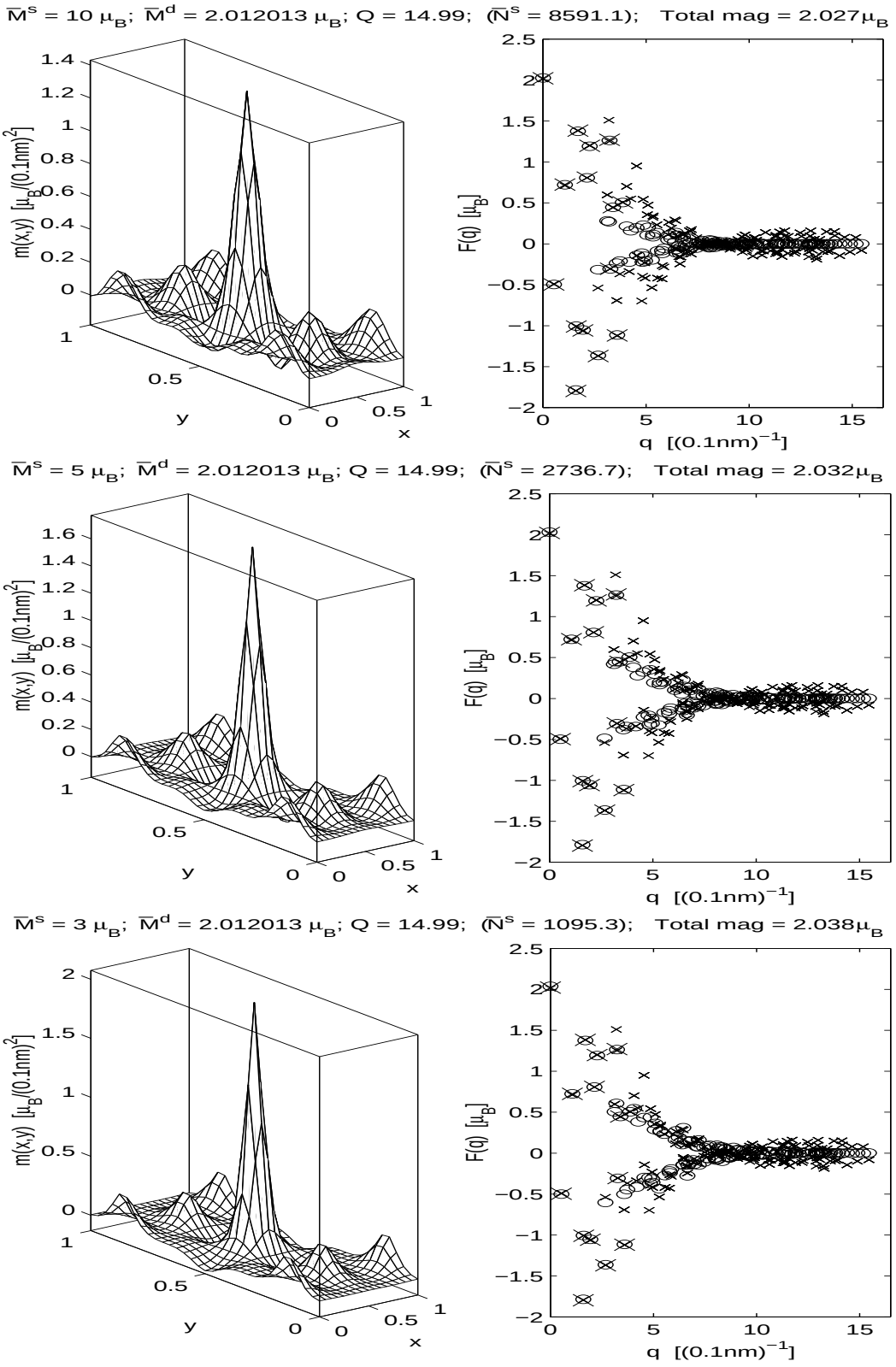

Figure 4.23: Same as in Fig. 4.11, but with $\bar{M}^d = 2.012013$.

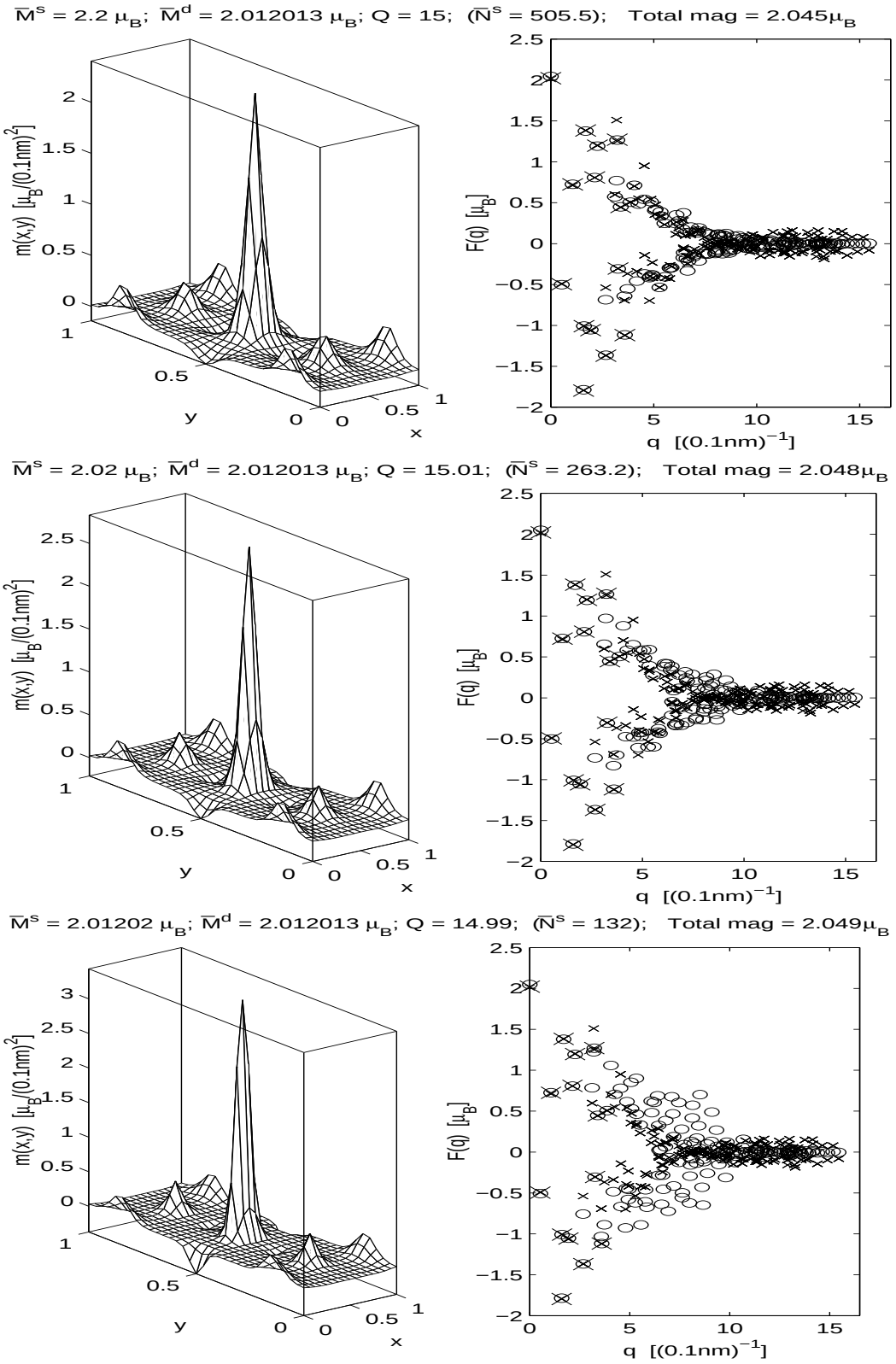

Figure 4.24: See caption Fig. 4.23.

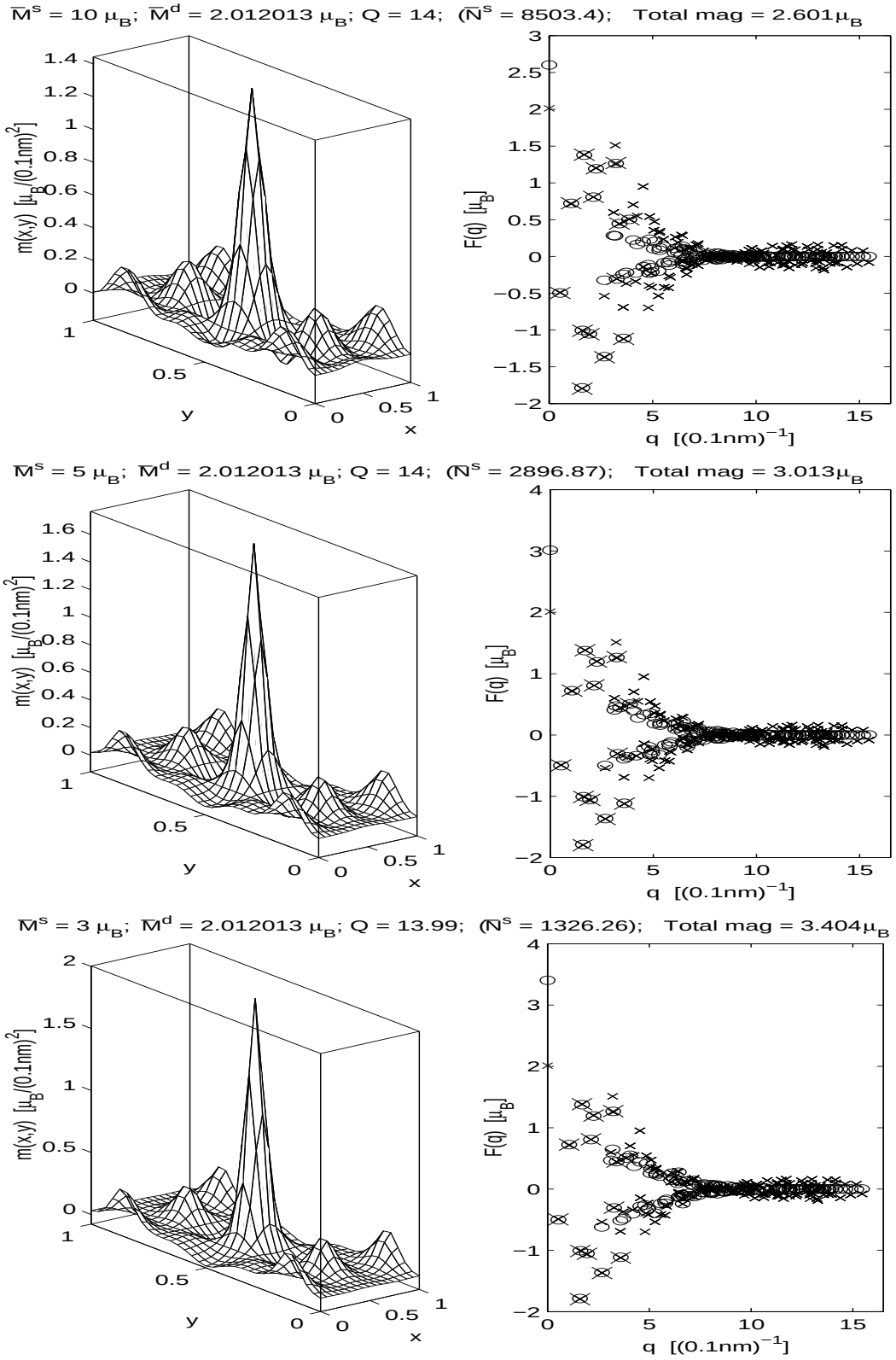

Figure 4.25: Same as in Fig. 4.11, but without the zero-q data point and with $\bar{M}^d = 2.012013$.

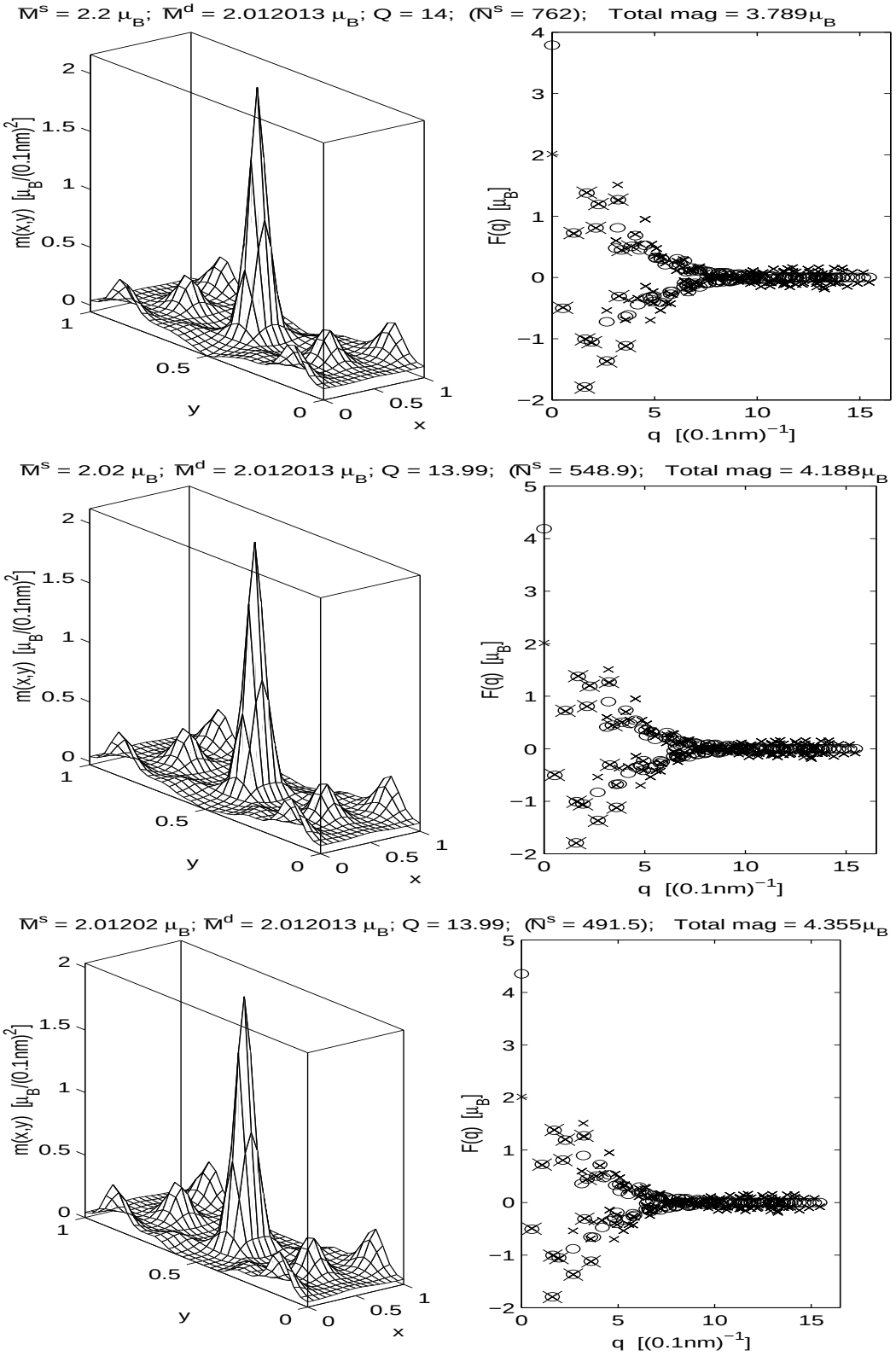

Figure 4.26: See caption Fig. 4.25.

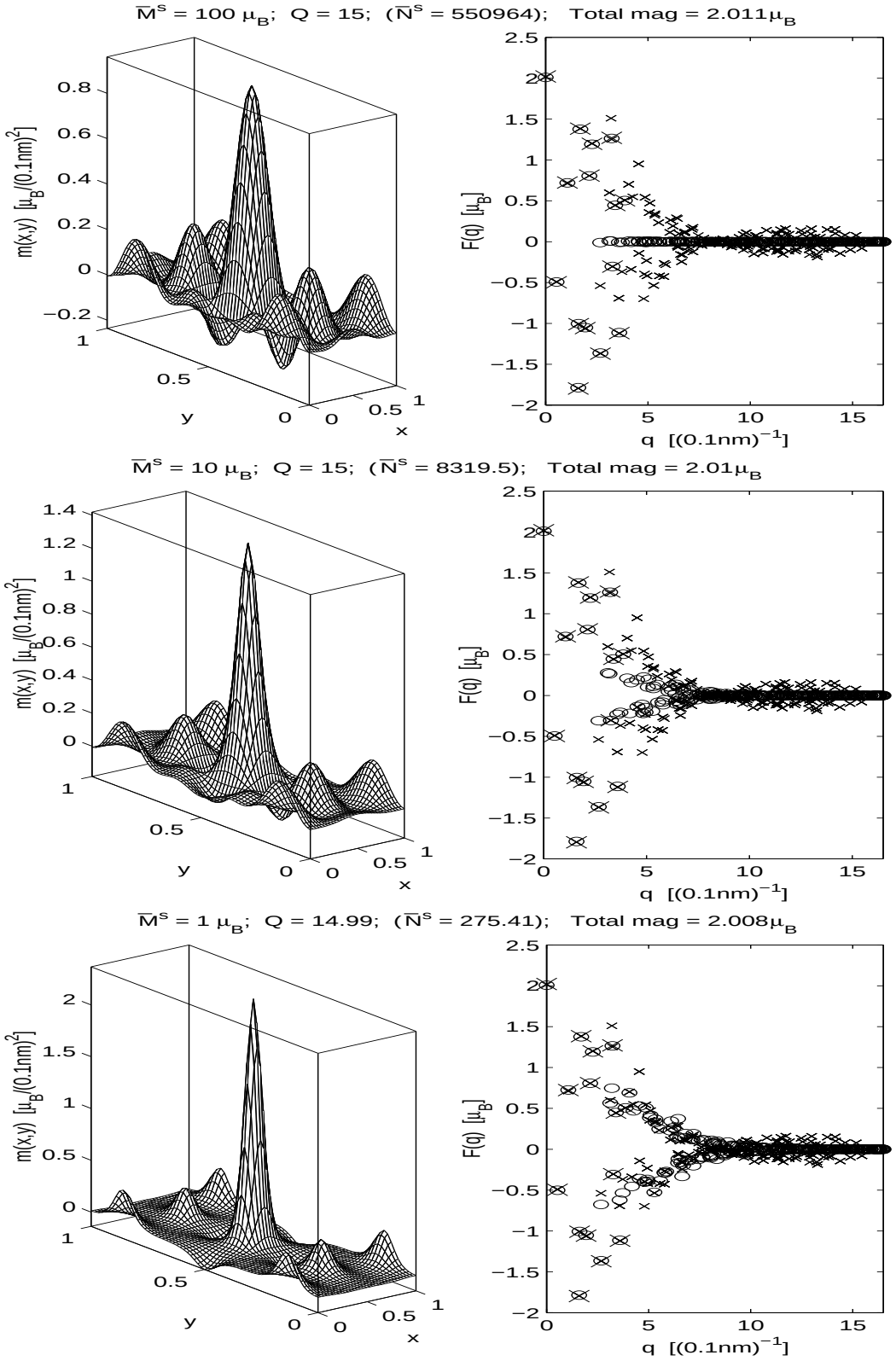

Figure 4.27: Same as in Fig. 4.11, but with the reconstructions made using a 32×64 image grid.

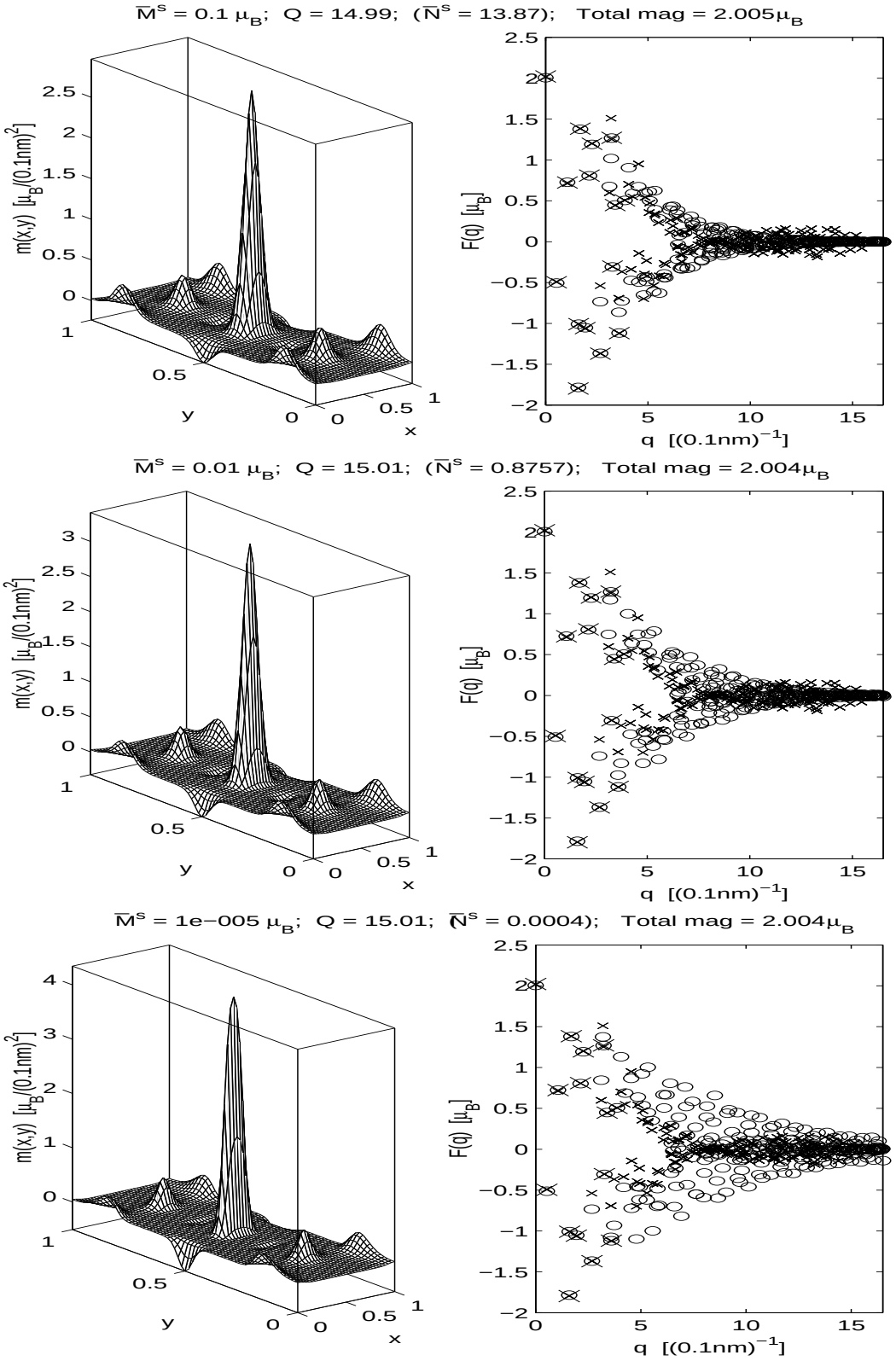

Figure 4.28: See caption Fig. 4.27.

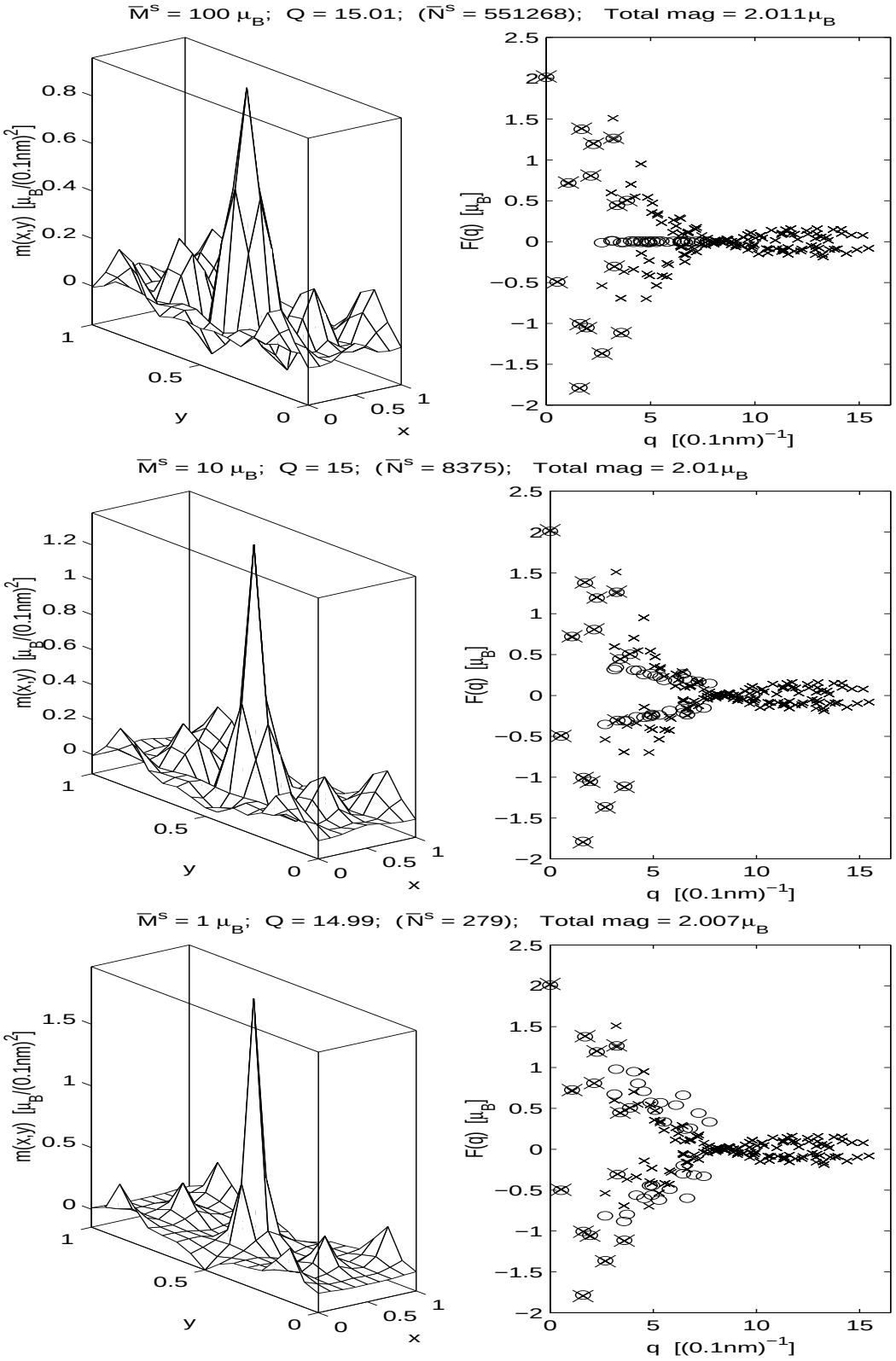

Figure 4.29: Same as in Fig. 4.11, but with the reconstructions made using a 8×16 image grid.

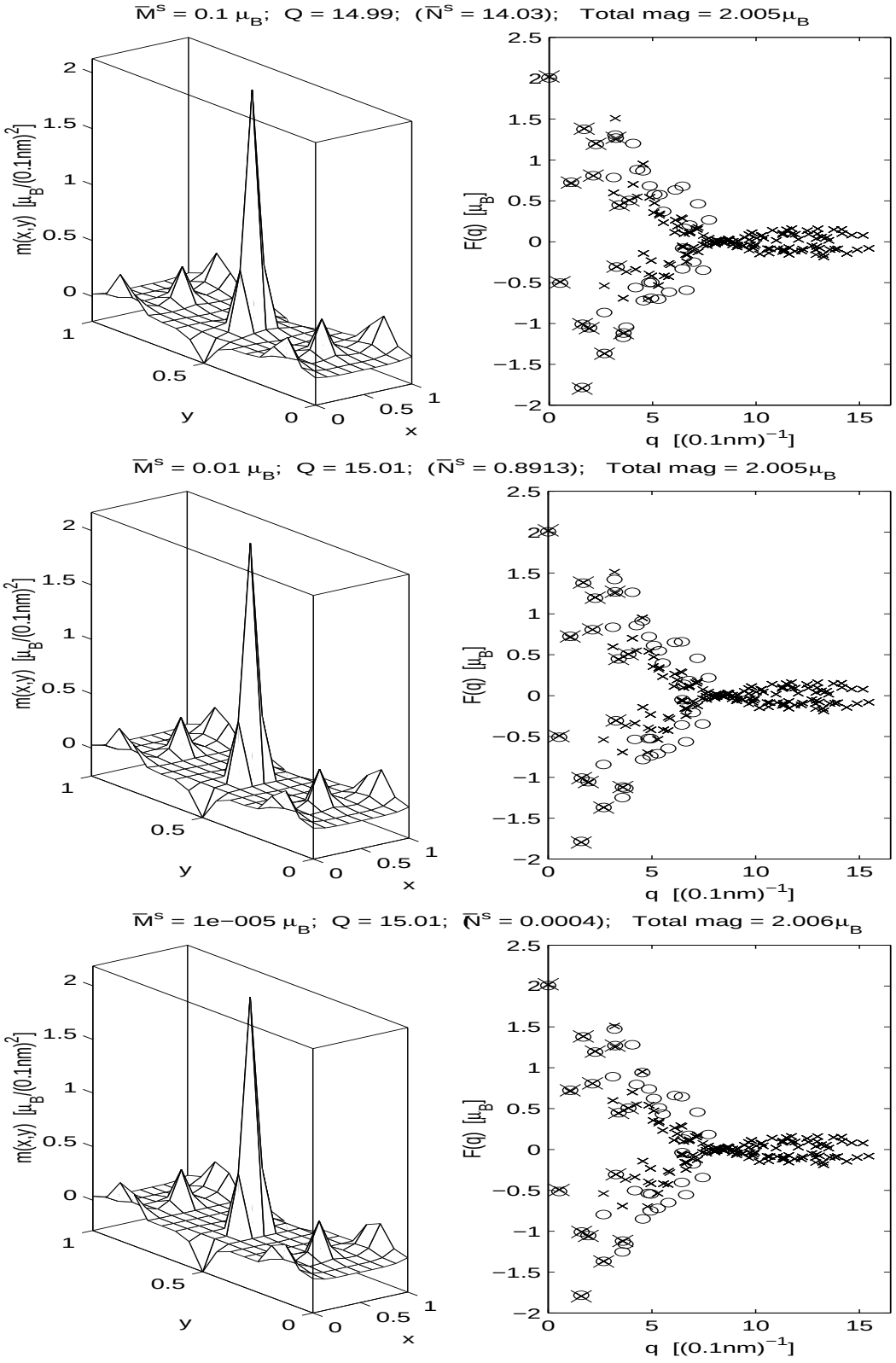

Figure 4.30: See caption Fig. 4.29.

Chapter 5

Polarised neutron diffraction data from $PrBa_2Cu_3O_{6+x}$

This chapter describes the analysis of a polarised neutron diffraction data set measured on a single crystal of $PrBa_2Cu_3O_{6.24}$. These data were discussed in the short paper [82]. Here, a more detailed discussion is presented using an improved data analysis technique for inverting the measured magnetic structure factors into magnetization density image, and additional theoretical calculations have been performed to help in the interpretation of the magnetization density images. The original purpose for investigating the magnetization density in $PrBa_2Cu_3O_{6+x}$ was to look for evidence of a hybridisation of the Pr $4f$ electrons with electronic states in the CuO_2 planes, as has been predicted by a number of theoretical works to cause the supression of superconductivity in $PrBa_2Cu_3O_{6+x}$. However, very little evidence is found which can support such an hybridisation. Of interest also, is that a number of other moments, with significant magnitude, are observed in the unit cell, which cannot be accounted for by existing theoretical work describing the suppression of superconductivity in $PrBa_2Cu_3O_{6+x}$. Such observed features may prove useful in helping theoreticians to suggest new theoretical models for $PrBa_2Cu_3O_{6+x}$.

5.1 Introduction to the $PrBa_2Cu_3O_{6+x}$ problem

It has been known for more than 10 years that the substitution of Pr for Y in $YBa_2Cu_3O_{6+x}$ causes a suppression of superconductivity in that compound [1]. Despite intensive research it is still not known what causes this effect, the main reason why it remains of strong interest is because of its relevance to the physics of high-T_c superconductivity. For a very recent review on $PrBa_2Cu_3O_{6+x}$ see Ref. [1]. Many aspects of $PrBa_2Cu_3O_{6+x}$ have been studied experimentially and many theoretical models have been proposed (see also the older review paper in Ref. [93]). For instance, the magnetic ordering of the Pr and Cu moments are now well understood experimentally, but still not theoretically [1]. The aspect of $PrBa_2Cu_3O_{6+x}$ which will be studied is its magnetic field induced magnetization density. The original purpose of this study was to make a comparison between the observed induced magnetization density and theoretical models which predict electric interactions between Pr and the superconducting CuO_2 planes.

A common theoretical model used for explaining suppression of superconductivity by Pr in $Pr_yY_{1-y}Ba_2Cu_3O_{6+x}$ involves the valence state of the Pr ions. Pr is situated at the centre of the unit cell in Fig. 5.1 and for superconducting isostructural compounds of $PrBa_2Cu_3O_{6+x}$ this site is normally occupied by ions with a 3+ oxidation state (like Y^{3+} or most other rare earths), but if Pr were more highly ionised then holes in the CuO_2 planes could be neutralised and the number of mobile charge carriers reduced below the level needed to sustain superconductivity. In this context the most influential work has been that of Fehrenbacher & Rice [94] who proposed a hybrid state containing stable Pr^{3+} and an intermediate valence Pr(IV) state made from a linear combination of Pr^{4+} and $Pr^{3+}\underline{L}$, where $\underline{L}$ denotes a ligand hole in the neighbouring oxygen $2p$ orbitals. Such a model involves changes in the electronic structure of the Pr ions as compared to that of Pr ions without interaction with the CuO_2 layers. Information on the spatial distribution of the 4f electrons of Pr would therefore be valuable to test models of

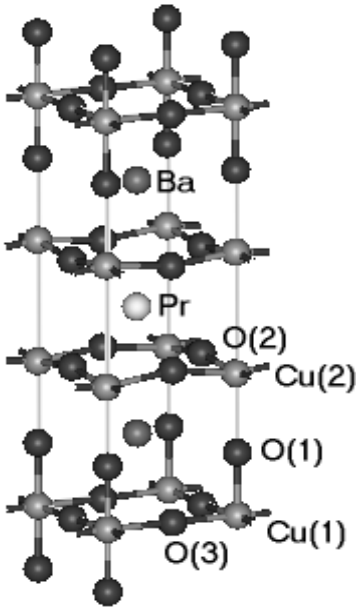

Figure 5.1: Shows the unit cell of $PrBa_2Cu_3O_{6+x}$ in its tetragonal phase. For the temperature parameters in table 5.1 it is assumed that the y-axis points in the O(3)-Cu(1) direction and the x-axis in the Cu(2)-O(2) direction.

this kind, this information can be obtained by polarised neutron diffraction. That was our original motivation for performing the PND measurements presented in this chapter.

5.2 Data collection for $PrBa_2Cu_3O_{6.24}$

A crystal of $PrBa_2Cu_3O_{6.24}$ with mass 23 mg was selected from the same batch as used in recent studies of the magnetic ordering[95]. To obtain a set of magnetic structure factors, unpolarised neutron diffraction and polarised neutron diffraction measurements were performed using the D9 and D3 instruments at Institut Laue-Langevin (ILL), Grenoble. The unpolarised measurements were done to refine the crystal structure and extinction parameters. The data was collected on a four-circle multidetector diffractometer, the detector being a gas-filled two-dimensional multiwire proportional counter. The crystal was mounted in an Air

Product two-stage Displex cryorefrigerator and the sample was cooled down to $T = 20K$. 989, 1234 and 697 reflections were scanned at wavelengths $0.8389\mathring{A}$, $0.70\mathring{A}$ and $0.54\mathring{A}$ respectively. Different wavelengths were used in order to estimate the amount of extinction. The three-dimensional arrays of counts observed around each scanned reflection were corrected for background and reduced to squared structure amplitudes using the three-dimensional integration method of Wilkinson et al. [96]. Contamination from neighbouring reflections, a potential problem at the shorter wavelenghts, was avoided by limiting the integration of each reflection to extend no further than half way to the neighbouring reflections. Averaging over repeated measurements and equivalent reflections gave 410, 648 and 412 reflections at the three wavelengths. After correcting for extinction and absorption a least squares refinement program (UPALS) was used to refine the nuclear structure against the data at all three wavelengths simultaneously[1]. The position, anisotropic temperature parameters and fractional occupancy of each site in the unit cell are listed in table 5.1.

The polarised neutron diffraction was also carried out at $T = 20K$, this is well above the antiferromagtic ordering temperature of the Pr sublattice ($T_N = 12K$ [95]). The setup of the polarised experiment is discussed in Chap. 1. The wavelength of the neutrons was $0.843\mathring{A}$ and an applied magnetic field $\mathbf{B}$ of 4.6 T was used over the sample. In order to access a large amount of reciprocal space, the experiment was carried out using two orientations of the crystal relative to the applied field: $\mathbf{B}$ parallel to [1 0 0], i.e. $\mathbf{B} \parallel$ [1 0 0], and $\mathbf{B}$ parallel to [1 1 0], i.e. $\mathbf{B} \parallel$ [1 1 0]. Because the crystal structure is tetragonal, the magnitude of the induced moment is expected to be parallel to the applied field and the same for both orientations on a macroscopic scale. With $\mathbf{B} \parallel$ [1 0 0] and $\mathbf{B} \parallel$ [1 1 0], 61 and 56 unique flipping ratios were measured respectively. To convert these flipping ratios into magnetic structure factors the CCSL program SORGAM was used.

[1]Many thanks to Garry McIntyre for doing this.

atom	position (z-component)	temperature parameters	site occupancy
Pr	0.5	$B_{11} = 0.0864(159)$	1
		$B_{33} = 0.1704(234)$	
Ba	0.19171(6)	$B_{11} = 0.3386(113)$	1
		$B_{22} = 0.4955(184)$	
Cu(1)	0	$B_{11} = 0.6158(128)$	1
		$B_{33} = 0.2695(161)$	
Cu(2)	0.35224(4)	$B_{11} = 0.1394(67)$	1
		$B_{33} = 0.3374(106)$	
O(1)	0.15288(7)	$B_{11} = 1.5016(170)$	1
		$B_{33} = 0.4888(167)$	
O(2)	0.36970(4)	$B_{11} = 0.2236(99)$	1
		$B_{22} = 0.3462(106)$	
		$B_{33} = 0.4983(100)$	
O(3)	0	$B_{11} = 0.876(190)$	0.1222(55)
		$B_{22} = 0.663(173)$	
		$B_{33} = 0.548(156)$	

Table 5.1: Refined fractional atomic coordinates, thermal parameters and site occupancies are listed. The anisotropic temperature factor has the form $\exp(-h^2B_{11}/4a^2 - k^2B_{22}/4b^2 - l^2B_{33}/4c^2)$, where the units of the B_{ii} are $\mathring{A}^2$. The space group of $PrBa_2Cu_3O_{6.24}$ is P4/mmm and the lattice parameters refined to $a = 3.8982(3)\mathring{A}$ and $c = 11.7976(10)\mathring{A}$. Residuals for the refinement were $R_w(F^2) = 7.7\%$, $R(F) = 3.6\%$ and goodness of fit = 2.52, see e.g. Ref [97],[98] for the formal definition of these quantities. The site occupations of Pr, Cu(1), Cu(2) and O(1) were also varied, but the values stayed within two esd's of the stoichiometric values. Notice that the displacement parameters of O(1) are surprisingly large. The parameters for O(3) are also large, but the site is only slightly occupied. The extinction parameter was 265.8 sec RMS mosaic.

SORGAM takes flipping ratios as input and structural information about the position of nuclei, extinction[2] and absorption parameters etc.. Structural information for $PrBa_2Cu_3O_{6.24}$ is listed in table 5.1. The resulting magnetic structure factors are listed in table 5.5 except for the 13 structure factors which have the same Miller indices for both **B** field orientations, these are listed separately in table 5.6. In table 5.6 an estimated 'average' value of each of these 13 Miller indices is listed in the last column, taken to be the arithmetic mean: $\bar{F}_{hkl} = (F_{hkl}^{B \parallel [1\ 0\ 0]} + F_{hkl}^{B \parallel [1\ 1\ 0]})/2$ and $\bar{\sigma}_{hkl} = (\sigma_{hkl}^{B \parallel [1\ 0\ 0]} + \sigma_{hkl}^{B \parallel [1\ 1\ 0]})/2$. For the two reflections with Miller indices $(0, 0, 11)$ and $(-1, 3, 0)$, $\bar{\sigma}_{hkl}$ is further increased to the value $\bar{\sigma}_{hkl} = |\bar{F}_{hkl} - F_{hkl}^{B \parallel [1\ 0\ 0]}|$. These 13 'averaged' structure factor values and standard deviations are estimated with the aim of accounting for possible differences in the induced moment distributions in the unit cell when applying **B** $\parallel$ [1 0 0] and **B** $\parallel$ [1 1 0]. In addition to the magnetic structure factors obtained from polarised neutron diffraction the zero-q structure factor (with Miller indices $(0, 0, 0)$), which is equal to the total moment in the unit cell, was determined from a bulk magnetisation (SQUID) measurement on the same crystal of $PrBa_2Cu_3O_{6.24}$ also at $T = 20K$ and $B = 4.6T$ [3]. The zero-q structure factor was measured to be $0.283 \pm 0.011\mu_B$. The complete set of unique magnetic structure factors for $PrBa_2Cu_3O_{6.24}$ will be analysed both as a whole and as two independent data sets.

5.3 Crystal field calculation and comparison with two data sets

The two independently measured PND data set for $PrBa_2Cu_3O_{6.24}$ with **B** $\parallel$ [1 0 0] and **B** $\parallel$ [1 1 0] are analysed separately in this section using the Bayesian image

[2]Many thanks to Bruce Forsyth (one of the authors of the CCSL suite) who helped with the insertion of the extinction parameter from the UPALS refinement into SORGAM.

[3]Many thanks to Steven Lister for doing this.

analysis method of Chap. 4, and the resulting magnetization density maps are compared. To help assess whether the density maps show evidence of an interaction between the Pr $4f$ electrons and electronic states in the CuO_2 planes a theoretical calculation is performed. The basis for the theoretical calculation is discussed in the next subsection, in subsection 5.3.2 the theoretical and experimental densities are compared.

5.3.1 Crystal field model for Pr^{3+}

The induced magnetization density in $PrBa_2Cu_3O_{6.24}$ will mainly originate from Pr. Pr is believed to be in a Pr^{3+} valence state (or in a mixture of Pr^{3+}/Pr^{4+} with the majority of Pr being in the 3+ valence state), Pr^{3+} is a magnetic ion with an open $4f$ electron shell containing two electrons. These two electrons may hybridise with electronic states of the CuO_2 planes and cause suppression of superconductivity as, for example, is claimed by the theoretical work in Ref. [94]. If this is true, such a hybridisation will result in a distortion of the Pr^{3+} $4f$ electron density from that expected for an isolated Pr^{3+} ion under the influence of the local crystal field. To assist such an investigation, the magnetization density of an isolated Pr^{3+} ion in the local crystal field of $PrBa_2Cu_3O_{6.24}$ is calculated.

The magnetization density of Pr^{3+} is calculated by evaluating the quantum expectation value of magnetization density operators with respect to a specified wave function. Thus, the first step is to determine the wave function for Pr^{3+} in $PrBa_2Cu_3O_{6.24}$ at $T = 20K$, with an applied field of $B = 4.6T$, from a model taking into account the local crystal field. Pr is a rare earth and is well described within the intermediate coupling angular momentum scheme because the mean radius of the $4f$ electrons is small and the dominating interaction is the Coulomb repulsion between the $4f$ electrons. With reference to a carefully measured experiment [99], an appropriate theoretical model for the crystal field which includes all $^{2S+1}L_J$ terms of the f^2 configuration of Pr^{3+} was deduced. The magnetic

	$\mathbf{B} \parallel [1\ 0\ 0]$		$B \parallel [1\ 1\ 0]$	
M	$\|0\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
-4	-0.002	0.003	$0.002 - 0.002i$	$-0.002 - 0.002i$
-3	0.645	-0.663	$0.645i$	0.663
-2	-0.163	0.002	$-0.114 - 0.114i$	$-0.001 + 0.001i$
-1	0.205	0.211	0.205	$0.210i$
0	-0.002	0.0	0.0	$-0.001 - 0.001i$
1	0.205	-0.211	$0.205i$	0.210
2	-0.163	-0.002	$-0.114 - 0.114i$	$0.001 - 0.001i$
3	0.645	0.663	0.645	$0.663i$
4	-0.002	-0.003	$-0.002 + 0.002i$	$-0.002 - 0.002i$

Table 5.2: $|0\rangle$ is the calculated ground state and $|1\rangle$ is the first excited state $0.12meV$ above the ground for both $\mathbf{B}$ field directions. At $T = 20K$ it is assumed that $|0\rangle$ and $|1\rangle$ are equally populated. The complete wave functions contain $\sim 10\%$ admixture of higher $^{2S+1}L_J$ terms, but only the components of the $^3H_4$ term are listed. The wave functions were calculated using a computer program written by Andrew Boothroyd and described in Ref. [99].

properties are found to be dominated by the ground state and 1^{st} excited state. The wave functions of these states are given in table 5.2 with $\mathbf{B} \parallel [1\ 0\ 0]$ and $\mathbf{B} \parallel [1\ 1\ 0]$.

Once the wave function of Pr^{3+} is specified the magnetization density is obtained by calculating the expected value of appropriate magnetization density operators with respect to this wave function. Take the magnetization density spin and orbital operators to be [100]

$$\mathbf{M}^S(\mathbf{r}) = -2\mu_B \sum_{i=1}^{n} \mathbf{s}_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{5.1}$$

and

$$\mathbf{M}^L(\mathbf{r}) = (1/[cr]) \int_r^\infty z\hat{\mathbf{r}} \times \mathbf{j}^L(z\hat{\mathbf{r}}) dz \quad , \tag{5.2}$$

where

$$\mathbf{j}^L(\mathbf{r}) = -(\mu_B c/\hbar)\sum_{i=1}^{n}[\mathbf{p}_i\delta(\mathbf{r}-\mathbf{r}_i)+\delta(\mathbf{r}-\mathbf{r}_i)\mathbf{p}_i] \quad . \tag{5.3}$$

$\mathbf{r}_i$, $\mathbf{p}_i$ and $\mathbf{s}_i$ are the position, momentum and spin of the i^{th} magnetic electron respectively, and $\hat{\mathbf{r}}$ is the unit vector of $\mathbf{r}$. With the operators in Eqs. (5.1-5.2) the relevant matrix elements can be written as [100]

$$\begin{aligned}
&\langle SLJM|M_q^S(\mathbf{r})|S'L'J'M'\rangle = -\frac{\mu_B}{\pi}\sum_{K,Q,K',Q'} Y_Q^K(\hat{\mathbf{r}})U^2(r)(-1)^{\frac{1}{2}+M+q-J'+L'+S'} \\
\times\ & \sqrt{3/2}\,[l][K']\sqrt{[K][S][S'][L][L'][J][J']}\; n\sum_{\bar{\theta}}(\theta\{|\bar{\theta})(\theta'\{|\bar{\theta})(-1)^{\bar{S}+\bar{L}} \\
\times\ & \left\{\begin{array}{ccc} S & 1 & S' \\ \frac{1}{2} & \bar{S} & \frac{1}{2}\end{array}\right\}\left\{\begin{array}{ccc} L & K & L' \\ l & \bar{L} & l\end{array}\right\}\left\{\begin{array}{ccc} 1 & K & K' \\ S' & L' & J' \\ S & L & J\end{array}\right\} \\
\times\ & \left(\begin{array}{ccc} J & K' & J' \\ -M & Q' & M'\end{array}\right)\left(\begin{array}{ccc} K & K' & 1 \\ Q & Q' & -q\end{array}\right)\left(\begin{array}{ccc} l & K & l \\ 0 & 0 & 0\end{array}\right)
\end{aligned} \tag{5.4}$$

and

$$\begin{aligned}
&\langle SLJM|M_q^L(\mathbf{r})|S'L'J'M'\rangle = -\frac{\mu_B}{\pi}\sum_{K,Q,K',Q'} Y_Q^K(\hat{\mathbf{r}})\frac{1}{r}\int_r^\infty U^2(z)dz \\
\times\ & (-1)^{q+M+L+L'+S}\delta_{SS'}[l]^{3/2}\sqrt{[J][J'][L][L'][K]l(l+1)}[K'] \\
\times\ & \left(\begin{array}{ccc} l & K & l \\ 0 & 0 & 0\end{array}\right)\left\{\begin{array}{ccc} l & K' & l \\ K & l & 1\end{array}\right\}\left\{\begin{array}{ccc} K' & L' & L \\ S & J & J'\end{array}\right\} n\sum_{\bar{\theta}}(\theta\{|\bar{\theta})(\theta'\{|\bar{\theta})(-1)^{\bar{L}} \\
\times\ & \left\{\begin{array}{ccc} K' & l & l \\ \bar{L} & L & L'\end{array}\right\}\left(\begin{array}{ccc} J & K' & J' \\ -M & Q' & M'\end{array}\right)\left(\begin{array}{ccc} K & K' & 1 \\ Q & Q' & -q\end{array}\right) \quad .
\end{aligned} \tag{5.5}$$

The notation used in Eqs. (5.4-5.5) originates from atomic physics [101]. The 2×3 arrays with round and curly brackets are named $3j$ and $6j$ symbols respectively, and the 3×3 array in Eq. (5.4) is a $9j$ symbol. θ, θ' and $\bar{\theta}$ are shorthand for SL, $S'L'$ and $\bar{S}\bar{L}$ respectively, where θ and θ' are possible LS coupling terms (or Russell-Saunders terms) which can be formed from the l^n electronic configuration and $\bar{\theta}$ possible coupling terms formed from the l^{n-1} configuration (called parent terms); l is the orbital quantum number of the open ionic shell and n is the number of electrons in that shell. As an example, for Pr^{3+}, which has two $4f$

electrons, then $l = 3$ and $n = 2$ and since only one LS coupling term can be formed from a one electron configuration ($l^{n-1} = 3^1$), Pr^{3+} has only one parent term with $\bar{S} = 1/2$ and $\bar{L} = 3$ (in spectroscopic notation ^{2}F). $(\theta\{|\bar{\theta})$ is a so-called coefficient of fractional percentage, see for instance Chap. 5, Sec. 6 of Ref. [101] for how to calculate such fractional percentage in general. For Pr^{3+}, which has only one parent term, the fractional percentage is simply $(\theta\{|\bar{\theta}) = 1$. Y_Q^K is the well-known spherical harmomic of rank K and order Q, and $U(r)$ is the radial wave function for an electron of the l^n configuration, assumed to be the same for all n electrons. For the calculation of the magnetization denisty of Pr^{3+} the radial wave function in Eq. (4.31), Sec., Chap. 4 is used. The total magnetization density at any point in space is clearly the sum of the spin and orbital magnetization density, i.e. $M_q = M_q^S + M_q^L$. The sub-index q of M_q^S and M_q^L refers to the spherical components and $q = 1, 0, -1$. These relate to cartesian vector components as follows $M_x = \frac{1}{\sqrt{2}}(M_{-1} - M_1)$, $M_y = \frac{1}{i\sqrt{2}}((M_{-1} + M_1)$ and $M_z = M_0$.

The expressions in Eqs. (5.4-5.5) may look complicated, but are relatively straight forward to derive, although this is somewhat painstaking to do. A number of authors have contributed to obtaining matrix expressions of this type. In this context, a theory suitable for relativistic calculations can be found in Ref. [102], and a general summary and introduction to these works can be found in Chap. 11 in Ref. [4].

Andrew Boothroyd and I have independently written computer programs for calculating the matrix elements $\langle\psi|M_i|\psi\rangle$, $i = x, y, z$, where $|\psi\rangle$ is a wave function from table 5.2. The calculation of the densities was carried out on a $65 \times 65 \times 65$ grid of a cubic box with length a (a being the lattice parameter a for $PrBa_2Cu_3O_{6.24}$). The grid point (n_x, n_y, n_z), where n_x, n_y, n_z are any integers $0, 1 \ldots 64$, is the center of a bin at $(n_x, n_y, n_z)\frac{a}{64}$ stretching from $\frac{n_i a}{64} - \frac{a}{2*64}$ to $\frac{n_i a}{64} + \frac{a}{2*64}$ along $i = x, y, z$. To make the comparison with experimental results easier the center of the Pr^{3+} density is positioned at $(\frac{a}{2}, \frac{a}{2}, \frac{a}{2})$ within the cubic

box. Since the orbital magnetization density diverges at $r = 0$, see Eq. (5.5), then special care need to be taken for obtaining a number representing the average density for the bin centered at $(\frac{a}{2}, \frac{a}{2}, \frac{a}{2})$ (in Eq. (5.5) at $r = 0$). For this purpose, notice that when integrating the spherical components of $\langle SLJM|\mathbf{M}(\mathbf{r})|SLJM'\rangle$ over a sphere with radius r_0 we obtain

$$\int_0^{r_0}\int_0^{\pi}\int_0^{2\pi} \langle SLJM|M_q(\mathbf{r})|SLJM'\rangle r^2 dr d\Omega \tag{5.6}$$
$$= \begin{cases} -\mu_B \tilde{g} M \delta_{MM'} & \text{for } q = 0 \\ -\mu_B \tilde{g}(-q)\frac{1}{\sqrt{2}}[(J - qM')(J + qM' + 1)]^{1/2}\,\delta_{M(M'+q)} & \text{for } q = -1, 1 \end{cases},$$

where

$$\tilde{g} = \int_0^{r_0} \left[r^2U^2(r) + f(r) + (r^2U^2(r) - f(r))[S(S+1) - L(L+1)]/(J(J+1))\right] dr\,, \tag{5.7}$$

and $f(r) = r\int_r^{\infty} U^2(z)dz$. Hence, an average density for the bin centered at $(\frac{a}{2}, \frac{a}{2}, \frac{a}{2})$ is obtained separately using the expression in Eq. (5.6) with $r_0 = \frac{a}{2*64}$ (and divided by $(\frac{a}{64})^3$ to obtain a density). Notice, with $r_0 = \infty$ then $\tilde{g}$ in Eq. (5.7) reduces to the expression for the well known Landé splitting factor for a paramagnetic ion[4].

Consider first the wave function for Pr^{3+} calculated with the applied field $\mathbf{B} \parallel [1\ 0\ 0]$ in table 5.2. The component of $\langle \mathbf{M}(\mathbf{r})\rangle_{[1\ 0\ 0]}$ [5]pointing in the [1 0 0] direction[6] is drawn as two isosurfaces in Fig. 5.2(e-f). What is suggested perhaps unexpectedly, by the theoretical calculation is that the tetragonal crystal symmetry of $PrBa_2Cu_3O_{6.24}$ is broken by an applied $\mathbf{B}$ field of $4.6T$. This is interesting because it stresses that the induced moment distribution in $PrBa_2Cu_3O_{6.24}$ with

[4]Use $\int_0^{\infty} r^2U^2(r)dr = 1$ and $\int_0^{\infty} f(r)dr = \frac{1}{2}$ then $\tilde{g}(\infty) = \frac{3}{2} + \frac{1}{2}[S(S+1) - L(L+1)]/(J(J+1))$. $U(r)$ is the radial part of the orbital function. In atomic units the radial part of the density ρ in Eq. (4.31) is related to U as $\rho = U^2$.

[5]where $\langle \mathbf{M}(\mathbf{r})\rangle$ is the thermal average $\langle \mathbf{M}(\mathbf{r})\rangle = \frac{\langle 0|\mathbf{M}|0\rangle + \langle 1|\mathbf{M}|1\rangle}{2}$ due to $|0\rangle$ and $|1\rangle$ being equally populated, see caption table 5.2.

[6]The other two components have integrated moments of zero.

different applied field directions may be slightly or even considerably different. It is for this reason that the data measured for $PrBa_2Cu_3O_{6.24}$ are analysed as two separate data sets here, one data set for each **B**-field orientation, before being analysed as an 'averaged' data set. Fig. 5.2(a-d) show the projection of $\langle M_x(\mathbf{r})\rangle_{[1\ 0\ 0]}$ onto the plane $[0\ 1\ 0]-[0\ 0\ 1]$ in Sec. 5.4. It is the spatial distribution of the magnetization density perpendicular to the applied field that can be observed most directly by polarised neutron diffraction, see Chap. 1.

In Fig. 5.3(e-f) the component of $\langle \mathbf{M}(\mathbf{r})\rangle_{[1\ 1\ 0]}$ parallel to [1 1 0] is shown as two isosurfaces, where the thermal average $\langle \mathbf{M}(\mathbf{r})\rangle_{[1\ 1\ 0]}$ is calculated using the wave functions given in table 5.2 (see also footnote 5). As for $\langle M_x(\mathbf{r})\rangle_{[1\ 0\ 0]}$ the theoretical calculation predicts that the applied **B** field breaks the tetragonal crystal symmetry. Otherwise, apart from being rotated by 45°, the calculated magnetization densities are quite similar, in particular when comparing the projected densities onto the plane perpendicular to the field direction.

5.3.2 Comparison with data

From the data measured with the applied field $\mathbf{B} \parallel [1\ 0\ 0]$ consider the structure factors in the $[0\ 1\ 0]-[0\ 0\ 1]$ plane. There are 52 of these in tables 5.5-5.6 and by adding the zero-q data point a data set consisting of 53 structure factors is generated. A reconstruction of this data set with[7] $\bar{M}^s = 0.1\mu_B$ ($Q = P = 53$) on a 64×256 grid is shown in Fig. 5.4 for four different cut-off values. a and c are the lattice parameters along [0 1 0] and [0 0 1] and the plane crystal symmetry is mm (same symmetry as used for the thesis model in Sec. 4.4.2, Chap. 4). A cut-off value means the value above which all magnetization density values are set equal to this value. Using cut-off values helps emphasize structures at different magnetization density levels in Fig. 5.4. Separately, in Fig. 5.6(a) a contour plot of the magnetization density without cut-off for a part of the unit

[7] The $\bar{M}^s$ value is obtained using the recipe in Sec. 4.7, Chap. 4, see also Fig. 5.9.

cell centered around the Pr site is shown. From this figure it is clear that at the highest contour levels the magnetization density of Pr is spherically distributed. A similar structure is observed from the crystal field calculation in Fig. 5.2(a) and comparing with Fig. 5.6(a) the only difference is a depletion of density at the center of the Pr density that is not observed experimentally. However, even if this feature was present it is probably not detected with the data in tables 5.5-5.6 because of its small spatial dimensions. Its spherical diameter is close to $\frac{a}{64}$, to reveal a structure of this scale it would require data with Miller indices of the order $(32, 32, 96)$, this is beyond the data in tables 5.5-5.6 [8].

By comparing Fig. 5.2(b) with Fig. 5.4(a), Fig. 5.2(c) with Fig. 5.4(b) and Fig. 5.2(d) with Fig. 5.4(c) a good comparison is seen between the shape and diameter of the calculated and experimentally obtained Pr density. This suggests that the theoretical model is adequate for describing the observed Pr density for density values ranging from the maximum density value to a factor of 100 or more below that value; only at this point does the observed Pr density start to show substantial derivations from the calculated density.

To obtain an idea of the significance of the individual features in Fig. 5.4(a-d), error estimates (standard derivations) are calculated using the covariance procedure outlined in Sec. 4.6, Chap. 4. These are listed in table 5.3 for the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction. The feature listed first represents an integration of the corner densities, presumably originating from an induced moment on Cu(1). It is found to be of size $0.0109 \pm 0.0028 \mu_B$. Such a moment is small relative to the induced moment of the Pr ion, which is found to be $0.2492 \pm 0.0028 \mu_B$, but significant within the uncertainty of the data. Its slightly odd shape suggests that

[8] As a rule of thumb, a Miller index h may give structural information on interval length $\frac{a}{2h}$. This rule of thumb is a simplified version of the well-known sampling theorem in Fourier analysis (see e.g. [103] page 230) which states that any band width limited function where for all $|h| > h_{max}$, $F_h = 0$ then the function in real space is fully defined by sampling at points with the interval equal to $\frac{a}{2h}$.

description of feature	position [Å]	area of integration [0 1 0] [a]	[0 0 1] [c]	multi-plicity	moment ± s.d. [μ_B]
Cu(1)	$(0, 0)$	$-\frac{13}{64} : \frac{13}{64}$	$-\frac{29}{256} : \frac{29}{256}$	1	0.0109 ± 0.0028
Pr	$(\frac{a}{2}, \frac{c}{2})$	$\frac{17}{64} : \frac{47}{64}$	$\frac{107}{256} : \frac{149}{256}$	1	0.2492 ± 0.0028
Pr on Ba site	$(\frac{a}{2}, 2.5)$	$\frac{21}{64} : \frac{43}{64}$	$\frac{43}{256} : \frac{73}{256}$	2	0.0164 ± 0.0032
probably Pr (or O(2))	$(\frac{a}{2}, 4.1)$	$\frac{25}{64} : \frac{39}{64}$	$\frac{77}{256} : \frac{97}{256}$	2	0.0021 ± 0.0014
negative feature	$(\frac{a}{2}, 0)$	$\frac{19}{64} : \frac{45}{64}$	$-\frac{19}{256} : \frac{19}{256}$	1	-0.0103 ± 0.0027
negative feature	$(0.5, 2.0)$	$\frac{1}{64} : \frac{21}{64}$	$\frac{33}{256} : \frac{57}{256}$	4	-0.0034 ± 0.0014
left/right of Pr	$(0, \frac{c}{2})$	$-\frac{9}{64} : \frac{9}{64}$	$\frac{111}{256} : \frac{145}{256}$	1	0.0036 ± 0.0020

Table 5.3: Summed moments over bins of the [0 1 0] − [0 0 1] reconstruction in Fig. 5.4. The area of integration defines the area which is summed over, and a and c are the refined lattice constants for $PrBa_2Cu_3O_{6.24}$ (see table 5.1). The multiplicity, calculated moment and standard deviation in the reconstruction of each integrated feature is listed in columns 4 and 5.

description of feature	position [Å]	area of integration [−1 1 0] [$\frac{a}{\sqrt{2}}$]	[0 0 1] [c]	multi-plicity	moment ± s.d. [μ_B]
Pr	$(\frac{a}{\sqrt{2}}, \frac{c}{2})$	$\frac{15}{64} : \frac{49}{64}$	$\frac{107}{256} : \frac{149}{256}$	1	0.2530 ± 0.0030
Pr on Ba site	$(\frac{a}{\sqrt{2}}, 2.5)$	$\frac{19}{64} : \frac{45}{64}$	$\frac{41}{256} : \frac{69}{256}$	2	0.0109 ± 0.0025
probably Pr (or Cu(2))	$(\frac{a}{\sqrt{2}}, 3.8)$	$\frac{23}{64} : \frac{41}{64}$	$\frac{73}{256} : \frac{87}{256}$	2	0.0020 ± 0.0013
negative feature	$(\frac{a}{2\sqrt{2}}, 2.1)$	$-\frac{13}{64} : \frac{13}{64}$	$\frac{37}{256} : \frac{55}{256}$	2	-0.0030 ± 0.0016
negative feature	$(\frac{a}{2\sqrt{2}}, 5.1)$	$-\frac{9}{64} : \frac{9}{64}$	$\frac{103}{256} : \frac{121}{256}$	2	-0.0012 ± 0.0012
left/right of Pr	$(\frac{a}{2\sqrt{2}}, \frac{c}{2})$	$-\frac{5}{64} : \frac{5}{64}$	$\frac{125}{256} : \frac{131}{256}$	1	0.0003 ± 0.0008
'deformation' of Pr	$(3.4, 4.8)$	$\frac{41}{64} : \frac{51}{64}$	$\frac{99}{256} : \frac{107}{256}$	4	0.0001 ± 0.0004

Table 5.4: Summed moments over bins of the [−1 1 0] − [0 0 1] reconstruction in Fig. 5.5

there may be an induced moment other than the Cu(1) moment contributing to this feature. This Cu(1) moment may be compared to induced moments for the same site in isostructural $YBa_2Cu_3O_{6+x}$. A number of PND experiments have been carried out on $YBa_2Cu_3O_{6+x}$ for a variety of different oxygen compositions, applied **B** fields and temperatures [74],[104],[105],[106],[107],[108]. However, a direct comparison to these works is difficult because the induced magnetic moments in $YBa_2Cu_3O_{6+x}$ show a strong temperature dependence [74] and variation with oxygen composition[9]. One notable difference, though, is that the Cu(1) moment is between a factor of 2 to 10 times larger here, than in these references.

The largest feature not situated at the Pr site has its center at $\sim (\frac{a}{2}, 2.5\mathring{A})$, see Fig. 5.2(a-c), close to the Ba nuclear position $(\frac{a}{2}, 0.19171 \cdot c \simeq 2.3\mathring{A})$ as determined from the unpolarised neutron diffraction experiment (see table 5.1). Since Ba is non-magnetic the feature centered at $(\frac{a}{2}, 2.5\mathring{A})$ cannot be explained by an induced moment originating from Ba. Instead it is evidence of Ba/Pr cation mixing, it shows the moment from the small fraction of Pr ions which have been substituted for Ba ions on the Ba site in the $PrBa_2Cu_3O_{6.24}$ single crystal used for our neutron experiments. From table 5.3 this moment is found to be $0.0164 \pm 0.0032\mu_B$, which is 6% of the Pr moment and is only an estimate taking into consideration that the susceptibility of Pr on the Ba site may not be the same as Pr on the Pr site. A nearby feature with center $\sim (\frac{a}{2}, 4.1\mathring{A})$ and moment $0.0021 \pm 0.0014\mu_B$ may also be originating from an induced moment on a tiny number of interstitual Pr ions on this site. Alternatively its position is near the O(2) nuclear site $(\frac{a}{2}, 0.3697 \cdot c = 4.4\mathring{A})$ hence it could be attributed to O(2).

A number of relatively strong negative features are present. The largest of these is positioned at $(\frac{a}{2}, 0)$, see Fig. 5.4(d), with moment $-0.0103 \pm 0.0027\mu_B$.

[9] In Ref. [106] it is argued that two crystals with the same oxygen composition held at the same temperature may also exhibit different Cu(1) moments due to specific oxygen ordering of each sample, depending on how the samples were prepared.

An obvious guess would be that this feature originates from an induced moment on the O(3) atom position at the crystallographic site $(1/2, 0, 0)$. However, later discussion suggests that it may be situated at the $(1/2, 1/2, 0)$ crystallographic site, or alternatively a combination of both $(1/2, 0, 0)$ and $(1/2, 1/2, 0)$. More importantly, this feature is definitely real within the uncertainty of the data. Close to $(0.5\mathring{A}, 2.0\mathring{A})$ in Fig. 5.4(d) is a smaller negative moment of $-0.0034 \pm 0.0014\mu_B$. Its multiplicity in the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction is 4 and therefore its total moment contribution to the unit cell is $4(-0.0034\mu_B) = -0.0136\mu_B$, and it will be discussed in some detail later. The last feature of table 5.3 is situated to the left and right of the Pr site with a moment of $0.0036 \pm 0.0020\mu_B$. It is quite a small moment and its origin is unknown.

The second independent data set to be discussed is obtained by selecting those structure factors measured with $\mathbf{B} \parallel [1\ 1\ 0]$ and with Miller indices in the $[-1\ 1\ 0] - [0\ 0\ 1]$ plane. There are 36 of those in tables 5.5-5.6 and by adding the zero-q structure factor a data set consisting of 37 structure factors is obtained. A MEMx reconstruction of this data set with $\bar{M}^s = 0.1\mu_B$ $(Q = P = 37)$ is shown in Fig. 5.5 and Fig. 5.6(b). The data are contained in the $[-1\ 1\ 0] - [0\ 0\ 1]$ plane, so the periodic lattice parameters are $\frac{a}{\sqrt{2}}$ and c respectively, and for convenience the $[-1\ 1\ 0]$ axis is displayed from $\frac{a}{2\sqrt{2}}$ to $\frac{3a}{2\sqrt{2}}$. As for the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction the plane symmetry is mm and the reconstruction is calculated on a 64×256 grid.

For the highest magnetization density values the experimentally obtained density around the Pr site in Fig. 5.6(b) takes the form of a sphere, in agreement with the theoretical calculated density in Fig. 5.5(a)[10]. This agrees with observations made for the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction and confirms that, with the available data and at the strongest magnetization density levels, the Pr density

[10]Except for fine features in the theoretical density not present in the experimental observed density. Even if these are present they are not expected to be revealed with the available data, because of their small spatial dimension - see discussion of Fig. 5.6(a) in the text above.

shows no evidence of deviating from the crystal field calculated density. For magnetization density values a factor of 100 or more lower than the highest density value there is evidence of small 'deformations' of the Pr density from, the shape predicted by the crystal field calculations. Of these deformations only one seems to be pointing in the direction of one of the known nuclear sites (other than the Pr site) as determined from the unpolarised neutron experiment (see table 5.1). This feature is most clearly seen in Fig. 5.5(c) close to $(3.4\mathring{A}, 4.8\mathring{A})$ and could possibly be pointing towards the O(2) nuclear positions at $(0, 4.4\mathring{A})$. It is relatively small in magnitude compared to the majority of other features in the reconstruction and an attempt to integrate over it[11] gave $0.0001 \pm 0.0004\mu_B$. To the left and right of the Pr site in Fig. 5.5(b-c) is another small feature with $0.0003 \pm 0.0008\mu_B$, this is also weak. Another structure close to the Pr site is the negative feature at $\sim (\frac{a}{2\sqrt{2}}, 5.1\mathring{A})$ with $-0.0012 \pm 0.0012\mu_B$ (see Fig. 5.5(d)), and is just about significant within the uncertainty of the data. The strongest negative moment for the $[-1\ 1\ 0] - [0\ 0\ 1]$ reconstruction appears at $\sim (\frac{a}{2\sqrt{2}}, 2.1\mathring{A})$ with $-0.0030 \pm 0.0016\mu_B$. Combined with the observed negative features for the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction, see Fig. 5.4(d), the strongest negative features appear to be in and below the region of the Ba site (as seen from the bottom Ba site). Apart from the moment on the Pr site, some of these contribute with some of the biggest moments in the unit cell, for instance the $(\frac{a}{2}, 0)$ and $(0.5, 2.0)$ features in the $[0\ 1\ 0] - [0\ 0\ 1]$ projection with total moments of $1 \times (-0.0103)\mu_B$ and $4 \times (-0.0034) = -0.0136\mu_B$ respectively. Under the assumption that the induced moments are distributed equally with $\mathbf{B} \parallel [1\ 0\ 0]$ and with $\mathbf{B} \parallel [1\ 1\ 0]$, the observation of smaller negative moments for the $[-1\ 1\ 0] - [0\ 0\ 1]$ projection may be explained in the region $z \simeq 0$ by a concentration of negative density close to the crystallographic site $(1/2, 1/2, 0)$ which could be cancelled out by a positive Cu(1) moment when projecting onto $[-1\ 1\ 0] - [0\ 0\ 1]$. Similarly, for $z \simeq 2\mathring{A}$, a

[11] Because it is not located at an isolated position but is part of a larger feature it is difficult to determine the area which is covered i.e. its area of integration in table 5.4.

concentration close to the O(1) nuclear site may cancel out parts of the positive density at the Ba site, originating from Pr/Ba cation mixing. This latter feature is estimated to be $0.0109 \pm 0.0025\mu_B$ in the $[-1\ 1\ 0] - [0\ 0\ 1]$ reconstruction, this is $0.0060\mu_B$ less than the same feature in the $[0\ 1\ 0] - [0\ 0\ 1]$ projection, and supports the explanation of a negative moment close to the O(1) site. The moment at the Pr site is found to be $0.2530 \pm 0.0030\mu_B$ in the $[-1\ 1\ 0] - [0\ 0\ 1]$ reconstruction, this is within the experimental error of the value for the same feature in the $[0\ 1\ 0] - [0\ 0\ 1]$ reconstruction. The remaining feature listed in table 5.4 is located at $\sim (\frac{a}{\sqrt{2}}, 3.8\mathring{A})$ with $0.0020 \pm 0.0013\mu_B$. This feature together with the $(\frac{a}{2}, 4.1\mathring{A})$ feature in table 5.3 may be evidence of a small fraction of Pr ions close to the crystallographic site $(a/2, a/2, 4\mathring{A})$. Alternatively, the nuclear position of Cu(2) in the $[-1\ 1\ 0] - [0\ 0\ 1]$ projection is $(\frac{a}{\sqrt{2}}, \sim 4.2\mathring{A})$ and the $(\frac{a}{2}, 4.1\mathring{A})$ feature could be originating from Cu(2).

5.4 Reconstruction of 'averaged' $PrBa_2Cu_3O_{6.24}$ data set

In this section all the 91 structure factors in table 5.5 and the 13 'averaged' structure factors in table 5.6 are considered as one data set. Adding the zero-q structure factor generates a data set consisting of 105 data points. In combining the data in tables 5.5-5.6 the assumption is (as also discussed in Sec. 5.2) that the induced moment distributions with $\mathbf{B} \parallel [1\ 0\ 0]$ and $\mathbf{B} \parallel [1\ 1\ 0]$ are very similar and the combined data set is analysed in space group P4/mmm, the known space group for $PrBa_2Cu_3O_{6.24}$. The extent to which this assumption is valid is difficult to assess, but judging from *i*) the similarity of the 13 structure factors in table 5.6 measured independently for both $\mathbf{B}$ field orientations (compare $F_{hkl}^{B \parallel [1\ 0\ 0]} \pm \sigma_{hkl}^{B \parallel [1\ 0\ 0]}$ and $F_{hkl}^{B \parallel [1\ 1\ 0]} \pm \sigma_{hkl}^{B \parallel [1\ 1\ 0]}$ in table 5.6); *ii*) the similarities between the reconstructions in Fig. 5.4 and Fig. 5.5; and *iii*) the fact

that $PrBa_2Cu_3O_{6.24}$ has tetragonal symmetry which means that an applied **B** field with any direction within the xy-plane is expected to induce the same total moment, the moment distributions with **B** $\parallel$ [1 0 0] and **B** $\parallel$ [1 1 0] are expected to be at least reasonably similar. It is therefore valid to analyse the combined data set. The data is reconstructed on a $16 \times 16 \times 64$ grid and the prior model constant $\bar{M}^s$ is found using the recipe in Sec. 4.7, Chap. 4. To illustrate the latter, a FOM_1 versus $\bar{M}^s$ plot is shown in Fig. 5.9, where the $\bar{M}^s$ reference point is found to be $\bar{M}^s{}_{ref} \simeq 0.4\mu_B$ and therefore a $\bar{M}^s$ value close to $0.1\mu_B$ is selected. Shown in Fig. 5.7 and continuing in Fig. 5.8 are 9 isosurfaces of the resulting reconstruction.

Starting from the top left-hand frame in Fig. 5.7, the isosurface at the contour level (CL) $1\mu_B\mathring{A}^{-3}$ is shown. This value is close to the highest magnetization density value, which is $3.08\mu_B\mathring{A}^{-3}$, and at this level, as expected, a single isosurface centered at the Pr site is observed. Moving down in contour levels the diameter of the isosurface enclosing the Pr site increases and at CL $= 0.05\mu_B\mathring{A}^{-3}$ the first feature not centered at the Pr site is seen. At CL $= 0.0025\mu_B\mathring{A}^{-3}$ a number of additional moments have appeared. All of these have been discussed in the previous section, but the isosurfaces offer additional information about the more precise spatial location in the unit cell, under the assumption that the induced moment distributions are very similar with the two **B** field orientations. For instance it is seen from the CL $= 0.0025\mu_B\mathring{A}^{-3}$ isosurface that the negative density in the plane $z \simeq 0$ seems to be centered at $(1/2, 1/2, 0)$ stretching towards the Ba site. The last two isosurface plots in Fig. 5.8 show the CL $= 0.00075\mu_B\mathring{A}^{-3}$ isosurface from two different perspectives. $0.00075\mu_B\mathring{A}^{-3}$ is more than a factor of 1000 below the highest magnetization density value and at this level many new features appear which are not supported within the uncertainty of the data. However, none of the isosurfaces show any evidence of the density surrounding the Pr site pointing towards any known nuclear positions, including nuclear sites within the CuO_2 planes. Only at CL $= 0.001\mu_B\mathring{A}^{-3}$ does the Pr density start to

deviate from a spherical symmetric form, here it appears to be stretching towards the $(0, 0, 1/2)$ crystallographic site, which is unexpected as there are no known atoms located at this site.

5.5 Polarised neutron diffraction data from $PrBa_2Cu_3O_7$

One other PND experiment on $PrBa_2Cu_3O_{6+x}$ has been carried out [109], this used a single crystal of $PrBa_2Cu_3O_7$. It would be interesting to compare the magnetization densities of $PrBa_2Cu_3O_7$ and $PrBa_2Cu_3O_{6.24}$. To achieve the best conditions for such a comparison the sub-sets of data from both data sets with common Miller indices are used. For $PrBa_2Cu_3O_7$ this subset is taken from Ref. [110], and both data subsets contain 30 structure factors which are listed for $PrBa_2Cu_3O_7$ in table 5.7 [12]. The reconstruction for the $PrBa_2Cu_3O_{6.24}$ sub-set with $\bar{M}^s = 0.1\mu_B$ (as used previously), on a 64×256 grid, yields a total calculated moment of $0.2265\mu_B$, and is displayed in Fig. 5.10. The reconstruction for $PrBa_2Cu_3O_7$ using the same grid and prior model constants yields a total calculated moment of $0.3223\mu_B$, it is scaled down by a factor $\frac{0.2265}{0.3223}$ to allow better comparison between the two data set, and is displayed in Fig. 5.11. There are differences between the two magnetization density maps in Fig. 5.10 and Fig. 5.11. However, around the area of the Pr site there is striking similarity for both positive and negative magnetization density values. This shows that the overall structure of the magnetization density distribution in the neighbourhood of the Pr site has little dependence on the oxygen content in $PrBa_2Cu_3O_{6+x}$ (at least when going from $x = 1$ to $x = 0.24$). This might not be surprising considering that the Pr site is furthest away from the Cu(1)-O(3) chains in the unit cell, where oxygen atoms can be taken out or inserted into the structure. Probably the most significant difference between the two $PrBa_2Cu_3O_{6+x}$ compounds is seen at the Cu(1) position, where a stronger moment is induced on Cu(1) in $PrBa_2Cu_3O_7$

[12] The equivalent sub-set for $PrBa_2Cu_3O_{6.24}$ can be obtained from tables 5.5-5.6.

relative to Cu(1) in $PrBa_2Cu_3O_{6.24}$. The magnetization density near the Ba site is also different. These differences may be caused by the use of single crystals grown in different crucibles. To grow the $PrBa_2Cu_3O_7$ and $PrBa_2Cu_3O_{6.24}$ single crystals, Al_2O_3 and MgO crucible were used respectively.

Another issue governs the interpretation of regions of a magnetization density image which have zero density. This may seem trivial but it is worth a discussion in the context of data presented in this chapter. Clearly the question is only relevant when dealing incomplete data, as is usually the case, and the question of whether an area of zero density means that there is nothing there, or whether the zero density is caused by a shortcoming of the data may arise. As a first example, the sub-set containing the 30 data points used to generate Fig. 5.10 is taken from the larger 37 point data set discussed in Sec. 5.3.2 and displayed in Fig. 5.5. By comparing Fig. 5.10 and Fig. 5.5 it is clear that notable differences are present. Looking more carefully at these figures it is also apparent that there are no features in Fig. 5.10 which are not seen in some form in Fig. 5.5. The opposite is not the case, as expected. This emphasizes that when data are taken out of the data set the picture of the resulting reconstruction may change. More specifically, compare the Pr density in Fig. 5.10 and Fig. 5.5. From Fig. 5.10 the Pr density looks perfectly circular. However, the reconstruction in Fig. 5.5 is circular but with a few additional small features.

For the special case of going from Fig. 5.5 to Fig. 5.10 most of the information lost can regained by simply adding two of the seven structure factors which are not part of the 30 point sub-set. This is illustrated in Fig. 5.12 where the the 0 0 1 and 0 0 2 structure factors have been inserted back to obtain a 32 structure factor sub-set. It is therefore tempting to conclude that if the 0 0 1 and 0 0 2 structure factor were measured for the $PrBa_2Cu_3O_7$ crystal a similar re-distribution of the negative and positive magnetization would be observed.

5.6 Discussion and conclusions

Studying the magnetization density reconstructions of $PrBa_2Cu_3O_{6.24}$ in Sec. 5.3-5.4 and comparing then with the magnetization density of $PrBa_2Cu_3O_7$ in Sec. 5.5, a number of very interesting features are revealed, in many cases the evidence of these features in the data is great. How do these new observations relate to previous experimental and theoretical work?

Focusing on the induced moment density of the Pr ion, it was seen in Sec. 5.3 that its overall geometry (spherical) and size could be explained with a local crystal field model for Pr^{3+}. This is not surprising since a number of magnetic excitation spectra on $PrBa_2Cu_3O_{6+x}$ have been measured [111],[112],[99],[113],[114], and although the excitation spectra shown unusual peak broadening as compared to spectra measured on $RBa_2Cu_3O_{6+x}$ (R=rare earth), the crystal field parameters describing the local environment of the Pr ion can be refined with good accuracy. Hence it is not surprising that the major shape of the induced Pr moment in $PrBa_2Cu_3O_{6.24}$ was found to be well described within such a crystal field model. However, crystal field parameters do not describe every detail of a rare earth local environment, for instance, effects which cause peak broadening are not accounted for. In particular the enhanced widths in the magnetic excitation spectrum of $PrBa_2Cu_3O_{6+x}$ are attributed to an increased hybridisation between $4f$ electrons and electronic states in the CuO_2 planes [111],[112],[99],[113], and such an hybridisation is said to be responsible for the T_C suppression in $PrBa_2Cu_3O_{6+x}$ by the majority of people working in this field (see review papers Refs. [1],[93] and references therein). It would therefore be of key interest if the structure of this hybridisation could be mapped out. A number of theoretical works, including Fehrenbacher & Rice model [94], predict the cause to involve a hybridisation of the $4f$ electrons of Pr and electronic states in the CuO_2 planes. An ideal experiment to perform in order to probe the exact spatial distribution of these $4f$ electrons is a PND experiment. This measures the Fourier components of the

magnetization density distribution in the unit cell and since only the $4f$ electrons of Pr are magnetic, it is only these Pr electrons which are detectable in a PND experiment, plus other non Pr magnetic contributions. A unit cell magnetization density may look different from a unit cell electron density (see Chap. 1), but a $\mathbf{B}$ field of magnitude $4.6T$ was used for the PND experiments reported here and these densities should be similar. In any case, to make a direct comparison to a PND experiment the expectation value of magnetization density operators may be calculated for any proposed theoretical model and then compared directly with the observed magnetization density image.

In summary the following has been observed.

1. In the region around the Pr site the magnetization density distribution was

   (a) found to be more or less independent of oxygen composition (at least when going from $x = 1$ to $x = 0.24$ for $PrBa_2Cu_3O_{6+x}$.

   (b) The crystal field model refined from magnetic excitation spectra was found to describe the features of the Pr density for magnetization density values ranging from the highest value to 1/100 of the highest value. However, from the discussions in Sec. 5.5 it should be remembered that if more structure factors are measured, the Pr density may gain more fine structure, although, the dominating Pr feature will remain the same.

   (c) Small deviations of this density were seen below 1/100, this puts a constraint on the degree of possible hybridisation and theoretical models like that of Fehrenbacher & Rice. In principle to make a full comparison would require the calculation of the magnetization density corresponding to the Fehrenbacher & Rice state (and equivalent states from other theoretical works) under the influence of an applied field, at a temperature just above the Neel temperature of the Pr moments (say $T = 20K$).

2. The observed magnetic structure in the region around the Ba site is also of interest, since some of these features could be key to the explanation of the suppression of superconductivity in $PrBa_2Cu_3O_{6+x}$, or may even be the reason for such suppression. The experiment which would clarify this uniquely, would be a PND experiment, as described in Sec. 5.2, using a single crystal of $PrBa_2Cu_3O_{6+x}$ grown under special conditions [115],[116], where superconductivity in this compound has been achieved. If the magnetization density distribution in the region of the Pr site is different for such a single crystal[13] then Pr on the Pr site does contribute to the suppression of superconductivity for a 'standard' $PrBa_2Cu_3O_{6+x}$ crystal. It could may also be possible that small amounts of Pr/Ba mixing trigger the Pr ions on the Pr sites to form a hybridisation state with O($2p$) levels in the CuO_2?

[13]The comparison would be best for crystal grown with an oxygen composition where $PrBa_2Cu_3O_{6+x}$ is in its tetragonal phase.

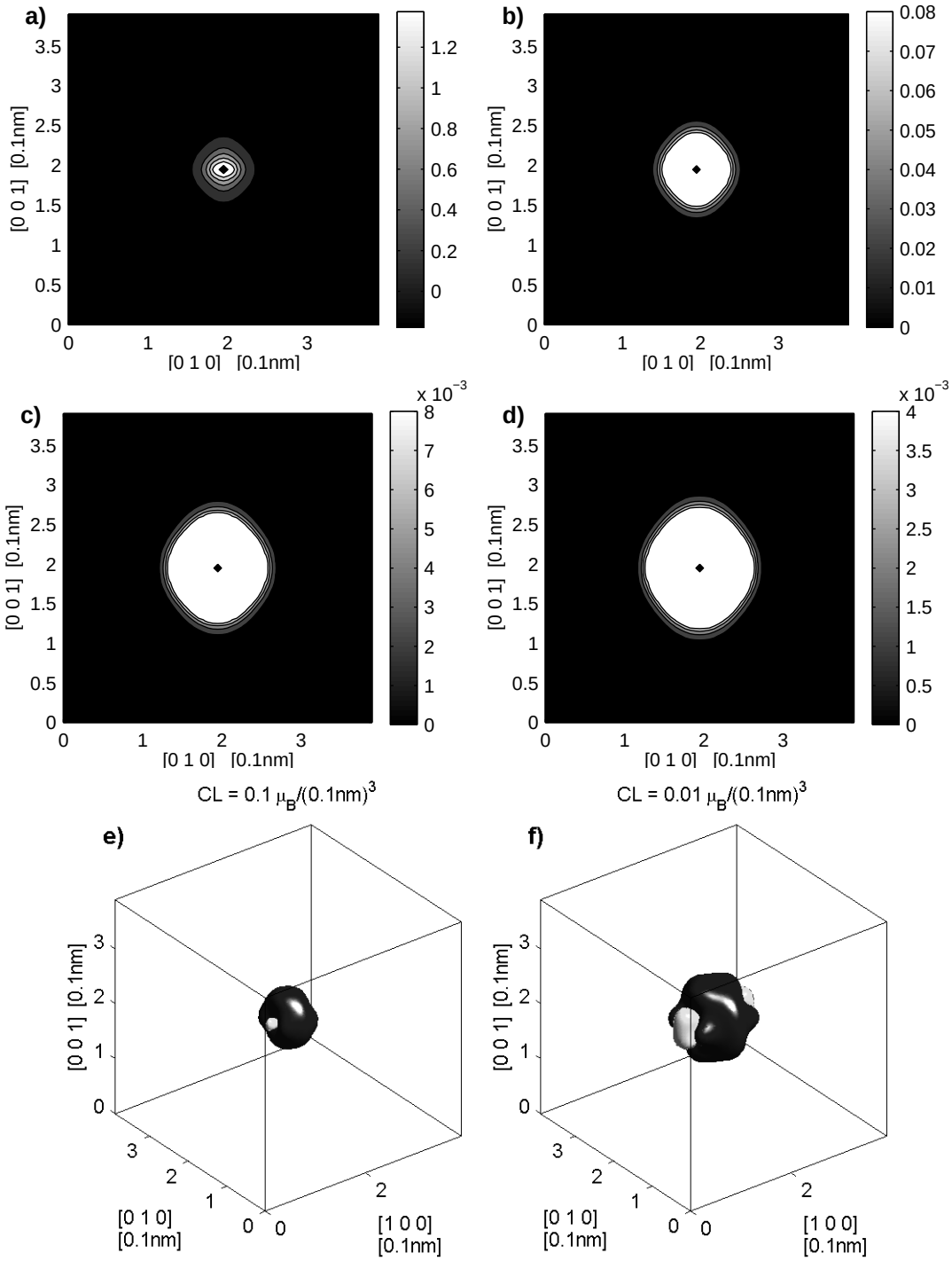

Figure 5.2: Shows $\langle M_x(\mathbf{r})\rangle_{[1\ 0\ 0]}$. Two isosurfaces illustrate $\langle M_x\rangle$ in (e-f), the lighter coloured surfaces indicate negative magnetization of the same contour level. In (a-d) a projection of $\langle M_x\rangle$ is shown where in (b-d) all negative density = 0 and in addition in (b) all positive density above 0.1 is cut-off, in (c) cut-off=0.01 and in (d) 0.005 $\mu_B \mathring{A}^{-2}$.

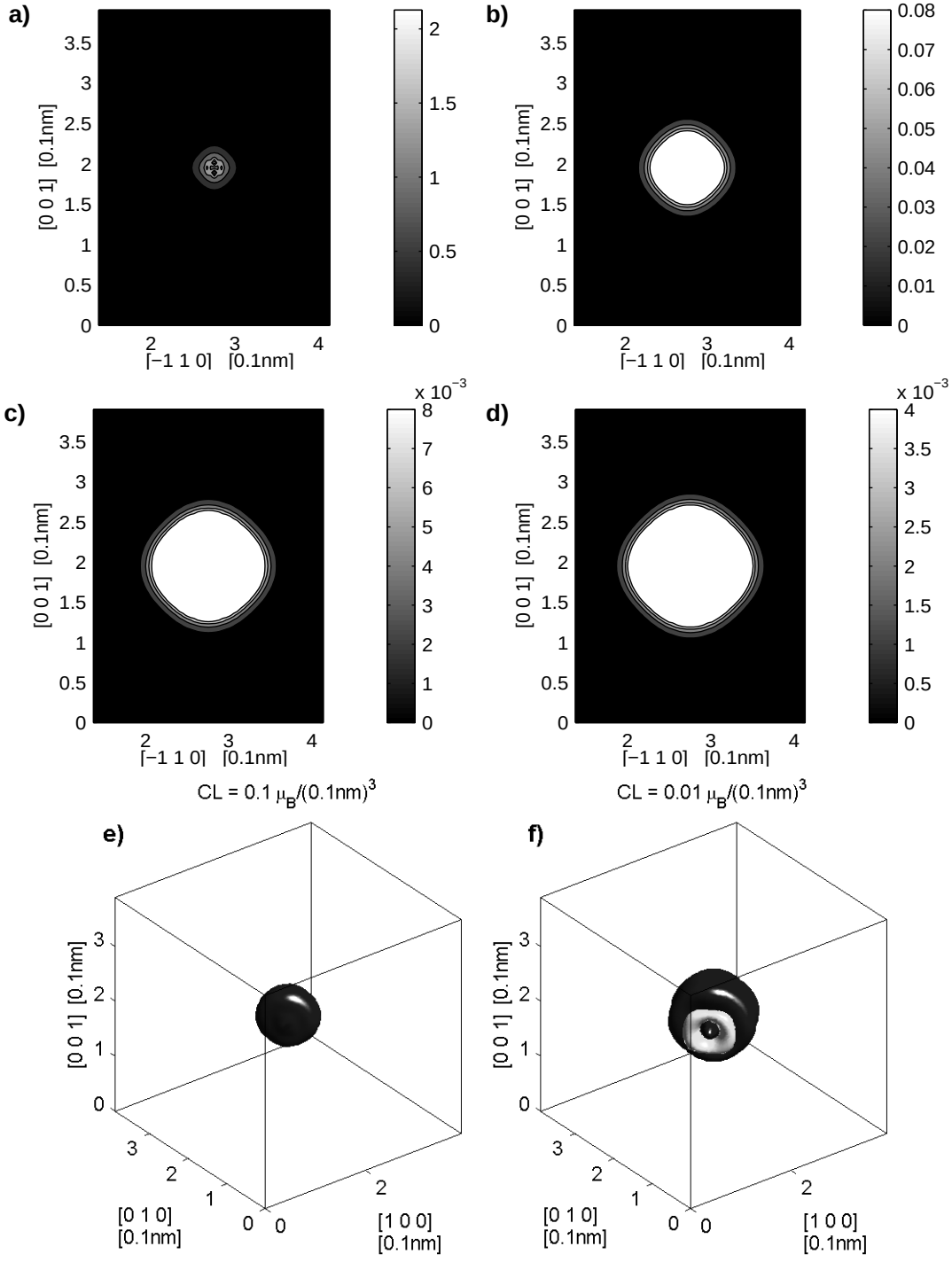

Figure 5.3: Same as in Fig. 5.2 but using the wave function with $\mathbf{B} \parallel [1\ 1\ 0]$ in table 5.2.

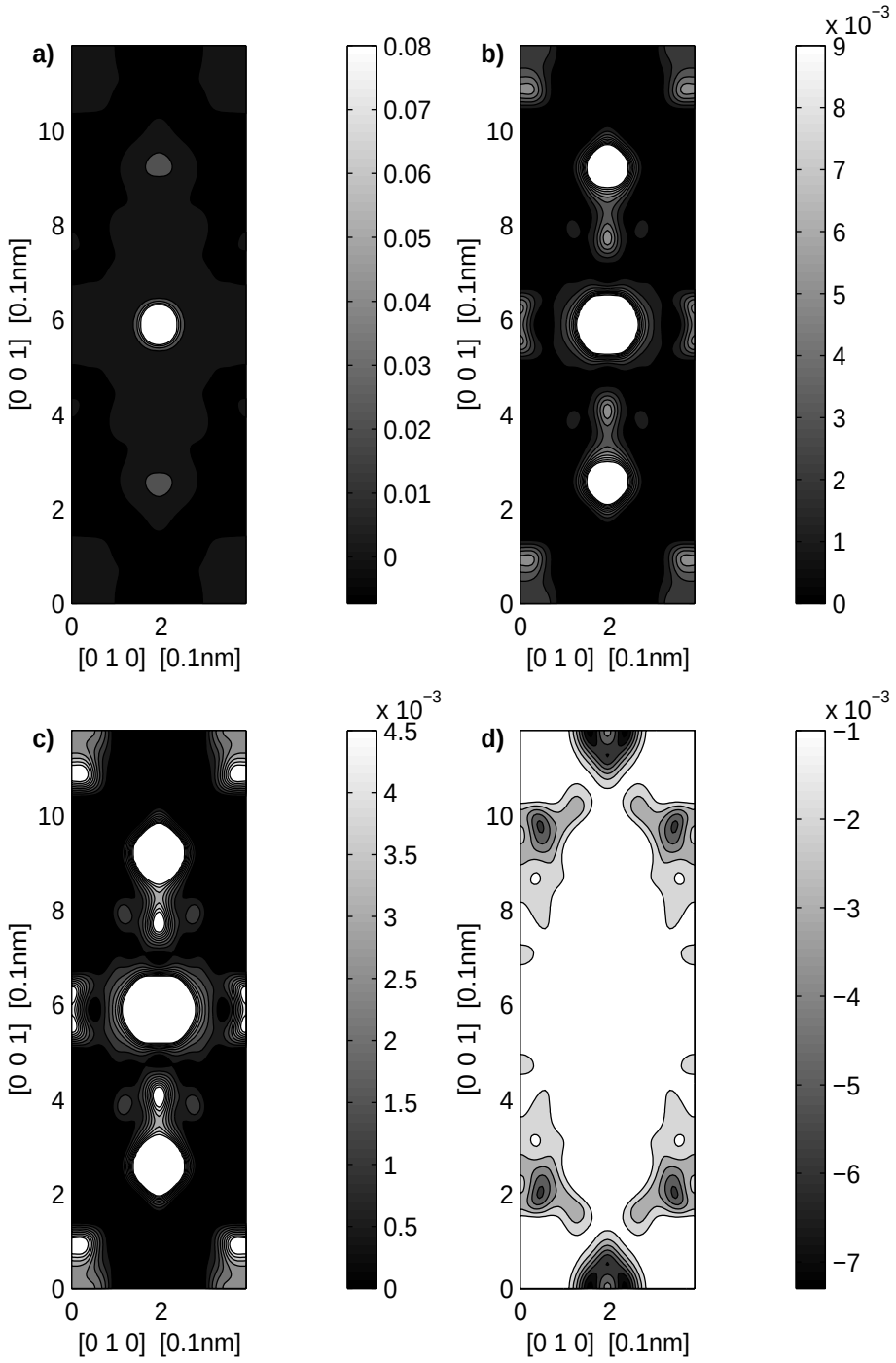

Figure 5.4: Shows a reconstruction of the data set consisting of the structure factors in tables 5.5-5.6 which are in the $[0\ 1\ 0] - [0\ 0\ 1]$ plane. In (a-d) all density above 0.1, 0.01, 0.005 and $0\mu\mathring{A}^{-2}$ respectively are cut-off. In addition in (b-c) all negative density = 0. $\bar{M}^s = 0.1\mu_B$ $(Q = P = 53)$.

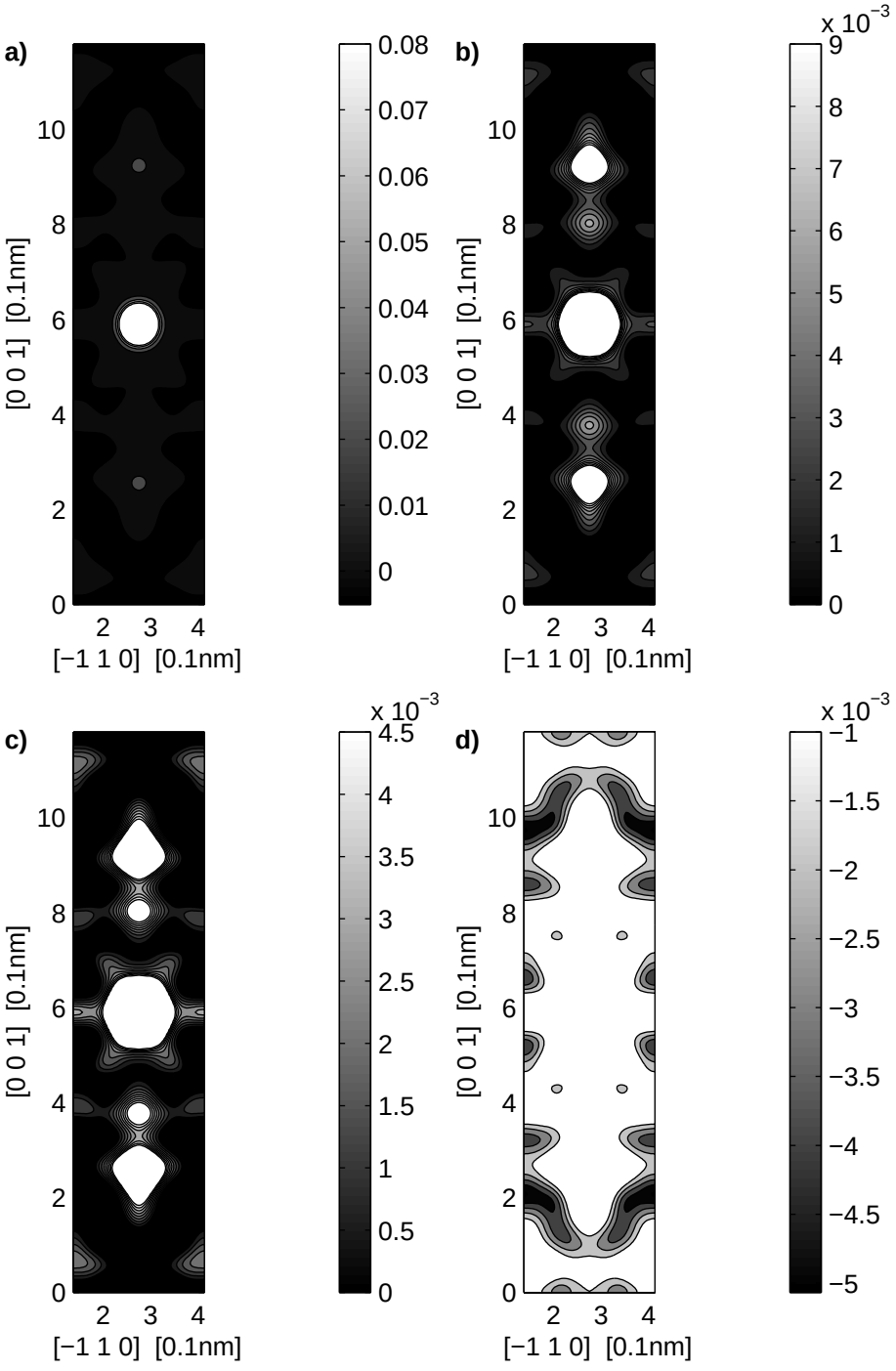

Figure 5.5: Same as in Fig. 5.4 but for the data in tables 5.5-5.6 which are in the $[1\ 1\ 0] - [0\ 0\ 1]$ plane. $\bar{M}^s = 0.1\mu_B$ $(Q = P = 37)$.

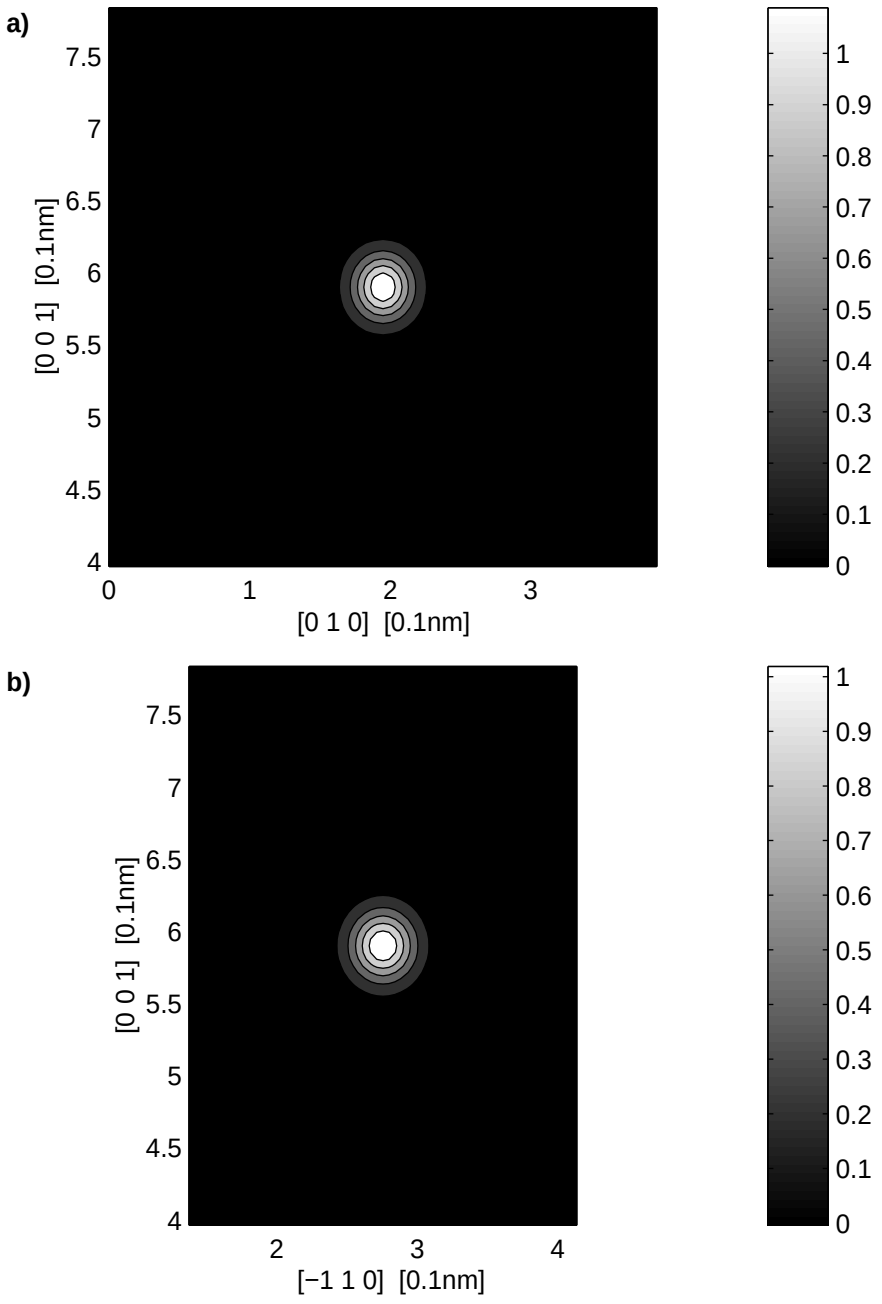

Figure 5.6: (a) same shows a contour plot of same reconstruction as described in Fig. 5.4 with no cut-off and for part of the unit to emphasize the density close to the Pr site. (b) same as (a) but for the reconstruction in Fig. 5.5.

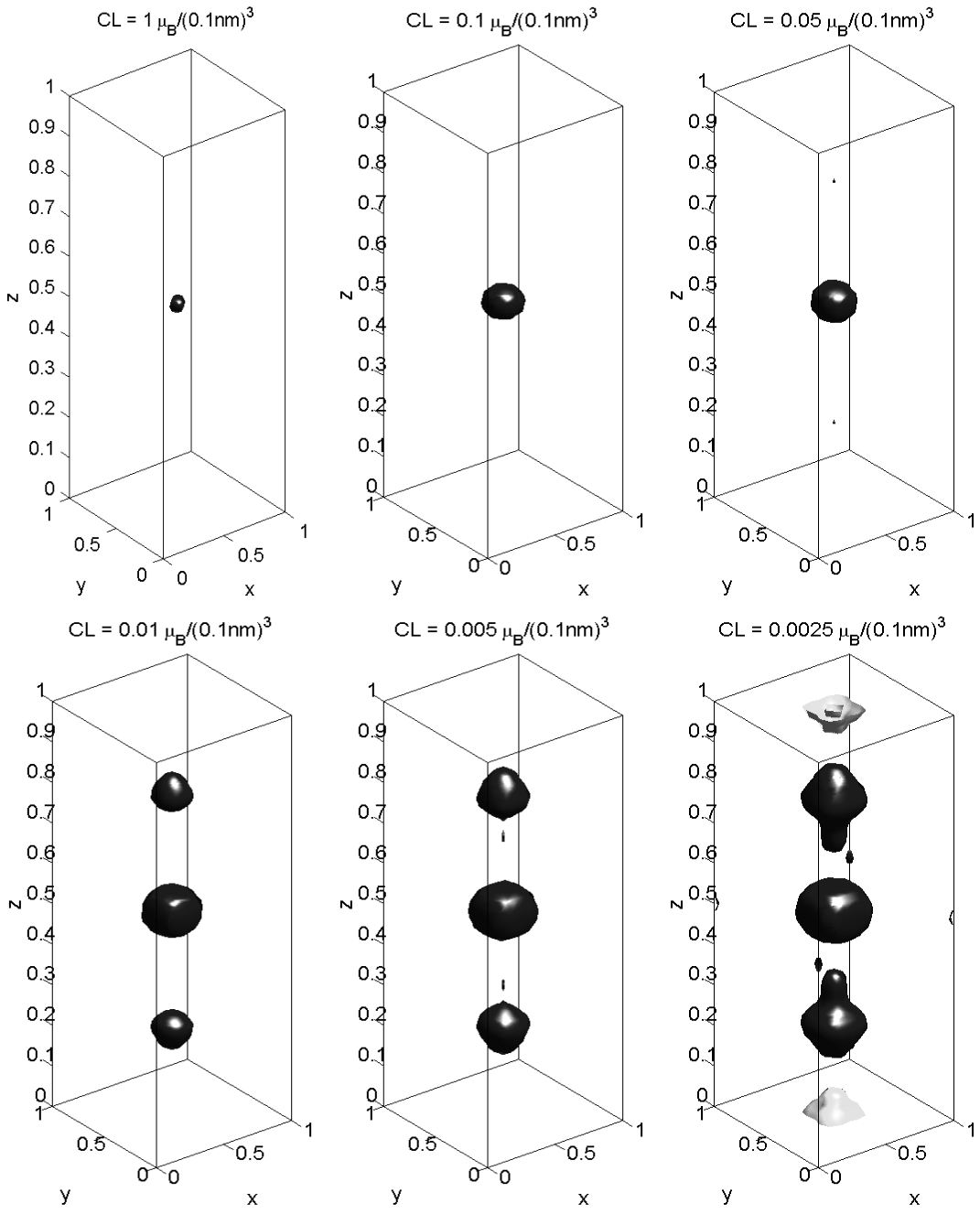

Figure 5.7: Reconstruction of the 'averaged' $PrBa_2Cu_3O_{6.24}$ data in tables 5.5-5.6. $\bar{M}^s = 0.1\mu_B$ and $Q = 105$. Continues on next page.

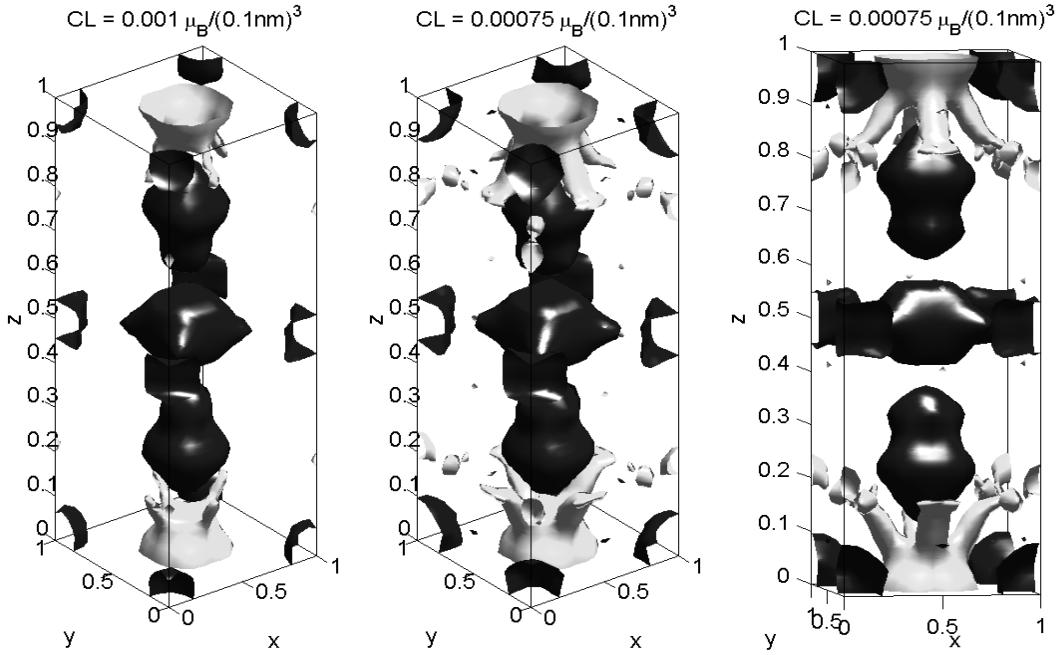

Figure 5.8: Continued from Fig. 5.7.

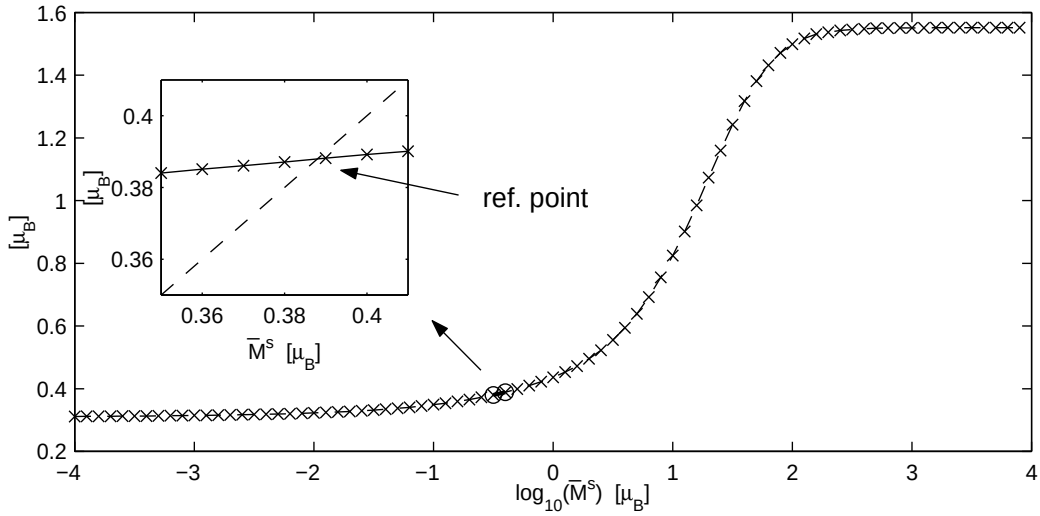

Figure 5.9: Shows the figure of merit $v\sum_i |m_i^{max}|$ as a function of $\bar{M}^s$ for the data in tables 5.5-5.6 ($Q = 105$). The inset represents a zoom-in on the region near the reference point $\bar{M}^s{}_{ref}$ estimated to $\bar{M}^s{}_{ref} \simeq 3.8\mu_B$. Using the recipe in Sec. 4.7, Chap. 4 this suggests that the prior constant $\bar{M}^s$ should be selected close to $\bar{M}^s = 0.1\mu_B$.

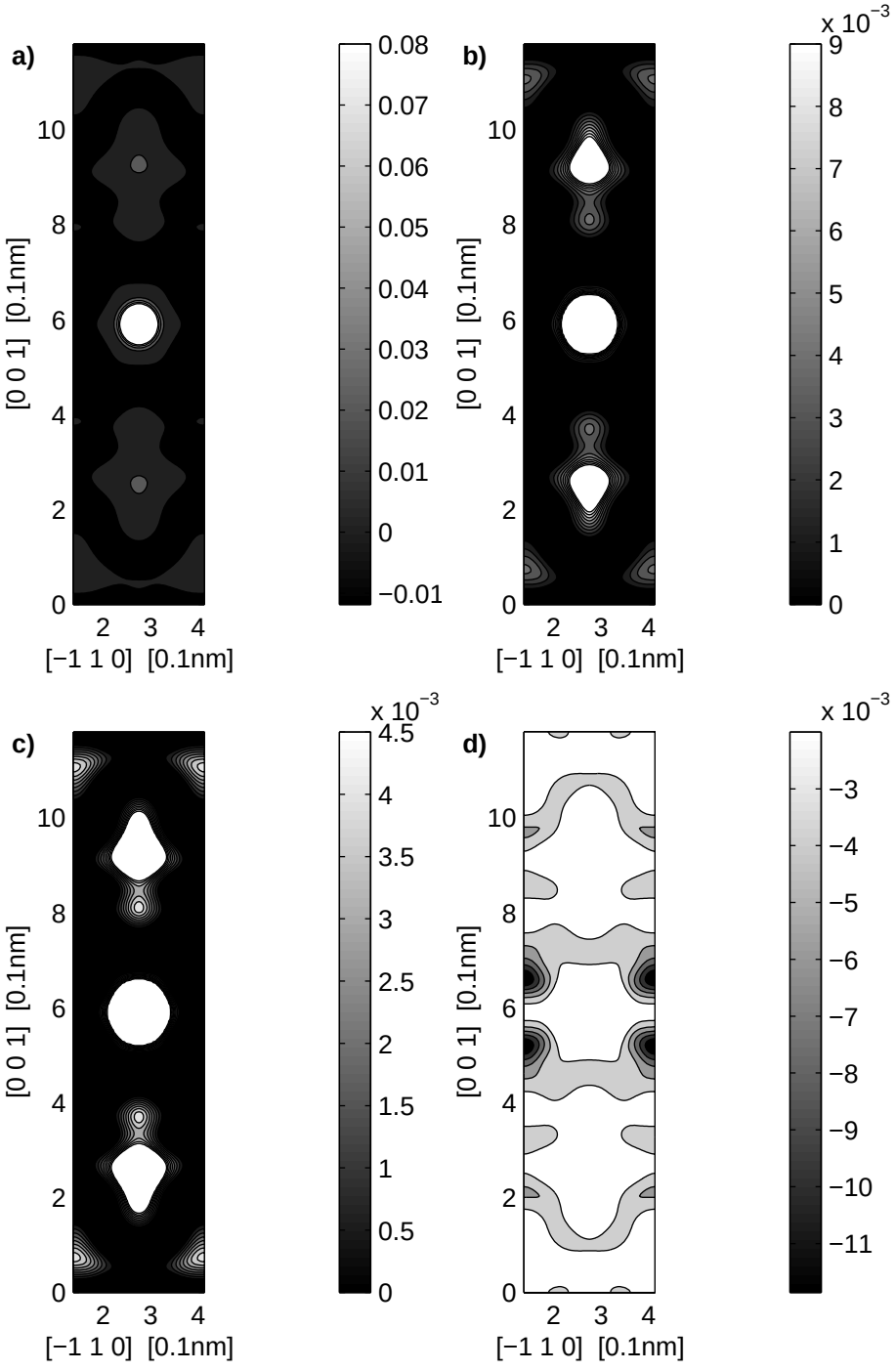

Figure 5.10: Same as in Fig. 5.4 but for the 30 structure factor data sub-set described in the text. $\bar{M}^s = 0.1\mu_B$ ($Q = P = 30$).

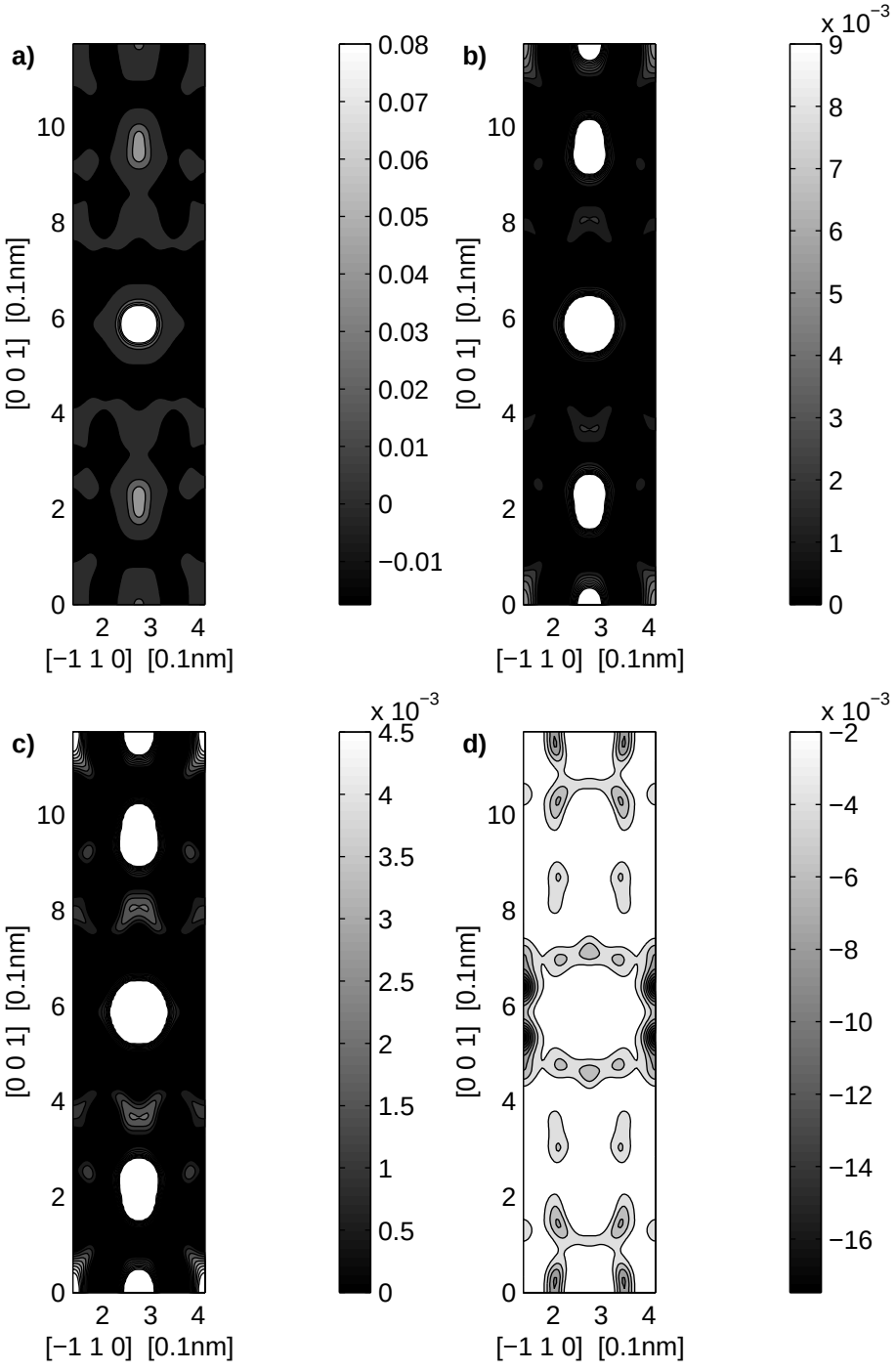

Figure 5.11: Same as in Fig. 5.4 but for the data given in table 5.7. $\bar{M}^s = 0.1\mu_B$ ($Q = P = 30$).

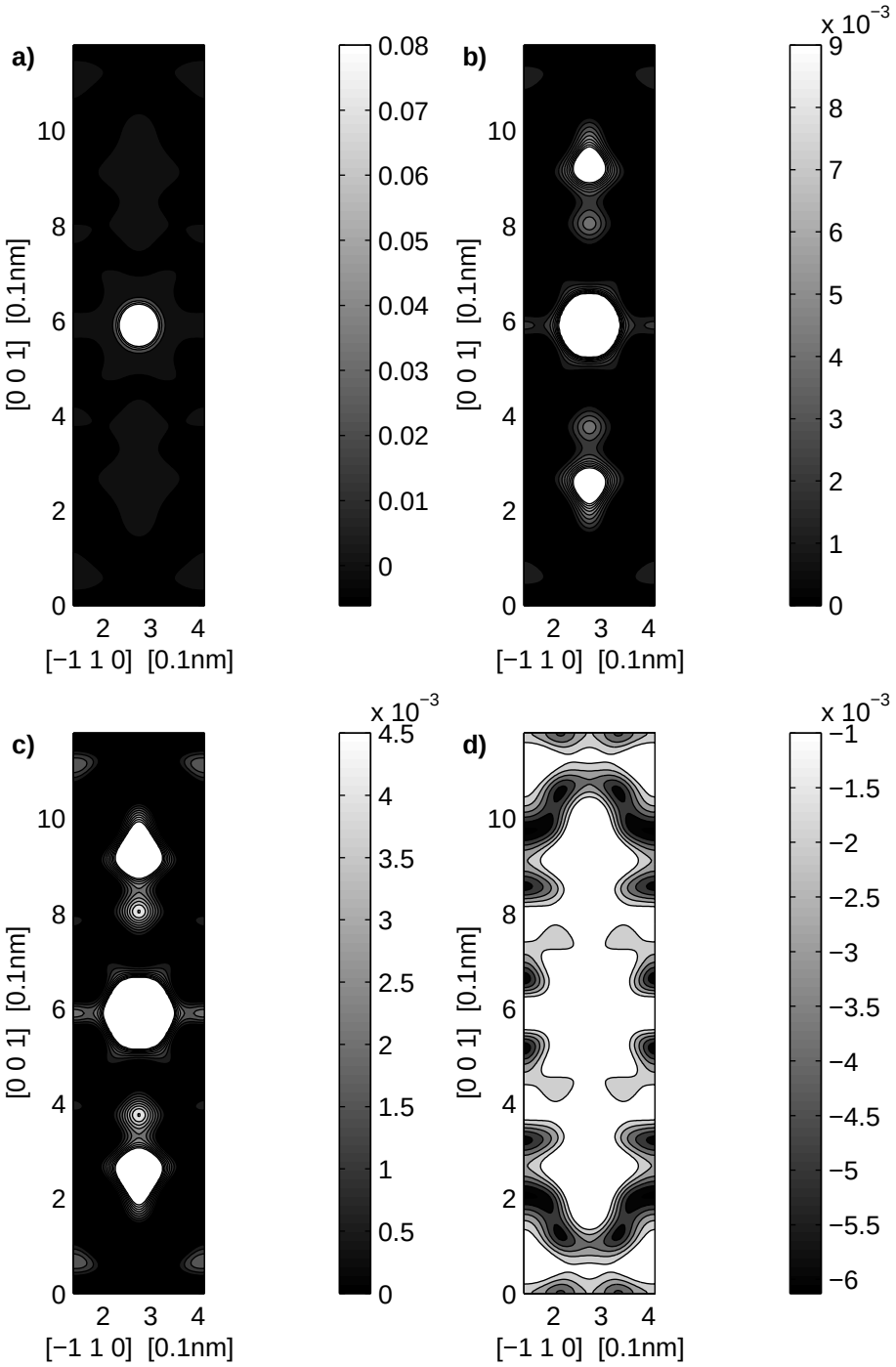

Figure 5.12: Same as in Fig. 5.10 but where the 0 0 1 and 0 0 2 structure factor in the $\mathbf{B} \parallel [1\ 1\ 0]$ columns of table 5.6 have been added to the data used to obtain the reconstruction in Fig. 5.10. $\bar{M}^s = 0.1\mu_B$ $(Q = P = 32)$.

Miller indices			**B** ∥ [1 1 0]		Miller indices			**B** ∥ [1 0 0]	
h	k	l	F_{hkl} [μ_B]	σ_{hkl} [μ_B]	h	k	l	F_{hkl} [μ_B]	σ_{hkl} [μ_B]
0	0	6	0.20277	0.00497	0	1	0	-0.26468	0.00324
0	0	14	0.08150	0.03194	0	1	1	0.25115	0.00381
-1	1	0	0.26083	0.00209	0	1	2	-0.15615	0.00665
-1	1	1	-0.23235	0.00361	0	1	3	0.24869	0.00342
-1	1	2	0.18454	0.00907	0	1	4	-0.22842	0.00574
-1	1	3	-0.23857	0.00468	0	1	5	0.16901	0.01679
-1	1	4	0.22956	0.00651	0	1	6	-0.19589	0.00642
-1	1	6	0.19071	0.00591	0	1	7	0.20587	0.00940
-1	1	7	-0.18871	0.02693	0	1	8	-0.17158	0.01657
-1	1	8	0.15241	0.01017	0	1	9	0.14497	0.01666
-1	1	12	0.09527	0.01862	0	1	10	-0.14689	0.01096
-1	1	13	-0.08031	0.04310	0	1	11	0.16102	0.02396
-1	2	0	-0.22128	0.00615	0	1	12	-0.11327	0.04215
-1	2	1	0.19502	0.00904	0	1	13	0.13327	0.06028
-1	2	2	-0.12092	0.03144	0	1	14	-0.10895	0.03671
-1	2	3	0.21010	0.00513	0	1	16	-0.05703	0.05714
-1	2	4	-0.19940	0.00991	0	2	0	0.22949	0.00281
-1	2	6	-0.17748	0.01020	0	2	1	-0.20126	0.00731
-1	2	7	0.15539	0.02223	0	2	2	0.16382	0.01579
-1	2	8	-0.12791	0.01024	0	2	3	-0.21254	0.00624
-2	2	0	0.19310	0.00613	0	2	4	0.19731	0.00596
-2	2	1	-0.16029	0.01265	0	2	5	-0.17360	0.01037
-2	2	3	-0.17447	0.00833	0	2	6	0.17202	0.01177
-2	2	4	0.15847	0.00818	0	2	7	-0.18728	0.01274
-2	2	5	-0.15544	0.00704	0	2	8	0.14866	0.01003
-2	2	6	0.15380	0.00781	0	2	11	-0.09824	0.01487
-2	2	7	-0.13972	0.00727	0	3	0	-0.17707	0.01896
-2	2	8	0.12727	0.00691	0	3	3	0.15865	0.00945
-2	2	11	-0.07750	0.03007	0	3	4	-0.15783	0.02683
-1	3	4	0.06409	0.04017	0	3	6	-0.14440	0.01756
-2	3	0	-0.16987	0.01342	0	3	7	0.02916	0.05050
-2	3	3	0.14865	0.00776	0	3	8	-0.15290	0.02333
-2	3	6	-0.12437	0.03527	0	3	9	0.09669	0.08361
-3	3	0	0.13211	0.02068	0	3	10	-0.11438	0.02763
-3	3	3	-0.11202	0.04944	0	4	0	0.13762	0.00882
-3	3	6	0.10140	0.02731	0	4	1	-0.13746	0.01996
-2	4	0	0.11051	0.01689	0	4	4	0.11550	0.02277
-2	4	3	-0.08728	0.04690	0	4	5	-0.11624	0.02787
-2	4	4	0.11038	0.06397	0	4	6	0.13621	0.01684
-3	4	0	-0.12230	0.03199	0	5	3	-0.00554	0.04764
-3	4	3	0.07594	0.02354	0	6	0	0.06006	0.03051
-4	4	0	0.06281	0.03504	0	6	6	0.07999	0.02675
-4	4	6	0.00814	0.05156	-1	4	0	-0.13049	0.03646
					-1	5	0	0.12043	0.04355
					-1	5	1	-0.02056	0.09670
					-1	4	3	0.12088	0.03028
					-1	3	6	0.15335	0.01162
					-1	5	6	0.03401	0.05516

Table 5.5: Shows the structure factors measured with **B** ∥ [1 1 0] and **B** ∥ [1 0 0] which do not have common Miller indices.

Miller indices			**B** ∥ [1 1 0]		**B** ∥ [1 0 0]		Average	
h	k	l	F_{hkl} [μ_B]	σ_{hkl} [μ_B]	F_{hkl} [μ_B]	σ_{hkl} [μ_B]	$\bar{F}_{hkl}$ [μ_B]	$\bar{\sigma}_{hkl}$ [μ_B]
0	0	1	-0.25995	0.00257	-0.26136	0.00390	-0.26066	0.00324
0	0	2	0.24081	0.01055	0.23386	0.00661	0.23734	0.00858
0	0	3	-0.24396	0.00604	-0.24463	0.00415	-0.24430	0.00510
0	0	4	0.25283	0.00537	0.25112	0.00612	0.25198	0.00575
0	0	5	-0.20652	0.00447	-0.21557	0.00827	-0.21105	0.00637
0	0	7	-0.21725	0.00648	-0.22942	0.00964	-0.22334	0.00806
0	0	8	0.16286	0.01319	0.16807	0.01134	0.16547	0.01227
0	0	9	-0.09085	0.05025	-0.10146	0.02917	-0.09616	0.03971
0	0	10	0.16137	0.02497	0.18291	0.04203	0.17214	0.00335
0	0	11	-0.15583	0.00879	-0.12644	0.01349	-0.14114	0.01470
-1	3	0	0.15186	0.02465	0.21622	0.01853	0.18404	0.03218
-1	3	1	-0.17214	0.02491	-0.14432	0.03295	-0.15823	0.02893
-1	3	3	-0.17616	0.02494	-0.17237	0.02146	-0.17427	0.02320

Table 5.6: Shows the structure factors measured with **B** ∥ [1 1 0] and **B** ∥ [1 0 0] which have common Miller indices, and an estimated 'averaged' structure factor value and standard deviation value for each hkl set.

Miller indices					Miller indices				
h	k	l	F_{hkl} [μ_B]	σ_{hkl} [μ_B]	h	k	l	F_{hkl} [μ_B]	σ_{hkl} [μ_B]
0	0	3	-0.326	0.003	-1	1	12	0.153	0.009
0	0	4	0.31	0.007	-1	1	13	-0.137	0.01
0	0	5	-0.234	0.004	-2	2	0	0.306	0.008
0	0	6	0.314	0.003	-2	2	1	-0.174	0.006
0	0	7	-0.277	0.009	-2	2	3	-0.235	0.007
0	0	8	0.203	0.008	-2	2	4	0.238	0.011
0	0	11	-0.173	0.008	-2	2	5	-0.171	0.008
0	0	14	0.153	0.012	-2	2	6	0.235	0.009
-1	1	0	0.364	0.005	-2	2	7	-0.173	0.007
-1	1	1	-0.272	0.003	-2	2	8	0.164	0.011
-1	1	2	0.264	0.007	-2	2	11	-0.146	0.011
-1	1	3	-0.335	0.006	-3	3	0	0.205	0.011
-1	1	4	0.294	0.011	-3	3	3	-0.191	0.012
-1	1	6	0.291	0.008	-3	3	6	0.144	0.011
-1	1	8	0.193	0.006	-4	4	0	0.084	0.012

Table 5.7: Shows the structure factors measured on $PrBa_2Cu_3O_7$ from Ref. [110], page 74, which have common Miller indices with any of the data in tables 5.5-5.6.

Chapter 6

A practical example of comparing MaxEnt algorithms

A comparison will be made between the MEMx reconstructions presented in Dobrzynski et al. [80],[81] using the algorithms presented by Sakata et al. [67] and Papoular et al. [66], and reconstructions obtained using MEMx in Def. 7, Chap. 4. This comparison of MEMx algorithms gives a fine illustration of the sort of difficulties that can appear when comparing MEMx related publications. The data analysed in [80],[81] is polarised neutron diffraction data of Ni (and Fe), and this chapter provides support for features observed by Dobrzynski et al. and others, and may provide additional insight into clarifying the magnetization density of nickel.[1]

6.1 Introduction

MEMx

Although the name MEMx should refer to one method, authors of papers involving the MEMx method often ignore or find it difficult to understand each other's

[1] Getting this far (and near the end) the reader almost deserves a good bottle of wine! Such as maybe a Brunello di Montalcino, 1993.

work. The papers by Dobrzynski et al. [80],[81] give a fine illustration of this. Here two data analysis methods with the same name (maximum entropy) are used to analyse the same data. In the introduction, Dobrzynski et al. [80],[81] give a short description of the two maximum entropy methods. The notation P-MEM and S-MEM are used respectively to refer to the MEMx algorithm of Papoular et al. [66] and Sakata et al. [67]. Clearly both Papoular et al. and Sakata et al. claim to be the authors of a MEMx method, so what is different? In Dobrzynski et al. [80],[81] differences are explained in terms of 'soft' and 'hard' constraints. This is a notation which, to my opinion, indicates that the authors do not totally understand the differences between these two MEMx algorithms. This will become apparent when reading Sec. 6.2 and Sec. 6.3 where the algorithms of Sakata et al. and Papoular et al. are discussed in detail. It is found that the algorithm of Papoular et al. is more appropriate for the analysis of PND data as compared to the algorithm of Sakata et al. for the particular case of PND data because of the physical information which is available in such data.

The aim of the above discussion is not to criticise the individual works of Sakata et al. and Papoular et al. but rather to present a criticism of the MEMx literature in general, which can at times be obscure and hard to compare, particularly for anyone who simply wants to apply the method.

Nickel problem

For the majority of crystal structures only a small subset of the electrons of the structure will be magnetic. Using polarised neutron diffraction the magnetization density of these electrons is measured. In addition the electron and magnetization densities are in general closely related and are often close to being proportional. Polarised neutron diffraction has the ability to probe only the magnetic electrons, and is the best known experimental method for obtaining information about the electron density of these magnetic electrons (see also Chap. 1). Of interest here are the transition metals, where the interpretation of experimentally measured

structure factors is still not unique. As an example, the magnetic structure factors (or equivalently form factors) calculated for free atoms are often scaled and fitted to the data (see Ref. [80],[81] and references therein). For transition metals most of the magnetic scattering comes from unpaired $3d$ electron spins. There is also an orbital contribution to the magnetization, but, since this contribution is small it introduces only a small uncertainty to the interpretation of the magnetization density being proportional to the electron density of the unpaired electrons (see Chap. 1).

Nickel is one of the most studied compounds using polarised neutron diffraction. As, for example, in the carefully performed experiments by Mook [117] and Maniawski et al. [118]. These data showed evidence of small negative magnetization in the regions between the Ni sites in addition to the dominating positive moments at the Ni sites. The theoretical model which has most frequently been used for modelling such data takes the negative magnetization density to be uniformly distributed throughout the unit cell, and the positive magnetization to be modelled from free atom form factors. Despite being simple, this picture has been successfully used to model experimental data. Alternative models have been proposed with the aim of adding more details into the description of transition metal magnetization densities. Discussion of such models can be found in Dobrzynski et al. [80],[81] and references therein. Dobrzynski et al. use the MEMx method to reanalyse the data of Mook and Maniawski et al., and observe new features which had not been revealed previously. The most significant of these being the depletion of magnetization density at the nuclei positions.

The main theme of this chapter is a discussion of the MEMx algorithms of Sakata et al. and Papoular et al.. The algorithm used by Sakata et al. is not directly applicable to reconstruction of magnetization densities from PND data, since it is designed for physical problems where the total amount of positive and negative integrated material of the unit cell are separately known, see Sec. 6.2. On the contrary it is found, in Sec. 6.3, that direct comparison can be made

between the reconstructions generated using the code of Papoular et al. and the reconstructions obtained using MEMx in Def. 7, Chap. 4. This requires that a specific choice is made for the value of the prior model constant, $\bar{M}^s$, in the MEMx method of Chap. 4, which turns out to be significantly different from the one recommended using the recipe for finding such a constant in Chap. 4. The mode of the posterior for the nickel data for two different $\bar{M}^s$ values is discussed in Sec. 6.4. In addition error estimates are calculated for some of the nickel features in Sec. 6.4. These results support the observations made by Dobrzynski et al. of depletion of magnetization density at the Ni sites.

6.2 The approach of Sakata et al.

Sakata et al. study charge densities in Ref. [119],[120] and nuclear densities in Ref. [67]. A description of the MEMx algorithm which has been used in their studies can be found in Ref. [121] and an analysis of it in Ref. [87].

A charge density is positive everywhere and the authors take the approach of interpreting the charge density as a probability distribution. Using the notation in Ref. [120] let ρ'_i be the probability density associated with the i^{th} bin (or pixel) of the image (charge density). Let ρ_i be the actual electron density, then ρ'_i and ρ_i are related by $\rho'_i = \rho_i / \sum_i \rho_i$. Interpreting an image (charge density) in this way (as a probability distribution) was first done, in the context of a maximum entropy data analysis method, by Frieden, see Chap. 3. Often when this approach is followed a 'prior' is associated with the 'image' probability distribution. It is not a prior in the sense of being a prior in Bayes' theorem but in the sense of the definition of MEM in Def. 7, Chap. 3. Sakata et al. use the notation τ'_i to denote the prior density for ρ'_i.

In Ref. [67] Sakata et al. apply MEMx to obtain the nuclear density for TiO_2. The mean scattering lengths of Ti and O atoms are denoted by b_{Ti} and b_{O} respectively. Since these two scattering lengths are of opposite sign the nuclear

density of TiO_2 will contain regions of positive and negative density. Sakata et al. treat the nuclear density originating from Ti and O separately and associate the probability density ${\rho'}_{\rm Ti}$ with the Ti atoms and ${\rho'}_{\rm O}$ with the O atoms. Likewise so-called priors ${\tau'}_{\rm Ti}$ and ${\tau'}_{\rm O}$ are introduced. An entropy is then defined as

$$S = -\sum_i \sum_{j=1}^{I} {\rho'}_{ij} \ln({\rho'}_{ij}/{\tau'}_{ij}) \quad , \tag{6.1}$$

where j is the sum over the image bins and i is the sum over the different types of atoms in the unit cell; for TiO_2 that is i =Ti and i =O. According to the broad definition given in Def. 7, Chap. 3, MEMx is the maximization of entropy subject to a chi-squared like statistic. The chi-squared like statistic used by Sakata et al. is for non-overlapping and centro-symmetric data

$$C = \frac{1}{P} \sum_{i=1}^{P} \frac{(F_i^{obs} - F_i)^2}{\sigma_i^2} \quad , \tag{6.2}$$

where P is the number of data points, and F_i^{obs} and F_i are the $i^{\rm th}$ observed and calculated structure factors respectively.

Introducing the Lagrangian multiplier λ and combining Eq. (6.1) and Eq. (6.2) Sakata et al. aim to optimize

$$f = S - \lambda C/2 \quad . \tag{6.3}$$

The independent variables in Eq. (6.3) are taken to be the ${\rho'}_i$'s and f is maximized with respect to these variables[2]. Sakata et al. assume that the total amount of material contributed from each atom is known and denote these amounts $n_i b_i$, where b_i is nuclear scattering length and n_i is the number of atoms of type i. Then the nuclear density and associated probability density for the $i^{\rm th}$ atom are related by $\rho_{ij} v = n_i b_i {\rho'}_{ij}$ and it is therefore possible to write the image-to-data Fourier

[2]To be mathematical correct, f in Eq. (6.3) should have the form $Q = S - \lambda C/2 - \sum_i \lambda_i \sum_{j=1}^{I} {\rho'}_{ij}$, since presumably we must require the associated probability distributions ${\rho'}_{ij}$ to be normalized for each type of atom i.

transform simply in terms of the ${\rho'}_{ij}$'s, as $F_l = \sum_i \sum_{j=1}^{I} n_i b_i {\rho'}_{ij} \exp(-2\pi i \mathbf{k}_l \cdot \mathbf{r}_j)$ [3]. Hence, taking the derivative of f with respect to the ${\rho'}_{ij}$'s and using $\partial S/\partial {\rho'}_{ij} = -1 - \ln({\rho'}_{ij}/{\tau'}_{ij})$ (and ignoring the -1 term) Sakata et al. in Ref. [67] obtain

$$0 = -\ln \frac{{\rho'}_{ij}}{{\tau'}_{ij}} - \frac{\lambda}{2} \frac{\partial C}{\partial {\rho'}_{ij}} \quad , \tag{6.4}$$

where

$$\frac{\partial C}{\partial {\rho'}_{ij}} = (2 n_i b_i / P) \sum_{l=1}^{P} \exp(-2\pi i \mathbf{k}_l \cdot \mathbf{r}_j)(F_l - F_l^{obs})/\sigma_l^2 \quad .$$

Let ${\tau'}_{ij}$ be related to a 'prior' nuclear density τ_{ij} as $\tau_{ij} v = n_i b_i {\tau'}_{ij}$ and Eq. (6.4) may be written in terms of nuclear densities by substituting $\ln \frac{{\rho'}_{ij}}{{\tau'}_{ij}}$ with $\ln \frac{\rho_{ij}}{\tau_{ij}}$ and likewise substituting ${\rho'}_{ij}$ in $\frac{\partial C}{\partial {\rho'}_{ij}}$ with $\frac{\rho_{ij} v}{n_i b_i}$.

In the computer algorithm used for finding the solution to Eq. (6.4) Sakata et al. use the following procedure. Rather than solving for Eq. (6.4) for fixed ${\tau'}_{\mathrm{Ti}}$ and ${\tau'}_{\mathrm{O}}$ (and data) by, for example, using one of the search algorithms described in Sec. 4.4.1 the MEMx solution is found within an approximation referred to as the "zeroth-order single-pixel approximation". The effect of employing this approximation is that the so-called priors ${\tau'}_{\mathrm{Ti}}$ and ${\tau'}_{\mathrm{O}}$ do not stay fixed in the optimization process, but are updated under each iteration circle. Hence, what that means is that Sakata et al. must assume that the MEMx solution which is obtained is not significantly dependent on the values of ${\tau'}_{\mathrm{Ti}}$ and ${\tau'}_{\mathrm{O}}$. Also, these values are never quoted in their publications.

The MEMx algorithm developed by Sakata et al. is not directly applicable to the analysis of PND data. This is because the equivalent values of the nuclear

[3] Write the Fourier transform between density and structure factor as

$$F_l = \int_{cell} \rho(\mathbf{r}) \exp(-2\pi i \mathbf{k}_l \cdot \mathbf{r}) d\mathbf{r} \quad .$$

This integral is in its discrete Fourier series approximation given as

$$F_l = \sum_i \sum_{j=1}^{I} {\rho'}_{ij} \exp(-2\pi i \mathbf{k}_l \cdot \mathbf{r}_j) \quad .$$

scattering lengths, b_i (for TiO_2 that is b_{Ti} and b_{O}), are not available for a PND experiment. These input values for the algorithm of Sakata et al. impose delta function constraints on the reconstructed image. Such a situation does not comply with the physical conditions of a PND experiment, where the total positive magnetic moment and total negative magnetic moment are not known, and cannot be measured separately[4]. The algorithm of Sakata et al. works well with the physical situation of obtaining the nuclear density from nuclear diffraction data but clearly not for analysising PND data. It is therefore concluded that the algorithm of Sakata et al. should probably not have been used for analysis of the PND data in Dobrzynski et al. because it forces information which is not available into the data analysis process. For that reason the Sakata et al. reconstructions in Dobrzynski et al. are not discussed further in Sec. 6.4.

6.3 The approach of Papoular et al.

Papoular et al. apply MEMx to the study of PND data in Refs. [66],[70],[122],[123]. MEMx is also used in Ref. [124] for the retrival of deformation electron densities and is applied in Ref. [125] to the reconstruction of Patterson and Fourier densities in orientationally disordered molecules.

Like Sakata et al., Papoular et al. use the approach of Frieden and associate a probability distribution with the physical density to be reconstructed. Using the notation in Ref. [124], ρ_i^+ and ρ_i^- denote positive and negative material respectively of the i^{th} bin and p_i^+ and p_i^- the associated probabilities. The relation

[4]Information about the total magnetic moment can be obtained (e.g. from a magnetic susceptibility measurement) but not in the form represented probabilistically by a delta function. The total moment is equal to the difference between the magnitude of the total positive moment and the magnitude of the total negative moment.

between these quantities are written as

$$\begin{aligned} p_i^+ &= \rho_i^+ / \sum_{i=1}^{I} (\rho_i^+ + \rho_i^-) \\ p_i^- &= \rho_i^- / \sum_{i=1}^{I} (\rho_i^+ + \rho_i^-) \quad . \end{aligned} \tag{6.5}$$

The Papoular et al. definition of the associated probability distributions in Eqs. (6.5) differs slightly from the Sakata et al. definition by using the denominator $\sum_{i=1}^{I}(\rho_i^+ + \rho_i^-)$ instead of $\sum_{i=1}^{I} \rho_i^+$ and $\sum_{i=1}^{I} \rho_i^+$. Notice that, in Eqs. (6.5) neither p_i^+ nor p_i^- sum up to one and are therefore not probabilities according to the usual definition of a discrete probability distribution. Similar to the definition for the entropy function by Sakata et al. in Eq. (6.1), Papoular et al. define entropy as

$$S = -\sum_{i=1}^{I} [p_i^+ \ln(p_i^+ / p_{0i}^+) + p_i^- \ln(p_i^- / p_{0i}^-)] \quad , \tag{6.6}$$

where p_{0i}^+ and p_{0i}^- [5] are the 'prior' probabilities to the 'image' probabilities p_i^+ and p_i^-, and p_{0i}^+ and p_{0i}^- are related to 'prior' material densities ρ_{0i}^+ and ρ_{0i}^- as in Eqs. (6.5). In Ref. [124] Papoular et al. define the chi-squared constraint function as (assuming data to be centro-symmetric)

$$C = \sum_{i=1}^{P} \frac{(F_i^{obs} - F_i)^2}{\sigma_i^2} \quad , \tag{6.7}$$

which is the same function as Q in Eq. (4.22), Chap. 4 and equal to C used by Sakata et al. in Eq. (6.2) multiplied by P, where P is the number of data points.

The main difference between the work of Papoular et al. and the work of Sakata et al., apart from differences in the search algorithms they use, is that Papoular et al. do not assume the values of $\sum_i \rho_i^+$ and $\sum_i \rho_i^+$ are known. So, when the derivative of the entropy in Eq. (6.6) is taken with respect to ρ_i^+ we obtain

$$\frac{\partial S}{\partial \rho_i^+} = \frac{-\ln(\rho_i^+ / \rho_{0i}^+) + A'}{\sum_{j=1}^{I}(\rho_j^+ + \rho_j^-)} \quad , \tag{6.8}$$

[5] Papoular et al. [124] use the notation m_{0i}^+ and m_{0i}^- for these quantities but to avoid confusion with notation already used in this thesis these are denoted here by p_{0i}^+ and p_{0i}^-.

where

$$A' = \sum_{j=1}^{I} [p_j^+ \ln(\rho_j^+/\rho_{0j}^+) + p_j^- \ln(\rho_j^-/\rho_{0j}^-)] \quad . \tag{6.9}$$

In particular when $\rho_{0i}^+ = \rho_{0i}^- = \rho_0$ for all $i = 1, 2, \ldots, I$, the $2I$ system of equations which are to be solved are

$$\begin{aligned} 0 &= -\ln \rho_i^+ + A - \lambda_0 [\sum_{j=1}^{I} (\rho_j^+ + \rho_j^-)](\partial C/\partial \rho_i^+) \\ 0 &= -\ln \rho_i^- + A - \lambda_0 [\sum_{j=1}^{I} (\rho_j^+ + \rho_j^-)](\partial C/\partial \rho_i^-) \end{aligned} \tag{6.10}$$

for $i = 1, 2, \ldots, I$, and $A = \sum_{j=1}^{I} [p_j^+ \ln \rho_j^+ + p_j^- \ln \rho_j^-]$ [6]. What is interesting is that if Eqs. (6.10) are to be solved exactly as written the resulting solution(s) would be of no physical use. The main reason for this is the introduction of A and its particular dependence on the ρ_i^+'s and ρ_i^-'s. If A had been a constant, i.e. independent of the ρ_i's, then the solution to Eqs. (6.10) would be a MEMx solution as will be demonstrated later. The relatively simple structure of the equations in Eqs. (6.10) enables the following straightforward derivations of a couple of properties which must be satisfied by any solution to Eqs. (6.10). C in Eq. (6.7) depends only on the difference densities $\rho_i^+ - \rho_i^-$, $i = 1, 2, \ldots, I$ because F_i^{obs} and F_i in Eq. (6.7) are by definition Fourier components of the difference density, and therefore $\partial C/\partial \rho_i^+ = -\partial C/\partial \rho_i^-$ for all i. Thus, adding together in pairs the equations in Eqs. (6.10) with the same index number i the following I equations are obtained

$$0 = -\ln \rho_i^+ - \ln \rho_i^- + 2A \tag{6.11}$$

for $i = 1, 2, \ldots, I$. From Eqs. (6.11) it is clear that $\ln(\rho_i^+ \rho_i^-) = 2A$ for all i. Multiply Eqs. (6.11) on both sides with $\sum_{j=1}^{I} (\rho_i^+ + \rho_i^-)$ to obtain

$$0 = -\sum_{j=1}^{I} (\rho_j^+ + \rho_j^-) \ln(\rho_i^+ \rho_i^-) + 2 \sum_{j=1}^{I} (\rho_j^+ \ln \rho_j^+ + \rho_j^- \ln \rho_j^-) \tag{6.12}$$

[6] Comparing with Eq. (9b) in Ref. [124] then A here is equal to $\ln A$ in that equation.

Using the fact that $\rho_i^+\rho_i^- = \rho_j^+\rho_j^-$ for all $j = 1, 2, \ldots I$ the first sum in Eq. (6.12) can be written as $\sum_{j=1}^{I}(\rho_j^+ + \rho_j^-)\ln(\rho_j^+\rho_j^-)$ and Eq. (6.12) can be reduced to

$$0 = \sum_{i=1}^{I}(\rho_i^+ - \rho_i^-)\ln(\rho_i^+/\rho_i^-) \quad . \tag{6.13}$$

Observe that each of the terms in the sum in Eq. (6.13) is always positive or zero and for this reason any solution to this equation, and the $2I$ equations in Eqs. (6.10), must satisfy $\rho_i^+ = \rho_i^-$ for all $i = 1, 2, \ldots I$, in addition to the relations $\ln(\rho_i^+\rho_i^-) = 2A$. This clearly shows that finding a solution to Eqs. (6.10) is ambiguous, because any such solution must satisfy $\rho_i^+ - \rho_i^- = 0$ for all i, which means that any solution to Eqs. (6.10) has to have an absolute density of zero everywhere in the unit cell.

In practice, Papoular et al. fix some value for A in Eqs. (6.10) and then find a solution for these equations. This way a solution to Eqs. (6.10) no longer has to satisfy $\rho_i^+ = \rho_i^-$ for all $i = 1, 2, \ldots I$ and a meaningful solution can be obtained. Papoular et al. suggest ways of determining a value for A in Refs. [124],[125] in the light of the expression for A in Eq. (6.9). However, since this expression has the effect of making the system of equations in Eqs. (6.10) meaningless, then these estimates of A probably have limited value. The notation in Eqs. (6.10) is (partly) taken from Collins Ref. [126] and others (see references in Ref. [125]) and the observations made in this section also apply to these works.

With A considered a fixed MEMx constant, Eqs. (6.10) can be mapped onto the MEMx in Def. 7, Chap. 4. Consider the expression in Eq. (4.28), Chap. 4 with $\bar{M}^+ = \bar{M}^-$, and use $\bar{M}^+ = \bar{M}^- = \bar{M}^s/2$ (see Eqs. (4.5)) then Eq. (4.28), Chap. 4 reads

$$\text{logP} = \frac{v}{\epsilon}\sum_{i=1}^{I}[m_i^+ + m_i^- - m_i^+\ln(\frac{2Vm_i^+}{\bar{M}^s}) - m_i^-\ln(\frac{2Vm_i^-}{\bar{M}^s})] - \frac{1}{2}Q \tag{6.14}$$

where m_i^+ and m_i^- are the separate positive and negative magnetization densities

of the i^{th} bin. The extremum of Eq. (6.14) is found by solving the $2I$ equations

$$\begin{aligned} 0 &= -\ln(2Vm_i^+/\bar{M}^s) - (\epsilon/(2v))(\partial Q/\partial m_i^+) \\ 0 &= -\ln(2Vm_i^-/\bar{M}^s) - (\epsilon/(2v))(\partial Q/\partial m_i^-) \quad . \end{aligned} \tag{6.15}$$

Identifying $\epsilon/(2v)$ with $\lambda_0[\sum_{j=1}^{I}(\rho_j^+ + \rho_j^-)]$ in Eqs. (6.10) and $\bar{M}^s/(2V)$ with $\exp(A)$ it is seen that these two systems of equations are identical provided A is treated as a constant in Eqs. (6.10). Further, we know from the discussion in Sec. 4.3, Chap. 4 that finding the solution to Eqs. (6.15) is equivalent to solving for[7]

$$0 = -\sinh^{-1}(Vm_i/\bar{M}^s) - (\epsilon/(2v))(\partial Q/\partial m_i) \ , \ i = 1, 2, \ldots, I \ , \tag{6.16}$$

where $m_i = m_i^+ - m_i^-$ and provided Q only depends on the m_i's and that it is only the m_i's that we are aiming to evaluate. Solving for Eqs. (6.16) returns the output of the MEMx in Def. 7, Chap. 4 and it has hereby been shown that by choosing a value of $\bar{M}^s$ in Eqs. (6.16) which matches A in Eqs. (6.10) [8] then these two MEMx's are identical. This will be illustrated in the next section, where the reconstruction generated using the Papoular et al. algorithm in Dobrzynski et al. is reproduced.

6.4 Nickel data in Dobrzynski et al.

Table 6.1 lists 28 unique form factors (or structure factors) of nickel. Nickel is FCC (space group $Fm\bar{3}m$) and its form factors and structure factors are related by $F_{hkl}^{obs} = 4\langle\mu_{\text{Ni}}\rangle f_{hkl}^{obs}$, where $\langle\mu_{\text{Ni}}\rangle$ is the number of Bohr magnetons per Ni atom at room temperature. Using the numbers in Mook [117] then $\langle\mu_{\text{Ni}}\rangle = 0.606\mu_B \cdot 0.946 \simeq 0.5733\mu_B$. In comparison Maniawski et al. [118] measured the total magnetic moment of nickel to be $\langle\mu_{\text{Ni}}\rangle = 0.579(5)\mu_B$ by bulk magnetometry.

[7] Eqs. (6.16) is the derivative of Eq. (4.26), Chap. 4 with respect to m_i and with $\bar{M}^d = 0$.

[8] The Q values automatically match because both MEMx methods use $Q = P$.

The added standard deviation for the zero-q data point in table 6.1 is included to reflect the discrepancy between the values of total magnetic moment of Mook and Maniawski et al. and the errorbar ($0.005\mu_B$) of Maniawski et al.. A test was carried out in where a smaller errorbar was used, in order to determine whether or not this would make a difference to the reconstruction; $\sigma_{000} = 0.001$ was used instead of $\sigma_{000} = 0.01$ and no differences great enough to affect the comparisons drawn to the papers by Dobrzynski et al. were observed. The remaining 27 form factors in table 6.1 are copied from Ref. [80], and these form factors are an average of the data of Mook [117] and Maniawski et al. [118].

By trying out different values of prior model constant, $\bar{M}^s$, it becomes clear that choosing a value near $\bar{M}^s = 0.00065\mu_B$ reproduces the reconstructions presented by Dobrzynski et al. using the algorithm of Papoular et al.. In Dobrzynski et al. [80],[81] three 2D contour plots show the magnetization density of nickel obtained using the Papoular et al. algorithm as two slices through the unit cell at $z = 0$ displaying the positive and negative density separately and one slice through $z = 1/4$ displaying the negative density. To make the comparison to the reconstruction with $\bar{M}^s = 0.00065\mu_B$ using MEMx in Def. 7 this reconstruction is displayed as follows: one contour plots shows a slice through the unit cell at $z = 0$ (top frame in Fig. 6.4), another a slice at $z = 1/4$ (bottom frame in Fig. 6.3) and a third $z = 0$ but with the positive density values put to zero to emphasize negative density features[9]. There is excellent agreement between Fig. 6.3 and Fig. 6.4(top frame) and the plots in Dobrzynski et al. [80],[81] generated from the Papoular et al. code. This justifies the claim that with $\bar{M}^s = 0.00065\mu_B$ the

[9]Recall from the previous section that Papoular et al. calculate two densities for the unit cell (the same applies to Sakata et al.'s MEMx); one for the positive density and one for the negative density. When the $\sinh^{-1}$ prior is used (as is the case for the MEMx of Def. 7, Chap. 4) only one density is calculated which for the mode of the posterior is equal to the difference between the two densities used by Papoular et al. (providing matching prior model constants have been used).

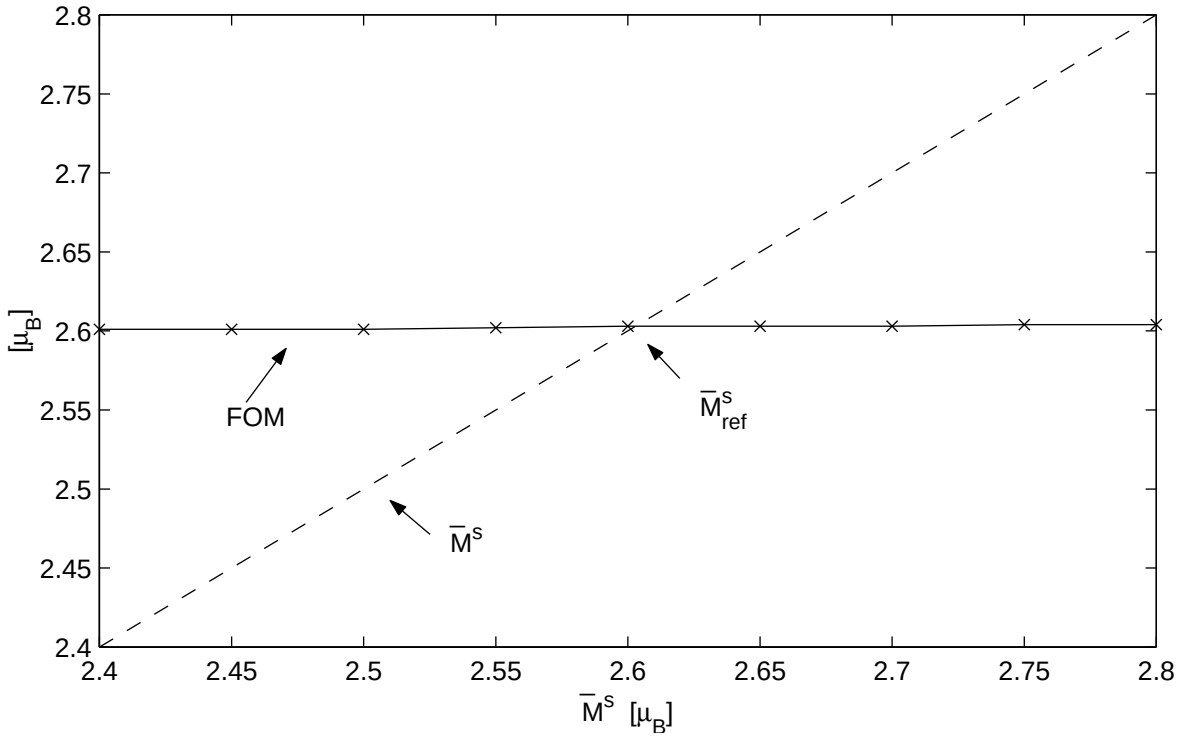

Figure 6.1: Shows a figure of merit $\mathrm{FOM}_1 = v \sum_{i=1}^{I} |m_i^{max}|$ versus $\bar{M}^s$ plot as the solid line for the nickel data in table 6.1 (Q=P). To help identify when FOM_1 is equal to $\bar{M}^s$, an $\bar{M}^s$ versus $\bar{M}^s$ plot is included as the dotted line.

plots using the Papoular et al. algorithm in Dobrzynski et al. are reproduced.

Choosing a value of $\bar{M}^s$ using the recipe in Sec. 4.7, Chap. 4 results in a value of $\bar{M}^s$ significantly different from $\bar{M}^s = 0.00065\mu_B$. With help from Fig. 6.1 the reference point for $\bar{M}^s$ is found to be approximately $\bar{M}^s{}_{ref} = 2.6\mu_B$ and therefore from the recipe of Sec. 4.7, Chap. 4 chose $\bar{M}^s \simeq \bar{M}^s{}_{ref}/4 = 0.65\mu_B$. Equivalent slices to those used to illustrate the $\bar{M}^s = 0.00065\mu_B$ reconstruction are shown for the reconstruction with $\bar{M}^s = 0.65\mu_B$ in Fig. 6.4(bottom frame) and Fig. 6.5. Observe in Fig. 6.4(bottom frame) the same main positive magnetization density features: Firstly, the non-spherical distribution of the nickel density pointing towards the corner atoms indicating an interaction between the nickel atoms. Secondly, the depletion of magnetization density at the Ni sites, as observed by

Dobrzynski et al.. By comparing the negative magnetization densities of the reconstruction with $\bar{M}^s = 0.00065\mu_B$ in Fig. 6.3 with $\bar{M}^s = 0.65\mu_B$ in Fig. 6.5 significant differences are clearly observed. These differences can be explained by the fact that a smaller value of $\bar{M}^s$ will tend to reduce the areas with non-zero density, i.e. to 'sharpen up' the features in the image (see Chap. 4). Therefore the $\bar{M}^s = 0.00065\mu_B$ reconstruction appears more 'spiky' than the $\bar{M}^s = 0.65\mu_B$ reconstruction as seen by comparing the two $z = 1/4$ contour plots. The mode of the posterior with $\bar{M}^s = 0.65\mu_B$ is also illustrated by the 4 isosurfaces in Fig. 6.2. The advantage of visualizing the reconstruction as a series of isosurfaces is that a better overview of the positions and shapes of the individual features in the unit cell is given. The isosurface plot with contour level (CL) values $\pm 0.02\mu_B \mathring{A}^{-3}$ shows that the largest negative feature is situated at the crystallographic site $(1/4, 1/4, 1/4)$ in the unit cell. It first appears at CL $\simeq -0.028\mu_B$ (see also Fig. 6.5(bottom frame)) which is almost a factor of 100 less than the largest positive feature of the image which occurs at CL $= 1.83\mu_B$. For isosurfaces below CL $= -0.02\mu_B \mathring{A}^{-3}$ the negative magnetization density takes on a complicated pattern in the interstitial regions of the nickel atoms, and it is approximately uniformly distributed. This supports the theoretical model proposed by Mook (and Moon see reference in Mook [117]) in which the negative magnetization density is treated as a uniform distribution throughout the unit cell.

With the Bayesian image analysis method of Def. 7 error estimates of selected regions of the unit cell can be calculated. Of interest is the negative density feature centered at $(1/4, 1/4, 1/4)$. Integration over the bins in the cell corresponding to the volume between 0.17969 and 0.32031, along the directions of the three axes, gives the total moment $-0.00154 \pm 0.00388\mu_B$ for that volume. According to this error estimate it is almost insignificant[10]. Integrating instead over all bins with negative magnetization density gives the total negative mo-

[10]The negative feature is repeated 4 times in the unit cell. Hence taking the crystal symmetry into account, the error estimate for this feature is $0.00388/\sqrt{4}\mu_B = 0.00194\mu_B$.

ment $-0.1519 \pm 0.02233\mu_B$, which is significant. Modelling the negative density as a uniform distribution; Mook [117] found the negative moment of the unit cell to be $-0.105\mu_B$, which is close to $-0.1519\mu_B$ considering the differences in approach and also the differences between the structure factors listed in Dobrzynski et al. [80] and those used by Mook.

The mode of the posterior (reconstruction) with $\bar{M}^s = 0.65\mu_B$ is also plotted in 'data space' (i.e. as structure factors) in the plot to the left hand side in Fig. 6.6. As expected from the selected $\bar{M}^s$ value (see Chap. 4) a smooth transition is seen between the 'observed' calculated structure factors and the 'unobserved' calculated structure factors in the region close to $q \simeq 14\mathring{A}^{-1}$. It is of interest in relation to the results in Sec. 4.4.4-4.4.5 because the data in table 6.1 provide an example of a 'good quality' data set containing a large number of data points with small σ_i/F_i^{obs} ratios. Say, on average, $\sigma/F^{obs} < 0.05$. Applying Eq. (4.36) it is expected that the mode of the posterior will be robust to changes of Q for Q values smaller than about $Q = P(0.1 * 0.05)^2 = 112$. This was confirmed by examining the reconstructions with $Q = 0.1$, $Q = 1$, $Q = 28$ and $Q = 112$, keeping $\bar{M}^s = 0.65\mu_B$. As an example the reconstructions using $Q = 0.1$ and $Q = 28$ are shown in 'data space' in Fig. 6.6. In particular observe the almost exact agreement between the F_i's in both plots in Fig. 6.6 for $q > 14\mathring{A}^{-1}$. The values of these calculated structure factors are determined almost exclusively by the value of the prior model constant $\bar{M}^s$ for all $Q < 112$ in the case of the nickel data in table 6.1.

6.5 Conclusions of chapter

A discussion of the MEMx algorithms used in Dobrzynski et al. was presented and compared to the MEMx of Chap. 4. This gave a fine illustration of some of the problems of the MEMx literature and the content of this chapter may help overcome some of these difficulties. The magnetization density of nickel was also

presented using a value for $\bar{M}^s$ selected according to the recipe in Sec. 4.7. These results mainly confirmed the observations made by Dobrzynski et al. [80],[81], Mook [117] and Maniawski et al. [118] but, provided the accuracy of the data, these results gives additional insight into the overall picture of the magnetization density of nickel.

Miller indices				
h	k	l	f_{hkl}^{obs}	σ_{hkl}
0	0	0	1.0000	0.0100
1	1	1	0.7876	0.0040
2	0	0	0.6985	0.0040
2	2	0	0.4437	0.0033
1	1	3	0.3189	0.0033
2	2	2	0.3090	0.0026
4	0	0	0.1558	0.0026
3	3	1	0.1670	0.0026
4	2	0	0.1307	0.0026
2	2	4	0.1070	0.0033
1	1	5	0.0357	0.0033
3	3	3	0.1089	0.0026
4	4	0	0.0574	0.0033
5	3	1	0.0317	0.0033
6	0	0	-0.0244	0.0026
4	4	2	0.0522	0.0033
6	2	0	-0.0092	0.0033
3	3	5	0.0357	0.0033
2	2	6	0.0059	0.0033
4	4	4	0.0370	0.0033
1	1	7	-0.0469	0.0033
5	5	1	0.0092	0.0033
6	4	0	-0.0007	0.0033
6	4	2	0.0007	0.0033
7	3	1	-0.0264	0.0033
5	5	3	0.0119	0.0033
8	0	0	-0.0627	0.0033
3	3	7	-0.0165	0.0033

Table 6.1: 28 unique magnetic structure factors for nickel which have been scale such that the zero-q structure factor is one. The total moment of the unit cell is estimated to $2.2931\mu_B$. See text for further description.

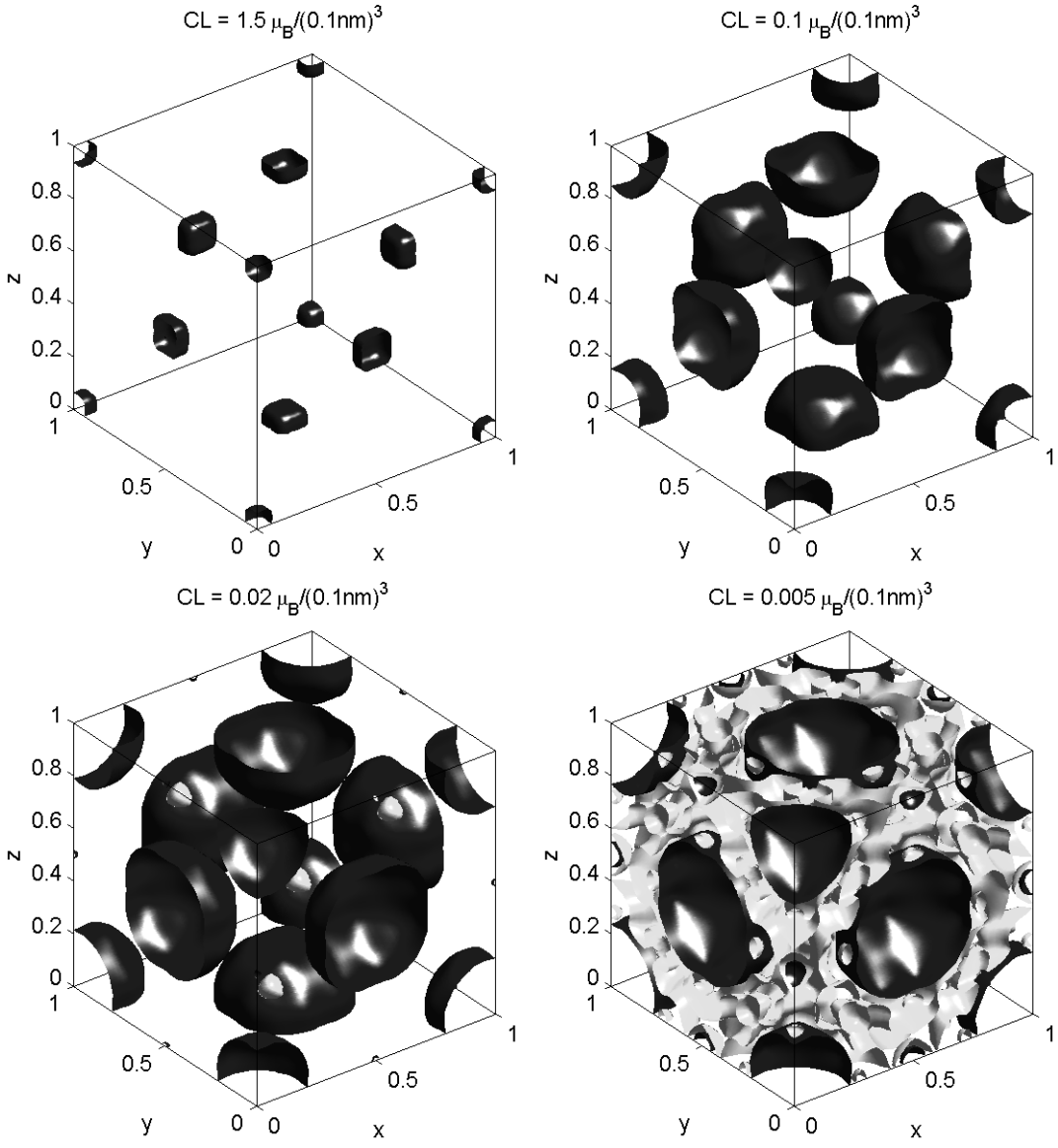

Figure 6.2: Reconstruction with $\bar{M}^s = 0.65\mu_B$ $(Q = P)$ illustrated at 4 different contour levels (CL). The lighter coloured features show isosurfaces at -CL.

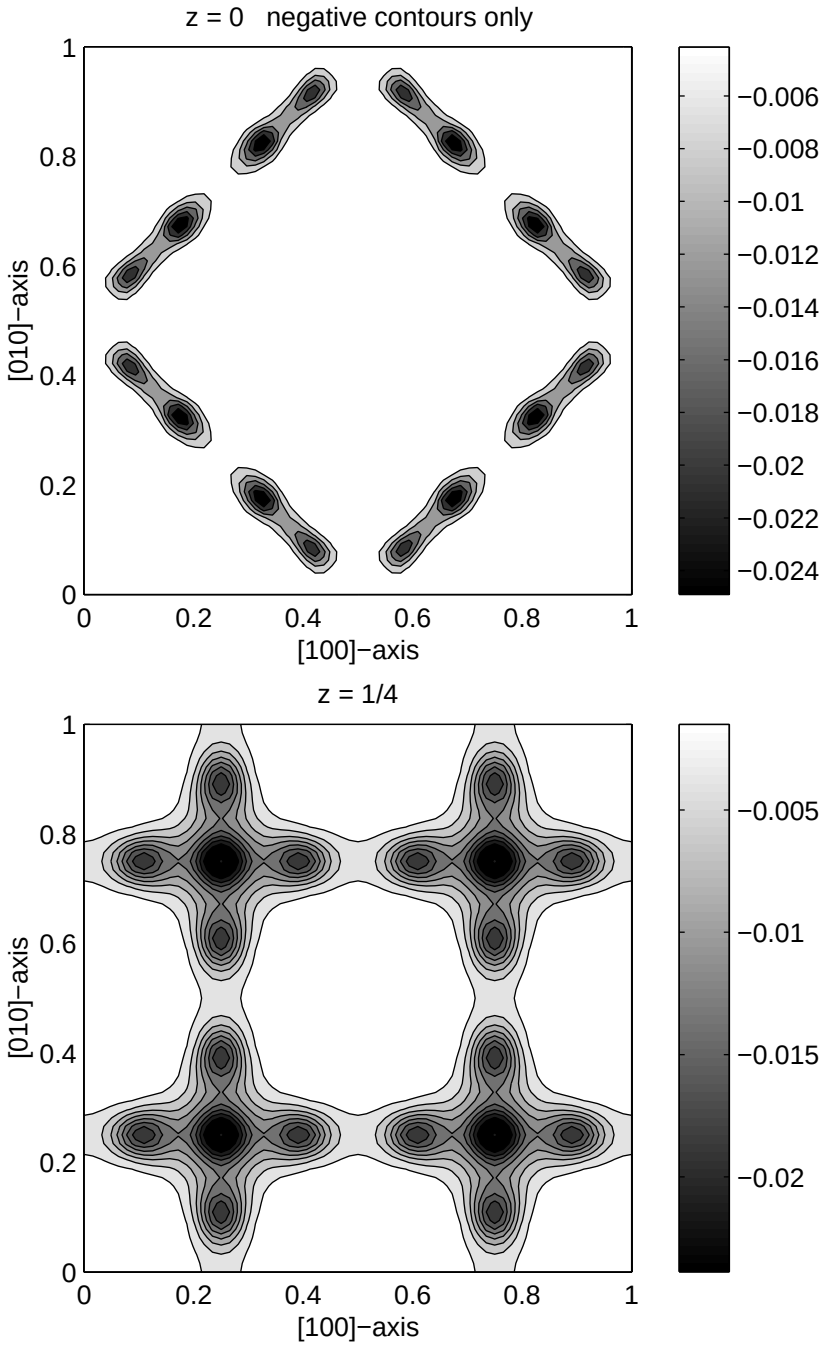

Figure 6.3: Reconstruction with $\bar{M}^s = 0.00065\mu_B$ ($Q = P$). Illustrated here with two slices through the unit cell at $z = 0$ and $z = 1/4$. In the top plot all positive magnetization density has been put to zero to emphasize the negative magnetization density in the $z = 0$ plane. The unit of the colour code bars is $\mu_B \mathring{A}^{-3}$.

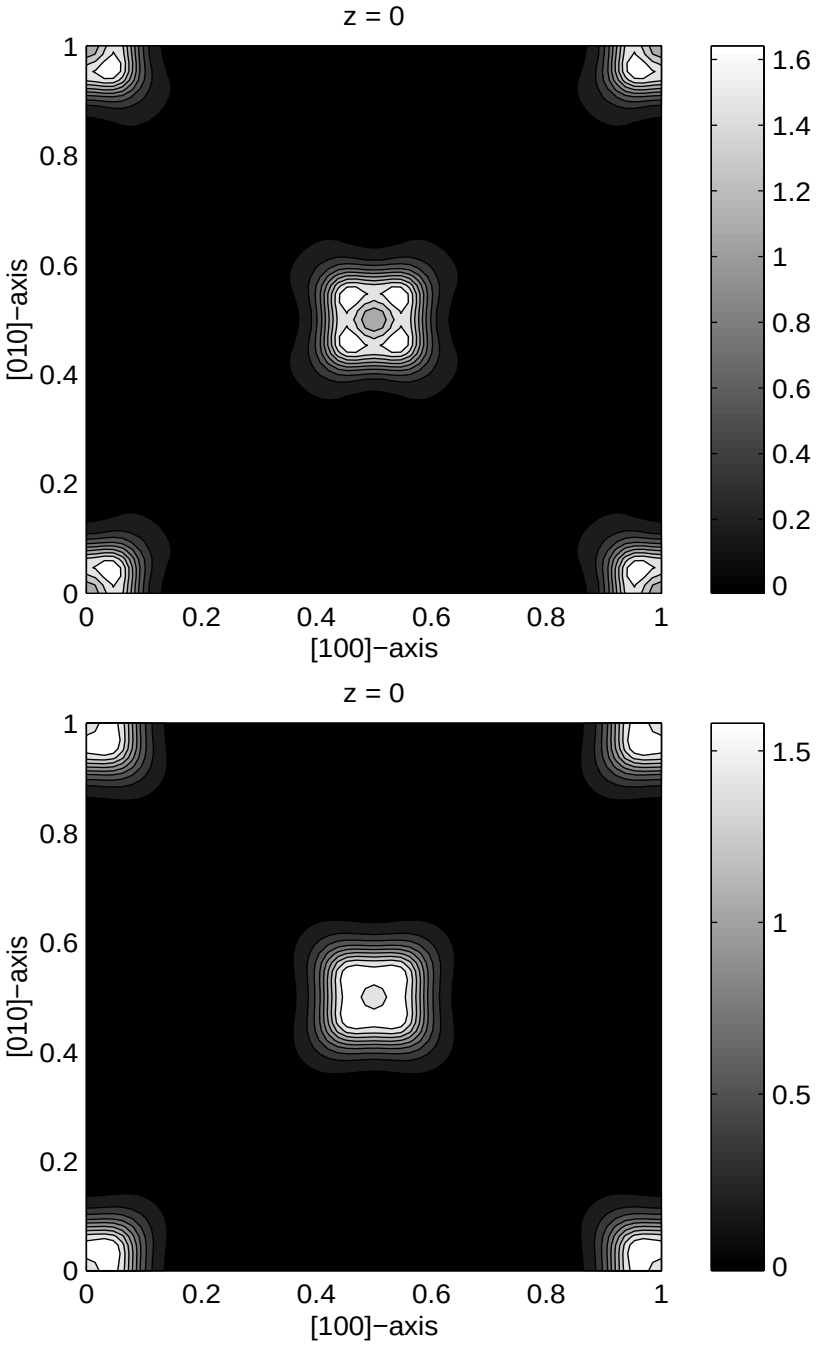

Figure 6.4: Top plot shows reconstruction with $\bar{M}^s = 0.00065\mu_B$ and bottom plot with $\bar{M}^s = 0.65\mu_B$. Both show the slice through the unit cell at $z = 0$. The unit of colour code bars is $\mu_B \mathring{A}^{-3}$.

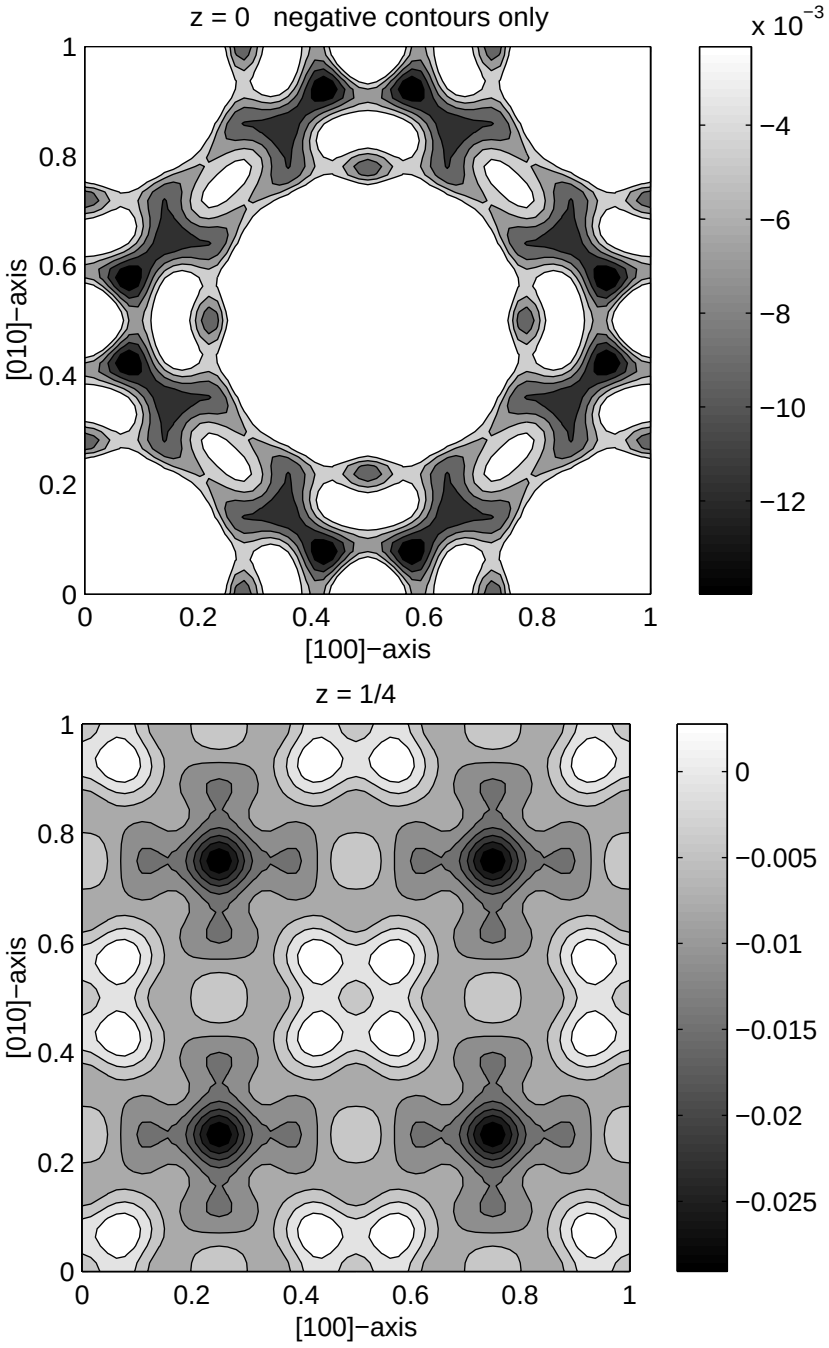

Figure 6.5: Reconstruction with $\bar{M}^s = 0.65\mu_B$, illustrated as in Fig. 6.3.

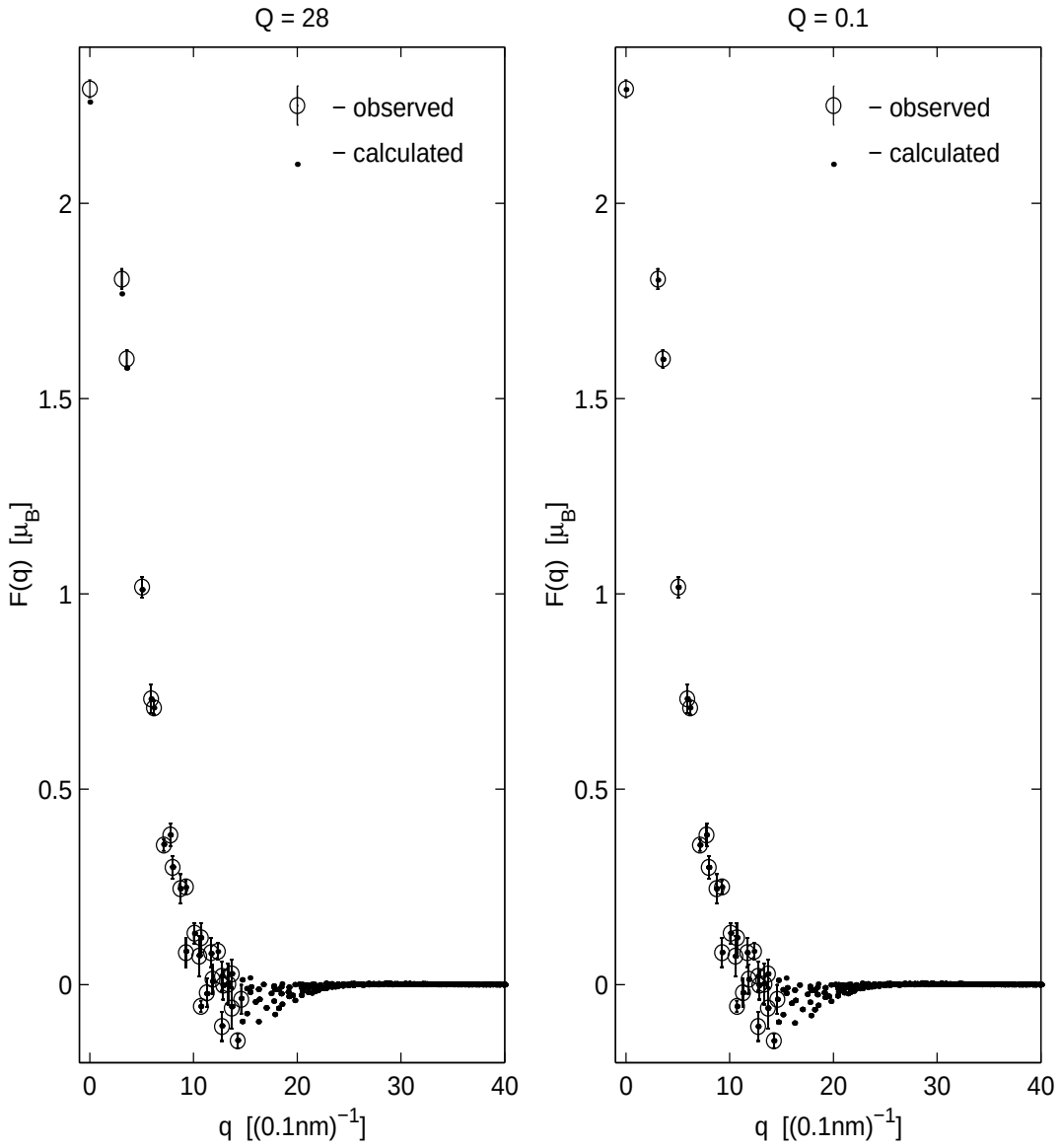

Figure 6.6: In the left-hand side plot the mode of the posterior (reconstruction) with $\bar{M}^s = 0.65\mu_B$ and $Q = P = 28$ is shown in 'data space' as calculated structure factors (the dots). In the right-hand side plot the same reconstruction is shown but with $Q = 0.1$. In both figures the observed data are shown as the circles with errorbars.

Conclusions and further work

The main accomplishments of this thesis are 1) the mapping of the induced magnetization density in $PrBa_2Cu_3O_{6+x}$; 2) An improved recipe for applications of the MaxEnt technique to PND data and possibly other types of data with a similar information content.

A brief summary of each of these two contributions is as follows:

1. In Chap. 5 the induced magnetization density of $PrBa_2Cu_3O_{6+x}$ revealed a number of interesting new features which are summarized in Sec. 5.6. One of these features is the small 'deformation' of the Pr induced density, the possibility that this is caused by a hybridisation between the Pr $4f$ electrons and electronic states in the CuO_2 planes has yet to be fully justified, but the possibility exists for work to be carried out in order to investigate this further. This further investigation could include;

   (a) Measurement of structure factors further out in reciprocal space, these may help define the structure of the existing Pr 'deformations' with greater detail.

   (b) The calculation of the magnetization density under the experimental conditions for existing and new proposed theoretical models. This would enable a detailed comparison between theoretical models and the experimentally observed magnetization density, and such calculations should be feasible to do with the current available computer power.

(c) An experiment, as described end of Sec. 5.6, but on a 'superconducting' single crystal of $PrBa_2Cu_3O_{6+x}$. A comparison with the results in this thesis on 'non-superconducting' $PrBa_2Cu_3O_{6+x}$ could help explain why some samples are superconducting and others not.

A rich structure is also revealed in the region around the Ba site. Except for the positive feature at the Ba, which shows evidence of Pr/Ba cation mixing, the origin of the remaining features is uncertain, but they may contain crucial information for the understanding of why $PrBa_2Cu_3O_7$ for standard grown single crystals do not superconduct.

2. The MaxEnt method was critically analysed in Chap. 4. A Bayesian robustness analysis was performed to understand how the choice of prior model constants affects the output. The results are summarized in Sec. 4.7 and provide the basis for an improved MaxEnt method, tuned specifically to the analysis of PND data. In Chap. 6 MaxEnt algorithms from the literature are compared using a practical example with PND data and here for PND data of Nickel. Chap. 6 provides a detailed picture of these algorithms and how they relate to the work in Chap. 4.

   The program used for the MaxEnt reconstructions is freely available. At present it uses a rather primitive interface and therefore is not very user friendly, however this will be improved in the future. Meanwhile, the code for finding the mode of the posterior in Eq. (4.24) (equal to MaxEnt output) is normally easy to write (see Sec. 4.4.1), it is more difficult to write code for a program that can handle all possible space group symmetries together with a proper graphical user interface. If a serious attempt is made to produce such a program in the future the end product will remain freely available for academics.

Appendices

A Boltzmann's reasoning in the limit $N_0 \to \infty$

In this appendix we look more closely at the consequences of the Boltzmann statistical model given in Def. 5. The purpose is to present the intuitive reasons why as well as proof, that as $N_0 \to \infty$, the probability of the gas being in any finite interval around Boltzmann's most probable macro-state tends to unity. For more details see Ref. [27] and references therein.

Let $(x_1, x_2, \ldots, x_{N_0})$ be a sequence of r.v. each with range $\mathcal{X} = \{a_1, a_2, \ldots, a_I\}$. We will use the notation $\mathbf{x}$ to denote a sequence $x_1, x_2, \ldots, x_{N_0}$, and we will also refer to $\mathbf{x}$ as a micro-state; in this context think of the elements of the set $\mathcal{X}$ as the possible phase-space cells a molecule can be positioned in: $a_1 =$ cell 1, $a_2 =$ cell 2 etc. As in the main text, define n_i to be equal to the number of variables in $\mathbf{x}$ which take on the value a_i; e.g. if x_j represents the j^{th} molecule and a_i the i^{th} phase-space cell, then n_i is the number of molecules in cell i. Let $\mathbf{n_x}$ be the macro-state corresponding to a particular micro-state $\mathbf{x}$, and let $\mathcal{N}_{N_0}$ be the total number of possible macro-states which can be formed from N_0 r.v. $x_1, x_2, \ldots, x_{N_0}$. The upper limit for $\mathcal{N}_{N_0}$ is $(N_0 + 1)^I$, since there are I components in the vector $\mathbf{n}_x$ and each component can take on only $N_0 + 1$ values. So there are at most $(N_0 + 1)^I$ choices for the macro-state vector. Of course, these choices are not quite independent, for instance the sum of the n_i's must add up to N_0, but a sufficiently good upper bound here is $\mathcal{N}_{N_0} \leq (N_0 + 1)^I$. Define $T(\mathbf{n})$ to be the

collection of all micro-states having $\mathbf{n}$ as their macro-state. Mathematically we write this macro-class as $T(\mathbf{n}) = \{\mathbf{x} \in \mathcal{X}^{N_0} : \mathbf{n}_x = \mathbf{n}\}$. Hopefully it is clear that the number of micro-states in each macro-class is just $|T(\mathbf{n})| = N_0!/(n_1! \cdots n_I!)$, i.e. a multinomial coefficient. As has been shown in Eq. (3.12) when N_0 is large $|T(\mathbf{n})| \simeq e^{N_0 H_e(\frac{\mathbf{n}}{N_0})}$, or more accurately it can be shown that (see e.g. [27])

$$\frac{1}{(N_0+1)^I} e^{N_0 H_e(\frac{\mathbf{n}}{N_0})} \leq |T(\mathbf{n})| \leq e^{N_0 H_e(\frac{\mathbf{n}}{N_0})} \ . \tag{A.1}$$

An important argument in this appendix is that as N_0 rises there is only a polynomial increase in the number of possible macro-states $\mathcal{N}_{N_0}$, whereas the number of micro-states in each macro-class increases exponentially with $\sim e^{N_0 H_e}$. The set of all possible micro-states can be imagined as a staircase, where each step represents one of the many possible values that $H_e(\mathbf{n})$ can take and where the height of each step is scaled by the number of micro-states having entropy $H_e(\frac{\mathbf{n}}{N_0})$. The bottom step will then represent the micro-states for which all the r.v. $x_1, x_2, \ldots x_{N_0}$ take on the same value (e.g. $x_1 = a_1, x_2 = a_1, \ldots, x_{N_0} = a_1$) and therefore $H_e = 0$; the height of this step is I micro-states. As we move up in entropy we move up the staircase and the top step represent the macro-class of micro-states where H_e is maximum; the height of this step is $|T(\mathbf{n}^{bol})|$. We can now use this staircase to visualize what happens as N_0 rises. The number of steps will increase approximately as $(N_0 + 1)^I$, and at the same time the distances between the steps will increase, such that, as N_0 becomes larger the number of micro-states in the step below will start to become vanishingly small compared to the number of micro-states in the step above it. In the mathematical limit $N_0 \to \infty$ we have: "The ratio of the number of micro-states in any finite neighbourhood of steps to the number of micro-states in any other finite region of steps laying above it goes towards zero as N_0 goes to infinity."

To justify the above picture we use the following theorem (for a proof see e.g. Ref. [27])

Theorem 1 *Let $\mathcal{P}$ be the set of all possible macro-states which can be formed*

for a fixed value of N_0. Members of this set(or space if you like) are all macrostates $\mathbf{n} = (n_1, n_2, \ldots, n_I)$, where each of the I coordinates n_i can take the values $n_i = 0, 1, \ldots, N_0$ and satisfy $\sum_{i=1}^{I} n_i = N_0$. Let $\mathcal{E}$ be a subset of $\mathcal{P}$ and in general $\mathcal{E} = \left\{\mathbf{n} : \sum_{a\in\mathcal{X}} g(a)n(a) \geq \alpha\right\}$. The maximum entropy in the subset $\mathcal{E}$ is $H_e^{max} = \max_{\mathbf{n}\in\mathcal{E}} H_e(\frac{\mathbf{n}}{N_0})$. The claim is then that for every fixed δ

$$Pr\left(|\mathbf{n} - \mathbf{n}^{bol}| < \delta \,\middle|\, \mathbf{n} \in \mathcal{E}\right) \to 1 \tag{A.2}$$

as N_0 goes to infinity.

B The modified Bessel prior

Consider for simplicity and without loss of generality the case where we have just one bin, then

$$p(n^+, n^-|\bar{N}^+, \bar{N}^-) = \frac{(\bar{N}^+)^{n^+}}{n^+!}\frac{(\bar{N}^-)^{n^-}}{n^-!}e^{-(\bar{N}^+ + \bar{N}^-)} \quad . \tag{B.1}$$

Make the change of variables $n = n^+ + n^-$ and $q = n^+ - n^-$. Since the map between the space spanned by (n^+, n^-) and (n, q) is one-to-one and onto, the distribution in Eq. (B.1) in terms of the coordinates (n, q) is obtained by substituting the parameters (n, q) directly into that equation, thus

$$p(n, q|\bar{N}^+, \bar{N}^-) = \left(\frac{\bar{N}^+}{\bar{N}^-}\right)^{q/2} \frac{(\bar{N}^+\bar{N}^-)^{n/2}}{[(n+q)/2]![(n-q)/2]!}e^{-(\bar{N}^+ + \bar{N}^-)} \quad , \tag{B.2}$$

where $q = \ldots, -1, 0, 1, \ldots$ and $n = |q|, |q| + 2, |q| + 4, \ldots$. Introduce the dummy index

$$\begin{aligned} l &= 1/2(n - q) \quad \text{when } q \geq 0 \\ l &= 1/2(n + q) \quad \text{when } q < 0 \end{aligned} \quad . \tag{B.3}$$

If q is positive or zero, then in terms of (l, q) Eq. (B.2) becomes

$$p(l, q|\bar{N}^+, \bar{N}^-) = \left(\frac{\bar{N}^+}{\bar{N}^-}\right)^{q/2} \frac{(\bar{N}^+\bar{N}^-)^{l+q/2}}{(l+q)!l!}e^{-(\bar{N}^+ + \bar{N}^-)} \quad , \tag{B.4}$$

where $l = 0, 1, 2, \ldots$ and $q = \ldots, -1, 0, 1, \ldots$. Noting that a modified Bessel function can be expressed in term of the ascending series $\mathrm{I}_q(z) = \sum_{l=0}^{\infty}(z/2)^{2l+q}/(\Gamma(l+1+q)l!)$, we see that marginalising over l in Eq. (B.4) gives,

$$p(q|\bar{N}^+, \bar{N}^-) = \sum_{l=0}^{\infty} p(l, q|\bar{N}^+, \bar{N}^-) = \left(\frac{\bar{N}^+}{\bar{N}^-}\right)^{q/2} \mathrm{I}_q(2\sqrt{\bar{N}^+\bar{N}^-})e^{-(\bar{N}^+ + \bar{N}^-)} \ . \quad \text{(B.5)}$$

For negative q we also arrive at the distribution in the above equation, and Eq. (B.5) is the distribution in Eq. (4.4), Sec. 4.1.

Since the moments of the modified Bessel distribution in Eq. (B.5) appear not to have been calculated in the literature, the first two moments will be calculated here.

Changing to the prior model constants $\bar{N}^s = \bar{N}^+ + \bar{N}^-$ and $\bar{N}^d = \bar{N}^+ - \bar{N}^-$, the aim is to calculate

$$\langle q \rangle = \sum_{q=-\infty}^{\infty} q \left(\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d}\right)^{q/2} \mathrm{I}_q(\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}})e^{-\bar{N}^s} \quad . \quad \text{(B.6)}$$

For notational convenience we introduce $z = \sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}$ and $\kappa = \sqrt{\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d}}$ and make use of the modified Bessel relation $\mathrm{I}_q(z) = \frac{z}{2q}[\mathrm{I}_{q-1}(z) - \mathrm{I}_{q+1}(z)]$. Eq. (B.6) then implies

$$e^{\bar{N}^s}\langle q \rangle = \sum_{q=-\infty}^{\infty} \kappa^q \frac{z}{2}[\mathrm{I}_{q-1}(z) - \mathrm{I}_{q+1}(z)] = \left[\kappa - \frac{1}{\kappa}\right] \frac{z}{2} \sum_{q=-\infty}^{\infty} \kappa^q \mathrm{I}_q(z) \ . \quad \text{(B.7)}$$

Obviously the sum $\sum_{q=-\infty}^{\infty} \kappa^q \mathrm{I}_q(z)$ is equal to the normalization constant $e^{\bar{N}^s}$ of the distribution (this is also clear from the well-known relation $\exp[\frac{z}{2}(\kappa + \frac{1}{\kappa})] = \sum_{q=-\infty}^{\infty} \kappa^q \mathrm{I}_q(z)$), and Eq. (B.7) implies

$$\langle q \rangle = \left[\sqrt{\frac{\bar{N}^s + \bar{N}^d}{\bar{N}^s - \bar{N}^d}} - \sqrt{\frac{\bar{N}^s - \bar{N}^d}{\bar{N}^s + \bar{N}^d}}\right] \frac{\sqrt{\bar{N}^{s\,2} - \bar{N}^{d\,2}}}{2} = \bar{N}^d \ . \quad \text{(B.8)}$$

Hence the mean value of the distribution is $\bar{N}^d$.

Next we aim to calculate the second moment of Eq. (B.5),

$$\langle q^2 \rangle = \sum_{q=-\infty}^{\infty} q^2 \kappa^q \mathrm{I}_q(z) e^{-\bar{N}^s} \ . \quad \text{(B.9)}$$

Using $\mathrm{I}_q(z) = \frac{z}{2q}[\mathrm{I}_{q-1}(z) - \mathrm{I}_{q+1}(z)]$ in succession we find that

$$q^2\mathrm{I}_q(z) = \frac{z^2}{4}[\mathrm{I}_{q-2}(z) + \mathrm{I}_{q+2}(z) - 2\mathrm{I}_q(z)] + \frac{z}{2}[\mathrm{I}_{q-1} + \mathrm{I}_{q+1}(z)] \ . \tag{B.10}$$

Inserting the last term in Eq. (B.10) into Eq. (B.9) and comparing with the step from Eq. (B.7) to Eq. (B.8) simply gives

$$\sum_{q=-\infty}^{\infty} \kappa^q \frac{z}{2}[\mathrm{I}_{q-1}(z) - \mathrm{I}_{q+1}(z)] = \bar{N}^s \ . \tag{B.11}$$

Inserting the first term of Eq. (B.10) into Eq. (B.9) and using again $\sum_{q=-\infty}^{\infty} \kappa^q \mathrm{I}_q(z) = e^{\bar{N}s}$ gives

$$\sum_{q=-\infty}^{\infty} \kappa^q \frac{z^2}{4}[\mathrm{I}_{q-2}(z) + \mathrm{I}_{q+2}(z) - 2\mathrm{I}_q(z)] = \frac{z^2}{4}[\kappa^2 + \frac{1}{\kappa^2} - 2] = \bar{N}^{d^2} \ . \tag{B.12}$$

Summing up Eq. (B.11) and Eq. (B.12) we have

$$\langle q^2 \rangle = \bar{N}^{d^2} + \bar{N}^s \ , \tag{B.13}$$

and the variance of the distribution in Eq. (B.5) is then

$$\langle (q - \langle q \rangle)^2 \rangle = \langle q^2 \rangle - \langle q \rangle^2 = \bar{N}^s \ . \tag{B.14}$$

The above equation and Eq. (B.8) are Eqs. (4.7,4.8) in Sec. 4.1.

C Comment on Skilling & Bryan algorithm and notation by Skilling

In Sec. 4.4.1 we discussed various options for finding the mode of the posterior in Eq. (4.12) numerically. One of these options is to use the algorithm in Ref. [84] by Skilling & Bryan. They found that the optimization algorithm in Ref. [84] was improved by incorporating the idea of a metric into the search space. Using a metric in non-linear optimization problems is not uncommon, see for example Ref. [127]. It is stated in Ref. [84] that using a specific form for the metric "is the

single most important key to the development of a robust algorithm". It is also clear from the work in Ref. [84] that the metric is considered as a purely technical option used with the purpose of improving the overall computing speed.

Perhaps because of the emphasis placed on the choice of metric in Ref. [84], Skilling in Ref. [90] manages to include the metric in Ref. [84] into the expression for the prior function itself. What becomes unclear at this stage is whether this metric should now be interpreted as a real physical quantity or as a technical option as in the previous paper [84]. If it is a real quantity, then it must imply that Skilling's MEMx version in Ref. [90] is different from previous MEMx versions (including his own in Ref. [84]) and will result in a different output when applied. Since this is probably not the intention the prior in Ref. [90] is referred to as a notation.[11]

D Calculating $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ with Gaussian prior

Changing to reciprocal space coordinates and using $\sum_{i=1}^{I} m_i^2 = \frac{I}{V^2}\sum_{i=1}^{I}(F_i^{cal})^2$ the Gaussian prior in Eq. (4.33) becomes

$$p(\mathbf{F}^{cal}|\bar{M}^s, \epsilon) = \frac{1}{\sqrt{2\pi\epsilon\bar{M}^s}^I}\exp\left(-\frac{1}{2}\sum_{i=1}^{I}\frac{(F_i^{cal})^2}{\epsilon\bar{M}^s}\right) . \tag{D.1}$$

We aim to find $p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon)$ from the expression in Eq. (4.40),

$$\begin{aligned} p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon) &= \frac{1}{\sqrt{2\pi\epsilon\bar{M}^s}^I}\frac{1}{\sqrt{2\pi}^P\prod_{i=1}^{P}\sigma_i} \\ &\int \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\frac{(F_i^{cal})^2}{\epsilon\bar{M}^s} - \frac{1}{2}\sum_{i=1}^{P}\frac{(F_i^{obs}-F_i^{cal})^2}{\sigma_i^2}\right) d\mathbf{F}^{cal} . \end{aligned} \tag{D.2}$$

We can immediately integrate out the F's not in $\mathbf{F}^{obs}$ (these are the F's present in the first sum but not the second sum in the integrand in Eq. (D.2)) using

[11] To obtain a prior expression which includes a metric, Skilling in Ref. [90] follows a procedure which may be compared to parts of Sec. 4.1, but when stepping from a discrete to a continuous prior distribution he includes an extra term in the Stirling approximation to obtain a prior expression which is discussed below Eq. (4.17).

$\int_{-\infty}^{\infty} \exp(-ax^2)dx = \sqrt{\pi/a}$. Using the relation

$$\frac{(F_i^{cal})^2}{\epsilon\bar{M}^s} + \frac{(F_i^{obs} - F_i^{cal})^2}{\sigma_i^2} = \frac{\sigma_i^2 + \epsilon\bar{M}^s}{\epsilon\bar{M}^s\sigma_i^2}\left(F_i^{cal} - \frac{\epsilon\bar{M}^s F_i^{obs}}{\sigma_i^2 + \epsilon\bar{M}^s}\right)^2 + \frac{(F_i^{obs})^2}{\sigma_i^2 + \epsilon\bar{M}^s} \ ,$$

the remaining F's can likewise be eliminated by integration to obtain

$$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon) = \frac{1}{\sqrt{2\pi}^P \prod_{i=1}^P \sqrt{\sigma_i^2 + \epsilon\bar{M}^s}} \exp\left(-\frac{1}{2}\sum_{i=1}^P \frac{(F_i^{obs})^2}{\sigma_i^2 + \epsilon\bar{M}^s}\right) \quad , \qquad \text{(D.3)}$$

equal to Eq. (4.43).

When the unit cell has a space group symmetry other than the trivial space group symmetry[12] the number of independent variables is reduced. Denote by I_{sym} the number of calculated symmetry unique structure factors , and s_i the number of symmetry equivalent structure factors of each unique structure factor. Then in reciprocal coordinates (structure factor coordinates) the prior including symmetry constraints is

$$p(\mathbf{F}^{cal}|\bar{M}^s, \epsilon, S_{group}) = \frac{1}{\prod_{i=1}^{\mathrm{I}_{sym}} \sqrt{2\pi\epsilon\bar{M}^s/s_i}} \exp\left(-\frac{1}{2}\sum_{i=1}^{\mathrm{I}_{sym}} \frac{s_i(F_i^{cal})^2}{\epsilon\bar{M}^s}\right) \quad , \qquad \text{(D.4)}$$

where S_{group} denotes the space group of the unit cell. Again we aim to find

$$\begin{aligned} p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon, S_{group}) &= \frac{1}{\prod_{i=1}^{\mathrm{I}_{sym}} \sqrt{2\pi\epsilon\bar{M}^s/s_i}} \frac{1}{\sqrt{2\pi}^{P_{sym}} \prod_{i=1}^{P_{sym}} \sigma_i/\sqrt{s_i}} \\ &\times \int \exp\left(-\frac{1}{2}\sum_{i=1}^{\mathrm{I}_{sym}} \frac{s_i(F_i^{cal})^2}{\epsilon\bar{M}^s} - \frac{1}{2}\sum_{i=1}^{P_{sym}} \frac{s_i(F_i^{obs} - F_i^{cal})^2}{\sigma_i^2}\right) d\mathbf{F}^{cal} \ , \qquad \text{(D.5)} \end{aligned}$$

where P_{sym} is equal to the number of unique observed structure factors. Now comparing the above integral with the integral in Eq. (D.2) we see that

$$p(\mathbf{F}^{obs}|\bar{M}^s, \epsilon, S_{group}) = \frac{1}{\sqrt{2\pi}^{P_{sym}} \prod_{i=1}^{P_{sym}} \sqrt{(\sigma_i^2 + \epsilon\bar{M}^s)/s_i}} \exp\left(-\frac{1}{2}\sum_{i=1}^{P_{sym}} \frac{s_i(F_i^{obs})^2}{\sigma_i^2 + \epsilon\bar{M}^s}\right) \ . \qquad \text{(D.6)}$$

[12] I.e. the group containing only the identity element. In crystallography this space group is P1.

If $\sigma_i = \sigma,\ \forall\ i = 1, 2, \ldots, P_{sym}$ then searching for the point at which the derivative of Eq. (D.6) with respect to $\epsilon\bar{M}^s$ is zero we obtain

$$\frac{1}{2}\frac{\sum_{i=1}^{P_{sym}} s_i(F_i^{obs})^2}{(\sigma^2 + \epsilon\bar{M}^s)^2} - \frac{P_{sym}}{2}\frac{1}{\sigma^2 + \epsilon\bar{M}^s} = 0 \tag{D.7}$$

which implies

$$\epsilon\bar{M}^s = \epsilon^2\bar{N}^s = \frac{1}{P_{sym}}\sum_{i=1}^{P_{sym}}(s_i(F_i^{obs})^2 - \sigma^2) = \frac{1}{P_{sym}}\sum_{i=1}^{P}(F_i^{obs})^2 - \sigma^2\ , \tag{D.8}$$

which is Eq. (4.45).

Comment

If, instead of the Gaussian prior in Eq. (4.33), we had used

$$p(\mathbf{m}|\epsilon, \bar{M}^s, \bar{M}^d) = \frac{1}{\sqrt{2\pi\epsilon\bar{M}^s/(v^2 I)}^I}\exp\left(-\frac{v^2 I}{2\epsilon\bar{M}^s}\sum_{i=1}^{I}(m_i - \bar{M}^d/V)^2\right) \tag{D.9}$$

to calculate $p(\mathbf{F}^{obs}|\bar{M}^s, \bar{M}^d, \epsilon)$, we would have obtained exactly the same expression as in Eq. (D.6) but with the zero-q structure factor substituted by $F_0 - \bar{M}^d$. Hence for most cases according to $p(\mathbf{F}^{obs}|\bar{M}^s, \bar{M}^d, \epsilon)$ we should select $\bar{M}^d = 0$ when the zero-q structure factor is absent and $\bar{M}^d = F_0$ when it is present. The latter choice, we know from Sec. 4.4.6, results in a reconstruction which is worse than the reconstruction obtained using $\bar{M}^d = 0$.

E Calculating $|\mathbf{\Sigma}|$

We can write $\mathbf{A}^T\mathbf{\Sigma}_0^{-1}\mathbf{A} = (\mathbf{\Sigma}_0^{-1/2}\mathbf{A})^T(\mathbf{\Sigma}_0^{-1/2}\mathbf{A})$ and

$$\mathbf{\Sigma}^{-1} = \mathbf{D}^{1/2}(\mathbf{M}^T\mathbf{M} + \mathbf{I})\mathbf{D}^{1/2}\ , \tag{E.1}$$

where $\mathbf{M} = \mathbf{\Sigma}_0^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ and is of size $P \times I$. Introduce the $I \times I$ orthogonal matrix $\mathbf{U}$ whose columns are the eigenvectors of $\mathbf{M}^T\mathbf{M}$, and write Eq. (E.1) as

$$\mathbf{\Sigma}^{-1} = \mathbf{D}^{1/2}\mathbf{U}(\mathbf{U}^T\mathbf{M}^T\mathbf{M}\mathbf{U} + \mathbf{I})\mathbf{U}^T\mathbf{D}^{1/2}\ , \tag{E.2}$$

where $\mathbf{U}^T\mathbf{M}^T\mathbf{M}\mathbf{U}$ is the eigenvalue matrix of $\mathbf{M}^T\mathbf{M}$. Denote these eigenvalues by $\lambda_1, \lambda_2, \ldots, \lambda_I$. When $I > P$, then clearly the symmetric and positive definite matrix $\mathbf{M}^T\mathbf{M}$ is singular and its non-zero eigenvalues are identical to the eigenvalues of the $P \times P$ symmetric matrix $\mathbf{M}\mathbf{M}^T$. Therefore the determinant of the matrix in Eq. (E.2) is

$$|\mathbf{\Sigma}^{-1}| = \prod_{i=i}^{I} \mathbf{D}_{ii} \prod_{i=1}^{P} (\lambda_i + 1) \ , \tag{E.3}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_P$ are the non-zero eigenvalues of the $I \times I$ matrix $\mathbf{M}^T\mathbf{M}$, which are identical to the eigenvalues of the $P \times P$ matrix $\mathbf{M}\mathbf{M}^T$. Hence we can calculate the determinant of $\mathbf{\Sigma}$ as a $P \times P$ eigenvalue problem. It is very rare that more than about 150 structure factors are measured in a PND experiment. Therefore calculating $|\mathbf{\Sigma}|$ and $p(\mathbf{F}^{obs}|\bar{M}^S, \epsilon)$ in Eq. (4.41) takes little time on a standard PC or more powerful computer.

F Calculating $\mathbf{\Sigma}$

Not surprisingly, the calculation of $\mathbf{\Sigma}$ is more computer intensive than the calculation of $|\mathbf{\Sigma}|$ in App. E, nevertheless it is still manageable. In calculating the matrix $\mathbf{\Sigma}$ from $\mathbf{\Sigma}^{-1}$ the procedure outlined in [51] and in chapter 3 in [16] will be followed.

Taking the inverse of $\mathbf{\Sigma}^{-1}$ in Eq. (E.2) gives

$$\mathbf{\Sigma} = \mathbf{D}^{-1/2}\mathbf{U}(\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{U}^T\mathbf{D}^{-1/2} \ , \tag{F.1}$$

where $\mathbf{\Lambda}$ is the eigenvalue matrix $\mathbf{U}^T\mathbf{M}^T\mathbf{M}U$. Now add and subtract an $I \times I$ identity matrix in Eq. (F.1)

$$\mathbf{\Sigma} = \mathbf{D}^{-1/2}\mathbf{U}[\mathbf{I} - (\mathbf{\Lambda} + \mathbf{I})(\mathbf{\Lambda} + \mathbf{I})^{-1} + (\mathbf{\Lambda} + \mathbf{I})^{-1}]\mathbf{U}^T\mathbf{D}^{-1/2} \ . \tag{F.2}$$

The identical matrix is added and subtracted in the square bracket. Eq. (F.2) reduces to

$$\mathbf{\Sigma} = \mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{U}[\mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{I})^{-1}]\mathbf{U}^T\mathbf{D}^{-1/2} \ . \tag{F.3}$$

When $I > P$ $\mathbf{\Lambda}$, is singular and at most P of its I eigenvalues are non-zero. The diagonal matrix in the square bracket of Eq. (F.3) has the diagonal elements $\frac{\lambda_1}{\lambda_1+1}, \frac{\lambda_2}{\lambda_2+1}, \ldots, \frac{\lambda_P}{\lambda_P+1}$ as its first P diagonal elements, the rest being equal to zero. We may therefore write Eq. (F.3) as

$$\mathbf{\Sigma} = \mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{U}_{IP}[\mathbf{\Lambda}_{PP}(\mathbf{\Lambda}_{PP} + \mathbf{I}_{PP})^{-1}]\mathbf{U}_{IP}^T\mathbf{D}^{-1/2} \ , \tag{F.4}$$

where, for example, $\mathbf{\Lambda}_{PP}$ is the submatrix of $\mathbf{\Lambda}$ consisting of the first P rows and P columns. Hence it is possible to solve for $\mathbf{\Sigma}$ as a $I \times P$ SVD-problem. A well-known theorem of linear algebra says that, any $I \times P$ matrix, $\mathbf{M}^T$, with $I \geq P$, can be written as the product of an $I \times P$ column-orthogonal matrix $\mathbf{U}_{IP}$, a $P \times P$ diagonal matrix $\mathbf{\Lambda}_{PP}^{1/2}$ and the transpose of an orthogonal matrix $\mathbf{V}$,

$$\mathbf{M}^T = \mathbf{U}_{IP}\mathbf{\Lambda}_{PP}^{1/2}\mathbf{V}^T \ . \tag{F.5}$$

Hence $\mathbf{M}\mathbf{M}^T = \mathbf{V}\mathbf{\Lambda}_{PP}\mathbf{V}^T$ and $\mathbf{M}^T\mathbf{M} = \mathbf{U}_{IP}\mathbf{\Lambda}_{PP}\mathbf{U}_{IP}^T$, and the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{M}\mathbf{M}^T$, and the columns of $\mathbf{U}_{IP}$ are the eigenvectors of the non-zero eigenvalues of $\mathbf{M}^T\mathbf{M}$. We see from Eq. (F.5) that $\mathbf{M}^T\mathbf{V} = \mathbf{U}_{IP}\mathbf{\Lambda}_{PP}^{1/2}$ and we may therefore rewrite $\mathbf{\Sigma}$ in Eq. (F.4) as

$$\mathbf{\Sigma} = \mathbf{D}^{-1} - \mathbf{D}^{-1/2}\mathbf{M}^T\mathbf{V}(\mathbf{\Lambda}_{PP} + \mathbf{I}_{PP})^{-1}\mathbf{V}^T\mathbf{M}\mathbf{D}^{-1/2} \ . \tag{F.6}$$

The second term in Eq. (F.6) clearly takes the most time to compute. Calculating the eigenvalues and eigenvectors of $\mathbf{M}\mathbf{M}^T$ takes of the order $o(P^3)$ computer operations. A matrix multiplication like $\mathbf{M}^T\mathbf{V}$ takes of the order $o(IP^2)$ operations and the matrix multiplication of $\mathbf{M}^T\mathbf{V}$ with $(\mathbf{\Lambda}_{PP} + \mathbf{I}_{PP})^{-1}\mathbf{V}^T\mathbf{M}$ of the order $o(I^2P)$[13]. Thus when $I > P$ the most computer intensive part is not the calculation of eigenvalues and eigenvectors, but is the matrix multiplication. However, often we will only be interested in calculating the diagonal elements of $\mathbf{\Sigma}$, in which case even when I is significantly larger than P the cost of computing $\mathbf{\Sigma}$ is

[13]Of course we can exploit the fact the final matrix is symmetric, even so it will still take of the order $o(I^2P)$ computer operations.

simply a case of finding the eigenvalues and eigenvectors of a symmetric matrix, this is very similar to the cost of calculating $|\mathbf{\Sigma}|$ in the previous appendix.

Bibliography

[1] A. T. Boothroyd, J. Alloy. Compd. **303**, 489 (2000).

[2] J. P. Burg, Maximum entropy spectrum analysis, Oklahoma City (1967), paper presented at the 37th Ann. Intern. Meeting Soc. of Explor. Geophys.; reprinted in [35].

[3] G. L. Squires, *Introduction to the theory of thermal neutron scattering* (Cambridge University Press, Cambridge, 1978).

[4] S. W. Lovesey, *Theory of neutron scattering from condensed matter* (Oxford University Press, Oxford, 1987).

[5] W. G. Williams, *Polarised neutrons* (Clarendon Press, Oxford, 1988).

[6] G. Casella and R. L. Berger, *Statistical Inference* (Wadsworth & Brooks/Cole, Pacific Grove, California, 1990).

[7] T. M. Apostol, *Calculus*, 2 ed. (John Wiley & Sons, Singapore, 1969), Vol. 2.

[8] C. P. Robert, *The Bayesian choice*, *Springer text in statistics* (Springer-Verlag, New York, 1994).

[9] J. M. Bernardo and A. F. M. Smith, *Bayesian theory*, *Wiley series in probability and mathematical statistics* (John Wiley & Sons, Chichester, 1994).

[10] J. A. Rice, *Mathematical statistics and data analysis*, 2 ed. (Duxbury Press, Belmont, Califonia, 1995).

[11] M. H. Degroot, *Probability and statistics*, 2 ed. (Addison-Wesley Pub, Reading, Mass, 1986).

[12] J. O. Berger, *Statistical decision theory and Bayesian analysis*, *Springer series in statistics*, 2 ed. (Springer-Verlag, New York, 1985).

[13] *Encyclopedia of statistical sciences*, edited by K. Samuel, N. L. Johsson, C. B. Read, and D. L. Banks (John Wiley & Sons, New York, 1982–), vol. 1-9, update vol. 1-2.

[14] D. S. Sivia, *Data analysis : a Bayesian tutourial* (Clarendon Press, Oxford, 1996).

[15] J. C. Johnston, Ph.D. thesis, Oxford University, Trinity term 1996.

[16] V. A. Macaulay, Ph.D. thesis, Oxford University, Trinity term 1992.

[17] S. M. Ross, *Introduction to probability models*, 4 ed. (Academic Press, Boston, 1989).

[18] A. Tarantola, *Inverse problem theory* (Elsevier, Amsterdam, 1987).

[19] A. Papoulis, *Probability, random variables, and stochastic processes*, 3 ed. (McGraw-Hill, New York, 1991).

[20] D. J. C. Mackay, Ph.D. thesis, California institute of technology, 1992.

[21] E. T. Jaynes, fragments available from the author: Wayman Crow Professor of Physics, Washington University, St. Louis, 63130, USA. *ftp://bayes.wustl.edu/* (unpublished).

[22] W. T. Grandy, *Foundations of statistical mechanics*, No. 19 in *Fundamental theories of physics* (Reidel, Dordrecht, 1987).

[23] M. Plischke and B. Bergersen, *Equilibrium statistical physics*, 2 ed. (World Scientific, Singapore, 1994).

[24] R. Balian, *From microphysics to macrophysics* (Springer-Verlag, Berlin; London, 1991), Vol. 1.

[25] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[26] C. E. Shannon, Bell System Tech. J. **27**, 623 (1948).

[27] T. M. Cover and J. A. Thomas, *Elements of information theory*, *Wiley series in telecommunications* (John Wiley & Sons, New York, 1991).

[28] E. T. Jaynes, in *Statistical physics*, edited by K. W. Ford (W. A. Benjamin, Inc., New York, 1963), Vol. 3, p. 182, reprinted in [31], p. 39.

[29] E. T. Jaynes, IEEE Trans. Syst. Sci. Cybern **4**, 227 (1968).

[30] S. Kullback, *Information theory and statistics*, *Dover books on mathematics* (Dover, Mineola, New York, 1997).

[31] E. T. Jaynes, in *Papers on probability, statistics and statistical physics*, *Synthese library*, edited by R. D. Rosenkrantz (Reidel, Dordrecht, 1983).

[32] L. Boltzmann, *Lectures on gas theory* (Dover, New York, 1995), translated by S. G. Brush.

[33] K. Huang, *Statistical mechanics*, 2 ed. (John Wiley & Sons, New York, 1987).

[34] C. E. Shannon, Bell System Tech. J. **27**, 379 (1948).

[35] D. G. Childers, *Modern spectrum analysis* (IEEE Press, New York, 1978).

[36] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).

[37] J. Skilling and S. F. Gull, SIAM-AMS Proceedings **14**, 167 (1984).

[38] R. Gordon, R. Bender, and G. T. Herman, J. Theor. Biol. **29**, 471 (1970).

[39] R. Gordon and G. T. Herman, Quarterly Bull. Center for Theor. Biol., SUNY at Buffalo **4(1)**, 71 (1971).

[40] B. R. Frieden, J. Opt. Soc. Am. **62**, 511 (1972).

[41] S. Goldman, *Information Theory* (Prentice-Hall, New York, 1955), appendix II.

[42] B. R. Frieden, in *Picture proceeding and digital filtering*, Vol. 6 of *Topics in applied physics*, edited by T. S. Huang and H. C. Andrews (Springer-Verlag, Berlin, 1975), p. 177, there also exist a 2nd corrected and revised edition of this reference: Same title and authors, Springer-Verlag, Berlin, 1979.

[43] B. R. Frieden, Science **191**, 1237 (1976).

[44] L. R. D'Addario, Maximum a posteriori probability and maximum entropy reconstruction, Stanford (1975), presented at Image processing for 2D and 3D reconstruction from projection: Theory and practice in medicine and the physical sciences meeting.

[45] L. R. D'Addario, Proceedings of the Society of Photo Optical Instrumentation Engineers **231**, 2 (1980).

[46] S. J. Wernecke and L. R. D'Addario, IEEE Trans. Comput. **C-26**, 351 (1977).

[47] D. S. Stephenson, Progress in NMR Spectroscopy **20**, 515 (1988).

[48] R. Narayan, Ann. Rev. Astron. Astrophys. **24**, 127 (1986).

[49] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, J. R. Statist. Soc. B **54**, 41 (1992).

[50] B. Buck and V. A. Macaulay, in *Maximum entropy and Bayesian methods*, edited by R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer, Dordrecht, 1992), p. 241.

[51] V. A. Macaulay and B. Buck, Nuclear Phys. A **591**, 85 (1995).

[52] K. Maisinger, M. P. Hobson, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **290**, 313 (1997).

[53] A. W. Jones *et al.*, Mon. Not. R. Astron. Soc. **294**, 582 (1998).

[54] G. J. Daniell and P. J. Hore, J. Magn. Reson. **84**, 515 (1989).

[55] W. I. F. David, Nature **346**, 731 (1990).

[56] J. C. Hoch, A. S. Stern, D. L. Donoho, and I. M. Johnstone, J. Magn. Reson. **86**, 236 (1990).

[57] J. C. Hoch and A. S. Stern, *NMR data processing* (Wiley-Liss, Inc., New York, 1996).

[58] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate discrete distributions*, 2 ed. (John Wiley & Sons, New York, 1992).

[59] A. W. Kemp, in *Encyclopedia of statistical sciences*, edited by S. Kotz, C. B. Read, and D. L. Banks (John Wiley & Sons, New York, 1997), Vol. updata vol. 1, p. 93.

[60] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas* (jw, New York, 1972).

[61] H. Bateman and A. Erdélyi, *Higher transcendental functions* (McGraw-Hill, New York, 1953-55), Vol. 1-3.

[62] G. N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed. (University Press, Cambridge, 1966).

[63] D. M. Titterington, Astron. Astrophys **144**, 381 (1985).

[64] D. M. Titterington, Int. Statist. Rev. **53**, 141 (1985).

[65] E. D. Laue, J. Skilling, and J. Staunton, J. Magn. Reson. **63**, 418 (1985).

[66] R. J. Papoular and B. Gillon, Europhys. Lett. **13**, 429 (1990).

[67] M. Sakata, T. Uno, M. Takata, and C. J. Howard, J. Appl. Cryst. **26**, 159 (1993).

[68] E. D. Laue, M. R. Mayger, J. Skilling, and J. Staunton, J. Magn. Reson. **68**, 14 (1986).

[69] W. I. F. David, Inst. Phys. Conf. Ser. **107**, 93 (1990).

[70] R. J. Papoular and B. Gillon, in *Neutron scattering data analysis*, No. 107 in *Institute of Physics conference series*, edited by M. W. Johnson (Institute of Physics, Bristol, 1990), p. 101.

[71] P. J. Hore, *Nuclear magnetic resonance* (Oxford University Press, Inc., New York, 1995).

[72] P. J. Hore, D. S. Grainger, S. Wimperis, and G. J. Daniell, J. Magn. Reson. **89**, 415 (1990).

[73] J. A. Jones and P. J. Hore, J. Magn. Reson. **92**, 276 (1991).

[74] J. X. Boucherle *et al.*, Physica B **192**, 25 (1993).

[75] A. Vernière, J. X. Boucherle, P. Lejay, and B. Gillon, Physica B **267-8**, 41 (1999).

[76] A. Hiess *et al.*, J. Phys.:Condens. Matter **12**, 829 (2000).

[77] L. Paolasini *et al.*, J. Phys.: Condens. Matter **5**, 8905 (1993).

[78] A. Muñoz *et al.*, J. Phys.: Condens. Matter **7**, 8821 (1995).

[79] A. Hiess *et al.*, Physica B **230-232**, 687 (1997).

[80] L. Dobrzynski, R. J. Papoular, and M. Sakata, J. Phys. Soc. Jpn. **65**, 255 (1996).

[81] L. Dobrzynski, R. J. Papoular, and M. Sakata, J. Magn. Magn. Mater. **140-144**, 53 (1995).

[82] A. J. Markvardsen *et al.*, J. Magn. Magn. Mat. **177**, 502 (1998).

[83] W. H. Press, *Numerical recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[84] J. Skilling and R. K. Bryan, Mon. Not. R. astr. Soc. **211**, 111 (1984).

[85] K. Kubota, M. Takata, and M. Sakata, J. Appl. Cryst. **26**, 453 (1995).

[86] K. Burger, Powder Diffr. **13**, 117 (1998).

[87] S. Kumazawa, M. Takata, and M. Sakata, Acta Cryst. A **51**, 47 (1995).

[88] K.-B. Li, A. S. Stern, and J. C. Hoch, J. Magn. Reson. **134**, 161 (1998).

[89] A. J. Freeman and R. E. Watson, Phys. Rev. **127**, 2058 (1962).

[90] J. Skilling, in *Maximum entropy and Bayesian methods*, edited by J. Skilling (Kluwer Academic Pub., London, 1989), p. 45.

[91] S. F. Gull, in *Maximum entropy and Bayesian methods*, edited by J. Skilling (Kluwer Academic Pub., London, 1989), p. 53.

[92] S. F. Gull and J. Skilling, Technical report, Maximum Entropy Data Consultants Ltd, Meldreth, Royston SG8 6NR, England (unpublished).

[93] H. B. Radousky, J. Mater. Res. **7**, 1917 (1992).

[94] R. Fehrenbacher and T. M. Rice, Phys. Rev. Lett. **70**, 3471 (1993).

[95] A. T. Boothroyd *et al.*, Phys. Rev. Lett. **78**, 130 (1997).

[96] C. Wilkinson, H. W. Khamis, R. F. D. Stansfield, and G. J. McIntyre, J. Appl. Cryst. **21**, 471 (1988).

[97] A. Renault *et al.*, J. Physique **48**, 1407 (1987).

[98] G. J. McIntyre, A. Renault, and G. Collin, Phys. Rev. B **37**, 5148 (1988).

[99] A. T. Boothroyd, S. M. Doyle, and R. Osborn, Physica C **217**, 425 (1993).

[100] E. Balcar, J. Phys. C **8**, 1975 (1975).

[101] E. U. Condon and H. Odabaşi, *Atomic structure* (Cambridge University Press, New York, 1980).

[102] C. Stassis and H. W. Deckman, Phys. Rev. B **13**, 4934 (1976).

[103] G. B. Folland, *Fourier analysis and its applications* (Brooks/Cole Pub. Com., Pacific Grove, 1992).

[104] J. X. Boucherle *et al.*, Physica C **162-4**, 1285 (1989).

[105] J. Y. Henry *et al.*, Physica C **235-240**, 1659 (1994).

[106] B. Gillon, Physica B **174**, 340 (1991).

[107] B. Gillon *et al.*, Physica C **162-4**, 1275 (1989).

[108] B. Gillon *et al.*, Physica B **156-7**, 851 (1989).

[109] M. P. Nutley, A. T. Boothroyd, and G. J. McIntyre, J. Magn. Magn. Mat. **104-7**, 623 (1992).

[110] M. P. Nutley, Ph.D. thesis, Warwick University, March 1994.

[111] G. L. Goodman, C. K. Loong, and L. Soderholm, J. Phys.: Condens. Matter **3**, 49 (1991).

[112] H.-D. Jostarndt *et al.*, Phys. Rev. B **46**, 14872 (1992).

[113] G. Hilscler *et al.*, Phys. Rev. B **49**, 535 (1994).

[114] S. J. S. Lister *et al.*, Physica C **317-8**, 572 (1999).

[115] Z. Zou *et al.*, Phys. Rev. Lett. **80**, 1074 (1998).

[116] J. Ye *et al.*, Phys. Rev. B **58**, R619 (1998).

[117] H. A. Mook, Phys. Rev. **148**, 495 (1966).

[118] F. Maniawski, B. v. Laar, and S. Kaprzyk, J. Phys. F.: Metal Phys. **11**, 1895 (1981).

[119] M. Sakata and M. Sato, Acta. Cryst. A **46**, 263 (1992).

[120] M. Sakata, T. Uno, M. Takata, and R. Mori, Acta. Cryst. B **48**, 591 (1992).

[121] S. Kumazawa, Y. Kubota, M. Takata, and M. Sakata, J. Appl. Cryst. **26**, 453 (1993).

[122] R. J. Papoular, E. Ressouche, J. Schweizer, and A. Zheludev, in *Maximum entropy and Bayesian methods*, edited by A. Mohammad-Djafari and G. Demoments (Kluwer, Dordrecht, 1993), p. 311.

[123] R. J. Papoular and A. Delapalme, Phys. Rev. Lett. **72**, 1486 (1994).

[124] R. J. Papoular, Y. Vekhter, and P. Coppers, Acta. Cryst. **A52**, 397 (1996).

[125] R. J. Papoular, W. Prandl, and J. Schweizer, in *Maximum entropy and Bayesian methods*, edited by C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer, Dordrecht, 1992), p. 359.

[126] D. M. Collins, Nature **298**, 49 (1982).

[127] R. W. H. Sargent, in *Numerical methods for constrained optimization*, edited by P. E. Gill, W. Murray, and R. Fletcher (Academic Press, London, 1974), p. 149.