Extracting Ramachandran torsional angle distributions from 2D NMR data using Tikhonov regularization

by

Amirmohamad Keyvanloo Shahrestanaky

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Mathematics and Sciences

Department of Physics

BROCK UNIVERSITY

January 4, 2005

Abstract

Solid state nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for studying structural and dynamical properties of disordered and partially ordered materials, such as glasses, polymers, liquid crystals, and biological materials. In particular, two-dimensional(2D) NMR methods such as ^{13}C–^{13}C correlation spectroscopy under the magic-angle-spinning (MAS) conditions have been used to measure structural constraints on the secondary structure of proteins and polypeptides. Amyloid fibrils implicated in a broad class of diseases such as Alzheimer's are known to contain a particular repeating structural motif, called a β-sheet. However, the details of such structures are poorly understood, primarily because the structural constraints extracted from the 2D NMR data in the form of the so-called Ramachandran (backbone torsion) angle distributions, $g(\phi, \psi)$, are strongly model-dependent.

Inverse theory methods are used to extract Ramachandran angle distributions from a set of 2D MAS and constant-time double-quantum-filtered dipolar recoupling (CTDQFD) data. This is a vastly underdetermined problem, and the stability of the inverse mapping is problematic. Tikhonov regularization is a well-known method of improving the stability of the inverse; in this work it is extended to use a new regularization functional based on the Laplacian rather than on the norm of the function itself. In this way, one makes use of the inherently two-dimensional nature of the underlying Ramachandran maps. In addition, a modification of the existing numerical procedure is performed, as appropriate for an underdetermined inverse problem.

Stability of the algorithm with respect to the signal-to-noise (S/N) ratio is examined using a simulated data set. The results show excellent convergence to the true angle distribution function $g(\phi, \psi)$ for the S/N ratio above 100.

Contents

List of Figures

Acknowledgements

I would like to thank my supervisor, Prof. Edward Sternin, for the help and support he has given me throughout this work. I thank Prof. S. K. Bose for giving me so much motivation and encouragement throughout the courses I had with him, and, as one of my committee members, for his assistance. I appreciate Prof. B. Mitrovic for teaching me many body theory and for his guidance as a member of my committee. I am also grateful to Prof. K. Samokhin , a member of my committee, for useful discussions.

I gratefully acknowledge prof. R. Tycko for providing simulated 2D MAS exchange and CTDQFDR data to us. My special thanks to Dr. H. Schaefer for sharing the codes and for great discussions. I also appreciate my external examiner, Prof. D. Weliky, for careful reviewing of my thesis and for many valuable discussions. I thank Prof. J. S. Hartman for very important suggestions about the details.

I thank F. Benko and F. Boseglav and F. Meffe for their kindful support. Also, I would like to acknowledge S. Sadeghi for reviewing of my thesis and for useful suggestions.

I am grateful to my parents, Parvaneh Azari and Ebrahim Keyvanloo, and my wife, Fatemeh Haghighat, for their constant patience, support, and encouragement.

Chapter 1

Introduction

Solid-state nuclear magnetic resonance (NMR) can be used to study the structural and dynamical parameters of disordered and partially ordered materials such as glasses, polymers, liquid crystals, and biological materials. These parameters are not sharply defined but characterized by probability distribution functions instead. Such distributions are mapped into distributions of NMR parameters, such as the chemical shift, the dipolar and the quadrupolar interaction, relaxation, spin diffusion, and exchange rates and others. Then, the structural and dynamical information contained in NMR spectra can be extracted by means of "inverse theory". A wide variety of inversion techniques have been developed and applied to solid-state NMR [1, 2, 3, 4, 5]. The extraction of a distribution in one parameter from NMR data, which may be multi-dimensional, is concerned with one-dimensional inverse problems. On the other hand, two-dimensional inverse problems have been addressed to determine distributions in two parameters like T_1 & T_2 relaxation times in porous materials [6], torsion angles and orientational angles in polymers [7], chemical shift and quadrupolar parameters in inorganic disordered materials [8]. The extraction of the Ramachandran angles distribution function, $g(\phi, \psi)$, is an example of two-dimensional inverse problems that we discuss in this thesis. Here, we discuss conformational properties of polypeptide chains very briefly and then introduce the Ramachandran plot.

Various three-dimensional arrangements of atoms that are inter-convertible without breaking covalent bonds are generally described as **conformations**. Different conformations of a protein molecule must have the same chirality of atoms, *i.e.* the same **configuration**. Three-dimensional aspects of structure are especially important for macromolecules, in which many bonds can rotate and can make many conformations. Macromolecules tend in general to be very flexible, so no one conformation predominates. Proteins, on the other hand, have used this flexibility to adopt relatively fixed conformations that are determined by non-covalent interactions among atoms that are distant in the covalent structure. There is a hierarchy of levels of protein structure. The **primary structure** is the covalent structure, which is defined by the amino acid (residue) sequence. The **secondary structure**, which is the topic of our discussion, is the local conformation of the polypeptide backbone and exhibits the folding of the polypeptide chain in space. The **tertiary structure** is the folded conformation of proteins.

How to define a conformation is not obvious. Even a simple molecule might be considered to exist in an infinite number of conformations if the positions of the atoms are defined with sufficient accuracy, because bond lengths vary by $\pm\, 0.05\, A°$ and bond angles by about $\pm\, 5°$ at 0°K temperature. For this reason, only the energetically most stable arrangements are usually classified as individual conformations. In the case of proteins, each amino acid residue in a polypeptide chain can exist in a number of conformations, perhaps eight on average. Therefore, a small polypeptide chain of 100 residues might be able to adopt up to

8^{100} conformations. Some of these theoretical conformations are not allowed because they would have atoms overlapping in space (the excluded volume effect). It is not yet possible to calculate the number of conformations that are indeed possible, and only rough estimates can be given. Since so many conformations are possible, the conformational properties of random polypeptides are best calculated statistically using the mathematical procedures developed for synthetic polymers. Such calculations require detailed knowledge of the conformational properties of the monomeric unit of the polymer; *i.e.* the relative energies of all its possible conformations.

A portion of the backbone of a polypeptide chain is shown in Fig. 1.1, illustrating the conventions used in describing polypeptide conformation. The peptide bond is usually planar

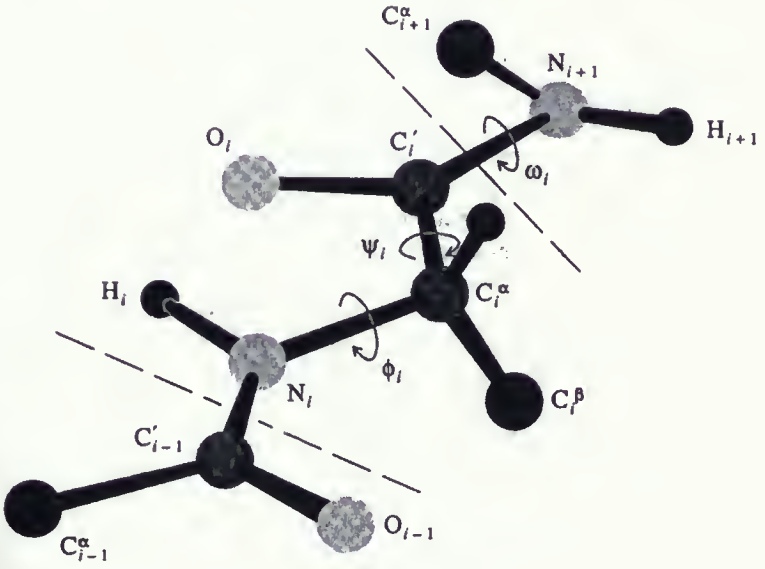


Figure 1.1: Perspective drawing of a segment of polypeptide chain comprising two peptide units. The limits of a single residue (number i of the chain) are indicated by the dashed lines. The polypeptide chain is shown in the fully extended conformation, where $\phi = \psi = \omega = 180°$. Reproduced from [9].

because of its partial double bond, and the group of atoms shown in Fig. 1.2 usually acts as a rigid unit. For this reason, this group is often designated a **peptide unit**. The unit more

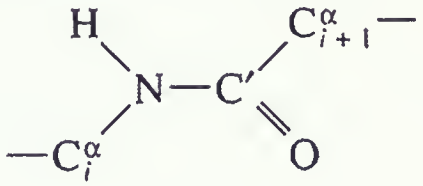


Figure 1.2: Peptide unit. Reproduced from [9].

commonly used is the **residue**, in which all the atoms originate from the same amino acid.

Rotations about the bonds are described as **torsion** or **dihedral** angles, which usually lie in the range $-180°$ to $+180°$. Rotation about the N$-$C$^{\alpha}$ bond of the peptide backbone is denoted by the torsion angle ϕ, rotation about the C$^{\alpha}-$C$'$ bond by ψ and that about the peptide bond (C$'-$N) by ω. The maximum value of 180° is given to each of the torsion angles in the maximally extended chain, as shown in Fig. 1.1, when the N, C$^{\alpha}$ and C$'$ atoms are all *trans* to each other.

Two best known and most easily recognized secondary structures of the polypeptides are the right-handed α-helix and the β-sheets. Fig. 1.3 shows the structure of right-handed α-helix. This residue has 3.6 residues per turn and a translation per residue of 1.50 A°,

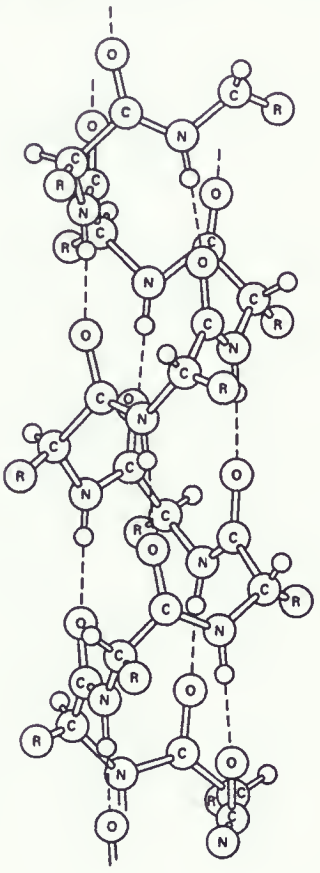


Figure 1.3: The classical right-handed α-helix. Reproduced from [9].

which gives a translation of 5.41 A° per turn. After the α-helix, the second most regular and identifiable conformation adopted by homopolypeptides is the β-sheet, Fig. 1.4. The basic unit is the β-**strand**, with the polypeptide almost fully extended; this can be considered a special type of helix with 2.0 residues per turn and a translation of 3.4 A° per residue. The β-strand conformation is stable only when incorporated into a β-**sheet**. There are several other type of conformations for polypeptides, such as the left-handed α-helix, the right-handed π-helix and others.

The possible values of the torsion angles, ϕ and ψ, are constrained geometrically due to steric clashes between non-neighboring atoms. The permitted values of ϕ and ψ were

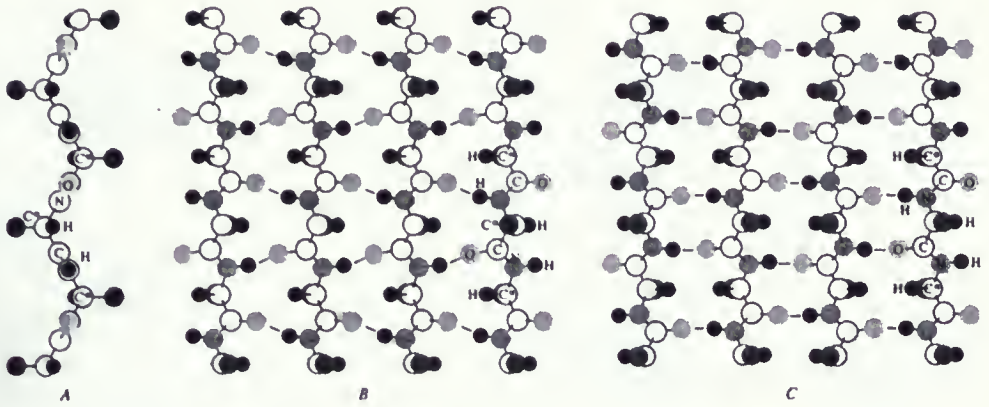


Figure 1.4: A single β-strand (A) and its incorporation into flat parallel (B) and antiparallel (C) β-sheets. Reproduced from [9].

first determined by Ramachandran and colleagues, using hard-sphere models of the atoms and fixed geometries of the bonds. The permitted values of ϕ and ψ are usually indicated on a two-dimensional map of the $\phi - \psi$ plane, called **Ramachandran plot**. For instance, Fig. 1.5 A illustrates the Ramachandran plot for an Ala(nine) residue. The normally allowed values, for which there is no steric overlap, are shaded and the regions enclosed by a solid line are the partially allowed regions. The connecting regions enclosed by the dashed lines are permissible with slight flexibility of bond angles. The part of the total area that is fully allowed is about 7.5 % and the part that is partially allowed is about 22.5% , which gives a quantitative measure of the limitations on flexibility of the polypeptide chain. Gly(cine) residues have no C^{β} atom and so the restrictions on allowed conformations are much less severe, as shown in Fig. 1.5 B. The fully allowed area is 45% of the total area and 61% area is partially allowed. Fig. 1.6 shows a Ramachandran plot on which the positions of different types of the polypeptides conformations are illustrated. The β-sheet conformation of peptides and proteins plays an important role in the strucural investigations of Alzheimer's plaque fibrils. Amyloid fibrils are filamentous structures, with typical diameters of 10nm and lengths up to several micrometers, formed by numerous peptides and proteins with different sequences and molecular weights. Biomedical interest in amyloid fibrils arises from their occurrence in amyloid diseases, including Alzheimer's disease, type 2 diabetes, Huntington's disease and prion disease. X-ray fiber diffraction shows that amyloid fibrils contain cross-β strutural motifs, *i.e.* extended β-sheets in which the β-strand segment run approximately perpendicular to, and the intermolecular hydrogen bonds run approximately parallel to, the long axis of the fibril. In the case of fibrils formed by the full-length β-amyloid peptide associated with Alzheimer's disease ($A\beta_{1-40}$), several molecular models have been proposed. A strucural model for $A\beta_{1-40}$ that is recently developed based on experimental constraints from solid state NMR [10] is shown in Fig. 1.7. Also, Fig. 1.8 a illustrates transmission electron microscope images of $A\beta_{1-40}$ fibrils. In Fig. 1.8 (b-d) 2D $^{13}C-^{13}C$ chemical shift correlation spectrum of $A\beta_{1-40}$ fibril is shown. Backbone torsion angles ϕ and ψ in $A\beta_{1-40}$ samples have been investigated and the presence of non-β-strand conformations has been detected. The data obtained through experiments are compared with the simulations based

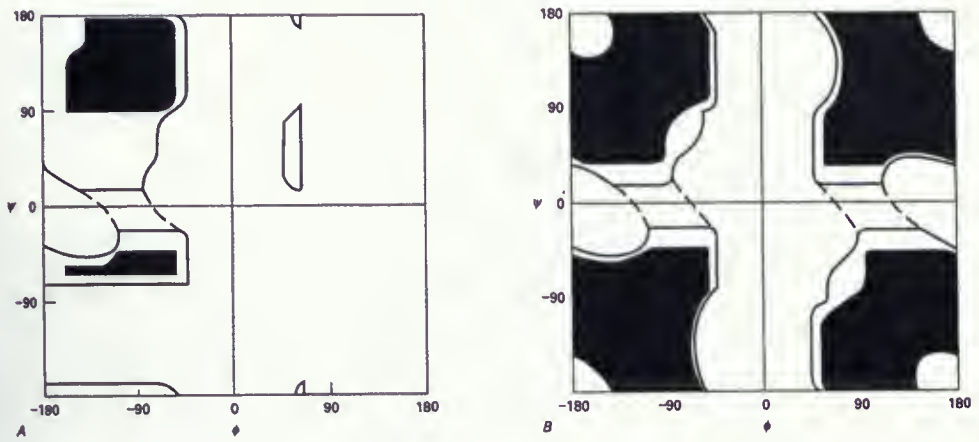


Figure 1.5: Ramachandran plots of the permitted values of ϕ and ψ angles. Each two-dimensional plot is continuous at the edges, because a rotation of $-180°$ is the same as one of $+180°$. The original plots that considered only repulsions between hard-sphere atoms are shown in A and B for Ala and Gly residues, respectively. Reproduced from [9].

on "constant-time finite-pulse radio-frequency-driven recoupling" (fpRFDR-CT), "double-quantum chemical shift anisotropy" (DQCSA) and "two-dimensional magic-angle spinning exchange" (2D MAS exchange). Simulated data are obtained for single pairs of ϕ and ψ angles and are fitted on top of the spectrum Fig. 1.9. The discrepancy between the measured data and the simulated data, specially in fpRFDR-CT measurements, Fig. 1.9 a, for each pair of (ϕ,ψ) indicates that a distribution of torsion angles, $g(\phi, \psi)$, is needed to best characterize the secondary structure of peptides and proteins. We adopt a different approach to extract this distribution. instead of fitting the calculated data to the measured spectrum, we will try to extract the possible orientations of ϕ and ψ angles through inversion. For this purpose, one can calculate the basis spectra by employing suitable solid state NMR techniques and construct the kernel $K(\omega; \phi, \psi)$, each column in the kernel is a basis spectra. Then the data, $s(\omega)$ can be measured through experiments that correspond to these NMR techniques. Finally, by applying appropriate inversion techniques, the distribution of the torsion angles, $g(\phi, \psi)$ can be obtained by solving the following equation

$$s(\omega) = \int_{\psi} \int_{\phi} K(\omega; \phi, \psi)\, g(\phi, \psi)\, d\phi\, d\psi\,. \tag{1.1}$$

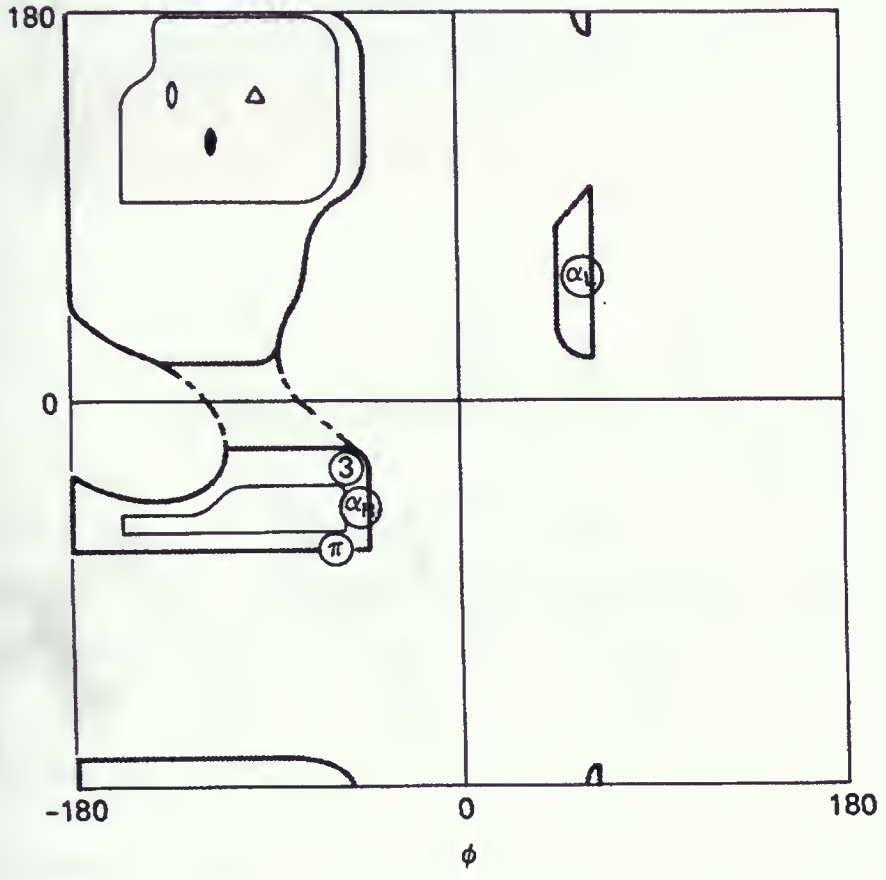


Figure 1.6: The positions of the regular conformations of polypeptides on a Ramachandran plot. The regular conformations are α_R, the right-handed α-helix; α_L, the left-handed α-helix; ○, the antiparallel $\beta - sheet$; ●, the parallel $\beta - sheet$; 3, the right-handed 3_{10}-helix; π, the right-handed π-helix; △, polyPro I, polyPro II, and polyGly II. Reproduced from [9].

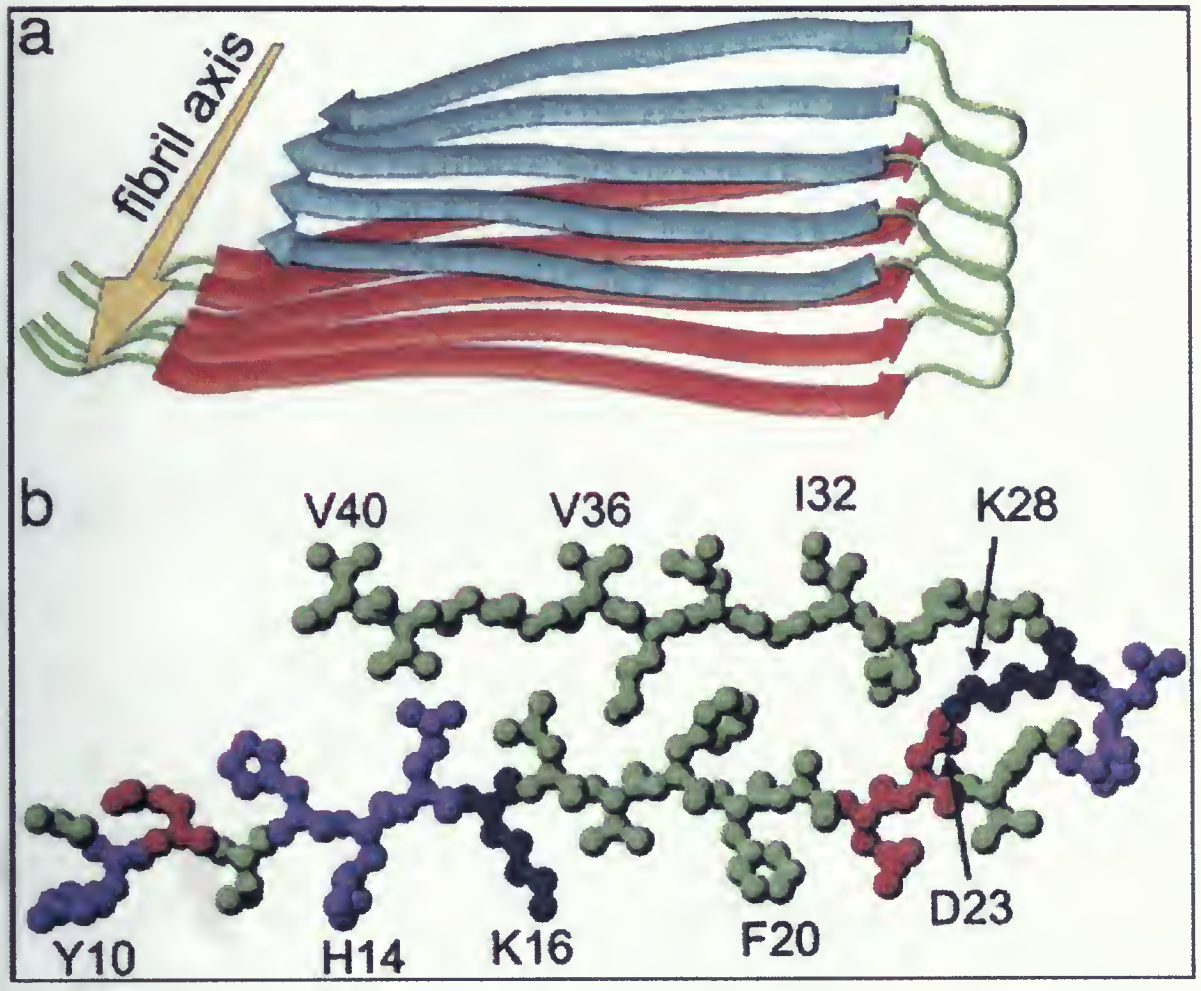


Figure 1.7: Structural model for $A\beta_{1-40}$ fibrils, consistent with solid state NMR constraints on the molecular conformation and intermolecular distances and incorporating the cross-β motif common to all amyloid fibrils. Residues 1-8 are considered fully disordered and are omitted. (a) Schematic representation of a single molecular layer, or cross- unit. The yellow arrow indicates the direction of the long axis of the fibril, which coincides with the direction of intermolecular backbone hydrogen bonds. The cross- unit is a double-layered structure, with in-register parallel -sheets formed by residues 12-24 (orange ribbons) and 30-40 (blue ribbons). (b) Central $A\beta_{1-40}$ molecule from the energy-minimized, five-chain system, viewed down the long axis of the fibril. Residues are color-coded according to their sidechains as hydrophobic (green), polar (magenta), positive (blue), or negative (red). Reproduced from [10].

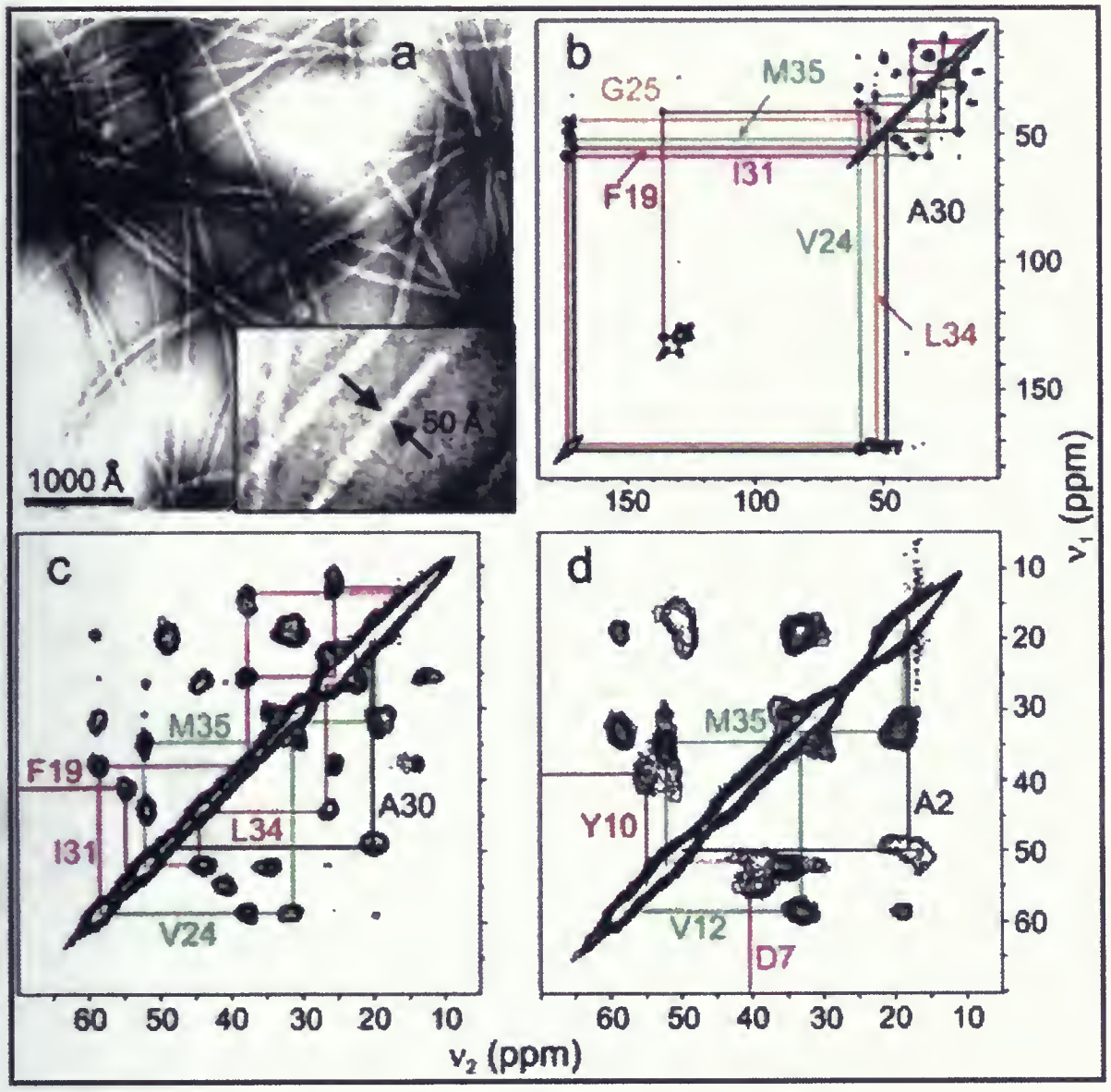


Figure 1.8: (a) Transmission electron microscope images of negatively stained amyloid fibrils after 14-day incubation of a 0.5 mM $A\beta_{1-40}$ solution. A 3× expansion (Inset) shows fibrils with the smallest diameters observed. (b) 2D $^{13}C-^{13}C$ chemical shift correlation spectrum of $A\beta_{1-40}$ fibril sample SU7, showing resonance assignment paths for the seven uniformly ^{15}N- and ^{13}C-labeled residues in this sample. (c) Expansion of the aliphatic region of the 2D spectrum of SU7. (d) Aliphatic region of the 2D $^{13}C-^{13}C$ chemical shift correlation spectrum of $A\beta_{1-40}$ fibril sample SU6. Reproduced from [10].

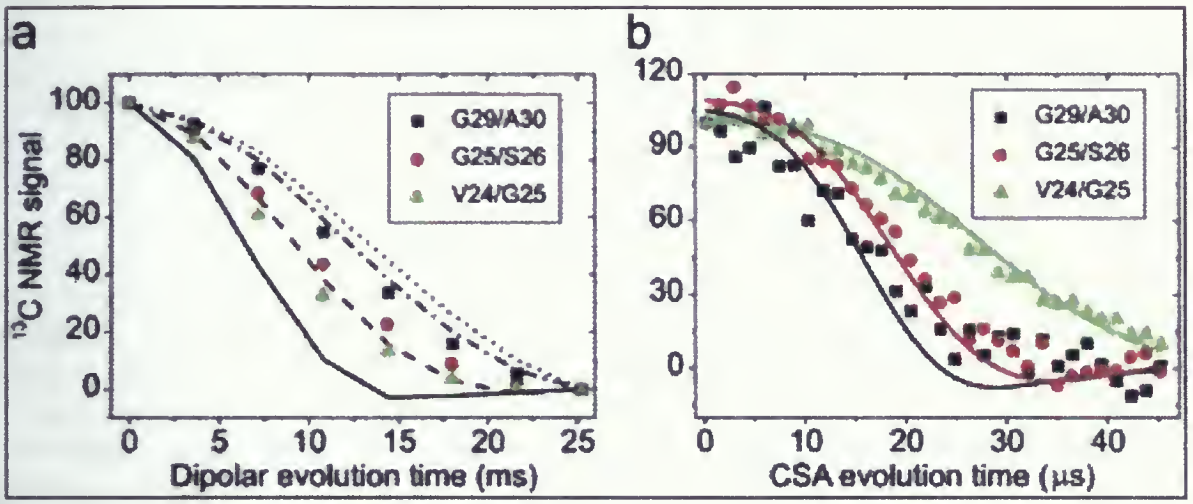


Figure 1.9: Solid state NMR data on DLn Aβ_{1-40} fibril samples with ^{13}C labels at the indicated backbone carbonyl sites. These data constrain the ϕ and ψ angles of the second labeled residue. (a) fpRFDR-CT data and simulations for $\phi = 40°$ (solid line), 80° (dashed line), 120° (dot-dashed line), and 160° (dotted line). Simulations are scaled and baseline-corrected to match the first and last experimental data points. (b) DQCSA data and simulations for , ϕ, ψ = -70°, -40° (green); 70°, -65° (red); and -165°, 135 (black). Reproduced from [10].

Chapter 2

Inverse Theory

2.1 Defining inverse problems and ill-posedness

To introduce and illuminate inverse problems, we will use a simple $1D$ example. Consider an NMR experiment that yields a single lineshape with the lineshape function $K(\omega, x)$, called the kernel, and its position is defined by a parameter x. A superposition of n lines with different x_i's, weighted by their relative abundances $g_i = g(x_i)$, generates a spectrum. The spectrum $s(\omega)$ can be measured at m discrete frequency points ω_j and is given by

$$s(\omega_j) = \sum_{i=1}^{n} K(\omega_j, x_i) g(x_i) \,, \tag{2.1}$$

The components g_i's form the vector of the relative abundances $\mathbf{g}$, the data vector $\mathbf{s}$ has the components s_j's and $\mathbf{K}$ is an $m \times n$ matrix that provides the mapping between $\mathbf{g}$ and $\mathbf{s}$. In practice we are interested in determining quantities like $g(x)$, which is not directly accessible by experiment, from the knowledge of its mapping $s(\omega)$ and the basis spectra in $K(\omega, x)$, *i.e.* inverting the equation $\mathbf{s} = \hat{\mathbf{K}}\mathbf{g}$. A least squares (LS) approach minimizes the distance between $\mathbf{s}$ and $\hat{\mathbf{K}}\,\mathbf{g}$, *i.e.*

$$\| \hat{\mathbf{K}}\,\mathbf{g} - \mathbf{s} \|^2 \longrightarrow \min \,, \tag{2.2}$$

where $\|\mathbf{v}\|$ denotes the Euclidean norm of the vector $\mathbf{v}$. The operator $\hat{\mathbf{K}}$ often represents the continuous version of Eq. 2.1 stated as a Fredholm integral equation of the first kind (FIE):

$$s(\omega) = \int_x g(x) K(\omega, x) dx \,, \tag{2.3}$$

i.e., the distribution function $g(x)$ is mapped linearly onto the spectrum $s(\omega)$ by the (real or complex) kernel function $K(\omega, x)$ which is "well-behaved" :

$$\int_w \int_x | K(\omega, x) |^2 \, dw \, dx = \text{const.} < \infty \tag{2.4}$$

In practice determination of g from s through FIE of the first kind is an ill-posed problem in the sense introduced by Hadamard [11]: at least one of the three criteria of existence, uniqueness, and stability is not satisfied:

Existence, meaning that for any s there exists a solution g, is not satisfied if the experimental data $s(\omega)$ is not within the range of $\hat{K}(\omega, x)$. This happens because the kernel $\hat{K}(\omega, x)$ does not describe random noise, while in experimental data analysis we often measure noisy data

$$s_j^{\sigma} = s(\omega_j) + \sigma \, \sigma_j \, \epsilon_j \,, \quad j = 1, 2, ..., m \tag{2.5}$$

where $\epsilon_1, \epsilon_2, \ldots, \epsilon_m$ are m independent standard normally distributed random variables, σ is a usually unknown scaling factor of the error and $\sigma_1, \sigma_2, \ldots, \sigma_m$ are given by an error model. (If for example $\sigma_i = 1$, the data are affected by absolute errors of the size σ. If *e.g.* $\sigma_i = g(x_i)$ the data are affected by relative errors of size σ.) Thus, to fulfill the existence requirement, we write

$$s(\omega) = \hat{K}(\omega, x) g(x) + \sigma(\omega) , \tag{2.6}$$

and calculate a function g^σ which minimizes the discrepancy between s and Kg, *i.e.*

$$\min \left\{ \| \hat{K}(\omega, x)\, g(x) - s(\omega) \| \right\} = \| \hat{K}(\omega, x)\, g^\sigma(x) - s(\omega) \| \tag{2.7}$$

This is called the generalized, or pseudo-inverse approach that determines "the best fit" to the data. Obviously, the function g^σ found according to pseudo-inverse approach is just an approximation to the true function $g(x)$.

Uniqueness, saying that for any s there exists one and only one solution g, is violated by the coexistence of several solutions $g_i^\sigma(x)$ ($g_a^\sigma(x) \neq g_b^\sigma(x)$ if $a \neq b$), with similar misfits for a given spectrum $s(\omega)$. To obtain a unique solution, an additional constraint could be added to the problem and inverse solution which has the minimal norm be chosen.

Stability, ensuring that the solution g is continuously dependent on s, is not fulfilled if two "adjacent" mappings s_1, s_2 do not have "adjacent" origins g_1, g_2. If so, then the fact that

$$\| \hat{\mathbf{K}}\, g_1^\sigma - \hat{\mathbf{K}}\, g_2^\sigma \| = \| s_1 - s_2 \| \tag{2.8}$$

is a small number does not ensure us that

$$\| g_1^\sigma - g_2^\sigma \| \tag{2.9}$$

is also small. Lack of stability is a distinguishing feature of ill-posed inverse problems and must be dealt with.

2.2 Restoring Stability Via Regularization Methods

To stabilize the solution of an ill-posed inverse problem, we have to impose extra restrictions on g and therefore to enforce that the quantity $\| g_1^\sigma - g_2^\sigma \|$ becomes small whenever the quantity $\| s_1 - s_2 \|$ is small, while solving Eq. 2.1. This is called a regularization [12] of Eq. 2.1. There are several methods of regularization such as truncated singular-value decomposition (SVD), Miller's regularization, maximum entropy, projection, iteration, Tikhonov regularization, non-negativity, and others. Here, we discuss just three techniques of stabilization: Discretization, Truncated SVD, and Smoothness (Tikhonov regularization). Out of all different methods, Tikhonov regularization is the best-known one and there exists much mathematical literature [13] about it.

2.2.1 Singular Values

A singular value expansion of the kernel $K(\omega, x)$ is

$$K(\omega, x) = \sum_{l=1}^{\infty} \zeta_l \, u_l(\omega) \, v_l(x) \tag{2.10}$$

where the singular values ζ_l and the singular functions $u_l(\omega)$ and $v_l(x)$, which are orthogonal, satisfy the condition $\int_x K(\omega, x) v_l(x) dx = \zeta_l u_l(\omega)$. The solution to the LS problem of Eq. 2.7, $g^\sigma(x)$, can be characterized by projecting the spectrum onto the singular function as

$$g^\sigma(x) = \sum_{l=1}^{\infty} \frac{1}{\zeta_l} \left(u_l(\omega), s(\omega)\right) \, v_l(x) \tag{2.11}$$

Here $(\ldots,\ldots)$ stands for the scalar product. Suppose the spectra $s_1(\omega)$ and $s_2(\omega)$ are produced by the distributions $g_1^\sigma(x)$ and $g_2^\sigma(x)$, then the LS discrepancy in the parameter domain is

$$\| \, g_1^\sigma(x) - g_2^\sigma(x) \, \|^2 = \sum_{l=1}^{\infty} \frac{1}{\zeta_l^2} \left\{u_l(\omega), (s_1(\omega) - s_2(\omega))\right\}^2 \tag{2.12}$$

It is obvious that small singular values ζ_l increase the risk of an unstable solution. For small ζ_l, $\| \, g_1^\sigma(x) - g_2^\sigma(x) \, \|$ is not small even if $\| \, s_1(\omega) - s_2(\omega) \, \|$ is. Therefore, the decay of the singular values can be a good measure for ill-posedness. In the discrete case, the singular value decomposition (SVD) of the matrix $\mathbf{K}$ is

$$\mathbf{K} = \mathbf{U} \, \mathbf{O} \, \mathbf{V}^T \tag{2.13}$$

where $\mathbf{U}$ and $\mathbf{V}$ are made up of the column vectors $\mathbf{u}_l$ and $\mathbf{v}_l$ which are the eigenvectors of $\mathbf{K}\mathbf{K}^T$ and $\mathbf{K}^T\mathbf{K}$, respectively. $\mathbf{U}$ is an $m \times n$ sized column-orthogonal matrix, $\mathbf{O}$ an $n \times n$ sized diagonal matrix of singular values $o_1, o_2, ..., o_n$, and $\mathbf{V}$ is an $n \times n$ sized orthogonal matrix. In this case, Eq. 2.11 becomes

$$\mathbf{g}^\sigma = \sum_{l=1}^{n} \frac{1}{o_l} \left(\mathbf{v}_l \, . \, \mathbf{u}_l^T\right) \, \mathbf{s} \tag{2.14}$$

here, n is the number of basis spectra in the kernel.

2.2.2 Discretization

By using a discrete grid x_i and ω_j and noting that $\Delta(x_i)$ is the interval size at x_i the parameter domain can be discretized and the continuous Eq. 2.3 be approximated by

$$s(\omega_j) = \sum_{i=1}^{n} K(\omega_j, x_i) \, g(x_i) \, \Delta(x_i) \tag{2.15}$$

Such discretization strongly restrict the solution g^σ and already provide regularization. By changing the grid density, *i.e.*, changing n, we can control the degree of stabilization. When

n is too large, we end up with artificial peaks in the structure of the approximated $g^{\sigma}(x)$. As the density of the grid decreases, the solution becomes stable (*i.e.* the fluctuations in $g^{\sigma}(x)$ becomes small). But, there is an optimal value for the grid density beyond which the result is oversmoothed and some of the features of the solution may be lost. Therefore it is inevitable to choose a suitable degree of stabilization for an underlying $s(\omega)$.

2.2.3 Truncation

Instead of using a lower grid density to stabilize the solution $g^{\sigma}(x)$, we may keep the grid density, n, unchanged but truncate the series of Eq. 2.15 at a certain threshold $n_{th} < n$. Again,by decreasing n_{th} the solution can be stabilized, and beyond an optimal value, some features of the solution will not be retrievable.

2.2.4 Smoothness

Generally speaking, regularization can be done by generalizing the minimization problem, Eq. 2.7, with additional constraints, $T\left(g(x)\right)$, *i.e.*

$$\| \hat{K}(\omega, x)\, g(x) - s(\omega) \|^2 + \lambda\, T\left(g(x)\right) \longrightarrow \min . \tag{2.16}$$

Here λ is called regularization parameter, which is responsible for the quality of the result given by a regularization method. This parameter is special in the sense that for different noise levels different optimum values of lambda are necessary to keep the information on $g(x)$ and the compatibility with the data $s(\omega)$. The very well-known and powerful functional $T\left(g(x)\right)$ is $\| \hat{\mathbf{L}}\, g(x) \|^2$ where $\hat{\mathbf{L}}$ is an operator for which the identity or the second derivative is frequently used. The choice of $\hat{\mathbf{L}} = \hat{\mathbf{I}}$, *i.e.* minimizing the norm of g, is suggested by Tikhonov [14]. Minimizing the norm of the second derivative of g, *i.e.* $\hat{\mathbf{L}} = \frac{d^2}{dx^2}$, is another possibility that provides smooth solutions and is due to Phillips [15]. Usually, the term "Tikhonov regularization" refers to both. For an arbitrary function f The quantity $\| f \|^2$ is given by

$$\| f \|^2 = \int f^2(x)\, dx \tag{2.17}$$

Maximum entropy method (MEM) is another popular regularization method for which the regularization term is, for example, the Shannon entropy [16]:

$$T\left(g(x)\right) = -\int g(x)\, \log\left[g(x)\right]\, dx \tag{2.18}$$

Tikhonov regularization, which asks for a small norm of $g(x)$, modifies Eq. 2.14 to

$$\mathbf{g}_{\lambda}^{\sigma} = \sum_{l=1}^{n} \frac{o_l}{(o_l^2 + \lambda)} \left(\mathbf{v}_l \,.\, \mathbf{u}_l^T\right) \mathbf{s} \tag{2.19}$$

Here, the degree of smoothing not only depends on n but also depends on the regularization parameter λ. Obviously, increasing λ leads to a higher degree of smoothness of $\mathbf{g}_{\lambda}^{\sigma}$ that stabilizes the solution, since the existence of λ in $\frac{o_l}{o_l^2+\lambda}$ blocks the fluctuations of $\mathbf{g}_{\lambda}^{\sigma}$ due to

small singular values. If this parameter is chosen too small, the solution $\mathbf{g}_\lambda^\sigma$ will show artificial peaks. Otherwise, if it is chosen too large, the solution will be oversmoothed. Therefore a robust procedure for the determination of an appropriate value for the regularization parameter is necessary.

2.3 Determination of the regularization parameter

It is essential to develop and apply a procedure to obtain an appropriate value for the regularization parameter λ. Many different methods are developed and studied to determine optimum values of this parameter. For example, the discrepancy method [17] in which one looks for a regularization parameter that is large enough to provide the most stabilized solution but that still leads to sufficient agreement between the data and the fit. There are other methods like the predictive minimum mean square error method [18] and the Provencher method [19, 20]. However, the self-consistent (SC) method developed recently by Honerkamp and Weese [18, 21] has been by far the most powerful procedure. References [18, 22] contain a comparison of the SC-method with other ones which are usually used for the determination of the regularization parameter. This comparison shows that the results obtained with the SC-method are much better and more reliable than the results obtained with the other methods. Also, various procedures used for the determination of the regularization parameter require the calculation of intermediate solutions for several values of this parameter, before the final solution is reached. For every step in this calculation matrix inversions are required, leading to enormous computational effort. This difficulty does not arise in SC-method, since a generalized singular-value decomposition [23] is used. Consequently, time and memory needed for the calculations is reduced by this decomposition and the computer program based on this procedure runs on a personal computer.

In this section, we explain the idea of the SC-method and present the notations used in this method. In section 2.5, we shall discuss how Honerkamp and Weese justify that the determination of an optimum value of the regularization parameter λ is done using the SC-method. This will followed by calculating the solution of Fredholm integral equations of the first kind based on Tikhonov regularization.

In order to define the SC-method, consider the error of the estimate g_λ^σ obtained by Tikhonov regularization, *i.e.*

$$d_g = \|g - g_\lambda^\sigma\|^2 \tag{2.20}$$

This error is a realization of a random variable D_g whose expectation value, ED_g depends on the function g, the regularization parameter λ and the size of the error σ,

$$ED_g = ED_g(g, \lambda, \sigma), \tag{2.21}$$

and the best estimate for the regularization parameter should minimize this quantity

$$\frac{\partial}{\partial \mu} ED_g(g, \mu, \sigma)\Big|_{\mu=\lambda} = 0\,. \tag{2.22}$$

In practice this equation can not be solved, because the function g is not known. The idea of the SC-method is to replace the function g in Eq. 2.22 by g_λ^σ, the best estimate that can

be obtained from the data. Therefore, the equation that defines an optimum value λ_{SC} for the regularization parameter is

$$\frac{\partial}{\partial \mu} ED_g(g_\lambda^\sigma, \mu, \sigma)\Big|_{\mu=\lambda} = 0\,. \tag{2.23}$$

If the value λ_{SC} leads to a good estimate g_λ^σ for the function g, the replacement of the function g by g_λ^σ is justified and Eq. 2.23 should define a good value for the regularization parameter. Thus, by solving Eq. 2.23 the regularization parameter is determined in a self-consistent manner. The Monte Carlo simulations proposed in refs. [18, 22] show that the values for the regularization parameter obtained with this procedure are only slightly different from the optimal value defined by the minimum of $ED_g(g, \lambda, \sigma)$.

In many cases the scaling factor σ is not known and in order to apply the SC-method to those cases the scaling factor σ in Eq. 2.23 should be replaced by a proper estimate $\hat{\sigma}$. In order to get an estimate for the scaling factor, we consider the misfit

$$d_s = \|s^\sigma - K g_\lambda^\sigma\|^2 \tag{2.24}$$

as a realization of a random variable depending on the function g, the regularization parameter λ and the scaling factor σ:

$$ED_s = ED_s(g, \lambda, \sigma)\,. \tag{2.25}$$

By comparing the misfit d_s with the expectation value ED_s, an estimate for the scaling factor can be defined as

$$d_s = ED_s(g, \lambda, \hat{\sigma})\,. \tag{2.26}$$

Again one should replace the function g in Eq. 2.26 by g_λ^σ, since in practice the function g is not known. And, therefore, the estimate $\hat{\sigma}$ for the scaling factor σ is determined by solving the equation

$$d_s = ED_s(g_\lambda^\sigma, \lambda, \hat{\sigma})\,. \tag{2.27}$$

Now the scaling factor in Eq. 2.23 can be replaced by the estimate $\hat{\sigma}$ and the optimum value for the regularization parameter is determined by solving the equation

$$\frac{\partial}{\partial \mu} ED_g(g_\lambda^\sigma, \mu, \hat{\sigma})\Big|_{\mu=\lambda} = 0\,. \tag{2.28}$$

2.4 Error intervals

Since the solution of Fredholm integral equations is obtained from noisy data, it is necessary to estimate the influence of the data errors on the result. This can be done by calculating confidence intervals, which are defined in such a way that the true result is in the interval with a probability of 68%.

The result obtained by a regularization method is affected by two errors, one is a bias caused by the regularization and the other is a statistical error caused by the data error. Since the bias can not be estimated, confidence intervals can not be defined for the result obtained

by a regularization method. In spite of this fact, for any fixed regularization parameter, error intervals which describe the influence of the data error on the result are useful for the interpretation of the result.

Such error intervals can be defined by the influence of errors of the size $\sigma\sigma_i$ in the data s_i^σ on the result. For the result g_λ^σ these intervals are given by

$$[g_\lambda^\sigma(x) - \sigma\left(g_\lambda^\sigma(x)\right), g_\lambda^\sigma(x) + \sigma\left(g_\lambda^\sigma(x)\right)] \,, \tag{2.29}$$

where the quantity $\sigma\left(g_\lambda^\sigma(x)\right)$ must be calculated for any fixed regularization parameter.

2.5 Theory

2.5.1 Description of the solution method

The solution of Fredholm integral equations of the first kind, Eq. 2.3, is defined as the function g_λ^σ, which minimizes the quantity

$$V(\lambda) = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left(s_i^\sigma - \left(\int K(\omega, x) g(x) dx \right) \right)^2 + \lambda \parallel \mathbf{L}\, g \parallel^2 , \tag{2.30}$$

where m is the number of data points. In this minimization if the function g is known to be positive, the constraints

$$g(x) \geq 0 \,, \tag{2.31}$$

must be considered. For simplicity, the vector s'^σ and the operator $\mathbf{K}'$ are introduced by

$$s_i'^\sigma = \frac{1}{\sigma_i} s_i^\sigma, \; i = 1, \ldots, m \,, \tag{2.32}$$

$$(\mathbf{K}' g)_i = \frac{1}{\sigma_i} \int K(\omega, x) g(x) dx \,, \; i = 1, \ldots, m \,. \tag{2.33}$$

With these notations $V(\lambda)$ reduces to

$$V(\lambda) = \parallel \mathrm{s}'^\sigma - \mathbf{K}' g \parallel^2 + \lambda \parallel \mathbf{L}\, g \parallel^2 . \tag{2.34}$$

If no constraints must be considered, the function g_λ^σ, given by the minimum of $V(\lambda)$, can be written explicitly as

$$g_\lambda^\sigma = \mathbf{K}'^{-1}(\lambda)\, \mathrm{s}'^\sigma \,, \tag{2.35}$$

where $\mathbf{K}'^{-1}(\lambda)$ is given by

$$\mathbf{K}'^{-1}(\lambda) = \left(\mathbf{K}'^t \mathbf{K}' + \lambda\, \mathbf{L}^t \mathbf{L} \right)^{-1} \mathbf{K}'^t \,. \tag{2.36}$$

And, explicit expressions can be obtained for the quantities $\sigma\left(g_\lambda^\sigma(x)\right)$, $ED_g(g, \lambda, \sigma)$ and $ED_s(g, \lambda, \sigma)$:

$$\sigma^2\left(g_\lambda^\sigma(x)\right) = \sigma^2 \left(\mathbf{K}'^{-1}(\lambda) \left(\mathbf{K}'^{-1}(\lambda) \right)^t \right)(x, x) \,, \tag{2.37}$$

$$ED_g(g,\lambda,\sigma) = \langle \| g - \mathbf{G}_\lambda^\sigma \|^2 \rangle \tag{2.38}$$
$$= \| g - \mathbf{K}'^{-1}(\lambda)\mathbf{K}'g\|^2 + \sigma^2 \operatorname{Tr}\left\{\mathbf{K}'^{-1}(\lambda)\left(\mathbf{K}'^{-1}(\lambda)\right)^t\right\},$$
$$ED_s(g,\lambda,\sigma) = \langle \| \mathbf{S}'^\sigma - \mathbf{K}'\mathbf{G}_\lambda^\sigma \|^2 \rangle \tag{2.39}$$
$$= \left\| \left(\mathbf{I} - \mathbf{K}'\mathbf{K}'^{-1}(\lambda)\right)\mathbf{K}'g \right\|^2 + \sigma^2 \operatorname{Tr}\left\{\left(\mathbf{I} - \mathbf{K}'\mathbf{K}'^{-1}(\lambda)\right)\left(\mathbf{I} - \mathbf{K}'\mathbf{K}'^{-1}(\lambda)\right)^t\right\}.$$

Here, $\mathbf{G}_\lambda^\sigma$ and $\mathbf{S}'^\sigma$ are the random variables belonging to g_λ^σ and s'^σ respectively.

If the constraints in Eq. 2.31 must be added, Eq. 2.35 no longer holds for the result. Also, the SC-method can not be applied in the way described so far, because the expectation values shown above can not be calculated explicitly. Thus, a modified SC-method will be introduced for this case. In addition, no explicit expressions for the error intervals can be obtained and the calculation of these intervals by a Monte Carlo simulation requires too much time. Therefore, the calculation of the error intervals will be simplified in this case.

2.5.2 Approximation by finite-dimensional quantities

In order to determine the solution of Fredholm integral equation of the first kind, we must approximate all functions and operators by finite-dimensional vectors and matrices. Although the approximation can be done in several ways, a simple approximation method with a low degree of convergence is sufficiently accurate. This is because the error caused by these approximations is small compared with the error caused by the regularization and the data error. Therefore, the quantities $(\mathbf{K}'g)$, $\| \mathbf{L}g \|^2$ and $\| g - g_\lambda^\sigma \|^2$ are approximated by

$$(\mathbf{K}'g)_i \approx \frac{h}{\sigma_i}\sum_{j=1}^{n} K(\omega_i, x_j)g(x_j), \quad i = 1,\ldots,m, \tag{2.40}$$

$$\| \mathbf{L}g \|^2 \approx h\sum_{i=1}^{n'}\left(\sum_{j=1}^{n} L_{ij}g(x_j)\right)^2, \tag{2.41}$$

$$\| g - g_\lambda^\sigma \|^2 \approx h\sum_{j=1}^{n}\left(g(x_j) - g_\lambda^\sigma(x_j)\right)^2, \tag{2.42}$$

With

$$x_j = x_{\min} + (j-1)\,h, \quad j = 1,\ldots,n,$$
$$h = (x_{\max} - x_{\min})/(n-1). \tag{2.43}$$

The values of the coefficients L_{ij} in Eq. 2.41 depend on the operator $\mathbf{L}$. If $\mathbf{L}$ is the identity, the coefficients L_{ij} are given by

$$L_{ij} = \delta_{ij}, \quad i = 1,\ldots,n', \quad j = 1,\ldots,n,$$
$$n' = n. \tag{2.44}$$

If $\mathbf{L}$ is the second derivative, the coefficients L_{ij} are given by

$$\begin{aligned} L_{ij} &= \frac{1}{h^2}\left(\delta_{ij} - 2\delta_{(i+1)j} + \delta_{(i+2)j}\right), \quad i = 1, \ldots, n', \quad j = 1, \ldots, n, \\ n' &= n - 2. \end{aligned} \tag{2.45}$$

If $\mathbf{L}$ is the second derivative with zeros at the boundaries, the coefficients L_{ij} are given by

$$\begin{aligned} L_{ij} &= \frac{1}{h^2}\left(\delta_{ij} - 2\delta_{(i+1)j} + \delta_{i(j+2)}\right), \quad i = 1, \ldots, n', \quad j = 1, \ldots, n, \\ n' &= n + 2. \end{aligned} \tag{2.46}$$

Now, we can consider the elements of the vectors $\mathbf{g}$ and $\mathbf{g}_\lambda^\sigma$ and of the matrices $\mathbf{K}'$ and $\mathbf{L}'$ as

$$g_j = g(x_j), \qquad j = 1, \ldots, n, \tag{2.47}$$

$$g_{\lambda j}^\sigma = g_\lambda^\sigma(x_j), \qquad j = 1, \ldots, n, \tag{2.48}$$

$$K'_{ij} = \frac{h}{\sigma_i} K(\omega_i, x_j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{2.49}$$

$$L'_{ij} = \sqrt{h} L_{ij}, \qquad i = 1, \ldots, n', \quad j = 1, \ldots, n. \tag{2.50}$$

2.5.3 Calculation of the result without constraints

As mentioned in chapter one, the result for Fredholm integral equation of the first kind is determined by the minimum of the quantity

$$V(\lambda) = \| \mathbf{s}'^\sigma - \mathbf{K}'\mathbf{g} \|^2 + \lambda \| \mathbf{L}'\mathbf{g} \|^2, \tag{2.51}$$

where the vector $\mathbf{g}$ and the matrices $\mathbf{K}'$ and $\mathbf{L}'$ are defined in Eqs. 2.47 to 2.50. The minimum of $V(\lambda)$ is given by the solution of the equation

$$\left(\mathbf{K}'^t\mathbf{K}' + \lambda\, \mathbf{L}'^t\mathbf{L}'\right) \mathbf{g}_\lambda^\sigma = \mathbf{K}'^t\mathbf{s}'^\sigma, \tag{2.52}$$

which can be written as

$$\mathbf{g}_\lambda^\sigma = \left(\mathbf{K}'^t\mathbf{K}' + \lambda\, \mathbf{L}'^t\mathbf{L}'\right)^{-1} \mathbf{K}'^t\mathbf{s}'^\sigma. \tag{2.53}$$

To calculate the result, the matrix $\mathbf{K}'^{-1}(\lambda) = \left(\mathbf{K}'^t\mathbf{K}' + \lambda\, \mathbf{L}'^t\mathbf{L}'\right)^{-1} \mathbf{K}'^t$ must be evaluated. The calculation of this matrix can be done in three steps. In the first two steps, the matrix $\mathbf{K}'^{-1}(\lambda)$ is evaluated for a given value of the regularization parameter. In the third step, a generalized singular value decomposition [23] is obtained and the matrix can be evaluated for different values of the regularization parameter without further matrix decompositions.

The first step can be divided into two cases:

Overdetermined problem, where the rank of the matrix $\mathbf{K}'$ is n, *i.e.* $n < m$. The QR-decomposition [24] of the matrix $\mathbf{K}'$ is used to reduced the number m of data points:

$$\mathbf{K}' = \mathbf{Q}\,\mathbf{K}'', \tag{2.54}$$

here, $\mathbf{Q}$ is an $m \times n$ matrix satisfying

$$\mathbf{Q}^t\mathbf{Q} = \mathbf{I}, \tag{2.55}$$

and $\mathbf{K}''$ is an upper triangular matrix with size $n \times n$. With these quantities $\mathbf{K}'^{-1}(\lambda)$ can be written as

$$\mathbf{K}'^{-1}(\lambda) = \left(\mathbf{K}''^t\mathbf{K}'' + \lambda\,\mathbf{L}'^t\mathbf{L}'\right)^{-1}\mathbf{K}''^t\mathbf{Q}^t \tag{2.56}$$

Underdetermined problem, where the rank of the matrix $\mathbf{K}'$ is m, *i.e.* $m < n$. In this case Eq. 2.54 holds with

$$\mathbf{Q} = \mathbf{I}, \tag{2.57}$$
$$\mathbf{K}'' = \mathbf{K}', \tag{2.58}$$

In the case of $m = n$ either approach is valid.

In the second step the QR-decomposition of the matrix

$$\begin{pmatrix} \mathbf{K}'' \\ s_l\,\mathbf{L}' \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix}\mathbf{R} \tag{2.59}$$

is introduced. In this equation the scaling factor s_l is given by

$$s_l = \frac{\|\mathbf{K}'\|_\infty}{\|\mathbf{L}'\|_\infty} \tag{2.60}$$

or by

$$s_l = \sqrt{\lambda_0} \tag{2.61}$$

if an appropriate value λ_0 for the regularization parameter λ is known. $\mathbf{R}$ is an upper triangular $n \times n$ matrix, $\mathbf{Q}_1$ is a $(\min(m,n)) \times n$ matrix and $\mathbf{Q}_2$ is an $n' \times n$ matrix. Furthermore, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ satisfy the relation

$$\mathbf{Q}_1^t\mathbf{Q}_1 + \mathbf{Q}_2^t\mathbf{Q}_2 = \mathbf{I} \tag{2.62}$$

With this decomposition the matrix $\mathbf{K}'^{-1}(\lambda)$ becomes

$$\mathbf{K}'^{-1}(\lambda) = \mathbf{R}^{-1}\left((1 - \frac{\lambda}{s_l^2})\mathbf{Q}_1^t\mathbf{Q}_1 + \frac{\lambda}{s_l^2}\mathbf{I}\right)^{-1}\mathbf{Q}_1^t\mathbf{Q}^t\,. \tag{2.63}$$

If the regularization parameter is known, the scaling factor s_l is given by Eq. 2.61 and the result together with the error intervals can be calculated for the regularization parameter λ_0:

$$\mathbf{g}_\lambda^\sigma = \mathbf{R}^{-1}\mathbf{Q}_1^t\mathbf{Q}^t\mathbf{s}'^\sigma\,, \tag{2.64}$$
$$\sigma^2\left(g_{\lambda i}^\sigma\right) = \sigma^2\left(\mathbf{R}^{-1}\mathbf{Q}_1^t\mathbf{Q}_1\left(\mathbf{R}^{-1}\right)^t\right)_{ii}\,,\quad i = 1,\ldots,n\,. \tag{2.65}$$

Otherwise, if the regularization parameter must be determined with the SC-method, a third step should be considered. In this step the SV-decomposition [25] of $\mathbf{Q}_1$ is introduced:

$$\mathbf{Q}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^t\,. \tag{2.66}$$

Here, $\mathbf{U}$ is a $(\min(m,n)) \times (\min(m,n))$ matrix satisfying

$$\mathbf{U}^t\mathbf{U} = \mathbf{I}\,, \tag{2.67}$$

and $\mathbf{V}$ is a $n \times (\min(m,n))$ matrix satisfying

$$\mathbf{V}^t\mathbf{V} = \mathbf{I}\,. \tag{2.68}$$

The matrix $\mathbf{W}$ is a diagonal matrix containing the singular values of $\mathbf{Q}_1$. Replacing $\mathbf{Q}_1$ by $\mathbf{UWV}^t$ in Eq. 2.63, the final expression for the matrix $\mathbf{K'}^{-1}(\lambda)$ is obtained:

$$\mathbf{K'}^{-1}(\lambda) = \mathbf{R}^{-1}\,\mathbf{V}\,\mathbf{W}^{-1}(\lambda)\,\mathbf{U}^t\,\mathbf{Q}^t\,, \tag{2.69}$$

where

$$W_{ij}^{-1}(\lambda) = \delta_{ij}\,W_{ij} \Big/ \left\{\left(1 - \frac{\lambda}{s_l^2}\right) W_{ii}^2 + \frac{\lambda}{s_l^2}\right\} \quad i,j = 1,\ldots,\min(m,n)\,. \tag{2.70}$$

Then, for any regularization parameter λ, the solution of Fredholm integral equations of the first kind together with the error intervals can be obtained:

$$\mathbf{g}_\lambda^\sigma = \mathbf{K'}^{-1}(\lambda)\,\mathbf{s'}^\sigma = \mathbf{R}^{-1}\,\mathbf{V}\,\mathbf{W}^{-1}(\lambda)\,\mathbf{U}^t\,\mathbf{Q}^t\,\mathbf{s'}^\sigma\,. \tag{2.71}$$

$$\sigma^2\,(g_{\lambda i}^\sigma) = \sigma^2 \left(\mathbf{R}^{-1}\mathbf{V}\mathbf{W}^{-1}(\lambda)\left(\mathbf{R}^{-1}\mathbf{V}\mathbf{W}^{-1}(\lambda)\right)^t\right)_{ii}\,, \quad i = 1,\ldots,n\,. \tag{2.72}$$

Now, an optimum value for the regularization parameter must be determined. This can be done by calculating $ED_g(g,\mu,\sigma)$ via Eq. 2.38 and then by solving Eq. 2.28. The function g and the operators $\mathbf{K'}$ and $\mathbf{K'}^{-1}(\mu)$ are first replaced by corresponding vector and matrices. Then, approximations Eqs. 2.40 and 2.42 result in the following:

$$\begin{aligned} ED_g(\mathbf{g}_\lambda^\sigma,\mu,\sigma) &= \left\langle \| \mathbf{g}_\lambda^\sigma - \mathbf{G}_\mu^\sigma \|^2 \right\rangle \\ &= h \| \mathbf{g}_\lambda^\sigma - \mathbf{K'}^{-1}(\mu)\mathbf{K'}\mathbf{g}_\lambda^\sigma\|^2 + h\,\sigma^2\,\mathrm{Tr}\left\{\mathbf{K'}^{-1}(\mu)\left(\mathbf{K'}^{-1}(\mu)\right)^t\right\} \end{aligned} \tag{2.73}$$

By using Eqs. 2.54, 2.59 and 2.66 the matrix $\mathbf{K'}$ can be expressed as

$$\mathbf{K'} = \mathbf{Q}\,\mathbf{U}\,\mathbf{W}\,\mathbf{V}^t\,\mathbf{R} \tag{2.74}$$

And the final expression for $ED_g(\mathbf{g}_\lambda^\sigma,\mu,\sigma)$ can be obtained with help of the Eqs. 2.69, 2.71 and 2.74:

$$\begin{aligned} ED_g(\mathbf{g}_\lambda^\sigma,\mu,\sigma) &= h \left\| \mathbf{R}^{-1}\mathbf{V}\left(\mathbf{I} - \mathbf{W}^{-1}(\mu)\mathbf{W}\right)\mathbf{W}^{-1}(\lambda)\mathbf{U}^t\mathbf{Q}^t\mathbf{s'}^\sigma \right\|^2 \\ &+ h\,\sigma^2\,\mathrm{Tr}\left\{\mathbf{R}^{-1}\mathbf{V}\mathbf{W}^{-1}(\mu)\left(\mathbf{R}^{-1}\mathbf{V}\mathbf{W}^{-1}(\mu)\right)^t\right\}\,. \end{aligned} \tag{2.75}$$

After the value λ_{SC} for the regularization parameter λ is calculated by solving Eq. 2.28, the error intervals and the result itself are obtained by replacing λ in Eqs. 2.71 and 2.72 by λ_{SC}. If the scaling factor σ is not known, an estimate $\hat{\sigma}$ must be calculated before the result

and the error intervals are obtained. Through similar process as to calculate $ED_g(\mathbf{g}_\lambda^\sigma, \mu, \sigma)$, explicit expressions can be obtained for $ED_s(\mathbf{g}_\lambda^\sigma, \lambda, \sigma)$ and d_s:

$$\begin{aligned} ED_s(\mathbf{g}_\lambda^\sigma, \lambda, \sigma) &= \left\| \left(\mathbf{I} - \mathbf{W}\mathbf{W}^{-1}(\lambda)\right) \mathbf{W}\mathbf{W}^{-1}(\lambda)\mathbf{U}^t\mathbf{Q}^t\mathbf{s}'^\sigma \right\|^2 \\ &+ \sigma^2 \operatorname{Tr}\left\{ \left(\mathbf{I} - \mathbf{W}\mathbf{W}^{-1}(\lambda)\right) \left(\mathbf{I} - \mathbf{W}\mathbf{W}^{-1}(\lambda)\right)^t \right\} + \sigma^2 (m+n) , \end{aligned} \tag{2.76}$$

$$d_s = \|\mathbf{s}'^\sigma\|^2 - \|\mathbf{U}^t\mathbf{Q}^t\mathbf{s}'^\sigma\|^2 + \left\| \left(\mathbf{I} - \mathbf{W}\mathbf{W}^{-1}(\lambda)\right) \mathbf{U}^t\mathbf{Q}^t\mathbf{s}'^\sigma \right\|^2 . \tag{2.77}$$

By solving Eq. 2.27 an estimate $\hat{\sigma}$ for the scaling factor σ can be obtained.

Finally, here is a summary of various steps necessary to compute the result:

If an appropriate value λ_0 for the regularization parameter is known, the calculation can be performed in 5 steps:

1. initialize m, n, n', $\mathbf{K}'$, $\mathbf{L}'$, σ and λ_0;
2. if $m \leq n$, set $\mathbf{K}'' = \mathbf{K}'$ and $(\mathbf{Q}^t\mathbf{s}'^\sigma) = \mathbf{s}'^\sigma$,
   if $m \geq n$, determine $\mathbf{K}''$ and $(\mathbf{Q}^t\mathbf{s}'^\sigma)$ by QR-decomposition (Eq. 2.54);
3. set s_l (Eq. 2.61) and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59);
4. calculate $(\mathbf{R}^{-1}\mathbf{Q}_1^t)$;
5. calculate $\mathbf{g}_\lambda^\sigma$ (Eq. 2.64) and $\sigma\left(\mathbf{g}_{\lambda i}^\sigma\right)$ for $i \notin M$ (Eq. 2.65);

If an appropriate value for the regularization parameter is not known, the calculation can be performed in 7 steps:

1. initialize m, n, n', $\mathbf{K}'$, $\mathbf{L}'$ and σ, if known;
2. if $m \leq n$, set $\mathbf{K}'' = \mathbf{K}'$ and $(\mathbf{Q}^t\mathbf{s}'^\sigma) = \mathbf{s}'^\sigma$,
   if $m \geq n$, determine $\mathbf{K}''$ and $(\mathbf{Q}^t\mathbf{s}'^\sigma)$ by QR-decomposition (Eq. 2.54);
3. set s_l (Eq. 2.60) and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59);
4. determine $\mathbf{V}$, $\mathbf{W}$ and $(\mathbf{U}^t\mathbf{Q}^t\mathbf{s}'^\sigma)$ by SV-decomposition (Eq. 2.66);
5. calculate $(\mathbf{R}^{-1}\mathbf{V})$;
6. determine λ_{SC} by solving Eq. 2.23 or Eq. 2.28, if an estimate for the scaling factor σ must be calculated;
7. calculate $\mathbf{g}_\lambda^\sigma$ (Eq. 2.71) and $\sigma\left(\mathbf{g}_{\lambda i}^\sigma\right)$ for $i \notin M$ (Eq. 2.72).

2.5.4 Calculation of the result with constraints

If additional constraints should be considered, the same QR-decomposition , Eq. 2.54 in the case of overdetermined problem and Eqs. 2.57 and 2.58 in the case of underdetermined problem, is used to reduce the number m of data points. With the quantities introduced there the result is obtained by minimizing

$$V(\lambda) = \| \mathbf{Q}^t\mathbf{s}'^\sigma - \mathbf{K}''\mathbf{g} \|^2 + \lambda \| \mathbf{L}'\,\mathbf{g} \|^2 + \| \mathbf{s}'^\sigma \|^2 - \| \mathbf{Q}^t\mathbf{s}'^\sigma \|^2 \tag{2.78}$$

subject to the constraints

$$g_i \geq 0, \quad i = 1, \ldots, n\,. \tag{2.79}$$

In order to determine an optimum value for the regularization parameter λ, the SC-method must be modified. According to the modified SC-method, the determination of the regularization parameter can be done in three steps:

Step 1, a value λ'_{SC} for the regularization parameter λ and , if necessary, an estimate $\hat{\sigma}$ for the scaling factor σ is calculated in the absence of the constraints Eq. 2.79.

Step 2, an intermediate result is determined by minimizing $V(\lambda'_{SC})$ subject to the constraints Eq. 2.79. This minimization can be done by a QP-algorithm [26] and leads to a set M' of active constraints:

$$g^{\sigma}_{\lambda i} = 0, \quad i \in M'\,. \tag{2.80}$$

Step 3, the final value λ_{SC} for the regularization parameter λ is determined. For this purpose the set of active constraints M' is taken into account by removing the columns from the matrices $\mathbf{K}''$ and $\mathbf{L}'$ which belong to an active constraint.

If a value for the regularization parameter is given, λ_0, or has been calculated, λ_{SC}, with the modified SC-method, the final result is obtained by minimizing $V(\lambda_0)$ or $V(\lambda_{SC})$ subject to the constraints Eq. 2.79. At this stage, a set M of active constraints for the final result is calculated. This set is used for the determination of the error intervals by removing the columns of the matrices $\mathbf{K}''$ and $\mathbf{L}'$ which belong to an active constraint. Then the QR-decomposition introduced in Eq. 2.59 is obtained using these modified matrices. Finally, the error intervals can be calculated according to Eq. 2.65.

Again, various steps necessary to calculate the result are summarized:

If an appropriate value λ_0 for the regularization parameter is known, the calculation can be performed in 7 steps:

1. initialize m, n, n', $\mathbf{K}'$, $\mathbf{L}'$, σ and λ_0;
2. if $m \leq n$, set $\mathbf{K}'' = \mathbf{K}'$ and $(\mathbf{Q}^t \mathbf{s}'^{\sigma}) = \mathbf{s}'^{\sigma}$,
   if $m \geq n$, determine $\mathbf{K}''$ and $(\mathbf{Q}^t \mathbf{s}'^{\sigma})$ by QR-decomposition (Eq. 2.54);
3. determine $\mathbf{g}^{\sigma}_{\lambda}$ and the set M of active constraints by minimizing $V(\lambda_0)$ subject to the constraints Eq. 2.79;
4. determine $\mathbf{K}''_m$ and $\mathbf{L}'_m$ by removing the columns from $\mathbf{K}''$ and $\mathbf{L}'$ which belong to an active constraint;
5. set s_l (Eq. 2.61) and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59 with $\mathbf{K}''$ and $\mathbf{L}'$ replaced by $\mathbf{K}''_m$ and $\mathbf{L}'_m$);
6. calculate $(\mathbf{R}^{-1}\mathbf{Q}_1^t)$;
7. calculate $\mathbf{g}^{\sigma}_{\lambda}$ (Eq. 2.64) and $\sigma\,(\mathbf{g}^{\sigma}_{\lambda i})$ for $i \notin M$ (Eq. 2.65).

If an appropriate value for the regularization parameter is not known, the calculation can be performed in 17 steps:

1. initialize m, n, n', $\mathbf{K}'$, $\mathbf{L}'$ and σ, if known;
2. if $m \leq n$, set $\mathbf{K}'' = \mathbf{K}'$ and $(\mathbf{Q}^t \mathbf{s}'^{\sigma}) = \mathbf{s}'^{\sigma}$,
   if $m \geq n$, determine $\mathbf{K}''$ and $(\mathbf{Q}^t \mathbf{s}'^{\sigma})$ by QR-decomposition (Eq. 2.54);
3. set s_l (Eq. 2.60) and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59);
4. determine $\mathbf{V}$, $\mathbf{W}$ and $(\mathbf{U}^t \mathbf{Q}^t \mathbf{s}'^{\sigma})$ by SV-decomposition (Eq. 2.66);
5. calculate $(\mathbf{R}^{-1}\mathbf{V})$;
6. determine the preliminary value λ'_{SC} for the regularization parameter λ by solving Eq. 2.23 or Eq. 2.28, if in addition an estimate for the scaling factor σ must be calculated.
7. determine the set M' of active constraints by minimizing $V(\lambda'_{SC})$ subject to the constraints Eq. 2.79;
8. determine $\mathbf{K}''_m$ and $\mathbf{L}'_m$ by removing the columns from $\mathbf{K}''$ and $\mathbf{L}'$ which belong to an active constraint;
9. set s_l (Eq. 2.60) and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59 with $\mathbf{K}''$ and $\mathbf{L}'$ replaced by $\mathbf{K}''_m$ and $\mathbf{L}'_m$);
10. determine $\mathbf{V}$, $\mathbf{W}$ and $(\mathbf{U}^t \mathbf{Q}^t \mathbf{s}'^{\sigma})$ by SV-decomposition (Eq. 2.66);
11. calculate $(\mathbf{R}^{-1}\mathbf{V})$;
12. determine the final value λ_{SC} for the regularization parameter λ by solving Eq. 2.23;
13. determine $\mathbf{g}^{\sigma}_{\lambda}$ and the set M of active constraints by minimizing $V(\lambda_{SC})$ subject to the constraints Eq. 2.79;
14. determine $\mathbf{K}''_m$ and $\mathbf{L}'_m$ by removing the columns from $\mathbf{K}''$ and $\mathbf{L}'$ which belong to an active constraint;
15. set $s_l = \sqrt{\lambda_{SC}}$ and determine $\mathbf{Q}_1$ and $\mathbf{R}$ by QR-decomposition (Eq. 2.59 with $\mathbf{K}''$ and $\mathbf{L}'$ replaced by $\mathbf{K}''_m$ and $\mathbf{L}'_m$);
16. calculate $(\mathbf{R}^{-1}\mathbf{Q}_1^t)$;
17. calculate $\sigma\,(\mathbf{g}^{\sigma}_{\lambda i})$ for $i \notin M$ (Eq. 2.65).

2.6 Two-dimensional problem

The solution of Fredholm integral equation of the first kind described in the previous section is a one-dimensional vector $\mathbf{g}$ with elements $g_j = g(x_j)$, where $j = 0, \ldots, (n-1)$. The kernel $\mathbf{K}$ that provides the mapping between $\mathbf{g}$ and the data vector $\mathbf{s}$ is a matrix with elements $K_{ij} = K(\omega_i, x_j)$. If the function g is a two-dimensional object $g(x, y)$, *e.g.* distribution of ϕ and ψ angles that characterize peptide backbone conformations, its discrete vector $\mathbf{g}$ has

the components g_{kl}:

$$\begin{aligned} g_{kl} &= g(x_k, y_l)\,, \\ x_k &= x_{\min} + k\,h_x\,, \quad k = 0, \ldots, (n_x - 1)\,, \quad n_x = \frac{x_{\max} - x_{\min}}{h_x} + 1\,, \\ y_l &= y_{\min} + l\,h_y\,, \quad l = 0, \ldots, (n_y - 1)\,, \quad n_y = \frac{y_{\max} - y_{\min}}{h_y} + 1\,, \end{aligned} \tag{2.81}$$

where h_x and h_y are the step size of discretization along x and y directions respectively. One can number the two dimensions of grid points in a single one-dimensional sequence by transforming the double index kl to a single index j according to

$$j = k\,n_y + l\,, \quad j = 0, \ldots, (n-1)\,, \quad \text{where} \quad n = n_x\,n_y\,. \tag{2.82}$$

Accordingly, the pair of arguments (x_k, y_l) is lumped into a single component η_j. Then the minimization problem Eq. 2.34 can be solved to obtain the optimum value for the regularization parameter and to calculate the solution g. If the operator $\mathbf{L}$ is chosen to be the identity, the functional to be minimized is given by

$$\Psi(g) = \| s^\sigma(\omega) - \mathbf{K}(\omega, \eta)\, g(\eta) \|^2 + \lambda \| g(\eta) \|^2 \tag{2.83}$$

If $\mathbf{L}$ is the second derivative, the functional $\Psi(g)$ is given by

$$\Psi(g) = \| s^\sigma(\omega) - \mathbf{K}(\omega, \eta)\, g(\eta) \|^2 + \lambda \left\| \frac{d^2}{d\eta^2} g(\eta) \right\|^2 \tag{2.84}$$

In both cases the operator $\mathbf{L}$ introduces no coupling between the two dimensions. The space of the two dimensions is unwrapped into a one-dimensional space and consequently the function g is treated as if it were a one-dimensional object. An appropriate replacement for the operator $\mathbf{L}$ can be the "Laplacian" defined according to the underlying system of coordinates. The Laplacian has intrinsic dependence on the operators acting on both dimensions, thereby introduces a cross-coupling between the two dimensions. This coupling should recover the shape of the two-dimensional function $g(x, y)$ better and result in more stable solutions. In the next chapter, we present discrete implementation of the Laplacian in both Cartesian and Spherical coordinates systems.

Chapter 3

NMR backgrounds and set up of calculations

In this chapter we present the numerical details necessary for the extraction of the Ramachandran angles distribution function $g(\phi,\psi)$ via two-dimensional inverse theory. This includes the calculation of the kernel, the simulation of the noisy data and the discrete implementation of the Laplacian. The distribution of the Ramachandran angles, ϕ and ψ, can be obtained by solving the Fredholm integral equation of the first kind:

$$s(\omega) = \int\int K(\omega;\phi,\psi)\, g(\phi,\psi)\, d\phi\, d\psi\,. \tag{3.1}$$

As discussed in chapter 2, finite-dimensional vectors can be introduced for the functions $s(\omega)$ and $g(\phi,\psi)$, and a matrix $\mathbf{K}$ for the kernel-function $K(\omega;\phi,\psi)$. The data vector s consists of m discrete data points $s_1,\ldots,s_m$. Also, the solution vector $\mathbf{g}$ is an n dimensional column vector with elements $g_1,\ldots,g_n$. Each g_j is connected to a pair of (ϕ_k,ψ_l) through

$$g_j = g_{(kn_\psi+l)} = g_{k,l} = g(\phi_k,\psi_l)\,. \tag{3.2}$$

Therefore Eq. 3.1 is approximated by

$$s_i = h_\phi\, h_\psi \sum_{j=1}^{n} K_{i,j}\, g_j\,, \quad i=1,\ldots,m\,. \tag{3.3}$$

The kernel $\mathbf{K}$ is the set of basis spectra covering the parameter domain (ϕ,ψ); each column in the matrix $\mathbf{K}$ is a basis spectrum and there are a total of $n = n_\phi n_\psi$ such spectrum. The kernel is known from the theory and once the data points are obtained from either the experiment or simulation, the solution vector $\mathbf{g}$ can be calculated by inverting Eq. 3.3.

3.1 Kernel

The kernel $\mathbf{K}$ is obtained through simulations based on two different measurements. The first 154 rows of the kernel $\mathbf{K}$ contains simulated data points based on "constant-time double-quantum-filtered dipolar recoupling"(CTDQFD) measurements. The other 36 rows of the kernel $\mathbf{K}$ contains simulated data points based on "2D MAS exchange" measurements. On each row, each number corresponds to a given (ϕ,ψ) where ϕ ranges from $-180°$ to $0°$, inclusive, in increments of 5 degrees and ψ ranges from $-180°$ to $180°$, inclusive, in increments of 5 degrees. Thus, there are a total of 2701 pairs of (ϕ,ψ) and for each (ϕ,ψ) there is a total of 190 simulated data points. For $5°$ resolution, therefore, the size of the inversion problem

is 190 × 2701 which is a **vastly underdetermined** problem. To speed up the calculations and to examine the stability of the result for lower grid densities, the decimated versions of the kernel **K** are also used, using every second, or third, or even fourth columns in **K**, to simulate 10°, 15° and 20° resolutions in (ϕ, ψ) space.

3.1.1 Two-dimensional (2D) nuclear magnetic resonance (NMR) exchange spectroscopy with magic angle spinning (MAS)

In general, 2D NMR exchange spectroscopy [27] permits the measurement of correlations between NMR frequencies measured in two time periods, t_1 and t_2, separated by an exchange period τ. In the simple case of a static, unoriented, molecular solid with spin-$\frac{1}{2}$ labels at two sites A and B that are separated by several angstroms, the one-dimensional (1D) NMR spectrum is typically a superposition of the powder pattern line shapes of the two sites. The powder pattern line shapes result from the orientation dependence of the individual NMR frequencies, which is principally due to the chemical shielding anisotropy (CSA) [28]. In the case of an axially symmetric interaction or in the presence of rapid molecular motion which is symmetric with respect to one axis, the orientation dependence of an NMR frequency is given by

$$\omega = \omega(\theta) = x\,\frac{(3\cos^2\theta - 1)}{2} = x\,P_2(\cos\theta)\,, \tag{3.4}$$

where x corresponds to $\theta = 0$, $x = \omega(0)$, and θ is the angle between the externally applied magnetic field and the molecular axis of symmetry. An oriented spectrum for $\theta = 0$ is a direct representation of the anisotropy distribution, $g(x)$, while the powder spectrum $S(\omega)$ is the superposition of contributions from all possible angles θ:

$$S(\omega) = \int_0^{\frac{\pi}{2}} \left[g(x)\frac{\partial x}{\partial \omega}\right] p(\theta) d\theta\,, \tag{3.5}$$

where $p(\theta) = \sin\theta$ is the solid angle weighting factor. Fig. 3.1 illustrates the relationship between a Gaussian spectrum and corresponding powder pattern. In a 2D exchange measurement nuclear magnetization can be exchanged between sites A and B during τ, due to the weak dipole-dipole couplings between the labeled nuclei [29, 30, 31]. If τ is comparable to or greater than the time scale for magnetization exchange, the 2D spectrum shows off-diagonal intensity that is a direct measurement of the correlations between the orientation-dependent NMR frequencies of sites A and B. These frequency correlations are determined by the relative orientation of the CSA tensors of A and B within a labeled molecule. If, as is often the case, the orientation of the CSA principal axes relative to the local bonding geometries at sites A and B are known, then the relative orientation of the chemical groups that contain A and B can be determined from or strongly constrained by the 2D spectrum.

2D exchange spectra have been used to investigate the structures of molecular solids, synthetic polymers and peptides in the manner described above [32, 33, 34]. Fig. 3.2 illustrates a basic rf pulse sequence for 2D NMR exchange spectroscopy. Two significant problems with static 2D exchange spectroscopy, which are particularly severe in biochemical applications and other applications where sample quantities are limited and spectra are complex, are the low sensitivity and the poor resolution associated with the broad, powder pattern line

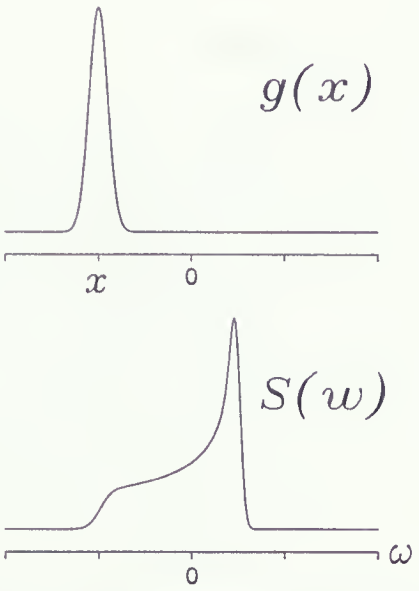


Figure 3.1: A schematic representation of a powder pattern.

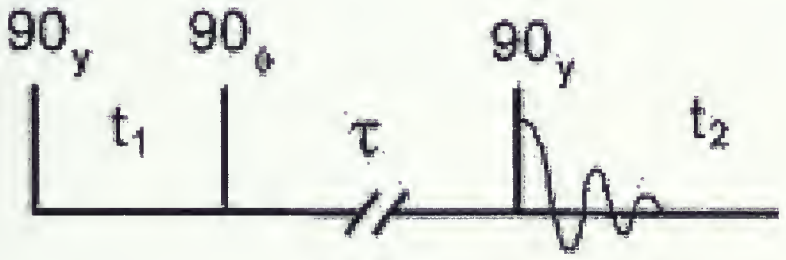


Figure 3.2: A simple rf pulse sequence for 2D NMR exchange spectroscopy. The notation θ_ϕ represents a pulse that rotates nuclear spin angular momenta by $\theta°$ about axis ϕ in the rotating frame. Reproduced from [35]

shapes. These problems can be alleviated to a large extend by magic angle spinning (MAS). Magic angle is the angle for which the spectral line becomes infinitely narrow and is obtained by setting "$\omega(\theta)$" to zero. In other words

$$\frac{(3\cos^2\theta - 1)}{2} = 0\,, \longrightarrow \theta_{\text{magic}} = 54.6^\circ\,. \tag{3.6}$$

At moderate speeds, MAS narrows CSA powder patterns to comparatively sharp lines, called spinning sideband lines, at the isotropic chemical shift frequencies plus integral multiples of the spinning frequency. As shown by Veeman *et al.* [36] and Spiess *et al.* [37, 38], 2D exchange measurements can be carried out with MAS in such a way that off-diagonal NMR signals appear only if exchange processes are present, by proper synchronization of the radio-frequency (rf) pulse sequence with the sample rotation. Compared with nonspinning 2D exchange measurements, rotor-synchronized 2D MAS NMR exchange measurements have dramatically improved sensitivity and resolution because the off-diagonal signals are concentrated into sharp peaks that connect the spinning sideband frequencies in t_1 and t_2.

Detailed descriptions of the theory behind rotor-synchronized 2D MAS NMR exchange spectroscopy are presented by Robert Tycko *et al.* [35]. A new class of technique, called orientationally weighted 2D MAS exchange measurements, is introduced and shown to augment the structural information contained in 2D MAS exchange data. The special case of investigations of peptide backbone conformations, characterized by torsion angles ϕ and ψ, using 2D ^{13}C MAS NMR exchange spectroscopy is treated in depth, in both theory and experiment.

Rotor-synchronized pulse sequences that produce purely absorptive (real) orientationally weighted 2D MAS NMR exchange spectra are shown in Fig. 3.3. Note that the synchronization of the pulse sequence with the sample rotation is given by $\tau_n = n\,\tau_R$; where n is an integer, $\tau_R = \frac{2\pi}{\omega_R}$ and ω_R is the sample rotation frequency. Also, Fig. 3.4 shows rf pulse sequence for obtaining experimental 2D MAS exchange spectra. The 2D frequency-domain spectrum consists of a set of delta-function signals at frequencies $\omega^i_{iso} + M\omega_R$ in the first frequency domain (ω_1) and $\omega^j_{iso} + N\omega_R$ in the second frequency domain (ω_2); where $\omega^i_{iso} = -\omega_0\,\sigma^i_{iso}$ is the isotropic chemical shift for site i in rad/s, ω_0 is the nuclear Larmor frequency and σ^i_{iso} is the isotropic chemical shielding for site i. These signals are called "spinning sideband crosspeaks." The total spinning sideband crosspeak amplitudes due to exchange from site i to site j in molecules are

$$A^{ij}_{MN} = \frac{1}{4\pi}\int_0^{2\pi} d\alpha \int_0^{\pi} d\beta\, \sin\beta\, H^i_M(\alpha,\beta)\, F^j_N(\alpha,\beta)\, G^{ij}_{M-N}(\alpha,\beta)\,, \tag{3.7}$$

where

$$F^i_N(\alpha,\beta) = \frac{1}{2\pi}\int_0^{2\pi} d\zeta\, e^{-iN\zeta}\, f_i(\zeta)\,, \tag{3.8}$$

$$G^{ij}_N(\alpha,\beta) = \frac{1}{2\pi}\int_0^{2\pi} d\zeta\, e^{-iN\zeta}\, f_i(\zeta)\, f^*_j(\zeta)\,, \tag{3.9}$$

$$H^i_N(\alpha,\beta) = \frac{1}{2\pi}\int_0^{2\pi} d\gamma\, e^{iN\gamma}\, W(\alpha,\beta,\gamma)\, f^*_i(\alpha,\beta,\gamma)\,. \tag{3.10}$$

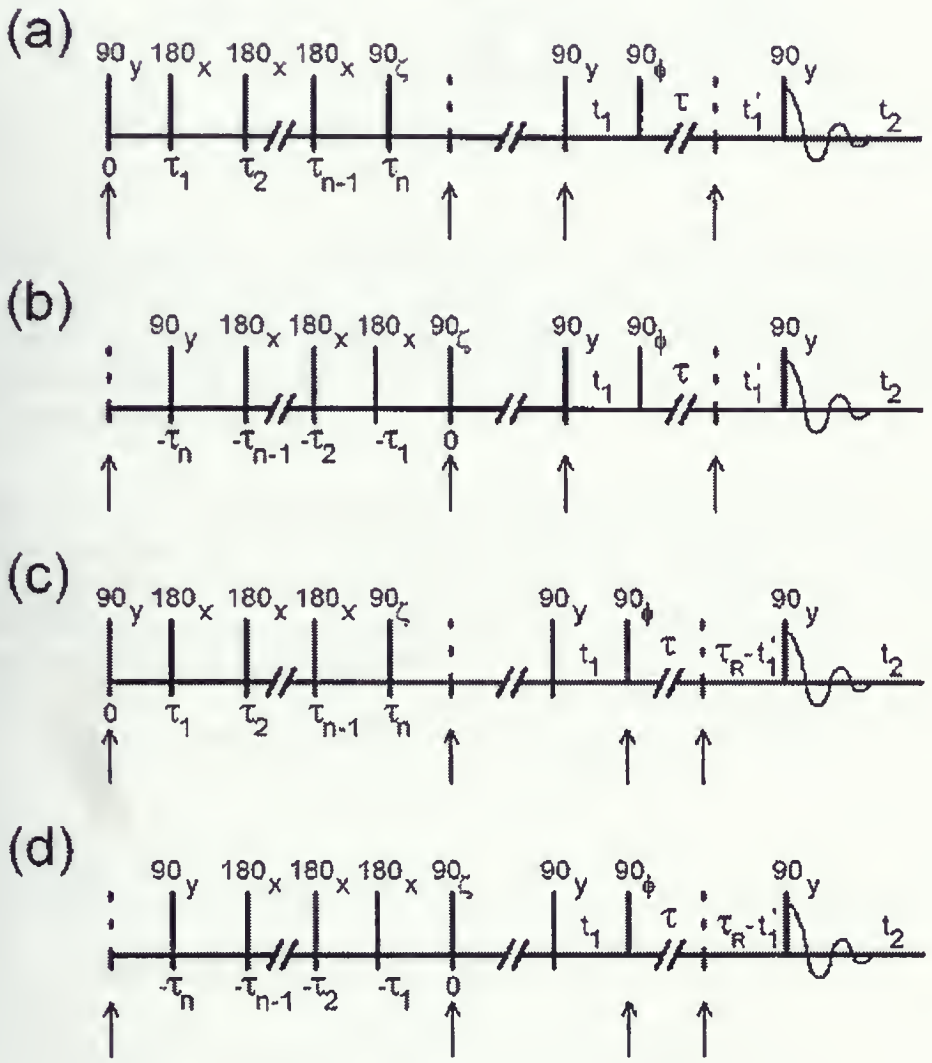


Figure 3.3: General form for rotor-synchronized pulse sequences that produce purely absorptive orientationally weighted 2D MAS NMR exchange spectra. Arrows indicate time points that are synchronized with the beginning of sample rotation periods. Unweighted spectra are obtained by omitting pulses up to and including 90_ζ. Reproduced from [35]

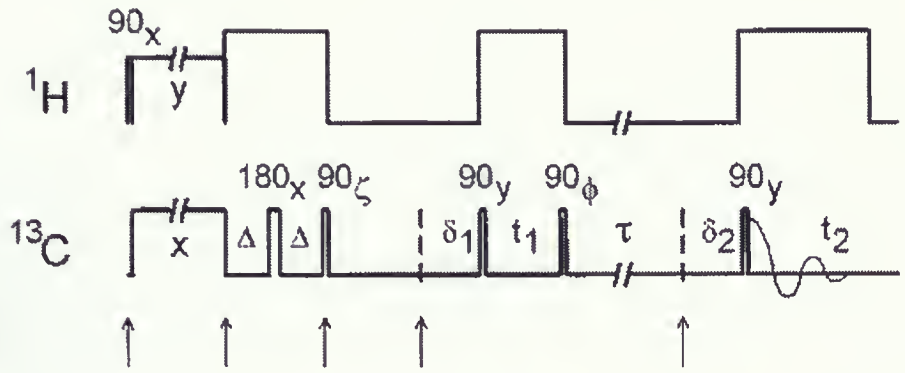


Figure 3.4: RF pulse sequence to obtain experimental 2D MAS exchange spectra. Orientationally weighted spectra are obtained with the sequence as shown, with $\zeta = \pm y$ or $\pm x$ and alternate addition and subtraction of signals. Unweighted spectra are obtained by setting Δ to a value close to zero (0.5 μs) and setting $\zeta = \pm y$. Reproduced from [35]

The function $W(\alpha, \beta, \gamma)$ weights the spinning sideband crosspeak signals at the orientation (α, β, γ) and comes into effect if the initial nuclear spin magnetization distribution is a function of the molecular orientation. The functions $f_i(\zeta) = f_i(\alpha, \beta, \zeta)$ that appropriately express the NMR signals in a rotor-synchronized 2D exchange experiment, as shown in Refs. [37] and [38], have the following form

$$f_i(\zeta) = \exp[i(a_i \sin 2\zeta - b_i \cos 2\zeta + c_i \sin\zeta - d_i \cos\zeta)], \tag{3.11}$$

where

$$\begin{aligned} a_i(\alpha,\beta) &= -\frac{\omega_0}{2\omega_R}\left\{\frac{1}{2}(\sigma^i_{zz}-\sigma^i_{iso})\sin^2\beta - \left[\frac{1}{12}(\sigma^i_{yy}-\sigma^i_{xx})\cos 2\alpha - \frac{1}{6}\sigma^i_{xy}\sin 2\alpha\right]\right. \\ &\quad \left.\times(\cos 2\beta + 3) - \frac{1}{3}(\sigma^i_{xz}\cos\alpha + \sigma^i_{yz}\sin\alpha)\sin 2\beta\right\}, \end{aligned} \tag{3.12}$$

$$\begin{aligned} b_i(\alpha,\beta) &= -\frac{\omega_0}{2\omega_R}\left\{\left[\frac{1}{3}(\sigma^i_{yy}-\sigma^i_{xx})\sin 2\alpha + \frac{2}{3}\sigma^i_{xy}\cos 2\alpha\right]\cos\beta\right. \\ &\quad \left.+\frac{2}{3}(\sigma^i_{xz}\sin\alpha - \sigma^i_{yz}\cos\alpha)\sin\beta\right\}, \end{aligned} \tag{3.13}$$

$$\begin{aligned} c_i(\alpha,\beta) &= -\frac{\omega_0}{\omega_R}\left\{-\frac{1}{\sqrt{2}}(\sigma^i_{zz}-\sigma^i_{iso})\sin 2\beta - \left[\frac{1}{3\sqrt{2}}(\sigma^i_{yy}-\sigma^i_{xx})\cos 2\alpha - \frac{\sqrt{2}}{3}\sigma^i_{xy}\sin 2\alpha\right]\right. \\ &\quad \left.\times\sin 2\beta + \frac{2\sqrt{2}}{3}(\sigma^i_{xz}\cos\alpha + \sigma^i_{yz}\sin\alpha)\sin 2\beta\right\}, \end{aligned} \tag{3.14}$$

$$d_i(\alpha,\beta) = -\frac{\omega_0}{\omega_R}\left\{\left[\frac{\sqrt{2}}{3}(\sigma^i_{yy}-\sigma^i_{xx})\sin 2\alpha + \frac{2\sqrt{2}}{3}\sigma^i_{xy}\cos 2\alpha\right]\sin\beta\right.$$

$$- \frac{2\sqrt{2}}{3} (\sigma^i_{xz} \sin\alpha - \sigma^i_{yz} \cos\alpha) \cos\beta \Big\} . \tag{3.15}$$

In these expressions, σ^i_{jk} are the elements of the CSA tensor for nuclear spin site i represented in a molecule-fixed Cartesian axis system with axes $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. σ^i_{iso} is equal to the average of the CSA principal values σ^i_{11}, σ^i_{22} and σ^i_{33}. The molecule-fixed axis system is mapped to a rotor-fixed axis system $\{\mathbf{x}', \mathbf{y}', \mathbf{z}'\}$ by Euler angles α, β and γ such that α and β are spherical coordinates of the $\mathbf{z}'$ direction in the molecule-fixed axis system. The axis of sample rotation is taken to be $\mathbf{z}'$. The external magnetic field $\mathbf{B}$ of the NMR spectrometer is taken to lie in the $\mathbf{x}'\mathbf{z}'$ plane at time $t = 0$, at the magic angle $\theta_m = \cos^{-1}(1/\sqrt{3})$ to $\mathbf{z}'$ and at angle $\pi/2 - \theta_m$ to $\mathbf{x}'$.

The total spinning sideband crosspeak amplitudes are functions of the anisotropy and asymmetry of the CSA tensors of sites i and j, the ratio ω_0/ω_R, and the relative orientation of the principal axis systems of the two CSA tensors, as described by a set of Euler angles. It is the dependence on the relative orientation of CSA principal axis systems of the exchanging sites that makes rotor-synchronized 2D exchange spectroscopy a useful probe of molecular conformations. In order to relate the spinning sideband crosspeak amplitudes to molecular structure, consider two sites, $i = 1$ and $j = 2$. If the molecular axis system $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is taken to be the principal axis system for the CSA tensor of site 1, $\{\mathbf{x_1}, \mathbf{y_1}, \mathbf{z_1}\}$, then for this site σ^i_{xx}, σ^i_{yy} and σ^i_{zz} in Eqs. 3.12 through 3.15 are equal to the CSA principal values σ^1_{11}, σ^1_{22} and σ^1_{33} and $\sigma^i_{xy} = \sigma^i_{xz} = \sigma^i_{yz} = 0$. For site 2, the CSA elements in Eqs. 3.12 through 3.15 become

$$\sigma^2_{\mu\nu} = \sigma^2_{11} \cos\theta_{x\mu} \cos\theta_{x\nu} + \sigma^2_{22} \cos\theta_{y\mu} \cos\theta_{y\nu} + \sigma^2_{33} \cos\theta_{z\mu} \cos\theta_{z\nu} , \tag{3.16}$$

in terms of the CSA principal values σ^2_{11}, σ^2_{22} and σ^2_{33} for site j and the direction cosines of the CSA principal axis of site 2, $\{\mathbf{x_2}, \mathbf{y_2}, \mathbf{z_2}\}$, in the principal axis system of site 1, *e.g.* $\cos\theta_{xz} = \mathbf{x_2}.\mathbf{z_1}$. In terms of the Euler angles α_{12}, β_{12} and γ_{12} that relate the two CSA principal axis systems,

$$\begin{aligned}
\cos\theta_{xx} &= \cos\gamma_{12} \cos\beta_{12} \cos\alpha_{12} - \sin\gamma_{12} \sin\alpha_{12} , && (3.17)\\
\cos\theta_{xy} &= \cos\gamma_{12} \cos\beta_{12} \sin\alpha_{12} + \sin\gamma_{12} \cos\alpha_{12} , && (3.18)\\
\cos\theta_{xz} &= -\cos\gamma_{12} \sin\beta_{12} , && (3.19)\\
\cos\theta_{yx} &= -\sin\gamma_{12} \cos\beta_{12} \cos\alpha_{12} - \cos\gamma_{12} \sin\alpha_{12} , && (3.20)\\
\cos\theta_{yy} &= -\sin\gamma_{12} \cos\beta_{12} \sin\alpha_{12} + \cos\gamma_{12} \cos\alpha_{12} , && (3.21)\\
\cos\theta_{yz} &= \sin\gamma_{12} \sin\beta_{12} , && (3.22)\\
\cos\theta_{zx} &= \sin\beta_{12} \cos\alpha_{12} , && (3.23)\\
\cos\theta_{zy} &= \sin\beta_{12} \sin\alpha_{12} , && (3.24)\\
\cos\theta_{zz} &= \cos\beta_{12} . && (3.25)
\end{aligned}$$

The Euler angles are defined so that $R_{x_1}(\alpha_{12})R_{y_1}(\beta_{12})R_{z_1}(\gamma_{12})\{\mathbf{x_1}, \mathbf{y_1}, \mathbf{z_1}\} = \{\mathbf{x_2}, \mathbf{y_2}, \mathbf{z_2}\}$, where for example $R_{y_1}(\beta_{12})$ is the right-handed rotation about $\mathbf{y_1}$ by β_{12}.

Fig. 3.5 illustrates a portion of the molecular structure. To a good approximation, the peptide backbone is comprised of a series of planar peptide bond units with relative orientations specified by the dihedral angles ϕ and ψ. Experiments carried out on model compounds

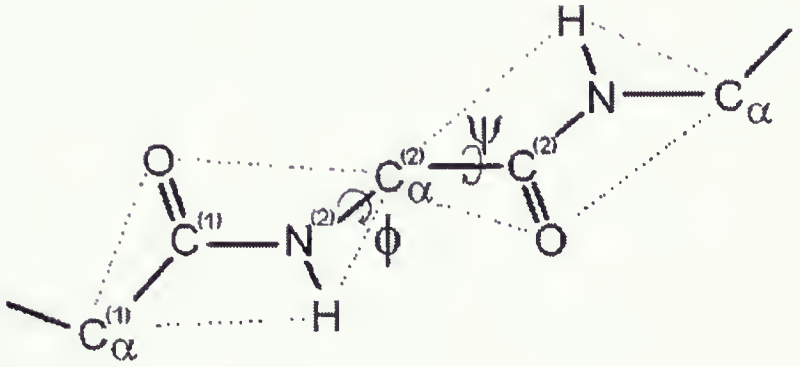


Figure 3.5: Peptide backbone, illustrating the dihedral angles ϕ and ψ that define the backbone conformation at amino acid residue 2. Peptide planes are indicated by dashed lines. The all-trans conformation is shown, for which $\phi = \psi = 180°$. In the experiments on L-alanylglycylglycine (AGG), ^{13}C labels are introduced at carbonyl carbons 1 and 2. Reproduced from [35]

with ^{13}C labels at two successive carbonyl sites in the peptide backbone have shown that the carbonyl ^{13}C CSA principal axis system in a peptide is oriented with **z** (corresponding to σ_{33}) perpendicular to the peptide plane and **y** (corresponding to σ_{22}) nearly parallel to the C–O bond, making an angle $\chi \approx 130°$ with the C–N bond. Assuming this geometry and defining δ_1 to be the angle between the $C^{(1)}$–$N^{(2)}$ bond and the $N^{(2)}$–$C^{(2)}_\alpha$ bond ($\delta_1 \approx 58°$), δ_2 to be the angle between the $N^{(2)}$–$C^{(2)}_\alpha$ bond and the $C^{(2)}_\alpha C^{(2)}$ bond ($\delta_2 \approx 69°$) and δ_3 to be the angle between the $C^{(2)}_\alpha C^{(2)}$ bond and the $C^{(2)}N^{(3)}$ bond ($\delta_3 \approx 64°$), the relation between the Euler angles α_{12}, β_{12} and γ_{12} and the dihedral angles ϕ and ψ can be expressed as

$$\alpha_{12} = \bar{\alpha} + \chi_1 - \delta_1 , \tag{3.26}$$

$$\cos\beta_{12} = \sin\phi \sin\psi \cos\delta_2 - \cos\phi\cos\psi , \tag{3.27}$$

$$\gamma_{12} = \bar{\gamma} - \chi_2 + \delta_3 , \tag{3.28}$$

where χ_1 and χ_2 are the values of χ for sites 1 and 2, and $\bar{\alpha}$ and $\bar{\gamma}$ satisfy

$$\sin\bar{\alpha}\sin\beta_{12} = \sin\psi\sin\delta_2 , \tag{3.29}$$

$$\cos\bar{\alpha}\sin\beta_{12} = -\cos\phi\sin\psi\cos\delta_2 - \sin\phi\cos\psi , \tag{3.30}$$

$$\sin\bar{\gamma}\sin\beta_{12} = -\sin\phi\sin\delta_2 , \tag{3.31}$$

$$\cos\bar{\gamma}\sin\beta_{12} = -\sin\phi\cos\psi\cos\delta_2 - \cos\phi\sin\psi . \tag{3.32}$$

Eqs. 3.16 through 3.32 permit numerical calculations of the 2D spinning sideband crosspeak amplitudes for any values of the dihedral angles ϕ and ψ that specify the peptide backbone conformation.

3.1.2 Constant-time double-quantum-filtered dipolar recoupling

Solid state NMR spectra in different experiments are determined by various nuclear spin interactions, such as Zeeman interaction H_Z, the chemical shift H_{CS}, homonuclear and het-

eronuclear magnetic dipole-dipole interactions H_{II} and H_{IS}, and interactions with applied radio-frequency (rf) fields H_{RF}. NMR experiments are analyzed in the rotating frame, where the Hamiltonians corresponding to these interactions have the following forms

$$H_Z = \Delta\omega\, I_z\,, \tag{3.33}$$

$$H_{CS} = \omega_0 \left(\delta_{iso} + \delta_{11}\sin^2\theta\cos^2\phi + \delta_{22}\sin^2\theta\sin^2\phi + \delta_{33}\cos^2\theta \right), \tag{3.34}$$

$$H_{II} = \frac{-\gamma_I^2\hbar}{r_{12}^3}\frac{(3\cos^2\theta' - 1)}{2}\left(3\, I_{z1}\, I_{z2} - I_1.I_2 \right), \tag{3.35}$$

$$H_{IS} = \frac{-2\gamma_I\gamma_S\hbar}{r_{IS}^3}\frac{(3\cos^2\theta'' - 1)}{2}\, I_z\, S_z\,, \tag{3.36}$$

$$H_{RF} = \omega_1(t)\left[I_x\cos\chi(t) + I_y\sin\chi(t) \right]. \tag{3.37}$$

In these equations: I and S are spin angular momentum vector operators for two different spin species. $\Delta\omega$ is the resonance offset (the difference between the nuclear Larmor frequency ω_0 and the rf carrier frequency ω). δ_{iso} is the isotropic chemical shift and δ_{11}, δ_{22} and δ_{33} are the principal values of the chemical shift anisotropy (CSA) tensor. θ and ϕ angles specify the direction of the magnetic field in the CSA principal axis system. γ_I is the I spin magnetogyric ratio, r_{12} is the distance between the two coupled I spins, and θ' is the angle between the internuclear displacement vector and the externally applied magnetic field (directed along the laboratory z axis). Similar definitions hold for γ_S, r_{IS} and θ''. $\omega_1(t)$ and $\chi(t)$ are the time- dependent rf field amplitude and phase, with respect to a constant-phase carrier signal at frequency ω. In system of many spins, H_{CS}, H_{II} and H_{IS} would contain many terms with different values of the various parameters for each type of spin or spin pair. In heteronuclear spin system, H_Z and H_{RF} would contain one term for each spin species.

Obviously, anisotropic chemical shifts and nuclear magnetic dipole-dipole couplings contain structural information through the dependence on internuclear distances and orientations. The orientation dependences of H_{CS}, H_{II} and H_{IS} result in a strong orientation dependence of the NMR frequencies. Therefore, NMR spectra of polycrystalline and noncrystaline solids, in which the molecules in the sample take on all possible orientations (with respect to the external magnetic field) randomly, exhibit inhomogeneously broadened lines, called powder patterns. Fig. 3.6 a shows a simple example of a powder pattern lineshape for the case of a ^{13}C-labeled compound in polycrystalline form. Although powder pattern lineshapes can contain structural information, broad lineshapes result in lower resolution and sensitivity. To achieve high resolution and sensitivity, solid state NMR experiments on unoriented samples are carried out with magic-angle spinning (MAS). The line narrowing effect of MAS on solid state NMR spectra is shown in Fig. 3.6 b-d. At the magic angle the orientation-dependent parts of H_{CS}, H_{II} and H_{IS} average to zero over one rotation period τ_R. When the MAS rotation frequency ν_R greatly exceeds the inhomogeneous linewidth, the powder pattern collapses into a single, much sharper line at the isotropic chemical shift frequency. Since the total area of the lineshape is unchanged, the signal-to-noise ratio increases as the MAS rotation frequency, ν_R, increases.

There are two important problems with the MAS technique. First, since the orientation-dependent parts of the dipole-dipole couplings are averaged out, the structural information

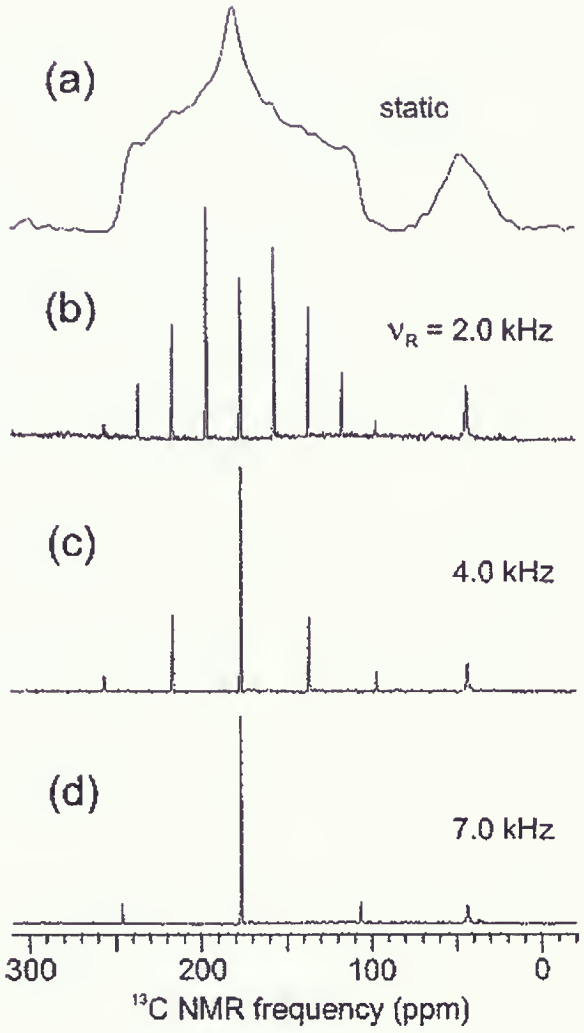


Figure 3.6: (a) ^{13}C NMR spectrum of static, polycrystalline glycine with approximately 5% ^{13}C enrichment at the carboxyl site. A chemical shift anisotropy powder pattern lineshape is observed for the carboxyl carbons, spanning the 100-250 ppm range. (b-d) Spectra obtained with magic-angle spinning at the indicated spinning frequencies ν_R. Lines at 177 ppm and 43 ppm are centerbands at the isotropic chemical shifts of the labeled carboxyl and natural-abundance α-carbon sites. Lines whose frequencies vary with ν_R are spinning sidebands. Spectra obtained at 100.4 MHz ^{13}C NMR frequency. Reproduced from [39].

contained in these couplings is lost. Second, the line narrowing is quite often insufficient to produce a spectrum in which resonances from inequivalent nuclei are resolved, *e.g.* in ^{13}C MAS spectra of large molecules such as peptides. In such situations, the spectrum should be simplified by selecting certain resonances of interest and suppressing the others.

Regarding the first problem, a wide variety of techniques have been devised to maintain both the high resolution of MAS and the information contained in homonuclear and heteronuclear dipole-dipole couplings. A large number of references can be found in [40] (page 9444) and [39] (page 582) regarding both homonuclear and heteronuclear dipolar recoupling techniques. The simplest technique is to spin slowly. As shown in Fig. 3.4, the spinning sidebands become more numerous and intense as ν_R decreases. Most of the structural information contained in powder pattern lineshapes is also contained in spinning sidebands under slow MAS (*i.e.* $2\pi\nu_R < \omega_0|\delta_{11} - \delta_{33}|$). Another technique is to combine MAS with simple resonant radio-frequency (rf) pulse sequences in synchrony with the sample rotation [41]. Interference between the time dependence induced by MAS and the time dependence induced by pulse sequences can prevent anisotropic spin interactions from being averaged to zero, even under fast MAS. This phenomenon has come to be called "recoupling".

"Dipolar recovery at the magic angle" (DRAMA) [40] is a well-known recoupling technique in the case of homonuclear dipole-dipole coupling, *e.g.* ^{13}C–^{13}C or ^{15}N–^{15}N couplings. The basic element of this technique is the pulse sequence illustrated in Fig. 3.7 a. MAS alone makes the $(3\cos^2\theta - 1)$ term in Eq. 3.35 time dependent, with terms that oscillate at ν_R and $2\nu_R$ and therefore average to zero. The rf pulses in DRAMA technique introduce an additional time dependence of the spin operator term $(3\,I_{z1}\,I_{z2} - I_1.I_2)$ by rotating this term about x by 90° in the period between the two pulses. The Hamiltonian H_{II} then reduces to

$$\widetilde{H}_{II}(t) = \frac{-\gamma_I^2\hbar}{r_{12}^3}(C_1\cos\omega_R t + S_1\sin\omega_R t + C_2\cos 2\omega_R t + S_2\sin 2\omega_R t)\,\widetilde{T}(t)\,, \tag{3.38}$$

where

$$\widetilde{T}(t) = \begin{cases} 3I_{z1}\,I_{z2} - I_1.I_2\,, & 0 < t < \tau' \\ 3\,I_{y1}\,I_{y2} - I_1.I_2\,, & \tau' < t < \tau' + \tau \\ 3\,I_{z1}\,I_{z2} - I_1.I_2\,, & \tau' + \tau < t < \tau_R \end{cases} \tag{3.39}$$

The coefficients C_1, S_1, C_2 and S_2 are functions of the internuclear direction in an axis system fixed in the MAS rotor. Because of the time dependence of the operator $\widetilde{T}(t)$, the time average of $\widetilde{H}_{II}(t)$, which determines the effect of $\widetilde{H}_{II}(t)$ to lowest order becomes

$$\left\langle\widetilde{H}_{II}(t)\right\rangle = \frac{-\gamma_I^2\hbar}{r_{12}^3}\left[C_1\,\frac{\sin\omega_R(\tau'+\tau) - \sin\omega_R\tau'}{2\pi} + C_2\,\frac{\sin 2\omega_R(\tau'+\tau) - \sin 2\omega_R\tau'}{4\pi}\right](3\,I_{y1}\,I_{y2} - 3\,I_{z1}\,I_{z2})\,, \tag{3.40}$$

which is generally nonzero. Because C_1 and C_2 are orientation dependent, spectra acquired with DRAMA (or other recoupling techniques) exhibit powder pattern lineshapes that are dependent on the details of the pulse sequence, as shown in Fig. 3.7 b.

"Radio-frequency-driven recoupling" (RFDR) [42, 43, 44] is another class of recoupling techniques in the case of dipole-dipole homonuclear spin systems with large chemical shift or

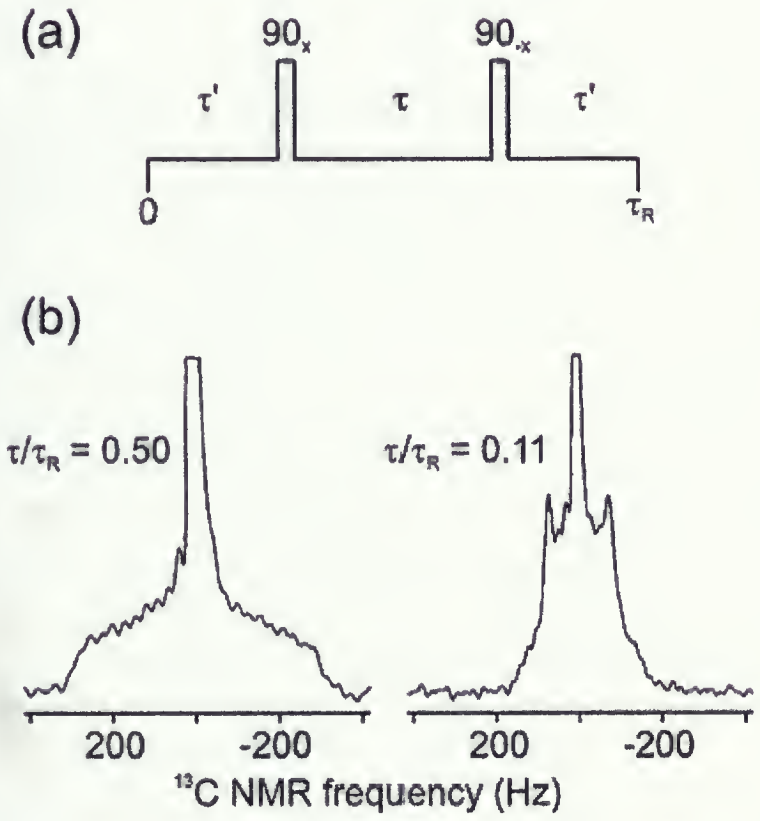


Figure 3.7: (a) Radio-frequency pulse sequence for dipolar recovery at the magic angle (DRAMA), consisting of two $\pi/2$ pulses with phases x and $-$x per MAS rotation period R. (b) ^{13}C NMR spectra of a doubly ^{13}C-labeled model compound (bisulfite adduct of acetone) under the DRAMA sequence. The powder pattern lineshapes illustrate the recoupling effect, i.e. the recovery of ^{13}C-^{13}C dipole-dipole couplings that would be averaged to zero by MAS in the absence of the rotation-synchronized pulses. Spectra obtained at 25.3 MHz for ^{13}C NMR frequency, with $R = 3.33$ kHz. Reproduced from [39].

CSA differences. In this class of techniques recoupling occurs owing to interference between time dependences induced by MAS and chemical shift differences, rather than by rf pulse sequences. RFDR has the advantages of being insensitive to inhomogeneous broadening of the NMR lines and extremely sparse, so that effects of pulse imperfections and signal losses due to insufficient proton decoupling are minimized. The average dipole-dipole coupling Hamiltonian under RFDR has the form $\left\langle \widetilde{\mathrm{H}}_{II}(t) \right\rangle = d(\mathrm{I}_{1+}\mathrm{I}_{2-} + \mathrm{I}_{1-}\mathrm{I}_{2+})$, where I_{i+} is the angular momentum raising operator for spin i and d is the effective coupling constant [42, 43, 44].

Both DRAMA and RFDR sequences can be incorporated into a double-quantum filtering [40, 45, 46, 47] technique that address the second problem with the MAS technique. Double-quantum (DQ) filtering makes it possible to pick out only those resonances that arise from dipole-coupled pairs of spins in an MAS spectrum and to suppress resonances from isolated, uncoupled spins. In a DQ-filtered measurement, DQ-coherences (*i.e.* superpositions of the $|++>$ and $|-->$ spin states of a two-spin system) are prepared during the preparation period and are allowed to evolve during the evolution period. The basic sequence for double-quantum filtering has the form $\mathrm{CP}_{90+\zeta}-(\mathrm{RFDR})_L-90_\zeta-90_0-(\mathrm{RFDR})_M-\mathrm{FID}$, where CP represents cross-polarization from protons to ^{13}C nuclei, 90_ζ represents a $\pi/2$ pulse with rf phase ζ, $(\mathrm{RFDR})_L$ represents an RFDR train lasting L MAS rotor periods τ_R, and FID represents the detection of free-induction-decay signals. FIDs are recorded with ζ =0°, 90°, 180° and 270° and coadded after multiplication by $\mathrm{e}^{2i\zeta}$. In order to reduce the effect of residual carbon-proton couplings and transverse relaxation processes on the dipolar evolution curves, constant-time (CT) version of DQ-filtering (CTDQFD) is introduced by extending the sequence to $\mathrm{CP}_{90+\zeta}-(\mathrm{RFDR})_L-90_\zeta-90_0-(\mathrm{RFDR})_M-90_{180}-90_{90}-(\mathrm{RFDR})_N-\mathrm{FID}$. For a pair of spin-1/2 nuclei with coupling $\widetilde{\mathrm{H}}_{II}(t)$, the second pair of $\pi/2$ pulses can be shown to refocus dipolar evolution. In a real system, the extended sequence allows the total recoupling period $T = (L+M+N)\tau_R$ to be kept constant throughout an experiment while the effective dipolar evolution periods are varied by varying L, M and N.

No analytical expressions exist for calculating the amplitudes of double-quantum filtered NMR signals. Instead, the amplitudes can be calculated by carrying out quantum dynamical simulations based on density operators [40, 48].

3.2 Simulated data

The distribution function $g(\phi, \psi)$, to be calculated through inversion, comprised of two Gaussian-shaped blobs in the (ϕ, ψ) plane. One blob is centered at $(\phi, \psi) = -60°, -40°$, the other at $(\phi, \psi) = -130°, +130°$. The first has a width of 30, the second has a width of 40. The first has a peak amplitude of 0.1, the second has a peak amplitude of 0.05. After this weighting function is built up, the data vector s is generated by doing forward calculations:

$$\mathbf{s} = h_\phi \, h_\psi \, \mathbf{K} \, \mathbf{g} \, . \tag{3.41}$$

Then, several noise levels, 0.1%, 0.5%, 1% and 5%, are added to the data.

3.3 Discrete implementation of the Laplacian

3.3.1 Cartesian Coordinates System

Considering the Laplacian in Cartesian system of coordinates, the additional constraint in solving the LS problem contains the term $\mathbf{L}g(x,y)$ which can be written as

$$\nabla^2 g(x,y) = \frac{\partial^2}{\partial x^2} g(x,y) + \frac{\partial^2}{\partial y^2} g(x,y) \tag{3.42}$$

Consider the function $g(\eta)$ whose discrete counterpart, vector $\mathbf{g}$, has the elements $g_i = g(\eta_i)$, where $i = 0, \ldots .(n-1)$. The second derivative of $\mathbf{g}$ has elements

$$g_i'' = \frac{1}{h^2} \left(g_{i-1} - 2\, g_i + g_{i+1} \right) , \tag{3.43}$$

where h is the step-size of discretization. The special cases, g_{i-1} for $i = 0$ and g_{i+1} for $i = (n-1)$, must be treated by appropriate boundary conditions. Also, the second derivative operator is a matrix with elements

$$\begin{aligned} D_{i,j} &= \frac{1}{h^2} \left(\delta_{i-1,j} - 2\, \delta_{i,j} + \delta_{i+1,j} \right) \\ i &= 0, \ldots, (n-1) - 2 , \quad j = 0, \ldots, (n-1) . \end{aligned} \tag{3.44}$$

For a two-dimensional function $g(x,y)$ the vector $\mathbf{g}$ is defined according to Eq. 3.2. The partial derivative in x-direction, $\frac{\partial^2}{\partial x^2}$, will move the k-index of g_{kl} and the partial derivative in y-direction, $\frac{\partial^2}{\partial y^2}$, will move the l-index of g_{kl} according to the rule given in Eq. 3.43. These will result in

$$\begin{aligned} \left(\nabla^2 g(x,y)\right)_{k,l} &= \frac{1}{h_x^2} \left(g_{k-1,l} - 2\, g_{k,l} + g_{k+1,l} \right) + \frac{1}{h_y^2} \left(g_{k,l-1} - 2\, g_{k,l} + g_{k,l+1} \right) \\ k &= 0, \ldots, (n_x - 1) , \quad l = 0, \ldots, (n_y - 1) . \end{aligned} \tag{3.45}$$

In this equation the boundary points must be considered to define appropriate values for $g_{k-1,l}$ at $k = 0$, for $g_{k+1,l}$ at $k = (n_x - 1)$, for $g_{k,l-1}$ at $l = 0$ and for $g_{k,l+1}$ at $k = (n_y - 1)$. If the grid spacing is the same in both dimensions, the elements of $\nabla^2 g(x,y)$ can be expressed as

$$\left(\nabla^2 g(x,y)\right)_{k,l} = \frac{4}{h^2} \left(\frac{g_{k-1,l} + g_{k+1,l} + g_{k,l-1} + g_{k,l+1}}{4} - g_{k,l} \right) , \tag{3.46}$$

The quantity within the brackets is the difference between the value of $g(x,y)$ at a certain point and its average at the nearest neighbors. This is called the *finite-difference method.*

In order to make a vector out of $(\nabla^2 g)_{k,l}$, we number the two dimensions of grid points in a single one-dimensional sequence according to Eq. 2.82. Hence, the components of the underlying vector are given by

$$\begin{aligned} \left(\nabla^2 g\right)_i &= \frac{1}{h_x^2} \left(g_{i-n_y} - 2\, g_i + g_{i+n_y} \right) + \frac{1}{h_y^2} \left(g_{i-1} - 2\, g_i + g_{i+1} \right) \\ i &= 0, \ldots, (n-1) , \quad n = n_x\, n_y . \end{aligned} \tag{3.47}$$

which can be written as

$$\begin{aligned}(\nabla^2 g)_i &= \frac{1}{h_x^2}\left(\delta_{i-n_y,j} - 2\,\delta_{i,j} + \delta_{i+n_y,j}\right) g_j + \frac{1}{h_y^2}\left(\delta_{i-1,j} - 2\,\delta_{i,j} + \delta_{i+1,j}\right) g_j\,,\\ i &= 0,\ldots,(n-1)\,,\\ j &= 0,\ldots,(n-1)\,,\end{aligned} \tag{3.48}$$

and results in an expression for the matrix structure of the Laplacian in Cartesian coordinates system:

$$\begin{aligned}(\nabla^2)_{i,j} &= \frac{1}{h_x^2}\left(\delta_{i-n_y,j} - 2\,\delta_{i,j} + \delta_{i+n_y,j}\right) + \frac{1}{h_y^2}\left(\delta_{i-1,j} - 2\,\delta_{i,j} + \delta_{i+1,j}\right)\,,\\ i &= 0,\ldots,(n-1)\,,\\ j &= 0,\ldots,(n-1)\,,\end{aligned} \tag{3.49}$$

The boundary conditions originally defined for g_{kl} must now be considered for g_i:

$$\begin{aligned} g_{k-1,l} \quad &\text{for } k=0 \iff g_{i-n_y} \quad \text{for } i=0,\ldots,(n_y-1)\,,\\ g_{k+1,l} \quad &\text{for } k=(n_x-1) \iff g_{i+n_y} \quad \text{for } i=(n-n_y),\ldots,(n-1)\,,\\ g_{k,l-1} \quad &\text{for } l=0 \iff g_{i-1} \quad \text{for } i=0,n_y,\ldots,(n_x-1)n_y\,,\\ g_{k,l+1} \quad &\text{for } l=(n_y-1) \iff g_{i+1} \quad \text{for } i=(n_y-1),(2n_y-1),\ldots,(n_xn_y-1)\end{aligned} \tag{3.50}$$

For instance, in the case of periodic boundary conditions different cases to be considered are as follows

1. $0 \le i \le (n_y-1)$,
which arises from $k=0$. $g_{k-1,l}$ should be replaced by $g_{k-1+n_x,l}$ and consequently g_{i-n_y} by g_{i-n_y+n}. Hence, in Eq. 3.49, $\delta_{i-n_y,j}$ should be replaced by $\delta_{i-n_y+n,j}$.

2. $(n-n_y) \le i \le (n-1)$,
which arises from $k=(n_x-1)$. $g_{k+1,l}$ should be replaced by $g_{k+1-n_x,l}$ and consequently g_{i+n_y} by g_{i+n_y-n}. Hence, in Eq. 3.49, $\delta_{i+n_y,j}$ should be replaced by $\delta_{i+n_y-n,j}$.

3. i **is a multiple of** n_y ,
which arises from $l=0$. $g_{k,l-1}$ should be replaced by $g_{k,l-1+n_y}$ and consequently g_{i-1} by g_{i-1+n_y}. Hence, in Eq. 3.49, $\delta_{i-1,j}$ should be replaced by $\delta_{i-1+n_y,j}$.

4. $(i+1)$ **is a multiple of** n_y ,
which arises from $l=(n_y-1)$. $g_{k,l+1}$ should be replaced by $g_{k,l+1-n_y}$ and consequently g_{i+1} by g_{i+1-n_y}. Hence, in Eq. 3.49, $\delta_{i+1,j}$ should be replaced by. $\delta_{i+1-n_y,j}$

3.3.2 Spherical Coordinates System

In spherical coordinates system, the function g is a distribution of the polar angles ϕ and ψ and the Laplacian has singularities at $\sin\psi = 0$. Therefore in solving the LS-problem the poles must be excluded from the range of ψ:

$$\nabla^2 = \frac{1}{\sin^2\psi}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial\psi^2} + \cot\psi\,\frac{\partial}{\partial\psi}\,, \quad \sin\psi \neq 0\,. \tag{3.51}$$

The two-dimensional grid points and the components of the vector **g** are defined as

$$\begin{aligned} g_{kl} &= g(\phi_k, \psi_l)\,, \\ \phi_k &= \phi_0 + k\,h_\phi\,, \quad k = 0, \ldots, (n_\phi - 1)\,, \quad n_\phi = \frac{\phi_{n_\phi - 1} - \phi_0}{h_\phi} + 1\,, \\ \psi_l &= \psi_0 + l\,h_\psi\,, \quad l = 1, \ldots, (n_\psi - 2)\,, \quad n_\psi = \frac{\phi_{n_\psi - 1} - \psi_0}{h_\psi} + 1\,, \end{aligned} \tag{3.52}$$

where $\phi_0 = \psi_0 = -180°$, $\phi_{n_\phi - 1} = +180°$ and $\phi_{n_\psi - 1} = 0°$. Note that in Eq. 3.52 the possibilities $l = 0$ and $l = n_\psi - 1$ are excluded from the range of l. Because, for those values of l the quantity $\sin(\psi_0 + l\,h_\psi)$ is zero and the Laplacian is not defined. The partial derivatives with respect to ψ, *i.e.* $\frac{\partial^2}{\partial \psi^2}$ and $\frac{\partial}{\partial \psi}$, move the l index in $g_{k,l}$. While, The partial derivative $\frac{\partial^2}{\partial \phi^2}$ will move the k index in $g_{k,l}$. The rule for the second derivative is the same as in Eq. 3.43, for example

$$\left(\frac{\partial^2}{\partial \phi^2} g(\phi, \psi)\right)_{k,l} = \frac{1}{h_\phi^2}\,(\,g_{k-1,l} - 2\,g_{k,l} + g_{k+1,l}\,) \tag{3.53}$$

For the first derivative $\frac{\partial}{\partial \psi} g(\phi, \psi)$ the components are given by

$$\left(\frac{\partial}{\partial \psi} g(\phi, \psi)\right)_{k,l} = \frac{1}{2\,h_\psi}\,(\,g_{k,l+1} - g_{k,l-1}\,)\,. \tag{3.54}$$

Therefore the components of the quantity $\nabla^2 g(\phi, \psi)$ in spherical coordinates are given by

$$\begin{aligned} (\nabla^2 g)_{k,l} &= \frac{1}{h_\phi^2 \sin^2(\psi_0 + l\,h_\psi)}\,(\,g_{k-1,l} - 2\,g_{k,l} + g_{k+1,l}\,) + \frac{1}{h_\psi^2}\,(\,g_{k,l-1} - 2\,g_{k,l} + g_{k,l+1}\,) \\ &+ \cot(\psi_0 + l\,h_\psi)\,\frac{1}{2\,h_\psi}\,(\,g_{k,l+1} - g_{k,l-1}\,)\,, \end{aligned} \tag{3.55}$$

which can be written as

$$\begin{aligned} (\nabla^2 g)_{k,l} &= \frac{1}{h_\phi^2 \sin^2(\psi_0 + l\,h_\psi)}\,(\,g_{k-1,l} + g_{k+1,l}\,) \\ &+ \left(\frac{1}{h_\psi^2} - \frac{\cot(\psi_0 + l\,h_\psi)}{2\,h_\psi}\right) g_{k,l-1} + \left(\frac{1}{h_\psi^2} + \frac{\cot(\psi_0 + l\,h_\psi)}{2\,h_\psi}\right) g_{k,l+1} \\ &- 2\left(\frac{1}{h_\psi^2} + \frac{1}{h_\phi^2 \sin^2(\psi_0 + l\,h_\psi)}\right) g_{k,l}\,. \end{aligned} \tag{3.56}$$

Now, in order to make a vector out of $(\nabla^2 g)$ the double index (k, l) can be transformed into a single index (i) through $i = k\,(n_\psi - 2) + l$, where $i = 1, \ldots, (n - 2n_\phi)$ and $n = n_\phi n_\psi$. Then, the index 'l' is equal to 'i mod $(n_\psi - 2)$' and the components of the underlying vector are

$$(\nabla^2 g)_i = \frac{1}{h_\phi^2 \sin^2(\psi_0 + (i \bmod (n_\psi - 2))\,h_\psi)}\,(\,g_{(i-n_\psi+2)} + g_{(i+n_\psi-2)}\,)$$

$$+ \left(\frac{1}{h_\psi^2} - \frac{\cot(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)}{2\, h_\psi} \right) g_{i-1}$$
$$+ \left(\frac{1}{h_\psi^2} + \frac{\cot(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)}{2\, h_\psi} \right) g_{i+1}$$
$$- 2 \left(\frac{1}{h_\psi^2} + \frac{1}{h_\phi^2 \sin^2(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)} \right) g_i \,. \qquad (3.57)$$

From this, the matrix representation of the Laplacian in spherical coordinates system is obtained:

$$\begin{aligned}
(\nabla^2)_{i,j} &= \frac{1}{h_\phi^2 \sin^2(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)} \left(\delta_{(i-n_\psi+2),j} + \delta_{(i+n_\psi-2),j} \right) \\
&+ \left(\frac{1}{h_\psi^2} - \frac{\cot(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)}{2\, h_\psi} \right) \delta_{i-1,j} \\
&+ \left(\frac{1}{h_\psi^2} + \frac{\cot(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)}{2\, h_\psi} \right) \delta_{i+1,j} \\
&- 2 \left(\frac{1}{h_\psi^2} + \frac{1}{h_\phi^2 \sin^2(\psi_0 + (i \bmod (n_\psi - 2))\, h_\psi)} \right) \delta_{i,j} \,, \\
i &= 1, \ldots, (n - 2n_\phi) \,, \quad j = 1, \ldots, (n - 2n_\phi) \,. \qquad (3.58)
\end{aligned}$$

As in the case of the Cartesian coordinates system, appropriate boundary conditions must be considered to treat the boundary points properly, they are

$$\begin{aligned}
g_{(i-n_\psi+2)} &\quad \text{for} \quad i = 1, \ldots, (n_\psi - 2) \,, \\
g_{(i+n_\psi-2)} &\quad \text{for} \quad i = (n - 2n_\phi - n_\psi + 3), \ldots, (n - 2n_\phi) \,, \\
g_{i-1} &\quad \text{for} \quad i = 1, (n_\psi - 2) + 1, 2(n_\psi - 2) + 1, \ldots, (n_\phi - 1)(n_\psi - 2) + 1 \,, \\
g_{i+1} &\quad \text{for} \quad i = (n_\psi - 2), 2(n_\psi - 2), \ldots, n_\phi(n_\psi - 2) \,. \qquad (3.59)
\end{aligned}$$

In the case of periodic boundary conditions different cases to be considered are as follows

1. $1 \leq i \leq (n_\psi - 2)$,
which arises from $k = 0$. The term $g_{k-1,l}$ should be replaced by $g_{(k-1+n_\phi),l}$ and consequently $g_{(i-n_\psi+2)}$ by $g_{(i-n_\psi+2+n-2n_\phi)}$. Hence, in Eq. 3.58, $\delta_{(i-n_\psi+2),j}$ should be replaced by $\delta_{(i-n_\psi+2+n-2n_\phi),j}$.

2. $(n - 2n_\phi - n_\psi + 3) \leq i \leq (n - 2n_\phi)$,
which arises from $k = (n_\phi - 1)$. The term $g_{k+1,l}$ should be replaced by $g_{(k+1-n_\phi),l}$ and consequently $g_{(i+n_\psi-2)}$ by $g_{(i+n_\psi-2-n+2n_\phi)}$. Hence, in Eq. 3.58, $\delta_{(i+n_\psi-2),j}$ should be replaced by $\delta_{(i+n_\psi-2-n+2n_\phi),j}$.

3. $(i - 1)$ **is a multiple of** $(n_\psi - 2)$,
which arises from $l = 1$. The term $g_{k,l-1}$ should be replaced by 0 and consequently g_{i-1} by 0. Hence, in Eq. 3.58, $\delta_{i-1,j}$ should be replaced by 0.

4. i **is a multiple of** $(n_\psi - 2)$,
which arises from $l = (n_\psi - 2)$. The term $g_{k,l+1}$ should be replaced by 0 and consequently g_{i+1} by 0. Hence, in Eq. 3.58, $\delta_{i+1,j}$ should be replaced by 0.

Chapter 4

Results and discussion

As we should recall, our purpose is to extract the Ramachandran angles distribution function $g(\phi, \psi)$ via inverse theory. As described in chapter 3, our kernel $\mathbf{K}$ is an $m \times n$ matrix whose elements, $K_{i,j}$'s, are calculated based on "CTDQFDR" and "2D MAS exchange" experiments. This kernel maps an assumed function of ϕ and ψ angles onto a set of data points. Noise is added to these data points and finally a simulated data vector, $\mathbf{s}^\sigma$, is obtained. Then, in order to recover the shape of the distribution function via inverse theory, three different approaches are examined and compared together. The quantities necessary for this purpose are the data vector $\mathbf{s}'^\sigma$, the matrix $\mathbf{K}'$ and the operator $\mathbf{L}'$ introduced in chapter 2:

$$s_i'^\sigma = \frac{1}{\sigma_i} s_i^\sigma, \quad i = 1, \ldots, m, \tag{4.1}$$

$$K'_{ij} = \frac{h_\phi \, h_\psi}{\sigma_i} K(\omega_i, x_j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{4.2}$$

$$L'_{ij} = \sqrt{h_\phi \, h_\psi} \, L_{ij}, \quad i = 1, \ldots, n', \quad j = 1, \ldots, n. \tag{4.3}$$

4.1 Non-Negative Least Squares (NNLS)

The simplest way of solving Fredholm integral equations of the first kind is to ask for a solution that provides the highest compatibility to the data, *i.e.* the lowest misfit. The misfit is the discrepancy between the measured, or simulated, data, $\mathbf{s}'^\sigma$, and the recalculated data, $\mathbf{K}' \, \mathbf{g}^\sigma$. In other words, one looks for a vector $\mathbf{g}^\sigma$ that minimizes

$$\| \mathbf{s}'^\sigma - \mathbf{K}' \, \mathbf{g} \|^2 . \tag{4.4}$$

This procedure is called non-negative least-squares (NNLS) approach. The result, $\mathbf{g}^\sigma$, obtained through this procedure is given at the top of the Fig. 4.1 for 5° resolution using 0.1% noise. The results for 10° and 15° resolutions are given in Fig. 4.2, and Fig. 4.3 (top) shows the result for 20°, all using 0.1% noise level. For a large grid density, $\Delta\phi = \Delta\psi = 5°$, the problem is vastly underdetermined and as one can see in Fig. 4.1 (top) the calculated $\mathbf{g}^\sigma$ shows artificial peaks which are physically meaningless and as the grid density decreases, $\Delta\phi = \Delta\psi = 20°$, the problem is less underdetermined and the calculated $\mathbf{g}^\sigma$ gets smoother and it recovers the true vector $\mathbf{g}$ better.

In Fig. 4.1 (bottom) the misfit function is shown for 5° resolution, while Fig. 4.3 (bottom) shows the one for 20°. Notice that for 5° resolution there is a huge misfit for the first couple of points on the x-axis and as we go to a lower resolution the misfit gets smaller. The influence of the underdetermination of the problem on the quality of the fit can be seen by comparing the standard deviation of the two misfit functions, $\sigma = 5.963$ for 5° and $\sigma = 0.3796$ for 20°.

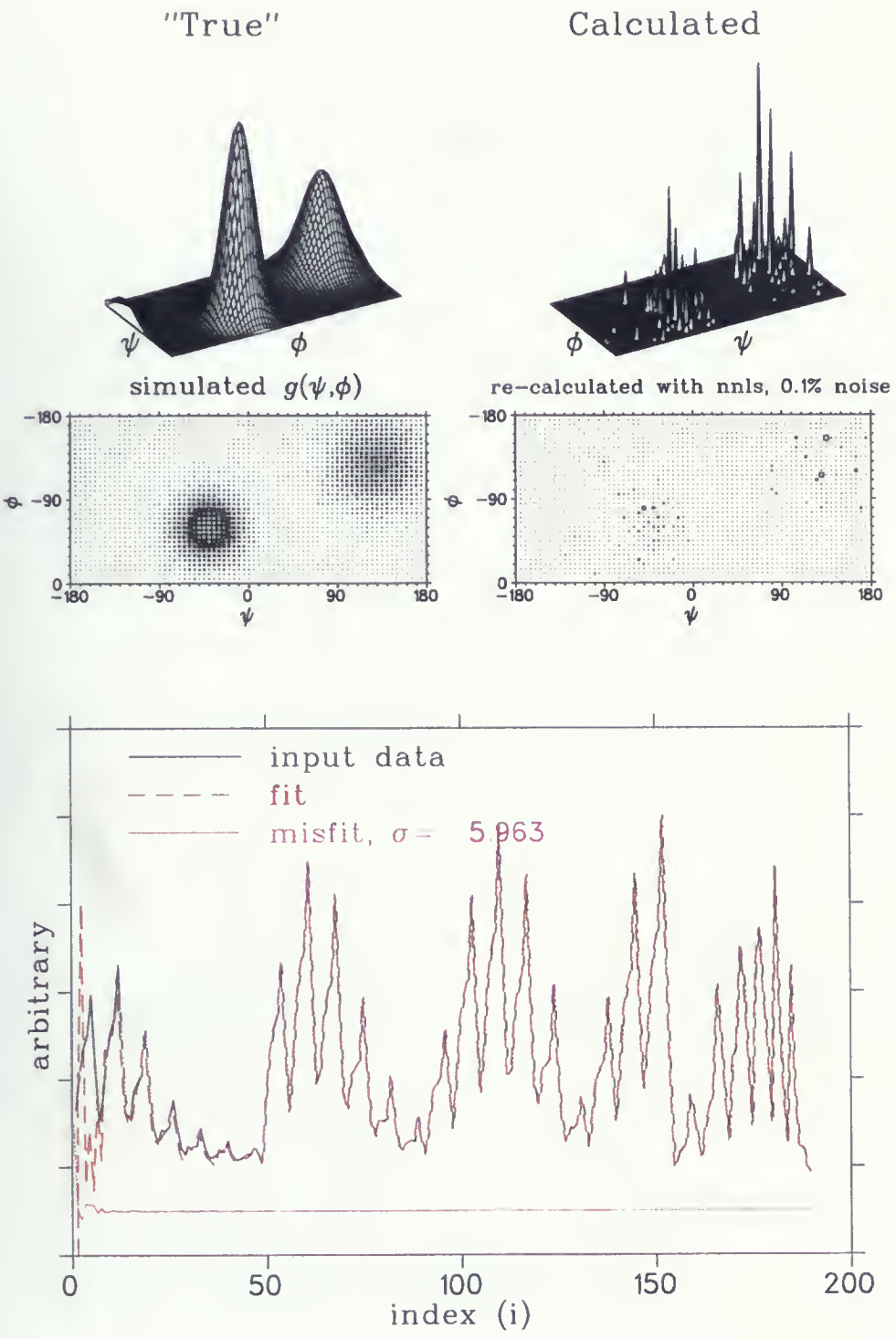


Figure 4.1: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, using NNLS. **Top**: 0.1% noise level (fit-5-0.1.nnls); **bottom**: misfit for the above (misfit-5-0.1.nnls).

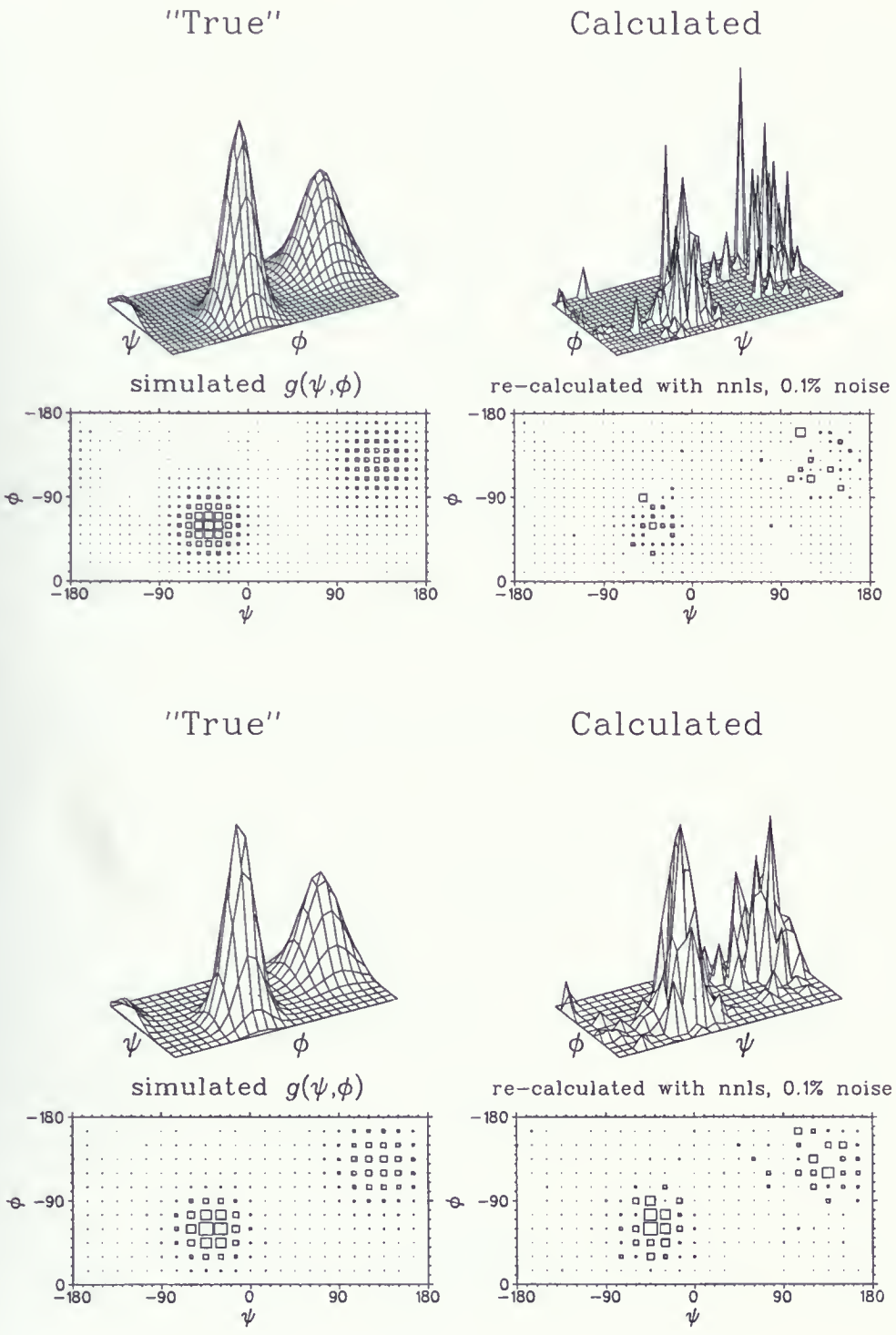


Figure 4.2: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, using NNLS. **Top**: $\Delta\phi = \Delta\psi = 10°$ (fit-10-0.1.nnls); **bottom**: $\Delta\phi = \Delta\psi = 15°$ (fit-15-0.1.nnls)

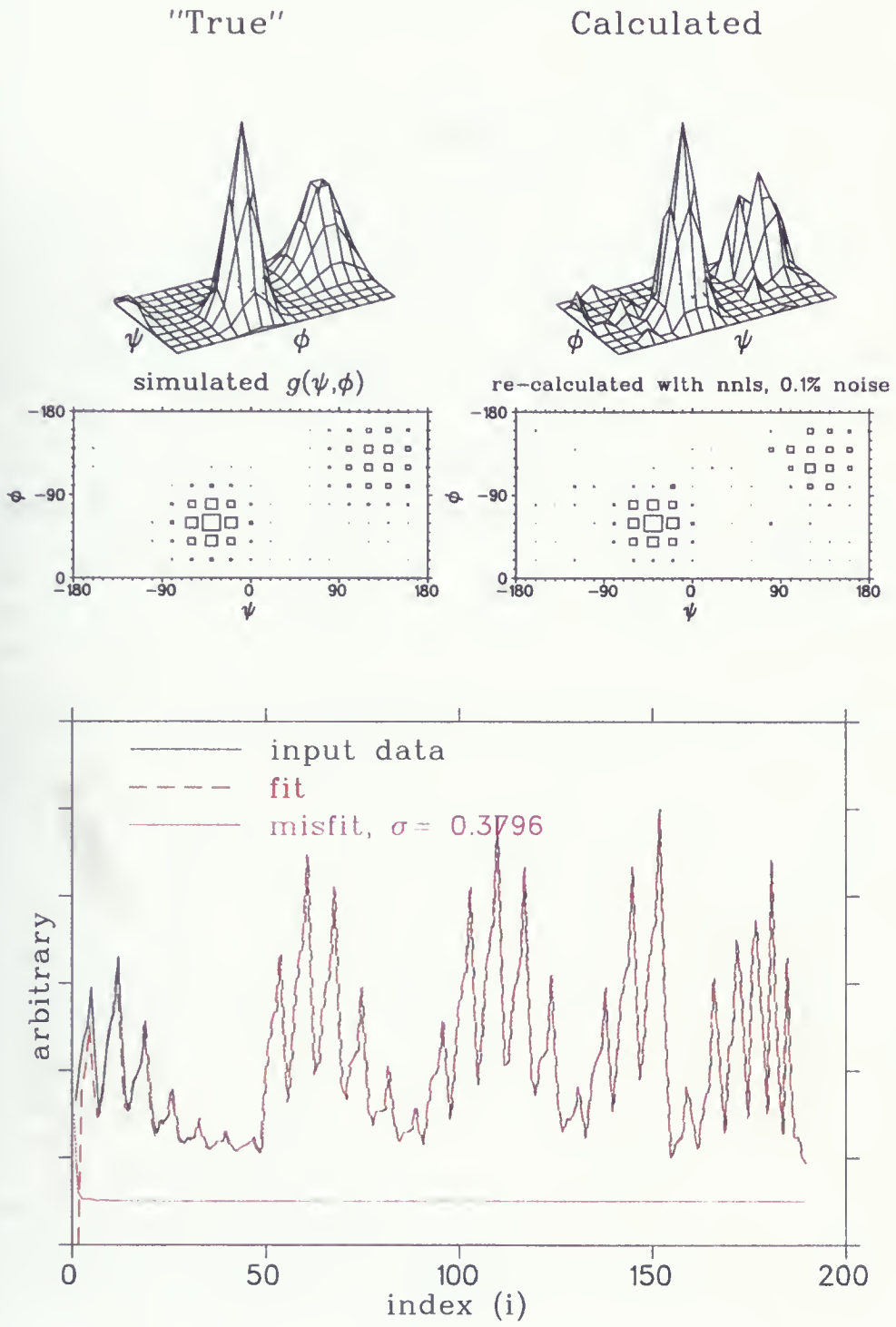


Figure 4.3: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 20°$, using NNLS. **Top**: 0.1% noise level (fit-20-0.1.nnls); **bottom**: misfit for the above (misfit-20-0.1.nnls).

4.2 Tikhonov regularization using the unity operator (Tr1)

As mentioned in NNLS approach, the solution of Fredholm integral equations of the first kind, $\mathbf{g}^{\sigma}$, is obtained by minimizing the misfit function. This ensures us that the highest compatibility to the data is maintained, although the result may not have enough stability. Lack of stability can be overcome by imposing an additional constraint on the solution. Tikhonov regularization approach restrict the norm of the quantity $\mathbf{L}\,\mathbf{g}$ to be minimum, where $\mathbf{L}$ is an operator for which the unity or the second derivative are frequently used. For our two-dimensional case, we examine the regularization effects of the unity and the Laplacian operators. If $\mathbf{L}$ is the unity operator, one asks for the minimum Euclidean norm of the vector $\mathbf{g}$ in the minimization problem:

$$\| \mathbf{s}'^{\sigma} - \mathbf{K}'\,\mathbf{g} \|^2 + \lambda \| \mathbf{g} \|^2 \implies \min. \tag{4.5}$$

Optimum values for the regularization parameter λ are estimated using the self-consistent (SC) method. A typical value of λ is, for example, 4.2951e-01 for 5° resolution using 0.1% noise. As the grid resolution decreases smaller values of λ are obtained at a particular noise level. On the other hand, at a particular grid resolution larger values for the regularization parameter are obtained for higher noise levels.

For $\Delta\phi = \Delta\psi = 5°$, the results and the corresponding misfit functions are presented in Figs. 4.4, 4.5 and 4.6 using 0.1%, 0.5% and 1% noise levels respectively. By comparing Figs. 4.4 (top) and 4.1 (top) one can see that the structure of the true function $g(\phi, \psi)$ is recovered much better by applying Tikhonov regularization technique to the standard least-squares problem. Unlike the misfit function obtained by the standard NNLS-method, Fig. 4.1 (bottom), the misfit function obtained by Tikhonov regularization technique, Fig. 4.4 (bottom), is random and no systematic features appear. The standard deviation of the misfit for the Tr1-method, $\sigma = 0.4682e - 02$, is much smaller than the one for NNLS-method, $\sigma = 5.963$. This means that for high signal-to-noise the compatibility between the recalculated data and the simulated data is much better in Tikhonov regularization method. On the other hand, Figs. 4.5 (top) and 4.6 (top) show that for lower signal-to-noise some of the features of the true distribution function $g(\phi, \psi)$ can not be recovered. At the noise levels above about 0.5% systematic features show up and the stability of the result, $\mathbf{g}_{\lambda}^{\sigma}$, is not satisfactory.

Figs. 4.7 and 4.8 show the results for $\Delta\phi = \Delta\psi = 10°$ using different noise levels. Figs. 4.9 and 4.10 contain the results for 15°-resolution, while Figs. 4.11 and 4.12 use 20° resolution. By going to lower grid resolutions the problem is less underdetermined and as one can expect the quality of the fit should be better. For example, for $\Delta\phi = \Delta\psi = 20°$ the standard deviation of the misfit, $\sigma = 0.3025e - 03$, is noticeably smaller than the standard deviation of the misfit, $\sigma = 0.4682e - 02$ for $\Delta\phi = \Delta\psi = 5°$. Also, Fig. 4.11 (top) shows that for a less underdetermined problem the shape of the true function g is retrieved better. Notice that all of these aspects appear at a high signal-to-noise ratio. Regardless of how less underdetermined the problem is, for a low signal-to-noise ratio systematic features appear and as can be seen in Figs. 4.8, 4.10 and 4.12 (bottom) the results do not contain enough stability to recover the exact shape of the true distribution function.

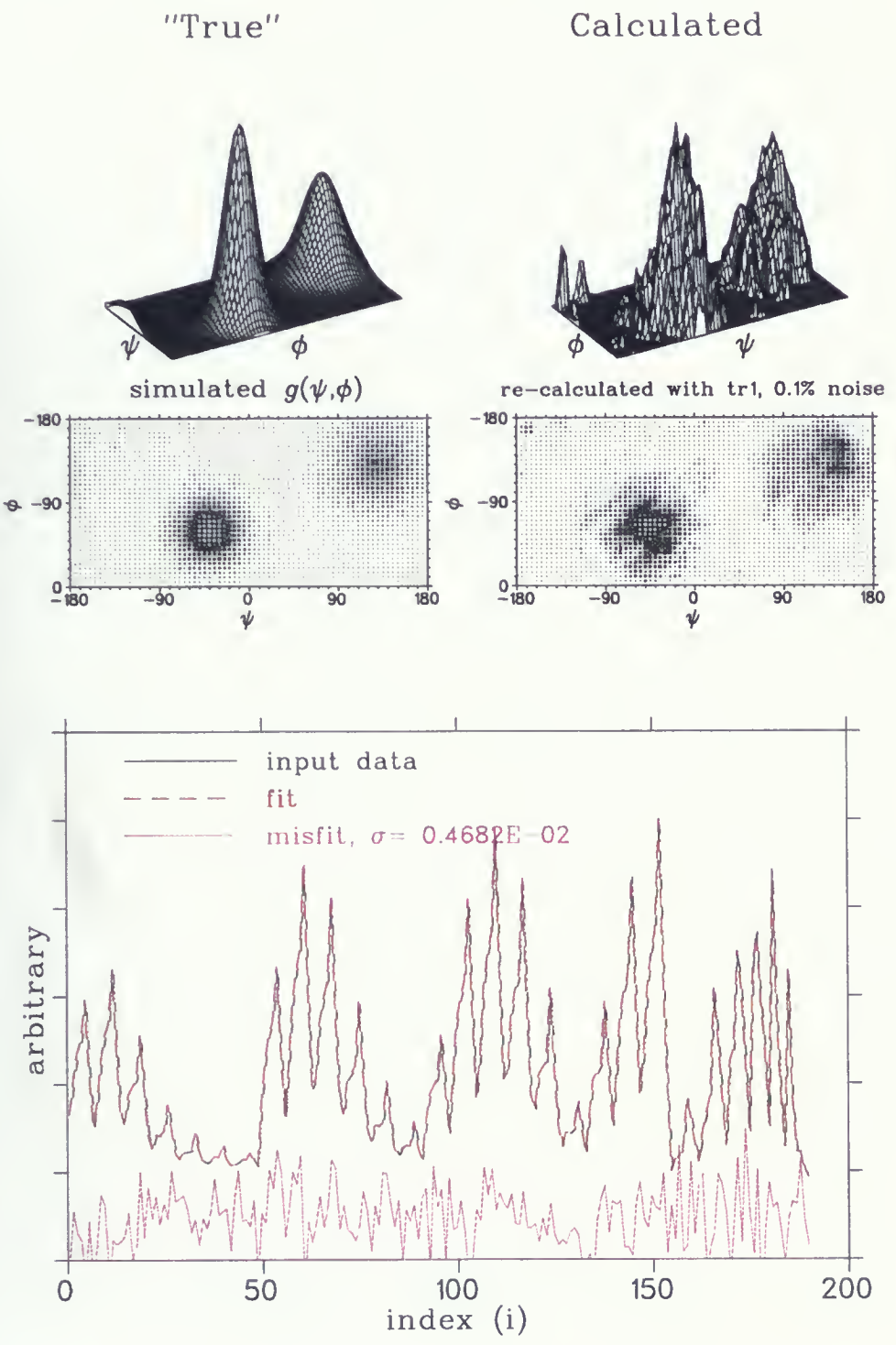


Figure 4.4: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the identity. **Top**: 0.1% noise level (fit-5-0.1.tr1); **bottom**: misfit for the above (misfit-5-0.1.tr1).

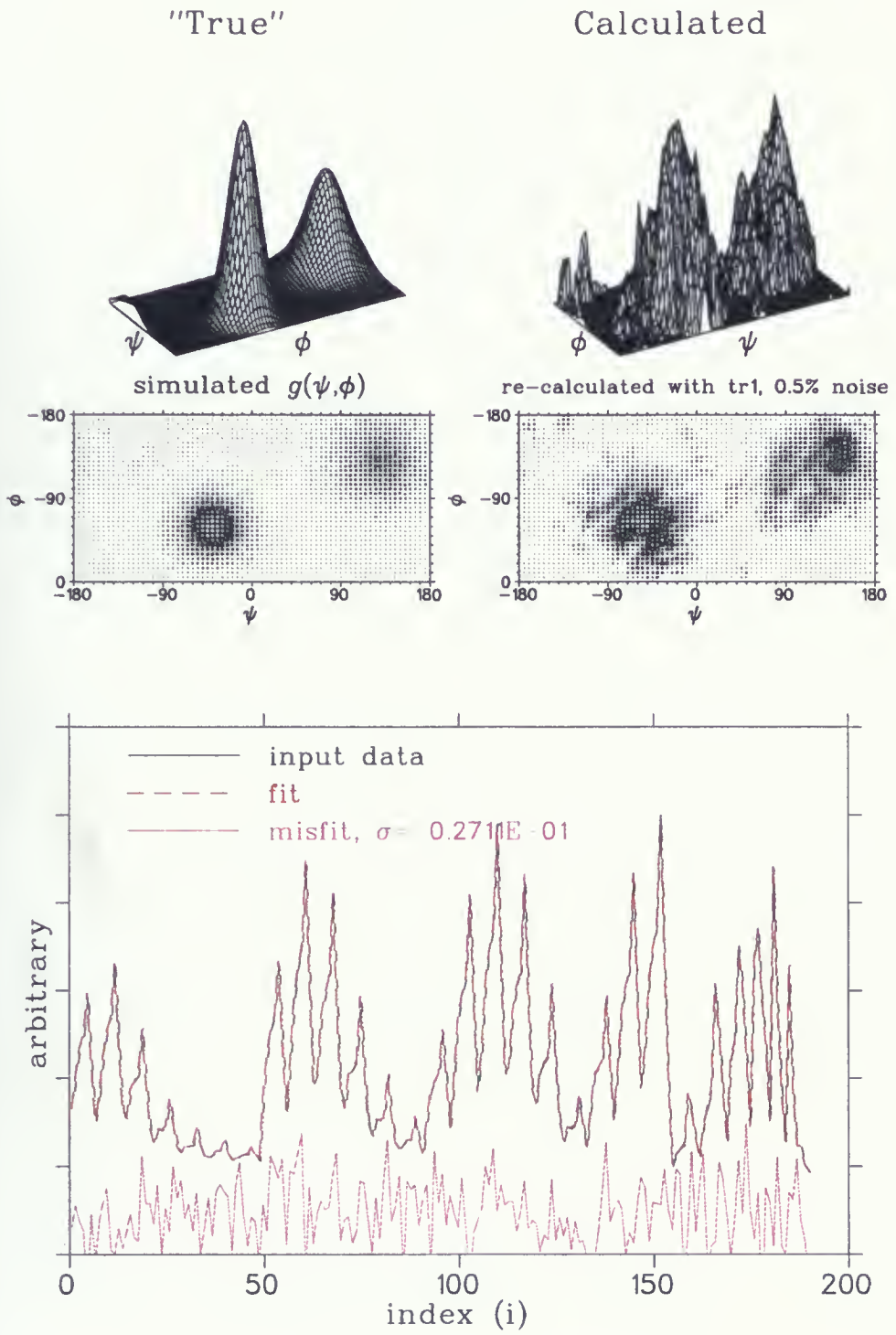


Figure 4.5: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the identity. **Top**: 0.5% noise level (fit-5-0.5.tr1); **bottom**: misfit for the above (misfit-5-0.5.tr1).

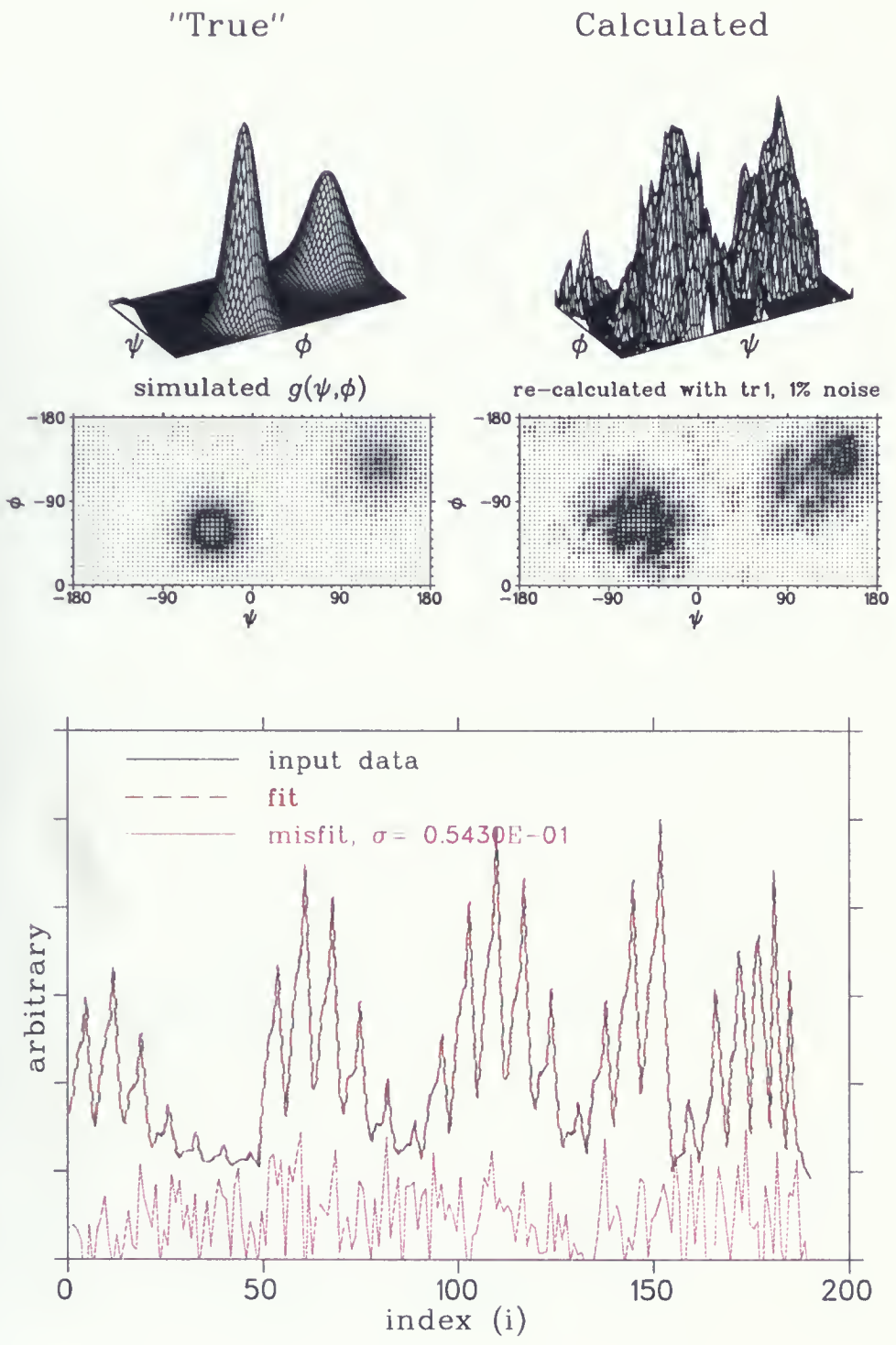


Figure 4.6: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the identity. **Top**: 1% noise level (fit-5-1.tr1); **bottom**: misfit for the above (misfit-5-1.tr1).

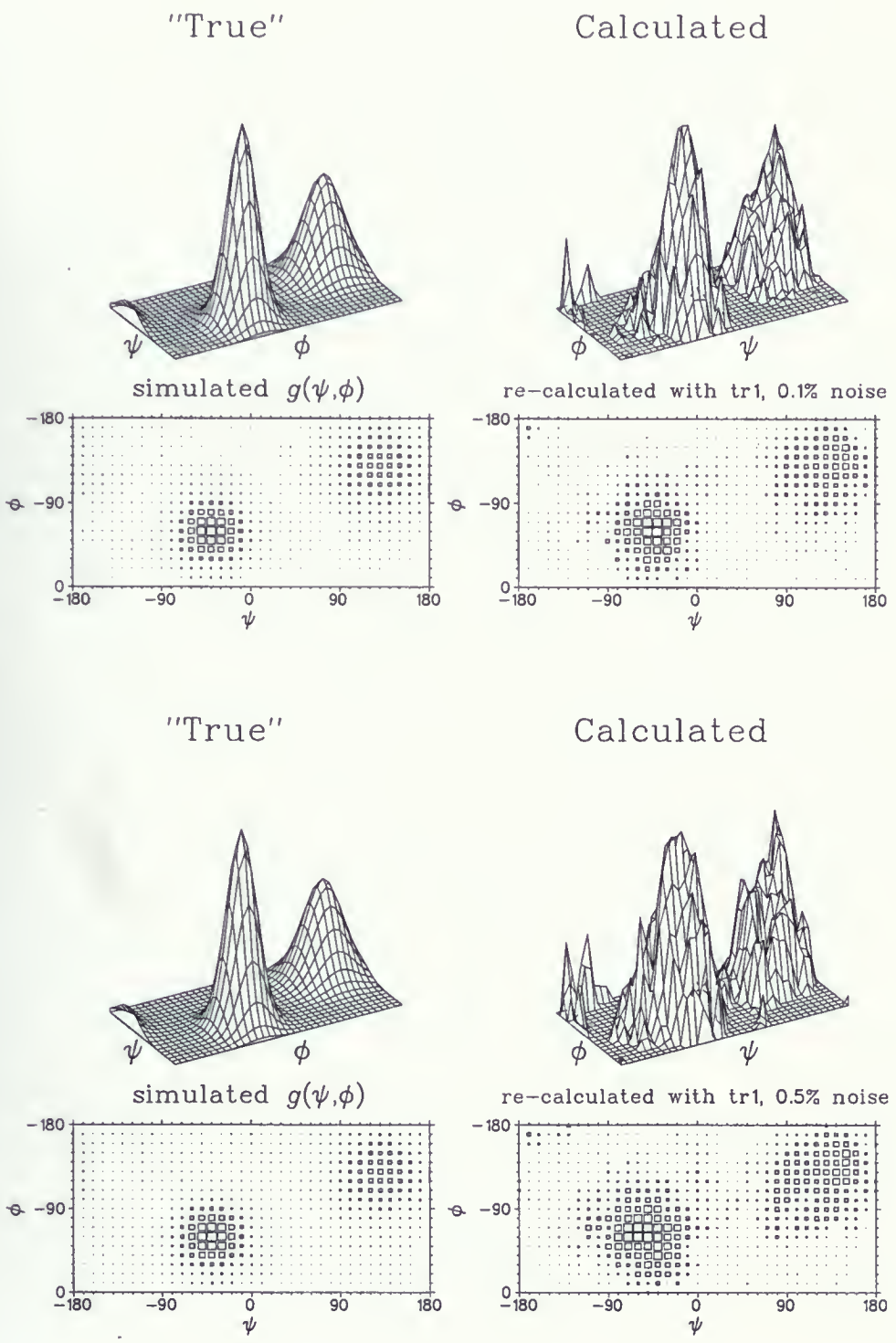


Figure 4.7: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 10°$, Tikhonov regularization using the identity. **Top**: 0.1% noise level (fit-10-0.1.tr1); **bottom**: 0.5% noise level (fit-10-0.5.tr1).

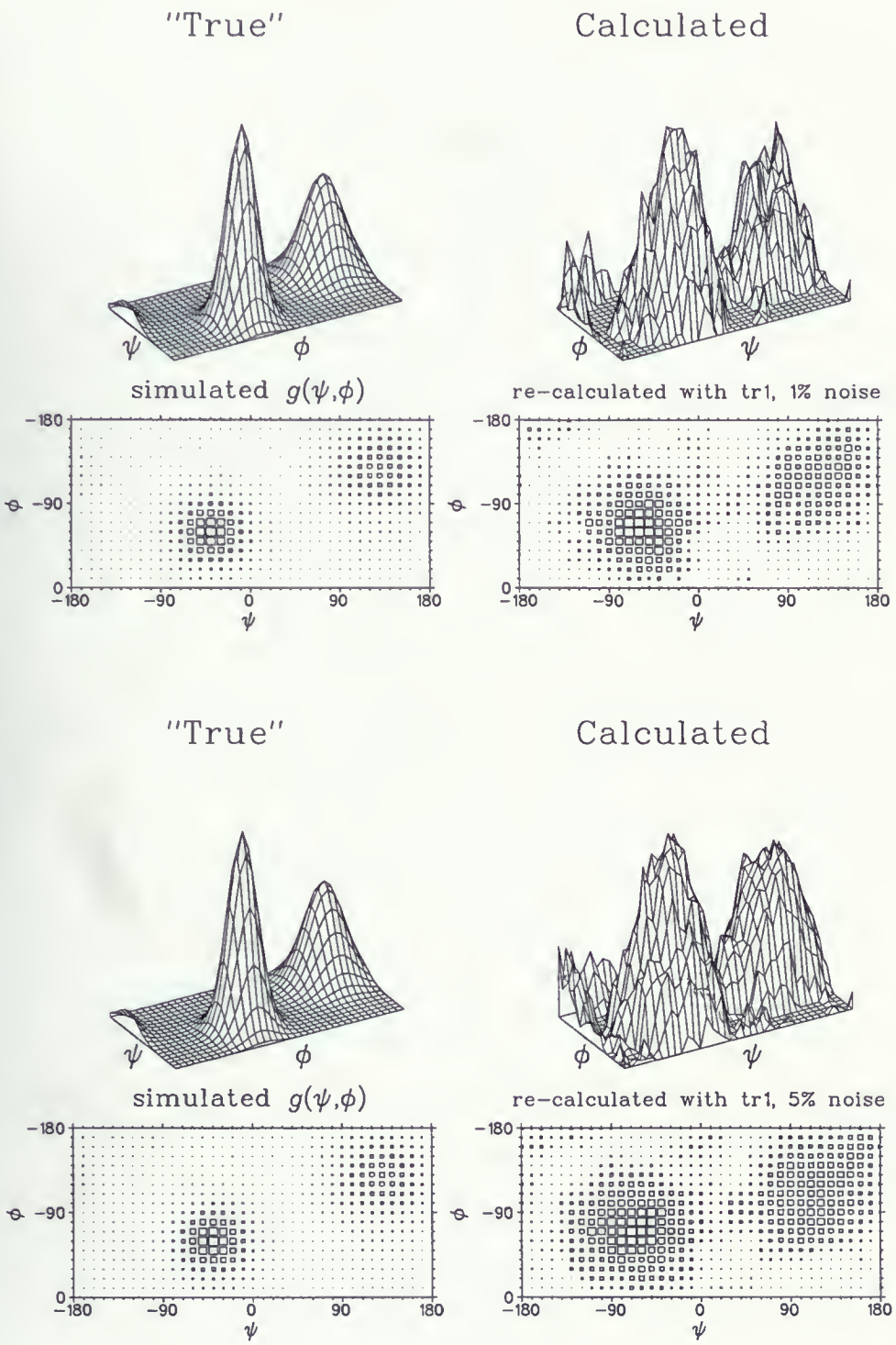


Figure 4.8: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 10°$, Tikhonov regularization using the identity. **Top**: 1% noise level (fit-10-1.tr1); **bottom**: 5% noise level (fit-10-5.tr1).

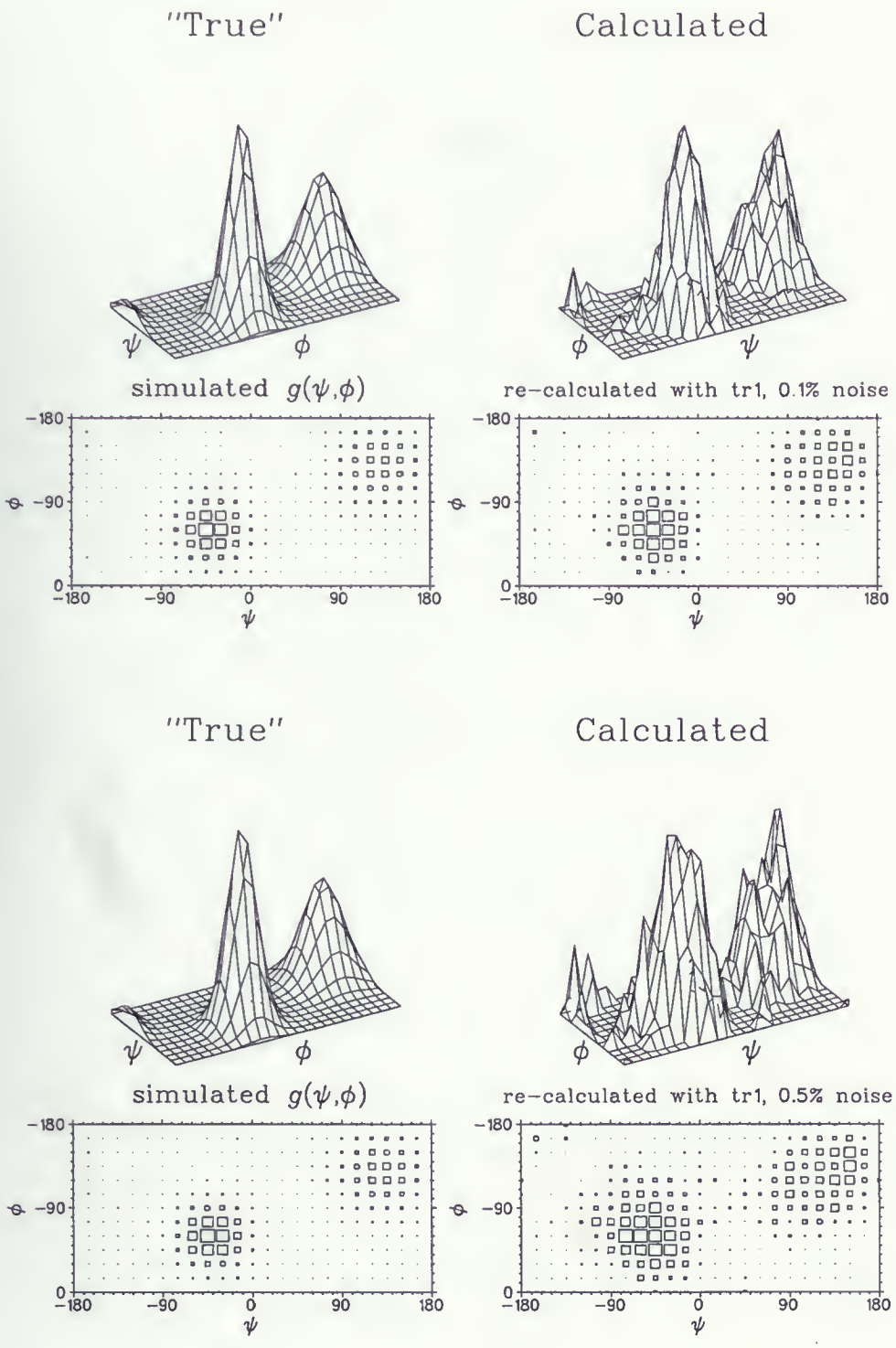


Figure 4.9: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 15°$, Tikhonov regularization using the identity. **Top:** 0.1% noise level (fit-15-0.1.tr1); **bottom:** 0.5% noise level (fit-15-0.5.tr1).

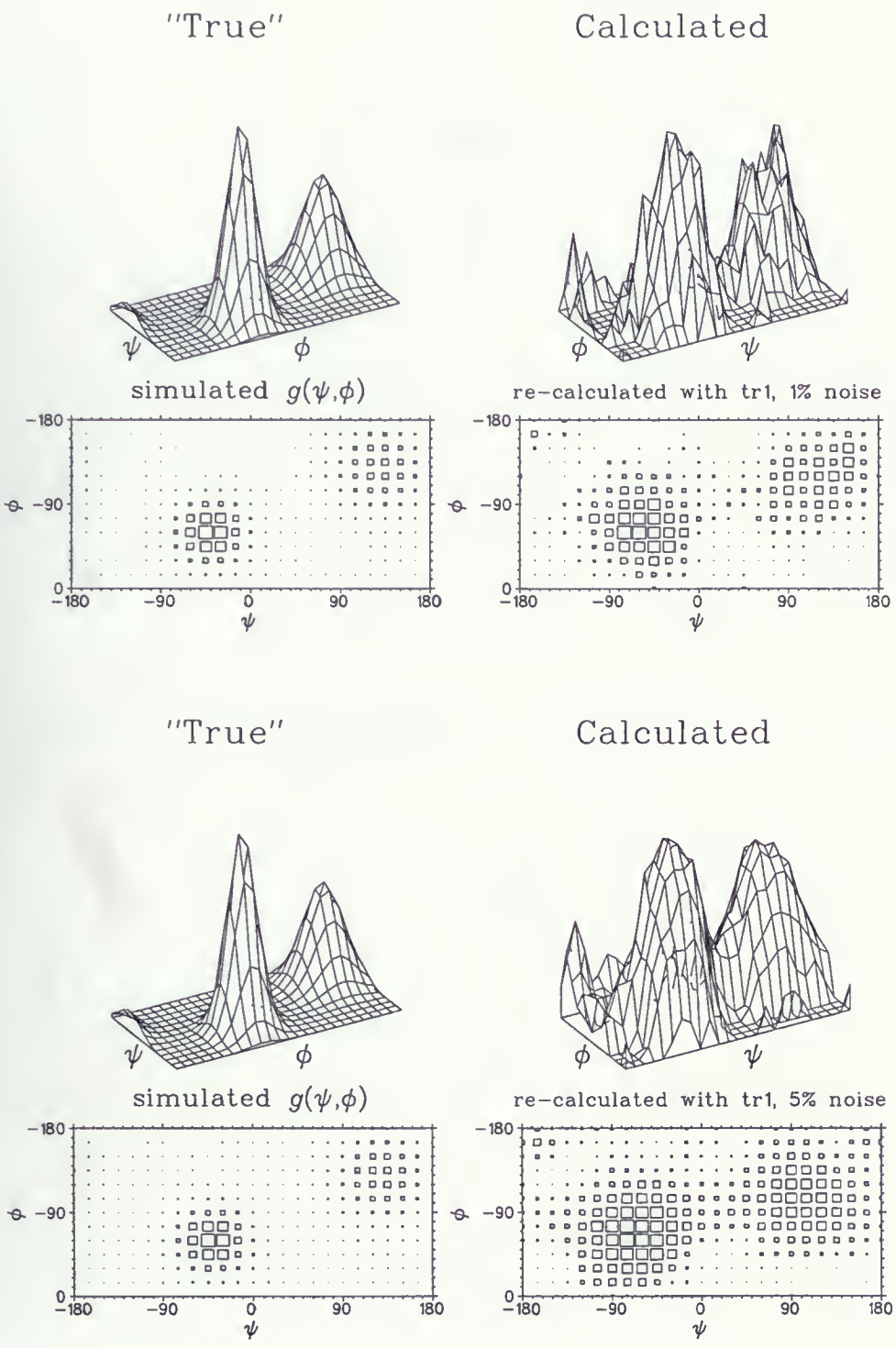


Figure 4.10: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 15°$, Tikhonov regularization using the identity. **Top**: 1% noise level (fit-15-1.tr1); **bottom**: 5% noise level (fit-15-5.tr1).

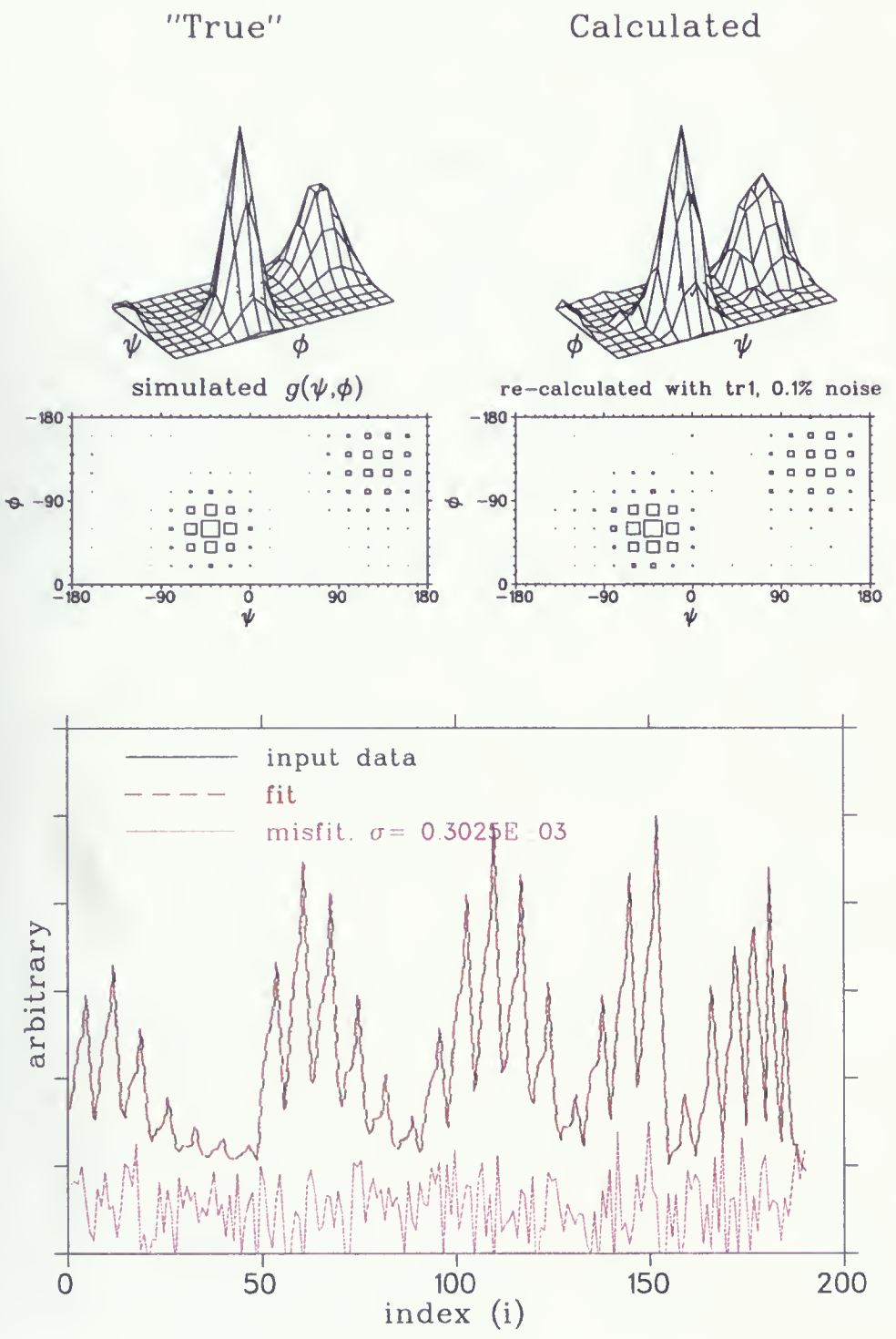


Figure 4.11: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 20°$, Tikhonov regularization using the identity. **Top**: 0.1% noise level (fit-20-0.1.tr1); **bottom**: misfit for the above (misfit-20-0.1.tr1).

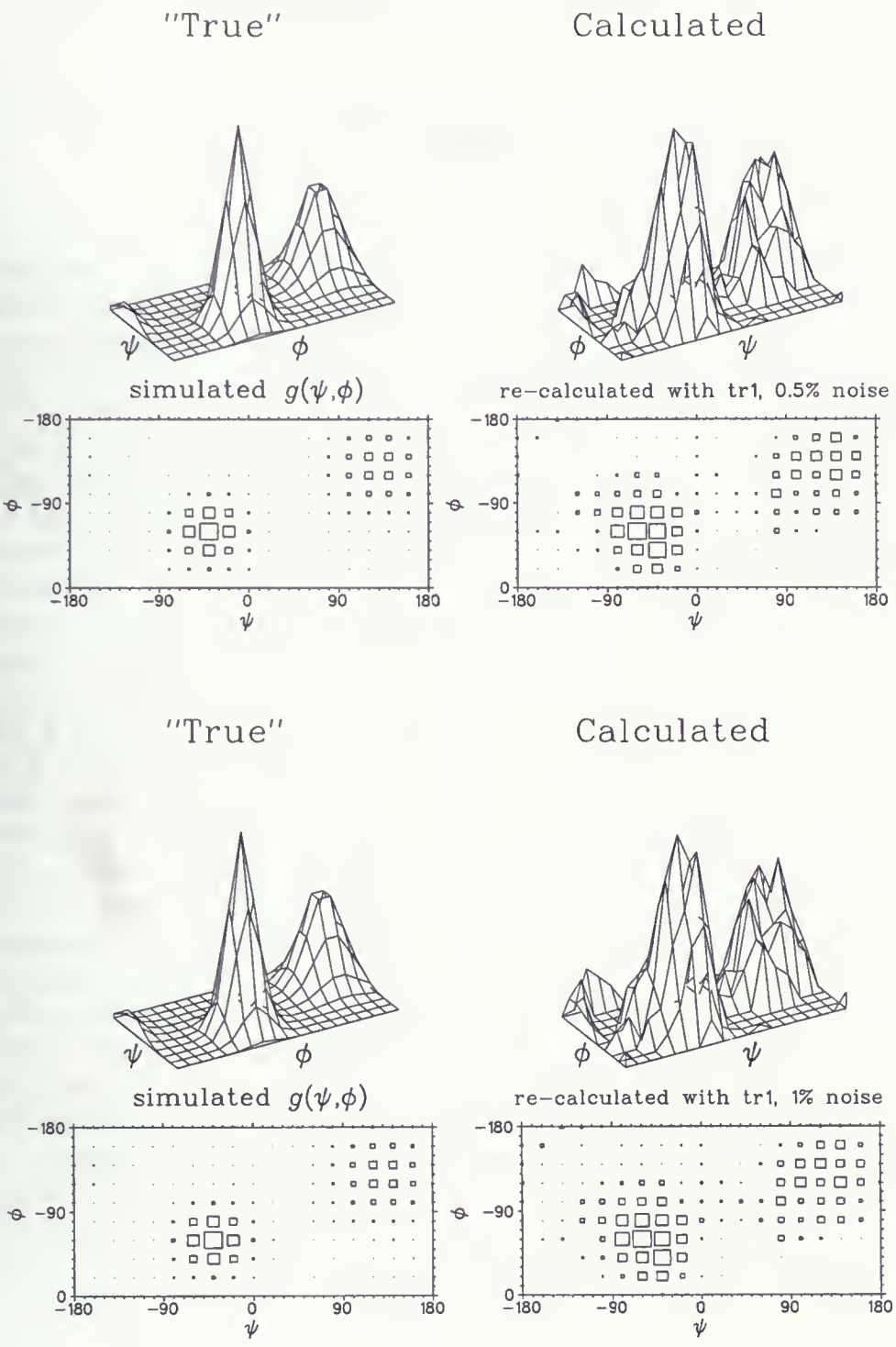


Figure 4.12: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 20°$, Tikhonov regularization using the identity. **Top**: 0.5% noise level (fit-20-0.5.tr1); **bottom**: 1% noise level (fit-20-1.tr1).

4.3 Tikhonov regularization using the Laplacian operator (Tr2)

The operator **L** mentioned in the previous section is now chosen to be the Laplacian in spherical coordinates system and the result, $\mathbf{g}_\lambda^\sigma$, is the solution of the following problem:

$$\| \mathbf{s}'^{\sigma} - \mathbf{K}'\,\mathbf{g} \|^2 + \lambda \| \nabla^2\,\mathbf{g} \|^2 \implies \min. \tag{4.6}$$

As we should recall, the torsion angle ϕ ranges from $-180°$ to $0°$, inclusive, and the torsion angle ψ ranges from $-180°$ to $+180°$, inclusive. The whole range of the ϕ-angle can not be covered through the calculations based on the Laplacian, because in implementing the Laplacian in spherical coordinates system, as described in chapter 3, the poles must be excluded. Therefore the structure of the true function $g(\phi,\psi)$ can not be recovered for the regions $(\phi,\psi) = (-180°,\psi)$ and $(\phi,\psi) = (0°,\psi)$.

Again, the SC-method is used to determine optimum values for the regularization parameter λ. As an example, For $5°$ resolution using 0.1% noise the value obtained for λ is 2.0972e-04 which is much smaller than the value, 4.2951e-01, obtained by the Tr1-method. The results and the misfit functions for $5°$ are given in Figs. 4.13, 4.14 and 4.15. Like Tr1-method, the misfit functions obtained via the Tr2-method are random and no systematic features appear. Also, for a particular noise level, the standard deviation of the misfit for the Tr2-method, *e.g.* $\sigma = 0.4214e-02$ for 0.1% noise, is slightly less than the one for the Tr1-method, *e.g.* $\sigma = 0.4682e-02$ for 0.1% noise. This means that in terms of the compatibility to the data both methods are highly effective.

In terms of the recovering the shape of the true distribution function $g(\phi,\psi)$, the Tr2-method is much more successful. By comparing Figs. 4.5 (top) and 4.14 (top), one can see that the 2D-structure of the true function g is missing in Tr1-method, while it is perfectly recovered in Tr2-method. This is because the Tr2-method introduces a coupling between the two dimensions in the parameter space, while in Tr1-method the two-dimensional (ϕ,ψ) space is unwrapped into a long one-dimensional space.

Several other results for $10°$ resolution are given in Figs. 4.16 and 4.17. Figs 4.15 (top) and 4.17 (top) show that the Tr2-method, compared to the Tr1-method, is able to recover the shape of the true function g and to provide a high stability for the result even for a low signal-to-noise ratio. For a very low signal-to-noise ratio, as can be seen in Fig 4.17 (bottom), the result is oversmoothed and the center of the peaks are shifted, although there is no systematic features showing up and the shape of the true function g is reasonably recovered. The results for $15°$ resolution are presented in Figs. 4.18 and 4.19.

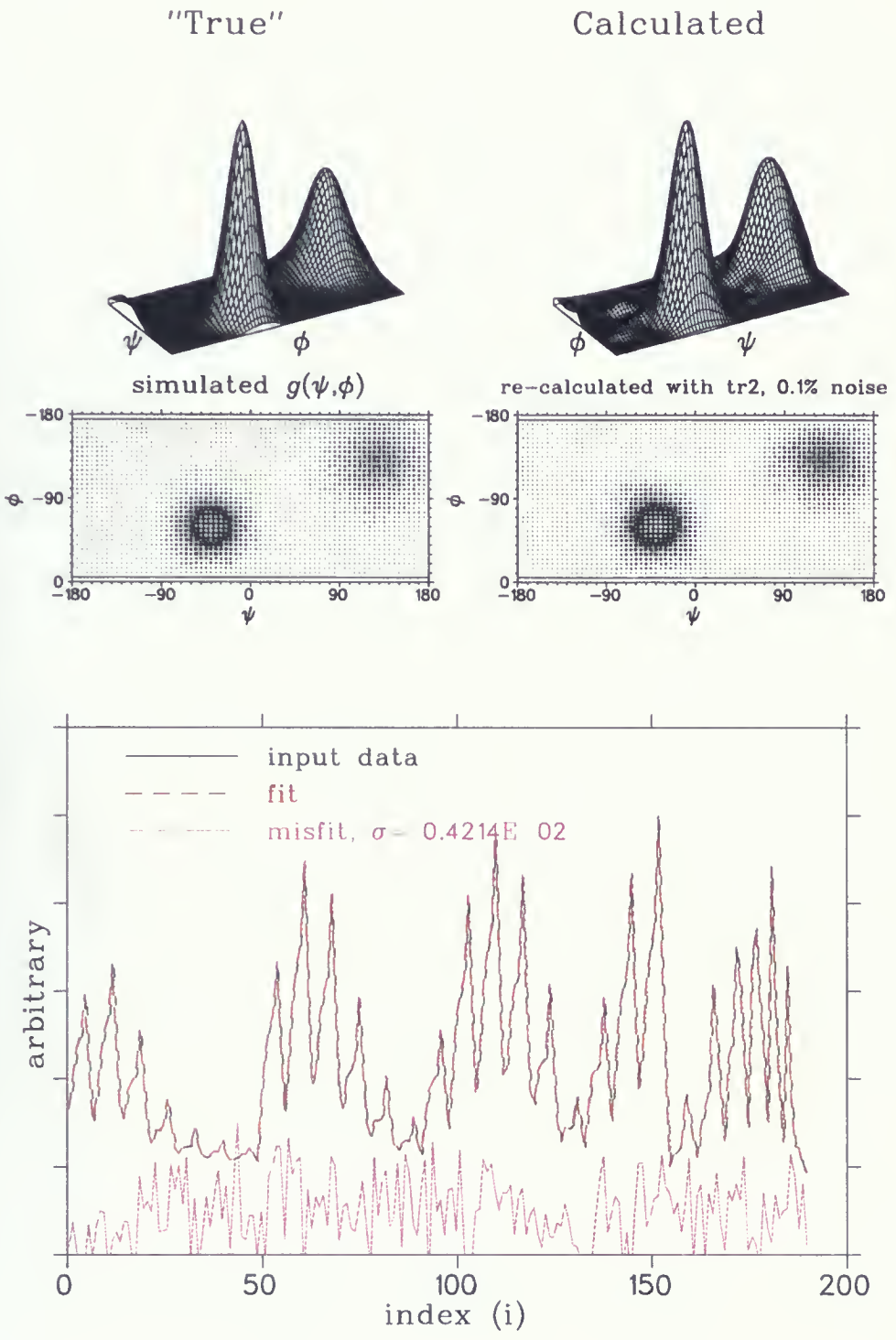


Figure 4.13: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the Laplacian. **Top**: 0.1% noise level (fit-5-0.1.tr2); **bottom**: misfit for the above (misfit-5-0.1.tr2).

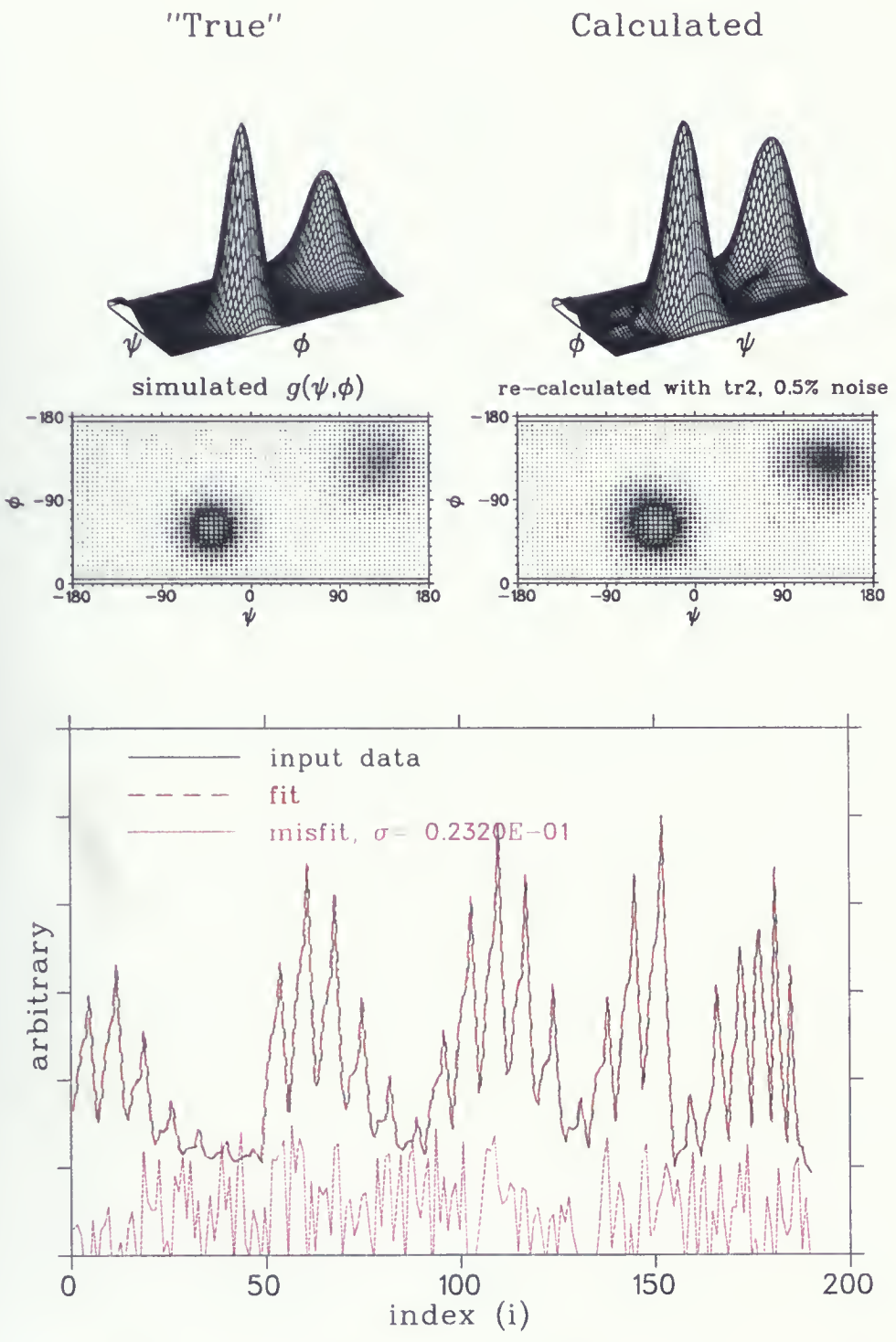


Figure 4.14: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the Laplacian. **Top**: 0.5% noise level (fit-5-0.5.tr2); **bottom**: misfit for the above (misfit-5-0.5.tr2).

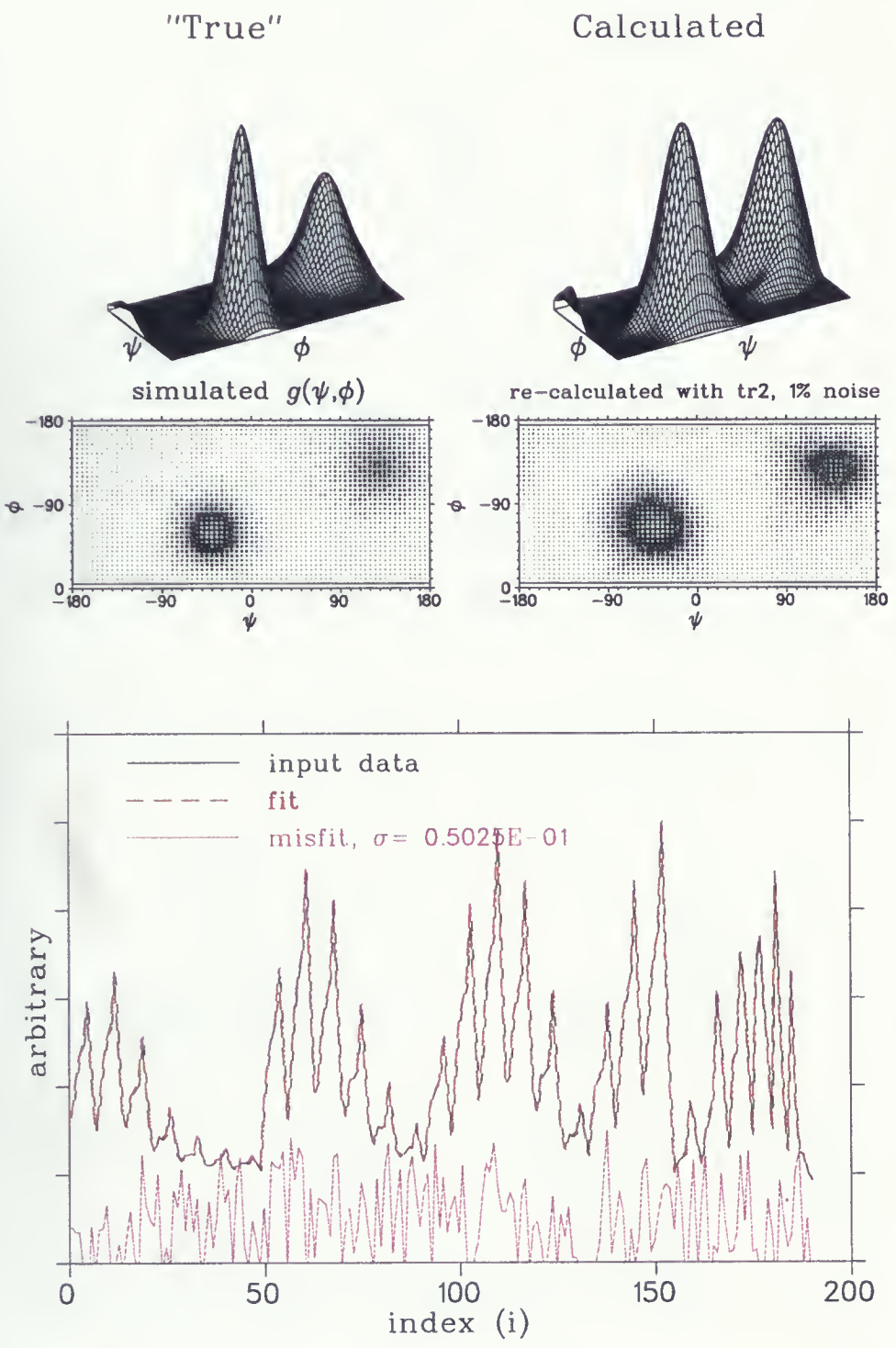


Figure 4.15: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 5°$, Tikhonov regularization using the Laplacian. **Top**: 1% noise level (fit-5-1.tr2); **bottom**: misfit for the above (misfit-5-1.tr2).

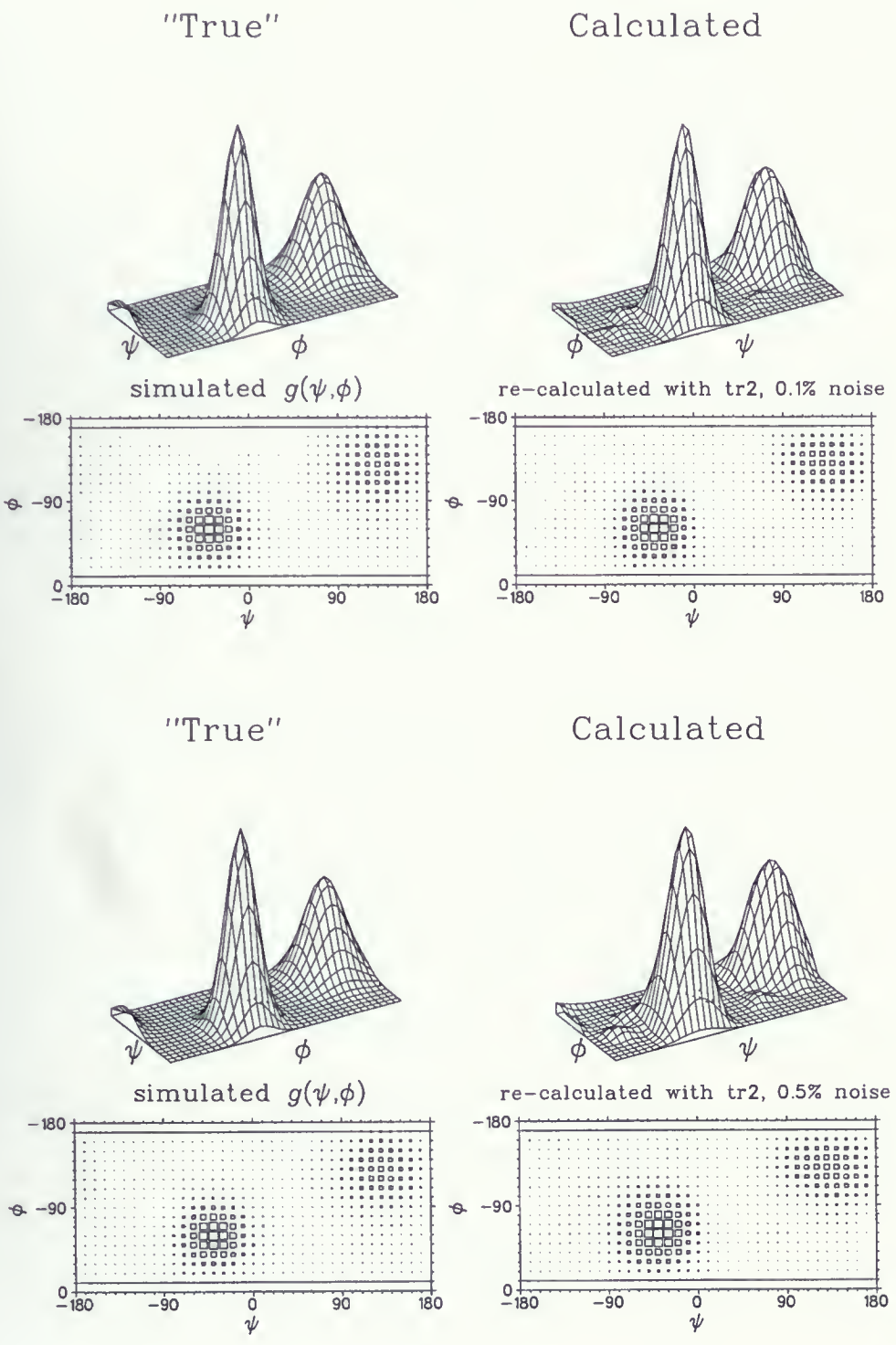


Figure 4.16: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 10°$, Tikhonov regularization using the Laplacian. **Top**: 0.1% noise level (fit-10-0.1.tr2); **bottom**: 0.5% noise level (fit-10-0.5.tr2).

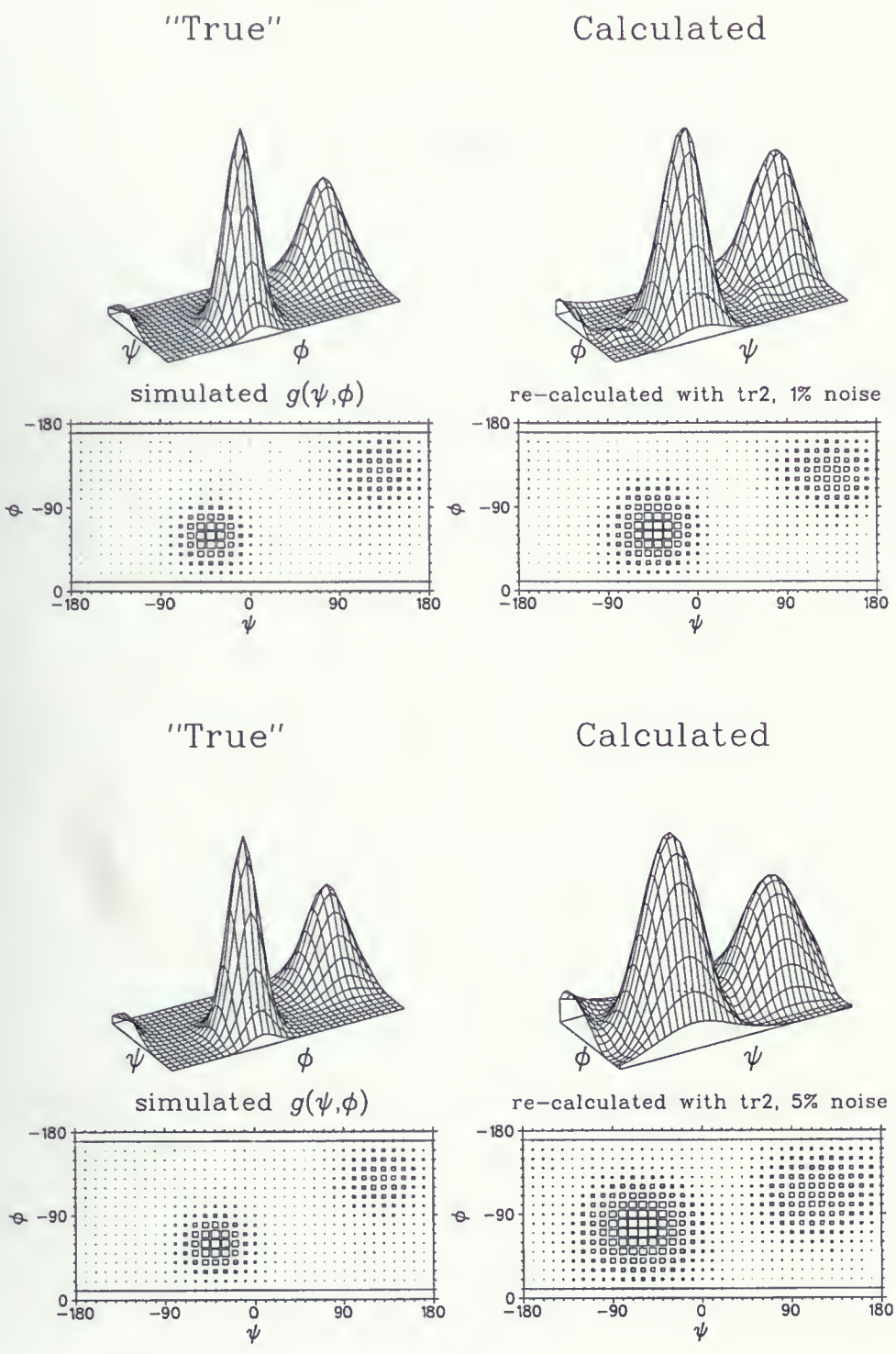


Figure 4.17: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 10°$, Tikhonov regularization using the Laplacian. **Top**: 1% noise level (fit-10-1.tr2); **bottom**: 5% noise level (fit-10-5.tr2).

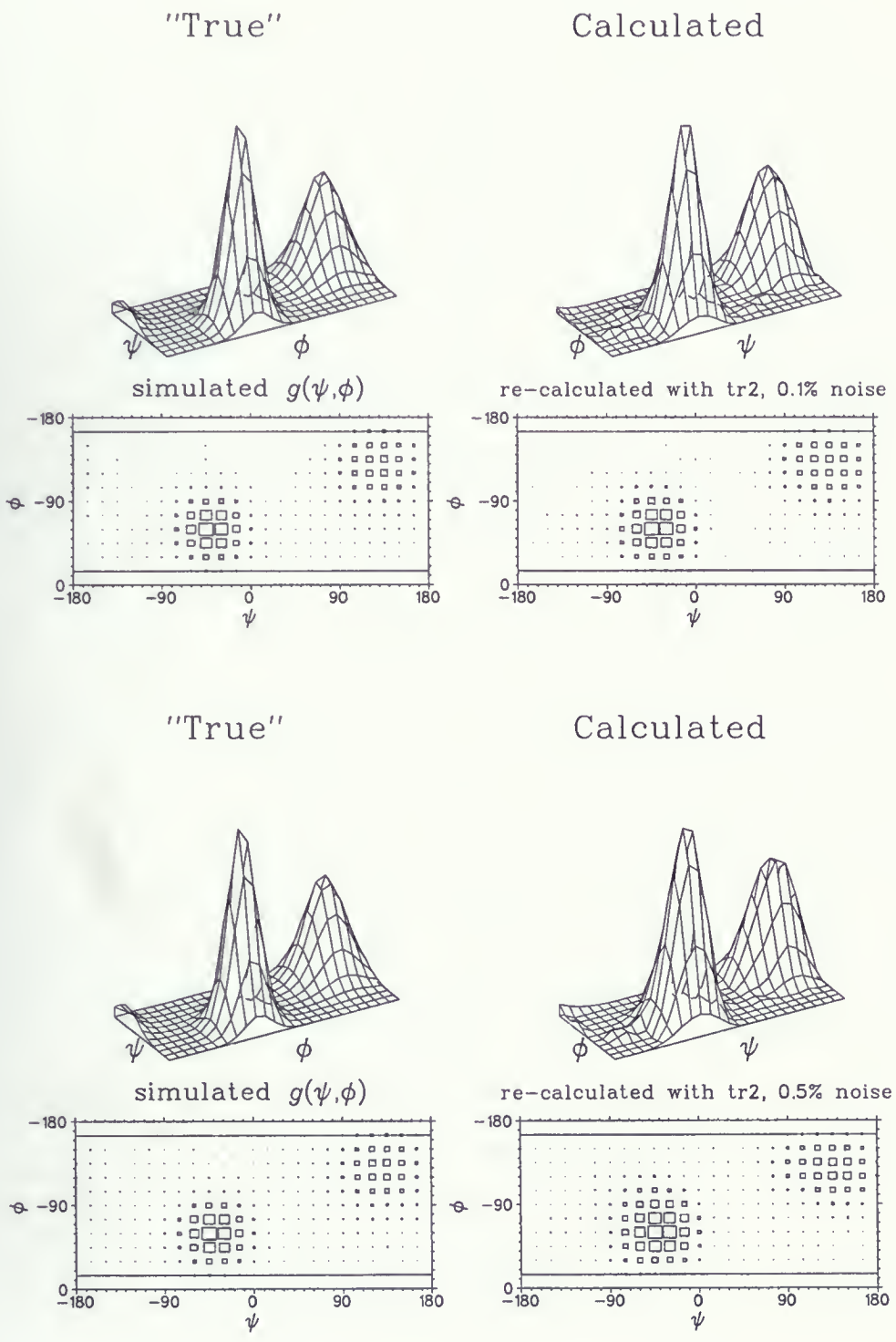


Figure 4.18: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 15°$, Tikhonov regularization using the Laplacian. **Top**: 0.1% noise level (fit-15-0.1.tr2); **bottom**: 0.5% noise level (fit-15-0.5.tr2).

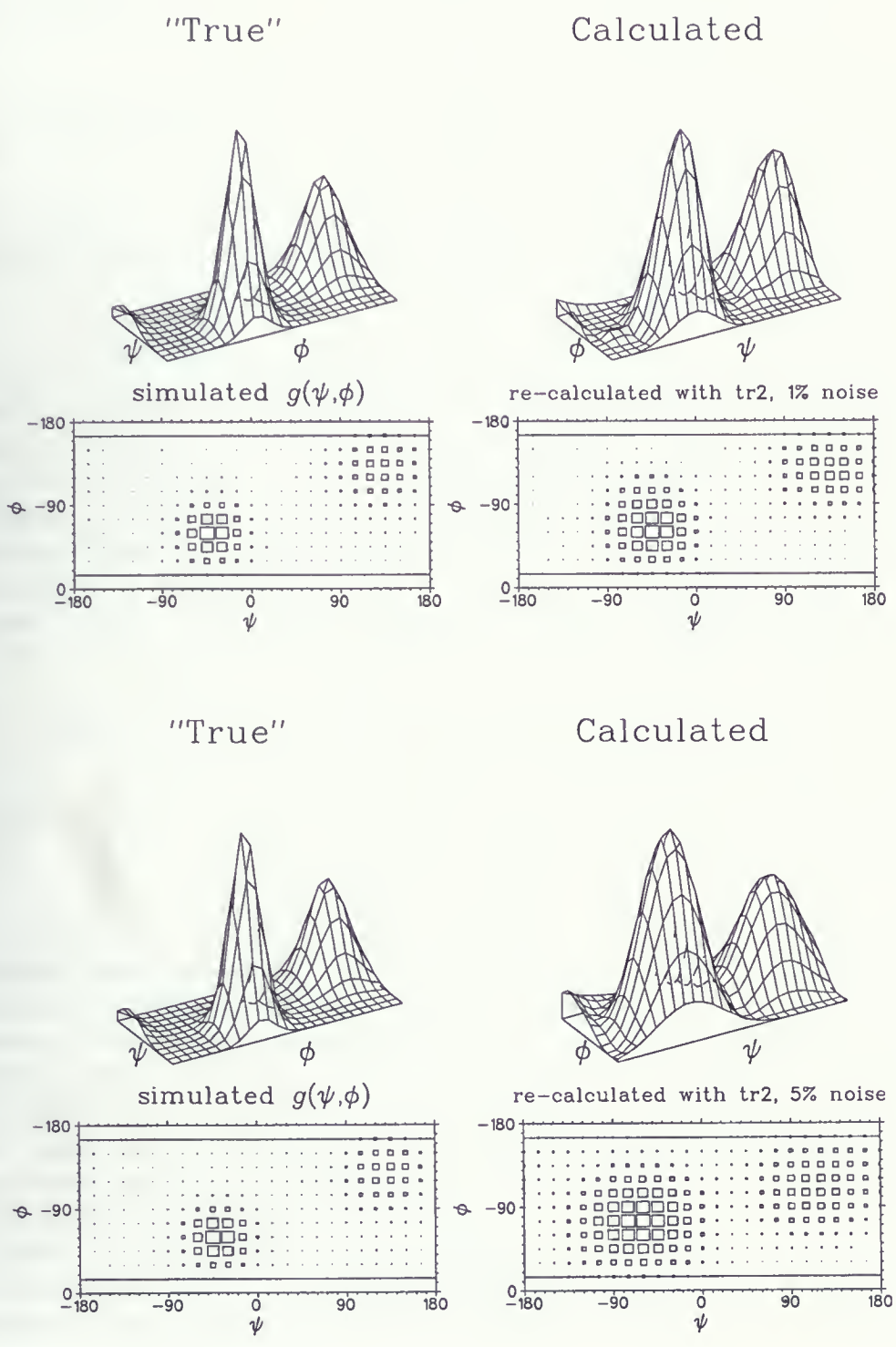


Figure 4.19: Ramachandran plot from simulated CTDQFD & 2D MAS exchange measurements, for $\Delta\phi = \Delta\psi = 15°$, Tikhonov regularization using the Laplacian. **Top:** 1% noise level (fit-15-1.tr2); **bottom:** 5% noise level (fit-15-5.tr2).

Chapter 5

Conclusions

Two-dimensional Fredholm integral equations (FIE) of the first kind

$$s(\omega) = \int_{\psi} \int_{\phi} K(\omega; \phi, \psi)\, g(\phi, \psi)\, d\phi\, d\psi \tag{5.1}$$

can be solved to extract structural information in the form of distribution functions that are mapped into NMR data. This is an ill-posed inverse problem and the stability of the solution is an important issue. Regularization techniques are developed to stabilize the inverse mapping; out of them Tikhonov regularization is known to be the most effective one. It imposes an additional constraint on the solution g in the form of $\| \hat{\mathbf{L}}\, g(\phi, \psi) \|^2$.

We applied inverse theory techniques to extract the local conformations of peptides in the form of distributions of (ϕ, ψ) angle from a set of 2D MAS and CTDQFD data (simulations). A simple non-negative least-squares (NNLS) approach is first taken to obtain the best fit to the data

$$\Psi_{\mathrm{LS}}(g) = \left\| \int \int K(\omega; \phi, \psi) g(\phi, \psi) d\phi d\psi - s(\omega) \right\|^2 \longrightarrow \min . \tag{5.2}$$

Then Tikhonov regularization (TR) is applied to stabilize the approximate g_λ, using $\hat{\mathbf{L}} = \hat{1}$ and $\hat{\mathbf{L}} = \nabla^2$

$$\Psi(g) = \Psi_{\mathrm{LS}}(g) + \lambda \| \hat{\mathbf{L}}\, g \|^2 \longrightarrow \min . \tag{5.3}$$

It is observed, Fig 5.1, that the TR-methods improve the stability of the solution. $\|\nabla^2 g\|^2$ is an excellent regularization functional. Besides maintaining high compatibility with the data, it takes into account the 2D nature of the distribution function $g(\phi, \psi)$, by making use of the inherently two-dimensional nature of the underlying Ramachandran maps. In Fig 5.2 a comparison between the misfits is shown for NNLS and TR-methods, using different grid densities. Notice that TR methods are greatly insensitive to grid density, even at high grid density the misfit has no systematic features and remains small. TR is thus suitable for greatly underdetermined problems such as the CTDQFD experiments [49] where few data points (typically 10 points) are available.

The signal-to-noise ratio (SNR) is found to play a crucial rule in the quality of the best-fit. For signal-to-noise ratio above 100:1 the results show excellent convergence to the true angle distribution function $g(\phi, \psi)$. While, for SNR below 50:1 only a qualitative agreement is seen.

It is observed that the TR-method, using $\|\nabla^2 g\|^2$, is not sensitive to noise level, Fig 5.3. Even at low SNR, despite the broadening, the relative population of α-helix and β-sheet conformations is preserved (within 11% error), suitable for real data. Also, the ratio of the populations of α-helix to β-sheet conformations is greatly insensitive to grid density (close to true value 1.5, within 7% error), suitable when very few data points are available.

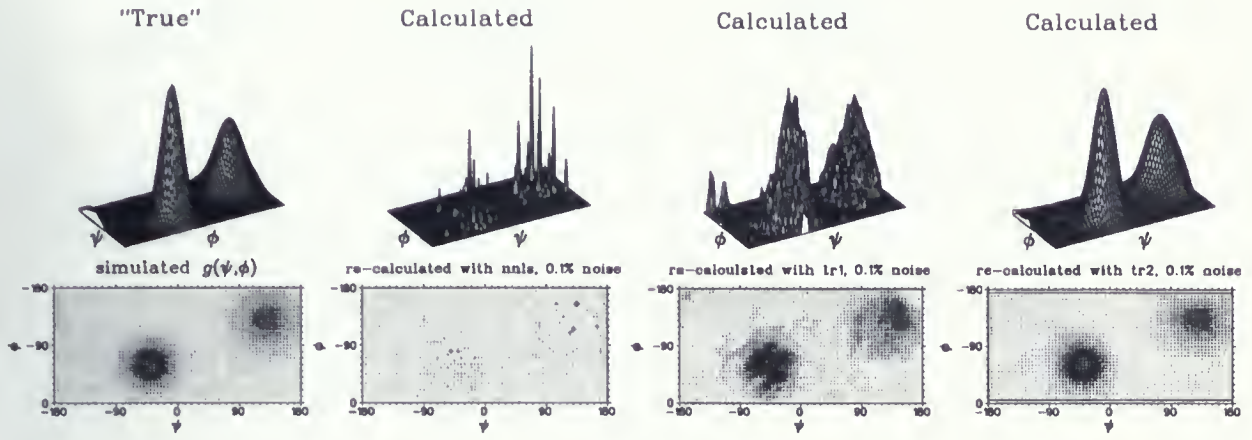
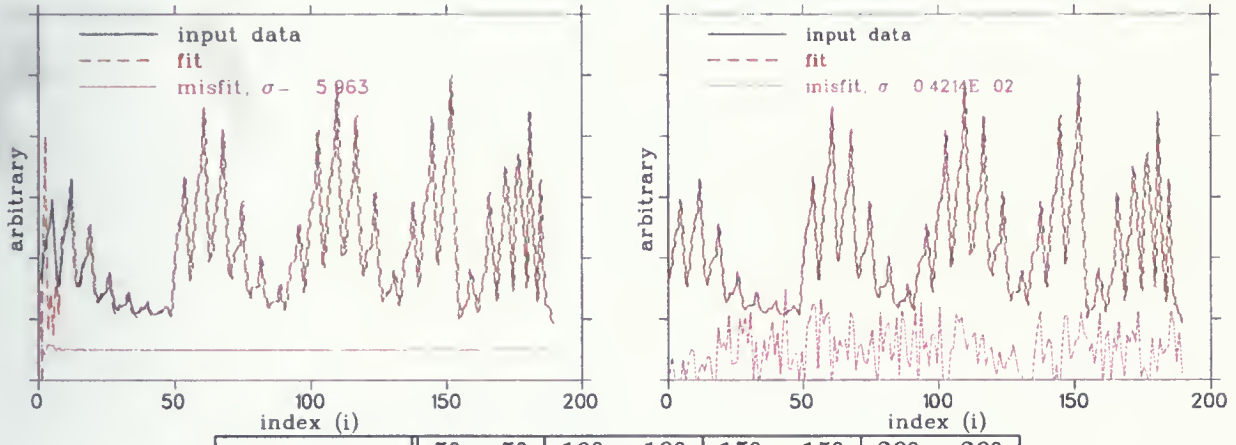
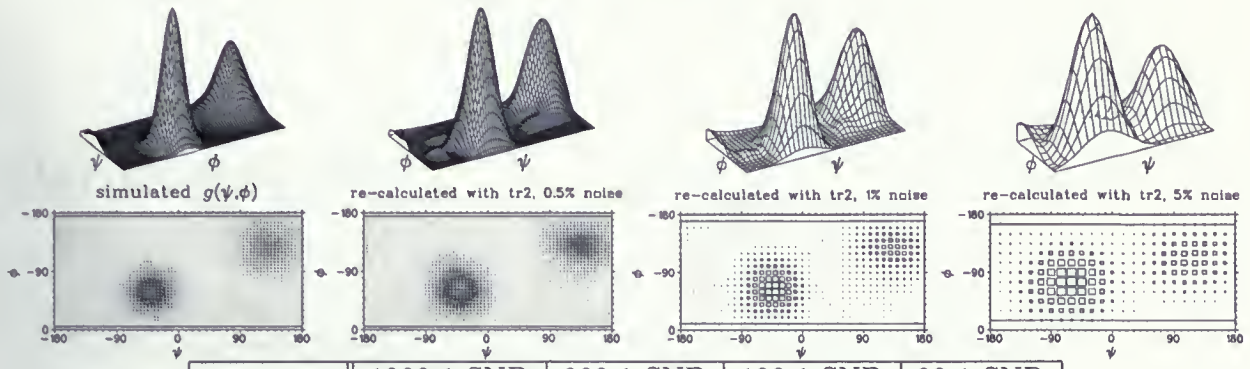


Figure 5.1: While NNLS favours discontinuous solutions, the additional constraints introduced by the TR term help recover the original continuous shape of the distribution function.



	$5° \times 5°$	$10° \times 10°$	$15° \times 15°$	$20° \times 20°$
NNLS	5.96	1.50	0.67	0.380
TR, $\|g\|^2$	4.68e-3	1.24e-3	0.55e-3	0.303e-3
TR, $\|\nabla^2 g\|^2$	4.21e-3	1.09e-3	0.5e-3	—

Figure 5.2: Misfit at various grid densities: TR methods remain robust even when very few data points are available. All data for 1000:1 SNR.



	1000:1 SNR	200:1 SNR	100:1 SNR	20:1 SNR
$5° \times 5°$	1.37	1.34	1.34	—
$10° \times 10°$	1.40	1.50	1.45	1.55
$15° \times 15°$	1.37	1.51	1.54	1.63

Figure 5.3: Even at low SNR the relative population of α-helix and β-sheet conformations is well preserved — suitable for real data.

The aforementioned powers of the new regularization functional $\|\nabla^2 g\|^2$ suggest that even for low SNR, TR framework is well-suited to incorporating additional physical constraints into the minimization problem. *E.g.* if it is known that only α-helix and β-sheet conformations are present, one can restrict the range of (ϕ, ψ) pairs close to $(-57°, -47°)$ and $(-119°, 113°)$. This is left as future work when applications to experimental data, rather than simulations, will be studied.

Bibliography

[1] E. Sternin, M. Bloom, and A. L. MacKay. De-Pake-ing of NMR spectra. *J. Magn. Reson.*, 55:274–282, 1983.

[2] K. Whittall, E. Sternin, M. Bloom, and A.L. MacKay. Time– and frequency–domain "dePakeing" using inverse theory. *J. Mag. Res.*, 84:64–71, 1989.

[3] H. Schäfer, B. Mädler, and F. Volke. De-pake-ing of NMR powder spectra by non-negative least squares analysis with Tikhonov regularization. *J. Magn. Reson. A*, 116:145–149, 1995.

[4] H. Schäfer, B. Mädler, and E. Sternin. Determination of orientational order parameters from ^{2}H NMR spectra of magnetically partially oriented lipid bilayers. *Biophys. J.*, 74:1007–1014, 1998.

[5] H. Schäfer and R. Stannarius. Calculation of orientational distributions of partially ordered samples from NMR spectra. *J. Magn. Reson. B*, 106:14–23, 1995.

[6] Y.Q. Song, L. Venkataramanan, M.D. Hürlimann, M. Flaum, P. Frulla, and C. Straley. *J. Magn. Res.*, 154:261–268, 2002.

[7] M. Utz. *J. Chem. Phys.*, 109:6110–6124, 1998.

[8] F. Angeli, T. Charpentier, P. Faucon, and J.C. Petit. *J. Phys. Chem.*, B 103:10356–10364, 1999.

[9] T.E. Creighton. *Proteins: structures and molecular properties.* W.H. Freeman and Co., New York, 1993.

[10] A.T. Petkova, Y. Ishii, J.J. Balbach, O.N. Antzutkin, R.D. Leapman, F. Delaglio, and R. Tycko. A strutural model for alzheimer's b-amyloid fibrils based on experimental constraints from solid state nmr. *PNAS*, 99:16742–16747, 2002.

[11] J. Hadamard. *Lectures of the Cauchy Problem in Linear Partial Differential Equation*, volume 3. Yale University Press, New Haven, 1923.

[12] Heinz W. Engl, Martin Hanke, and Andreas Neubrecht. *Regularization of Inverse Problems.* Kluwer, Dordrecht, 1996.

[13] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Pitman, London, 1984.

[14] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* J. Wiley, New York, 1977.

[15] A. K. Luis. *Invers und schlecht gestellte Probleme.* Teubner, Stuttgart, 1989.

[16] J. Skilling. *Maximum Entropy and Bayesian Methods.* Kluwer, Dordrecht, 1989.

[17] V.A. Morozov. *Methods for solving incorrectly posed problems.* Springer, New York, 1984.

[18] J. Honerkamp and J. Weese. Tikhonov's regularization method for ill-posed problems. A comparison of different methods for the determination of the regularization parameter. *Contin. Mech. Thermodyn.*, 2:17–30, 1990.

[19] S.W. Provencher. A constrained regularization method for inverting data represented by linear algebraic or integral equations. *Comput. Phys. Commun.*, 27:213–227, 1982.

[20] S.W. Provencher. Contin: A general purpose constrained regularization program inverting noisy linear algebraic or integral equations. *Comput. Phys. Commun.*, 27:229–242, 1982.

[21] J. Weese. A reliable and fast method for the solution of Fredholm integral equations of the first kind based on Tikhonov regularization. *Comp. Phys. Commun.*, 69:99–111, 1992.

[22] J. Weese. PhD thesis, Universität Freiburg, 1989.

[23] C. van Loan. *SIAM J. Numer. Anal.*, 13:76, 1976.

[24] J. Stoer. *Numerische Mathematik 1*, volume 1. Springer, Berlin, 1989.

[25] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T.Vetterling. *Numerical Recipes.* Cambridge Univ. Press, Cambridge, 1986.

[26] J. Stoer. *SIAM J. Numer. Anal.*, 8:382, 1971.

[27] J. Jeener, B.H. Meier, P. Bachmann, and R.R. Ernst. *J. Chem. Phys.*, 71:4546, 1979.

[28] M. Mehring. *principles of High Resolution NMR in Solids.* Springer-Verlag, Heidelberg, 1983.

[29] R. Tycko. *Israel J. Chem.*, 32:179, 1992.

[30] D. Suter and R.R. Ernst. *Phys. Rev.*, B 32:5608, 1986.

[31] P.M. Henrichs and M. Linder. *J. Chem. Phys.*, 85:7077, 1986.

[32] R. Tycko and G. Dabbagh. Nuclear magnetic resonance crystallography: molecular orientational ordering in three forms of solid methanol. *J. Am. Chem. Soc.*, 113:3592–3593, 1991.

[33] G. Dabbagh, D.P. Weliky, and R. Tycko. *Macromolecules*, 27:6183, 1994.

[34] P. Robyr, B.H. Meier, P. Fischer, and R.R. Ernst. *J. Am. Chem. Soc.*, 116:5315, 1994.

[35] R. Tycko, D.P. Weliky, and A.E. Berger. Investigation of molecular structure in solids by two-dimensional NMR exchange spectroscopy with magic angle spinning. *J. Chem. Phys.*, 105:7915–7930, 1996.

[36] A.P.M. Kentgens, E. de Boer, and W.S. Veeman. *J. Chem. Phys.*, 87:6859, 1987.

[37] A. Hagemeyer, K. Schmidt-Rohr, and H.W. Spiess. *Adv. Magn. Reson.*, 13:85, 1989.

[38] Z. Luz, H.W. Spiess, and J.J. Titman. *Israel J. Chem.*, 32:145, 1992.

[39] R. Tycko. Biomolecular solid state NMR: advances in structural methodology and applications to peptide and protein fibrils. *Annu. Rev. Phys. Chem.*, 52:575–606, 2001.

[40] R. Tycko and G. Dabbagh. Double-quantum filtering in magic-angle-spinning NMR spectroscopy: an approach to spectral simplification and molecular structure determination. *J. Am. Chem. Soc.*, 113:9444–9448, 1991.

[41] R. Tycko and G. Dabbagh. *Chem. Phys. Lett.*, 173:461–465, 1990.

[42] A.E. Bennett, R.G. Griffin, and S. Vega. *NMR Basic Prin. Prog.*, 33:1–77, 1994.

[43] A.E. Bennett, J.H. Ok, S. Vega, and R.G. Griffin. *J. Chem. Phys.*, 96:8624–8627, 1992.

[44] T. Gullion and S. Vega. *Chem. Phys. Lett.*, 194:423–428, 1992.

[45] R. Tycko and S.O. Smith. *J. Chem. Phys.*, 98:932–943, 1993.

[46] S. Vega and A. Pines. *J. Chem. Phys.*, 66:5624–5644, 1977.

[47] A. Wokaun and R.R. Ernst. *Chem. Phys. Lett.*, 52:407–412, 1977.

[48] R. Tycko. *J. Chem. Phys.*, 92(10):5776–5793, 1990.

[49] A.E. Bennett, D.P. Weliky, and R. Tycko. Quantitative conformational measurements in solid state nmr by constant-time homonuclear dipolar recoupling. *J. Am. Chem. Soc.*, 120:4897–4898, 1998.