

Parametric Estimation of NMR Signals in the Time Domain

SIMON HULSE

Balliol College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term, 2023

Supervisors:

Mohammadali Foroozandeh

Timothy Claridge

Fay Probert

Dedication goes here

ACKNOWLEDGEMENTS

ABSTRACT

ACRONYMS

1D one-dimensional

2D two-dimensional

2DJ 2D J-resolved

AIC Akaike information criterion

ALPESTRE a linear predictive estimation of signal time reversal

API application programming interface

AR autoregressive

ARMA autoregressive moving average

AWGN additive white gaussian noise

BIRD bilinear rotation decoupling

COSY correlation spectroscopy

CPMG Carl-Purcell-Meiboom-Gill

CPU central processing unit

CUPID computer-assisted undiminished-sensitivity protocol for ideal decoupling

DFT density functional theory

DMSO Dimethyl sulfoxide, $(\text{H}_3\text{C})_2\text{SO}$

DMSO-d₆ Deuterated DMSO

DOSY diffusion-ordered spectroscopy

EYM Eckart-Young-Mirsky

FID free induction decay

FFT fast Fourier transform

FS frequency-swept

FT Fourier transform

GN Gauss-Newton

GUI graphical user interface

IFT inverse Fourier transform

INEPT insensitive nuclei enhancement by polarization transfer

LM Levenberg-Marquardt

LP linear prediction

LPSVD linear prediction singular value decomposition

MDL minimum description length

MEMPM matrix enhancement and matrix pencil method

MLE maximum likelihood estimate

MMEMPM modified matrix enhancement and matrix pencil method

MPM matrix pencil method

NLP non-linear programming

NMR nuclear magnetic resonance spectroscopy

NMR-EsPy NMR estimation in Python

NOESY nuclear overhauser effect spectroscopy

pdf probability density function

PFG pulsed field gradient

PGSE pulsed gradient spin echo

PGSTE pulsed gradient stimulated echo

PGSTEBP pulsed gradient stimulated echo with bipolar gradients

PSYCHE pure shift yeilded by chirp excitation

RAM random access memory

RF radio-frequency

SNR signal-to-noise ratio

ST Steihaug-Toint

SVD singular value decomposition

TSE-PSYCHE triple spin echo PSYCHE

TROSY transverse relaxation optimised spectroscopy

VE virtual echo

WURST wideband, uniform rate, smooth truncation

ZS Zanger-Sterk

NOMENCLATURE

General Mathematics

$\mathcal{A} := \mathcal{B}$	\mathcal{A} is defined to be equal to \mathcal{B}
$\lfloor a \rfloor$	The nearest integer to a , such that $\lfloor a \rfloor \leq a$
$\lceil a \rceil$	The nearest integer to a , such that $\lceil a \rceil \geq a$
$\lfloor a \rfloor$	The nearest integer to a
$a \bmod b$	a modulo b , given by $a \bmod b = a - \left\lfloor \frac{a}{b} \right\rfloor$

Complex Numbers

i	The imaginary unit, $i := \sqrt{-1}$
$\Re(z)$	The real component of z
$\Im(z)$	The imaginary component of z
$ z $	The absolute value of z , given by $ z = \sqrt{\Re(z)^2 + \Im(z)^2}$

Number sets

\mathbb{C}	The set of complex numbers
\mathbb{N}	The set of natural numbers with zero excluded
\mathbb{N}_0	The set of natural numbers with zero included
\mathbb{R}	The set of real numbers
$\mathbb{R}_{>0}$	The set of positive real numbers
\mathbb{Z}	The set of integers

Probability

$x \sim \mathcal{X}$	x behaves according to distribution \mathcal{X}
$A \perp\!\!\!\perp B$	A and B are conditionally independent
$\mathcal{N}(\mu, \sigma^2)$	Normal (Gaussian) distribution with mean μ and variance σ^2 :

$$\mathcal{N}(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

$\mathcal{U}(l, r)$	Uniform distribution with bounds l and r :
	$\mathcal{U}(x l, r) = \begin{cases} \frac{1}{r-l} & l \leq x \leq r \\ 0 & \text{otherwise} \end{cases}$

$\mathcal{N}_C(\mu, \sigma^2)$	Complex normal distribution with mean μ and variance $\sigma^2/2$
--------------------------------	---

Vectors, Matrices, and Arrays

.T	Transpose
.†	Conjugate transpose, equivalent to $.*^T$
. $\circlearrowleft(d)$	Right circular rotation by along axis d of the array by one element. For example, for a 2D matrix $\mathbf{A} \in \mathbb{F}^{N^{(1)} \times N^{(2)}}$, $\mathbf{A}^{\circlearrowleft(1)}$ is given by
	$\begin{bmatrix} \mathbf{A}[N^{(1)} - 1, 0] & \mathbf{A}[N^{(1)} - 1, 1] & \dots & \mathbf{A}[N^{(1)} - 1, N^{(2)} - 1] \\ \mathbf{A}[0, 0] & \mathbf{A}[0, 1] & \dots & \mathbf{A}[0, N^{(2)} - 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}[N^{(1)} - 2, 0] & \mathbf{A}[N^{(1)} - 2, 1] & \dots & \mathbf{A}[N^{(1)} - 2, N^{(2)} - 1] \end{bmatrix}$

\mathbf{A}^+	The Moore-Penrose pseudo inverse of $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a matrix $\in \mathbb{C}^{n \times m}$ given by
	$\mathbf{A}^+ = (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger.$

diag(\mathbf{v})	Given a vector $\mathbf{v} \in \mathbb{F}^N$, this creates a matrix $\in \mathbb{F}^{N \times N}$ of the form
	$\text{diag}(\mathbf{v}) = \begin{bmatrix} \mathbf{v}[0] & 0 & \dots & 0 \\ 0 & \mathbf{v}[1] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{v}[N - 1] \end{bmatrix}$

$\langle \mathbf{A}, \mathbf{B} \rangle$ The inner product of \mathbf{A} and \mathbf{B} , two D -dimensional arrays over the field \mathbb{C} with the same dimensions, given by

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{n^{(1)}=0}^{N^{(1)}-1} \cdots \sum_{n^{(D)}=0}^{N^{(D)}-1} \left(\mathbf{A} \begin{bmatrix} n^{(1)}, \dots, n^{(D)} \end{bmatrix}^* \mathbf{B} \begin{bmatrix} n^{(1)}, \dots, n^{(D)} \end{bmatrix} \right)$$

$\|\mathbf{A}\|$ The norm of \mathbf{A} , given by

$$\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$$

$\nabla f(\mathbf{x})$ The gradient of a differentiable, scalar value function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, given by

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}[0]} & \frac{\partial f}{\partial \mathbf{x}[1]} & \cdots & \frac{\partial f}{\partial \mathbf{x}[n-1]} \end{bmatrix}^T$$

$\nabla^2 f(\mathbf{x})$ The Hessian of a twice-differentiable, scalar value function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, given by

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial \mathbf{x}[0]^2} & \frac{\partial^2 f}{\partial \mathbf{x}[0] \partial \mathbf{x}[1]} & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}[0] \partial \mathbf{x}[n-1]} \\ \frac{\partial^2 f}{\partial \mathbf{x}[1] \partial \mathbf{x}[0]} & \frac{\partial^2 f}{\partial \mathbf{x}[1]^2} & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}[1] \partial \mathbf{x}[n-1]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \mathbf{x}[n-1] \partial \mathbf{x}[0]} & \frac{\partial^2 f}{\partial \mathbf{x}[n-1] \partial \mathbf{x}[1]} & \cdots & \frac{\partial^2 f}{\partial \mathbf{x}[n-1]^2} \end{bmatrix}$$

Other symbols

$f_{\text{off}}^{(d)}$ Transmitter offset in dimension d

$f_{\text{sw}}^{(d)}$ Sweep width (spectral window) in dimension d

$\Delta_t^{(d)}$ Sampling rate in dimension d

TABLE OF CONTENTS

1	INTRODUCTION	1
1.1	Introducing NMR	1
1.1.1	A Brief History of NMR	1
1.1.2	Nuclear Spin and Magnetism	2
1.1.3	The NMR Spectrometer	5
1.1.4	The Acquisition and Structure of NMR Data	6
1.2	An Overview of NMR Data Analysis	9
1.2.1	Conventional NMR Analysis	10
1.2.2	Conventional analysis for multidimensional datasets	14
1.2.3	Estimation Techniques for NMR Analysis	16
1.2.4	Signal-Noise Subspace Separation Techniques	17
1.2.5	Bayesian Methods	18
1.2.6	Iterative Schemes	18
2	THEORY	21
2.1	Outline of the Problem	21
2.2	Generating an Initial Guess: Matrix Pencil Method	22
2.2.1	1D Matrix Pencil Method	23
2.2.2	2D Matrix Enhancement and Matrix Pencil Method	27
2.2.3	Model Order Selection	30
2.3	Non-linear programming for NMR estimation	32
2.3.1	An overview of non-linear programming	32
2.3.2	Non-linear programming applied to FID estimation	35
2.3.3	Approximating the Hessian	36
2.3.4	Phase Variance Minimisation	37
2.3.5	Estimation Errors	40
2.3.6	Visualisation of a simple example	41
2.4	Profiling the MPM and NLP	43

2.5 Frequency Filtration	45
2.5.1 The virtual echo	45
2.5.2 The filtering process	48
2.6 Summary	51
3 RESULTS AND APPLICATIONS	53
3.1 Evaluating the Estimation Routine	53
3.1.1 1D Datasets	53
3.2 2D Datasets	58
3.3 Pure-Shift Spectra via 2D J-resolved Estimation	58
3.3.1 An Overview of Pure Shift NMR	59
3.3.2 An outline of CUPID	63
3.3.3 Results Using CUPID	67
3.4 Amplitude-attenuated Datasets	74
3.4.1 Relaxation experiments	74
3.4.2 Diffusion experiments	75
3.4.3 Methodology	79
3.4.4 Results	83
3.5 Phased broadband spectra from chirp excitation	85
3.5.1 Chirp excitation	88
3.5.2 Methodology	89
3.5.3 Results	90
4 THE NMR-EsPy PACKAGE	93
4.1 The Structure of NMR-EsPy	93
4.1.1 Why PYTHON?	93
4.1.2 ???	94
5 CONCLUSIONS AND FUTURE WORK	97
A ADDITIONAL THEORY	109
A.1 Additional Algorithms	109

B CODE LISTINGS	115
B.1 Matrix pencil methods	115
B.1.1 MDL	115
B.1.2 MPM	116
B.1.3 MMEMPM	117
B.2 NLP	121
B.2.1 Trust Region Algorithm	121
B.2.2 Computing \mathcal{F}_ϕ , $\nabla \mathcal{F}_\phi$, and $\nabla^2 \mathcal{F}_\phi$	124
B.2.3 The main routine	126
C INFORMATION ON DATASETS	129
C.1 Simulated datasets	129
C.1.1 SPINACH Simulations	129
C.2 Experimental datasets	134
C.2.1 Structures	134
C.2.2 Datasets related to CUPID	135
D MISCELLANEOUS TOPICS OF INTEREST	139
D.0.1 The Nyquist Frequency	139
D.0.2 Hessian Validation	139
D.0.3 Statistical Definitions	140

LIST OF FIGURES

1.1	The variation of energy of the spin states of ^1H , ^7Li , and ^{17}O with external magnetic field strength.	4
1.2	An simple illustration of quadrature detection.	7
1.3	An illustration of the influence of the four parameters associated with an oscillator in both the time-domain nels and Fourier-domain.	11
1.4	Spectra acquired from amplitude- and phase-modulated 2D signals.	15
2.1	An visualisation of the behaviour of the MDL for three different FIDs comprising the same model, but with different noise variances.	31
2.2	A visualisation of the trajectory of a 2-parameter optimisation involving a simulated FID comprising a single resonance.	42
2.3	TODO	44
2.4	An illustration of the filtering procedure applied to a 1D FID.	49
3.1	The result of estimating a series of 5 simulated signals comprising 20 oscillators, using solely the MPM and also with phase variance-regularised NLP afterwards.	54
3.2	Result of applying the estimation routine to selected regions of a pulse-acquire dataset of andrographolide.	57
3.3	Example of a simple 2DJ spectrum derived from an AMX spin system.	60
3.4	The pulse sequences of some common pure shift experiments.	62
3.5	An illustration of the reasoning behind the name “ -45° signal”.	65
3.6	An illustration of the filtering procedure for 2D J-resolved (2DJ) data.	67
3.7	The result of applying CUPID to 5 instances of simulated 2DJ datasets with 4 heavily overlapping multiplet structures.	68
3.8	Application of CUPID on a simulated sucrose 2DJ dataset.	69
3.9	Application of CUPID on the non-aromatic regions of a quinine 2DJ dataset.	70
3.10	Application of CUPID on a camphor dataset.	71
3.11	Application of CUPID on a dexamethasone dataset.	73
3.12	Application of CUPID on a 17β -estradiol dataset.	74
3.13	Pulse sequences used for the determination of translational diffusion constants.	76
3.14	Three examples of results generated on simulated inversion recovery datasets comprising five ddd multiplet structures.	84
3.15	Result of estimating a Oneshot DOSY dataset of andrographolide.	86

3.16	An illustration of an experiment comprising a single chirp pulse.	87
3.17	Comparison of quadratic phase correction vs frequency-dependent back-propagation in treating simulated single-chirp excitation data.	90
3.18	Comparison of quadratic phase correction vs frequency-dependent back-propagation in treating experimental single-chirp excitation data generated from a sample of Gd-doped H ₂ O in D ₂ O.	91
A1	The molecular structures of species giving rise to the experimental NMR datasets considered in this work.	134
A2	PSYCHE pulse sequence used for the acquisition of estradiol data.	136
A3	TSE-PSYCHE pulse sequence used for the acquisition of dexamethasone data. .	136

LIST OF TABLES

1.1	A table of regularly encountered nuclei in nuclear magnetic resonance spectroscopy (NMR), along with common nuclei which are not NMR active.	3
2.1	The number of first and second derivatives that are necessary to compute both the gradient vector and Hessian matrix of the fidelity for 1- 2- and 3-dimensional datasets, as well as a general D -dimensional dataset.	36
3.1	The various functional forms of \mathcal{A} according to the different amplitude-attenuating NMR experiments considered.	80
4.1	A complete list of estimator objects provided by NMR estimation in Python (NMR-EsPy) at the time of writing.	94
A1	The isotropic chemical shifts, scalar couplings and relaxation times associated with spin systems used in SPINACH simulations.	130
A2	Experiment parameters for 2DJ simulations run using SPINACH.	133
A3	Experiment parameters for inversion recovery simulations run using SPINACH. .	134
A4	Andrographolide multiplet structures	135
A5	Noteworthy experiment parameters for the 2D J-Reolved and PSYCHE experiments run.	137

LIST OF ALGORITHMS

1	The matrix pencil method.	26
2	Nonlinear programming routine employed in NMR-EsPy.	41
3	An overview of the estimation procedure outlined in this work.	52
4	Routine for estimating a sequence of one-dimensional (1D) free induction decays (FIDs) which exhibit variation in amplitudes across increments. NLPAMP denotes a routine which is akin to NLP (Algorithm 2), except only amplitudes are allowed to be altered, whilst phases, frequencies and damping factors are fixed.	81
A1	The MMEMPM.	110
A2	Steihaug-Toint method for determining an update for nonlinear programming.	111
A3	Filtering procedure for 1D data.	112
A4	Filtering procedure for 2D data.	113
A5	Filtering procedure for 2DJ data.	114
A6	An algorithm for multiplet assignment of a 2DJ estimation result.	114

INTRODUCTION

1

Since its conception almost 80 years ago, NMR has become a ubiquitous technique in chemistry, biochemistry and numerous other disciplines, thanks to the unique insights into structure and dynamics it can provide. In this chapter, after a very brief historical overview of the subject, an outline of the basic theory behind the experiment – without delving into the necessary quantum mechanics for a truly formal account – is given. Details of the means by which the raw experimental data, the FID is produced, and its structure, are given. Following on from this, an account is given of techniques used to extract quantitative information from the FID, which is the central motivation of this work.

1.1 Introducing NMR

1.1.1 A Brief History of NMR

The origins of NMR can be traced back to 1945, when independent work by Felix Bloch on water[1] and Edward Purcell on parrafin[2] gave rise to the first illustrations of nuclear magnetic resonances in condensed phases. The two hadn't met before their respective papers were published with about a month's separation[3]. Both received the Nobel Prize in Physics in 1952 for their pioneering work in the field. A notable mention should also be given to Yevgeny Zavoisky – the father of Electron Paramagnetic Resonance – who probably observed NMR as far back as 1941[4]. Alas, he dismissed his results as irreproducible. In 1949 and 1950, work investigating NMR spectra from compounds containing Cu, ^{31}P , ^{14}N , and ^{19}F nuclei illustrated the concept of the chemical shift[5–7], in which nuclei in different chemical environments exhibit non-identical resonant frequencies. Chemists regarded these results with great interest, as these findings suggested that NMR could give insights into molecular structure.

Russel Varian secured the first patent for a commercial NMR machine, and a 30 MHz spectrometer followed soon after. The first spectrometers were continuous wave, which functioned

by slowly sweeping the magnetic field, causing spins to come into resonance at different times. Richard Ernst and Weston Anderson, working at Varian Inc. at the time, proposed an alternative method: pulsed Fourier transform (FT) spectroscopy[8]. This was not seen as a fruitful endeavour by the company, largely because of the very long time it took to digitise the signal, and subsequently compute its FT[9]. Instead, the first commercial pulsed FT spectrometer was produced by Bruker Corp. in 1969, which revolutionised NMR. The emergence of the Cooley-Tukey's fast Fourier transform (FFT) algorithm[10] led to vast improvements in the speed with which experiments could be conducted, which incentivised the development of the new FT approach.

The idea of 2D NMR spectroscopy was proposed by Jean Jeener in 1971[11, 12], which Ernst and co-workers showcased a few years later in the form of a correlation spectroscopy (COSY) experiment[13]. The use of multiple dimensions to spread out signals enabled vastly more complex structures to be studies. In 1985, a report of the first protein assigned by NMR (using COSY and nuclear overhauser effect spectroscopy (NOESY)) was presented by Kurt Wüthrich and co-workers[14]. Over time, extensive developments in techniques for biomolecular systems have occurred, including the creation of 3D and 4D experiments[15, 16], as well transverse relaxation optimised spectroscopy (TROSY) experiments[17] for the study of large proteins.

NMR's significance as an analytical tool is evidenced by Nobel Prizes in Chemistry being awarded for work in the field on two separate occasions. First, Ernst received the prize in 1991 “for his contributions to the development of the methodology of high resolution nuclear magnetic resonance spectroscopy”[18]. In 2002, Wüthrich was recognised “for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution”[19].

1.1.2 Nuclear Spin and Magnetism

NMR relies on *spin*, an intrinsic property of certain nuclei (along with other elementary particles) which, along with orbital angular momentum, is one of the two sources of angular momentum in quantum mechanics. The angular momentum associated with a nuclear spin is characterised by the quantum number I , which may be integer or half-integer. Spin- $\frac{1}{2}$ nuclei are the most commonly studied in NMR, as those with $I > \frac{1}{2}$ often have very short-lived excited states, due to electric quadrupole effects. While a rigorous description of the theory behind NMR requires quantum mechanics[20–22], the fundamentals can be understood using a classical description of angular momentum.

The nuclear spin angular momentum $\mathbf{I} \in \mathbb{R}^3$ is a vector with the following (square) magnitude:

$$\mathbf{I}^2 = \mathbf{I} \cdot \mathbf{I} = \hbar I(I + 1) \quad (1.1)$$

where $\hbar = h/2\pi$ is the reduced Planck constant. Whilst it is not possible to specify multiple com-

Nucleus	I	γ (rad T $^{-1}$ s $^{-1}$)	Relative Abundance (%)
^1H	1/2	2.6752×10^8	99.9885
^2H	1	4.1066×10^7	0.0115
^6Li	1	3.9371×10^7	7.59
^7Li	3/2	1.0398×10^8	92.41
^{12}C	0	-	98.93
^{13}C	1/2	6.7283×10^7	1.07
^{14}N	1	1.9338×10^7	99.636
^{15}N	1/2	-2.7126×10^7	0.364
^{16}O	0	-	99.756
^{17}O	5/2	-3.6281×10^7	0.038
^{19}F	1/2	2.5162×10^8	100
^{31}P	1/2	1.0839×10^8	100
^{113}Cd	1/2	-5.9609×10^7	12.26

TABLE I.I: A table of regularly encountered nuclei in NMR, along with common nuclei which are not NMR active.

ponents of the angular momentum simultaneously due to the uncertainty principle, it is possible to specify one along with \mathbf{I}^2 . Conventionally, this is chosen to be the z-component, I_z , for which

$$I_z = \hbar m \quad (1.2)$$

where $m \in \{-I, -I + 1, \dots, +I\}$. Eqs. 1.1 & 1.2 imply that the orientation of the z-component may only adopt certain discrete values (i.e. it is quantised). A nucleus with non-zero spin has an associated *magnetic moment*, given by:

$$\boldsymbol{\mu} = \gamma \mathbf{I} \implies \mu_z = \gamma I_z = \gamma \hbar m \quad (1.3)$$

γ is a proportionality constant called the *gyromagnetic ratio*, which is dependent on the nucleus of interest. Table 1.1 provides the gyromagnetic ratios for a few nuclei commonly encountered in NMR.

Without the presence of an external field, the nuclear spin states of different m are degenerate. However, when subjected to a magnetic field, the energies of the different states diverge. This phenomenon is known as the *Zeeman effect*. the associated energy of a given magnetic moment is given by

$$E = -\boldsymbol{\mu} \cdot \mathbf{B} \quad (1.4)$$

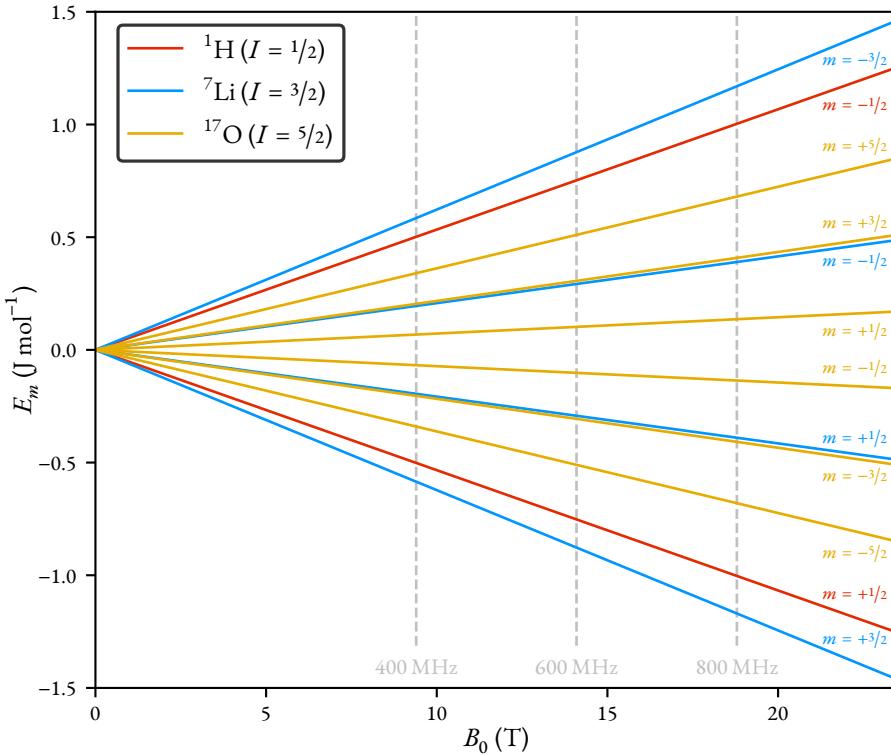


FIGURE 1.1: The variation of energy of the spin states of ^1H , ^7Li , and ^{17}O with external magnetic field strength (B_0), up to 23.5 T, which is approximately the strength of a 1 GHz NMR magnet. Three common field strengths for commercial NMR magnets are indicated: 9.40 T (400 MHz), 14.10 T (600 MHz), and 18.79 T (800 MHz).

where \mathbf{B} is the magnetic field vector. It is conventional to define the external field as directed along the laboratory z -axis, such that $B_x = B_y = 0$ and $B_z = B_0$ where B_0 is the magnetic field strength. The energies of the individual spin states are therefore

$$E_m = -\gamma I_z B_0 = -m\hbar\gamma B_0 \quad (1.5)$$

Figure 1.1 illustrates how the relative energies of different spin states vary with B_0 .

REVISIT THIS SECTION NMR samples comprise a vast ensemble of equivalent spin systems, and it is the macroscopic properties of the sample that are observed. Consider an ensemble of N spin- I nuclei. At thermal equilibrium, the various spin states will be disproportionately populated in accordance with the Boltzmann distribution. Assuming that $m\hbar\gamma B_0/k_B T \gg 1$, which is the case for conventional temperatures, the proportion of spins in state m can be shown to be given approximately by[23]

$$\frac{N_m}{N} \approx \frac{1}{2I+1} \left(1 + \frac{m\hbar\gamma B_0}{k_B T} \right) \quad (1.6)$$

where k_B is the Boltzmann constant. The ensemble has an associated bulk magnetic moment \mathbf{M} ,

given by the summation of all the individual spin magnetic moments:

$$\mathbf{M} = \gamma \mathbf{J} = \sum_{n=1}^N \boldsymbol{\mu}_n \quad (1.7)$$

where \mathbf{J} is the spin angular momentum of the ensemble. At equilibrium, the x - and y -components of the bulk magnetisation are zero, i.e. $\sum_{n=1}^N \mu_{x,n} = \sum_{n=1}^N \mu_{y,n} = 0$, such that magnetisation is colinear with the field direction: $\mathbf{M} = [0, 0, M_0]^T$, where M_0 is the magnitude of the bulk magnetism. The time evolution of the angular momentum is equivalently the torque, which is given by

$$\boldsymbol{\tau} = \frac{d\mathbf{J}(t)}{dt} = \mathbf{M}(t) \times \mathbf{B}(t) \quad (1.8)$$

Therefore, the rate of change of the bulk magnetism is

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma \mathbf{B}(t), \quad (1.9)$$

which results in the following expression under the assumption that $\mathbf{B}(t)$ is static and directed along the z -direction:

1.1.3 The NMR Spectrometer

A modern NMR spectrometer has the capability of conducting a plethora of experiments, thanks to the availability of high magnetic fields, sophisticated electronics for pulse generation, and signal acquisition. An overview of the key components of the spectrometer is presented here.

The Magnet

The resolution of NMR data increases linearly with B_0 , while sensitivity increases as $B_0^{3/2}$ [20], such that it is desirable to use magnets that produce the highest fields possible.* Conventional NMR magnets feature a superconducting solenoid immersed in liquid He, which has a boiling point of 4.2 K at atmospheric pressure. Common materials used for the solenoid include the type II superconductors Nb-Ti and Nb₃Sn. The dewar that the Helium is contained in is lined with a thermal radiation shield, which is surrounded by a larger dewar of liquid N₂, to minimise the extent of He evaporation. Passing through the magnet's z -direction is a bore, which is maintained at ambient temperature, or some other user specified temperature. Within this bore sits the sample, as well as the probe.

To maintain high spatial field homogeneity - a necessity for results of acceptable resolution - a series of coils called *shims* surround the sample. Each coil produces a weak magnetic field with a

*This relation isn't perfect, when features such as a nucleus' chemical shift anisotropy are considered.

specific spatial profile, which can cancel out any inhomogeneity inherent to the main magnet.

The Probe

The probe features the coil(s) used to pulse the sample with radio-frequency (RF) radiation, as well as receive the resultant signal. A relatively recent development in probe technology is the cryogenic probe[24–26]. The transmit/receive coil(s) and other probe circuitry are maintained at very low temperature ($\sim 25\text{K}$), which reduces Johnson-Nyquist noise. This manifests in a higher signal-to-noise ratio (SNR), which can be as large as 4 times that achievable with conventional probes.

The Transmitter

The transmitter is the collection of components which generates RF pulses. A synthesiser acts as an RF source, producing a continuous carrier wave at or very close to the Larmor frequency. Pulses are subsequently produced by a modulator, and are then amplified from a power of a few mW, to daW, or even hW[27].

The Receiver

The Pulse Programmer

1.1.4 The Acquisition and Structure of NMR Data

In NMR experiments, RF pulses manipulate nuclear magnetic spin states. While there is a plethora of different pulse sequences, the net effect of (virtually[†]) all of them is the generation of some magnetisation which is transverse to the field direction. The evolution of this magnetisation induces a time-varying electromotive force within the coil(s) of the probe, with the signal generated being amplified and digitised by the receiver, generating an FID. The FID is sampled at equally-spaced time intervals Δ_t , such that it is of the form

$$\mathbf{y} = \begin{bmatrix} y(t=0) & y(t=\Delta_t) & y(t=2\Delta_t) & \cdots & y(t=(N-1)\Delta_t) \end{bmatrix}^T \quad (1.10)$$

where $y(t)$ is the (continuous) variation of the generated signal with time, and N is the number of points sampled. The inverse of the sampling rate, $1/\Delta_t$ is the sweep width or spectral window f_{sw} which defines how wide the range of samplable frequencies is (see Appendix Section D.0.1 for more on this).

On modern spectrometers, a quadrature acquisition system is used[23: Appendix A.5] (see Figure 1.2), such that FIDs adopt the form of a summation of M complex exponentials (oscillators), where $M \in \mathbb{N}$ is the number of resonances contributing to the signal. Each resonance will

[†]In a pulse calibration experiment for example, the goal is to rotate the spin magnetisation by exactly 360° , leading to only longitudinal (undetectable) magnetisation.

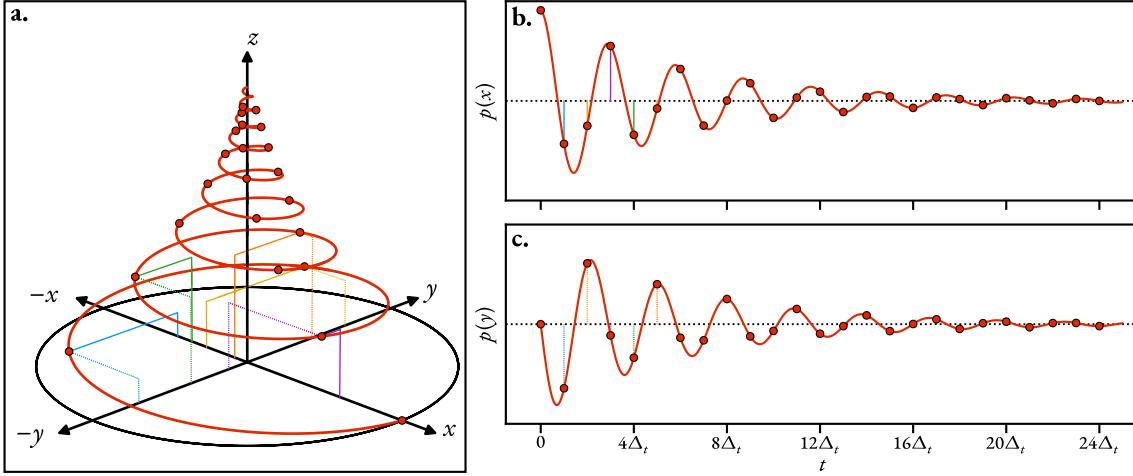


FIGURE 1.2: An illustration of quadrature detection using the vector model[28: Chapter 1]. **a.** An illustration of the evolution of the bulk magnetisation for a system comprising many identical spins, aligned initially along the laboratory x -axis (for example, immediately after a 90° - y pulse in a pulse-acquire experiment). At regular time intervals separated by Δ_t , the signal is sampled, such that the projection onto the x -axis, $p(x)$, and onto the y -axis, $p(y)$ is measured. For the first few samples, the projections onto the x - and y -axes are denoted by solid and dashed coloured lines, respectively. The complete projections are given in panels **b.** and **c..** The FID is given by the complex value $p(x) + ip(y)$.

be subjected to damping due to relaxation phenomena, which is typically exponential in nature. An FID acquired with N samples $\mathbf{y} \in \mathbb{C}^N$ therefore takes the form

$$\mathbf{y}[n] = \mathbf{x}[n] + \mathbf{w}[n] \quad \forall n \in \{0, 1, \dots, N-1\}, \quad (1.11a)$$

$$\mathbf{x}[n] = \sum_{m=0}^{M-1} \boldsymbol{\alpha}[m] \exp(i\phi[m]) \exp([2\pi i (\mathbf{f}[m] - f_{\text{off}}) - \boldsymbol{\eta}[m]] n\Delta_t). \quad (1.11b)$$

f_{off} is included to account for the transmitter offset, which is the difference between the carrier wave frequency of the transmitter and the spectrometer basic frequency. (1.11) indicates that an FID comprises contributions from the (modellable) evolution of the spin magnetisation \mathbf{x} and experimental noise \mathbf{w} (*vide infra*). Each oscillator which contributes to \mathbf{x} is defined by four parameters:

- ❖ $\alpha \in \mathbb{R}_{>0}$ - the amplitude,
- ❖ $\phi \in (-\pi, \pi]$ - the phase,
- ❖ $f \in [f_{\text{off}} - 1/2 f_{\text{sw}}, f_{\text{off}} + 1/2 f_{\text{sw}}]$ - the frequency,
- ❖ $\eta \in \mathbb{R}_{>0}$ - the damping factor.

an FID can therefore be parameterised by the vector $\boldsymbol{\theta} \in \mathbb{R}^{4M}$:

$$\boldsymbol{\theta} = [\boldsymbol{\alpha}^T \quad \boldsymbol{\phi}^T \quad \mathbf{f}^T \quad \boldsymbol{\eta}^T]^T, \quad (1.12)$$

where $\boldsymbol{a} \in \mathbb{R}^M$ is a vector of all amplitudes, $\boldsymbol{\phi} \in \mathbb{R}^M$ is a vector of all phases, etc.

Multidimensional experiments involve incrementing one or more experimental parameters (a common example being a particular delay period in the pulse sequence), in order to obtain an array of 1D FIDs. In a D -dimensional dataset, each contributing resonance is parameterised by an amplitude and phase as before, and also D distinct frequencies and corresponding damping factors, such that a general parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{2(1+D)M}$ is given by

$$\boldsymbol{\theta} = \left[\boldsymbol{a}^T \quad \boldsymbol{\phi}^T \quad [\boldsymbol{f}^{(1)}]^T \quad \dots \quad [\boldsymbol{f}^{(D)}]^T \quad [\boldsymbol{\eta}^{(1)}]^T \quad \dots \quad [\boldsymbol{\eta}^{(D)}]^T \right]^T, \quad (1.13)$$

where $\boldsymbol{f}^{(D)}$ and $\boldsymbol{\eta}^{(D)}$ are the frequencies and damping factors in the actively acquired (direct) dimension, and $\boldsymbol{f}^{(1)} - \boldsymbol{f}^{(D-1)}$ and $\boldsymbol{\eta}^{(1)} - \boldsymbol{\eta}^{(D-1)}$ are those for the indirect dimension(s). Indirect dimensions can exhibit different forms of evolution, depending on the precise nature of the pulse sequence. Two common functional forms exist[22: Section 4.3.4]. Signals of the form $\cos(2\pi ft)$ and $\sin(2\pi ft)$ modulate the amplitude of the direct dimension signal across increments, while those of the form $\exp(2\pi ift)$ or $\exp(-2\pi ift)$ modulate the phase. Amplitude- or phase-modulated signals are often acquired as pairs when possible, as this ensures that frequency-discriminated spectra can be generated with absorption-mode Lorentzian lineshapes in all dimensions (*vide infra*). In general, a D -dimensional FID $\mathbf{Y} \in \mathbb{C}^{N^{(1)} \times \dots \times N^{(D)}}$ can be expressed as

$$\begin{aligned} \mathbf{Y}[n^{(1)}, \dots, n^{(D)}] &= \sum_{m=0}^{M-1} \boldsymbol{a}[m] \exp(i\boldsymbol{\phi}[m]) \\ &\times \prod_{d=1}^D \zeta\left(2\pi\left(\boldsymbol{f}^{(d)}[m] - f_{\text{off}}^{(d)}\right)n^{(d)}\Delta_t^{(d)}\right) \exp\left(-\boldsymbol{\eta}^{(d)}[m]n^{(d)}\Delta_t^{(d)}\right) \end{aligned} \quad (1.14a)$$

$$+ \mathbf{W}[n^{(1)}, \dots, n^{(D)}] \quad \zeta(\cdot) \begin{cases} = \exp(i \cdot) & d = D \\ \in \{\cos(\cdot), \sin(\cdot), \exp(i \cdot) \exp(-i \cdot)\} & \text{otherwise} \end{cases}, \quad (1.14b)$$

Without any specific knowledge about the experiment, it is typical to assume that \mathbf{W} is an array of additive white gaussian noise (AWGN), meaning that the noise instances are described by a complex normal distribution with mean 0, and pairs of noise instances are statistically independent, regardless of their time separation:

$$\begin{aligned} \mathbf{W}[n^{(1)}, \dots, n^{(D)}] &:= w \sim \mathcal{N}_c(0, 2\sigma^2) \\ \implies \Re(w) &\perp\!\!\!\perp \Im(w), \Re(w) \sim \mathcal{N}(0, \sigma^2), \Im(w) \sim \mathcal{N}(0, \sigma^2). \end{aligned} \quad (1.15)$$

Double check the equations are right for the SNR below... The extent by which a signal is

corrupted by noise is given by the signal-to-noise ratio (SNR), the ratio of signal power and noise power: **definition of power in appendix**

$$\begin{aligned} \text{SNR}(\mathbf{Y}) &:= \frac{P_{\mathbf{X}}}{P_{\mathbf{W}}} \\ &= \frac{1}{2\mathfrak{N}\sigma^2} \sum_{n^{(1)}=0}^{N^{(1)}-1} \cdots \sum_{n^{(d)}=0}^{N^{(D)}-1} \left| \mathbf{X}[n^{(1)}, \dots, n^{(D)}] \right|^2, \end{aligned} \quad (1.16)$$

where $\mathfrak{N} := N^{(1)} \dots N^{(D)}$ is the total number of points the signal comprises. Due to the large dynamic range that the SNR can exhibit, it is common to express it using a logarithmic scale instead, in units of decibels (dB):

$$\text{SNR}_{\text{dB}} := 10 \log_{10} (\text{SNR}). \quad (1.17)$$

From this, the noise variance associated with a signal whose noiseless component is \mathbf{X} is given by

$$\sigma^2 = \frac{1}{20^{\frac{\text{SNR}_{\text{dB}}}{10}} \mathfrak{N}} \sum_{n^{(1)}=0}^{N^{(1)}-1} \cdots \sum_{n^{(d)}=0}^{N^{(D)}-1} \left| \mathbf{X}[n^{(1)}, \dots, n^{(D)}] \right|^2. \quad (1.18)$$

1.2 An Overview of NMR Data Analysis

To gain insights from NMR experiments on the chemical system of interest, extraction of the defining parameters θ is necessary, though the majority of NMR users are unlikely to think about the process of NMR analysis in this way. For example:

- ❖ An understanding of the chemical environments of atoms in a molecule can be gained by considering the chemical shifts of the various peaks in the spectra, which are a proxy for the FID frequencies f .
- ❖ The relative stoichiometries **does this make sense?** of a molecule can be elucidated by inspecting the integrals of spectral peaks, which are directly related to the FID amplitudes a .

- ❖ **Typical approach to analysing the data (FT, peak pick, integrate, baseline correction, window functions, zero-filling etc),**
- ❖ **Estimation techniques: LP, SVD techniques, iterative techniques (AMARES, VARPRO), Bayesian techniques (CRAFT), ML techniques**

1.2.1 Conventional NMR Analysis

The conventional route to interpret NMR experiments is to transform the raw output (the FID) into the frequency domain, producing an NMR spectrum. This is achieved through application of the Fourier transform (FT):

$$\text{Continuous case: } \text{FT} \left(x(t^{(1)}) \right) (F^{(1)}) = \int_0^{\infty} x(t^{(1)}) \exp(-2\pi i t^{(1)} F^{(1)}) dt \quad \forall t \in \mathbb{R}, \quad (1.19a)$$

$$\text{Discrete case: } \text{FT}(x)[n] = \sum_{k=0}^{N-1} x[k] \exp\left(-\frac{2\pi i k n}{N}\right) \quad \forall n \in \{0, \dots, N-1\}. \quad (1.19b)$$

The FT of a single exponentially-damped complex sinusoid takes the form of a Lorentzian[‡]:

$$s(F^{(1)}) = \text{FT} \left(\alpha \exp(i\phi) \exp((i\Delta f^{(1)} - \gamma^{(1)})t) \right) (F^{(1)}), \quad (1.20a)$$

$$s(F^{(1)}) = \frac{\alpha \exp(i\phi)}{\gamma^{(1)} + i\Delta f^{(1)}}, \quad (1.20b)$$

$$\Delta f^{(1)} = 2\pi (f^{(1)} - F^{(1)}). \quad (1.20c)$$

When $\alpha = \phi = 0$, this function is equivalent to the (unnormalised) probability density function (pdf) of the Cauchy distribution. The FT is a linear function, such that the FT of a summation of signals is equivalent to the summation of the FTs of each signal. A corollary is that an NMR spectrum comprises a series of Lorentzian “peaks”, located at the resonance frequencies in the FID:

$$s(F^{(1)}) = \sum_{m=0}^{M-1} \frac{\alpha[m] \exp(i\phi[m])}{\gamma^{(1)}[m] + i\Delta f_m^{(1)}}. \quad (1.21)$$

The FT is a very attractive means of processing NMR data, as it presents the data in a format which is human-interpretable, with the basic rules describing how chemical structure is mapped to NMR spectral properties being a fundamental skill that virtually all experimental chemists require[30]. Due to innate properties of the NMR experiment, as well as features arising from analysing a discrete signal, the raw output from an NMR experiment often leads to spectra with undesirable characteristics without any extra manipulation. Extra processing steps that are frequently applied to NMR data are now outlined, in the order that they are applied to the data. Note that apodisation and zero filling are applied to the FID (prior to FT), while phase correction and baseline correction are applied to the spectrum (after FT).

[‡]An upshot of only possessing data for $t \geq 0$ is that the FT of an FID features a vertical offset, such that the baseline does not sit at 0[29]. This is corrected by halving the initial point of the FID prior to FT.

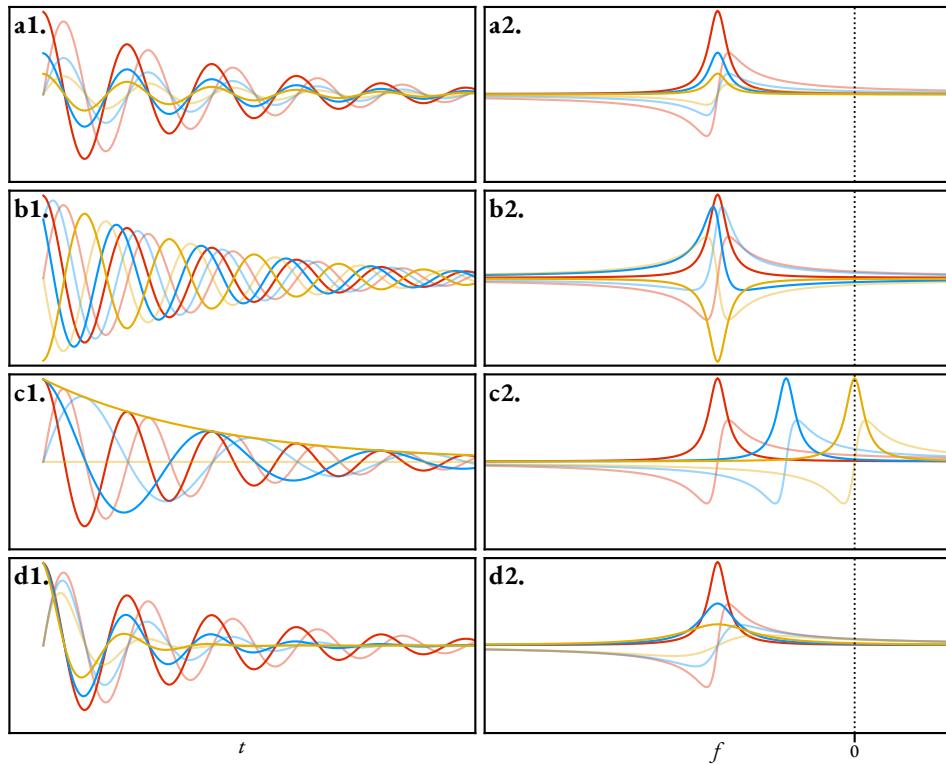


FIGURE 1.3: An illustration of the influence of the four parameters associated with an oscillator in both the time-domain (panels **a1.** – **d1.**) and Fourier-domain (panels **a2.** – **d2.**). The red signal is generated with the same parameters across all panels: $\alpha = \alpha_{\text{red}}$, $\phi = 0$, $f = f_{\text{red}}$, $\eta = \eta_{\text{red}}$. The blue and yellow signals were produced by altering one parameter out of the four. **a.** $\alpha_{\text{yellow}} = 1/2\alpha_{\text{blue}} = 1/4\alpha_{\text{red}}$. **b.** $\phi_{\text{blue}} = \pi/4$, $\phi_{\text{yellow}} = \pi$. **c.** $f_{\text{blue}} = 1/2f_{\text{red}}$, $f_{\text{yellow}} = 0$. **d.** $\eta_{\text{blue}} = 1/2\eta_{\text{red}}$, $\eta_{\text{yellow}} = 1/4\eta_{\text{red}}$. The real and imaginary components of each signal are plotted, with the imaginary component being paler than its real counterpart.

Apodisation

Apodisation refers to the process of mutating a signal by multiplying it with a specified function, often called a *window function*, in order to enhance a certain property. Apodisation is employed to improve either the sensitivity or the resolution of the final spectrum, albeit at the cost of worsening the other feature. There is no such thing as a free lunch after all.

Sensitivity Enhancement As an FID progresses with time, the contributions from the desirable signal and experimental noise becomes more weighted towards the noise, as spin relaxation phenomena dampen the signal. As such, multiplying the FID with a function which is initially large and gets progressively smaller with time can be used to enhance the SNR of the FID. The most common function to achieve this is the negative exponential such that a given point is multiplied by $\exp(-kn^{(1)}/N^{(1)} - 1)$ $\forall n^{(1)} \in \{0, \dots, N^{(1)} - 1\}$, with k often referred to as the line broadening factor, as the increased dampening applied to the signal causes the linewidth of the spectral peaks to increase (see panel d of Figure 1.3).

Resolution Enhancement By contrast, resolution enhancement can be achieved by applying a window function that artificially reduces the rate of oscillator decay, by suppressing points that are early in the FID. Popular examples of window functions for resolution enhancement are the Lorentz-Gauss function, which bestow a sharper, Gaussian shape to the peaks, and the sine-bell, which is commonly used in multidimensional experiments. Since the initial points in the FID are attenuated these window functions reduce the sensitivity of the resulting spectra.

By being defined to decay to 0 at the end of the FID, window functions are also able to suppress *truncation artefacts*, which appear when the FID still possesses appreciable signal amplitude at the end of the acquisition period. Truncated FIDs produce spectra with peaks of a form that is akin to the convolution of the FT of the untruncated FID, with the FT of a box-function, which takes the form of a sinc function ($\sin(x)/x$). The resulting artefacts in spectra are often referred to as *sinc wiggles* for this reason.

Zero filling

TODO

Phase correction

The real and imaginary components of (1.21) are as follows:

$$\Re(s(F^{(1)})) = \sum_m \boldsymbol{\alpha}[m] (\cos(\phi[m]) \mathcal{A}_m(F^{(1)}) + \sin(\phi[m]) \mathcal{D}_m(F^{(1)})), \quad (1.22a)$$

$$\Im(s(F^{(1)})) = \sum_m \boldsymbol{\alpha}[m] (\sin(\phi[m]) \mathcal{A}_m(F^{(1)}) - \cos(\phi[m]) \mathcal{D}_m(F^{(1)})), \quad (1.22b)$$

where \mathcal{A}_m and \mathcal{D}_m denote *absorption* and *dispersion* Lorentzians, respectively:

$$\mathcal{A}_m(F^{(1)}) = \frac{\gamma^{(1)}[m]}{(\gamma^{(1)}[m])^2 + (\Delta f_m^{(1)})^2}, \quad (1.23a)$$

$$\mathcal{D}_m(F^{(1)}) = \frac{\Delta f_m^{(1)}}{(\gamma^{(1)}[m])^2 + (\Delta f_m^{(1)})^2}. \quad (1.23b)$$

As illustrated most clearly in panel c2 of Figure 1.3, a peak with an absorption lineshape is far more desirable than one with a dispersion lineshape for two key reasons: (a) its maximum corresponds to the oscillator frequency, while a dispersion Lorentzian has a magnitude of 0 at the oscillator frequency (b) they decay more rapidly towards 0. Generating a spectrum where all peaks possess absorption Lorentzians is therefore desired, which is possible if all oscillators have a phase of 0° , such that:

$$\Re(s(F^{(1)})) = \sum_m \alpha[m] \mathcal{A}_m(F^{(1)}), \quad (1.24a)$$

$$\Im(s(F^{(1)})) = -\sum_m \alpha[m] \mathcal{D}_m(F^{(1)}). \quad (1.24b)$$

Fortunately, for the majority of NMR experiments[§], the phases of the contributing resonances possess a phase which depends linearly (to first-order) on their frequency, i.e.

$$\phi[m] = \phi_0 + \phi_1 f^{(1)}[m], \quad (1.25)$$

where $\phi_0 \in (-\pi, \pi]$ and $\phi_1 \in \mathbb{R}$ are zero- and first-order phase terms. A feature of all NMR processing platforms is the ability to perform phase-correction, in which the user effectively determines ϕ_0 and ϕ_1 by inspecting the appearance of the spectrum for different values of p_0 and p_1 , according to

$$s_\phi[n^{(1)}] = s[n^{(1)}] \exp\left(i\left(p_0 + \frac{p_1 n^{(1)}}{N^{(1)} - 1}\right)\right), \quad (1.26)$$

in an attempt to give all peaks in the spectrum absorption-mode lineshapes.

Baseline correction

The *baseline* of an NMR spectrum is used to describe regions where no discernible peaks reside (i.e. only experimental noise exists). Baseline distortion is used to describe scenarios when the baseline, rather than exhibiting a flat profile with a moving average of zero, has a distorted shape

[§]An example of an exception to this rule is the use of a frequency-swept, or “chirped” pulse, which generate datasets with quadratic phase behaviour. **refer to BBQCHILI section when written up.**

instead. There are a number of potential causes of baseline distortion, including acquisition starting at a time $\neq 0$, “clipping” of the initial points due to excessive receiver gain, and “baseline roll” due to the transient response of audio filters[22: Section 3.3, 29]. Whatever the cause(s), a corruption of the initial points in the FID typically leads to baseline distortion. It is common to apply baseline correction algorithm after phase correction has been undertaken to negate any distortion. This involves multiplying the spectrum with a high-order polynomial function, whose coefficients are determined by an appropriate algorithm. A popular algorithm involves the two-step procedure of (a) determining spectral regions which are part of the baseline (b) fitting the baseline regions to a polynomial[31, 32].

1.2.2 Conventional analysis for multidimensional datasets

As for 1D datasets, it is desirable that multidimensional NMR spectra feature peaks which are (a) frequency discriminated, and (b) comprise pure absorption-mode lineshapes, in each dimension. Frequency discrimination describes the ability to determine whether a particular resonance is at a frequency which higher or lower than the frequency of the transmitter, which is in the middle of the spectral window. As an illustration of how this can be achieved, two-dimensional (2D) signals comprising a single oscillator will be considered, with amplitude $\alpha = 1$, phase $\phi = 0^\circ$, frequencies $f^{(1)}$ and $f^{(2)}$, and damping factors $\gamma^{(1)}$ and $\gamma^{(2)}$.

Amplitude-modulated signals

It is clear that a cosine modulated signal, given by (1.14) with $D = 2$ and $\zeta^{(1)} = \cos(\cdot)$. cannot achieve frequency discrimination, on account of the relation

$$\cos(2\pi f^{(1)} t^{(1)}) = \frac{1}{2} (\exp(2\pi i f^{(1)} t^{(1)}) + \exp(-2\pi i f^{(1)} t^{(1)})) \quad (1.27)$$

FT of a cosine-modulated FID in both dimensions leads to a spectrum whose real component comprises two peaks, one at the true resonance frequency $(f^{(1)}, f^{(2)})$, and the other at the mirror-image frequency in the indirect dimension, $(f_{\text{off}}^{(1)} - f^{(1)}, f^{(2)})$. On top of this, the peaks possess a mixture of absorption and dispersion character, with the resultant peak shape often referred to as *phase twist*[33]. A spectrum of this form is presented in panel a of Figure 1.4. It is possible to generate pure absorption peaks by applying FT in the direct dimension, setting the imaginary component of the result to zero prior to FT in the indirect dimension, and finally retaining the real component (panel b), leading to a *double absorption* spectrum.

To achieve frequency discrimination, it is necessary to also possess the analogous sine-modulated signal, for which $\zeta = \sin(\cdot)$, as this effectively achieves quadrature detection in the indirect dimension. For numerous multidimensional experiments, this can be achieved by repeating a pulse se-

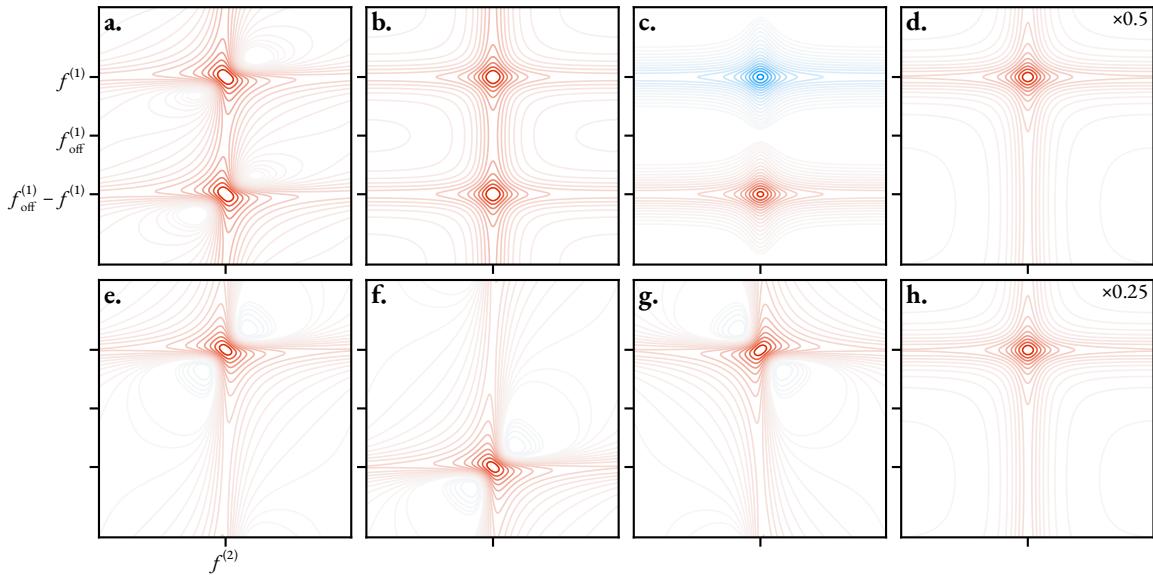


FIGURE 1.4: Spectra acquired from amplitude- and phase- modulated 2D signals. Red contour lines denote positive values, while blue contours denote negative values. **a.** The FT of a cosine-modulated FID, featuring peaks both at the true resonance frequency ($f^{(1)}$) and the mirrored frequency ($f_{\text{off}}^{(1)} - f^{(1)}$), and with phase-twist lineshapes. **b.** Double-absorption spectrum generated by applying FT in the direct dimension, setting the imaginary component to zero, applying FT in the indirect dimension, and retaining the real component. **c.** Spectrum acquired with the same processing method as in **b.** but with a sine-modulated FID, and with the imaginary component retained. **d.** The subtraction of the spectrum in **c.** from that in **b.** leads to a spectrum with frequency discrimination and a pure absorption lineshape. **e.** The FT of a positive phase-modulated FID, exhibiting frequency discrimination, but with a phase twist shape. **f.** The FT of a negative phase-modulated FID. **g.** The spectrum in **f.** inverted along the indirect axis, about $f_{\text{off}}^{(1)}$. **h.** The summation of **e.** and **g.** generates a spectrum with an absorption lineshape.

quence, with careful adjustments to the phases of particular pulses, in a process referred to as *phase cycling*[27: Chapter 11]. Applying the same processing as that which achieved the double absorption spectrum for the cosine-modulated case generates a spectrum whose imaginary component features two peaks, but with opposite signs (panel c) which is borne out of the sine function being odd. It then becomes possible to generate a frequency discriminated spectrum by subtracting the sine spectrum from the cosine spectrum (panel d).

Phase-modulated signals

A “positive” phase-modulated signal with the form $\zeta^{(1)} = \exp(i\cdot)$ (commonly referred to as *hypercomplex*) is frequency-discriminated due to its quadrature nature. However, direct FT of such a signal in both dimensions leads to a peaks with a phase-twist lineshape, with no means of separating the absorption and dispersion contributions (panel e). Certain pulse sequences, including 2DJ spectroscopy[34, 35] (see Section 3.3) and COSY[11–13] produce hypercomplex datasets, and the conventional means of processing is simply to display the absolute value of the spectrum,

which while removing the phase twist shapes, produces peaks with broad wings, due to the influence of the dispersive components. In scenarios where it is possible, it is desirable to acquire the equivalent “negative” signal ($\zeta^{(1)} = \exp(-i\cdot)$), whose FT leads to a peak with the same phase twist form, but centered at $f_{\text{off}}^{(1)} - f^{(1)}$ (panel f). Inverting this spectrum in $F^{(1)}$ (panel g), and summing with the positive spectrum nullifies the dispersive contributions, generating a spectrum with absorption lineshapes[36] (panel h).

1.2.3 Estimation Techniques for NMR Analysis

Linear prediction

Linear prediction (LP)[37, 38] is a procedure which is now widely used in NMR data analysis, often with the intention of (a) propagating the FID further in time in order to reduce the presence of truncation artefacts and (b) correct the commonly corrupted initial points of the FID, as a means of improving the spectral baseline. The concept of LP stems from the idea that a deterministic signal, such as a 1D FID can be described as an autoregressive (AR) process, such that a given sample from the dataset can be described by a linear combination of a certain number of previous samples:

$$\mathbf{y}[n^{(1)}] = \sum_{l=0}^{L-1} \mathbf{c}[l] \mathbf{y}[n^{(1)} - l - 1] + \mathbf{e}[l] \quad (1.28)$$

$\forall n^{(1)} \in \{L, L+1, \dots, N^{(1)} - 1\}$, where $L \in \mathbb{Z}$ defines the order of the linear estimator, and $\mathbf{c} \in \mathbb{R}^L$ is a set of *forward* LP coefficients. $\mathbf{e} \in \mathbb{R}^L$ is a set of parameters, sometimes called the innovations, which account of error in the LP model. A datapoint can also be described by a linear combination of a certain number of subsequent points, using the set of *backward* LP coefficients $\mathbf{b} \in \mathbb{R}^L$:

$$\mathbf{y}[n^{(1)}] = \sum_{l=0}^{L-1} \mathbf{b}[l] \mathbf{y}[n^{(1)} + l + 1] + \mathbf{e}[l] \quad (1.29)$$

$\forall n^{(1)} \in \{0, 1, \dots, N^{(1)} - L - 1\}$. Determining the LP coefficients enables the estimation of FID values beyond the data actually acquired ($n^{(1)} < 0$ and $n^{(1)} > N^{(1)} - 1$). It should be noted that (1.28) and (1.29) are only valid for FIDs without any corruption from experimental noise. Noisy datasets are instead an example of an autoregressive moving average (ARMA) process. Despite this, due to the greater simplicity of the AR model, it is far more common to employ this. The most common means of performing LP is by solving the Yule-Walker equations[39, 40], which describe the relationship between the signal autocorrelation coefficients and the LP coefficients[38: Section 3.3], with the Levinson-Durbin algorithm providing an efficient means of solving the equations[41, 42].

Stuff below here has been copy-pasted from old document. Will need editing!

1.2.4 Signal-Noise Subspace Separation Techniques

LPSVD

The Linear Prediction with Singular-Value Decomposition was proposed by Kumaresan and Tufts[43]. The core assumption of the method is that a given data point y_n may be expressed as a linear combination of L previous, or preceding points:

$$\begin{aligned} y_n &= b_1 y_{n+1} + b_2 y_{n+2} + \cdots + b_L y_{n+L} \\ &= b_{-1} y_{n-1} + b_{-2} y_{n-2} + \cdots + b_{-L} y_{n-L} \end{aligned} \quad (1.30)$$

where $b_l (b_{-l})$, $l \in \{1, 2, \dots, L\}$ are the backward(forward) prediction coefficients. For the backward prediction case, consider the following set of linear equations $\mathbf{A}\mathbf{b} = -\mathbf{b}$:

$$\begin{bmatrix} y_1^* & y_2^* & \cdots & y_L^* \\ y_2^* & y_3^* & \cdots & y_{L+1}^* \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-L}^* & y_{N-L+1}^* & \cdots & y_{N-1}^* \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_L \end{bmatrix} = - \begin{bmatrix} y_0^* \\ y_1^* \\ \vdots \\ y_{N-L-1}^* \end{bmatrix} \quad (1.31)$$

This equation can be augmented into the form $\mathbf{A}'\mathbf{b}' = \mathbf{0}$, with $\mathbf{A}' = [\mathbf{b} | \mathbf{A}]$ and $\mathbf{b}' = [1, \mathbf{b}^T]^T$. For the noiseless data case (i.e. $y_n = x_n$), any row of \mathbf{A}' can be written as the linear combination of J linearly independent vectors \mathbf{f}_j , $j \in \{1, 2, \dots, J\}$ [44]:

$$\mathbf{f}_j = \left[1, z_j^*, \left(z_j^2 \right)^*, \dots, \left(z_j^L \right)^* \right], \quad (1.32)$$

meaning that the rank of \mathbf{A}' is J , provided it possesses at least J rows ($J \leq N - L$). The fact that \mathbf{b}' lies in the null space of \mathbf{A}' implies that $\mathbf{f}_j \mathbf{b}' = 0$, such that the following holds:

$$1 + b_1 z_j^* + b_2 \left(z_j^2 \right)^* + \cdots + b_L \left(z_j^L \right)^* = 0 \quad (1.33)$$

It should be noted that L must satisfy $L \geq J$ to ensure the null space of \mathbf{A}' has at least a dimension of 1. Eqn. 1.33 leads to a means of deriving the signal poles z_j via polynomial $B(\zeta)$:

$$B(\zeta) = 1 + b_1 \zeta^{-1} + b_2 \zeta^{-2} + \cdots + b_L \zeta^{-L}, \quad (1.34)$$

the zeros of which will occur whenever $\zeta = \left(z_j^{-1} \right)^*$, $j \in \{1, 2, \dots, J\}$.

Of course, the prediction coefficient vector \mathbf{b} still needs to be determined. Eqn. 1.31 implies

that the optimal vector can be determined by simple linear least squares:

$$\hat{\mathbf{b}} = -\mathbf{A}^+ \mathbf{b}. \quad (1.35)$$

However, noise components of the signal will be incorporated into the solution of \mathbf{b} . To avoid this, \mathbf{A} is moved from the signal-noise subspace to the signal subspace via filtration using Singular-Valued Decomposition:

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^\dagger = \sum_{l=1}^R \sigma_l \mathbf{u}_l \mathbf{v}_l^\dagger \quad (1.36)$$

where $R =$

$\text{rank}(\mathbf{A}) = \min(L, N)$. The matrix $\tilde{\mathbf{A}}$ of rank J with the closest relationship to \mathbf{A} is a Frobenius norm sense is given by [Crozow 1988]:

$$\tilde{\mathbf{A}} = \sum_{l=1}^J \sigma_l \mathbf{u}_l \mathbf{v}_l^\dagger. \quad (1.37)$$

Using the following relationship between a matrix's pseudoinverse and its SVD:

$$\mathbf{A}^+ = \mathbf{V} \begin{bmatrix} \boldsymbol{\Sigma}^+ \\ \mathbf{0} \end{bmatrix} \mathbf{U}^\dagger, \quad (1.38)$$

The optimal vector of prediction coefficients is determined by:

$$\hat{\mathbf{b}} = -\tilde{\mathbf{A}}^+ \mathbf{b} = -\sum_{l=1}^J \sigma_l^{-1} \mathbf{v}_l [\mathbf{u}_l^\dagger \mathbf{b}] \quad (1.39)$$

HSVD

Matrix Pencil Method

1.2.5 Bayesian Methods

1.2.6 Iterative Schemes

VARPRO

The VARiable PROjection (VARPRO) method is an iterative approach to parameter estimation, developed by Golub and Pereyra [45], and applied to NMR signal analysis by van der Veen et

al.[46]. It relies on minimising the following quantity:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{x}(\boldsymbol{\theta})\|^2 \quad (1.40)$$

The vector \mathbf{x} can be expressed in the following format:

$$\mathbf{x} = \begin{bmatrix} z_1^0 & z_2^0 & \cdots & z_J^0 \\ z_1^1 & z_2^1 & \cdots & z_J^1 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_J^{N-1} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix} = \mathbf{Z}\boldsymbol{\alpha} \quad (1.41)$$

Using Eq. 1.41, Eq. 1.40 can be re-written as follows:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha}\|^2 \quad (1.42)$$

The complex amplitudes can be determined analytically as follows:

$$\hat{\boldsymbol{\alpha}} \approx (\mathbf{Z}^\dagger \mathbf{Z})^{-1} \mathbf{Z}^\dagger \mathbf{y} = \mathbf{Z}^+ \mathbf{y} \quad (1.43)$$

\mathbf{Z}^+ is the Moore-Penrose pseudoinverse of matrix \mathbf{Z} [47, 48]. The VARPRO method determines the frequencies and damping factors via an iterative minimisation of:

$$[\hat{\mathbf{f}}^T, \hat{\boldsymbol{\eta}}^T]^T = \arg \min_{[\mathbf{f}^T, \boldsymbol{\eta}^T]} \|\mathbf{y} - \mathbf{Z}\mathbf{Z}^+ \mathbf{y}\|^2 \quad (1.44)$$

which is achieved using the Levenberg–Marquardt algorithm[49, 50]. The amplitudes and phases are subsequently determined using Eq. 1.43. This means that the amplitudes and phases do not need initial guesses associated with them. Further specifications can be applied to the algorithm, reflecting the spectroscopist’s knowledge of the signal under inspection. For example, if it is known that a certain set of resonances constitute a multiplet, the relative frequency differences and amplitudes will be known.

AMARES

Advanced Method for Accurate, Robust, and Efficient Spectral fitting

THEORY

2

This chapter provides a detailed outline of an estimation routine which has been developed for the estimation of NMR data.

2.1 Outline of the Problem

For the purposes of this work, it is always assumed that an FID to be estimated is hypercomplex in form, meaning that it obeys (1.14) with $\zeta = \exp(i\cdot)$ $\forall d \in \{1, \dots, D\}$:

$$\mathbf{Y} \left[n^{(1)}, \dots, n^{(D)} \right] = \mathbf{X}(\boldsymbol{\theta}) \left[n^{(1)}, \dots, n^{(D)} \right] + \mathbf{W} \left[n^{(1)}, \dots, n^{(D)} \right] \quad (2.1a)$$

$$\mathbf{X}(\boldsymbol{\theta}) \left[n^{(1)}, \dots, n^{(D)} \right] = \sum_{m=0}^{M-1} \mathbf{a}[m] \exp(i\boldsymbol{\phi}[m]) \prod_{d=1}^D \exp \left(\left[2\pi i \left(\mathbf{f}^{(d)}[m] - f_{\text{off}}^{(d)} \right) - \boldsymbol{\gamma}^{(d)}[m] \right] n^{(d)} \Delta_t^{(d)} \right), \quad (2.1b)$$

$$\mathbf{W} \left[n^{(1)}, \dots, n^{(D)} \right] \sim \mathcal{N}_C(0, 2\sigma^2), \quad (2.1c)$$

where $\Delta_t^{(d)} = 1/f_{\text{sw}}^{(d)}$. Under this model, it is assumed that an FID consists of a summation of M damped complex sinusoids in the presence in AWGN. It is the goal of parametric estimation to establish the identity of all the quantities which describe the model component \mathbf{X} . These can be distilled into the vector $\boldsymbol{\theta} \in \mathbb{R}^{2(D+2)M}$:

Due to the assumed AWGN nature of the noise array, the pdf of an individual noise component $w := \mathbf{W} \left[n^{(1)}, \dots, n^{(D)} \right]$ is

$$p(w) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{|w|^2}{2\sigma^2} \right). \quad (2.2)$$

As the elements are independent and identically distributed, the joint pdf describing the entire

noise array is given by the product of each element's pdf:

$$\begin{aligned} p(\mathbf{W}) &= \prod_{n^{(1)}=0}^{N^{(1)}-1} \cdots \prod_{n^{(D)}=0}^{N^{(D)}-1} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|w|^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\mathfrak{N}}} \exp\left(-\frac{\|\mathbf{W}\|^2}{2\sigma^2}\right) \end{aligned} \quad (2.3)$$

As the noise array is the difference between the data and model, the likelihood function of θ given \mathbf{Y} , $\mathcal{L}(\theta|\mathbf{Y})$, is given by

$$\mathcal{L}(\theta|\mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{\mathfrak{N}}} \exp\left(-\frac{\|\mathbf{Y} - \mathbf{X}(\theta)\|^2}{2\sigma^2}\right). \quad (2.4)$$

It is common to consider instead the log-likelihood function of θ given \mathbf{Y} , $\ell(\theta|\mathbf{Y})$. As application of the logarithm is a monotonic transformation, the arguments of the maxima of \mathcal{L} and ℓ are equivalent.

$$\ell(\theta|\mathbf{Y}) = -\mathfrak{N} \ln(2\pi\sigma^2) - \frac{\|\mathbf{Y} - \mathbf{X}(\theta)\|^2}{2\sigma^2}. \quad (2.5)$$

Equation 2.5 implies that the optimal set of parameters $\theta^{(*)}$ is that which minimises the (square) norm of the difference between the data and model

$$\theta^{(*)} = \arg \max_{\theta \in \mathbb{R}^{2(D+2)M}} \ell(\theta|\mathbf{Y}) \equiv \arg \min_{\theta \in \mathbb{R}^{2(D+2)M}} \|\mathbf{Y} - \mathbf{X}(\theta)\|^2. \quad (2.6)$$

The application of NLP is a well-established approach to solve such a problem[51, 52]. The basic principle behind NLP is to iteratively explore - in a methodical way - how a function varies with its arguments, using information about the function and optionally its derivatives, and to terminate once conditions have been achieved which imply an extremum has been found. While derivative-free approaches to NLP do exist, (cite examples like simplex, simulated annealing, BOBYQA?) in scenarios where the function under consideration has well-defined, computationally tractable derivatives, the use of these can be valuable to solving optimisation problems. The problem outlined in Equation 2.6 is such an example.

2.2 Generating an Initial Guess: Matrix Pencil Method

In order for NLP to perform effectively, a large amount of *a priori* information is typically required, in the form of an initial guess $\theta^{(0)}$. Here, a description of one such approach to achieve this is presented blah blah blah...

2.2.1 1D Matrix Pencil Method

The matrix pencil method (MPM), developed by Hua and Sarkar[53–55], provides a route to extracting the signal poles of a 1D dataset, based on the assumption that the number of oscillators M is known.

Noiseless data

To motivate how it works, first consider a dataset which is devoid of noise, such that it is of the form $\mathbf{X}(\boldsymbol{\theta})$, given in Equation 2.1b, with $D = 1$:

$$\mathbf{X}(\boldsymbol{\theta}) \left[n^{(1)} \right] = \underbrace{\sum_{m=0}^{M-1} \boldsymbol{\alpha}[m] \exp(i\boldsymbol{\phi}[m])}_{\boldsymbol{\alpha}[m]} \underbrace{\exp \left([2\pi i (\mathbf{f}^{(1)}[m] - f_{\text{off}}) - \boldsymbol{\eta}^{(1)}[m]] n^{(1)} \Delta_t^{(1)} \right)}_{\boldsymbol{z}^{(1)}[m]^{n^{(1)}}}. \quad (2.7)$$

Consider the Hankel matrix $\mathbf{H}_X \in \mathbb{C}^{(N^{(1)} - L^{(1)}) \times (L^{(1)} + 1)}$:

$$\mathbf{H}_X = \begin{bmatrix} \mathbf{X}[0] & \mathbf{X}[1] & \dots & \mathbf{X}[L^{(1)}] \\ \mathbf{X}[1] & \mathbf{X}[2] & \dots & \mathbf{X}[L^{(1)} + 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}[N^{(1)} - L^{(1)} - 1] & \mathbf{X}[N^{(1)} - L^{(1)}] & \dots & \mathbf{X}[N^{(1)} - 1] \end{bmatrix}. \quad (2.8)$$

This matrix comprises windowed segments of the FID, with each row comprising the segment shifted to the right by one point relative to the row above. $L^{(1)} \in \mathbb{N}$ is the *pencil parameter*, which dictates the size of each window. Define the two matrices \mathbf{H}_{X1} and \mathbf{H}_{X2} , formed by the removal of the last or first column of \mathbf{H}_X , respectively:

$$\mathbf{H}_{X1} = \begin{bmatrix} \mathbf{X}[0] & \mathbf{X}[1] & \dots & \mathbf{X}[L^{(1)} - 1] \\ \mathbf{X}[1] & \mathbf{X}[2] & \dots & \mathbf{X}[L^{(1)}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}[N^{(1)} - L^{(1)} - 1] & \mathbf{X}[N^{(1)} - L^{(1)}] & \dots & \mathbf{X}[N^{(1)} - 2] \end{bmatrix}, \quad (2.9a)$$

$$\mathbf{H}_{X2} = \begin{bmatrix} \mathbf{X}[1] & \mathbf{X}[2] & \dots & \mathbf{X}[L^{(1)}] \\ \mathbf{X}[2] & \mathbf{X}[3] & \dots & \mathbf{X}[L^{(1)} + 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}[N^{(1)} - L^{(1)}] & \mathbf{X}[N^{(1)} - L^{(1)} + 1] & \dots & \mathbf{X}[N^{(1)} - 1] \end{bmatrix}. \quad (2.9b)$$

These matrices can be deconstructed into the following forms involving matrices containing the M signal poles and complex amplitudes that the data comprises:

$$\mathbf{H}_{X1} = \mathbf{Z}_L \mathbf{A} \mathbf{Z}_R, \quad (2.10a)$$

$$\mathbf{H}_{X2} = \mathbf{Z}_L \mathbf{A} \mathbf{Z}_D \mathbf{Z}_R, \quad (2.10b)$$

$$\mathbb{C}^{(N^{(1)} - L^{(1)}) \times M} \ni \mathbf{Z}_L = \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(1)} & \mathbf{z}^{(1)2} & \dots & \mathbf{z}^{(1)^{N^{(1)} - L^{(1)} - 1}} \end{bmatrix}^T, \quad (2.10c)$$

$$\mathbb{C}^{M \times L^{(1)}} \ni \mathbf{Z}_R = \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(1)} & \mathbf{z}^{(1)2} & \dots & \mathbf{z}^{(1)^{L^{(1)} - 1}} \end{bmatrix}, \quad (2.10d)$$

$$\mathbb{C}^{M \times M} \ni \mathbf{Z}_D = \text{diag}(\mathbf{z}^{(1)}), \quad (2.10e)$$

$$\mathbb{C}^{M \times M} \ni \mathbf{A} = \text{diag}(\boldsymbol{\alpha}). \quad (2.10f)$$

Description of what a matrix pencil is The matrix pencil $\mathbf{H}_{X2} - \lambda \mathbf{H}_{X1}$, $\lambda \in \mathbb{C}$ can therefore be expressed as

$$\mathbf{H}_{X2} - \lambda \mathbf{H}_{X1} = \mathbf{Z}_L \mathbf{A} (\mathbf{Z}_D - \lambda \mathbf{I}_M) \mathbf{Z}_R, \quad (2.11)$$

where $\mathbf{I}_M \in \mathbb{C}^{M \times M}$ is the identity matrix. Assuming that the following condition is met:

$$M \leq L^{(1)} \leq N^{(1)} - M, \quad (2.12)$$

the rank of the matrix pencil will be M . Equation 2.12 must be obeyed to ensure that both the number of rows and columns of the matrix pencil are at least M . Now consider the case when the scalar λ is equal to one of the signal poles i.e. $\lambda = \mathbf{z}[m] \forall m \in \{0, \dots, M-1\}$. The element $(\mathbf{Z}_D - \lambda \mathbf{I}_M)[m, m]$ will be set to 0, which will lead to the determinant of the matrix pencil being 0. The eigenvalues of the matrix pencil are the solution of the so-called *generalised eigenvalue problem*, and are defined as [56: Section 7.7]

$$\mathbf{z} = \{z \in \mathbb{C} : \det(\mathbf{H}_{X2} - z \mathbf{H}_{X1}) = 0\} \quad (2.13)$$

One means of finding the signal poles is by finding the eigenvalues of the matrix $\mathbf{H}_{X1}^+ \mathbf{H}_{X2}$. Deriving the corresponding complex amplitudes can then be achieved by solving the set of linear equations

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(1)} & \mathbf{z}^{(1)2} & \dots & \mathbf{z}^{(1)^{N^{(1)} - 1}} \end{bmatrix}^T \boldsymbol{\alpha} \equiv \mathbf{Z}^{(1)} \boldsymbol{\alpha} \implies \boldsymbol{\alpha} = \mathbf{Z}^+ \mathbf{X}. \quad (2.14)$$

Extraction of the amplitudes, phases, frequencies, and damping factors from the signal poles and complex amplitudes can then take place:

$$\boldsymbol{\alpha} = |\boldsymbol{\alpha}|, \quad (2.15a)$$

$$\phi = \arctan \left(\frac{\Im(\alpha)}{\Re(\alpha)} \right), \quad (2.15b)$$

$$f^{(1)} = \frac{f_{\text{sw}}^{(1)}}{2\pi} \Im(\ln z^{(1)}) + f_{\text{off}}, \quad (2.15c)$$

$$\eta^{(1)} = -f_{\text{sw}}^{(1)} \Re(\ln z^{(1)}). \quad (2.15d)$$

Noisy data

The presence of noise in the signal \mathbf{Y} complicates the process of determining the M signal poles, as \mathbf{H}_Y (\mathbf{H}_X 's equivalent with elements replaced by the noisy data) is likely to be full-rank ($\min(N^{(1)} - L^{(1)}, L^{(1)} + 1)$). To cope with this, it is necessary to generate a rank-reduced matrix $\tilde{\mathbf{H}}_Y$. By employing the Eckart-Young-Mirsky (EYM) theorem[56: Section 2.2], an appropriate matrix is can be obtained through singular value decomposition (SVD)**Appendix description of SVD:**

$$\tilde{\mathbf{H}}_Y = \mathbf{U}_M \boldsymbol{\Sigma}_M \mathbf{V}_M^\dagger, \quad (2.16a)$$

$$\mathbb{C}^{(N^{(1)} - L^{(1)}) \times M} \ni \mathbf{U}_M = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_M], \quad (2.16b)$$

$$\mathbb{C}^{(L^{(1)} + 1) \times M} \ni \mathbf{V}_M = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_M], \quad (2.16c)$$

$$\mathbb{C}^{M \times M} \ni \boldsymbol{\Sigma}_M = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_M). \quad (2.16d)$$

σ_m is the m^{th} largest singular value of \mathbf{H}_Y , $\mathbf{u}_m \in \mathbb{C}^{N^{(1)} - L^{(1)}}$ and $\mathbf{v}_m \in \mathbb{C}^{L^{(1)} + 1}$ are the corresponding left and right singular vectors, respectively. The EYM proves that $\tilde{\mathbf{H}}_Y$ is the closest matrix of rank M to \mathbf{H}_Y in a Frobenius norm sense, i.e.

$$\tilde{\mathbf{H}}_Y = \underset{\mathbf{A}: \text{rank}(\mathbf{A})=M}{\arg \min} \|\mathbf{A} - \mathbf{H}_Y\| \quad (2.17)$$

With a rank-reduced matrix produced from the noisy matrix, the signal poles can then be derived from the eigenvalues of $\tilde{\mathbf{H}}_{Y1}^+ \tilde{\mathbf{H}}_{Y2}$, where $\tilde{\mathbf{H}}_{Y1}$ and $\tilde{\mathbf{H}}_{Y2}$ have the same relation to $\tilde{\mathbf{H}}_Y$ as \mathbf{H}_{X1} and \mathbf{H}_{X2} do to \mathbf{H}_X . As a less expensive alternative, the same result can be achieved by computing the eigenvalues of $\mathbf{V}_{M1}^+ \mathbf{V}_{M2}$, with

$$\mathbf{V}_{M1} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_{M-1}], \quad (2.18a)$$

$$\mathbf{V}_{M2} = [\mathbf{v}_2 \ \mathbf{v}_3 \ \dots \ \mathbf{v}_M]. \quad (2.18b)$$

Algorithm 1 provides a pseudo-code description of the MPM as implemented in the NMR-EsPy package.

Discuss complexity of MPM, talking about reducing $N^{(1)}$ and $L^{(1)}$ as a means of reducing the cost.

Algorithm 1 The matrix pencil method.

```

1: procedure MPM( $\mathbf{Y} \in \mathbb{C}^{N^{(1)}}, M \in \mathbb{N}$ )
2:    $L \leftarrow \lfloor N^{(1)}/3 \rfloor;$ 
3:    $\mathbf{H}_Y \leftarrow \begin{bmatrix} \mathbf{Y}[0] & \mathbf{Y}[1] & \cdots & \mathbf{Y}[L^{(1)}] \\ \mathbf{Y}[1] & \mathbf{Y}[2] & \cdots & \mathbf{Y}[L^{(1)} + 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}[N^{(1)} - L^{(1)} - 1] & \mathbf{Y}[N^{(1)} - L^{(1)}] & \cdots & \mathbf{Y}[N^{(1)} - 1] \end{bmatrix};$ 
4:    $\mathbf{U}, \sigma, \mathbf{V}^\dagger \leftarrow \text{SVD}(\mathbf{H}_Y);$ 
5:    $\mathbf{V} \leftarrow [\mathbf{V}^\dagger]^\dagger;$ 
6:    $\mathbf{V}_M \leftarrow \mathbf{V}[:, :M];$  ▷ Retain first  $M$  right singular vectors.
7:    $\mathbf{V}_{M1}, \mathbf{V}_{M2} \leftarrow \mathbf{V}_M[:, :M - 1], \mathbf{V}_M[:, 1:];$  ▷ Remove last/first column
8:    $\mathbf{z}^{(1)} \leftarrow \text{EIGENVALUES}(\mathbf{V}_{M1}^\dagger \mathbf{V}_{M2});$ 
9:    $\mathbf{Z}^{(1)} \leftarrow \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(1)} & \mathbf{z}^{(1)2} & \cdots & \mathbf{z}^{(1)N^{(1)}} \end{bmatrix}^T;$ 
10:   $\alpha \leftarrow \mathbf{Z}^{(1)\dagger} \mathbf{Y};$ 
11:   $\alpha, \phi \leftarrow |\alpha|, \arctan\left(\frac{\Im(\alpha)}{\Re(\alpha)}\right);$ 
12:   $\mathbf{f}^{(1)} \leftarrow \frac{f_{sw}^{(1)}}{2\pi} \Im(\ln \mathbf{z}^{(1)}) + f_{off};$ 
13:   $\eta^{(1)} \leftarrow -f_{sw}^{(1)} \Re(\ln \mathbf{z}^{(1)});$ 
14:  if  $\eta^{(1)}$  contains negative values then ▷ Purge any oscillators with negative damping
15:    Remove these from  $\eta^{(1)}$ , and remove the corresponding values from  $\alpha, \phi$ , and  $\mathbf{f}^{(1)}$ ;
16:  end if
17:   $\theta^{(0)} \leftarrow [\alpha^T \quad \phi^T \quad [\mathbf{f}^{(1)}]^T \quad [\eta^{(1)}]^T]^T;$ 
18:  return  $\theta^{(0)};$ 
19: end procedure

```

2.2.2 2D Matrix Enhancement and Matrix Pencil Method

The MPM was extended for the consideration of 2D data by Hua with the matrix enhancement and matrix pencil method (MEMPM)[57]. The method centers around the enhanced matrix $\mathbf{E}_Y \in \mathbb{C}^{(L^{(1)}L^{(2)}) \times (N^{(1)} - L^{(1)} + 1)(N^{(2)} - L^{(2)} + 1)}$, a block Hankel matrix of the form

$$\mathbf{E}_Y = \begin{bmatrix} \mathbf{H}_{Y,0} & \mathbf{H}_{Y,1} & \cdots & \mathbf{H}_{Y,N^{(1)}-L^{(1)}} \\ \mathbf{H}_{Y,1} & \mathbf{H}_{Y,2} & \cdots & \mathbf{H}_{Y,N^{(1)}-L^{(1)}+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Y,L^{(1)}-1} & \mathbf{H}_{Y,L^{(1)}} & \cdots & \mathbf{H}_{Y,N^{(1)}-1} \end{bmatrix}, \quad (2.19a)$$

$$\mathbf{H}_{Y,n^{(1)}} = \begin{bmatrix} \mathbf{Y}[n^{(1)}, 0] & \mathbf{Y}[n^{(1)}, 1] & \cdots & \mathbf{Y}[n^{(1)}, N^{(2)} - L^{(2)}] \\ \mathbf{Y}[n^{(1)}, 1] & \mathbf{Y}[n^{(1)}, 2] & \cdots & \mathbf{Y}[n^{(1)}, N^{(2)} - L^{(2)} + 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}[n^{(1)}, L^{(2)} - 1] & \mathbf{Y}[n^{(1)}, L^{(2)}] & \cdots & \mathbf{Y}[n^{(1)}, N^{(2)} - 1] \end{bmatrix}. \quad (2.19b)$$

In a similar fashion to Equation 2.10, $\mathbf{H}_{Y,n^{(1)}}$ can be expressed as

$$\mathbf{H}_{Y,n^{(1)}} = \mathbf{Z}_L^{(2)} \mathbf{A} \left[\mathbf{Z}_D^{(1)} \right]^{n^{(1)}} \mathbf{Z}_R^{(2)}, \quad (2.20a)$$

$$\mathbf{Z}_L^{(2)} = \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(2)} & \mathbf{z}^{(2)2} & \cdots & \mathbf{z}^{(2)L^{(2)}-1} \end{bmatrix}^T, \quad (2.20b)$$

$$\mathbf{Z}_R^{(2)} = \begin{bmatrix} \mathbf{1} & \mathbf{z}^{(2)} & \mathbf{z}^{(2)2} & \cdots & \mathbf{z}^{(2)N^{(2)}-L^{(2)}} \end{bmatrix}, \quad (2.20c)$$

with $\mathbf{Z}_D^{(1)}$ and \mathbf{A} given by Equations 2.10e & 2.10f, respectively. This then leads to the enhanced matrix be expressed as

$$\mathbf{E}_Y = \mathbf{E}_L \mathbf{A} \mathbf{E}_R, \quad (2.21a)$$

$$\mathbb{C}^{L^{(1)}L^{(2)} \times M} \ni \mathbf{E}_L = \begin{bmatrix} \mathbf{Z}_L^{(2)} \\ \mathbf{Z}_L^{(2)} \mathbf{Z}_D^{(1)} \\ \vdots \\ \mathbf{Z}_L^{(2)} [\mathbf{Z}_D^{(1)}]^{L^{(1)}-1} \end{bmatrix}, \quad (2.21b)$$

$$\mathbb{C}^{M \times (N^{(1)} - L^{(1)} + 1)(N^{(2)} - L^{(2)} + 1)} \ni \mathbf{E}_R = \begin{bmatrix} \mathbf{Z}_R^{(2)} & \mathbf{Z}_D^{(1)} \mathbf{Z}_R^{(2)} & \cdots & [\mathbf{Z}_D^{(1)}]^{N^{(1)} - L^{(1)}} \mathbf{Z}_L^{(2)} \end{bmatrix}. \quad (2.21c)$$

As was the case in the 1D MPM, SVD can be utilised to generate a filtered matrix $\tilde{\mathbf{E}}_Y$ with its rank reduced to M :

$$\tilde{\mathbf{E}}_Y = \mathbf{U}_M \boldsymbol{\Sigma}_M \mathbf{V}_M^\dagger \quad (2.22)$$

If the conditions $N^{(d)} - L^{(d)} + 1 \geq M \forall d \in \{1, 2\}$ are met, $\text{range}(\mathbf{U}_M) = \text{range}(\mathbf{E}_L)$. This implies that there is some nonsingular matrix $\mathbf{T} \in \mathbb{C}^{M \times M}$ such that

$$\mathbf{U}_M = \mathbf{E}_L \mathbf{T}. \quad (2.23)$$

Now consider the following two matrices:

$$\mathbf{U}_{M1} = \mathbf{E}_{L1} \mathbf{T}, \quad (2.24a)$$

$$\mathbf{U}_{M2} = \mathbf{E}_{L1} \mathbf{Z}_D^{(1)} \mathbf{T}, \quad (2.24b)$$

$$\mathbb{C}^{L^{(1)}(L^{(2)}-1) \times M} \ni \mathbf{E}_{L1} = \begin{bmatrix} \mathbf{Z}_L^{(2)} \\ \mathbf{Z}_L^{(2)} \mathbf{Z}_D^{(1)} \\ \vdots \\ \mathbf{Z}_L^{(2)} [\mathbf{Z}_D^{(1)}]^{L^{(1)}-2} \end{bmatrix} \text{(c.f. Equation 2.21b).} \quad (2.24c)$$

\mathbf{U}_{M1} and \mathbf{U}_{M2} correspond to the \mathbf{U}_M with the last and first $L^{(2)}$ rows removed, respectively. The matrix pencil for \mathbf{U}_{M1} and \mathbf{U}_{M2} can be expressed as

$$\mathbf{U}_{M1} - \lambda \mathbf{U}_{M2} = \mathbf{E}_{L1} (\mathbf{Z}_D^{(1)} - \lambda \mathbf{I}_M) \mathbf{T}. \quad (2.25)$$

As seen previously, this matrix structure implies that $\mathbf{z}^{(1)}$ are the solutions to the generalised eigenvalue problem, such that they are the eigenvalues of $\mathbf{U}_{M1}^+ \mathbf{U}_{M2}$.

To extract the signal poles in the other dimension, $\mathbf{z}^{(2)}$, the permutation matrix is defined:

$$\mathbb{R}^{L^{(1)}L^{(2)} \times L^{(1)}L^{(2)}} \ni \mathbf{P} = \begin{bmatrix} \mathbf{e}(0)^T \\ \mathbf{e}(L^{(2)})^T \\ \vdots \\ \mathbf{e}((L^{(1)} - 1)L^{(2)})^T \\ \mathbf{e}(1)^T \\ \mathbf{e}(1 + L^{(2)})^T \\ \vdots \\ \mathbf{e}(1 + (L^{(1)} - 1)L^{(2)})^T \\ \vdots \\ \vdots \\ \mathbf{e}(L^{(2)} - 1)^T \\ \mathbf{e}(2L^{(2)} - 1)^T \\ \vdots \\ \mathbf{e}(L^{(1)}L^{(2)} - 1)^T \end{bmatrix}. \quad (2.26)$$

$\mathbf{e}(i) \in \mathbb{R}^{L^{(1)}L^{(2)}}$ corresponds to a unit vector comprising zeros except for $\mathbf{e}[i] = 1$. Multiplying \mathbf{E}_L by the permutation matrix leads to a matrix in which the roles of the two sets of signal poles are effectively swapped:

$$\mathbf{E}_{LP} := \mathbf{P}\mathbf{E}_L = \begin{bmatrix} \mathbf{Z}_L^{(1)} \\ \mathbf{Z}_L^{(1)}\mathbf{Z}_D^{(2)} \\ \vdots \\ \mathbf{Z}_L^{(1)}[\mathbf{Z}_D^{(2)}]^{L^{(2)}-1} \end{bmatrix}, \quad (2.27a)$$

$$\mathbf{Z}_L^{(1)} = [\mathbf{1} \ \mathbf{z}^{(1)} \ \mathbf{z}^{(1)2} \ \dots \ \mathbf{z}^{(1)L^{(1)}-1}]^T, \quad (2.27b)$$

$$\mathbf{Z}_D^{(2)} = \text{diag}(\mathbf{z}^{(2)}). \quad (2.27c)$$

Note the similarity of Equation 2.27a with Equation 2.21b, which implies that with the same reasoning as given above, $\mathbf{z}^{(2)}$ can be derived by extracting the eigenvalues of $\mathbf{U}_{MP1}^+ \mathbf{U}_{MP2}$, where \mathbf{U}_{MP1} and \mathbf{U}_{MP2} correspond to $\mathbf{P}\mathbf{U}_M$ with the last and first $L^{(1)}$ rows removed, respectively.

In the original account on the MEMPM, the final stage involved employing a pairing algorithm in order to assign the uncorrelated signal poles in $\mathbf{z}^{(1)}$ with $\mathbf{z}^{(2)}$ [57]. The modified matrix enhancement and matrix pencil method (MMEMPM) was developed in order to overcome two

issues with the pairing algorithm: (a) it is computationally expensive (b) it is fallible to return incorrect pairings[58]. As well as the generalised eigenvalues of $\mathbf{U}_{M1} - \lambda \mathbf{U}_{M2} (\mathbf{z}^{(1)})$, the MMEMPM requires extraction of the generalised eigenvectors too ($\mathbf{W}^{(1)}$). Assuming that there are no repeated poles in $\mathbf{z}^{(1)}$, the correctly paired second set of poles is then generated via

$$\mathbf{z}^{(2)} = \text{diag}(\mathbf{W}^{-1} \mathbf{U}_{MP1}^+ \mathbf{U}_{MP2} \mathbf{W}) \quad (2.28)$$

Talk about case of paired eigenvalues. See Algorithm A1

2.2.3 Model Order Selection

The MPM and MMEMPM require a specification of the model order M . There are various non-subjective criteria which have been established for estimating the model order of a given signal, with probably the two most prominent being the Akaike information criterion (AIC)[59] and minimum description length (MDL)[60, 61]. Both of these criteria consider a family of potential models which could describe a given set of observations, parametrised by the vector θ . For the purpose of FID estimation, the family of potential models comprise (2.1b), with variable M . Both the AIC and MDL take the same general form:

$$\mathcal{C}(k) = -c \ln \left(\mathcal{L}(\hat{\theta} | \mathbf{Y}) \right) + \mathcal{P}(k) \text{ with } \hat{\theta} \in \mathbb{R}^{4k}, \quad (2.29)$$

$\forall k \in \{0, 1, \dots\}$. $\mathcal{L}(\hat{\theta} | \mathbf{Y})$ is the likelihood function of a given model, with model order k , at the maximum likelihood estimate (MLE) $\hat{\theta}$ **define in stats appendix**, $c \in \mathbb{R}$ is a scaling constant, and \mathcal{P} is a penalising function, which acts to correct for bias. As the model order increases, the likelihood function at the MLE will increase in size, as a model with more parameters will be able to fit a given dataset more accurately. However, as the model order increases, there will become a point where practically all of the deterministic part of the signal has been incorporated into the model, and increasing the model order further leads to the model also accounting for noise. The penalising term, which is larger for higher k , is required in order to estimate a model order which is parsimonious. Wax and Kailath derived an expression for the likelihood at the MLE for models comprising a summation of complex sinusoids[62]:

$$\mathcal{L}(\hat{\theta} | \mathbf{Y}) = \left(\frac{\prod_{r=k}^{L^{(1)}-1} \sigma[r]^{1/L^{(1)}-k}}{\frac{1}{L^{(1)}-k} \sum_{r=k}^{L^{(1)}-1} \sigma[r]} \right)^{(L^{(1)}-k)N^{(1)}}, \quad (2.30)$$

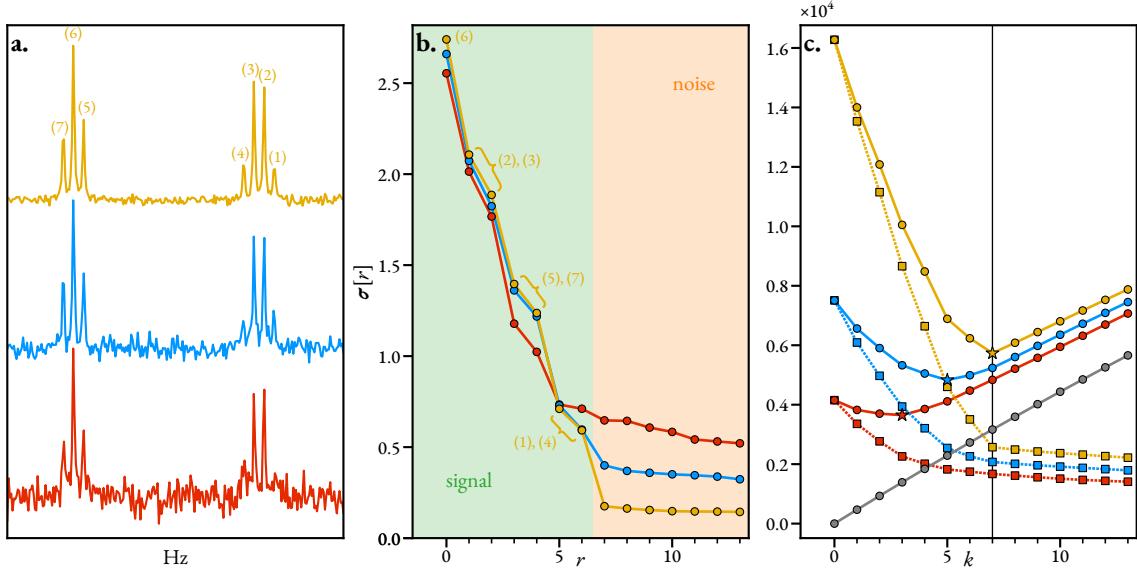


FIGURE 2.1: An visualisation of the behaviour of the MDL for three different FIDs comprising the same model, but with different noise variances. The model used to construct the FIDs feature 7 oscillators. The three SNRs used were 7 dB (red), 12 dB (blue), and 20 dB (yellow). The FIDs were generated with $N^{(1)} = 256$. **a.** Spectra of the three FIDs. **b.** The values of the 14 most significant singular values associated with the Hankel matrix \mathbf{H}_Y , with the pencil parameter $L^{(1)}$ set to $\lfloor N^{(1)}/3 \rfloor = 85$. **c.** Square points with dotted lines: The negative log-likelihood MLE, i.e. the first term of (2.31b). Grey line: the penalty component of the MDL, given by the second term in (2.31b). Circular points with solid lines: the MDL. Stars denote the minimum of the MDL for a given FID. The 20 dB signal is correctly deemed to have a model order of 7, while the other two are underestimated (predicted models orders are 5 and 3 for the 12 dB and 7 dB FIDs, respectively).

$\forall k \in \{0, 1, \dots, L^{(1)} - 1\}$. $\sigma \in \mathbb{R}^{L^{(1)}}$ is the set of singular values of \mathbf{H}_Y , in decreasing order. The forms of the AIC and MDL are given by

$$\text{AIC}(k) = -2 \ln \left(\mathcal{L}(\hat{\theta} | \mathbf{Y}) \right) + 2k(2L^{(1)} - k), \quad (2.31a)$$

$$\text{MDL}(k) = -\ln \left(\mathcal{L}(\hat{\theta} | \mathbf{Y}) \right) + \frac{1}{2}k(2L^{(1)} - k) \ln N^{(1)}. \quad (2.31b)$$

The AIC is to be inconsistent in that it tends to overestimate the model order as the number of samples increases[62]. For this reason, the MDL has found greater favour in signal processing applications. The estimated model order is then given by

$$M = \arg \min_{k \in \mathbb{N}_0: k < L^{(1)}} \text{MDL}(k). \quad (2.32)$$

Figure 2.1 illustrates the form of the MDL for three signals with equivalent underlying models, with $M = 7$, and noise instances with different variances. The first 14 singular values of \mathbf{H}_Y are plotted in panel b, where it can be seen that beyond the first 7, after singular values accounting for signal components are accounted for, the following singular values, associated with noise, decrease

at a far slower rate. The noise subspace for FIDs with higher SNRs have singular values which are (a) smaller in magnitude and (b) more consistent, such that distinguishing the noise and signal subspaces is an easier task (cf. the yellow line in panel b and the red line). As such, the MDL is more likely to provide a faithful estimate of the true number of components in the FID (panel c).

Mention that multidimensional criteria are available, though not implemented in this work.

2.3 Non-linear programming for NMR estimation

Before discussing the details of the specific NLP problem outlined above, an general overview is given.

2.3.1 An overview of non-linear programming

In an optimisation problem, the goal is to determine the minimum (or maximum) of a function $\mathcal{F}(\theta) : \mathbb{R}^n \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, of called the *cost function* or *fidelity*. This is typically with the goal of determining the argument $\theta^{(*)}$ at which the optimum is found:

$$\theta^{(*)} = \arg \min_{\theta \in \mathbb{R}^n} \mathcal{F}(\theta). \quad (2.33)$$

The above problem is unconstrained, as there are no limitations that the parameter vector is subjected to. Unless $\mathcal{F}(\theta)$ has particular properties, such as convexity*, it is generally only possible to determine a local minimum, rather than a global minimum.

Definition 1. $\theta^{(*)} \in \mathbb{R}^n$ is a local minimiser of $\mathcal{F}(\theta)$ if there is a neighbourhood $N \ni \theta^{(*)}$ such that $\mathcal{F}(\theta^{(*)}) \leq \mathcal{F}(\theta) \forall \theta \in N$

It can be shown that $\theta^{(*)}$ is a local minimiser of $\mathcal{F}(\theta)$ if the first- and second-order necessary conditions are satisfied. Furthermore, it is a unique (strict) local minimiser if the second-order sufficient condition is also satisfied.

Theorem 1 (First-order necessary condition). If $\mathcal{F}(\theta)$ is continuously differentiable, then the gradient vector $\mathbf{g}(\theta^{(*)}) := \nabla \mathcal{F}(\theta^{(*)})$ is the zero vector:

$$\mathbf{g}(\theta^{(*)}) = \mathbf{0} \in \mathbb{R}^n \quad (2.34)$$

Theorem 2 (Second-order necessary condition). If $\mathcal{F}(\theta)$ and $\mathbf{g}(\theta)$ are continuously differentiable, and $\theta^{(*)}$ is a local minimiser of $\mathcal{F}(\theta)$, then the Hessian $\mathbf{H}(\theta^{(*)}) := \nabla^2 \mathcal{F}(\theta^{(*)})$ is positive

*A convex function is one such that a line segment through any two points of the function lies above it.

semidefinite, i.e.

$$\mathbf{v}^T \mathbf{H}(\boldsymbol{\theta}^{(*)}) \mathbf{v} \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n. \quad (2.35)$$

Theorem 3 (Second-order sufficient condition). $\boldsymbol{\theta}^{(*)}$ is a strict local minimiser of $\mathcal{F}(\boldsymbol{\theta})$ if $\mathbf{H}(\boldsymbol{\theta}^{(*)})$ is positive definite:

$$\mathbf{v}^T \mathbf{H}(\boldsymbol{\theta}^{(*)}) \mathbf{v} > 0 \quad \forall \mathbf{v} \in \mathbb{R}^n. \quad (2.36)$$

A plethora of approaches have been established to determine local minima of scalar functions. One of the better-known strategies is *Newton's method*, in which a quadratic approximation of the fidelity is considered. For a given iteration $k \in \mathbb{N}_0$, the fidelity is approximated using

$$\mathcal{F}_Q(\boldsymbol{\theta}) = \mathcal{F}(\boldsymbol{\theta}^{(k)}) + \mathbf{b}^T \mathbf{g}(\boldsymbol{\theta}^{(k)}) + \frac{1}{2} \mathbf{b}^T \mathbf{H}(\boldsymbol{\theta}^{(k)}) \mathbf{b}, \quad (2.37)$$

where $\mathbf{b} = \boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}$. An updated prediction of the parameter vector is derived by finding the minimum of this quadratic approximation:

$$\begin{aligned} \frac{\partial \mathcal{F}(\boldsymbol{\theta})}{\partial \mathbf{b}} &= \mathbf{g}(\boldsymbol{\theta}^{(k)}) + \mathbf{H}(\boldsymbol{\theta}^{(k)}) \mathbf{b} \\ \implies 0 &= \mathbf{g}(\boldsymbol{\theta}^{(k)}) + \mathbf{H}(\boldsymbol{\theta}^{(k)}) (\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}) \\ \implies \boldsymbol{\theta}^{(k+1)} &= \boldsymbol{\theta}^{(k)} - [\mathbf{H}(\boldsymbol{\theta}^{(k)})]^{-1} \mathbf{g}(\boldsymbol{\theta}^{(k)}). \end{aligned} \quad (2.38)$$

As Theorem 1 indicates, a local minimum of $\mathcal{F}(\boldsymbol{\theta})$ is deemed to be reached when the magnitude of the gradient vector is smaller than a particular threshold $\epsilon > 0$, implying that it is close to the zero vector:

$$\|\mathbf{g}(\boldsymbol{\theta}^{(k)})\| \leq \epsilon. \quad (2.39)$$

In practice, Equation 2.38 tends not to be used as the update formula, with certain unattractive features, including the commonly expensive requirement to compute the inverse of the Hessian, and the possibility of the Newton update not being a minimising update if the Hessian is not positive definite. Two primary strategies have emerged for numerical optimisation:

- ❖ *Line search methods*[52: Chapter 3] determine an appropriate direction $\mathbf{p}^{(k)}$ along which the updated parameter vector is sourced. After this, an appropriate step length $\alpha^{(k)}$ is determined (typically in an efficient, though not optimal manner), leading to $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \alpha^{(k)} \mathbf{p}^{(k)}$
- ❖ *Trust region methods*[52: Chapter 4] define a radius $\Delta^{(k)} > 0$, and determine the minimum of Equation 2.37 subject to the constraint that $\|\mathbf{b}\| \leq \Delta^{(k)}$. This radius is typically programmed to be variable, such that if the quadratic model agrees well with the true value of the fidelity, it is increased for the next iteration, while the converse is true if there is poor agreement.

A trust region method is applied in this work, and as such further consideration of it will now be made.

Trust Region Methods

The structure of a typical trust region method is presented in Algorithm 2 (ignoring lines 19–22, which is a custom addition, see Section 2.3.4). An initial trust radius for the region $\Delta^{(0)}$ is defined, along with a maximum permitted trust radius Δ_{\max} , to ensure that excessively adventurous steps do not take place. For each iteration, a solution to the following sub-problem is sought:

$$\begin{aligned} \mathbf{p}^{(k)} = \arg \min_{\mathbf{p} \in \mathbb{R}^n} \mathcal{F}(\boldsymbol{\theta}^{(k)}) + \mathbf{p}^T \mathbf{g}(\boldsymbol{\theta}^{(k)}) + \frac{1}{2} \mathbf{p}^T \mathbf{H}(\boldsymbol{\theta}^{(k)}) \mathbf{p} \\ \text{subject to } \|\mathbf{p}\| \leq \Delta^{(k)}. \end{aligned} \quad (2.40)$$

This sub-problem is not usually minimised exactly, but instead an efficient means of determining a sufficiently good value for $\mathbf{p}^{(k)}$ is used. Common approaches include computing the Cauchy point and the Dogleg method, and a truncated conjugate-gradient approach, also known as the Steihaug-Toint (ST) method[52: Chapter 7], which is employed in this work (Algorithm A2). In the ST approach, iterates of the conjugate-gradient method are used, until either they go beyond the trust region, or negative curvature is discovered.

Once a step $\mathbf{p}^{(k)}$ is determined, a metric is considered which indicates how effectively the quadratic estimate at the proposed update $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \mathbf{p}^{(k)}$ agrees with the true value of the fidelity at this point:

$$\rho^{(k)} = \frac{\mathcal{F}(\boldsymbol{\theta}^{(k)}) - \mathcal{F}(\boldsymbol{\theta}^{(k)} + \mathbf{p}^{(k)})}{\mathcal{F}_Q(\boldsymbol{\theta}^{(k)}) - \mathcal{F}_Q(\boldsymbol{\theta}^{(k)} + \mathbf{p}^{(k)})}. \quad (2.41)$$

$\rho^{(k)}$ is the ratio between the actual reduction of the fidelity caused by taking the proposed step, and the predicted reduction based on the quadratic model. If $\rho^{(k)}$ is sufficiently close to 1, the quadratic model being used to generate new iterates is deemed to be acting well enough to warrant accepting the proposed update (lines 14-15). Furthermore, if $\rho^{(k)}$ is *particularly* close to 1, and the proposed update is at the boundary of the trust radius, it is appropriate to enlarge the radius of the trust region for the next iteration in an attempt to increase the rate of convergence (lines 9-10). On the other hand, a small value of $\rho^{(k)}$ implies that the quadratic model reflects the true fidelity poorly, such that the proposed update should be rejected (lines 16-17). As well as this, the trust region's radius should be decreased such that the model is more likely to behave faithfully (lines 7-8). The exact thresholds which dictate whether to accept an update, and whether to adjust the trust region radius are customisable. The hard-coded numerical values found in Algorithm 2 are the values used in this work, and the defaults in NMR-EsPy.

2.3.2 Non-linear programming applied to FID estimation

Returning to the problem of FID estimation, the fidelity $\mathcal{F}(\boldsymbol{\theta} | \mathbf{Y}) : \mathbb{C}^{N^{(1)} \times \dots \times N^{(D)}} \times \mathbb{R}^{2(1+D)M} \rightarrow \mathbb{R}$ is given by

$$\mathcal{F}(\boldsymbol{\theta} | \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X}(\boldsymbol{\theta})\|^2. \quad (2.42)$$

Remark 1. Prior to estimating the dataset, it is normalised, such that the signal actually under consideration is $\mathbf{Y}/\|\mathbf{Y}\|$. To make the result reflect the actual dataset, the final amplitudes $\boldsymbol{a}^{(*)}$ are multiplied by $\|\mathbf{Y}\|$.

The elements of the gradient vector $\mathbf{g}(\boldsymbol{\theta} | \mathbf{Y}) \in \mathbb{R}^{2(1+D)M}$ and the Hessian matrix $\mathbf{H}(\boldsymbol{\theta} | \mathbf{Y}) \in \mathbb{R}^{2(1+D)M \times 2(1+D)M}$ are then derived by taking the first and second partial derivatives of the fidelity with respect to the elements in $\boldsymbol{\theta}$, respectively:

$$\mathbf{g}[i] = -2\Re \left((\mathbf{Y} - \mathbf{X}), \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[i]} \right), \quad (2.43a)$$

$$\mathbf{H}[i,j] = 2\Re \left(\underbrace{\left(\frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[i]}, \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[j]} \right)}_{(1)} - \underbrace{\left((\mathbf{Y} - \mathbf{X}), \frac{\partial^2 \mathbf{X}}{\partial \boldsymbol{\theta}[i] \partial \boldsymbol{\theta}[j]} \right)}_{(2)} \right). \quad (2.43b)$$

$\forall i, j \in \{0, \dots, 2(1+D)M-1\}$. The complete set of first derivatives of a particular element of the model $x := \mathbf{X} \left[n^{(1)}, \dots, n^{(D)} \right]$ (see (2.1b)) is as follows $\forall m \in \{0, \dots, M-1\}$:

$$\frac{\partial x}{\partial \boldsymbol{\theta}[m]} \equiv \frac{\partial x}{\partial \boldsymbol{a}[m]} = \frac{x}{\boldsymbol{a}[m]}, \quad (2.44a)$$

$$\frac{\partial x}{\partial \boldsymbol{\theta}[m+M]} \equiv \frac{\partial x}{\partial \boldsymbol{\phi}[m]} = ix, \quad (2.44b)$$

$$\frac{\partial x}{\partial \boldsymbol{\theta}[m+(d+1)M]} \equiv \frac{\partial x}{\partial \mathbf{f}^{(d)}[m]} = 2\pi i \Delta_t^{(d)} n^{(d)} x, \quad (2.44c)$$

$$\frac{\partial x}{\partial \boldsymbol{\theta}[m+(d+D+1)M]} \equiv \frac{\partial x}{\partial \boldsymbol{\eta}^{(d)}[m]} = -\Delta_t^{(d)} n^{(d)} x. \quad (2.44d)$$

The complete set of the second derivatives is given by

$$\frac{\partial^2 x}{\partial \boldsymbol{\theta}[m]^2} \equiv \frac{\partial^2 x}{\partial \boldsymbol{a}[m]^2} = 0, \quad (2.45a)$$

$$\frac{\partial^2 x}{\partial \boldsymbol{\theta}[m] \partial \boldsymbol{\theta}[m+M]} \equiv \frac{\partial^2 x}{\partial \boldsymbol{a}[m] \partial \boldsymbol{\phi}[m]} = \frac{ix}{\boldsymbol{a}[m]}, \quad (2.45b)$$

$$\frac{\partial^2 x}{\partial \boldsymbol{\theta}[m] \partial \boldsymbol{\theta}[m+(d+1)M]} \equiv \frac{\partial^2 x}{\partial \boldsymbol{a}[m] \partial \mathbf{f}^{(d)}[m]} = \frac{2\pi i \Delta_t^{(d)} n^{(d)} x}{\boldsymbol{a}[m]}, \quad (2.45c)$$

$$\frac{\partial^2 x}{\partial \boldsymbol{\theta}[m] \partial \boldsymbol{\theta}[m+(d+D+1)M]} \equiv \frac{\partial^2 x}{\partial \boldsymbol{a}[m] \partial \boldsymbol{\eta}^{(d)}[m]} = \frac{-\Delta_t^{(d)} n^{(d)} x}{\boldsymbol{a}[m]}, \quad (2.45d)$$

dimensions	# 1 st derivatives	# 2 nd derivatives
1	$4MN^{(1)}$	$9MN^{(1)}$
2	$6MN^{(1)}N^{(2)}$	$20MN^{(1)}N^{(2)}$
3	$8MN^{(1)}N^{(2)}N^{(3)}$	$35MN^{(1)}N^{(2)}N^{(3)}$
D	$2(1+D)MN^{(1)}\dots N^{(D)}$	$((1+D)(2(1+D)+1)-1)MN^{(1)}\dots N^{(D)}$

TABLE 2.I: The number of first and second derivatives that are necessary to compute both the gradient vector and Hessian matrix of the fidelity for 1- 2- and 3-dimensional datasets, as well as a general D -dimensional dataset.

$$\frac{\partial^2 x}{\partial \theta [m+M]^2} \equiv \frac{\partial^2 x}{\partial \phi [m]^2} = -x, \quad (2.45e)$$

$$\frac{\partial^2 x}{\partial \theta [m+M] \partial \theta [m+(d+1)M]} \equiv \frac{\partial^2 x}{\partial \phi [m] \partial f^{(d)} [m]} = -2\pi \Delta_t^{(d)} n^{(d)} x, \quad (2.45f)$$

$$\frac{\partial^2 x}{\partial \theta [m+M] \partial \theta [m+(d+D+1)M]} \equiv \frac{\partial^2 x}{\partial \phi [m] \partial \eta^{(d)} [m]} = -i \Delta_t^{(d)} n^{(d)} x, \quad (2.45g)$$

$$\frac{\partial^2 x}{\partial \theta [m+(d+1)M]^2} \equiv \frac{\partial^2 x}{\partial f^{(d)} [m]^2} = -4\pi^2 (\Delta_t^{(d)} n^{(d)})^2 x, \quad (2.45h)$$

$$\frac{\partial^2 x}{\partial \theta [m+(d+1)M] \partial \theta [m+(d+D+1)M]} \equiv \frac{\partial^2 x}{\partial f^{(d)} [m] \partial \eta^{(d)} [m]} = -2\pi i (\Delta_t^{(d)} n^{(d)})^2 x, \quad (2.45i)$$

$$\frac{\partial^2 x}{\partial \theta [m+(d+D+1)M]^2} \equiv \frac{\partial^2 x}{\partial \eta^{(d)} [m]^2} = (\Delta_t^{(d)} n^{(d)})^2 x, \quad (2.45j)$$

$$\frac{\partial^2 x}{\partial \theta [i] \partial \theta [j]} = \frac{\partial^2 x}{\partial \theta [j] \partial \theta [i]}, \quad (2.45k)$$

$$\frac{\partial^2 x}{\partial \theta [i] \partial \theta [j]} = 0 \text{ if not specified above.} \quad (2.45l)$$

Equation 2.45l indicates that any second derivative with respect to two parameters which do not belong to the same oscillator will always be 0. This, along with the symmetrical nature of the second derivatives (Equation 2.45k) drastically reduces the required number of second derivatives to compute, from $4(1+D)^2 M^2$ per data-point to $(1+D)(3+2D)M$. Finally, Equation 2.45a indicates that another M second derivatives do not need to be computed. See Table 2.1 the total number of derivatives that need to be computed for 1D, 2D, and 3D signals.

2.3.3 Approximating the Hessian

Despite many of the model second derivatives being 0, computation of those that are not zero, and subsequently forming the Hessian matrix, is the most computationally expensive part of the

optimisation. There are a large number of optimisation problems where this too is the case, and as such there is considerable precedent for improving the efficiency of optimisation algorithms by generating less expensive approximations of the Hessian. Examples include the Gauss-Newton (GN) method and Levenberg-Marquardt (LM) algorithm, which are specifically for residual sum-of-squares problems[52: Chapter 10], and quasi-Newton methods such as the BFGS method[52: Chapter 6].

The GN and LM approaches replace the true Hessian matrix at each iteration with the following expression:

$$\mathbf{H}[i,j] \approx 2\Re \left(\frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[i]}, \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[j]} \right), \quad (2.46)$$

i.e. term ② in Equation 2.43b involving the second derivatives is neglected. The only thing that needs to be generated is the Jacobian $\mathbf{J} = \partial \mathbf{X} / \partial \boldsymbol{\theta}$, an array containing the $2(1+D)MN^{(1)}\dots N^{(D)}$ first derivatives of \mathbf{X} (Table 2.1). This brings a very large reduction in the computational cost, as no extra derivatives need to be computed for the Hessian at all, since the Jacobian is already required for generating the gradient vector (Equation 2.43a). In situations where the residuals between the data and model are small, term ① will tend to dominate term ②, and as such these methods often enjoy a convergence rate close to that of Newton's method close to local minima. Despite this, by invoking this approximation, the rate of convergence (i.e. the number of iterations required to reach $\boldsymbol{\theta}^{(*)}$) tends to be adversely affected. (See Section 2.3.6 for an example of this phenomenon)

2.3.4 Phase Variance Minimisation

While numerical optimisation procedures can return estimates $\boldsymbol{\theta}^{(*)}$ which can achieve highly accurate reconstructions of the original FID (i.e. \mathbf{Y} and $\mathbf{X}(\boldsymbol{\theta}^{(*)})$ are in close agreement), the estimate won't necessarily provide a faithful description of the physical phenomenon that has given rise to the data. For the purposes of FID estimation, the goal is to ensure that each resonance contributing to the FID is described by a single oscillator. In scenarios where the model order M used is greater than the true number of resonances, over-fitting of the data will occur, leading to spurious features in the estimation result, such as single resonances being fit by multiple oscillators and/or noise components being fit. To overcome this problem, it is desirable to include known information about the signal into the optimisation routine as a means of guiding the parameter vector to a more appropriate final value. One particular means of achieving this which has been found to be effective is the incorporation of the variance of oscillator phases into the fidelity, such that it becomes

$$\mathcal{F}_\phi(\boldsymbol{\theta} | \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X}(\boldsymbol{\theta})\|^2 + \text{Var}_o(\boldsymbol{\phi}), \quad (2.47)$$

where $\text{Var}_o(\boldsymbol{\phi}) := \text{Var}_o(\boldsymbol{\theta}[M : 2M])$ is the *circular* variance of the oscillator phases (*vide infra*).

Remark 2. The inclusion of the phase variance into the fidelity is one of the motivating reasons for normalising the data prior to estimation (see Remark 1). $\text{Var}_o(\phi)$ is constrained to the interval $[0, 1]$. If the data were not normalised, it is likely that $\|\mathbf{Y} - \mathbf{X}\|^2$ would dominate $\text{Var}_o(\phi)$ in Equation 2.47, such that the influence of the phase variance would be negligible.

For the phase variance to act effectively, it is necessary to apply phase-correction to the data prior to estimation, as the assumption that the phases of all contributing resonances are equal is typically valid. **Mention phase-variation of multiplet peaks in experiments where J-coupling evolves prior to acquisition? T_2 for example.** The inclusion of phase variance has also been found to be effective at purging excessive oscillators that may be present in the initial guess $\theta^{(0)}$, which be thought of in the following way:

- (i) Assume that the initial guess contains more oscillators than the true number of resonances. In this circumstance, it is common to find that true resonances are fit with an acceptable oscillator, whilst extra oscillators exist with spurious phases on account of over-fitting.
- (ii) Unconstrained numerical optimisation is now run, with the fidelity given by Equation 2.47. The phase variance will force oscillators with phases that are significantly different to the majority of oscillators to drastically alter their phases to match them. For some/all of these oscillators, it may be the case that the optimiser drives these to acquire a negative amplitude, such that they act as if they have a phase of π while ensuring that $\text{Var}_o(\phi)$ is small.
- (iii) This provides a criterion for detecting oscillators which are likely to be excessive. Such oscillators can be purged from θ by periodically checking whether any amplitudes (i.e. $\theta[:M]$) have become negative, purging these, and re-starting the optimisation.
- (iv) The numerical optimisation routine is re-started each time oscillators are purged. Termination is achieved once the routine converges and no negative-amplitude oscillators exist in θ .

Circular Variance

Oscillator phases are an example of a *circular variable*, in that all phases are wrapped within an interval of width 2π , for example $(-\pi, \pi]$. Given an unconstrained (unwrapped) phase $\tilde{\phi} \in \mathbb{R}$, the corresponding wrapped phase is given by

$$\phi = ((\tilde{\phi}[m] + \pi) \bmod 2\pi) - \pi. \quad (2.48)$$

This makes the conventional (linear) definition of variance, given by

$$\text{Var}_l(\boldsymbol{\phi}) = \frac{1}{M} \sum_{m=0}^{M-1} (\boldsymbol{\phi}[m] - \mu(\boldsymbol{\phi}))^2, \quad (2.49a)$$

$$\mu(\boldsymbol{\phi}) = \frac{1}{M} \sum_{m=0}^{M-1} \boldsymbol{\phi}[m], \quad (2.49b)$$

unsuitable for phases. Consider as a simple example a scenario where there are two oscillators with phases $\tilde{\boldsymbol{\phi}} = [\pi + \delta, \pi - \delta]^T$ for some small δ .[†] The phase variance is expected to be small as these two phases are similar. However, with the inclusion of wrapping (i.e. application of Equation 2.48), these phases would actually be set to $\boldsymbol{\phi} = [-\pi + \delta, \pi - \delta]^T$, and the conventional definition of phase variance would be large. It is therefore apparent that a definition of variance which accounts for the periodicity of the phases is needed. The *circular variance* is given by [63: Chapter 3]

$$[0, 1] \ni \text{Var}_o(\boldsymbol{\phi}) = 1 - \frac{R}{M}, \quad (2.50a)$$

$$R = \sqrt{c_\Sigma^2 + s_\Sigma^2}, \quad (2.50b)$$

$$c_\Sigma = \sum_{m=0}^{M-1} \cos \boldsymbol{\phi}[m], \quad (2.50c)$$

$$s_\Sigma = \sum_{m=0}^{M-1} \sin \boldsymbol{\phi}[m]. \quad (2.50d)$$

R is the length of the resultant vector produced by summing M unit vectors with the angles given by $\boldsymbol{\phi}$. In the case that all the vectors have the same angle, $R = M$, leading to the variance being 0 as expected. At the other extreme, with M vectors uniformly separated about the unit circle (with an angle $2\pi/M-1$ between all pairs of adjacent vectors), the vectors will perfectly cancel, leading to $R = 0$. In this case, the maximum variance of 1 is obtained.

The first and second derivatives of the circular variance are required for the computation of the gradient vector and Hessian matrix. These are given by

$$\frac{\partial \text{Var}_o(\boldsymbol{\phi})}{\partial \theta[i]} = \begin{cases} \frac{1}{MR} (c_\Sigma \sin \boldsymbol{\phi}[i-M] - s_\Sigma \cos \boldsymbol{\phi}[i-M]) & M \leq i < 2M \\ 0 & \text{otherwise} \end{cases} \quad (2.51a)$$

[†]The fact that the circular phase variance was necessary rather than the linear phase variance came rather late during my PhD. It was specifically realised while the estimation of inversion recovery datasets was being considered (see Section **InvRec reference**), as the initial increment of these datasets features resonances with phases close to π , such that the linear phase variance that was originally programmed in NMR-EsPy was giving far larger values than anticipated.

$$\frac{\partial^2 \text{Var}_o(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}[i] \partial \boldsymbol{\theta}[j]} = \begin{cases} \frac{1}{RM} \left[\frac{1}{R^2} (c_{\Sigma} \sin \boldsymbol{\phi}[i-M] - s_{\Sigma} \cos \boldsymbol{\phi}[i-M])^2 + c_{\Sigma} \cos \boldsymbol{\phi}[i-M] + s_{\Sigma} \sin \boldsymbol{\phi}[i-M] - 1 \right] & M \leq i, j < 2M, i = j \\ \frac{1}{RM} \left[\frac{1}{R^2} (c_{\Sigma} \sin \boldsymbol{\phi}[i-M] - s_{\Sigma} \cos \boldsymbol{\phi}[i-M]) \times (c_{\Sigma} \sin \boldsymbol{\phi}[j-M] - s_{\Sigma} \cos \boldsymbol{\phi}[j-M]) - \cos(\boldsymbol{\phi}[i-M] - \boldsymbol{\phi}[j-M]) \right] & M \leq i, j < 2M, i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (2.51b)$$

2.3.5 Estimation Errors

The errors associated with NLP can be extracted using information about the Hessian at $\boldsymbol{\theta}^{(*)}$. The vector of standard errors is equivalent to the square root of the diagonal elements of the inverse of the observed Fisher Information matrix for $\boldsymbol{\theta}^{(*)}$ [64: Section 2.7]:

$$\boldsymbol{\epsilon}(\boldsymbol{\theta}^{(*)}) = \sqrt{\text{diag}(\mathbf{I}(\boldsymbol{\theta}^{(*)})^{-1})}, \quad (2.52)$$

where the observed Fisher Information matrix contains the negative partial second derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$:

$$\mathbf{I}(\boldsymbol{\theta})[i,j] = -\frac{\partial^2 \ell(\boldsymbol{\theta} | \mathbf{Y})}{\partial \boldsymbol{\theta}[i] \partial \boldsymbol{\theta}[j]}. \quad (2.53)$$

Recalling the assumed form of the log-likelihood (Equation 2.5), elements of $\mathbf{I}(\boldsymbol{\theta})$ are

$$\mathbf{I}(\boldsymbol{\theta})[i,j] = -\frac{1}{\sigma^2} \Re \left(\left\langle \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[i]}, \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}[j]} \right\rangle - \left\langle (\mathbf{Y} - \mathbf{X}), \frac{\partial^2 \mathbf{X}}{\partial \boldsymbol{\theta}[i] \partial \boldsymbol{\theta}[j]} \right\rangle \right), \quad (2.54)$$

which very closely resembles the Hessian of $\boldsymbol{\theta}$:

$$\mathbf{I}(\boldsymbol{\theta})[i,j] = \frac{1}{2\sigma^2} (\mathbf{H}(\boldsymbol{\theta})[i,j]). \quad (2.55)$$

The standard errors therefore take the form

$$\boldsymbol{\epsilon}(\boldsymbol{\theta}^{(*)}) = \sqrt{2\sigma^2 \text{diag}([\mathbf{H}(\boldsymbol{\theta}^{(*)})]^{-1})}. \quad (2.56)$$

Considering that the mean and variance of the noise \mathbf{W} are 0 and $2\sigma^2$, respectively:

$$2\sigma^2 = \frac{1}{\mathfrak{N}-1} \sum_{n^{(1)}=0}^{N^{(1)}-1} \cdots \sum_{n^{(D)}=0}^{N^{(D)}-1} \left| \mathbf{W}[n^{(1)}, \dots, n^{(D)}] \right|^2 = \frac{1}{\mathfrak{N}-1} \|\mathbf{Y} - \mathbf{X}(\boldsymbol{\theta}^{(*)})\|^2, \quad (2.57)$$

so that finally a useable expression for the standard errors is arrived at:

$$\epsilon(\theta^{(*)}) = \sqrt{\frac{\mathcal{F}(\theta^{(*)}) \operatorname{diag}([\mathbf{H}(\theta^{(*)})]^{-1})}{N - 1}} \quad (2.58)$$

Algorithm 2 Nonlinear programming routine employed in NMR-EsPy. This makes use of Algorithms 4.1 & 7.2 in [52], with a extra check inserted to deal with any negative-amplitude oscillators which may be generated as the routine evolves.

```

1: procedure NLP( $\mathbf{Y} \in \mathbb{C}^{N^{(1)} \times \dots \times N^{(D)}}$ ,  $\theta^{(0)} \in \mathbb{R}^{2(D+1)M}$ )
2:    $\Delta^{(0)} \leftarrow 1/10 \|\mathbf{g}(\theta^{(0)} | \mathbf{Y})\|$ ;
3:    $\Delta_{\max} \leftarrow 16\Delta^{(0)}$ ;
4:   for  $k = 0, 1, \dots$  do
5:      $\mathbf{p}^{(k)} \leftarrow \text{STEIHAUGTOINT}(\mathbf{Y}, \theta^{(k)}, \Delta^{(k)})$ ; ▷ See Algorithm A2
6:      $\rho^{(k)} \leftarrow \frac{\mathcal{F}_{\phi}(\theta^{(k)}) - \mathcal{F}_{\phi}(\theta^{(k)} + \mathbf{p}^{(k)})}{\mathcal{F}_{\phi}(\theta^{(k)}) - \mathcal{F}_{\phi}(\theta^{(k)} + 2\mathbf{p}^{(k)})}$ ;
7:     if  $\rho^{(k)} < 1/4$  then
8:        $\Delta^{(k+1)} \leftarrow 1/4\Delta^{(k)}$ ;
9:     else if  $\rho^{(k)} > 3/4$  and  $\|\mathbf{p}^{(k)}\| = \Delta^{(k)}$  then
10:       $\Delta^{(k+1)} \leftarrow \min(2\Delta^{(k)}, \Delta_{\max})$ ;
11:    else
12:       $\Delta^{(k+1)} \leftarrow \Delta^{(k)}$ ;
13:    end if
14:    if  $\rho^{(k)} > 3/20$  then
15:       $\theta^{(k+1)} \leftarrow \theta^{(k)} + \mathbf{p}^{(k)}$ ;
16:    else
17:       $\theta^{(k+1)} \leftarrow \theta^{(k)}$ ;
18:    end if
19:    if  $k \bmod 25 = 0$  and  $\theta^{(k+1)}$  contains negative amplitudes then
20:       $\theta^{(0)} \leftarrow \theta^{(k+1)}$  with negative-amplitude oscillators removed;
21:       $\theta^{(*)}, \epsilon^{(*)} \leftarrow \text{NLP}(\mathbf{Y}, \theta^{(0)})$ ;
22:    end if
23:    if  $\|\mathbf{g}(\theta^{(k+1)})\| < 10^{-8}$  then
24:      break;
25:    end if
26:  end for
27:   $\theta^{(*)} \leftarrow \theta^{(k+1)}$ 
28:   $\epsilon^{(*)} \leftarrow \sqrt{\frac{\mathcal{F}(\theta^{(*)}) \operatorname{diag}([\mathbf{H}(\theta^{(*)})]^{-1})}{(N^{(1)} \dots N^{(D)}) - 1}}$ 
29:  return  $\theta^{(*)}, \epsilon^{(*)}$ ;
30: end procedure

```

2.3.6 Visualisation of a simple example

Figure 2.2 provides a visualisation of numerical optimisation applied on a simulated FID comprising a single resonance. The FID was constructed using Equation ?? with $D = 1$, $M = 1$, $N^{(1)} = 64$, $f_{\text{sw}}^{(1)} = 5.2 \text{ Hz}$ ($\Delta_t^{(1)} \approx 0.192 \text{ s}^{-1}$), and $f_{\text{off}}^{(1)} = 0 \text{ Hz}$. The resonance was parameterised by $\theta \in \mathbb{R}^4$ comprising $\alpha = 1$, $\phi = 0 \text{ rad}$, $f^{(1)} = 1 \text{ Hz}$, $\eta^{(1)} = 0.2 \text{ s}^{-1}$. White Gaussian noise was

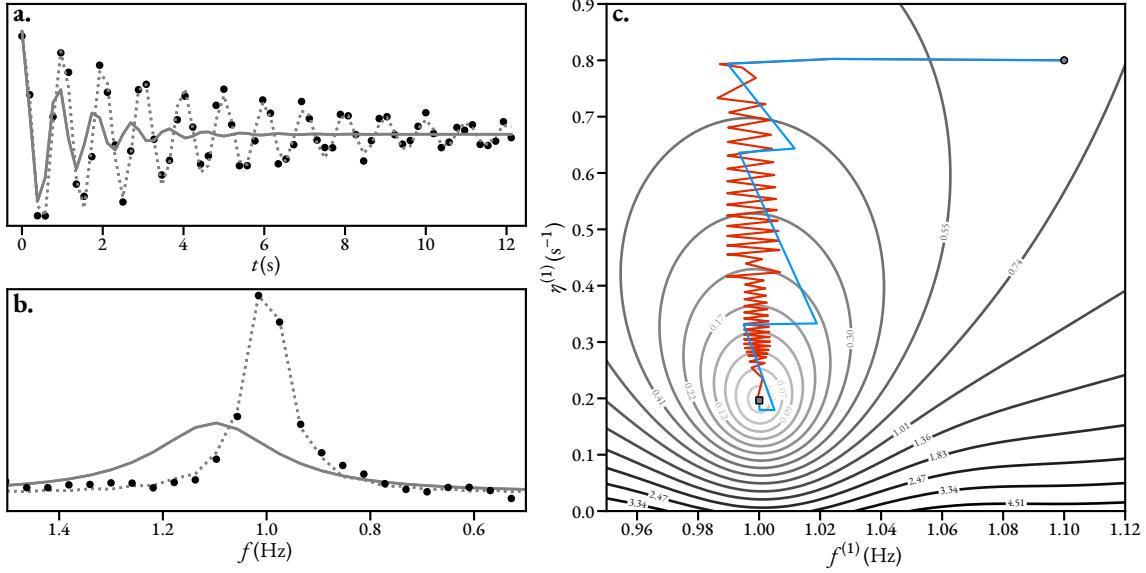


FIGURE 2.2: A visualisation of the trajectory of a 2-parameter optimisation involving a simulated FID comprising a single resonance. **a.** & **b.** Representations of the signal in the time domain and Fourier domain, respectively. Black dots: the synthetic signal \mathbf{Y} . Solid grey line: the model generated using the initial guess $\mathbf{X}(\boldsymbol{\theta}^{(0)})$. Dotted grey line: the model generated using the optimised result, $\mathbf{X}(\boldsymbol{\theta}^{(*)})$. **c.** A contour plot of the fidelity. Blue line: the trajectory of the parameter vector with the true Hessian matrix used in computing each update. Red line: the analogous trajectory using the Hessian approximation in place of the true Hessian.

added to the FID to give it an SNR of approximately 10 dB. As the visualisation of 5D space is beyond the scope of this work, only two parameters, the frequency and damping factor were optimised from an initial guess, with the amplitude and phase being fixed to their true values. The initial guess comprised a frequency of 1.1 Hz, and a damping factor of 0.8 s^{-1} , with the solid grey lines in panels a & b denoting the model generated using the initial guess in the time- and Fourier-domains, respectively. $\boldsymbol{\theta}^{(0)}$ was subjected to NLP twice. In the first instance, the exact Hessian matrix (Equation 2.43b) was used in order to compute each update step, while in the second the Hessian approximation given by Equation 2.46 was used. The initial radius of the trust region was set to $1/10$ of the gradient norm (≈ 0.3), which has a precedent in the literature[65]. The trajectories of the parameter vector are denoted as coloured lines in panel c. In both cases, the NLP routine successfully generated a result $\boldsymbol{\theta}^{(*)}$ which agreed with the true frequency and damping factor used to construct the FID. However, it is clear that using the true Hessian matrix (blue) in this example realised a far better rate of convergence compared with the approximated analogue (red), which exhibited “zig-zagging”, a phenomenon often seen in gradient search methods. 14 iterations were required to reach the convergence criterion $\epsilon \leq 10^{-8}$ when the true Hessian was used, while 81 were required for the approximated case. While an anecdotal example, this highlights that use of the true Hessian matrix tends to allow a better rate of convergence, however for FIDs comprising many resonances and far more points, the approximated form often requires a shorter time to

converge overall, as the cost of computing the Hessian second derivatives dominates the rate of convergence.

2.4 Profiling the MPM and NLP

Producing $\theta^{(*)}$ via generation of the initial guess using the MPM, and subjecting this to numerical optimisation involves operations which can be computationally demanding, both in terms of the amount of work done by the central processing unit (CPU), and space requirements (i.e. the amount of random access memory (RAM) required to store all the required information as the routine runs). For the MPM, the most demanding aspect is SVD calculations, which for numerical optimisation, it is generation of the Hessian matrix. Detailed accounts of the computational complexity of the MPM and MMEMPM has been described previously. However, it is useful to consider what the actual running times of these routines are. This is particularly useful, as a lot of the accounts on running the MPM are from decades before this work, and so the speed at which the estimation routine can be expected to take will have decreased a lot thanks to improvements in processing power. As an example, the account by Pines a coworkers from 1997 states that a signal with $N^{(1)} = 1024$ would take about 4.5 min to be processed by the MDL and MPM, using a 100 MHz CPU[66]. On the system used for all results generated for this work (see Remark 3), an equivalent computation takes about 100 ms.

Remark 3. All results generated in this work were acquired using a workstation featuring a Intel® Core™ i9-10900X CPU @ 3.7 GHz, and 32 GiB of RAM.

To gain a more concrete understanding of...

- ❖ **1D MPM: created signal with 10 oscs, varied $N^{(1)}$ from 512 → 8192 in steps of 512. $L^{(1)} = N^{(1)}/3$**
- ❖ **Total time very closely agreed with time to compute SVD of H_Y , especially for larger $N^{(1)}$: cubic dependence found.**
- ❖ **Panel a in plot: $N^{(1)}$ vs wall clock time, with cube root scale.**
- ❖ **Linear fit to circular points is shown (for square points (small $N^{(1)}$), other parts of the algorithm (notably EVD of VM1+VM2) had an appreciable impact on time.**
- ❖ **Peak memory requirement showed perfect quadratic relationship. Why?**
- ❖ **Extrapolated linear fit shows that unless very big signals considered, modern PCs should be able to cope.**

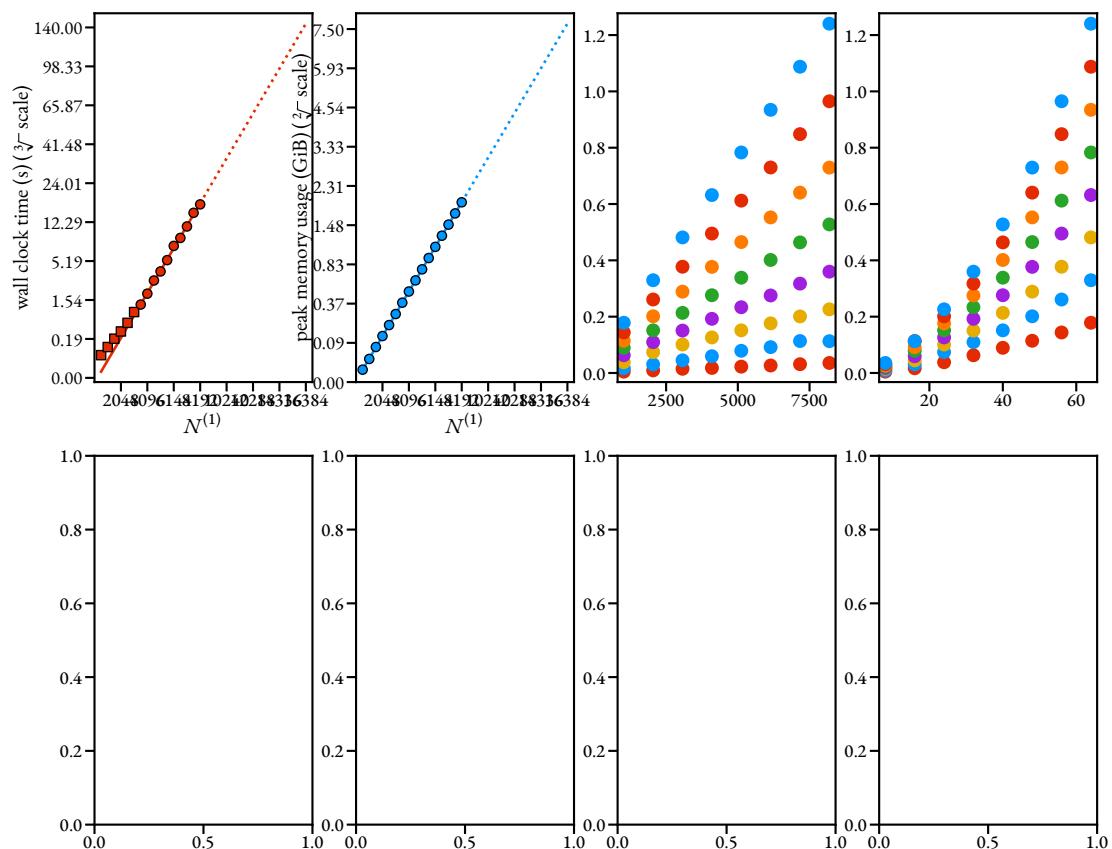


FIGURE 2.3: TODO

2.5 Frequency Filtration

The previous section provides motivation for finding ways to reduce the number of points in the signal and also the number of oscillators that the signal contains. This has led to work on a procedure for generated frequency-filtered “sub-FIDs” from the original data. A detailed description of the filtering procedure is presented in this section.

2.5.1 The virtual echo

In brief, the filtering procedure consists taking the FT of the FID, applying a band-pass filter on the spectral data to discard parts of not being considered, and returning the spectrum back to the time-domain by an inverse Fourier transform (IFT). For a filtered FID to be faithfully described by the model of a summation of damped complex sinusoids, it is necessary that the spectral peaks of interest lie effectively entirely within the filter region[‡]. For absorptive Lorentzians, due to their characteristically narrow linewidths, this is straightforward. However, for broader dispersive Lorentzians, this is far more challenging. For this reason, generating a spectrum in which only the real component is retained is desired. Assuming the data has been phase-corrected, this will produce a spectrum comprising only absorptive Lorentzians. The virtual echo (VE) has been employed here, which has found application in the field of compressed sensing NMR[67–69]. This is a signal with double the size as the original FID, with the key characteristic that its FT has a real component which is equivalent to its counterpart derived from an unaltered FID (except it has double the points), and an imaginary component of 0s.

The 1D virtual echo

Assuming that a 1D FID has been phase-corrected, such that $\phi = \mathbf{0} \in \mathbb{R}^M$, it can be denoted as

$$\mathbf{y} = \boldsymbol{\gamma} \odot (\mathbf{c}^{(1)} + i\mathbf{s}^{(1)}) + \mathbf{w}, \quad (2.59a)$$

$$\boldsymbol{\gamma} = \sum_m \boldsymbol{\alpha}[m] \exp(-\boldsymbol{\eta}^{(1)}[m] \boldsymbol{\tau}^{(1)}), \quad (2.59b)$$

$$\mathbf{c}^{(1)}/\mathbf{s}^{(1)} = \sum_m \cos / \sin \left(2\pi \mathbf{f}^{(1)}[m] \boldsymbol{\tau}^{(1)} \right), \quad (2.59c)$$

$$\boldsymbol{\tau}^{(1)} = \begin{bmatrix} 0 & \Delta_t^{(1)} & \dots & (N^{(1)} - 1) \Delta_t^{(1)} \end{bmatrix}^T \quad (2.59d)$$

[‡]Lorentzian lineshapes tend to, but don't reach zero, as the distance from the maximum tends to ∞ [29]. However, as long as a sufficiently wide filtering is employed, the regions of the Lorentzian which do not pass through the filter can be assumed to be negligible.

The frequency-dependence has been decomposed into its real and imaginary components. With this in mind, a conjugate pair of signals are defined:

$$\psi_{\pm} = \gamma \odot (c^{(1)} \pm i s^{(1)}) + w \equiv \Re(\mathbf{y}) \pm i \Im(\mathbf{y}) \quad (2.60)$$

Two vectors $\{\mathbf{t}_1, \mathbf{t}_2\} \in \mathbb{C}^{2N^{(1)}}$ are constructed using the conjugate pair:

- ⊕ \mathbf{t}_1 is given by ψ_+ padded with zeros from below:

$$\mathbf{t}_1 = \begin{bmatrix} \psi_+ \\ \mathbf{0} \in \mathbb{C}^{N^{(1)}} \end{bmatrix}, \quad (2.61)$$

- ⊕ \mathbf{t}_2 is given by ψ_- with its elements in reversed order ($\cdot^{\leftrightarrow(1)}$), padded with zeros from above, and finally subjected to a right circular shift by one element ($\cdot^{\circlearrowright(1)}$):

$$\mathbf{t}_2 = \begin{bmatrix} \mathbf{0} \in \mathbb{C}^{N^{(1)}} \\ \psi_-^{\leftrightarrow(1)} \end{bmatrix}^{\circlearrowright(1)}, \quad (2.62)$$

The VE \mathbf{y}_{ve} is then given by $\mathbf{t}_1 + \mathbf{t}_2$, with the first element divided by 2, which is equivalent to

$$\mathbf{y}_{ve} = \begin{bmatrix} \Re(\mathbf{y}[0]) & \mathbf{y}[1] & \cdots & \mathbf{y}[N^{(1)} - 1] & 0 & \mathbf{y}[N^{(1)} - 1]^* & \cdots & \mathbf{y}[1]^* \end{bmatrix}^T. \quad (2.63)$$

A signal of this form is illustrated in panel a of Figure 2.4. As eluded to already, the FT of \mathbf{y}_{ve} produces a spectrum \mathbf{s}_{ve} such that $\Im(\mathbf{s}_{ve}) = \mathbf{0}$, with $\Re(\mathbf{s}_{ve})$ featuring absorption Lorentzian peaks (panel b of Figure 2.4).

The 2D virtual echo

The VE concept can be generalised to any number of dimensions, assuming that a pair of amplitude-modulated signals exist for each indirect-dimension, thus requiring a set of 2^{D-1} signals for a D -dimensional dataset. **Maybe describe the general case in the appendix if you're bold enough?**

For the 2D case, this corresponds to the pair of signals $\{Y_{\cos}, Y_{\sin}\}$, given by (1.14) with $D = 2$ and $\zeta = \{\cos(\cdot), \sin(\cdot)\}$, taking the forms (with noise neglected)

$$Y_{\cos} = \boldsymbol{\Gamma} \odot \mathbf{C}^{(1)} \odot (\mathbf{C}^{(2)} + i \mathbf{S}^{(2)}), \quad (2.64a)$$

$$Y_{\sin} = \boldsymbol{\Gamma} \odot \mathbf{S}^{(1)} \odot (\mathbf{C}^{(2)} + i \mathbf{S}^{(2)}), \quad (2.64b)$$

$$\boldsymbol{\Gamma} = \sum_m \boldsymbol{\alpha}[m] (\exp(-\boldsymbol{\gamma}^{(1)}[m] \boldsymbol{\tau}^{(1)}) \otimes \exp(-\boldsymbol{\gamma}^{(2)}[m] \boldsymbol{\tau}^{(2)})), \quad (2.64c)$$

$$\mathbf{C}^{(1)}/\mathbf{S}^{(1)} = \sum_m \cos / \sin (2\pi \mathbf{f}^{(1)} [m] \boldsymbol{\tau}^{(1)}) \otimes \mathbf{1} \in \mathbb{C}^{N^{(2)}}, \quad (2.64d)$$

$$\mathbf{C}^{(2)}/\mathbf{S}^{(2)} = \mathbf{1} \in \mathbb{C}^{N^{(1)}} \otimes \sum_m \cos / \sin (2\pi \mathbf{f}^{(2)} [m] \boldsymbol{\tau}^{(2)}). \quad (2.64e)$$

Four matrices are then constructed of the form

$$\begin{aligned} \mathbf{Y}_{\pm\pm} &= \mathbf{I} \odot (\mathbf{C}^{(1)} \pm^{(1)} i\mathbf{S}^{(1)}) \odot (\mathbf{C}^{(2)} \pm^{(2)} i\mathbf{S}^{(2)}) \\ &\equiv \Re(\mathbf{Y}_{\cos}) \pm^{(1)} \pm^{(2)} - \Im(\mathbf{Y}_{\sin}) + i(\pm^{(1)} \Re(\mathbf{Y}_{\sin}) \pm^{(2)} \Im(\mathbf{Y}_{\cos})), \end{aligned} \quad (2.65)$$

from which the matrices $\mathbf{T}_{1-4} \in \mathbb{C}^{2N^{(1)} \times 2N^{(2)}}$ are generated:

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{Y}_{++} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (2.66a)$$

$$\mathbf{T}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Y}_{-+}^{\leftrightarrow(1)} & \mathbf{0} \end{bmatrix}^{\odot(1)}, \quad (2.66b)$$

$$\mathbf{T}_3 = \begin{bmatrix} \mathbf{0} & \mathbf{Y}_{+-}^{\leftrightarrow(2)} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^{\odot(2)}, \quad (2.66c)$$

$$\mathbf{T}_4 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{--}^{\leftrightarrow(1,2)} \end{bmatrix}^{\odot(1,2)}. \quad (2.66d)$$

Appendix for explicit forms of the matrices \mathbf{T}_{1-4} The virtual echo is then given by $\mathbf{Y}_{ve} = \sum_{i=1}^4 \mathbf{T}_i$, with the first row and column divided by two. For a full outline of the 2D filtering procedure, see Algorithm A4.

It is possible to construct a virtual echo using an appropriate set of phase-modulated signals too, which for the 2D case would be $\{\mathbf{Y}_{pos}, \mathbf{Y}_{neg}\}$, given by (1.14) with $D = 2$ and $\zeta = \{\exp(i\cdot), \exp(-i\cdot)\}$. These can be used to generate an amplitude modulated pair via

$$\mathbf{Y}_{cos} = \frac{\mathbf{Y}_{pos} + \mathbf{Y}_{neg}}{2}, \quad (2.67a)$$

$$\mathbf{Y}_{sin} = \frac{\mathbf{Y}_{pos} - \mathbf{Y}_{neg}}{2i}. \quad (2.67b)$$

Should give ζ a dimension index, i.e. $\zeta^{(1)}$

2.5.2 The filtering process

Having constructed a virtual echo $\mathbf{Y}_{\text{ve}} \in \mathbb{C}^{2N^{(1)} \times \dots \times 2N^{(D)}}$, a spectrum with absorption Lorentzians is produced with $\mathbf{S}_{\text{ve}} = \text{FT}(\mathbf{Y}_{\text{ve}})$. To filter the spectrum, it is subjected to element-wise multiplication with a super-Gaussian function. The super-Gaussian is defined by a centre $c^{(d)} \in \mathbb{R} : 0 < c^{(d)} < 2N^{(d)}$ and a bandwidth $b^{(d)} \in \mathbb{N} : b^{(d)} < 2N^{(d)}$ in each dimension (panel c of Figure 2.4):

$$\mathbf{G} = \bigotimes_{d=1}^D \mathbf{g}^{(d)}, \quad (2.68a)$$

$$\mathbf{g}^{(d)}[n^{(d)}] = \exp\left(-2^{p+1}\left(\frac{n^{(d)} - c^{(d)}}{b^{(d)}}\right)^p\right) \forall n^{(d)} \in \{0, \dots, 2N^{(d)} - 1\}. \quad (2.68b)$$

An example of a 1D super-Gaussian is given in panel c of Figure 2.4. The scalar $p \in \mathbb{R}_{>0}$ dictates the steepness of the filter at the boundaries, with the function becoming more “box-like” as it increases. It is set to 40 in this work and as the default in NMR-EsPy. Application of the super-Gaussian filter to \mathbf{S}_{ve} would lead to large sections of the filtered spectrum being 0. This has an undesired impact on the MDL, as noise that has passed through filter (i.e. the noise inside the region of interest) will now seem to resemble true signal, as its amplitude is infinitely greater than the zeroed regions. A massive over-estimation of model order therefore result. In order to obtain better results from the model order selection, an array of synthetic AWGN is added to the filtered spectrum. To achieve this, a region in \mathbf{S}_{ve} is specified which contains no discernible signal peaks (referred to as the *noise region*). The variance of this region σ^2 is determined, and used to construct an array of values sampled from a normal distribution with mean 0 and variance σ^2 , $\mathbf{W}_{\sigma^2} \in \mathbb{R}^{2N^{(1)} \times \dots \times 2N^{(D)}}$. The filtered spectrum is then given by

$$\tilde{\mathbf{S}}_{\text{ve}} = \mathbf{S}_{\text{ve}} \odot \mathbf{G} + \mathbf{W}_{\sigma^2} \odot (\mathbf{1} - \mathbf{G}). \quad (2.69)$$

Note that the noise array’s magnitude at each point is attenuated by the value of the super-Gaussian filter. Inside the region of interest ($\mathbf{G}[n^{(1)}, \dots, n^{(D)}] = 1$), the noise is nullified. See panel d of Figure 2.4 for an example.

After filtering, $\tilde{\mathbf{S}}_{\text{ve}}$ is returned to the time-domain by IFT. The IFT of a real-valued spectrum generates a conjugate-symmetric signal. This is sliced in half in each dimension, generating the final filtered sub-FID $\tilde{\mathbf{Y}} \in \mathbb{C}^{N^{(1)} \times \dots \times N^{(D)}}$:

$$\tilde{\mathbf{Y}} = \frac{1}{2^{D-1}} \text{IFT}(\tilde{\mathbf{S}}_{\text{ve}})[0:N^{(1)}, \dots, 0:N^{(D)}]. \quad (2.70)$$

The scaling factor in (2.70) is to account for how many signals have been combined to generate the VE.

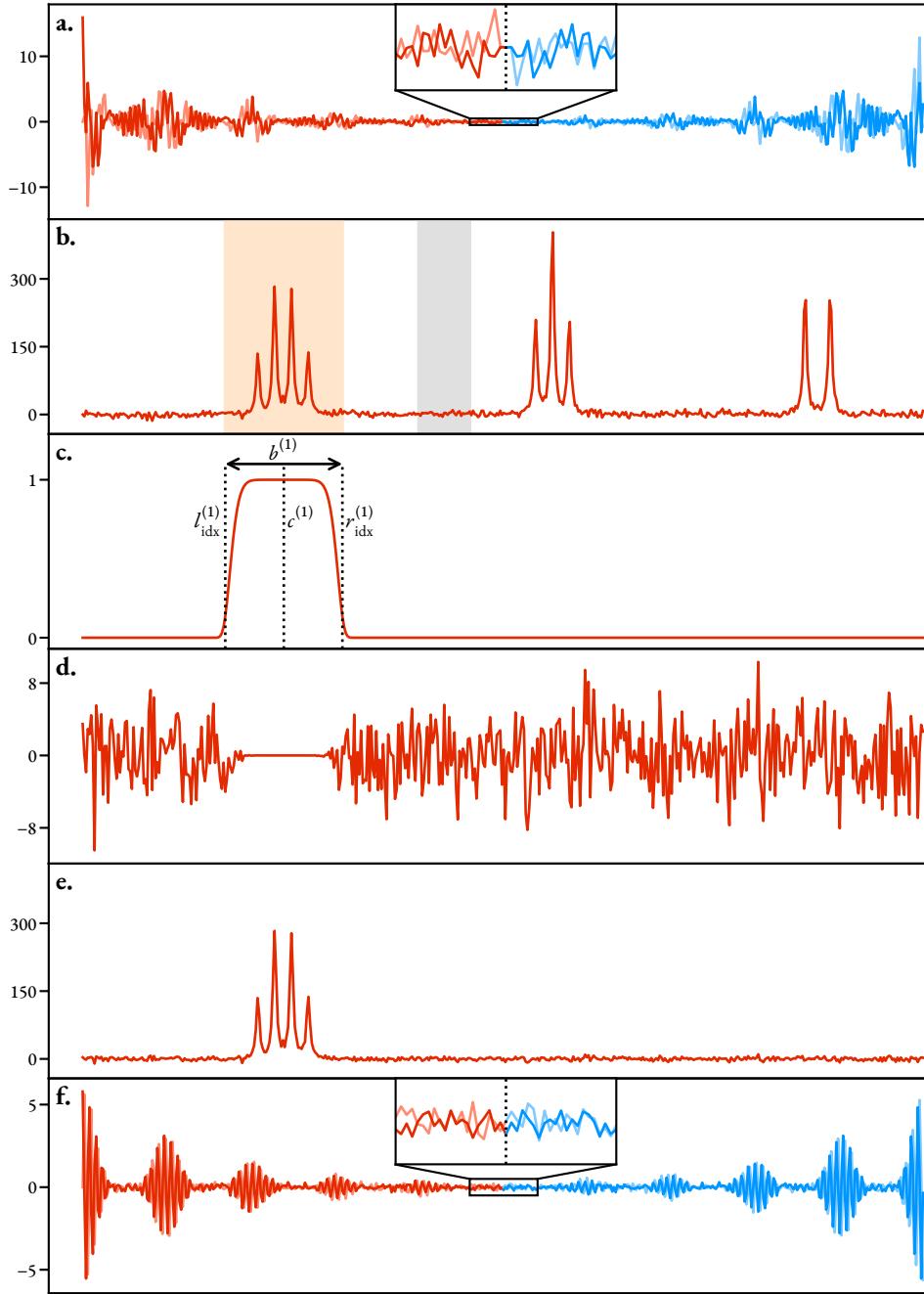


FIGURE 2.4: An illustration of the filtering procedure applied to a 1D FID. **a.** A VE y_{ve} , with the first and last $N^{(1)}$ points coloured red and blue, respectively. The middle of the VE is magnified to highlight its conjugate symmetry. **b.** The FT of the VE, s_{ve} . The region of interest (orange) and noise region (grey) are denoted. **c.** A super-Gaussian function used as a band-pass filter, g . **d.** AWGN vector to be added to the filtered spectrum. The magnitude of the signal at each point is dependent on the corresponding super-Gaussian value. **e.** The filtered spectrum \tilde{s}_{ve} , formed by applying the super-Gaussian filter, and adding the noise vector. **f.** The IFT of the filtered spectrum, \tilde{y}_{ve} , from which the final filtered signal \tilde{y} is obtained by extracting the first $N^{(1)}$ points.

Determining $c^{(d)}$ and $b^{(d)}$

The central index and bandwidth of the super-Gaussian filter function are given by the following expressions:

$$c^{(d)} = \frac{1}{2} \left(l_{\text{idx}}^{(d)} + r_{\text{idx}}^{(d)} \right), \quad (2.71\text{a})$$

$$b^{(d)} = l_{\text{idx}}^{(d)} - r_{\text{idx}}^{(d)}, \quad (2.71\text{b})$$

where $l_{\text{idx}}^{(d)}$ and $r_{\text{idx}}^{(d)}$ denote the desired indices where the filter's left and right bounds are located, respectively. Array indices can be obtained from the corresponding spectral frequency $f_{\text{Hz}}^{(d)}$ via

$$f_{\text{idx}}^{(d)} = \left\lfloor \frac{\left(2N^{(d)} - 1 \right) \left(f_{\text{sw}}^{(d)} + 2 \left(f_{\text{off}}^{(d)} - f_{\text{Hz}}^{(d)} \right) \right)}{2f_{\text{sw}}^{(d)}} \right\rfloor \quad (2.72)$$

$$\forall f_{\text{Hz}}^{(d)} \in \left[f_{\text{off}}^{(d)} - \frac{1}{2}f_{\text{sw}}^{(d)}, f_{\text{off}}^{(d)} + \frac{1}{2}f_{\text{sw}}^{(d)} \right].$$

Conversion from ppm to array indices can be achieved by replacing f_{Hz} in (2.72) with $f_{\text{ppm}}f_{\text{sfo}}$, where f_{sfo} is the transmitter frequency (MHz).

Spectrum slicing

Thus far, the method described is able to reduce the model order of a given signal, however the signal still comprises the same number of points. However it is clear that there are a large number of points outside the region of interest in $\tilde{\mathcal{S}}_{\text{ve}}$ that do not possess any meaningful information. Discarding such points should then lead to filtered FID with the same information about the resonances of interest, but with far fewer points. A slicing ratio is defined, $\xi \in \mathbb{R} : \xi > 1$, which dictates the left and right indices at which the spectrum should be sliced in each dimension:

$$l_{\text{slice}}^{(d)} = \begin{cases} c_{\text{idx}}^{(d)} - \left\lfloor \frac{b^{(d)}\xi}{2} \right\rfloor & \text{if } \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.73\text{a})$$

$$r_{\text{slice}}^{(d)} = \begin{cases} c_{\text{idx}}^{(d)} + \left\lceil \frac{b^{(d)}\xi}{2} \right\rceil & \text{if } \leq 2N^{(d)} - 1 \\ 2N^{(d)} - 1 & \text{otherwise} \end{cases} \quad (2.73\text{b})$$

The filtered spectrum is then sliced accordingly:

$$\tilde{\mathcal{S}}_{\text{ve, slice}} = \tilde{\mathcal{S}}_{\text{ve}} \left[l_{\text{slice}}^{(1)} : r_{\text{slice}}^{(1)} + 1, \dots, l_{\text{slice}}^{(D)} : r_{\text{slice}}^{(D)} + 1 \right]. \quad (2.74)$$

The equivalent process is applied to $\tilde{\mathbf{S}}_{\text{ve,slice}}$ to generate the final filtered sub-FID: IFT followed by slicing in half in each dimension. It is also necessary to scale the signal by the ratio of the number of points in the sliced spectrum and its unsliced counterpart.

$$\tilde{\mathbf{Y}} = \prod_{d=1}^D \left(\frac{r_{\text{slice}}^{(d)} - l_{\text{slice}}^{(d)}}{2N^{(d)}} \right) \text{IFT}(\tilde{\mathbf{S}}_{\text{ve, slice}}) [0 : N_{\text{slice}}^{(1)}, \dots, 0 : N_{\text{slice}}^{(D)}] \quad (2.75\text{a})$$

$$N_{\text{slice}}^{(d)} = \left\lfloor \frac{r_{\text{slice}}^{(d)} - l_{\text{slice}}^{(d)}}{2} \right\rfloor \quad (2.75\text{b})$$

Finally, the effective sweep widths and transmitter offsets will have been altered by this process. The corrected values can be computed using

$$f_{\text{sw,slice}}^{(d)} = \frac{r_{\text{slice}}^{(d)} - l_{\text{slice}}^{(d)}}{2N^{(d)} - 1} f_{\text{sw}}^{(d)} \quad (2.76\text{a})$$

$$f_{\text{off,slice}}^{(d)} = f_{\text{off}}^{(d)} + \frac{f_{\text{sw}}^{(d)}}{2} \left(1 - \frac{l_{\text{slice}}^{(d)} + r_{\text{slice}}^{(d)}}{2N^{(d)} - 1} \right) \quad (2.76\text{b})$$

2.6 Summary

The matrix pencil-based methods described above are well established as effective procedures for parametric estimation of time-domain signals in a number of disciplines, including radar detection[70], acoustics[TODO], and NMR[66], with specific application to relaxometry being a recently introduced application [71, 72]. Due to considerable advances computational processing power since the introduction of the technique, estimates can be acquired from NMR signals in reasonable times. One notable downside of the technique that has been realised while assessing its effectiveness in parametrising NMR FIDs is its propensity to return oscillators with unexpected phase behaviour, especially in scenarios involving resonances with close frequencies. For this reason, estimating phase-corrected FIDs, using the result of the MPM as an initial guess to feed into a phase-variance penalised NLP routine was proposed as a means of improving parameter estimates. The theory underpinning the procedure has been explored in this chapter.

The computational burden of running the procedure is large and often intractable for complete NMR signals, which often comprise thousands of samples, and at least hundreds of contributing resonances. This has been illustrated through profiling both the CPU time and the peak memory requirements for the 1D and 2D methods. **Say more when completed?** For this reason, a method to break FIDs into frequency-filtered sub-signals, by filtering spectra derived using virtual echoes, is introduced. This method enables the formation of signals with far fewer resonances and samples as compared to the full signal.

Algorithm 3 provides an overview of the principal steps involved in the 1D and 2D estimation procedures. Detailed algorithms for each step are presented elsewhere in this text.

Algorithm 3 An overview of the estimation procedure outlined in this work, for the consideration of 1D and 2D NMR signals.

```

1: procedure ESTIMATE1D( $\mathbf{y} \in \mathbb{C}^{N^{(1)}}$ ,  $\mathbf{r}_{\text{interest}} \in \mathbb{R}^2$ ,  $\mathbf{r}_{\text{noise}} \in \mathbb{R}^2$ ,  $M \in \mathbb{N}_0$ )
2:    $\tilde{\mathbf{y}} \leftarrow \text{FILTER1D } (\mathbf{y}, \mathbf{r}_{\text{interest}}, \mathbf{r}_{\text{noise}})$ ;                                 $\triangleright$  Algorithm A3
3:   if  $M = 0$  then
4:      $M \leftarrow \text{MDL } (\tilde{\mathbf{y}})$ ;                                               $\triangleright$  TODO
5:   end if
6:    $\theta^{(0)} \leftarrow \text{MPM } (\tilde{\mathbf{y}}, M)$ ;                                 $\triangleright$  Algorithm 1
7:    $\theta^{(*)}, \epsilon^{(*)} \leftarrow \text{NLP } (\tilde{\mathbf{y}}, \theta^{(0)})$ ;                       $\triangleright$  Algorithm 2
8:   return  $\theta^{(*)}, \epsilon^{(*)}$ ;
9: end procedure

10: procedure ESTIMATE2D( $\mathbf{Y}_{\text{cos}} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}$ ,  $\mathbf{Y}_{\text{sin}} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}$ ,  $\mathbf{R}_{\text{interest}} \in \mathbb{R}^{2 \times 2}$ ,  $\mathbf{R}_{\text{noise}} \in \mathbb{R}^{2 \times 2}$ ,  $M \in \mathbb{N}_0$ )
11:    $\tilde{\mathbf{Y}} \leftarrow \text{FILTER2D } (\mathbf{Y}_{\text{cos}}, \mathbf{Y}_{\text{sin}}, \mathbf{R}_{\text{interest}}, \mathbf{R}_{\text{noise}})$ ;           $\triangleright$  Algorithm A4
12:   if  $M = 0$  then
13:      $M \leftarrow \text{MDL } (\tilde{\mathbf{Y}}[:, 0])$ ;                                          $\triangleright$  Run the MDL on the first direct-dimension slice.
14:   end if
15:    $\theta^{(0)} \leftarrow \text{MMEMPPM } (\tilde{\mathbf{Y}}, M)$ ;                                 $\triangleright$  Algorithm A1
16:    $\theta^{(*)}, \epsilon^{(*)} \leftarrow \text{NLP } (\tilde{\mathbf{Y}}, \theta^{(0)})$ ;                       $\triangleright$  Algorithm 2
17:   return  $\theta^{(*)}, \epsilon^{(*)}$ ;
18: end procedure
```

Having established an estimation routine, the next chapter focusses on its performance, as well as applications which are possible through parametric estimation.

RESULTS AND APPLICATIONS

This chapter showcases numerous results generated using the estimation technique outlined in Chapter 2. Sections ?? and 3.2 focus on results generated on 1D and 2D amplitude/phase-modulated paired datasets, respectively. The subsequent three sections showcase three distinct applications that estimation facilitates. In Section 3.3, a means of generating broadband homodecoupled (pure shift) spectra through estimation of hypercomplex 2DJ datasets is provided. Section 3.4 highlights how a standard 1D estimation routine can be extended to datasets featuring a series of FIDs which feature attenuations in signal amplitudes across increments, such as inversion recovery (T_1) and diffusion experiments. Finally, a protocol to achieve ultra-broadband spectra through application of a single 90° frequency-swept pulse, which are devoid of quadratic phase dependencies as well as severe baseline distortions, is provided. All the results and applications presented here were acquired using the NMR-EsPy package, which is described in Chapter 4. Relevant information on the setup of simulations and experiments can be found in Appendix C. **Result metrics?**

3.1 Evaluating the Estimation Routine

3.1.1 1D Datasets

“Twenty signals”

In order to assess the estimation routine proposed in Chapter 2 – specifically the effectiveness of applying NLP using an initial guess generated using the MPM – a series of synthetic FIDs were constructed using (2.1) with $D = 1$. For each FID, a model order of $M = 20$ was used, the number of points sampled was $N^{(1)} = 1024$, the sweep width was $f_{\text{sw}}^{(1)} = 125 \text{ Hz}$, and the transmitter offset was $f_{\text{off}}^{(1)} = 0 \text{ Hz}$. Each oscillator was assigned a phase of 0° , while the amplitudes, frequencies and damping factors were drawn at random from the following distributions: $\alpha [m] \sim \mathcal{U}(1, 5)$, $f^{(1)} [m] \sim \mathcal{U}(-55 \text{ Hz}, 55 \text{ Hz})$, $\gamma^{(1)} [m] \sim \mathcal{U}(2 \text{ s}^{-1}, 8 \text{ s}^{-1}) \forall m \in \{0, \dots, 19\}$. An extra constraint

3 Results and Applications

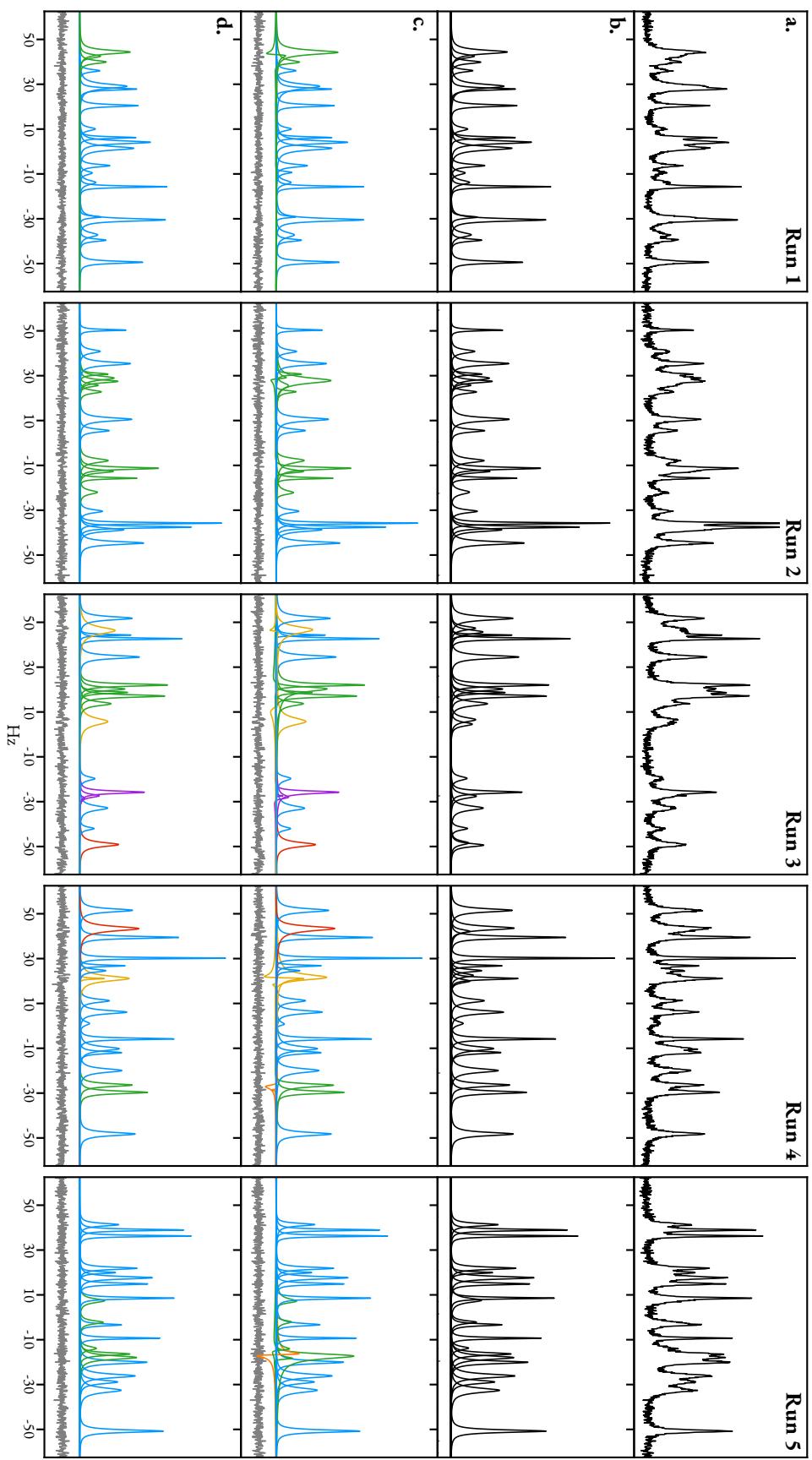


FIGURE 3:I: The result of estimating a series of 5 simulated signals comprising 20 oscillators (see the main text for details on how the datasets were constructed). **a.** Spectra of the datasets generated. **b.** Spectral lines corresponding to the true set of oscillators used to generate each dataset. **c.** Plots of spectral lines for each oscillator generated using the MPM. **d.** An equivalent plot for the result after applying NLP, with the MPM result being the initial guess. Also included in **c.** and **d.** is the residual between the data and the sum of the oscillator peaks (grey line). The colouring of oscillator lines in **c.** and **d.** is described in the main text.

was applied to the frequencies, such that no two oscillators were permitted to have frequencies that differed by less than $4f_{sw}^{(1)}/N^{(1)} \approx 0.49$ Hz. Each noiseless FID was then corrupted with AWGN, with a target SNR of 25 dB **Reference that this is the lowest SNR that the MPM considered to be effective**. The spectra of the simulated FIDs are presented in panel a of Figure 3.1, with the set of true oscillator peaks in panel b.

For each FID, the MPM was performed, assuming that the model order is 30, constituting a considerable over-fit. The MDL tended to produce considerable under-estimates of M for these FIDs, so a hard-coded value was used instead. Simulated signals featuring AWGN typically show a clean division between signal and noise components **better way of describing this?**, with noise components commonly being characterised by small amplitudes and/or very small damping factors. For this reason, prior to subjecting the MPM result to NLP, oscillators which satisfied either $\alpha [m] < 0.1$ or $\gamma^{(1)} [m] < 0.7 \text{ s}^{-1}$ were removed from the parameter set. The individual oscillators which make up the MPM result after purging noise components are displayed in panel c of Figure 3.1, along with the residual between the data and the summation of all the oscillator peaks (the model).

The MPM invariably generates a model with good agreement with the data, as evidenced by the residual. However, it can be seen that in several spectral regions across the datasets, especially ones that are highly crowded, oscillators possess parameters which deviate significantly from the true parameters, with the most notable feature being individual oscillator phases – these regularly stray far from 0° – and their associated amplitudes. The central motivation behind employing phase variance-regularised NLP is as a means of attempting to overcome this detrimental feature of the MPM. In panel c of Figure 3.1, the blue oscillators are those which agree very closely with a particular oscillator in the true set of parameters. Oscillators with other colours are not clearly mapped to a true oscillator, with the different colourings described shortly. The intention is for the NLP routine to adjust the parameters describing the non-blue oscillators in panel c such that they agree with oscillators found in the true set, while not affecting the blue oscillators. The results of NLP at convergence ($\epsilon = 10^{-8}$) are provided in panel d.

In discussing the outcome of the routine, it will be useful to introduce the concept of a *frequency neighbourhood*, a loose term which describes a small, continuous range of frequencies within the spectral window. As the NLP routine involves taking small steps through parameter space in an attempt to converge, it is unlikely that an oscillator which starts off with a frequency far away from a particular frequency neighbourhood will eventually enter it. As such, in order for the NLP routine to successfully estimate the region, sufficient oscillators need to present within the neighbourhood in the first place. Cases where the MPM generated enough oscillators for a given frequency neighbourhood, albeit with parameters which are noticeably off the true parameters are in either green or yellow. Green oscillators are those which the NLP routine was able to adjust in order to achieve agreement with the true result. As such, they indicate improvements to

the estimation result as opposed to the MPM being used by itself. Conversely, yellow oscillators denote cases where, though sufficient oscillators exists in the frequency neighbourhood in the initial guess, the NLP routine evolves such that at least one of the oscillators is driven by the phase variance constraint to acquire a negative amplitude, such that it is purged from the parameter set. This typically occurs when an oscillator has an initial phase which is considerably greater than 90°. Yellow oscillators therefore indicate cases where the final result has under-fit the dataset. There are a few instances, denoted by red oscillators, where the MPM assigned too few oscillators to a particular frequency region, and as such the NLP would not have been able to yield any improvement. The final two oscillator groupings, denoted by purple and orange, denote cases where the MPM generated more oscillators than are present in a given frequency neighbourhood (i.e. the data was over-fit in this region). Orange oscillators were purged by the NLP routine due to their acquiring negative amplitudes. This enabled a parsimonious fit of the frequency neighbourhood by the oscillators which remained. Finally, the purple oscillators denote the one occasion where an over-fit occurred, and the model order was not successfully reduced by the NLP routine.

Overall, it can be seen that the inclusion of NLP broadly improves the output of the MPM, where crowded regions often get assigned oscillators with spurious complex amplitudes (α and ϕ).

More discussion here...

Andrographolide

Figure 3.2 illustrates the outcome of applying the estimation routine to selected regions of a ^1H dataset of andrographolide (Figure A1.e), in DMSO-d₆. The NLP routine is effective at resolving the spurious phase-behaviour often generated by the MPM, while retaining reasonable fits to the data in most circumstances. Its ability to estimate parameters from resonances with high dynamic range is also evidenced by the fact that it was able to assign an intense, broad singlet, corresponding to water which has entered the sample over time, alongside a nearby quartet corresponding to the methylene of some residual ethanol in the sample.

Probably the most challenging aspect of estimating NMR signals is the fact that data frequently contains signals with incredibly similar frequencies, commonly due to the effect of scalar couplings. Molecules with fused ring systems such as andrographolide are prime examples of spin systems which generate such datasets, as they tend to have very dense coupling networks, and often exhibit appreciable long-range couplings (between spins separated by four or more bonds) alongside more prevalent two-bond (geminal) and three-bond (vicinal) couplings. Take the multiplet structure from spin (Q) as an example. This nucleus has separate 3-bond (vicinal) couplings to the diastereotopic protons (M) and (N), which are likely the greatest magnitude couplings associated with (Q). If these were the only couplings, a doublet of doublets (dd) structure would be expected, which is what has been fit by the estimation routine (see panel c in the region of 1.85 ppm

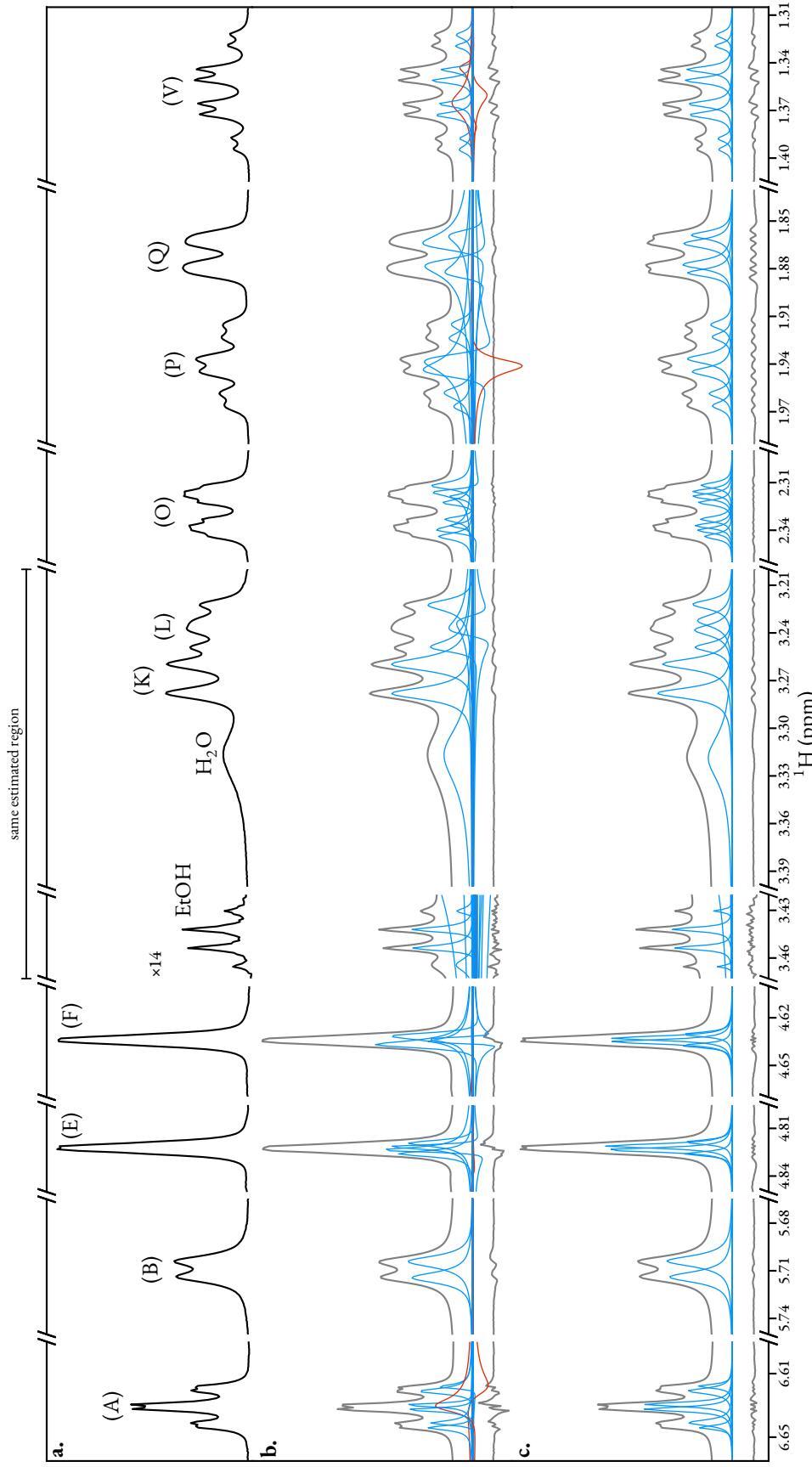


FIGURE 3.2: Result of applying the estimation routine to selected regions of a pulse-acquire dataset of andrographolide in DMSO-d_6 . **a.** Spectral data corresponding to the regions considered. **b.** The result of applying the MPM to the regions, with the model order predicted with the MDL. Blue/red lines: peaks of individual oscillators, grey line above: the model (sum of all oscillators), grey line below: the residual between the data and the model. **c.** The result after convergence of the NLP routine, again with the model above and residual below. Red peaks in panel b correspond to oscillators which acquire negative amplitudes during the NLP routine, and are subsequently purged. Note that one of the reasons estimated has been split in two in the figure to save space, with one half, featuring a signal from ethanol, being magnified.

to 1.88 ppm). However, a comparison of the data and the model indicates that there is clear discrepancy between the two, evidenced by systematic “wobbles” in the residual. An under-fit of the multiplet structure is therefore apparent. For comparison, note that the MPM generated oscillators with phases deviating far from 0° in order to agree as closely as possible to the data, however such a set of oscillators is unrealistic for a well-phased FID. Four-bond couplings with magnitudes that are large enough to influence the appearance of (Q)’s multiplet structure must therefore be present. Similar scenarios are apparent for spins (P) and (V), which from a cursory inspection appear to comprise triplet of doublets (dt) and quartet of doublets (dq) multiplet structures, respectively*, but which again have far more complex coupling structures due to the presence of small four-bond couplings.

3.2 2D Datasets

3.3 Pure-Shift Spectra via 2D J-resolved Estimation

Two key features of the NMR experiment for which improvements are constantly being sought are sensitivity and resolving power. There are numerous means of enhancing sensitivity, including technological advancements such as production of superconducting magnets with higher field strengths[73] ($\text{sensitivity} \propto B_0^{7/4}$) and cryogenic probes[24–26], as well as simply increasing the number of scans ($\text{sensitivity} \propto \sqrt{\text{no. scans}}$). However, few means of achieving better resolution exist beyond increased field strengths (resolution $\propto B_0$). **Perhaps mention improved field homogeneities (shimming)?** Significant interest has therefore been given to the development of techniques which generate broadband homodecoupled (*pure shift*) spectra, in which the effects of homonuclear scalar couplings are absent from the data. While often valuable for structural assignment purposes, their influence can lead to spectra which are too crowded for meaningful insights to be gleamed. Heteronuclear couplings are straightforward to decouple at the point of FID acquisition[74–76]. The removal of homonuclear couplings associated with a specific spin can also be achieved by irradiating said spin with a weak RF field[77: Section 4.2]. Broadband homodecoupling, in which all spins are simultaneously decoupled from each other, is a more challenging task. At the time of writing, there are a number of well-established pure shift experiments, which involve running a 2D pulse sequence, and concatenating the initial section of each FID. in a process referred to as chunking[78–80]. The key drawback of all of these techniques is the hit to sensitivity, a result of the fact that a subset **Does “subset” make sense?** of the available spin

*The major couplings associated with (P) are a geminal coupling to (O), a vicinal coupling to (V) with a dihedral angle of 180°, and another vicinal coupling to (R), with a dihedral angle of 60°. The (O) and (V) couplings are likely of similar magnitude, giving rise to the apparent triplet structure, with the (R) coupling being smaller, based on the Karplus equation CITE. For (V), a geminal coupling to (R), and vicinal couplings to (W) and (P), both with 180° dihedral angles, lead to the apparent quartet structure, with a smaller vicinal coupling to (O).

magnetisation contributes to the signal. At the time of writing, the state-of-the-art pure shift technique is arguably pure shift yeilded by chirp excitation (PSYCHE), both because of its sensitivity, and its ability to cope with strong couplings. **Describe strong couplings if this has not been done before now.** In this section, a method for deriving pure shift spectra indirectly via the estimation of 2DJ datasets is presented, named computer-assisted undiminished-sensitivity protocol for ideal decoupling (CUPID).

In this section, an overview of the key techniques which have been developed to generate pure shift spectra is given, starting with 2DJ spectroscopy and finishing up with PSYCHE, widely considered the most robust pure shift experiment in terms of sensitivity and tolerance to strong coupling. Subsequently, a new technique for generating pure shift spectra via parametric estimation, named computer-assisted undiminished-sensitivity protocol for ideal decoupling (CUPID) is presented.

3.3.1 An Overview of Pure Shift NMR

The 2D J-resolved Experiment

The 2DJ experiment[34, 35] provided the first means of achieving pure shift spectra. It has a simple pulse sequence, presented in Figure 3.4.a: after excitation of magnetisation onto the transverse plane, the indirect dimension evolution consists of a spin echo, with acquisition following immediately afterwards. Fourier transformation in both dimensions leads to a spectrum in which only scalar couplings contribute in $F^{(1)}$, as the chemical shifts are refocussed by the spin echo, while both scalar couplings and chemical shifts contribute in $F^{(2)}$. Peaks belonging to a particular multiplet lie along a line at 45° to the $F^{(1)}$ and $F^{(2)}$ axes, as seen in panel a. of Figure 3.3.

An FID generated by the 2DJ experiment is hypercomplex, taking the form of (1.14) with $D = 2$ and $\zeta = \exp(i\cdot)$, i.e.

$$\mathbf{Y}_{2\text{DJ}} \left[n^{(1)}, n^{(2)} \right] = \sum_{m=0}^{M-1} \boldsymbol{\alpha} [m] \exp(i\phi[m]) \exp \left((2\pi i \mathbf{f}^{(1)} [m] - \boldsymbol{\gamma}^{(1)} [m]) n^{(1)} \Delta_t^{(1)} \right) \times \\ \exp \left((2\pi i (\mathbf{f}^{(2)} [m] - f_{\text{off}}^{(2)}) - \boldsymbol{\gamma}^{(2)} [m]) n^{(2)} \Delta_t^{(2)} \right) + \mathbf{W} [n^{(1)}, n^{(2)}]. \quad (3.1)$$

The transmitter offset term has been neglected in the indirect dimension, since chemical shift evolution does not occur. A major downside of the 2DJ experiment is there is no means of generating a pair of phase- or amplitude-modulated signals which are the conventional route to frequency-discriminated spectra with absorption mode lineshapes, as no mixing time exists in the pulse sequence. The FT of $\mathbf{Y}_{2\text{DJ}}$ produces a spectrum $\mathbf{S}_{2\text{DJ}}$ with phase-twist peaks **Reference chapter 1 when 2D spectrum processing has been described**, which possess contributions from both absorption and dispersion Lorentzians. As with other experiments which produce hypercomplex

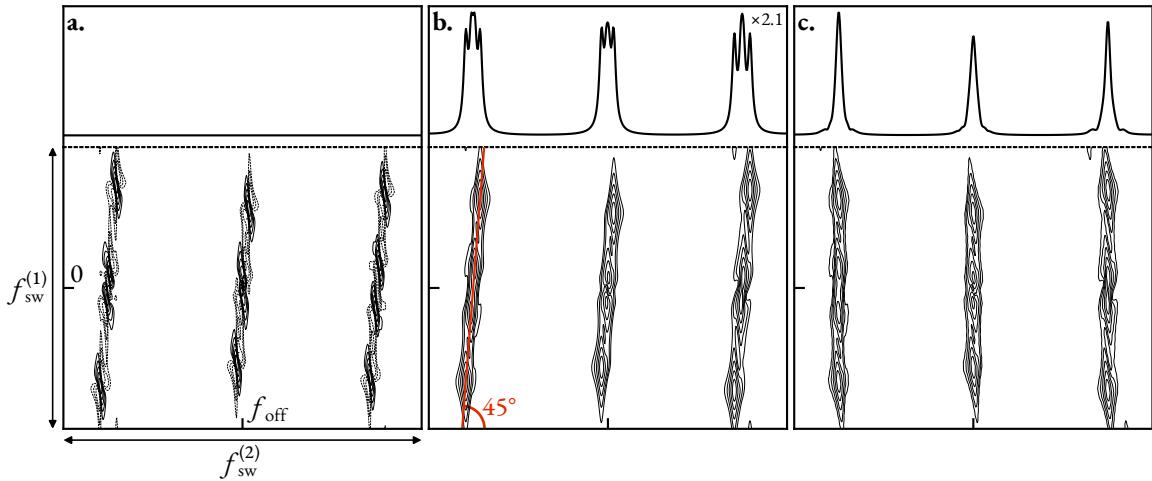


FIGURE 3.3: **a.** Contour plot of an absolute value mode 2DJ spectrum for an AMX spin system, produced by applying sine-bell apodisation on a hypercomplex FID before a FT in both dimensions. Each multiplet lies a long a line at 45° to the $F^{(1)}$ and $F^{(2)}$ axes. Note that this line often appears to make an angle that is greater than 45° with the $F^{(2)}$ axis when viewing spectra, since typically $f_{sw}^{(1)} \ll f_{sw}^{(2)}$. Above the 2DJ spectrum is a plot of its summation along the $F^{(1)}$ axis. **b.** Spectrum generated after application of a 45° shear, with its $F^{(1)}$ summation above.

signals, such as COSY, the data is displayed in magnitude-mode, in which the absolute value of each point in the spectrum is displayed.

There are two primary steps involved in obtaining a pure shift spectrum from $\mathbf{S}_{2\text{DJ}}$:

- (i) Perform a 45° shear (often called a tilt) on the spectrum array, leading to the separation of chemical shifts and scalar couplings onto orthogonal axes (panel b. in Figure 3.3). Each slice in $F^{(2)}$ is subjected to a right circular rotation such that

$$\mathbf{S}_{\text{tilt}} \left[n^{(1)}, n^{(2)} \right] = \mathbf{S}_{2\text{DJ}} \left[n^{(1)}, n_{\text{new}}^{(2)} \right], \quad (3.2\text{a})$$

$$n_{\text{new}}^{(2)} = \left(n^{(2)} + \left\lfloor \frac{f_{sw}^{(1)} N^{(2)}}{f_{sw}^{(2)} N^{(1)}} \left(\frac{N^{(1)}}{2} - n^{(1)} \right) \right\rfloor \right) \bmod N^{(2)}. \quad (3.2\text{b})$$

This achieves the mapping $\mathbf{S}_{2\text{DJ}} \left(f^{(1)}, f^{(2)} \right) \rightarrow \mathbf{S}_{2\text{DJ}} \left(f^{(1)}, f^{(2)} - f^{(1)} \right)$. The effectiveness of the shear is maximised when both $f_{sw}^{(2)}/f_{sw}^{(1)}$ and $N^{(2)}/N^{(1)}$ are powers of 2 **check this**.

- (ii) Sum the spectrum along $F^{(1)}$:

$$\mathbf{s}_{\text{PS}} \left[n^{(2)} \right] = \sum_{n^{(1)}=0}^{N^{(1)}-1} \mathbf{S}_{\text{tilt}} \left[n^{(1)}, n^{(2)} \right]. \quad (3.3)$$

If the spectrum wasn't in magnitude-mode, shearing and summing it would lead to the absorptive and dispersive components of the spectrum cancelling each other out. The process still leads to un-

desirable pure shift spectra with broad “wings” on account of the presence of dispersive character. The dispersive component can be suppressed by appropriate processing to make the FID envelope symmetric in both dimensions, such as with sine-bell apodisation or pseudo-echo reshaping[81], though this results in a significant reduction in sensitivity being incurred.

The Zangger-Sterk Method

Zangger and Sterk introduced a pulse sequence element which achieves *slice-selective excitation*, by applying a low RF power (weak) 180° pulse[†] in the presence of a pulsed field gradient (PFG) along the z -axis[83]. Such an element excites a given spin only in a narrow range of heights in the sample, as the PFG induces a shift in resonance frequency according to $\Delta\omega(z) = \gamma g z$, where g is the magnitude of the PFG. By placing a hard 180° pulse adjacent to the selective pulse, the “active” spin in a given slice is rotated by 360° (i.e. no net rotation), while all other (“passive”) spins are only rotated by 180° . Placing such an element in the middle of the $t^{(1)}$ evolution therefore achieves refocussing of the J-couplings associated with the active spin[84]. In order to achieve effective decoupling of any given pair of spins, it is necessary that the bandwidth of the selective π -pulse is smaller than the difference in their chemical shifts. However, with more selective pulses, a smaller proportion of the available spin magnetisation will contribute to the final FID, and hence sensitivity will be diminished[‡]. Therefore a trade-off exists between effective decoupling of all spins, and achieving the greatest sensitivity possible. In the case of strong coupling, the Zangger-Sterk (ZS) method tends to perform poorly relative to other options for this reason. The ZS element has been applied by Keeler and Pell in order to generate 2DJ datasets comprising phase-modulated pairs, enabling the generation of pure absorption-mode spectra[85], though such spectra suffer from poorer sensitivity than the conventional “full sensitivity” analogue.

The BIRD Method

The bilinear rotation decoupling (BIRD) pulse sequence element[86, 87], presented in Figure 3.4.c, also takes advantage of the idea of selectively inverting passive spins, while leaving active spins unaffected. However the active spins are those which are directly bound to ^{13}C nuclei, which naturally occur with an abundance of 1.1%, while the passive spins are those bound to far more abundant ^{12}C nuclei. The reduction in sensitivity of the experiment relative to a full-sensitivity experiment is therefore known and constant across samples. In scenarios where this strong coupling, BIRD can achieve improved sensitivity over ZS, since with the latter a very weak selective pulse would be required to ensure it is of a sufficiently small bandwidth. The BIRD method is particularly attractive in scenarios where the sensitivity penalty due to the involvement of a low-

[†]Conventionally, a R-SNOB pulse is used[82].

[‡]The reduction in sensitivity is $\propto f_B/\gamma G_z l_z$, where f_B is the selective pulse bandwidth, and l_z is the length of the sample lying within the receiver coil (≈ 1.5 cm).

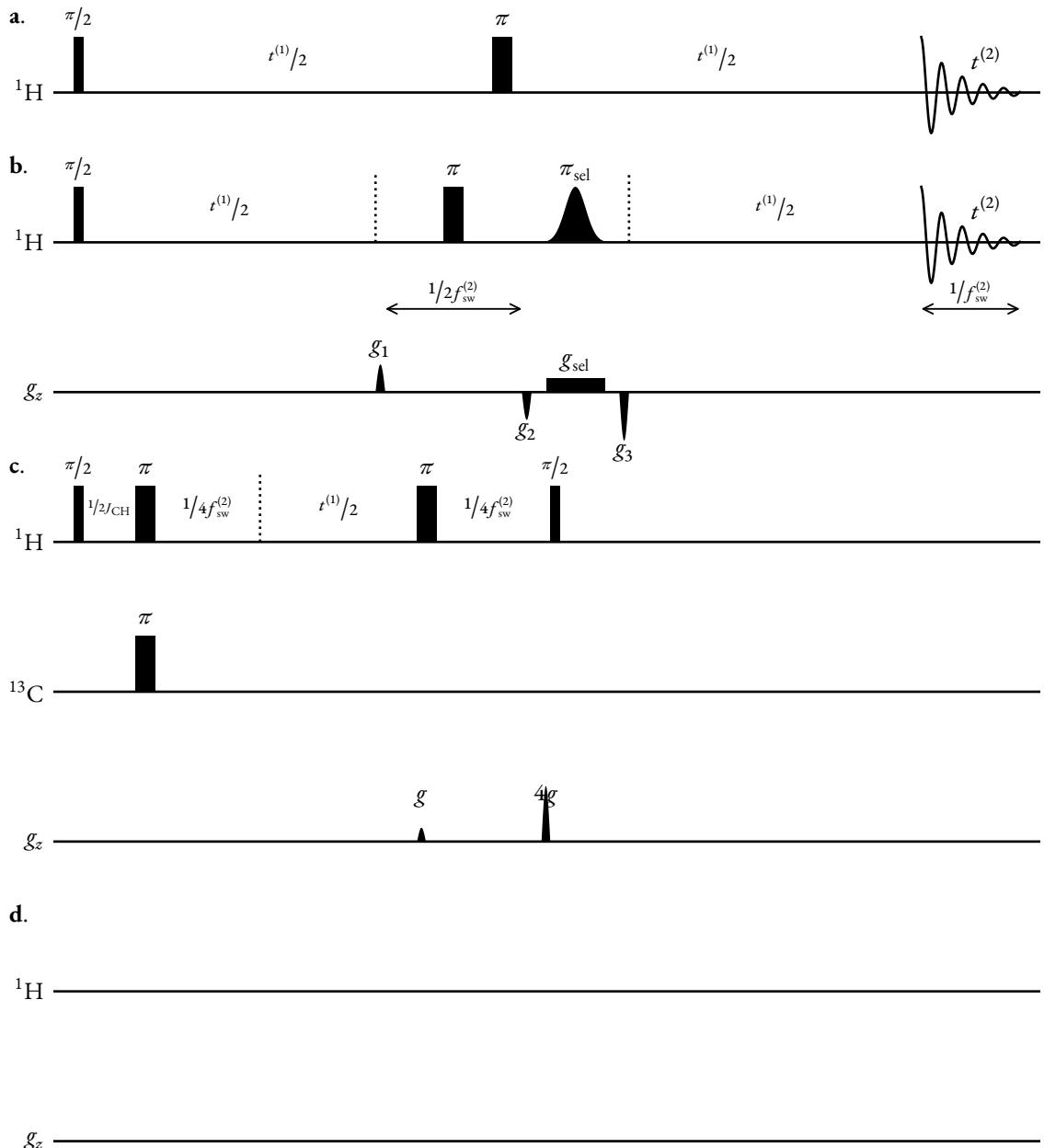


FIGURE 3.4: Work needed. Figure out the correct delays in ZS, BIRD, PSYCHE The pulse sequences of four of the most common pure shift experiments. **a.** 2DJ. **b.** The ZS method. **c.** The BIRD method. **d.** The PSYCHE method.

abundance nucleus has already been paid, for example in sequences where an insensitive nuclei enhancement by polarization transfer (INEPT) block is present.

PSYCHE

TODO Original paper[88], tutorial paper[89], PSYCHE-2DJ[90].

Pure Shift NMR via J-Resolved Post-Processing

There have been previous descriptions of acquiring pure-shift spectra from conventional 2DJ datasets via more sophisticated post-processing methods. Nuzillard introduced a linear predictive estimation of signal time reversal (ALPESTRE)[91, 92], in which the parameters of each indirect-dimension FID are estimated using linear prediction singular value decomposition (LPSVD), such that there is a set of parameters $\boldsymbol{\Theta} \in \mathbb{R}^{N^{(2)} \times 4M}$ with

$$\boldsymbol{\Theta} \left[n^{(2)} \right] = \left[\left[\boldsymbol{\alpha}_{n^{(2)}} \right]^T \quad \left[\boldsymbol{\phi}_{n^{(2)}} \right]^T \quad \left[\boldsymbol{f}_{n^{(2)}}^{(1)} \right]^T \quad \left[\boldsymbol{\eta}_{n^{(2)}}^{(1)} \right]^T \right]^T. \quad (3.4)$$

The parameters generated are used to propagate each FID backward into $-t^{(1)}$, producing a “full-echo”:

$$Y_{\text{full}} \left[n^{(1)}, n^{(2)} \right] = \sum_{m=0}^{M-1} \boldsymbol{\alpha}_{n^{(2)}} [m] \exp \left(i \boldsymbol{\phi}_{n^{(2)}} [m] \right) \exp \left(\left(2\pi i \boldsymbol{f}_{n^{(2)}}^{(1)} [m] n^{(1)} - \boldsymbol{\eta}_{n^{(2)}}^{(1)} [m] |n^{(1)}| \right) \Delta_t^{(1)} \right), \\ \forall n^{(1)} \in \{-N^{(1)} + 1, \dots, 0, \dots, N^{(1)} - 1\}, \forall n^{(2)} \{0, \dots, N^{(2)} - 1\}. \quad (3.5)$$

FT of (3.5) generates a spectrum whose real component comprises absorption-mode Lorentzian character in both dimensions. This opens up the means of producing pure-shift spectra from the 2DJ experiment with sharp lineshapes and without sensitivity loss. A similar approach proposed by Mutzenhardt et al. instead constructed full echoes via LP of each direct-dimension FID, and generation of a full echo by propagating into $-t^{(2)}$ [93].

3.3.2 An outline of CUPID

computer-assisted undiminished-sensitivity protocol for ideal decoupling (CUPID) aims to generate pure shift spectra by utilising the result of parametric estimation of 2DJ data, assumed to take the functional form of (3.1). Instead of estimating successive 1D FIDs, as has been described by Nuzillard and Mutzenhardt et al., the entire 2DJ signal is estimated as a whole, giving access to the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{6M}$. With knowledge of the frequencies and damping factors in both dimensions, it is possible to generate an FID which will produce a pure shift spectrum directly,

rather than constructing a full-echo 2DJ signal, and subsequently shearing and summing it. The desired signal is named the “ -45° signal” (Figure 3.5), and has the form:

$$\begin{aligned} \mathbf{y}_{-45^\circ}(\theta) [n^{(2)}] &= \sum_{m=0}^{M-1} \boldsymbol{\alpha}[m] \exp(i\phi[m]) \\ &\exp\left(\left(2\pi i (\mathbf{f}^{(2)}[m] - \mathbf{f}^{(1)}[m] - f_{\text{off}}^{(2)}) - \boldsymbol{\eta}^{(2)}[m]\right) n^{(2)} \Delta_t^{(2)}\right) \end{aligned} \quad (3.6)$$

$\forall n^{(2)} \in \{0, \dots, N^{(2)} - 1\}$. The -45° signal takes the same functional form as a typical 1D FID acquired by a pulse-acquire experiment, except that the frequency of each oscillator, which would usually be $f^{(2)}$, is replaced with $f^{(2)} - f^{(1)}$. In a 2DJ experiment, $f^{(1)}$ corresponds to the displacement of a given oscillator from the central frequency of the multiplet it is associated with. As such, the oscillators belonging to a given multiplet all provide a contribution with the same frequency to the -45° signal, namely the chemical shift of the relevant spin.

Multiplet Prediction

A holistic 2D estimation of the FID provides access to other useful information about the dataset. One such example is the grouping of oscillators on account of which multiplet structure they belong to. As has already been established, for oscillators which are associated with the same multiplet structure, the quantity $f^{(2)} - f^{(1)}$ should be equal. This provides a criterion in order to assess whether it is likely that two oscillators in the estimation result belong to the same multiplet:

$$|(\mathbf{f}^{(2)}[m_1] - \mathbf{f}^{(1)}[m_1]) - (\mathbf{f}^{(2)}[m_2] - \mathbf{f}^{(1)}[m_2])| < \epsilon \quad (3.7)$$

$\forall m_1, m_2 \in \{0, \dots, M - 1\}$, with $\epsilon \in \mathbb{R}_{>0}$ being a suitable threshold to account for error in the estimation. An appropriate value for ϵ is the spectral resolution in the more poorly resolved dimension, i.e. $\epsilon = \min(f_{\text{sw}}^{(1)}/N^{(1)}, f_{\text{sw}}^{(2)}/N^{(2)})$. In situations involving real 2DJ signals however, it is found that ϵ sometimes has to be increased to values slightly larger than this to achieve effective groupings (*vide infra*). Algorithm A6 provides a routine that can be used for multiplet prediction.

The ability to predict multiplet groupings can also assist in scenarios where the estimation result contains some oscillators with a spurious nature, typically due to overfitting. These typically possess either a very large damping factor or low amplitude, and are not associated with discernible peaks in the spectrum. Part of the reason that the variance of phases is included in the fidelity for NLP is to try and purge these oscillators, however this method is not infallible, and undesired oscillators can end up in the final result. An appreciable number of these can be removed in an automated fashion by noting that there should not be any oscillators in the estimation result of a 2DJ dataset which satisfy both of the following:

- (i) The oscillator is not grouped with any other oscillator as part of the multiplet assignment.

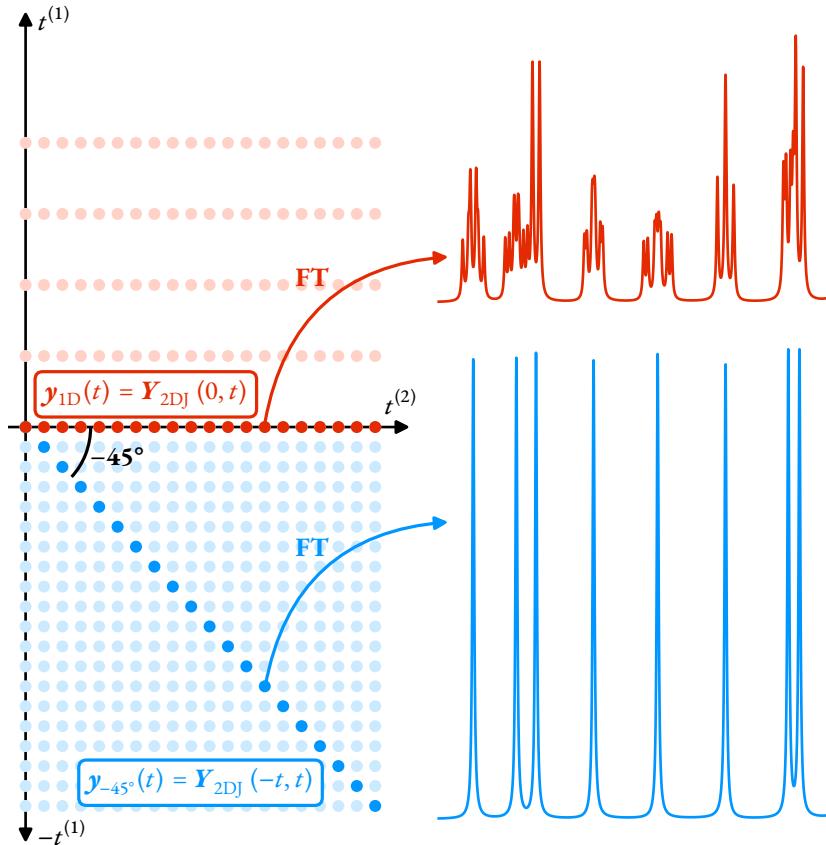


FIGURE 3.5: An illustration of the reasoning behind the name “ -45° signal”. The pale red dots denote a typical 2DJ FID, where the amount and rate of sampling in the direct dimension is greater than in the indirect dimension (i.e. $N^{(1)} \ll N^{(2)}$ and $f_{\text{sw}}^{(1)} \ll f_{\text{sw}}^{(2)}$). The bright red dots correspond to the first direct-dimension signal $\mathbf{Y}_{2\text{DJ}}(0, t^{(2)})$, which has the same form as an FID from a pulse-acquire experiment. A hypothetical signal generated by propagating the FID into $-t^{(1)}$, with the same rate of sampling in both dimensions, is denoted with pale blue dots. Taking the diagonal of this signal, such that it forms a -45° angle to the $t^{(2)}$ axis, yields an FID \mathbf{y}_{-45° which is homodecoupled. Note that there is a slight discrepancy between (3.6) and this description, in that the indirect-dimension damping factors $\eta^{(1)}$ are neglected in the former case.

- (ii) The magnitude of the indirect dimension frequency of the oscillator is appreciably greater than 0 Hz.

These criteria are borne out of the fact that it should not be possible to have oscillators in a 2DJ dataset which are not part of a multiplet structure, unless such oscillators are singlets. As no scalar couplings contribute, these singlets should have an indirect dimension frequency of 0 Hz.

Filtration of 2DJ data

Unlike the direct-dimension, which can often comprise sparsely distributed peaks in the Fourier domain, the indirect dimension of 2DJ datasets tends to be rather densely populated. As such, it is typically of little use in attempting to generate filtered sub-FIDs in the indirect dimension. The filtering procedure utilised for 2DJ data is therefore an extension of the filtration procedure for 1D data described in Section 2.5 (Figure 3.6). The filtering procedure involves the following steps:

- (i) The signal $\mathbf{Y}_{\text{ve}} \in \mathbb{C}^{N^{(1)} \times 2N^{(2)}}$ is constructed, such that a virtual echo is formed from each direct-dimension signal:

$$\mathbf{Y}_{\text{ve}}[n^{(1)}] = \begin{bmatrix} \Re(\mathbf{Y}[n^{(1)}, 0]) & \mathbf{Y}[n^{(1)}, 1] & \dots \\ \mathbf{Y}[n^{(1)}, N^{(2)} - 1] & 0 & \mathbf{Y}[n^{(1)}, N^{(2)} - 1]^* & \dots & \mathbf{Y}[n^{(1)}, 1]^* \end{bmatrix} \quad (3.8)$$

$$\forall n^{(1)} \in \{0, \dots, N^{(1)} - 1\}.$$

- (ii) \mathbf{Y}_{ve} is subjected to FT along the direct dimension to produce the spectrum \mathbf{S}_{ve} (panel a of Figure 3.6). This has a imaginary component of zeros.
- (iii) A super-Gaussian $\mathbf{G} \in \mathbb{R}^{N^{(1)} \times 2N^{(2)}}$ is constructed (panel b):

$$\mathbf{G} = \mathbf{1} \otimes \mathbf{g}^{(2)}, \quad (3.9)$$

where $\mathbf{1} \in \mathbb{R}^{N^{(1)}}$ is a vector of ones, and $\mathbf{g}^{(2)}$ is a super-Gaussian vector given by (2.68b) with $d = 2$.

- (iv) A matrix of additive noise is generated by extracting the variance σ^2 of a strip of \mathbf{S}_{ve} which is devoid of peaks, and generating an array $\mathbf{W}_{\sigma^2} \in \mathbb{R}^{N^{(1)} \times 2N^{(2)}}$ with values independently sampled from a normal distribution with mean 0 and variance σ^2 .
- (v) The spectrum is filtered according to (2.69), yielding $\tilde{\mathbf{S}}_{\text{ve}}$ (panel d).
- (vi) $\tilde{\mathbf{S}}_{\text{ve}}$ is subjected to IFT and is sliced in half in the direct dimension, yeilding the final filtered signal $\tilde{\mathbf{Y}}$:

$$\tilde{\mathbf{Y}} = \text{IFT}^{(2)}(\tilde{\mathbf{S}}_{\text{ve}})[:, :N^{(2)}]. \quad (3.10)$$

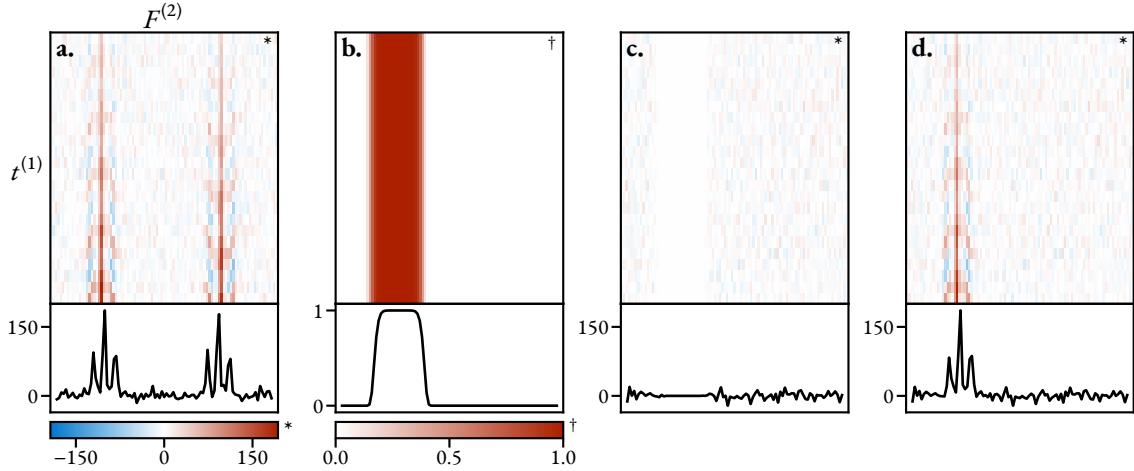


FIGURE 3.6: An illustration of the filtering procedure for 2DJ data. For each panel is a heat-map of the full 2D signal, as well as a plot underneath of the first slice of the signal in the direct dimension. **a.** The spectrum \mathbf{S}_{ve} , **b.** Super-Gaussian filter \mathbf{G} , **c.** Additive noise, attenuated by the super-Gaussian, $\mathbf{W}_{\sigma^2} \odot (\mathbf{1} - \mathbf{G})$, **d.** Filtered spectrum $\tilde{\mathbf{S}}_{\text{ve}}$. Panels **a.–d.** are analogous to panels **b.– e.** in Figure 2.4 for the 1D case.

3.3.3 Results Using CUPID

A number of examples of the application of CUPID are now provided. Initially, a few results are presented using data simulated using the Spinach MATLAB® library[94]. After this, examples are provided with experimental data. In a couple of these, comparison of the result acquired using CUPID is compared with a spectrum acquired using PSYCHE.

For additional details relating to generation of the simulated datasets, see section C.1 in the Appendix.

“Four Multiplets”

A series of simulated ^1H 2DJ datasets were generated such that within a known region of the spectrum, four ddd multiplet structures with significant overlap existed. To achieve this, a spin-system with 7 spins was formed, with the spins divided into 2 subsets:

- ✚ 4 of the spins (the “estimated spins”) were assigned random resonance frequencies sampled from $\mathcal{U}(-20 \text{ Hz}, 20 \text{ Hz})$.
- ✚ The remaining 3 spins (the “coupling spins”), were coupled to each of the estimated spins, with the values of the couplings randomly sampled from $\mathcal{U}(-10 \text{ Hz}, 10 \text{ Hz})$. The coupling spins were given chemical shifts such that they lay far from the estimated spins in the spectrum (i.e. their frequencies were $\gg 20 \text{ Hz}$).

AWGN noise was added to the FID, with a target SNR of 30 dB. A filtered sub-FID containing only the signals from the estimated spins was then generated using the filtering procedure

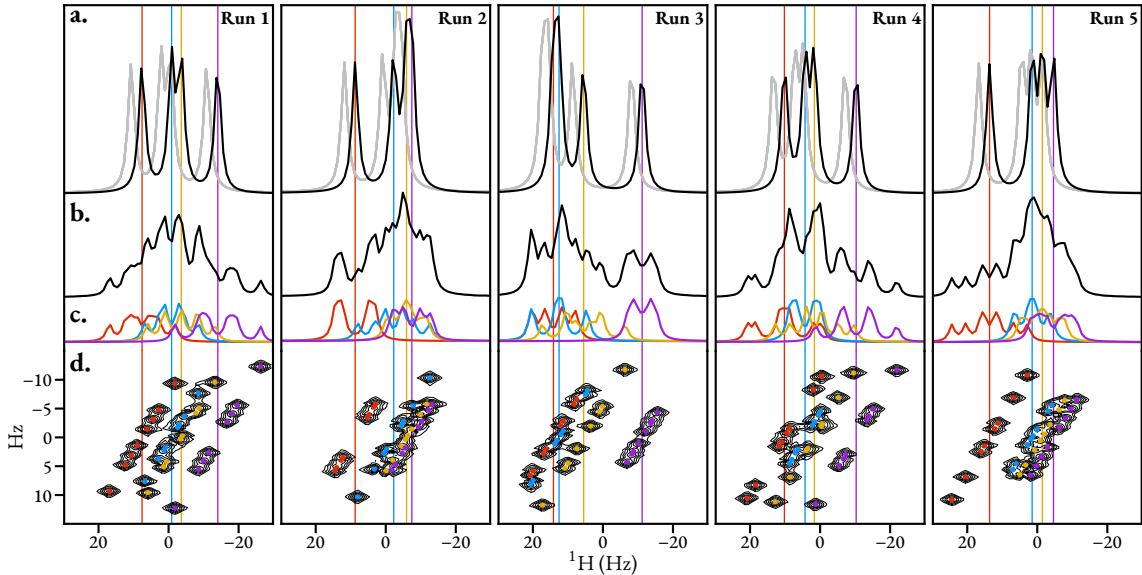


FIGURE 3.7: The result of applying CUPID to 5 instances of simulated 2DJ datasets with 4 heavily overlapping multiplet structures. **a.** Black: pure shift spectrum generated by CUPID. Grey: 1D spectrum simulated with Spinach, using the same spin system as was used to produce the 2DJ dataset, but with all scalar couplings sets to 0 Hz. This has been offset slightly for clarity. **b.** 1D spectrum of the dataset. **c.** Multiplet structures predicted, using a threshold $\epsilon = f_{sw}^{(2)}/N^{(2)} \approx 0.98$ Hz. **d.** Contour plot of the 2DJ spectrum in absolute-value mode. Coloured points denote the frequencies of oscillators in the estimation result. Coloured vertical lines denote the predicted central frequencies of each multiplet structure.

described above, with $l_{\text{Hz}}^{(2)} = 30$ Hz, $r_{\text{Hz}}^{(2)} = -30$ Hz. The resulting sub-FID was expected to comprise 32 (4×2^3) oscillators. To assess the estimation procedure's ability, a random integer from the range 33 – 40 was selected as the initial number of oscillators. Hence, the initial guess from the MMEMPM would comprise an excessive number of oscillators. The FID was subjected to estimation, yielding the result vector $\theta^{(*)}$. Spurious oscillators were checked for, using the criteria outlined above, with the threshold for multiplet assignment set to the spectral resolution in the direct dimension: $\epsilon = f_{sw}^{(2)}/N^{(2)}$. If spurious oscillators were found, these were removed, and NLP was run on the updated set of parameters.

Figure 3.7 illustrates the result achieved for 5 separate runs of this procedure. For each FID generated, the method was effective at producing an estimation result with 32 oscillators, as desired, despite the excessive number that were present in $\theta^{(0)}$. Most of the excessive oscillators were purged from $\theta^{(0)}$ through the NLP procedure. When spurious oscillators did remain[§], they were then detected when checking for spurious oscillators and subsequently removed. The pure-shift spectra generated using CUPID closely agree with pure-shift spectra generated by running a Spinach simulation on a spin system with the same chemical shifts, but with all scalar couplings set to 0 Hz.

[§]for 2 of the 5 datasets, the result after NLP comprised 33 oscillators

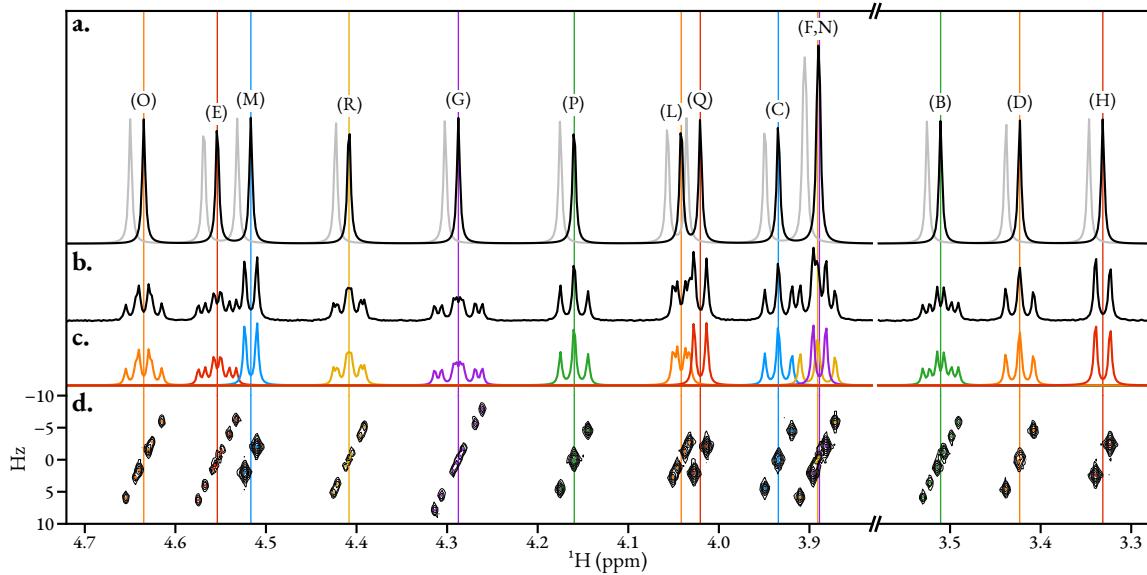


FIGURE 3.8: Application of CUPID on a simulated sucrose 2DJ dataset. **a.** Black: the spectrum generated from FT of the -45° signal. Grey: the spectrum of a simulated dataset with the same chemical shifts, with all scalar couplings set to 0 Hz. **b.** Conventional 1D spectrum. **c.** Multiplet structures assigned ($\epsilon \approx 0.27$ Hz). **d.** Contour plot of the absolute value mode 2DJ spectrum, with the locations of assigned oscillators given as coloured points.

“Sucrose”

As a second example of applying CUPID on simulated data, the chemical shifts and isotropic scalar couplings associated with a Gaussian[95] density functional theory (DFT) calculation of sucrose in a vacuum[¶] were used to construct a 2DJ dataset. AWGN was added with a target SNR of 20 dB. The CUPID procedure was applied to filtered sub-FIDs such that the resonances from all 22 spins were considered, though only the regions of the dataset with the most interesting multiplet structures are presented in Figure 3.8.

The estimation technique successfully assigned multiplet structures for all 22 multiplets in the dataset, including structures derived from two spins (F & N) with a 0.6 Hz difference in resonance frequency, approaching the spectral resolution in the direct dimension (0.537 Hz). The pure-shift spectrum generated via the -45° signal again showed close agreement with a 1D spectrum simulated using the same chemical shifts, with scalar couplings set to 0 Hz. There are particular multiplets where the number of oscillators fit using the estimation routine was less than the true number. Examples of this phenomenon are exhibited in the estimates of the multiplets for spins B & O, which are both ddd structures. The scalar couplings involved meant that certain oscillators were of such similar frequencies that they were separated by significantly less than the spectral

[¶]It is well known that isotropic chemical shift calculations using DFT are typically very inaccurate. The resulting spectrum is not typical of sucrose in the liquid state, though this doesn't really matter for assessing the performance of CUPID.

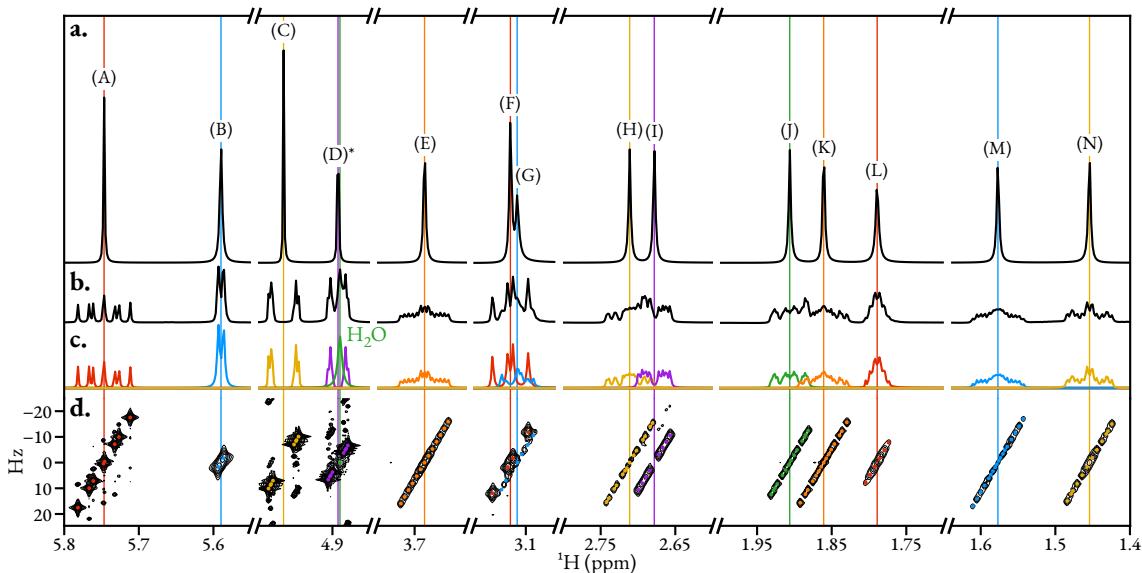


FIGURE 3.9: Application of CUPID on the non-aromatic regions of a quinine 2DJ dataset. **a.** The spectrum generated from FT of the -45° signal, with the green signal arising from water at about 4.89 ppm neglected. **b.** Conventional 1D spectrum. **c.** Multiplet structures assigned ($\epsilon = f_{sw}^{(2)}/N^{(2)} \approx 0.92$ Hz). **d.** Contour plot of the absolute value mode 2DJ spectrum, with the locations of assigned oscillators given as coloured points.

resolution, and thus resolving these was unrealistic. For example, there are two pairs of peaks in the spin-B multiplet which lie only 0.085 Hz apart. Under-fitting in this case had a negligible impact on the final pure shift spectrum. However there are circumstances which will be seen in the experimental examples below where more blatant cases of under-fitting lead to the generation of peaks in the pure shift spectrum which are noticeably broadened.

Quinine

Figure 3.9 illustrates the result of applying CUPID on a dataset generated from a sample comprising quinine in CD₃OD, with all non-aromatic protons considered. The method successfully generated a pure shift spectrum with distinct peaks for each ¹H environment. This example also highlights an added benefit of using CUPID: the ability to suppress unwanted signals in the pure shift spectrum. In this example, an intense, broad singlet at around 4.89 ppm was detected (see the green peak at this frequency in panel c). The singlet was due to the presence of water in the sample and was a hindrance due to it overlapping heavily with the multiplet structure corresponding to spin D. To obtain a clean singlet for spin D in the pure shift spectrum, the oscillator corresponding to the water signal was simply neglected from the parameter set used to generate the -45° signal.

Find reference for work talking about using estimation for solvent suppression.

As eluded to already, a few of the peaks in the pure-shift spectrum are rather broad on account of the estimation routine under-fitting the relevant multiplet structure. The most notable exam-

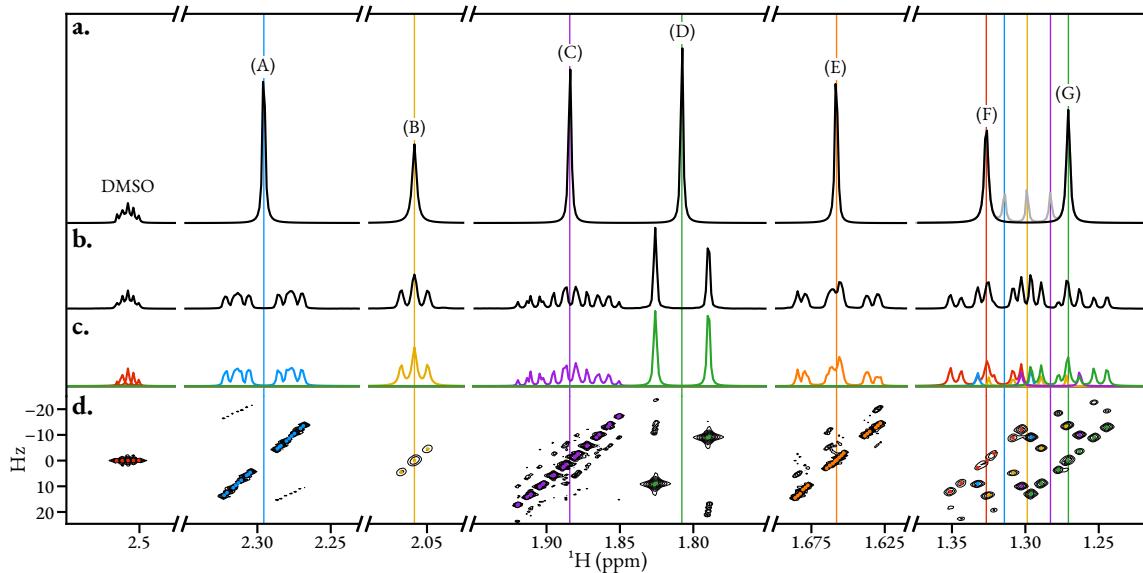


FIGURE 3.10: Application of CUPID on camphor 2DJ dataset. **a.** Black: the spectrum generated from FT of the -45° signal. Oscillators associated with strong coupling artefacts between spins (F) and (G) were neglected. Grey: spectrum generated without neglecting oscillators associated with strong coupling artefacts. **b.** 1D spectrum produced from the first direct-dimension FID in the dataset. Note that, unlike a conventional pulse-acquire spectrum, strong coupling artefacts are present. **c.** Multiplet structures assigned ($\epsilon = 2f_{sw}^{(2)}/N^{(2)} \approx 1.23 \text{ Hz}$). **d.** Contour plot of the absolute value mode 2DJ spectrum, with the locations of assigned oscillators given as coloured points.

ple of this phenomenon in the quinine example comes from the peak for spin G, where significant overlap with spin F has likely compounded the task of accurately estimating the relevant oscillators FID. With fewer than the true number of oscillators fitting a given multiplet structure, the NLP routine will compensate by giving the oscillators it does have at its disposal large amplitudes and damping factors, so that they can reasonably fit multiple similar-frequency oscillators. While affecting peak linewidths, this feature does not tend to significantly affect the integrals of the pure shift peaks. **Should probably calculate the integrals of these...**

Camphor

The application of CUPID to the non-methyl regions of a 2DJ dataset of camphor (Figure A1.c) in DMSO-d₆ is presented in Figure 3.10. As in the quinine case, there are instances of underfitting, which lead to broadening of the resulting pure shift peak, with the peak associated with spin (B) being the most drastic example of this. This example highlights the ability of CUPID to remove undesirable *strong coupling artefacts*^{||} from the final pure shift spectrum, which arise due to mixing effects induced by the 180° pulse in the 2DJ sequence[96, 97]. The effects of strong coupling lead to the presence of extra unexpected peaks which do not agree with the chemical shift of any spin

^{||}As stressed in [96], these are not strictly artefacts, but rather genuine signals, which are expected to be present in the 2DJ dataset. Despite this, the term is widespread in the literature.

associated with camphor. Such effects are observed between roughly 1.35 ppm to 1.25 ppm in the camphor spectrum, where artefacts associated with spins (F) and (G) exist. The estimation routine was able to determine parameters for the more intense oscillators which make up the strong coupling artefacts (these are coloured blue, yellow, and purple, while oscillators associated with the true multiplet structures for (F) and (G) are coloured red and green, respectively). Inclusion of all oscillators extracted by the estimation routine generates the spectrum in panel a, with the low-intensity grey peaks associated with the strong coupling effects included. However, in much the same way as the water signal in the quinine example could be neglected, it is trivial to construct the -45° signal with the oscillators associated with strong coupling neglected, which produces the black spectrum.

Dexamethasone

Double check mp thold

Figure 3.11 shows the result of applying CUPID on a dataset acquired from a sample dexamethasone in DMSO-d₆. A pure-shift spectrum was also acquired using the triple spin echo PSYCHE (TSE-PSYCHE) experiment[89, 98] for comparison. CUPID generated a pure-shift spectrum with overall excellent agreement with the TSE-PSYCHE spectrum. Certain multiplet structures in the spectrum exhibit splitting in F⁽²⁾, on account of heteronuclear couplings to ¹⁹F. Most notable are those derived from spins (D), (H) & (O). For the (D) multiplet, the magnitude of the heterocoupling is very small such that assigning these to separate oscillators is infeasible. For the spin (N) multiplet, two separate structures were successfully assigned (see the orange and green multiplets around 2.1 ppm). The estimation routine was unsuccessful at accurately estimating the structure associated with spin (H), where a severe under-fitting occurred. An under-fitting of this structure even occurred when the estimation was re-run using considerable over-estimation of the model order, with most oscillators in the initial guess being purged during the NLP procedure. The spin (H) multiplet provides an extreme example line-broadening in the pure shift spectrum on account of under-fitting. The most downfield peaks in the CUPID spectrum (corresponding to aromatic and hydroxyl protons) also appear to be noticeably broadened relative to their PSYCHE equivalents. This is also probably due to under-fitting of the relevant multiplet structures, though to a far less noticeable extent than for spin H. **Any other reason why this might be so?**

Estradiol

A final showcase of CUPID is provided by Figure 3.12, where a low concentration (2 mM) sample of 17 β -estradiol (Figure A1.

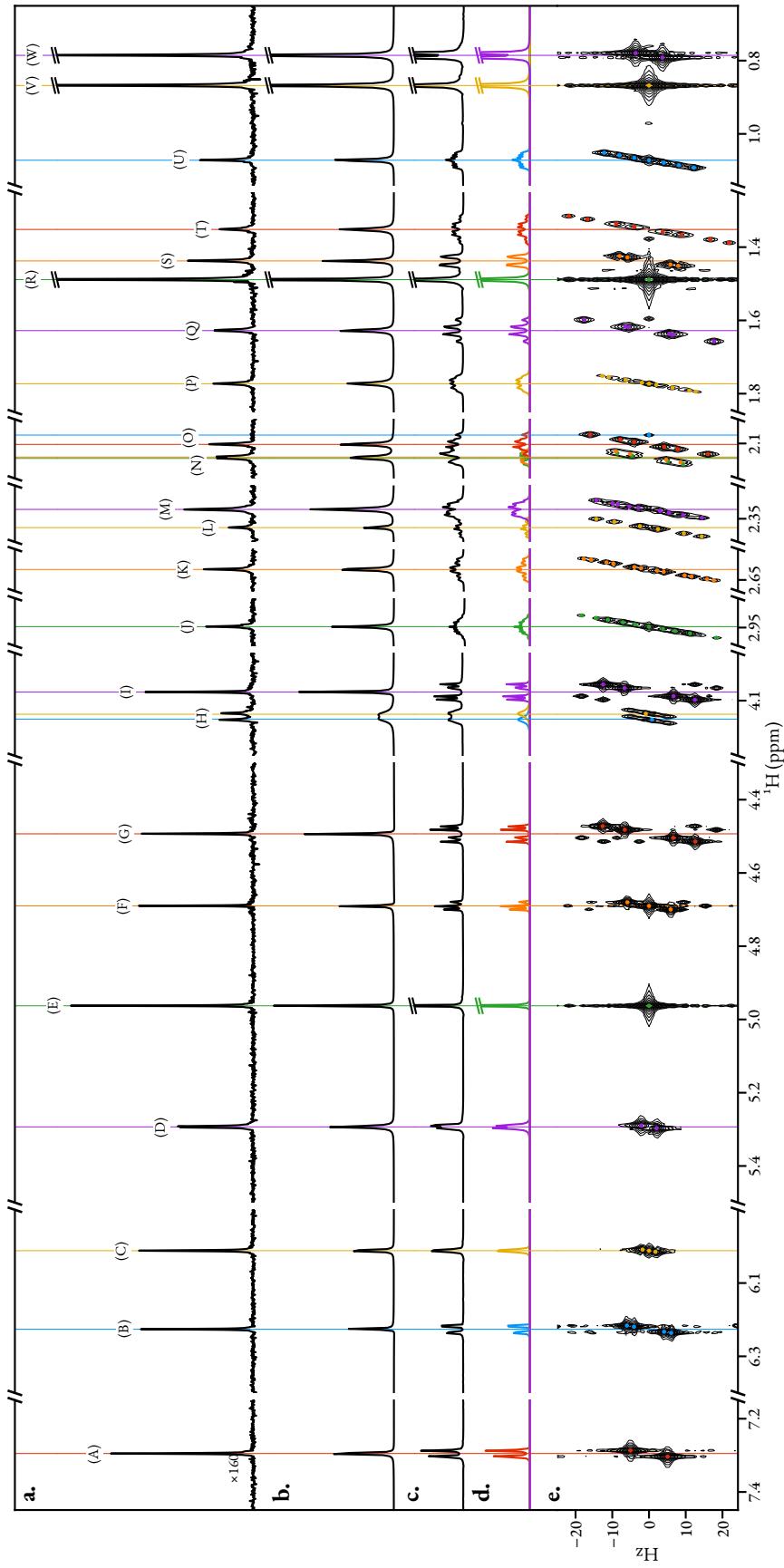


FIGURE 3.II: Fix magnification label. Application of CUPID on dexamethasone 2DJ dataset. **a.** TSE-PSYCHE spectrum of the sample. **b.** The spectrum generated from FT of the -45° signal. **c.** Conventional 1D spectrum. • Multiplet structures assigned ($\epsilon = f_{\text{sw}}^{(2)}/N^{(2)} \approx 0.92$ Hz). **d.** Contour plot of the absolute value mode 2DJ spectrum, with the locations of assigned oscillators given as coloured points.

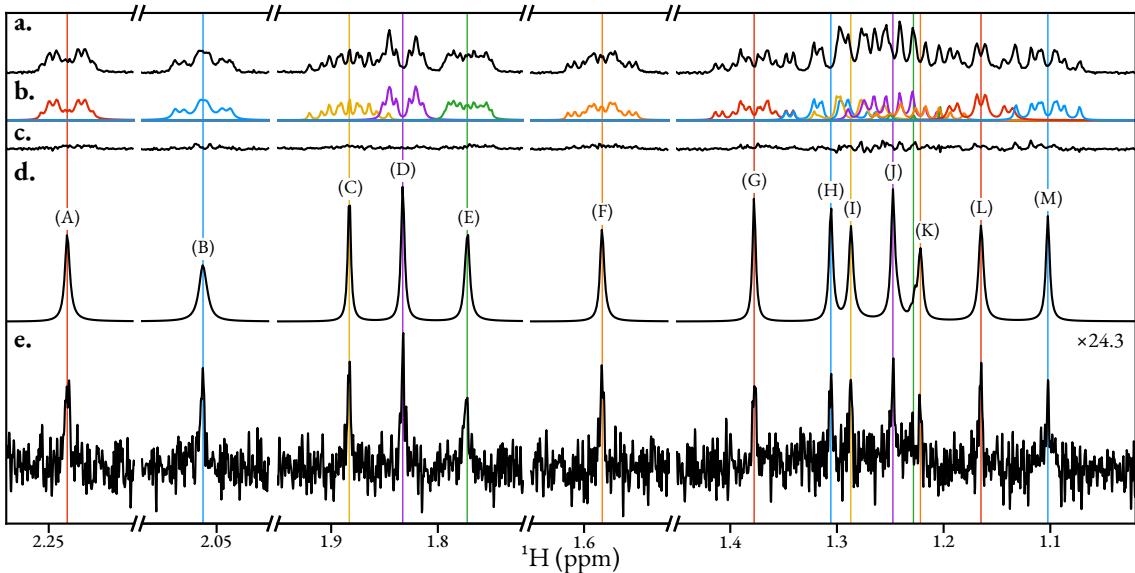


FIGURE 3.12: Application of CUPID on 2DJ dataset of 17β -estradiol in DMSO-d_6 . **a.** Spectrum of the first direct-dimension FID in the 2DJ dataset. **b.** Multiplet structures assigned ($\epsilon = f_{\text{sw}}^{(2)}/N^{(2)} \approx 2 \text{ Hz}$). **c.** The residual between the spectrum in panel a and the lines in panel b. **d.** The pure shift spectrum generated using CUPID. **e.** PSYCHE spectrum of the sample see Figure A2 for details on the pulse sequence. The spectrum has been scaled such that the maximum is of the same magnitude as the corresponding point in the CUPID spectrum.

3.4 Amplitude-attenuated Datasets

There are a number of 2D NMR experiments in which the variation of a parameter in the pulse sequence leads to the generation of FIDs of the same form except for an attenuation in their amplitudes across increments. These include experiments for the determination of translational diffusion rates and relaxation properties such as longitudinal and transverse decoherence rates. Here, an extension to the 1D estimation technique is described, facilitating the determination of these properties.

After a description of relevant experiments, the estimation procedure is p

3.4.1 Relaxation experiments

Integrals of peaks modelled with two parameters $\theta \in \mathbb{R}^2$:

$$\theta_1 = I_\infty = \lim_{\tau \rightarrow \infty} x \quad (3.11a)$$

$$\theta_2 = T_1 \quad (3.11b)$$

$$I(\theta, \tau) = I_\infty \left[1 - 2 \exp \left(-\frac{\tau}{T_1} \right) \right], \quad (3.12)$$

Fitting this function is achieved by minimising L2-norm (refer to it in previous discussion, and refer to grad and Hessian). The first and second derivatives of the model, required to construct the grad and Hess are

$$\frac{\partial I}{\partial I_\infty} = 1 - 2 \exp\left(-\frac{\tau}{T_1}\right) \quad (3.13a)$$

$$\frac{\partial I}{\partial T_1} = -\frac{2I_\infty\tau}{T_1^2} \exp\left(-\frac{\tau}{T_1}\right) \quad (3.13b)$$

$$\frac{\partial^2 I}{\partial I_\infty^2} = 0 \quad (3.13c)$$

$$\frac{\partial^2 I}{\partial I_\infty \partial T_1} = \frac{\partial^2 I}{\partial T_1 \partial I_\infty} = -\frac{2\tau}{T_1^2} \exp\left(-\frac{\tau}{T_1}\right) \quad (3.13d)$$

$$\frac{\partial^2 I}{\partial T_1^2} = \frac{2I_\infty\tau}{T_1^3} \exp\left(-\frac{\tau}{T_1}\right) \left(2 - \frac{\tau}{T_1}\right) \quad (3.13e)$$

3.4.2 Diffusion experiments

NMR is well established as a means of determining the rates of diffusion of chemical species[99, 100]. The first showcase for determining translational diffusion coefficients came from Stejskal and Tanner, in which they described the pulsed gradient spin echo (PGSE) pulse sequence[101] (Figure 3.13.a). The PGSE sequence consists of a conventional spin-echo ($90^\circ \xrightarrow{\tau} 180^\circ \xrightarrow{\tau}$ acquire), with PFGs applied after each of the RF pulses. As a simple overview of how the pulse sequence works, consider a single spin on resonance with the transmitter (i.e. its rotating frame frequency is zero) in a sample tube at position z along the axis collinear with the main field. After the 90° pulse, the magnetisation will be $-M_y$. During the first PFG, the spin's resonance frequency will become $\omega_{\text{PFG}} = -\gamma g z$, where g is the strength of the PFG. Assuming the gradient is applied for a time δ , the spin will precess by an angle of $\alpha = -\gamma g z \delta$. After the 180° pulse, the spin's magnetisation is as follows:

$$-M_y \xrightarrow{\text{PFG}} -M_y \cos(\alpha) + M_x \sin(\alpha) \xrightarrow{180^\circ} -M_y \cos(\alpha) - M_x \sin(\alpha).$$

Supposing that the spin has moved to a new position $z + \Delta_z$ between the end of the first gradient and the beginning of the second, application of the second gradient causes precession by the angle $\beta = -\gamma g(z + \Delta_z)\delta$:

$$\begin{aligned} & \xrightarrow{\text{PFG}} -M_y \cos(\alpha) \cos(\beta) + M_x \cos(\alpha) \sin(\beta) - M_x \sin(\alpha) \cos(\beta) - M_y \sin(\alpha) \sin(\beta) \\ &= -M_y \cos(\gamma g \delta \Delta_z) - M_x \sin(\gamma g \delta \Delta_z), \end{aligned}$$

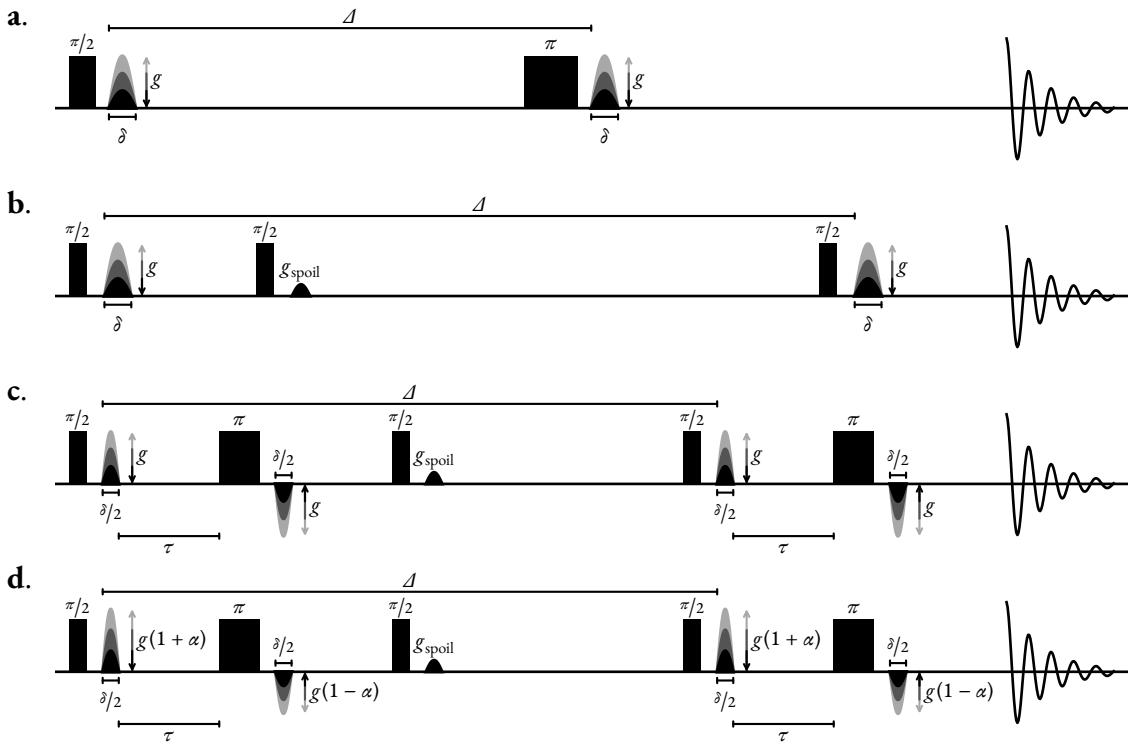


FIGURE 3.13: Pulse sequences used for the determination of translational diffusion constants. **a.** PGSE, **b.** PGSTE, **c.** PGSTEBP, **d.** One-shot DOSY. RF pulses are denoted by solid rectangles. Diffusion-encoding gradients are denoted by sine-bell shapes with varying shades, indicating that the intensity is incremented to create a 2D dataset. Spoiler gradients are denoted by solid black sine-bell shapes.

In the scenario that the spin has not translated in the z -direction between PFGs ($\Delta_z = 0$), the net effect of the pulse sequence is nothing (except for a loss of signal amplitude through T_2 relaxation). However, if translation does occur, the signal phase is adjusted, as a function of the extent of translation Δ_z . The gradients have effectively been employed to encode the change in position of the spin after a known amount of time. Extending this idea to a system of many identical spins, which will translate by different extents between the PFGs, individual spin contributions to the bulk magnetisation will become dephased, leading to an attenuation of the amplitude of the resulting FID.

Through consideration of the Bloch-Torrey equations[102], which extend the classic Bloch equations to account for the effects of diffusion on magnetisation, Stejskal and Tanner were able to derive the following equation for the variation of the amplitude of a resonance as a function of gradient strength, known widely as the Stejskal-Tanner equation:

$$\alpha(g) = \alpha_0 \exp\left(-\gamma^2 \delta^2 g^2 D \left(\Delta - \frac{\delta}{3}\right)\right), \quad (3.14)$$

where $\alpha_0 = \lim_{g \rightarrow 0} \alpha$, γ is the gyromagnetic ratio of the target nucleus (rad MHz T⁻¹), **Should**

check if radians are needed g is the gradient strength (T m^{-1})^{**}, δ is the duration of each PFG (s), Δ is the delay between the PFGs, often known as the diffusion time (s), and D is the translation diffusion constant of the species giving rise to the resonance ($\text{m}^2 \text{s}^{-1}$). While Equation 3.14 is widely stated in the literature, it is only strictly applicable when the PGSE sequence is used (or any other *monopolar* sequence, *vide infra*), and rectangular PFGs are applied ^{††}.

Tanner introduced a variant of the original PGSE experiment called pulsed gradient stimulated echo (PGSTE)[103] (Figure 3.13.b). Instead of the diffusion period including a 180° pulse, PGSTE features two 90° pulses, with the first being applied shortly after the initial PFG, and the second being applied just before the second PFG. The key difference between this and the PGSE experiment is that decoherence during the diffusion time is dictated by longitudinal relaxation (T_1) rather than transverse relaxation (T_2). PGSTE is therefore favoured in scenarios where $T_1 \ll T_2$, as improved sensitivity is attainable.

Both PGSE and PGSTE employ *monopolar* PFGs for diffusion encoding, in the sense that both diffusion-encoding PFGs are polarised in a single direction. Experiments also exist which employ *bipolar* gradient elements[104, 105], which consist of a PFG, followed by a 180° pulse, and then a second PFG with the opposite polarity to the first. A well-known example is the pulsed gradient stimulated echo with bipolar gradients (PGSTEBP) experiment (Figure 3.13.c). Bipolar gradient are useful in circumstances where it is important to purge the effects of static gradients in the sample, caused by field inhomogeneities. Morris and coworkers have also developed the One-shot DOSY experiment[106] (Figure 3.13.d), which requires a single transient per gradient strength (i.e. there is no requirement for a phase-cycling scheme). This is achieved through the use of bipolar gradients which comprise asymmetrical PFGs with relative powers $1 + \alpha : 1 - \alpha$ for some $0 > \alpha > 1$ (a common value is 0.2).

It is virtually always the case that the amplitudes of each resonance in the FID abide by the following general form of the Stejskal-Tanner equation:

$$\alpha(g) = \alpha_0 \exp(-cg^2D) \quad (3.15)$$

for some constant c (T s^{-2}). It is important to note that functional form of c is highly variable dependent on the type of experiment used, and its value is affected by the shape of the diffusion-encoding PFGs. A consideration of the Bloch-Torrey equation for a given experiment is necessary, with an extensive overview provided by Sinnaeve for most diffusion NMR experiments[107]. In general, c is as follows:

$$c = \gamma^2 \delta^2 \sigma^2 \Delta'. \quad (3.16)$$

^{**}Gradient strengths are often expressed in units of G cm^{-1} , which is equivalent to 10^{-2} T m^{-1} .

^{††}Rectangular PFGs (i.e. those in which there is an infinitesimal time to rise to full strength, and to fall back to zero) are in fact impossible to achieve as they would require gradient coils with zero inductance.

σ is the *shape factor* of the PFGs (*vide infra*), and \mathcal{A}' is the effective time that diffusion is allowed to occur. Examples of the value of \mathcal{A}' include:

$$\text{Monopolar gradients (PGSE, PGSTE)} \quad \mathcal{A} + 2(\kappa - \lambda)\delta, \quad (3.17\text{a})$$

$$\text{Bipolar gradients (PGSTEBP)} \quad \mathcal{A} + \frac{(2\kappa - 2\lambda - 1)\delta}{4} - \frac{\tau}{2}, \quad (3.17\text{b})$$

$$\text{One-shot} \quad \mathcal{A} + \frac{(\kappa - \lambda)(\alpha^2 + 1)\delta}{2} + \frac{(\delta + 2\tau)(\alpha^2 - 1)}{4}. \quad (3.17\text{c})$$

τ is the delay between the initial PFG and the 180° pulse in experiments with bipolar gradients. The factors σ , λ , and κ are related to the shape function $s(\epsilon) : \epsilon \in [0, 1]$ of the PFG, which describes the variation in the intensity of the gradient as a function of its progression. For a rectangular gradient, $s(\epsilon) = 1 \forall \epsilon$, whereas for a sine-bell gradient, $s(\epsilon) = \sin(\pi\epsilon)$. The cumulative distribution of the shape function is given by:

$$S(\epsilon) = \int_0^\epsilon s(\epsilon') d\epsilon' \quad \forall \epsilon \in [0, 1]. \quad (3.18)$$

The corresponding definition of S for the case of a gradient made of discrete steps with shape function $s \in \mathbb{R}^{N_g}$ is

$$S[n] = \frac{1}{n} \sum_{i=0}^n s[i] \quad \forall n \in \{0, \dots, N_g - 1\}, \quad (3.19)$$

where N_g is the number of points the gradient comprises. The three factors are given by

$$\sigma = S(1), \quad (3.20\text{a})$$

$$\lambda = \frac{1}{\sigma} \int_0^1 S(\epsilon) d\epsilon, \quad (3.20\text{b})$$

$$\kappa = \frac{1}{\sigma^2} \int_0^1 S^2(\epsilon) d\epsilon, \quad (3.20\text{c})$$

with their discrete counterparts being

$$\sigma = S[N_g - 1] \quad (3.21\text{a})$$

$$\lambda = \frac{1}{\sigma N_g} \sum_{n=0}^{N_g-1} S[n] = \frac{1}{\sigma N_g} \sum_{n=0}^{N_g-1} \frac{1}{n} \sum_{i=0}^n s[i] \quad (3.21\text{b})$$

$$\kappa = \frac{1}{\sigma^2 N_g} \sum_{n=0}^{N_g-1} S^2[n] = \frac{1}{\sigma^2 N_g} \sum_{n=0}^{N_g-1} \frac{1}{n^2} \left(\sum_{i=0}^n s[i] \right)^2 \quad (3.21\text{c})$$

For PFGs with a symmetrical shape, $\lambda = 1/2$. κ is typically equal to or close to $1/3$. It can now be seen that Equation 3.14 comes from plugging Equation 3.17a into Equation 3.16, with $\sigma = 1$, $\lambda = 1/2$, and $\kappa = 1/3$. In many situations, \mathcal{A} dominates in the expression of \mathcal{A}' , and so ensuring the

correct form of c could be seen as excessive. However, especially when \mathcal{A} is not orders of magnitude greater than δ , the exact form of \mathcal{A}' used in Equation 3.16 will be extremely important for accurate measurements of D .

3.4.3 Methodology

Give algorithm for fit of each oscillator. Include how initial guess is generated for invrec and diffusion, and initial trust radius.

Outline of the problem

All the experiments described above can be described in a general fashion. Suppose the experiment is run with $K \in \mathbb{N}$ increments, such that there is a vector $\mathbf{p} \in \mathbb{R}^K$ which gives the experimental variable for each increment. The complete FID that results from the experiment is expected to take the form $\mathbf{Y} \in \mathbb{C}^{K \times N^{(1)}}$ in which the value of the experimental parameter attenuates the amplitudes of the contributing resonances:

$$\mathbf{Y}[k, n^{(1)}] = \sum_{m=0}^{M-1} \mathbf{A}[k, m] \exp(i\phi[m]) \exp\left(\left(2\pi i \left(\mathbf{f}^{(1)}[m] - f_{\text{off}}^{(1)}\right) - \boldsymbol{\eta}^{(1)}[m]\right) n^{(1)} \Delta_t^{(1)}\right). \quad (3.22)$$

$\mathbf{A} \in \mathbb{R}^{K \times M}$ is a matrix of the amplitudes of each oscillator across each increment. A complete parameter vector for this signal is given by $\theta \in \mathbb{R}^{(K+3)M}$:

$$\theta = \begin{bmatrix} \mathbf{A}[0, :]^\top & \dots & \mathbf{A}[K-1, :]^\top & \boldsymbol{\phi}^\top & [\mathbf{f}^{(1)}]^\top & [\boldsymbol{\eta}^{(1)}]^\top \end{bmatrix}^\top. \quad (3.23)$$

The amplitudes are a function of the experiment parameter with the following form:

$$\mathbf{A}[k, m] = \mathbf{a}_0[m] \mathcal{A}(\psi[m] | \mathbf{p}[k]), \quad (3.24)$$

where $\mathbf{a}_0 \in \mathbb{R}^M$ is a vector of the “maximal” amplitude for each oscillator, and $\psi \in \mathbb{R}^M$ is a vector of the parameter of interest (diffusion coefficient, relaxation time etc.) for each oscillator. $\mathbf{a}_0[m]$ can be thought of as the largest possible amplitude that could be obtained for oscillator m with the given experiment:

- ❖ In an inversion recovery experiment, the largest amplitude is achieved as $\tau \rightarrow \infty$, as the spin system will have returned back to equilibrium prior to the 90° pulse.
- ❖ With Carl-Purcell-Meiboom-Gill (CPMG) experiments, the greatest amplitude will occur when $n_{\text{cycles}} = 0$, as no time is designated for transverse relaxation to take place.

Experiment	p	ψ	$\mathcal{A}(\psi p)$	$\frac{\partial \mathcal{A}(\psi p)}{\partial \psi}$	$\frac{\partial^2 \mathcal{A}(\psi p)}{\partial \psi^2}$
Inversion Recovery	τ	T_1	$(1 - 2 \exp(-\frac{\tau}{T_1}))$	$-\frac{2\tau}{T_1^2} \exp(-\frac{\tau}{T_1})$	$\frac{2\tau}{T_1^3} \exp(-\frac{\tau}{T_1}) (2 - \frac{\tau}{T_1})$
CPMG	n_{cycles}	T_2	?	?	?
Diffusion	g	D	$\exp(-cg^2D)$	$-cg^2 \exp(-cg^2D)$	$c^2 g^4 \exp(-cg^2D)$

TABLE 3.1: The various functional forms of \mathcal{A} according to the different amplitude-attenuating NMR experiments considered, along with its first and second derivatives, which are required to extract estimates of ψ using NLP.

- For diffusion experiments, the largest amplitude is achieved when $g = 0$, since no diffusion-induced dephasing of spins in different positions along the z -axis will have occurred.

The function \mathcal{A} describes how the amplitudes of resonances are attenuated by the experimental variable, and has a form which is intimately linked to the type of experiment. See Table 3.1 for specific examples.

Estimating amplitude-attenuated datasets

The close relationship between signals across increments means that completely estimating each signal in turn is not necessary. Instead, after the first increment is estimated from scratch, yielding $\theta_0 \in \mathbb{R}^{4M}$, the phases, frequencies and damping factors are fixed, and subsequent increments are determined by taking the parameter estimate of the previous iteration θ_{k-1} and subjecting it to NLP where only the amplitudes are allowed to be varied. Thus, determining parameter estimates for each iteration where $k \neq 0$ is reduced to the problem^{‡‡}

$$\boldsymbol{\alpha}[k,:] = \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^M} \mathcal{F}(\boldsymbol{\alpha} | \boldsymbol{\phi}, \boldsymbol{f}^{(1)}, \boldsymbol{\gamma}^{(1)}, \mathbf{Y}[k,:]). \quad (3.25)$$

A NLP routine can solve this very efficiently, typically in few iterations, on account of the linear dependence of the model with respect to the oscillator amplitudes. The linear dependence also means that second derivatives of the model are all zero (see (2.45a)), such that only first derivatives need to be computed to derive an exact Hessian matrix of the fidelity.

Remark 4. Due to the linear dependence on amplitudes, an alternative means of deriving ampli-

^{‡‡} \mathcal{F} in (3.25) reads as “the fidelity with respect to the amplitudes $\boldsymbol{\alpha}$, given phases $\boldsymbol{\phi}$, frequencies $\boldsymbol{f}^{(1)}$, damping factors $\boldsymbol{\gamma}^{(1)}$, and FID $\mathbf{Y}[k,:]$ ”. The expression has exactly the same mathematical form as (2.42), though it emphasises that the phases, frequencies and damping factors are no longer variables to be optimised, but fixed parameters.

Algorithm 4 Routine for estimating a sequence of 1D FIDs which exhibit variation in amplitudes across increments. NLPAMP denotes a routine which is akin to NLP (Algorithm 2), except only amplitudes are allowed to be altered, whilst phases, frequencies and damping factors are fixed.

```

1: procedure ESTIMATEAMPATTENUATED(  $\mathbf{Y} \in \mathbb{C}^{K \times N^{(1)}}$ ,  $\mathbf{r}_{\text{interest}} \in \mathbb{R}^2$ ,  $\mathbf{r}_{\text{noise}} \in \mathbb{R}^2$ ,  $M \in \mathbb{N}_0$  )
2:    $\theta_0, \epsilon_0 \leftarrow \text{ESTIMATE1D}(\mathbf{Y}[0, :], \mathbf{r}_{\text{interest}}, \mathbf{r}_{\text{noise}}, M)$ ;
3:    $M \leftarrow \text{len}(\theta_0)/4$ ;
4:    $\theta, \epsilon \leftarrow \mathbf{0} \in \mathbb{R}^{(K+3)M}, \mathbf{0} \in \mathbb{R}^{(K+3)M}$ ; ▷ Initialise complete parameter vector.
5:    $\theta[:M], \epsilon[:M] \leftarrow \theta_0[:M], \epsilon_0[:M]$ ; ▷ Amplitudes for first increment.
6:    $\theta[KM :], \epsilon[KM :] \leftarrow \theta_0[M :], \epsilon_0[M :]$ ; ▷ Phases, frequencies and damping factors, which are held constant across increments.
7:   for  $k = 1, \dots, K - 1$  do
8:      $\tilde{\mathbf{y}} \leftarrow \text{FILTER1D}(\mathbf{Y}[k, :], \mathbf{r}_{\text{interest}}, \mathbf{r}_{\text{noise}})$ ; ▷ Algorithm A3.
9:      $\theta_k, \epsilon_k \leftarrow \text{NLPAMP}(\tilde{\mathbf{y}}, \theta_{k-1})$ ;
10:     $\theta[kM : (k + 1)M], \epsilon[kM : (k + 1)M] \leftarrow \theta_k[:M], \epsilon_k[:M]$ ; ▷ Extract amplitudes.
11:   end for
12:   return  $\theta, \epsilon$ ;
13: end procedure

```

tudes for each increment is to determine the following $\forall k \in \{1, \dots, K - 1\}$:

$$\mathbf{A}[k, :] = \mathbf{A}^+ \mathbf{Y}[k, :], \quad (3.26a)$$

$$\mathbf{A} = \begin{bmatrix} \exp(i\phi[0]) & \cdots & \exp(i\phi[M-1]) \\ \exp(i\phi[0]) \mathbf{z}[0] & \cdots & \exp(i\phi[M-1]) \mathbf{z}[M-1] \\ \vdots & \ddots & \vdots \\ \exp(i\phi[0]) \mathbf{z}[0]^{N^{(1)}-1} & \cdots & \exp(i\phi[M-1]) \mathbf{z}[M-1]^{N^{(1)}-1} \end{bmatrix}, \quad (3.26b)$$

$$\mathbf{z}[m] = \exp\left(\left(2\pi i \left(f^{(1)}[m] - f_{\text{off}}^{(1)}\right) - \boldsymbol{\eta}^{(1)}[m]\right) \Delta_t^{(1)}\right). \quad (3.26c)$$

Determining the parameter of interest

Having generated a complete parameter estimate for the FID, focus subsequently moves to determining the parameters of interest ψ . For each oscillator, the maximal amplitude and parameter of interest are determined by solving the following problem:

$$\begin{bmatrix} \alpha_0^{(*)}[m] \\ \psi^{(*)}[m] \end{bmatrix} = \arg \min_{[\alpha_0 \ \psi]^T \in \mathbb{R}^2} \|\mathbf{A}[:, m] - \alpha_0 \mathcal{A}(\psi | \mathbf{p})\|^2, \quad (3.27)$$

which can also be solved using an NLP routine. The gradient vector and Hessian matrix of the fidelity take very similar functional forms to those for FID estimation (see (2.43a) and (2.43b)), as both problems involve solving residual sum-of-squares problems. They are as follows $\forall i, j \in$

$\{0, 1\}$:

$$\mathbf{g}[i] = -2 \left(\mathbf{A}[:, m] - \alpha_0 \mathcal{A}(\psi | \mathbf{p}), \frac{\partial \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \vartheta[i]} \right) \quad (3.28a)$$

$$\mathbf{H}[i, j] = 2 \left(\left(\frac{\partial \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \vartheta[i]}, \frac{\partial \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \vartheta[j]} \right) - \left(\mathbf{A}[:, m] - \alpha_0 \mathcal{A}(\psi | \mathbf{p}), \frac{\partial^2 \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \vartheta[i] \partial \vartheta[j]} \right) \right), \quad (3.28b)$$

$$\mathbb{R}^2 \ni \boldsymbol{\vartheta} = [\alpha_0 \ \psi]^T, \quad (3.28c)$$

with explicit expressions for the requisite first and second derivatives being

$$\frac{\partial \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \alpha_0} = \mathcal{A}(\psi | \mathbf{p}), \quad (3.29a)$$

$$\frac{\partial \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \psi} = \alpha_0 \frac{\partial \mathcal{A}(\psi | \mathbf{p})}{\partial \psi}, \quad (3.29b)$$

$$\frac{\partial^2 \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \alpha_0^2} = 0, \quad (3.29c)$$

$$\frac{\partial^2 \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \psi^2} = \alpha_0 \frac{\partial^2 \mathcal{A}(\psi | \mathbf{p})}{\partial \psi^2}, \quad (3.29d)$$

$$\frac{\partial^2 \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \alpha_0 \partial \psi} = \frac{\partial^2 \alpha_0 \mathcal{A}(\psi | \mathbf{p})}{\partial \psi \partial \alpha_0} = \frac{\partial \mathcal{A}(\psi | \mathbf{p})}{\partial \psi}. \quad (3.29e)$$

The functional forms of the first and second derivatives for the different experiments of interest of \mathcal{A} are given in Table 3.1.

Displaying results

Visualising the results from the routine described above can be done in a similar fashion to diffusion-ordered spectroscopy (DOSY) analysis. For each oscillator in the estimation result, a 2D array is generated, corresponding to the outer product of the FT of the oscillator (\mathbf{s}_m) and a distribution describing the predicted value of ψ (\mathbf{d}_m).

$$\mathbb{R}^{R \times N^{(1)}} \ni \mathbf{S} = \sum_{m=0}^{M-1} \mathbf{d}_m \otimes \mathbf{s}_m, \quad (3.30a)$$

$$\mathbf{s}_m = \Re(\text{FT}(\mathbf{x}_m)), \quad (3.30b)$$

$$\mathbf{x}_m \left[n^{(1)} \right] = \alpha_0[m] \exp(i\phi[m]) \exp \left(\left(2\pi i \left(\mathbf{f}^{(1)}[m] - f_{\text{off}}^{(1)} \right) - \eta_m^{(1)} \right) n^{(1)} \Delta_t^{(1)} \right), \quad (3.30c)$$

where $R \in \mathbb{N}$ is the number of samples to generate the distribution from. The exact form that the distribution should take is not set in stone, though it should indicate two key pieces of information: (i) its maximum should coincide with the predicted value of ψ and (ii) its width should indicate the certainty in the prediction. In this work, the distribution used is a Gaussian distribu-

tion with mean ψ and standard deviation $c\epsilon_\psi$, where ϵ_ψ is the estimation error associated with the parameter of interest,^{§§} and $c \in \mathbb{R}_{>0}$ is an arbitrary linewidth factor which can be chosen to ensure clear visibility of the peaks.^{¶¶}

$$\mathbf{d}_m[r] = -\frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{(p_r - \psi[m])^2}{2\sigma_m^2}\right) \quad \forall r \in \{0, \dots, R-1\}, \quad (3.31a)$$

$$p_r = p_{\min} + \frac{r(p_{\max} - p_{\min})}{R-1}, \quad (3.31b)$$

$$\sigma_m = c\epsilon[m]. \quad (3.31c)$$

p_{\min} and p_{\max} specify the range of values over which to generate the distribution.

3.4.4 Results

“Five multiplets”

Figure 3.14 shows the result of the described method in determining the T_1 values of resonances in a simulated inversion recovery datasets featuring 5 overlapping ddd multiplet structures. The spin systems used to generate the datasets were constructed in a similar way to that described for the “Four Multiplets” example in Section 3.3.3. In this case however, 5 estimated spins were present instead of 4. Some constraints were also placed on the shifts and couplings to ensure that no two oscillators would have frequencies with a difference less than $f_{\text{sw}}^{(1)}/N^{(1)}$. For each spin, a T_1 value was sampled from $\mathcal{U}(1 \text{ s}, 5 \text{ s})$, and a T_2 value was sampled from $\mathcal{U}(0.2 \text{ s}, 0.6 \text{ s})$. The inversion recovery experiment was simulated using SPINACH, with the relaxation phenomena described by the “Extended T_1/T_2 approximation”***. Each dataset was provided AWGN such that the target SNR of the datasets as a whole was 40 dB. The estimation routine outlined by Algorithm 4 was applied to the generate a parameter estimate of the 1.6 ppm to 1.2 ppm region. Despite heavy overlap between peaks, the routine was high successful at assigning each resonance with a T_1 value that agreed with the true value (see panel b.). As is to be expected, in scenarios where little multiplet overlap existed, T_1 predictions tended to be more accurate, with smaller associated errors (see for example the purple and orange multiplets in Run 1). Nevertheless, adequate estimates could still be obtained in cases of severe overlap, especially when the predicted T_1 s for all oscillators associated with a given multiplet are averaged. (see the red, blue and yellow multiplets in Run 1). Particular

^{§§}The errors can be extracted from the Hessian matrix once the NLP routine has reached convergence. See section 2.3.5 for more information.

^{¶¶}In cases where the estimation error is small, the distribution can be so sharp relative to the resolution used for plotting, that $\max(\mathbf{d}_m)$ can be 0. In these situations, errors arise due to attempted division by 0 (see (3.30a)). Broadening the distribution (i.e. setting $c > 1$) can resolve issues with plotting discrete samples.

^{***}As described in the Spinach documentation[108], longitudinal spin states are assigned the rate $1/T_1$ and transverse states are assigned the rate $1/T_2$. Multi-spin states are assigned rates which are the summation of each spin’s relevant rate.

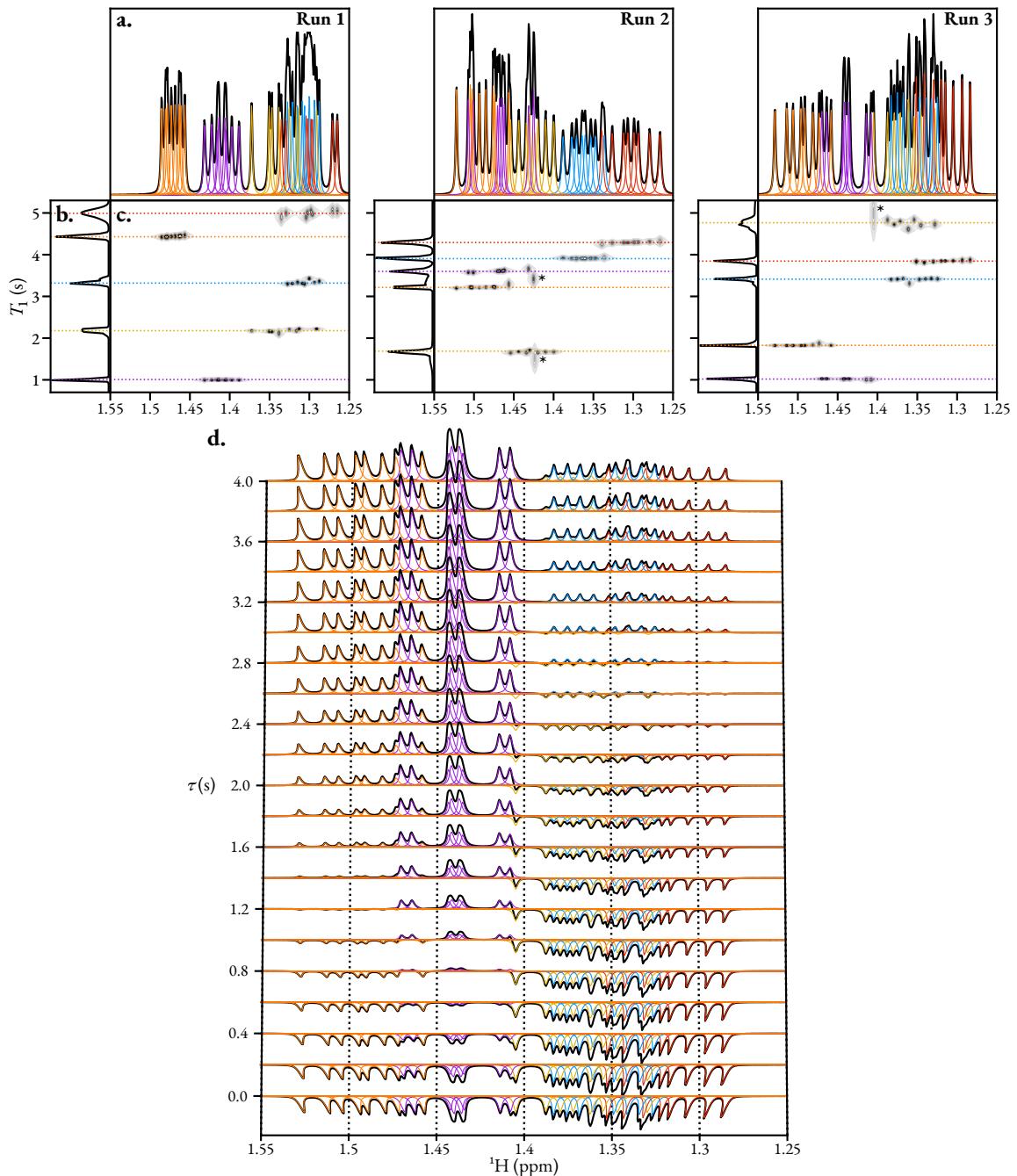


FIGURE 3.14: Three examples of results generated on simulated inversion recovery datasets comprising five ddd multiplet structures. **a.** Plot of the result generated for the first increment ($\tau = 0$ s), with each plot multiplied by -1 . Black: spectrum of the data. Coloured lines: spectra of individual oscillators generated by the estimation routine. Oscillators with the same colour are components of the same multiplet. **b.** Distribution of T_1 values, generated using (3.31), with $p_{\min} = 0.7$ s, $p_{\max} = 5.3$ s, $c = 40$, $R = 128$. **c.** DOSY-style contour plot of the result, generated using (3.30). Dashed horizontal lines denote the true T_1 values for each spin. **c.** Estimation result for each increment for Run 3, illustrating the evolution of the amplitudes of each oscillator with τ .

oscillators for which the estimate of T_1 is particularly far from the true value are often reassuringly associated with large errors, with examples denoted with an asterisk.

Andrographolide Diffusion

Figure 3.15 shows the result of applying the estimation technique on a oneshot DOSY dataset of andrographolide in unfresh DMSO-d₆ at 298 K. Exposure of the sample to water is evidenced by the broad peak around 3.3 ppm, estimated to have a diffusion constant of $4.57 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$. On top of this, the acidic hydroxyl protons (B, C, H) of andrographolide show significant line-broadening, and their estimated diffusion coefficients are considerably different compared with those of the non-hydroxyl protons, due chemical exchange with water in the sample[109]. The diffusion profile generated suggests a diffusion constant of andrographolide of $2.54 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$. The predicted diffusion constant for each estimated oscillator shows decent consistency, especially with oscillators of greater intensity. Lower intensity oscillators - especially those which significantly overlap with other oscillators - tended to be associated with less consistent diffusion constants and larger errors with examples of this phenomenon apparent in panel d of the figure. A few oscillators also show a significant deviation at around 2.5 ppm. This is likely due to the presence of resonances from partially protonated DMSO. As the data is insufficiently resolved to enable the separation of andrographolide and Dimethyl sulfoxide, (H₃C)₂SO (DMSO) resonances in the estimation result, estimated oscillators which significantly overlap with DMSO and andrographolide resonances will have an amplitude profile influenced by both species, leading to an aggregated diffusion constant. As $D_{\text{DMSO}} > D_{\text{andrographolide}}$, the affected oscillators show larger apparent diffusion constants.

Include result using TopSpin? Get similar value if i set sigma to 1 (it is 0.9)

3.5 Phased broadband spectra from chirp excitation

The are numerous nuclei of value in NMR with very wide chemical shift ranges, including ¹³C, ¹⁹F (of particular interest in the pharmaceutical industry) ³¹P, ¹⁹⁵Pt. Attaining spectra covering the entire chemical shift range of such spins for use in quantitative applications is challenging due to off-resonance effects, which severely alter the amplitudes and phases of resonances with frequencies far from the transmitter frequency[22: Section 3.4.1]. One popular means of achieving broadband excitation, in which a consistent amplitude- and phase-profile across a spectral window of tens or even hundreds of kHz is achieved, is to use frequency-swept (FS) pulses, during which the frequency of RF irradiation varies with time[110]. One of the most common classes of FS pulses are those where the variation of frequency with time is linear, with such pulses commonly referred to as *chirp* pulses. The application of a single 90° chirp pulse to achieve broadband exci-

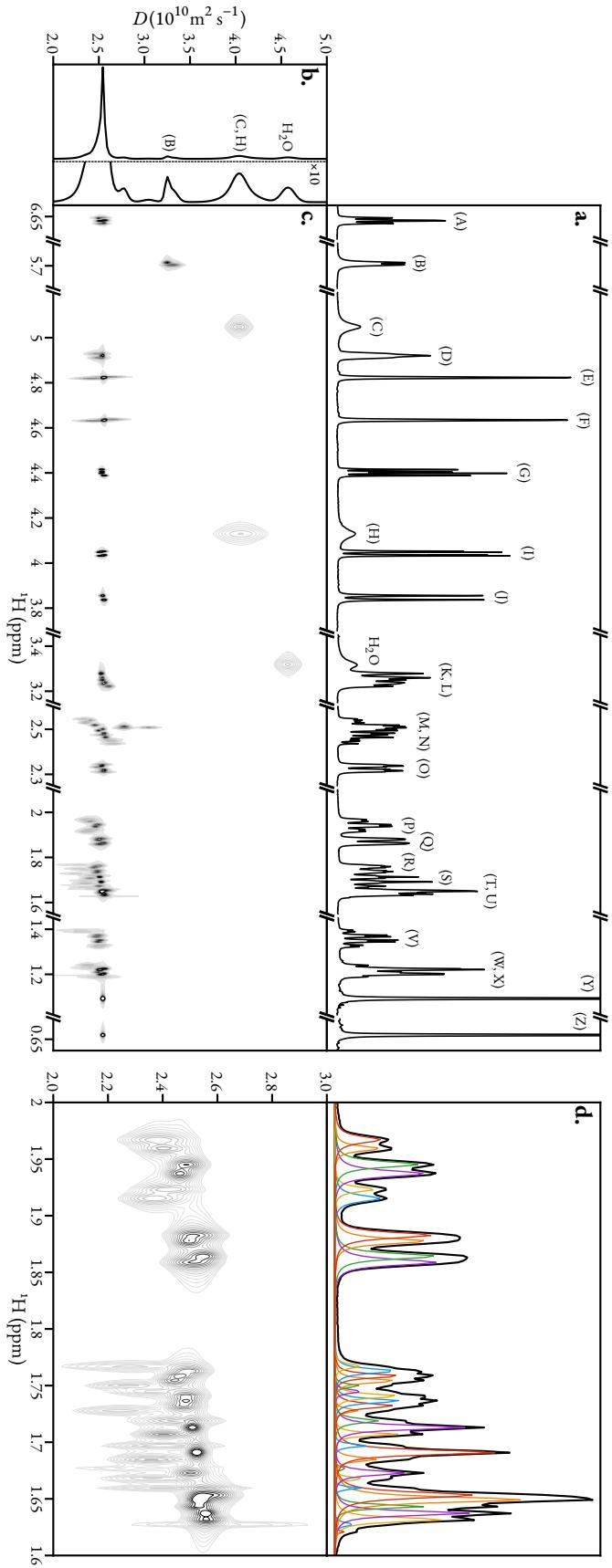


FIGURE 3.15: Result of estimating a Oneshot DOSY dataset of andrographolide in unfresh Deuterated DMSO (DMSO-d₆). **a.** 1D spectrum. **b.** Diffusion profile obtained by summing the contour plot in **c.** along the *x*-axis. **c.** Contour plot mapping estimated oscillators to diffusion constants, with $p_{\min} = 2 \times 10^{-10} \text{m}^2 \text{s}^{-1}$, $p_{\max} = 5 \times 10^{-10} \text{m}^2 \text{s}^{-1}$, $c = 2.5$, $R = 128$. **d.** Magnified view of the 2 ppm to 1.6 ppm spectral range, with estimated oscillator peaks plotted.

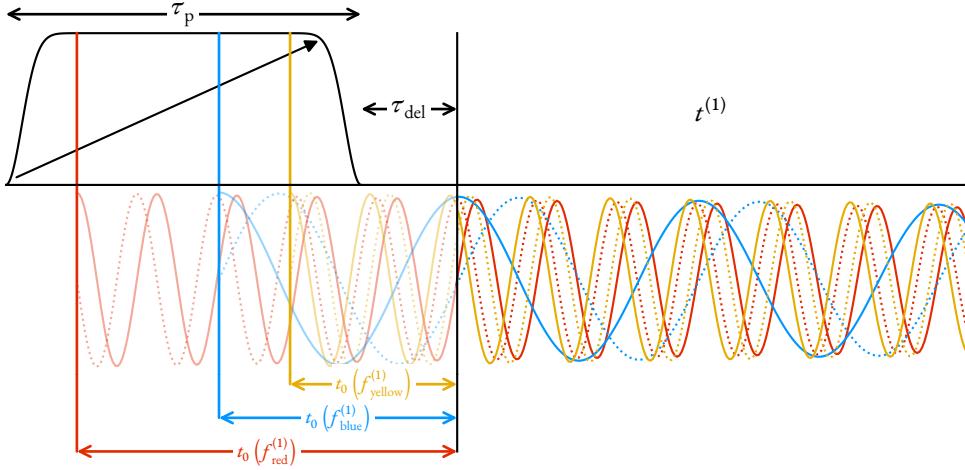


FIGURE 3.16: An illustration of an experiment comprising a single chirp pulse sweeping low to high frequencies of duration τ_p , followed by a pre-scan delay period or τ_{del} , prior to acquisition. The fate of three resonances with different frequencies is denoted, with $f_{\text{red}}^{(1)} < f_{\text{blue}}^{(1)} < f_{\text{yellow}}^{(1)}$. Each resonance is excited at different points in time, with lower frequency resonances being excited earlier, such that each resonance is allowed to evolve for different amounts of time prior to acquisition (t_0). The resulting FID possesses quadratic phase behaviour. Coloured oscillations denote the evolution of each resonance, with solid and dashed lines representing real and imaginary components, respectively. It is assumed that the 90° chirp rotates each resonance to be initially in phase with the receiver.

tation, while simple, yields spectra with undesirable phase behaviour, on account of resonances with different frequencies being excited at different moments in time. There are well-established methods for overcoming this using pulse sequences featuring an initial excitation, followed by one or more refocussing FS pulses[111–115].

With knowledge of the form of the chirp pulse, the expected phase of a particular resonance is determinable, and in this section, it will be shown that well-phased spectra can be obtained from excitation with a single FS pulse when appropriate post-processing of the FID is employed. The main advantage of being able to derive spectra with desirable features from a single chirp excitation experiment is the fact that ultra-broadband spectra can be generated using with a far shorter pulse sequence than state of the art methods such as **CHORUS!** (**CHORUS!**)[114, 115], where both a 90° chirp pulse, and two **180!** (**180!**) chirp pulses are applied. Spectra with greater intensity, as which could include broad resonances from slowly tumbling species could therefore be realised. Here, a description of the technique is presented, followed by an illustration of its performance on a simulated and an experimental dataset. This work is fairly nascent, and while the initial results presented here show promise, there is yet to be consideration on samples of greater interest.

3.5.1 Chirp excitation

Here, focus is limited to chirp pulses which sweep from low to high frequencies. Such a pulse is parameterised by its duration τ_p (s), excitation bandwidth ΔF (Hz), and RF amplitude ν_{RF} (Hz). The frequencies that the pulse sweeps through are in the range $[f_{off}^{(1)} - 1/2\Delta F, f_{off}^{(1)} + 1/2\Delta F]$, and the rate at which the frequency of the chirp is increased (the sweep rate) is given by $\Delta F/\tau_p$. Figure 3.16 provides an illustration of a single chirp excitation experiment. After application of the chirp pulse, there is typically a short pre-scan delay τ_{del} , usually on the order of a few μs , prior to the start of acquisition. While the pre-scan delay can be determined if an intimate knowledge of the spectrometer hardware is known, this varies from instrument to instrument, and is not trivial to ascertain.

The various pulse parameters are inter-related as follows:

$$\nu_{RF} = \sqrt{\frac{\Delta F Q}{2\pi\tau_p}}, \quad (3.32)$$

where $Q \in \mathbb{R}_{>0}$ is the *adiabaticity factor*[116]. For a pulse with flip angle $\beta < 180^\circ$, Q is related to β via

$$Q = \frac{2}{\pi} \ln \left(\frac{2}{\cos(\beta) + 1} \right), \quad (3.33)$$

ref? such that an appropriate pulse to achieve a flip angle of 90° requires selecting a combination of ν_{RF} , ΔF , and τ_p which satisfies $Q \approx 0.441^{\dagger\dagger\dagger}$. For a pulse with sufficiently low ν_{RF} , which requires a sufficiently large pulse duration for a given excitation bandwidth, it is reasonable to assume that the chirp pulse induces an instantaneous 90° rotation of a spin at the point of resonance, as illustrated in Figure 3.16. As such, resonances with different frequencies evolve for different amounts of time prior to the start of acquisition, according to **Double check this with Ali: his draft has + for last term, rather than -. I have -, as this makes t_0 larger for frequencies less than the transmitter, as I would expect for a low-to-high sweep.**

$$t_0(f^{(1)}) = \tau_{del} + \frac{\tau_p}{2} - \frac{(f^{(1)} - f_{off}^{(1)})\tau_p}{2\Delta F}. \quad (3.34)$$

$\tau_{del} + \tau_p/2$ is the amount of time between excitation and detection for an on-resonance oscillator, which is excited exactly halfway through the pulse. Resonances with a frequency smaller than the transmitter are excited earlier and hence have a larger t_0 , while the converse is true for resonances with greater frequencies. The resulting overall phase as a function of frequency can be

^{†††}The combination used in examples in this work are $\nu_{RF} \approx 16.8$ kHz, $\Delta F = 400$ kHz, $\tau_p = 100$ μs

approximated as[115]

$$\phi(f^{(1)}) = \phi_0 + 2\pi \left(\tau_{\text{del}} + \frac{\tau_p}{2} \right) (f^{(1)} - f_{\text{off}}^{(1)}) + 2\pi \left(\frac{\tau_p}{2\Delta F} \right) (f^{(1)} - f_{\text{off}}^{(1)})^2. \quad (3.35)$$

One might assume that it is possible to generate phased spectra by simply applying phase correction to the spectrum, via

$$s_\phi(f^{(1)}) = s(f^{(1)}) \exp(-i\phi(f^{(1)})). \quad (3.36)$$

While the quadratic phase behaviour of peaks is corrected by doing this, another issue with the dataset is not addressed. For any resonance, the signal that is detected can be thought of as the difference between two signals: (a) the “complete” signal, which starts at the time of excitation, and (b) a “truncated” signal which is identical to the complete signal before acquisition, and which comprises zeros once acquisition has begun. The linear nature of the FT dictates that the resulting delayed-acquisition spectrum comprises the difference between the FTs of the complete signal and the truncated signal. The FT of a severely truncated FID is well approximated as a broad sinc wiggle with its maximum at the resonance frequency. The form of the wiggle depends on the delay between excitation and acquisition, with resonances of lower frequencies, for which more of the signal is missed, displaying deeper, and narrower artefacts. The result of applying quadratic phase correction is therefore a spectrum of well-phased peaks, but with major baseline distortion, particularly to the low frequency end. Figures 3.17 & 3.18 both provide an example of this phenomenon (panel b).

3.5.2 Methodology

Both quadratic phase and missing point-derived baseline distortions can be resolved if an estimate of the FIDs parameters is obtained. This enables the construction of an FID featuring oscillators which are back-propagated, such that they begin not at the point of acquisition, but at the point of excitation. The appropriate start time for an oscillator with frequency $f^{(1)}$ is therefore given by $-t_0$, with t_0 defined in (3.34). The resulting corrected FID \mathbf{y}_{corr} is as follows:

$$\begin{aligned} \mathbf{y}_{\text{corr}} \left[n^{(1)} \right] = & \sum_{m=0}^{M-1} \boldsymbol{\alpha}[m] \exp(i\phi[m]) \times \\ & \exp \left(- \left(2\pi i (f^{(1)}[m] - f_{\text{off}}^{(1)}) - \boldsymbol{\gamma}^{(1)}[m] \right) t_0(f^{(1)}[m]) n^{(1)} \Delta_t^{(1)} \right). \end{aligned} \quad (3.37)$$

In this work so far, it has been assumed that all oscillators that make up the data are of the same phase. Of course this isn’t the case for signals generated from single-chirp excitation. As such, it is inappropriate to incorporate the variance of oscillator phases in fidelity for NLP. For the

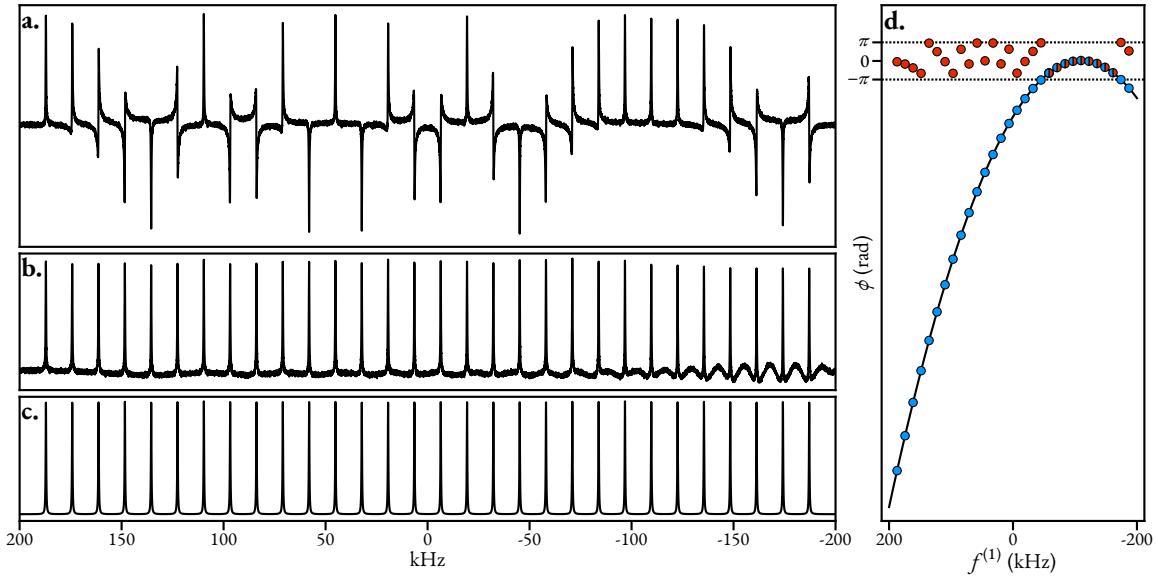


FIGURE 3.17: Comparison of quadratic phase correction vs frequency-dependent back-propagation in treating simulated single-chirp excitation data. **a.** Simulated spectrum for a spin system comprising 30 spins with uniformly-separated resonance frequencies. The data was generated with $N^{(1)} = 2^{14}$, $f_{\text{sw}}^{(1)} = 400 \text{ kHz}$, $f_{\text{off}}^{(1)} = 0 \text{ Hz}$, $\tau_p = 100 \mu\text{s}$, $\tau_{\text{del}} = 0 \text{ s}$, $\Delta F = 400 \text{ kHz}$. **b.** Spectrum generated using quadratic phase correction, with (3.36). **c.** Spectrum generated from estimation using the MPM, and back-propagation. **d.** Estimated phases of each oscillator as a function of frequency. Red points: phases wrapped within the range $(-\pi, \pi]$. Blue points: the same phases, adjusted by addition of a suitable multiple of 2π to each red point in order to display the quadratic dependence of the phases. Black curve: quadratic fit of the blue points.

examples presented in this work, NLP was not applied: the direct output of the MPM was used as the estimate of the FIDs parameters.

3.5.3 Results

Figure 3.17 provides a comparison of quadratic phase correction and estimation-based back-propagation on a simulated dataset comprising 30 non-interacting spins with evenly spaced resonance frequencies (relevant simulation parameters are provided in the figure caption). For the purpose of clarity, a very large damping factor (1000 s^{-1}) was assigned to each oscillator, as this augments the baseline distortions in the spectrum. As explained above, more intense, narrower artefacts are associated with low-frequency resonances, where a longer time between excitation and acquisition exists. The MPM was used to estimate the signal parameters, with only the first 2048 points considered. Performing the MPM on a signal with 2^{14} points would take (a) a long time, and (b) require a very large amount of RAM (see Figure 2.3). Consideration of the first 2048 in an FID is justifiable here, since all resonance frequencies are spaced reasonably far apart, such that oscillators in the signal become distinguishable early on into evolution. The spectrum generated via back-propagation is presented in panel c, where a well-phased spectrum without baseline distortion

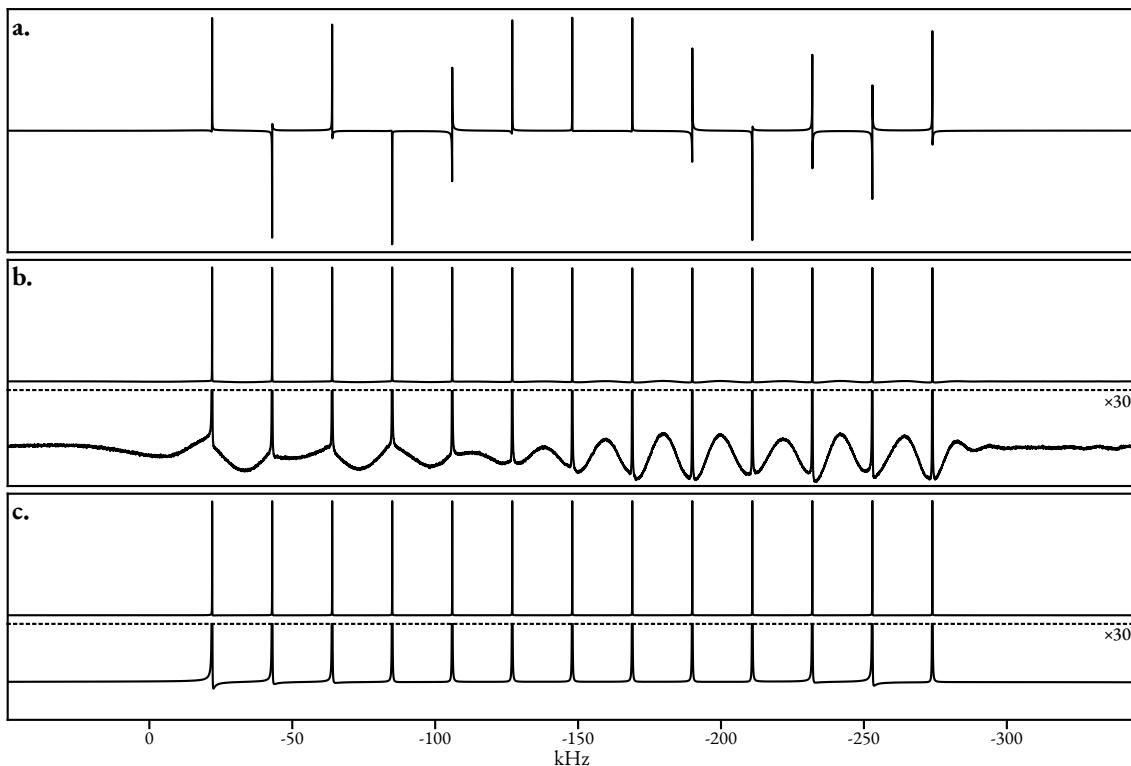


FIGURE 3.18: Comparison of quadratic phase correction vs frequency-dependent back-propagation in treating experimental single-chirp excitation data generated from a sample of 1% Gd-doped H₂O in D₂O. **a.** Spectrum generated directly from the acquired FID. **b.** Spectrum generated using quadratic phase correction. **c.** Spectrum generated from estimation using the MPM, followed by frequency-dependent back-propagation.

was achieved. An illustration of the variation of the estimated oscillator phases against their frequencies is presented in panel d, with the quadratic dependence clearly illustrated. Fitting the blue points in panel d to a quadratic function yielded a second-order coefficient of -7.85×10^{-10} rad s², in agreement with $-2\pi (\tau_p/2\Delta F)$ as expected.

Ask Ali for the original data (info on Chirp pulse, pulse sequence: assume its a sequential experiment where O1 is adjusted across increments) A second illustration of the procedure is provided for a real dataset, derived from a sample of 1% Gd-doped H₂O in D₂O.

Why dope with Gd? Is it to reduce the T₂ of water, in order to create more severe baseline distortions? The dataset was acquired using a 2D experiment, in which the transmitter offset was adjusted for each increment. As such, FIDs with a single resonance of differing frequency from H₂O are produced across the increments, which when summed lead to the spectrum in panel a of Figure 3.18. As with the simulated case, severe baseline distortions are apparent when quadratic phase correction is applied (panel b), with the most severe dips exhibited by low-frequency resonances. Again, estimation-based back-propagation yields a far cleaner spectral baseline.

THE NMR-EsPy PACKAGE

4

The estimation routine and applications emerging from it that have been presented in the previous two chapters are accessible via the NMR-EsPy package. NMR-EsPy aims to provide a feature-rich yet simple interface in order perform estimation on datasets of interest. In this chapter, a description of the package is given, including the design choices, a basic description of the package structure, and a description of the accompanying graphical user interface (GUI).

A rigorous description of the usage of NMR-EsPy, including details on installation, walk-throughs for different data types, and a reference for the application programming interface (API) are given in the documentation for NMR-EsPy. A hard-copy of the version 2.0 documentation can be found in Section **TODO** of the Appendix. The most up-to-date HTML version of the documentation can be found at <https://foroozandehgroup.github.io/NMR-EsPy/>. The source code for NMR-EsPy is hosted on the Foroozandeh group's GitHub page at <https://github.com/foroozandehgroup/NMR-EsPy>.

4.1 The Structure of NMR-EsPy

4.1.1 Why PYTHON?

There a number of reasons why PYTHON was the chosen programming language for NMR-EsPy:

- ❖ It has a large user-base, particularly within the scientific community.
- ❖ The SCIPY ecosystem[117], including the packages NUMPY[118] and MATPLOTLIB[119] is a powerful tool which enables high-performance scientific computation in PYTHON
- ❖ Being a scripting language makes PYTHON ideal for exploring datasets in a step-by-step fashion. This is useful in the context of NMR estimation, as the user may want to (a) inspect and pre-process then data, ten (b) determine the regions they wish to estimate, then (c) setup the estimation routine, and finally (d) output the estimation result. This can be achieved rather

easily by hacking and re-running PYTHON scripts or by using “notebook” environments, such as JUPYTER.

- ❖ It is free and open-source, as opposed to well-known scientific computing platforms such as MATLAB® and MATHEMATICA.
- ❖ PYTHON supports sophisticated object-oriented programming features, such as multiple levels of inheritance.

Probably the biggest drawback of PYTHON is its slow performance on account of it being an interpreted*, dynamically typed†, language with relies on garbage collection‡ for memory management. While NUMPY provides interfaces to run fast computations with pre-compiled C-code, a significant performnace benefit would likely be realised if a low-level compiled language like C, C++, or RUST were used. However the development time in writing programs with these lower-level languages is typically a lot greater than with a language with a higher level of abstraction like PYTHON.

4.1.2 ???

The fundamental user-facing object that NMR-EsPy provides is the `Estimator` class and its numerous inherited classes, which facilitate the estimation of different types of NMR data. A complete list of estimator objects at the time of writing is given in Table 4.1

TABLE 4.1: A complete list of estimator objects provided by NMR-EsPy at the time of writing. **Maybe indicate that the inv rec and diffusion estimators share similar functionality, except for the amplitude fitting. Diffusion experiments only differ in definition of fitting constant.**

Type	Description
<code>Estimator1D</code>	For consideration of 1D datasets.
<code>Estimator2D</code>	For consideration of 2D datasets comprising a pair of States (amplitude-modulated) signals.
<code>Estimator2DJ</code>	For consideration of 2DJ datasets. This object provides the functionality to generate pure shift spectra and assign multiplet structures using CUPID.

*With interpreted languages, the source code is processed line-by-line at the time of running by an interpreter. This differs from compiled languages, where prior to being run, the source code is converted to machine-readable byte-code.

†Dynamically typed languages, as opposed to statically typed languages like C, C++, JAVA, RUST etc. allow for a variable which is initially assigned a given type to be re-assigned to a completely different type.

‡Garbage collection involves a program routinely checking for any memory that has become dereferenced, and clearing this memory up.

<code>EstimatorInvRec</code>	For consideration of inversion recovery (T_1) experiments.
<code>EstimatorDiffusionMonopolar</code>	For consideration of monopolar gradient diffusion experiments.
<code>EstimatorDiffusionBipolar</code>	For consideration of bipolar gradient diffusion experiments.
<code>EstimatorDiffusionOneshot</code>	For consideration of one-shot DOSY experiments.

CONCLUSIONS AND FUTURE WORK

5

BIBLIOGRAPHY

- [1] F. Bloch, W. W. Hansen, and M. Packard. “Nuclear Induction”. In: *Phys. Rev.* 69 (3-4 Feb. 1946), pp. 127–127.
- [2] E. M. Purcell, H. C. Torrey, and R. V. Pound. “Resonance absorption by nuclear magnetic moments in a solid”. In: *Physical review* 69.1-2 (1946), p. 37.
- [3] E. D. Becker. “A brief history of Nuclear Magnetic Resonance”. In: *Analytical Chemistry* 65.6 (1993), 295A–302A.
- [4] G. R. Eaton, S. S. Eaton, and K. M. Salikhov. *Foundations of modern EPR*. Singapore ; London: World Scientific, 1998.
- [5] W. D. Knight. “Nuclear Magnetic Resonance Shift in Metals”. In: *Phys. Rev.* 76 (8 Oct. 1949), pp. 1259–1260.
- [6] W. G. Proctor and F. C. Yu. “The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound”. In: *Phys. Rev.* 77 (5 Mar. 1950), pp. 717–717.
- [7] W. C. Dickinson. “Dependence of the ^{19}F Nuclear Resonance Position on Chemical Compound”. In: *Phys. Rev.* 77 (5 Mar. 1950), pp. 736–737.
- [8] R. R. Ernst and W. A. Anderson. “Application of Fourier Transform Spectroscopy to Magnetic Resonance”. In: *Review of Scientific Instruments* 37.1 (1966), pp. 93–102.
- [9] R. Freeman and G. A. Morris. “The Varian story”. In: *Journal of Magnetic Resonance* 250 (2015), pp. 80–84.
- [10] J. W. Cooley and J. W. Tukey. “An Algorithm for the Machine Calculation of Complex Fourier Series”. In: *Mathematics of Computation* 19.90 (1965), pp. 297–301.
- [11] J. Jeener. “Ampere international summer school”. In: *Basko Polje, Yugoslavia* 197.1 (1971).
- [12] J. Jeener and A. G. ““Pulse pair technique in high resolution NMR” a reprint of the historical 1971 lecture notes on two-dimensional spectroscopy”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 94-95 (2016), pp. 75–80.
- [13] W. P. Aue, E. Bartholdi, and R. R. Ernst. “Two-dimensional spectroscopy. Application to nuclear magnetic resonance”. In: *The Journal of Chemical Physics* 64.5 (1976), pp. 2229–2246.

- [14] M. P. Williamson, T. F. Havel, and K. Wüthrich. "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry". In: *Journal of Molecular Biology* 182.2 (1985), pp. 295–315.
- [15] D. Marion et al. "Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta." In: *Biochemistry* 28.15 (1989), pp. 6150–6156.
- [16] L. Kay et al. "Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution". In: *Science* 249.4967 (1990), pp. 411–414.
- [17] K. Pervushin et al. "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution". In: *Proceedings of the National Academy of Sciences* 94.23 (1997), pp. 12366–12371.
- [18] R. R. Ernst. "Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture)". In: *Angewandte Chemie International Edition in English* 31.7 (1992), pp. 805–823.
- [19] K. Wüthrich. "NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture)". In: *Angewandte Chemie International Edition* 42.29 (2003), pp. 3340–3363.
- [20] P. Abragam and A. Abragam. *The Principles of Nuclear Magnetism*. Comparative Pathobiology - Studies in the Postmodern Theory of Education. Clarendon Press, 1961.
- [21] M. Goldman. *Quantum description of high-resolution NMR in liquids*. eng. International series of monographs on chemistry ; 15. Oxford: Clarendon Press, 1988.
- [22] J. Cavanagh. *Protein NMR spectroscopy : principles and practice*. eng. 2nd ed. Burlington, Mass. ; London: Academic Press, 2007.
- [23] M. H. Levitt. *Spin dynamics : basics of nuclear magnetic resonance*. eng. Second edition. Chichester, 2007.
- [24] P. Styles et al. "A high-resolution NMR probe in which the coil and preamplifier are cooled with liquid helium". In: *Journal of Magnetic Resonance (1969)* 60.3 (1984), pp. 397–404.
- [25] P. Styles, N. F. Soffe, and C. A. Scott. "An improved cryogenically cooled probe for high-resolution NMR". In: *Journal of Magnetic Resonance (1969)* 84.2 (1989), pp. 376–378.

- [26] H. Kovacs, D. Moskau, and M. Spraul. “Cryogenically cooled probes—a leap in NMR technology”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 46.2 (2005), pp. 131–155.
- [27] J. Keeler. *Understanding NMR spectroscopy*. eng. 2nd ed. Chichester: Wiley, 2010.
- [28] P. J. Hore, J. A. Jones, and S. Wimperis. *NMR : the toolkit : how pulse sequences work*. eng. Second edition. Oxford chemistry primers. Oxford, 2015.
- [29] C. Tang. “An Analysis of Baseline Distortion and Offset in NMR Spectra”. In: *Journal of Magnetic Resonance, Series A* 109.2 (1994), pp. 232–240.
- [30] P. J. Hore. *Nuclear magnetic resonance*. eng. Second edition. Oxford chemistry primers. Oxford, 2015.
- [31] W. Dietrich, C. H. Rüdel, and M. Neumann. “Fast and precise automatic baseline correction of one- and two-dimensional nmr spectra”. In: *Journal of Magnetic Resonance (1969)* 91.1 (1991), pp. 1–11.
- [32] J. Carlos Cobas et al. “A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data”. In: *Journal of Magnetic Resonance* 183.1 (2006), pp. 145–151.
- [33] J. Keeler and D. Neuhaus. “Comparison and evaluation of methods for two-dimensional NMR spectra with absorption-mode lineshapes”. In: *Journal of Magnetic Resonance (1969)* 63.3 (1985), pp. 454–472.
- [34] W. Aue, J. Karhan, and R. Ernst. “Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy”. In: *The Journal of Chemical Physics* 64.10 (1976), pp. 4226–4227.
- [35] G. A. Morris. “Two-Dimensional J -Resolved Spectroscopy”. In: *eMagRes*. John Wiley & Sons, Ltd, 2009.
- [36] A. L. Davis et al. “Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients”. In: *Journal of Magnetic Resonance (1969)* 98.1 (1992), pp. 207–216.
- [37] D. S. Stephenson. “Linear prediction and maximum entropy methods in NMR spectroscopy”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 20.6 (1988), pp. 515–626.
- [38] P. Koehl. “Linear prediction spectral analysis of NMR data”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 34.3 (1999), pp. 257–299.

- [39] G. U. YULE. “On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers”. eng. In: *Philosophical transactions of the Royal Society of London. Series A, Containing papers of a mathematical or physical character* 226 (1927), pp. 267–298.
- [40] G. T. Walker. “On periodicity in series of related terms”. eng. In: *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* 131.818 (1931), pp. 518–532.
- [41] N. Levinson. “The Wiener (Root Mean Square) Error Criterion in Filter Design and Prediction”. In: *Journal of Mathematics and Physics* 25.1-4 (1946), pp. 261–278.
- [42] J. Durbin. “The Fitting of Time-Series Models”. eng. In: *Revue de l’Institut international de statistique* 28.3 (1960), p. 233.
- [43] R. Kumaresan and D. Tufts. “Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 30.6 (1982), pp. 833–840.
- [44] R. Kumaresan. “On the zeros of the linear prediction-error filter for deterministic signals”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 31.1 (1983), pp. 217–220.
- [45] G. H. Golub and V. Pereyra. “The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate”. In: *SIAM Journal on Numerical Analysis* 10.2 (1973), pp. 413–432.
- [46] J. W. C. van der Veen et al. “Accurate quantification of in vivo ^{31}P NMR signals using the variable projection method and prior knowledge”. In: *Magnetic Resonance in Medicine* 6.1 (1988), pp. 92–98.
- [47] R. Penrose. “A generalized inverse for matrices”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (1955), pp. 406–413.
- [48] G. Strang. *Linear Algebra and Its Applications*. 5 ed. Florence: Cengage Learning, Inc, 2018.
- [49] K. Levenberg. “A method for the solution of certain non-linear problems in least squares”. In: *Quarterly of Applied Mathematics* 2.2 (1944), pp. 164–168.
- [50] D. W. Marquardt. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”. In: *Journal of the Society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.
- [51] R. Fletcher. *Practical methods of optimization*. eng. 2nd ed. Chichester: Wiley, 1987.

- [52] J. Nocedal and S. J. Wright. *Numerical optimization*. 2nd ed. Springer series in operations research. New York: Springer, 2006.
- [53] Y. Hua and T. Sarkar. “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise”. In: *IEEE Trans. on Acoust. Speech Signal Process.* 38.5 (1990), pp. 814–824.
- [54] Y. Hua and T. Sarkar. “Matrix pencil and system poles”. In: *Signal Processing* 21.2 (1990), pp. 195–198.
- [55] Y. Hua and T. Sarkar. “On SVD for estimating generalized eigenvalues of singular matrix pencil in noise”. In: *IEEE Trans. Signal Process.* 39.4 (1991), pp. 892–900.
- [56] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
- [57] Y. Hua. “Estimating two-dimensional frequencies by matrix enhancement and matrix pencil”. eng. In: *IEEE transactions on signal processing* 40.9 (1992), pp. 2267–2280.
- [58] F.-J. Chen et al. “Estimation of Two-Dimensional Frequencies Using Modified Matrix Pencil Method”. eng. In: *IEEE transactions on signal processing* 55.2 (2007), pp. 718–724.
- [59] H. Akaike. “A new look at the statistical model identification”. In: *IEEE Trans. Automat. Control* 19.6 (1974), pp. 716–723.
- [60] G. Schwarz. “Estimating the Dimension of a Model”. In: *Ann. Stat.* 6.2 (1978), pp. 461–464.
- [61] J. Rissanen. “Modeling by shortest data description”. In: *Automatica* 14.5 (1978), pp. 465–471.
- [62] M. Wax and T. Kailath. “Detection of signals by information theoretic criteria”. In: *IEEE Trans. Acoust. Speech Signal Process.* 33.2 (1985), pp. 387–392.
- [63] N. I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, 1993.
- [64] Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford science publications. OUP Oxford, 2001.
- [65] N. I. Gould et al. “Sensitivity of trust-region algorithms to their parameters”. eng. In: *4OR* 3.3 (2005), pp. 227–241.
- [66] Y.-Y. Lin et al. “A Novel Detection–Estimation Scheme for Noisy NMR Signals: Applications to Delayed Acquisition Data”. In: *J. Magn. Reson.* 128.1 (1997), pp. 30–41.
- [67] M. Mayzel, K. Kazimierczuk, and V. Y. Orekhov. “The causality principle in the reconstruction of sparse NMR spectra”. In: *Chem. Commun.* 50.64 (2014), pp. 8947–8950.

- [68] D. Gołowicz, P. Kasprzak, and K. Kazimierczuk. “Enhancing Compression Level for More Efficient Compressed Sensing and Other Lessons from NMR Spectroscopy”. In: *Sensors (Basel)* 20.5 (2020), p. 1325.
- [69] J. Luo et al. “Fast reconstruction of non-uniform sampling multidimensional NMR spectroscopy via a deep neural network”. In: *Journal of Magnetic Resonance* 317 (2020), p. 106772.
- [70] Y. Hua. “High resolution imaging of continuously moving object using stepped frequency radar”. In: *Signal Processing* 35.1 (1994), pp. 33–40.
- [71] S. Fricke et al. “Data processing in NMR relaxometry using the matrix pencil”. In: *Journal of Magnetic Resonance* 313 (2020), p. 106704.
- [72] D. Wörtge et al. “Quantitative stray-field T1 relaxometry with the matrix pencil method”. In: *Journal of Magnetic Resonance* 351 (2023), p. 107435.
- [73] H. Maeda and Y. Yanagisawa. “Future prospects for NMR magnets: A perspective”. In: *Journal of Magnetic Resonance* 306 (2019), pp. 80–85.
- [74] A. Shaka et al. “An improved sequence for broadband decoupling: WALTZ-16”. In: *Journal of Magnetic Resonance (1969)* 52.2 (1983), pp. 335–338.
- [75] A. Shaka, J. Keeler, and R. Freeman. “Evaluation of a new broadband decoupling sequence: WALTZ-16”. In: *Journal of Magnetic Resonance (1969)* 53.2 (1983), pp. 313–340.
- [76] A. Shaka, P. Barker, and R. Freeman. “Computer-optimized decoupling scheme for wide-band applications and low-level operation”. In: *Journal of Magnetic Resonance (1969)* 64.3 (1985), pp. 547–552.
- [77] T. D. Claridge. *High-Resolution NMR Techniques in Organic Chemistry: Third Edition*. eng. 2016.
- [78] N. H. Meyer and K. Zangger. “Simplifying Proton NMR Spectra by Instant Homonuclear Broadband Decoupling”. In: *Angewandte Chemie International Edition* 52.28 (2013), pp. 7143–7146.
- [79] R. W. Adams. “Pure Shift NMR Spectroscopy”. In: *eMagRes*. John Wiley & Sons, Ltd, 2014, pp. 295–310.
- [80] K. Zangger. “Pure shift NMR”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 86-87 (2015), pp. 1–20.
- [81] A. Bax, R. Freeman, and G. A. Morris. “A simple method for suppressing dispersion-mode contributions in NMR spectra: The “pseudo echo””. In: *Journal of Magnetic Resonance (1969)* 43.2 (1981), pp. 333–338.

- [82] E. Kupce, J. Boyd, and I. Campbell. “Short Selective Pulses for Biochemical Applications”. In: *Journal of Magnetic Resonance, Series B* 106.3 (1995), pp. 300–303.
- [83] K. Zanger and H. Sterk. “Homonuclear Broadband-Decoupled NMR Spectra”. In: *Journal of Magnetic Resonance* 124.2 (1997), pp. 486–489.
- [84] J. A. Aguilar et al. “Pure Shift 1H NMR: A Resolution of the Resolution Problem?” In: *Angewandte Chemie International Edition* 49.23 (2010), pp. 3901–3903.
- [85] A. J. Pell and J. Keeler. “Two-dimensional J-spectra with absorption-mode lineshapes”. In: *Journal of Magnetic Resonance* 189.2 (2007), pp. 293–299.
- [86] J. Garbow, D. Weitekamp, and A. Pines. “Bilinear rotation decoupling of homonuclear scalar interactions”. In: *Chemical Physics Letters* 93.5 (1982), pp. 504–509.
- [87] A. Bax. “Broadband homonuclear decoupling in heteronuclear shift correlation NMR spectroscopy”. In: *Journal of Magnetic Resonance* (1969) 53.3 (1983), pp. 517–520.
- [88] M. Foroozandeh et al. “Ultrahigh-Resolution NMR Spectroscopy”. In: *Angewandte Chemie International Edition* 53.27 (2014), pp. 6990–6992.
- [89] M. Foroozandeh, G. A. Morris, and M. Nilsson. “PSYCHE Pure Shift NMR Spectroscopy”. In: *Chemistry – A European Journal* 24.53 (2018), pp. 13988–14000.
- [90] P. Kiraly et al. “Anatomising proton NMR spectra with pure shift 2D J-spectroscopy: A cautionary tale”. In: *Chemical Physics Letters* 683 (2017). Ahmed Zewail (1946-2016) Commemoration Issue of Chemical Physics Letters, pp. 398–403.
- [91] J.-M. Nuzillard. “Time-Reversal of NMR Signals by Linear Prediction. Application to Phase-Sensitive HomonuclearJ-Resolved Spectroscopy”. In: *Journal of Magnetic Resonance, Series A* 118.1 (1996), pp. 132–135.
- [92] A. Martinez et al. “High-resolution and high-sensitivity 2D homonuclear J-resolved NMR spectroscopy”. In: *Magnetic Resonance in Chemistry* 50.1 (2012), pp. 28–32.
- [93] P. Mutzenhardt, F. Guenneau, and D. Canet. “A Procedure for Obtaining Pure Absorption 2D J-Spectra: Application to Quantitative Fully J-Decoupled Homonuclear NMR Spectra”. In: *Journal of Magnetic Resonance* 141.2 (1999), pp. 312–321.
- [94] H. Hogben et al. “Spinach – A software library for simulation of spin dynamics in large spin systems”. eng. In: *Journal of magnetic resonance* (1997) 208.2 (2011), pp. 179–194.
- [95] M. J. Frisch et al. *Gaussian03 Revision E.01*. Gaussian Inc. Wallingford CT. 2003.
- [96] M. J. Thrippleton, R. A. Edden, and J. Keeler. “Suppression of strong coupling artefacts in J-spectra”. In: *Journal of Magnetic Resonance* 174.1 (2005), pp. 97–109.

- [97] G. Wider et al. “Strong spin-spin coupling in the two-dimensional J-resolved 360-MHz ^1H NMR spectra of the common amino acids”. In: *Journal of Magnetic Resonance (1969)* 42.1 (1981), pp. 73–87.
- [98] M. Foroozandeh et al. “Measuring couplings in crowded NMR spectra: Pure shift NMR with multiplet analysis”. eng. In: *Chemical communications (Cambridge, England)* 51.84 (2015), pp. 15410–15413.
- [99] C. Johnson. “Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 34.3 (1999), pp. 203–256.
- [100] G. A. Morris. “Diffusion-Ordered Spectroscopy”. In: *eMagRes*. John Wiley & Sons, Ltd, 2009.
- [101] E. O. Stejskal and J. E. Tanner. “Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient”. In: *The Journal of Chemical Physics* 42.1 (1965), pp. 288–292.
- [102] H. C. Torrey. “Bloch Equations with Diffusion Terms”. In: *Phys. Rev.* 104 (3 Nov. 1956), pp. 563–565.
- [103] J. E. Tanner. “Use of the Stimulated Echo in NMR Diffusion Studies”. In: *The Journal of Chemical Physics* 52.5 (1970), pp. 2523–2526.
- [104] R. Cotts et al. “Pulsed field gradient stimulated echo methods for improved NMR diffusion measurements in heterogeneous systems”. In: *Journal of Magnetic Resonance (1969)* 83.2 (1989), pp. 252–266.
- [105] D. Wu, A. Chen, and C. Johnson. “An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses”. In: *Journal of Magnetic Resonance, Series A* 115.2 (1995), pp. 260–264.
- [106] M. D. Pelta et al. “A one-shot sequence for high-resolution diffusion-ordered spectroscopy”. In: *Magnetic Resonance in Chemistry* 40.13 (2002), S147–S152.
- [107] D. Sinnaeve. “The Stejskal–Tanner equation generalized for any gradient shape—an overview of most pulse sequences measuring free diffusion”. In: *Concepts in Magnetic Resonance Part A* 40A.2 (2012), pp. 39–65.
- [108] *Spinach Documentation*. https://spindynamics.org/wiki/index.php?title=Main_Page. Accessed 11-05-2023.
- [109] A. Chen et al. “Chemical Exchange in Diffusion NMR Experiments”. eng. In: *Journal of the American Chemical Society* 120.35 (1998), pp. 9094–9095.

- [110] M. Foroozandeh. “Spin dynamics during chirped pulses: applications to homonuclear decoupling and broadband excitation”. In: *Journal of Magnetic Resonance* 318 (2020), p. 106768.
- [111] J.-M. Bohlen, M. Rey, and G. Bodenhausen. “Refocusing with chirped pulses for broadband excitation without phase dispersion”. In: *Journal of Magnetic Resonance (1969)* 84.1 (1989), pp. 191–197.
- [112] J. Bohlen and G. Bodenhausen. “Experimental Aspects of Chirp NMR Spectroscopy”. In: *Journal of Magnetic Resonance, Series A* 102.3 (1993), pp. 293–301.
- [113] K. E. Cano, M. A. Smith, and A. Shaka. “Adjustable, Broadband, Selective Excitation with Uniform Phase”. In: *Journal of Magnetic Resonance* 155.1 (2002), pp. 131–139.
- [114] J. E. Power et al. “Increasing the quantitative bandwidth of NMR measurements”. In: *Chem. Commun.* 52 (14 2016), pp. 2916–2919.
- [115] M. Foroozandeh, M. Nilsson, and G. A. Morris. “Improved ultra-broadband chirp excitation”. In: *Journal of Magnetic Resonance* 302 (2019), pp. 28–33.
- [116] E. Kupce and R. Freeman. “Adiabatic Pulses for Wideband Inversion and Broadband Decoupling”. In: *Journal of Magnetic Resonance, Series A* 115.2 (1995), pp. 273–276.
- [117] P. Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nat. Methods* 17 (2020), pp. 261–272.
- [118] C. R. Harris et al. “Array programming with NumPy.” In: *Nature* 585.7825 (2020), p. 357.
- [119] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95.
- [120] S. Berger and S. Braun. *200 and more NMR experiments : a practical course.* eng. [3rd rev. and expanded ed.] Weinheim: Wiley-VCH, 2004.
- [121] C. P. Butts, C. R. Jones, and J. N. Harvey. “High precision NOEs as a probe for low level conformers-a second conformation of strychnine”. eng. In: 47.4 (2011), pp. 1193–1195.
- [122] L. A. O’Dell. “The WURST kind of pulses in solid-state NMR”. In: *Solid State Nuclear Magnetic Resonance* 55-56 (2013), pp. 28–41.
- [123] C. E. Shannon. “Communication in the Presence of Noise”. In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21.

ADDITIONAL THEORY

A

Add any extra theory stuff here: linear algebra definitions, probability stuff etc.

A.1 Additional Algorithms

Algorithm A1 The MMEMPM.

1: **procedure** MMEMPM($\mathbf{Y} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}, M \in \mathbb{N}$)
2: $L^{(1)}, L^{(2)} \leftarrow \lfloor N^{(1)}/2 \rfloor, \lfloor N^{(2)}/2 \rfloor;$
3: **for** $n^{(1)} \leftarrow \{0, \dots, N^{(1)} - 1\}$ **do**
4: $\mathbf{H}_{Y,n^{(1)}} \leftarrow \begin{bmatrix} \mathbf{Y}[n^{(1)}, 0] & \mathbf{Y}[n^{(1)}, 1] & \dots & \mathbf{Y}[n^{(1)}, N^{(2)} - L^{(2)}] \\ \mathbf{Y}[n^{(1)}, 1] & \mathbf{Y}[n^{(1)}, 2] & \dots & \mathbf{Y}[n^{(1)}, N^{(2)} - L^{(2)} + 1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}[n^{(1)}, L^{(2)} - 1] & \mathbf{Y}[n^{(1)}, L^{(2)}] & \dots & \mathbf{Y}[n^{(1)}, N^{(2)} - 1] \end{bmatrix}$
5: **end for**;
6: $\mathbf{E}_Y \leftarrow \begin{bmatrix} \mathbf{H}_{Y,0} & \mathbf{H}_{Y,1} & \dots & \mathbf{H}_{Y,N^{(1)}-L^{(1)}} \\ \mathbf{H}_{Y,1} & \mathbf{H}_{Y,2} & \dots & \mathbf{H}_{Y,N^{(1)}-L^{(1)}+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Y,L^{(1)}-1} & \mathbf{H}_{Y,L^{(1)}} & \dots & \mathbf{H}_{Y,N^{(1)}-1} \end{bmatrix};$
7: $\mathbf{U}_M, \Sigma_M, \mathbf{V}_M^\dagger \leftarrow \text{TRUNCATEDSVD}(\mathbf{E}_Y, M);$
8: $\mathbf{P} \leftarrow \mathbf{0} \in \mathbb{C}^{L^{(1)}L^{(2)} \times L^{(1)}L^{(2)}},$
9: $r \leftarrow 0$
10: **for** $i = 0, \dots, L^{(2)} - 1$ **do**
11: **for** $j = 0, \dots, L^{(1)} - 1$ **do**
12: $c \leftarrow i + jL^{(2)}$;
13: $\mathbf{P}[r, c] \leftarrow 1$;
14: $r = r + 1$;
15: **end for**
16: **end for**
17: $\mathbf{U}_{M1}, \mathbf{U}_{M2} \leftarrow \mathbf{U}_M[:, L^{(1)}(L^{(2)} - 1)], \mathbf{U}_M[L^{(2)} :];$ ▷ Last/First $L^{(2)}$ rows deleted
18: $\mathbf{z}^{(1)}, \mathbf{W}^{(1)} \leftarrow \text{EIGENDECOMPOSITION}(\mathbf{U}_{M1}^+ \mathbf{U}_{M2});$
19: $\mathbf{f}^{(1)}, \boldsymbol{\eta}^{(1)} \leftarrow (f_{sw}^{(1)}/2\pi) \Im(\ln \mathbf{z}^{(1)}) + f_{off}^{(1)}, -f_{sw}^{(1)} \Re(\ln \mathbf{z}^{(1)});$
20: $\mathbf{U}_{MP} \leftarrow \mathbf{P} \mathbf{U}_M;$
21: $\mathbf{U}_{MP1}, \mathbf{U}_{MP2} \leftarrow \mathbf{U}_{MP}[:, (L^{(1)} - 1)L^{(2)}], \mathbf{U}_{MP}[L^{(1)} :];$ ▷ Last/First $L^{(1)}$ rows deleted
22: $\mathbf{z}^{(2)} \leftarrow \text{diag}([\mathbf{W}^{(1)}]^{-1} \mathbf{U}_{MP1}^+ \mathbf{U}_{MP2} \mathbf{W}^{(1)});$
23: $\mathbf{f}^{(2)}, \boldsymbol{\eta}^{(2)} \leftarrow (f_{sw}^{(2)}/2\pi) \Im(\ln \mathbf{z}^{(2)}) + f_{off}^{(2)}, -f_{sw}^{(2)} \Re(\ln \mathbf{z}^{(2)});$
24: $\mathbf{Z}_L^{(2)} = [\mathbf{1} \quad \mathbf{z}^{(2)} \quad \mathbf{z}^{(2)2} \quad \dots \quad \mathbf{z}^{(2)L^{(2)}-1}]^T;$
25: $\mathbf{Z}_R^{(2)} \leftarrow [\mathbf{1} \quad \mathbf{z}^{(2)} \quad \mathbf{z}^{(2)2} \quad \dots \quad \mathbf{z}^{(2)N^{(2)}-L^{(2)}}];$
26: $\mathbf{Z}_D^{(1)} \leftarrow \text{diag}(\mathbf{z}^{(1)});$
27: $\mathbf{E}_L \leftarrow \begin{bmatrix} \mathbf{Z}_L^{(2)} \\ \mathbf{Z}_L^{(2)} \mathbf{Z}_D^{(1)} \\ \vdots \\ \mathbf{Z}_L^{(2)} [\mathbf{Z}_D^{(1)}]^{L^{(1)}-1} \end{bmatrix};$
28: $\mathbf{E}_R \leftarrow [\mathbf{Z}_R^{(2)} \quad \mathbf{Z}_D^{(1)} \mathbf{Z}_R^{(2)} \quad \dots \quad [\mathbf{Z}_D^{(1)}]^{N^{(1)}-L^{(1)}} \mathbf{Z}_R^{(2)}];$
29: $\boldsymbol{\alpha} \leftarrow \text{diag}(\mathbf{E}_L^+ \mathbf{E}_Y \mathbf{E}_R^+);$
30: $\boldsymbol{\alpha}, \boldsymbol{\phi} \leftarrow |\boldsymbol{\alpha}|, \arctan\left(\frac{\Im(\boldsymbol{\alpha})}{\Re(\boldsymbol{\alpha})}\right);$
31: $\boldsymbol{\theta}^{(0)} \leftarrow [\boldsymbol{\alpha}^T \quad \boldsymbol{\phi}^T \quad [\mathbf{f}^{(1)}]^T \quad [\mathbf{f}^{(2)}]^T \quad [\boldsymbol{\eta}^{(1)}]^T \quad [\boldsymbol{\eta}^{(2)}]^T]^T;$
32: **return** $\boldsymbol{\theta}^{(0)}$
33: **end procedure**

Algorithm A2 Steihaug-Toint method for determining an update for nonlinear programming.
This is equivalent to Algorithm 7.2 in [52].

```

1: procedure STEIHAUGTOINT( $Y \in \mathbb{C}^{N^{(1)} \times \dots \times N^{(D)}}$ ,  $\theta^{(k)} \in \mathbb{R}^{2(1+D)M}$ ,  $\Delta^{(k)} \in \mathbb{R}_{>0}$ )
2:    $\mathbf{g} \leftarrow \nabla \mathcal{F}_\phi(\theta^{(k)} | Y)$ ;                                 $\triangleright$  Grad vector: (2.43a)
3:    $\mathbf{H} \leftarrow \nabla^2 \mathcal{F}_\phi(\theta^{(k)} | Y)$ ;                             $\triangleright$  Hessian matrix, either exact: (2.43b) or approximate: (2.46)
4:    $\epsilon^{(k)} \leftarrow \min(1/2, \sqrt{\|\mathbf{g}\|}) \|\mathbf{g}\|$ ;
5:    $\mathbf{z}^{(0)} \leftarrow \mathbf{0} \in \mathbb{R}^{6M}$ ;
6:    $\mathbf{r}^{(0)} \leftarrow \mathbf{g}$ ;
7:    $\mathbf{d}^{(0)} \leftarrow -\mathbf{r}^{(0)}$ ;
8:   if  $\|\mathbf{r}^{(0)}\| < \epsilon^{(k)}$  then
9:     return  $\mathbf{z}^{(0)}$ ;
10:  end if
11:  for  $j = \{0, 1, \dots\}$  do
12:    if  $\mathbf{d}^{(j)\top} \mathbf{H} \mathbf{d}^{(j)} \leq 0$  then
13:      Find  $\tau$  such that  $\mathbf{p}^{(k)} = \mathbf{z}^{(j)} + \tau \mathbf{d}^{(j)}$  minimises  $\mathcal{F}_{\phi Q}(\theta^{(k)} + \mathbf{p}^{(k)})$ , subject to  $\|\mathbf{p}^{(k)}\| = \Delta^{(k)}$ ;
14:      return  $\mathbf{p}^{(k)}$ ;
15:    end if
16:     $\alpha^{(j)} \leftarrow \frac{\mathbf{r}^{(j)\top} \mathbf{r}^{(j)}}{\mathbf{d}^{(j)\top} \mathbf{H} \mathbf{d}^{(j)}}$ ;
17:     $\mathbf{z}^{(j+1)} \leftarrow \mathbf{z}^{(j)} + \alpha^{(j)} \mathbf{d}^{(j)}$ ;
18:    if  $\|\mathbf{z}^{(j+1)}\| < \epsilon^{(k)}$  then
19:      Find  $\tau \in \mathbb{R}_{>0}$  such that  $\mathbf{p}^{(k)} = \mathbf{z}^{(j)} + \tau \mathbf{d}^{(j)}$  satisfies  $\|\mathbf{p}^{(k)}\| = \Delta^{(k)}$ ;
20:      return  $\mathbf{p}^{(k)}$ ;
21:    end if
22:     $\mathbf{r}^{(j+1)} \leftarrow \mathbf{r}^{(j)} + \alpha^{(j)} \mathbf{H} \mathbf{d}^{(j)}$ ;
23:    if  $\|\mathbf{r}^{(j+1)}\| < \epsilon^{(k)}$  then
24:      return  $\mathbf{z}^{(j+1)}$ ;
25:    end if
26:     $\beta^{(j+1)} \leftarrow \frac{\mathbf{r}^{(j+1)\top} \mathbf{r}^{(j+1)}}{\mathbf{r}^{(j)\top} \mathbf{r}^{(j)}}$ ;
27:     $\mathbf{d}^{(j+1)} \leftarrow -\mathbf{r}^{(j+1)} + \beta^{(j+1)} \mathbf{d}^{(j)}$ ;
28:  end for
29: end procedure

```

Algorithm A3 Filtering procedure for 1D data. $\mathbf{r}_{\text{interest}}$ is a vector of length 2 containing the indices of the left and right bounds of the region of interest. These would typically be provided in units of Hz or ppm by a user. Conversion to array indices can be carried out using (2.72). $\mathbf{r}_{\text{noise}}$ contains the left and right bounds of the region used to estimate the noise variance. RANDOMSAMPLE indicates taking a random sample from the given distribution.

```

1: procedure FILTER1D( $\mathbf{y} \in \mathbb{C}^{N^{(1)}}$ ,  $\mathbf{r}_{\text{interest}} \in \mathbb{N}_0^2$ ,  $\mathbf{r}_{\text{noise}} \in \mathbb{N}_0^2$ )
2:    $\mathbf{s}_{\text{ve}} \leftarrow \text{VIRTUALECHO1D}(\mathbf{y});$ 
3:    $\mathbf{s}_{\text{ve}} \leftarrow \text{FT}(\mathbf{y}_{\text{ve}});$ 
4:    $l_{\text{idx}}^{(1)}, r_{\text{idx}}^{(1)} \leftarrow \mathbf{r}_{\text{interest}}[0], \mathbf{r}_{\text{interest}}[1];$ 
5:    $l_{\text{idx}, \text{noise}}^{(1)}, r_{\text{idx}, \text{noise}}^{(1)} \leftarrow \mathbf{r}_{\text{noise}}[0], \mathbf{r}_{\text{noise}}[1];$ 
6:    $c_{\text{idx}}^{(1)} \leftarrow (l_{\text{idx}}^{(1)} + r_{\text{idx}}^{(1)})/2;$ 
7:    $b_{\text{idx}}^{(1)} \leftarrow r_{\text{idx}}^{(1)} - l_{\text{idx}}^{(1)};$ 
8:    $\mathbf{g} \leftarrow \text{SUPERGAUSSIAN1D}\left(N^{(1)}, c_{\text{idx}}^{(1)}, b_{\text{idx}}^{(1)}\right);$ 
9:    $\mathbf{s}_{\text{noise}} \leftarrow \mathbf{s}_{\text{ve}} \left[ l_{\text{idx}, \text{noise}}^{(1)} : r_{\text{idx}, \text{noise}}^{(1)} + 1 \right];$ 
10:   $\sigma^2 \leftarrow \text{Var}(\mathbf{s}_{\text{noise}});$ 
11:   $\mathbf{w}_{\sigma^2} \leftarrow \mathbf{0} \in \mathbb{R}^{2N^{(1)}};$ 
12:  for  $n^{(1)} = 0, \dots, N^{(1)} - 1$  do
13:     $\mathbf{w}_{\sigma^2} \left[ n^{(1)} \right] \leftarrow \text{RANDOMSAMPLE}(\mathcal{N}(0, \sigma^2));$ 
14:  end for
15:   $\tilde{\mathbf{s}}_{\text{ve}} \leftarrow \mathbf{s}_{\text{ve}} \odot \mathbf{g} + \mathbf{w}_{\sigma^2} \odot (\mathbf{1} - \mathbf{g});$ 
16:   $\tilde{\mathbf{y}}_{\text{ve}} \leftarrow \text{IFT}(\tilde{\mathbf{s}}_{\text{ve}});$ 
17:   $\tilde{\mathbf{y}} \leftarrow \tilde{\mathbf{y}}_{\text{ve}} \left[ : N^{(1)} \right];$ 
18:  return  $\tilde{\mathbf{y}};$ 
19: end procedure

20: procedure VIRTUALECHO1D( $\mathbf{y} \in \mathbb{C}^{N^{(1)}}$ )
21:    $\mathbf{t}_1 \leftarrow \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \in \mathbb{C}^{N^{(1)}} \end{bmatrix};$ 
22:    $\mathbf{t}_2 \leftarrow \begin{bmatrix} \mathbf{0} \in \mathbb{C}^{N^{(1)}} \\ \mathbf{y}^* \leftrightarrow (1) \end{bmatrix} \circledcirc (1);$ 
23:    $\mathbf{y}_{\text{ve}} \leftarrow \mathbf{t}_1 + \mathbf{t}_2;$ 
24:    $\mathbf{y}_{\text{ve}}[0] \leftarrow \mathbf{y}_{\text{ve}}[0]/2;$ 
25:   return  $\mathbf{y}_{\text{ve}};$ 
26: end procedure

27: procedure SUPERGAUSSIAN1D( $N \in \mathbb{N}$ ,  $c_{\text{idx}} \in \mathbb{R}_{>0}$ ,  $b_{\text{idx}} \in \mathbb{N}$ )
28:    $\mathbf{g} \leftarrow \mathbf{0} \in \mathbb{R}^N;$ 
29:   for  $n = 0, \dots, N - 1$  do
30:      $\mathbf{g}[n] \leftarrow \exp\left(-2^{41} \left(\frac{n - c_{\text{idx}}}{b_{\text{idx}}}\right)^{40}\right);$   $\triangleright p$  in (2.68b) has been set to 40.
31:   end for
32:   return  $\mathbf{g}$ 
33: end procedure

```

Algorithm A4 Filtering procedure for 2D data.

```

1: procedure FILTER2D( $\mathbf{Y}_{\cos} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}, \mathbf{Y}_{\sin} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}, \mathbf{R}_{\text{interest}} \in \mathbb{N}_0^{2 \times 2}, \mathbf{R}_{\text{noise}} \in \mathbb{N}_0^{2 \times 2}$ )
2:    $\mathbf{Y}_{\text{ve}} \leftarrow \text{VIRTUALECHO2D}(\mathbf{Y}_{\cos}, \mathbf{Y}_{\sin});$ 
3:    $\mathbf{S}_{\text{ve}} \leftarrow \text{FT}(\mathbf{Y}_{\text{ve}});$ 
4:    $l_{\text{idx}}, r_{\text{idx}}, l_{\text{idx}}, r_{\text{idx}} \leftarrow \mathbf{R}_{\text{interest}}[0, 0], \mathbf{R}_{\text{interest}}[0, 1], \mathbf{R}_{\text{interest}}[1, 0], \mathbf{R}_{\text{interest}}[1, 1];$ 
5:   for  $d = 1, 2$  do
6:      $c_{\text{idx}}^{(d)} \leftarrow (l_{\text{idx}}^{(d)} + r_{\text{idx}}^{(d)})/2;$ 
7:      $b_{\text{idx}}^{(d)} \leftarrow r_{\text{idx}}^{(d)} - l_{\text{idx}}^{(d)};$ 
8:      $\mathbf{g}^{(d)} \leftarrow \text{SUPERGAUSSIAN1D}(2N^{(d)}, c_{\text{idx}}^{(d)}, b_{\text{idx}}^{(d)});$ 
9:      $\mathbf{G} \leftarrow \mathbf{g}^{(1)} \otimes \mathbf{g}^{(2)};$ 
10:    end for
11:     $l_{\text{idx,noise}}, r_{\text{idx,noise}}, l_{\text{idx,noise}}, r_{\text{idx,noise}} \leftarrow \mathbf{R}_{\text{noise}}[0, 0], \mathbf{R}_{\text{noise}}[0, 1], \mathbf{R}_{\text{noise}}[1, 0], \mathbf{R}_{\text{noise}}[1, 1];$ 
12:     $\mathbf{S}_{\text{noise}} \leftarrow \mathbf{S}_{\text{ve}} \left[ l_{\text{idx,noise}}^{(1)} : r_{\text{idx,noise}}^{(1)} + 1, l_{\text{idx,noise}}^{(2)} : r_{\text{idx,noise}}^{(2)} + 1 \right]$ 
13:     $\sigma^2 \leftarrow \text{Var}(\mathbf{S}_{\text{noise}});$ 
14:     $\mathbf{W}_{\sigma^2} \leftarrow \mathbf{0} \in \mathbb{R}^{2N^{(1)} \times 2N^{(2)}};$ 
15:    for  $n^{(1)} = 0, \dots, 2N^{(1)} - 1$  do
16:      for  $n^{(2)} = 0, \dots, 2N^{(2)} - 1$  do
17:         $\mathbf{W}_{\sigma^2}[n^{(1)}, n^{(2)}] \leftarrow \text{RANDOMSAMPLE}(\mathcal{N}(0, \sigma^2));$ 
18:      end for
19:    end for
20:     $\tilde{\mathbf{S}}_{\text{ve}} \leftarrow \mathbf{S}_{\text{ve}} \odot \mathbf{G} + \mathbf{W}_{\sigma^2} \odot (\mathbf{1} - \mathbf{G});$ 
21:     $\tilde{\mathbf{Y}}_{\text{ve}} \leftarrow \text{IFT}(\tilde{\mathbf{S}}_{\text{ve}});$ 
22:     $\tilde{\mathbf{Y}} \leftarrow \tilde{\mathbf{Y}}_{\text{ve}}[:, N^{(1)}, : N^{(2)}];$ 
23:    return  $\tilde{\mathbf{Y}};$ 
24: end procedure

25: procedure VIRTUALECHO2D( $\mathbf{Y}_{\cos} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}, \mathbf{Y}_{\sin} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}$ )
26:    $\mathbf{Y}_{++} \leftarrow \Re(\mathbf{Y}_{\cos}) - \Im(\mathbf{Y}_{\sin}) + i(\Im(\mathbf{Y}_{\cos}) + \Re(\mathbf{Y}_{\sin}));$ 
27:    $\mathbf{Y}_{+-} \leftarrow \Re(\mathbf{Y}_{\cos}) + \Im(\mathbf{Y}_{\sin}) + i(\Re(\mathbf{Y}_{\sin}) - \Im(\mathbf{Y}_{\cos}));$ 
28:    $\mathbf{Y}_{-+} \leftarrow \Re(\mathbf{Y}_{\cos}) + \Im(\mathbf{Y}_{\sin}) + i(\Im(\mathbf{Y}_{\cos}) - \Re(\mathbf{Y}_{\sin}));$ 
29:    $\mathbf{Y}_{--} \leftarrow \Re(\mathbf{Y}_{\cos}) - \Im(\mathbf{Y}_{\sin}) - i(\Im(\mathbf{Y}_{\cos}) + \Re(\mathbf{Y}_{\sin}));$ 
30:    $\mathbf{Z} \leftarrow \mathbf{0} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}$ 
31:    $\mathbf{T}_1 \leftarrow \begin{bmatrix} \mathbf{Y}_{++} & \mathbf{Z} \\ \mathbf{Z} & \mathbf{Z} \end{bmatrix};$ 
32:    $\mathbf{T}_2 \leftarrow \begin{bmatrix} \mathbf{Z} & \mathbf{Y}_{+-}^{(2)} \\ \mathbf{Z} & \mathbf{Z} \end{bmatrix}^{\odot (2)};$ 
33:    $\mathbf{T}_3 \leftarrow \begin{bmatrix} \mathbf{Z} & \mathbf{Z} \\ \mathbf{Y}_{-+}^{(1)} & \mathbf{Z} \end{bmatrix}^{\odot (1)};$ 
34:    $\mathbf{T}_4 \leftarrow \begin{bmatrix} \mathbf{Z} & \mathbf{Z} \\ \mathbf{Z} & \mathbf{Y}_{--}^{(1,2)} \end{bmatrix}^{\odot (1,2)};$ 
35:    $\mathbf{Y}_{\text{ve}} \leftarrow \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \mathbf{T}_4;$ 
36:   for  $n^{(1)} = 0, \dots, 2N^{(1)} - 1$  do
37:      $\mathbf{Y}_{\text{ve}}[n^{(1)}, 0] \leftarrow \mathbf{Y}_{\text{ve}}[n^{(1)}, 0]/2;$ 
38:   end for
39:   for  $n^{(2)} = 0, \dots, 2N^{(2)} - 1$  do
40:      $\mathbf{Y}_{\text{ve}}[0, n^{(2)}] \leftarrow \mathbf{Y}_{\text{ve}}[0, n^{(2)}]/2;$ 
41:   end for
42:   return  $\mathbf{Y}_{\text{ve}};$ 
43: end procedure

```

Algorithm A5 Filtering procedure for 2DJ data.

```

1: procedure FILTER2DJ(  $\mathbf{Y} \in \mathbb{C}^{N^{(1)} \times N^{(2)}}$ ,  $\mathbf{r}_{\text{interest}} \in \mathbb{N}_0^2$ ,  $\mathbf{r}_{\text{noise}} \in \mathbb{N}_0^2$  )
2:    $\mathbf{Y}_{\text{ve}} \leftarrow \mathbf{0} \in \mathbb{C}^{N^{(1)} \times 2N^{(2)}}$ ;
3:   for  $n^{(1)} = 0, \dots, N^{(1)} - 1$  do
4:      $\mathbf{Y}_{\text{ve}}[n^{(1)}, :] \leftarrow \text{VIRTUALECHO1D}(\mathbf{Y}[n^{(1)}, :])$ ;
5:   end for
6:    $\mathbf{S}_{\text{ve}} \leftarrow \text{FT}^{(2)}(\mathbf{Y}_{\text{ve}})$ ;
7:    $l_{\text{idx}}^{(2)}, r_{\text{idx}}^{(2)} \leftarrow \mathbf{r}_{\text{interest}}[0], \mathbf{r}_{\text{interest}}[1]$ ;
8:    $l_{\text{idx}, \text{noise}}^{(2)}, r_{\text{idx}, \text{noise}}^{(2)} \leftarrow \mathbf{r}_{\text{noise}}[0], \mathbf{r}_{\text{noise}}[1]$ ;
9:    $c_{\text{idx}}^{(2)} \leftarrow (l_{\text{idx}}^{(2)} + r_{\text{idx}}^{(2)})/2$ ;
10:   $b_{\text{idx}}^{(2)} \leftarrow r_{\text{idx}}^{(2)} - l_{\text{idx}}^{(2)}$ ;
11:   $\mathbf{g}^{(1)} \leftarrow \mathbf{1} \in \mathbb{R}^{N^{(1)}}$ ;
12:   $\mathbf{g}^{(2)} \leftarrow \text{SUPERGAUSSIAN1D}(N^{(2)}, c_{\text{idx}}^{(2)}, b_{\text{idx}}^{(2)})$ ;
13:   $\mathbf{G} \leftarrow \mathbf{g}^{(1)} \otimes \mathbf{g}^{(2)}$ ;
14:   $\mathbf{S}_{\text{noise}} \leftarrow \mathbf{S}_{\text{ve}}[:, l_{\text{idx}, \text{noise}}^{(2)} : r_{\text{idx}, \text{noise}}^{(2)} + 1]$ ;
15:   $\sigma^2 \leftarrow \text{Var}(\mathbf{S}_{\text{noise}})$ ;
16:   $\mathbf{W}_{\sigma^2} \leftarrow \mathbf{0} \in \mathbb{R}^{N^{(1)} \times 2N^{(2)}}$ ;
17:  for  $n^{(1)} = 0, \dots, N^{(1)} - 1$  do
18:    for  $n^{(2)} = 0, \dots, 2N^{(2)} - 1$  do
19:       $\mathbf{W}_{\sigma^2}[n^{(1)}, n^{(2)}] \leftarrow \text{RANDOMSAMPLE}(\mathcal{N}(0, \sigma^2))$ ;
20:    end for
21:  end for
22:   $\tilde{\mathbf{S}}_{\text{ve}} \leftarrow \mathbf{S}_{\text{ve}} \odot \mathbf{G} + \mathbf{W}_{\sigma^2} \odot (\mathbf{1} - \mathbf{G})$ ;
23:   $\tilde{\mathbf{Y}}_{\text{ve}} \leftarrow \text{IFT}^{(2)}(\tilde{\mathbf{S}}_{\text{ve}})$ ;
24:   $\tilde{\mathbf{Y}} \leftarrow \tilde{\mathbf{Y}}_{\text{ve}}[:, : N^{(2)}]$ ;
25:  return  $\tilde{\mathbf{Y}}$ ;
26: end procedure

```

Algorithm A6 An algorithm for multiplet assignment of a 2DJ estimation result.

```

1: procedure MULTIPLETASSIGN( $\theta \in \mathbb{R}^{6M}$ ,  $\epsilon \in \mathbb{R}_{>0}$ )
2:    $\mathbf{f}^{(1)}, \mathbf{f}^{(2)} \leftarrow \theta[2M : 3M], \theta[3M : 4M]$ ;
3:   MAP  $\leftarrow \text{HASHMAP}[\mathbb{R}, \text{VECTOR}[\mathbb{N}_0]]$ ;
4:   for  $m = \{0, \dots, M - 1\}$  do
5:      $f_c = \mathbf{f}^{(2)}[m] - \mathbf{f}^{(1)}[m]$ ;
6:     ASSIGNED  $\leftarrow \text{FALSE}$ ;
7:     for  $f_{\text{mp}}, i$  in MAP do
8:       if  $|f_c - f_{\text{mp}}| < \epsilon$  then
9:          $i \leftarrow [i^T \ m]^T$ ;
10:        ASSIGNED  $\leftarrow \text{TRUE}$ ;
11:        break;
12:       end if
13:     end for
14:     if ASSIGNED = FALSE then
15:       MAP.insert( $(f_c, [m])$ );
16:     end if
17:   end for
18:   return MAP;
19: end procedure

```

CODE LISTINGS

B

NEED TO ADD TEXT TO DESCRIBE THINGS The following is all the required imports:

```
1 import itertools
2 from typing import Any, Iterable, Optional, Tuple, Union
3 import numpy as np
4 import scipy as sp
```

B.1 Matrix pencil methods

B.1.1 MDL

The `mdl` function computes an estimate of the number of oscillators present in a 1D an FID (Section 2.2.3). The first relative minimum in `mdl_vec` is determined to be the estimate of M , rather than the global minimum, since the presence of very small singular values can lead to division errors.[More precise here?](#)

```
1 def mdl(sigma: np.ndarray, N: int, L: int) -> int:
2     """Compute the Minimum Description Length
3
4     Parameters
5     -----
6     sigma
7         Vector of singular values associated with the data matrix.
8
9     N
10        Number of points the signal comprises.
11
12    L
13        Pencil parameter.
```

```

14      """
15      mdl_vec = np.zeros(L)
16      for k in range(L):
17          mdl_vec[k] = (
18              -N * np.sum(np.log(sigma[k:])) +
19              N * (L - k) * np.log(np.sum(sigma[k:])) / (L - k) +
20              k * np.log(N) * (2 * L - k) / 2
21      )
22      M = sp.signal.argrelextrema(mdl_vec, np.less)[0][0]
23      return M

```

B.1.2 MPM

```

1 def mpm(Y: np.ndarray, sw: float, offset: float, M: int = 0) -> np.array:
2     """Matrix Pencil Method for estimating a 1D FID.
3
4     Parameters
5     -----
6     Y
7         FID.
8
9     sw
10        Sweep width (Hz).
11
12    offset
13        Transmitter offset (Hz).
14
15    M
16        Number of oscillators. If 0, this is estimated using the MDL.
17    """
18    N = Y #  $N^{(1)}$ 
19    L = N // 3 #  $L^{(1)}$ : pencil parameter
20    norm = np.linalg.norm(Y) #  $\|Y\|$ 
21    Y /= norm #  $Y/\|Y\|$ 
22    col = Y[:N - L] # First column of  $H_Y$ 
23    row = Y[N - L - 1 :] # Last row of  $H_Y$ 
24    HY = sp.linalg.hankel(col, row) #  $H_Y$ 
25    _, S, Vh = np.linalg.svd(HY) #  $\sigma$  and  $V^\dagger$ 
26    V = Vh.T #  $V$ 
27
28    if M == 0:
29        M = mdl(sigma, N, L)
30

```

```

31     VM = V[:, :M] #  $\mathbf{V}_M$ 
32     VM1 = Vm[:-1, :] #  $\mathbf{V}_{M1}$ 
33     VM2 = Vm[1:, :] #  $\mathbf{V}_{M2}$ 
34     VM1inv = np.linalg.pinv(VM1) #  $\mathbf{V}_{M1}^+$ 
35     VM1invVM2 = VM1inv @ VM2 #  $\mathbf{V}_{M1}^+ \mathbf{V}_{M2}$ 
36     z, _ = np.linalg.eig(VM1invVM2) #  $\mathbf{z}^{(1)}$ : signal poles
37     Z = np.power.outer(z, np.arange(N)).T #  $\mathbf{Z}^{(1)}$ 
38     alpha = np.linalg.pinv(Z) @ Y #  $\alpha$ : complex amplitudes
39
40     # Extract amplitude, phase, frequency and damping factor
41     amp = np.abs(alpha) * norm #  $\alpha$ 
42     phase = np.arctan2(np.imag(alpha), np.real(alpha)) #  $\phi$ 
43     freq = (sw / (2 * np.pi)) * np.imag(np.log(z)) + offset #  $f^{(1)}$ 
44     damp = -sw * np.real(np.log(z)) #  $\gamma^{(1)}$ 
45     theta = np.vstack((amp, phase, freq, damp)).T #  $\theta$ , as a  $M \times 4$  array
46
47     # Remove negative damping factors
48     neg_damp_idx = np.nonzero(damp < 0.0)[0]
49     theta = np.delete(theta, neg_damp_idx, axis=0)
50
51     theta = theta[np.argsort(theta[:, 2])] # order by  $f^{(1)}$ 
52     return theta

```

B.1.3 MMEMPM

```

1 def mmempm(
2     Y: np.ndarray, sw1: float, sw2: float,
3     offset1: float, offset2: float, M: int = 0,
4 ) -> np.ndarray:
5     """Modified Matrix Enhancement Matrix Pencil Method for estimating a
6     2D FID.
7
8     Parameters
9     -----
10    Y
11        FID.
12
13    sw1
14        Sweep width in first (indirect) dimension (Hz).
15
16    sw2
17        Sweep width in second (direct) dimension (Hz).
18

```

```

19     offset1
20         Transmitter offset in first dimension (Hz).
21
22     offset2
23         Transmitter offset in second dimension (Hz).
24
25     M
26         Number of oscillators. If 0, this is estimated by applying the MDL
27         to the first direct dimension FID (i.e. Y[0])
28 """
29 N1, N2 = Y.shape #  $N^{(1)}, N^{(2)}$ 
30 L1, L2 = N1 // 2, N2 // 2 #  $L^{(1)}, L^{(2)}$ : pencil parameters
31 norm = np.linalg.norm(Y) #  $\|Y\|$ 
32 Y /= norm #  $Y/\|Y\|$ 
33
34 if M == 0:
35     # Extract first direct dimension signal, and estimate model order
36     # with the MDL
37     L_mdl = N2 // 3
38     Y_mdl = Y[0]
39     col_mdl = Y_mdl[: N2 - L_mdl]
40     row_mdl = Y_mdl[N2 - L_mdl - 1 :]
41     HY_mdl = sp.linalg.hankel(col_mdl, row_mdl)
42     _, sigma_mdl, _ = np.linalg.svd(HY_mdl)
43     M = mdl(sigma_mdl, N2, L_mdl)
44
45 # === Construct block Hankel  $E_Y$  ===
46 row_size = L2
47 col_size = N2 - L2 + 1
48 EY = np.zeros(
49     (L1 * L2, (N1 - L1 + 1) * (N2 - L2 + 1)),
50     dtype="complex",
51 )
52 for n1 in range(N1):
53     # Construct  $H_{Y,n^{(1)}}$ , and assign to appropriate positions in  $E_Y$ 
54     col = Y[n1, :L2]
55     row = Y[n1, L2 - 1:]
56     HYn1 = sp.linalg.hankel(col, row)
57     for n in range(n1 + 1):
58         r, c = n, n1 - n
59         if r < L1 and c < N1 - L1 + 1:
60             EY[
61                 r * row_size : (r + 1) * row_size,
62                 c * col_size : (c + 1) * col_size
63             ] = HYn1

```

```

64      # Make  $\mathbf{E}_Y$  sparse: SciPy only supports truncated SVD for sparse
65      # matrices
66      EY = sp.sparse.csr_matrix(EY)
67      UM, *_ = sp.sparse.linalg.svds(EY, k=M)  #  $\mathbf{U}_M$ 
68
69      # === Construct permutation matrix  $\mathbf{P}$  ===
70      P = np.zeros((L1 * L2, L1 * L2))
71      r = 0
72      for l2 in range(L2):
73          for l1 in range(L1):
74              c = l1 * L2 + l2
75              P[r, c] = 1
76              r += 1
77
78      UM1 = UM[:, :L1 * (L2 - 1)]  # Last  $L^{(2)}$  rows deleted:  $\mathbf{U}_{M1}$ 
79      UM2 = UM[:, L2:]  # First  $L^{(2)}$  rows deleted:  $\mathbf{U}_{M2}$ 
80      z1, W1 = np.linalg.eig(np.linalg.pinv(UM1) @ UM2)  #  $\mathbf{z}^{(1)}, \mathbf{W}^{(1)}$ 
81
82      UMP = P @ UM  #  $\mathbf{U}_{MP}$ 
83      UMP1 = UMP[:, : (L1 - 1) * L2]  # Last  $L^{(1)}$  rows deleted:  $\mathbf{U}_{MP1}$ 
84      UMP2 = UMP[:, L1:]  # First  $L^{(1)}$  rows deleted:  $\mathbf{U}_{MP2}$ 
85      Z2 = np.linalg.inv(W1) @ np.linalg.pinv(UMP1) @ UMP2 @ W1  #  $\mathbf{Z}^{(2)}$ 
86      z2 = np.diag(Z2).copy()  #  $\mathbf{z}^{(2)}$ : copy needed as slice is readonly
87
88      # === Check for and deal with similar frequencies ===
89      freq1 = (0.5 * sw1 / np.pi) * np.imag(np.log(z1)) + offset1  #  $f^{(1)}$ 
90      threshold = sw1 / N1  #  $f_{sw}^{(1)} / N^{(1)}$ 
91      groupings = {}
92      # Iterate through values in  $f^{(1)}$  and group any with
93      # similar frequencies together
94      for idx, f1 in enumerate(freq1):
95          assigned = False
96          for group_f1, indices in groupings.items():
97              if np.abs(f1 - group_f1) < threshold:
98                  indices.append(idx)
99                  n = len(indices)
100                 indices = sorted(indices)
101                 # Get new mean freq of the group
102                 new_group_f1 = (n * group_f1 + f1) / (n + 1)
103                 groupings[new_group_f1] = groupings.pop(group_f1)
104                 assigned = True
105                 break
106             if not assigned:
107                 groupings[f1] = [idx]
108

```

```

109     for indices in groupings.values():
110         n = len(indices)
111         if n != 1:
112             A_slice = tuple(zip(*itertools.product(indices, repeat=2)))
113             A = Z2[A_slice].reshape(n, n)
114             new_group_z2, _ = np.linalg.eig(A)
115             z2[indices] = new_group_z2
116
117     # === Construct  $\mathbf{E}_L$  and  $\mathbf{E}_R$  ===
118     ZL2 = np.power.outer(z2, np.arange(L2)).T #  $\mathbf{Z}_L^{(2)}$ 
119     ZR2 = np.power.outer(z2, np.arange(N2 - L2 + 1)) #  $\mathbf{Z}_R^{(2)}$ 
120     Z1D = np.diag(z1) #  $\mathbf{Z}_D^{(1)}$ 
121
122     EL = np.zeros((L1 * L2, M), dtype="complex")
123     Z2LZ1D = ZL2
124     for i in range(L1):
125         EL[i * row_size : (i + 1) * row_size] = Z2LZ1D
126         Z2LZ1D = Z2LZ1D @ Z1D
127     ER = np.zeros((M, (N1 - L1 + 1) * (N2 - L2 + 1)), dtype="complex")
128     Z1DZ2R = ZR2
129     for i in range(N1 - L1 + 1):
130         ER[:, i * col_size : (i + 1) * col_size] = Z1DZ2R
131         Z1DZ2R = Z1D @ Z1DZ2R
132     #  $\alpha = \text{diag}(\mathbf{E}_L^+ \mathbf{E}_Y \mathbf{E}_R^+)$ 
133     alpha = np.diag(np.linalg.pinv(EL) @ EY @ np.linalg.pinv(ER))
134
135     amp = np.abs(alpha) * norm #  $\alpha$ 
136     phase = np.arctan2(np.imag(alpha), np.real(alpha)) #  $\phi$ 
137     freq2 = (0.5 * sw2 / np.pi) * np.imag(np.log(z2)) + offset2 #  $f^{(2)}$ 
138     damp1 = -sw1 * np.real(np.log(z1)) #  $\gamma^{(1)}$ 
139     damp2 = -sw2 * np.real(np.log(z2)) #  $\gamma^{(2)}$ 
140     #  $\theta$ , as a  $M \times 6$  array
141     theta = np.vstack((amp, phase, freq1, freq2, damp1, damp2)).T
142
143     # Remove negative damping factors
144     neg_damp_idx1 = list(np.nonzero(damp1 < 0.))[0]
145     neg_damp_idx2 = list(np.nonzero(damp2 < 0.))[0]
146     neg_damp_idx = list(set(neg_damp_idx1 + neg_damp_idx2))
147     theta = np.delete(theta, neg_damp_idx, axis=0)
148
149     theta = theta[np.argsort(theta[:, 3])] # order by  $f^{(2)}$ 
150     return theta

```

B.2 NLP

B.2.1 Trust Region Algorithm

```
1 def trust_steihaug_toint(
2     theta0: np.ndarray,
3     function_factory: FunctionFactory,
4     args: Iterable[Any] = (),
5 ) -> Tuple[np.ndarray, np.ndarray, bool]:
6     """Trust Region algorithm with Steihaug-Toint subroutine.
7
8     Parameters
9     -----
10    theta0
11        Initial guess, with shape (M, 4).
12
13    function_factory
14        Object for computing the objective, gradient and Hessian.
15
16    args
17        Extra arguments required for computing the objective and its
18        derivatives.
19
20    Returns
21    -----
22    theta
23        Parameter vector at termination.
24
25    errors
26        Errors associated with parameter vector.
27
28    negative_amps
29        Flag indicating whether or not termination occurred because
30        negative amplitudes were detected.
31    """
32    theta = theta0
33    M = theta.shape[0] // 4
34    factory = function_factory(theta, *args)
35
36    # === Define relevant parameters ===
37    # These have been hard-coded, though in NMR-EsPy they are all
38    # configurable.
39    eta = 0.15,
```

```

40     initial_trust_radius = 0.1 * factory.gradient_norm
41     max_trust_radius = 16 * initial_trust_radius
42     epsilon: float = 1.e-8,
43     max_iterations: int = 200,
44     check_neg_amps_every: int = 25,
45
46     k = 0
47     while True:
48         # === Steihaug-Toint ===
49         epsi = min(0.5, np.sqrt(factory.gradient_norm)) *
50             ↵ factory.gradient_norm
51         z = np.zeros_like(theta)
52         r = factory.gradient
53         d = -r
54         while True:
55             Bd = factory.hessian @ d
56             dBd = d.T @ Bd
57             if dBd <= 0:
58                 ta, tb = get_boundaries(z, d, trust_radius)
59                 pa = z + ta * d
60                 pb = z + tb * d
61                 p = min(factory.model(pa), factory.model(pb))
62                 hits_boundary = True
63                 break
64             r_sq = r.T @ r
65             alpha = r_sq / dBd
66             z_next = z + alpha * d
67             if sp.linalg.norm(z_next) >= trust_radius:
68                 _, tb = get_boundaries(z, d, trust_radius)
69                 p = z + tb * d
70                 hits_boundary = True
71                 break
72             r_next = r + alpha * Bd
73             r_next_sq = r_next.T @ r_next
74             if np.sqrt(r_next_sq) < epsi:
75                 hits_boundary = False
76                 p = z_next
77                 break
78             beta_next = r_next_sq / r_sq
79             d_next = -r_next + beta_next * d
80             z = z_next
81             r = r_next
82             d = d_next
83
84         # === Assess effectiveness of update ===

```

```
84     predicted_value = factory.model(p)
85     theta_proposed = theta + p
86     factory_proposed = function_factory(theta_proposed, *args)
87     actual_reduction = factory.objective - factory_proposed.objective
88     predicted_reduction = factory.objective - predicted_value
89     if predicted_reduction <= 0:
90         # No improvement could be found: terminate
91         negative_amps = False
92         break
93     rho = actual_reduction / predicted_reduction
94     if rho < 0.25:
95         # Quadratic model performing poorly: reduce TR
96         trust_radius *= 0.25
97     elif rho > 0.75 and hits_boundary:
98         # Quadratic model performing well: increase TR
99         trust_radius = min(2 * trust_radius, max_trust_radius)
100    if rho > eta:
101        # Accept update: new iteration
102        theta = theta_proposed
103        factory = factory_proposed
104        k += 1
105
106    # === Check for termination criteria ===
107    if (k % check_neg_amps_every == 0):
108        neg_amps = np.where(theta[amp_slice] <= 0)[0]
109        print(neg_amps)
110        if neg_amps.size > 0:
111            # Negative amps found: this run in order to purge
112            negative_amps = True
113            break
114        if factory.gradient_norm < epsilon:
115            # Convergence
116            negative_amps = False
117            break
118        if k == max_iterations:
119            # Maximum allowed iterations reached
120            negative_amps = False
121            break
122
123    # Routine terminated: compute errors and return parameter array
124    errors = np.sqrt(
125        factory.objective *
126        np.abs(np.diag(np.linalg.inv(factory.hessian))))
127    )
128    return theta, errors, negative_amps
```

```

129
130
131 def get_boundaries(
132     z: np.ndarray, d: np.ndarray, trust_radius: float
133 ) -> Tuple[float, float]:
134     """Determine the intersections of the search direction and the trust
135     region."""
136     a = d.T @ d
137     b = 2 * z.T @ d
138     c = (z.T @ z) - (trust_radius ** 2)
139     aux = b + np.copysign(
140         np.sqrt(b * b - 4 * a * c),
141         b,
142     )
143     return sorted([-aux / (2 * a), -(2 * c) / aux])

```

B.2.2 Computing \mathcal{F}_ϕ , $\nabla \mathcal{F}_\phi$, and $\nabla^2 \mathcal{F}_\phi$

```

1 class FunctionFactory:
2     """Object which computes and memoises the objective, gradient and
3     hessian for a given set of parameters."""
4     def __init__(self, theta: np.ndarray, fun: callable, *args) -> None:
5         self.theta = theta
6         self._new_theta = True
7         self.fun = fun
8         self.obj = None
9         self.grad = None
10        self.hess = None
11        self.args = args
12
13    def _compute_if_needed(self):
14        """Compute the obj, grad and Hess if they haven't been yet"""
15        if self.obj is None:
16            self.obj, self.grad, self.hess = self.fun(self.theta,
17                *self.args)
18
19    def model(self, p) -> float:
20        return self.objective + self.gradient @ p + 0.5 * (p.T @
21            self.hessian @ p)
22
23    @property
24    def objective(self) -> float:
25        self._compute_if_needed()

```

```

24         return self.obj
25
26     @property
27     def gradient(self) -> np.ndarray:
28         self._compute_if_needed()
29         return self.grad
30
31     @property
32     def gradient_norm(self) -> float:
33         return sp.linalg.norm(self.gradient)
34
35     @property
36     def hessian(self) -> np.ndarray:
37         self._compute_if_needed()
38         return self.hess
39
40
41 class FunctionFactoryGaussNewton1D(FunctionFactory):
42     def __init__(self, theta: np.ndarray, *args) -> None:
43         super().__init__(theta, obj_grad_gauss_newton_hess_1d, *args)
44
45
46 def obj_grad_gauss_newton_hess_1d(
47     theta: np.ndarray, *args: Tuple[int, int, np.ndarray],
48 ) -> Tuple[float, np.ndarray, np.ndarray]:
49     Y, tp = args # Unpack args: FID, timepoints
50     N = Y.shape[0]
51     M = theta.shape[0] // 4
52     X_per_osc = np.exp(
53         np.outer(tp, (2j * np.pi * theta[2 * M : 3 * M] - theta[3 * M :]))
54     ) * (theta[:M] * np.exp(1j * theta[M : 2 * M]))
55
56     # Jacobian: all first partial derivatives,  $\partial X / \partial \theta$ 
57     jac = np.zeros((N, 4 * M), dtype="complex")
58     jac[:, :M] = X_per_osc / theta[:M] #  $\partial X / \partial \alpha$ 
59     jac[:, M : 2 * M] = 1j * X_per_osc #  $\partial X / \partial \phi$ 
60     jac[:, 2 * M : 3 * M] = \
61         np.einsum("ij,i->ij", X_per_osc, 2j * np.pi * tp) #  $\partial X / \partial f^{(1)}$ 
62     jac[:, 3 * M :] = np.einsum("ij,i->ij", X_per_osc, -tp) #  $\partial X / \partial \eta^{(1)}$ 
63
64     X = np.einsum("ij->i", X_per_osc)
65     Y_minus_X = Y - X
66
67     obj = (Y_minus_X.conj().T @ Y_minus_X).real #  $\mathcal{F}(\theta)$ 
68     grad = -2 * (Y_minus_X.conj().T @ jac).real #  $\nabla \mathcal{F}(\theta)$ 

```

```

69     hess = 2 * (jac.conj().T @ jac).real #  $\nabla^2 \mathcal{F}(\theta)$ 
70
71     # === Determine phase variance and derivatives ===
72     phi = theta[M : 2 * M]
73     cos_phi = np.cos(phi)
74     cos_sum = np.sum(cos_phi)
75     sin_phi = np.sin(phi)
76     sin_sum = np.sum(sin_phi)
77     R = np.sqrt(cos_sum ** 2 + sin_sum ** 2)
78     pv_obj = 1 - (R / M)
79     pv_grad = (sin_phi * cos_sum - cos_phi * sin_sum) / (M * R)
80     x = (sin_phi * cos_sum) - (cos_phi * sin_sum)
81     term_1 = np.outer(x, x) / (R ** 2)
82     phi_array = np.zeros((M, M))
83     phi_array[:] = phi
84     term_2 = -np.cos(phi_array.T - phi_array)
85     pv_hess = term_1 + term_2
86     pv_hess[np.diag_indices(M)] += cos_phi * cos_sum + sin_phi * sin_sum
87     pv_hess /= M * R
88     obj += pv_obj #  $\mathcal{F}_\phi(\theta)$ 
89     grad[M : 2 * M] += pv_grad #  $\nabla \mathcal{F}_\phi(\theta)$ 
90     hess[M : 2 * M, M : 2 * M] += pv_hess #  $\nabla^2 \mathcal{F}_\phi(\theta)$ 
91
92     return obj, grad, hess

```

B.2.3 The main routine

```

1 def nlp(
2     Y: np.ndarray,
3     sw: float,
4     offset: float,
5     theta0: np.ndarray,
6 ) -> Tuple[np.ndarray, np.ndarray]:
7     """Nonlinear programming routine for 1D FID estimation.
8
9     Parameters
10     -----
11     Y
12         FID.
13
14     sw
15         Sweep width (Hz).
16

```

```

17     offset
18         Transmitter offset (Hz).
19
20     theta0
21         Initial guess, of shape (M, 4)
22
23     Returns
24     -----
25     theta
26         Parameter estimate.
27
28     errors
29         Errors associated with theta.
30     """
31     norm = np.linalg.norm(Y)
32     Y /= norm
33     N = Y.shape[0]
34     M = theta0.shape[0]
35     # Flatten parameter array: Fortran (column-wise) ordering
36     theta0_vec = theta0.flatten(order="F")
37     theta0_vec[:M] /= norm # Normalise amplitudes
38     # Remove transmitter offset from frequencies
39     theta0_vec[2 * M : 3 * M] -= offset
40
41     # Extra arguments needed to compute the objective, grad, and Hessian:
42     # FID and timepoints sampled
43     opt_args = [Y, np.linspace(0, float(N - 1) / sw, N)]
44
45     while True:
46         theta_vec, errors_vec, negativeamps = trust_ncg(
47             theta0=theta0_vec,
48             function_factory=FunctionFactoryGaussNewton1D,
49             args=opt_args,
50         )
51
52         if negativeamps:
53             # Negative amps exist: remove these
54             negative_idx = list(np.where(theta_vec[:M] <= 0.)[0])
55             slice_ = []
56             for idx in range(negative_idx):
57                 slice_.extend([i * M + idx for i in range(4)])
58             theta_vec = np.delete(theta_vec, slice_)
59             M -= len(negative_idx) # New model order
60
61     else:

```

```
62         # No negative amps: routine complete
63         break
64
65     # Reshape parameter array back to (M, 4)
66     theta = theta_vec.reshape((M, 4), order="F")
67     errors = errors_vec.reshape((M, 4), order="F")
68     errors_vec /= N - 1
69     theta[:, 2] += offset # Re-add transmitter offset to frequencies
70     theta[:, 0] *= norm # Re-scale amplitudes
71     errors[:, 0] *= norm
72     theta[:, 1] = (theta[:, 1] + np.pi) % (2 * np.pi) - np.pi # Wrap
    ↵   phases
73
74     return theta, errors
```

INFORMATION ON DATASETS

C

C.1 Simulated datasets

C.1.1 SPINACH Simulations

Many of the simulated datasets presented in this work were generated using the SPINACH MATLAB® package[94]. In each case, the dataset was generated via a call to the `new_spinach` method associated with the relevant `Estimator` object in NMR-EsPy. `new_spinach` works by instantiating a MATLAB engine from PYTHON, which then runs a SPINACH simulation of the relevant experiment with the specifications provided. The FID generated is then stored within in a new `Estimator` object. Table A1 provides a specification of the chemical shifts and scalar couplings that made up the spin systems used to construct simulated data with SPINACH. Tables A2 and A3 specify key parameters used in each 2DJ and inversion recovery simulation, respectively.

Sucrose

The chemical shifts and couplings for sucrose were extracted from a GAUSSIAN[95] logfile which ships with SPINACH at the path `SPINACHROOT/examples/standard_systems/sucrose.log`. Isotropic chemical shifts were determined for each spin i via

$$\delta_i = \frac{\text{Tr}(\sigma_i)}{3}, \quad (\text{C.1})$$

where $\sigma_i \in \mathbb{R}^{3 \times 3}$ is the computed chemical shift tensor of spin i .

Strychnine

The strychnine spin system was derived from the SPINACH function `<SPINACHROOT>/etc/strychnine.m`, which returns a spin system specification using chemical shifts and scalar couplings from[120: Ap-

pendix 5], and atomic coordinates from [121: Supplementary Material]. To determine T_1 and T_2 values for each spin, the relaxation superoperator \mathbf{R} was constructed according to Redfield theory, under the assumption that the molecule was undergoing spherical isotropic rotation with a rotational correlation time of 200 ps, in a magnetic field of 700 MHz. Individual T_1 s and T_2 s were then extracted using

$$T_{(1/2),i} = \frac{1}{R_{(1/2),i}}, \quad (\text{C.2a})$$

$$R_{1,i} = -\Re(\mathbf{I}_{z,i}^\dagger \mathbf{R} \mathbf{I}_{z,i}), \quad (\text{C.2b})$$

$$R_{2,i} = -\Re(\mathbf{I}_{+,i}^\dagger \mathbf{R} \mathbf{I}_{+,i}), \quad (\text{C.2c})$$

where $\mathbf{I}_{z,i}$ is the state vector of the Hilbert-space operator \hat{L}_z for spin i , and $\mathbf{I}_{+,i}$ is the corresponding vector for the \hat{L}_+ operator.

TABLE A1: The isotropic chemical shifts (δ), corresponding rotating frame frequencies (ω_{rot}), scalar couplings (J), and relaxation times (T_1, T_2 , if applicable) associated with spin systems used in SPINACH simulations.

Spin	δ (ppm)	ω_{rot} (Hz)	J (Hz)	T_1 (s)	T_2 (s)
Four Multiplets, Run 1					
A	-2.78×10^{-2}	-13.93	E: -9.627, F: -8.202, G: 6.742	—	—
B	-7.11×10^{-3}	-3.56	E: -4.491, F: 5.333, G: 9.303	—	—
C	-1.63×10^{-3}	-0.81	E: 3.953, F: 5.422, G: 5.914	—	—
D	1.53×10^{-2}	7.66	E: -7.902, F: -4.556, G: 6.217	—	—
Four Multiplets, Run 2					
A	-1.48×10^{-2}	-7.38	E: -7.492, F: 0.917, G: 2.933	—	—
B	-1.18×10^{-2}	-5.88	E: -4.304, F: -1.815, G: 5.420	—	—
C	-4.32×10^{-3}	-2.16	E: 4.832, F: 7.573, G: 8.268	—	—
D	1.76×10^{-2}	8.80	E: -9.244, F: -1.816, G: -0.478	—	—
Four Multiplets, Run 3					
A	-2.23×10^{-2}	-11.17	E: -5.347, F: -1.851, G: 1.407	—	—
B	1.13×10^{-2}	5.66	E: 6.425, F: 7.291, G: 9.806	—	—
C	2.53×10^{-2}	12.64	E: -8.640, F: 0.613, G: 6.998	—	—
D	2.84×10^{-2}	14.21	E: -8.613, F: 0.782, G: 3.830	—	—

Continues on next page...

Spin	δ (ppm)	ω_{rot} (Hz)	J (Hz)	T_1 (s)	T_2 (s)
Four Multiplets, Run 4					
A	-2.03×10^{-2}	-10.16	E: -8.646, F: 6.719, G: 7.921	-	-
B	3.53×10^{-3}	1.77	E: -8.857, F: 4.314, G: 9.197	-	-
C	8.61×10^{-3}	4.30	E: -0.620, F: 1.767, G: 6.567	-	-
D	2.06×10^{-2}	10.30	E: -9.060, F: 2.355, G: 9.810	-	-
Four Multiplets, Run 5					
A	-9.16×10^{-3}	-4.58	E: -8.281, F: 1.621, G: 3.229	-	-
B	-2.79×10^{-3}	-1.40	E: 1.655, F: 4.219, G: 6.998	-	-
C	3.00×10^{-3}	1.50	E: -4.280, F: 1.045, G: 5.896	-	-
D	2.74×10^{-2}	13.72	E: -9.316, F: -8.322, G: -3.938	-	-
Sucrose					
A	6.005	1801.6	B: 2.285	-	-
B	3.510	1053.1	A: 2.285, C: 4.657, H: 4.828	-	-
C	3.934	1180.2	B: 4.657, D: 4.326	-	-
D	3.423	1027.0	C: 4.326, E: 4.851	-	-
E	4.554	1366.1	D: 4.851, F: 5.440, G: 2.288	-	-
F	3.891	1167.4	E: 5.440, G: -6.210	-	-
G	4.287	1286.2	E: 2.288, F: -6.210, K: 7.256	-	-
H	3.332	999.5	B: 4.828	-	-
I	1.908	572.3	-	-	-
J	1.555	466.6	-	-	-
K	0.644	193.3	G: 7.256	-	-
L	4.042	1212.5	M: -4.005, S: 1.460	-	-
M	4.517	1355.0	L: -4.005	-	-
N	3.889	1166.7	O: 4.253	-	-
O	4.635	1390.4	N: 4.253, P: 4.448, U: 3.221	-	-
P	4.160	1248.0	O: 4.448, R: 4.733	-	-
Q	4.021	1206.2	R: -4.182	-	-
R	4.408	1322.4	P: 4.733, Q: -4.182, V: 1.350	-	-

Continues on next page...

Spin	δ (ppm)	ω_{rot} (Hz)	J (Hz)	T_1 (s)	T_2 (s)
S	0.311	93.3	L: 1.460	—	—
T	1.334	400.2	—	—	—
U	0.893	267.9	O: 3.221	—	—
V	0.150	45.0	R: 1.350	—	—
Five Multiplets, Run 1					
A	1.33	665.99	F: 10.965, G: 12.657, H: 17.070	2.178	—
B	1.47	735.44	F: 3.610, G: 2.543, H: 8.448	4.430	—
C	1.31	653.87	F: 10.630, G: 6.282, H: 3.012	3.319	—
D	1.41	705.01	F: 8.101, G: 4.589, H: 9.068	1.007	—
E	1.30	650.23	F: 3.014, G: 16.537, H: 15.587	4.992	—
Five Multiplets, Run 2					
A	1.47	733.07	F: 19.488, G: 18.279, H: 3.147	3.600	—
B	1.36	681.25	F: 11.924, G: 8.400, H: 5.515	3.905	—
C	1.30	651.61	F: 13.672, G: 6.543, H: 16.275	4.291	—
D	1.43	713.48	F: 12.007, G: 5.141, H: 9.981	1.687	—
E	1.49	744.53	F: 8.715, G: 14.309, H: 9.805	3.214	—
Five Multiplets, Run 3					
A	1.32	658.87	F: 4.984, G: 18.119, H: 10.642	3.846	—
B	1.44	719.50	F: 13.518, G: 14.381, H: 3.074	1.018	—
C	1.37	683.34	F: 8.758, G: 16.689, H: 12.956	4.766	—
D	1.35	676.98	F: 17.648, G: 7.514, H: 3.918	3.414	—
E	1.49	746.76	F: 16.396, G: 7.455, H: 11.352	1.827	—
Strychnine					
A	7.167	2150.1	B: 7.490, C: 1.080, D: 0.230	1.74	1.26
B	7.098	2129.4	A: 7.490, C: 7.440, D: 0.980	2.47	1.78
C	7.255	2176.5	A: 1.080, B: 7.440, D: 7.900	2.50	1.80
D	8.092	2427.6	A: 0.230, B: 0.980, C: 7.900	5.12	3.69
E	3.860	1158.0	I: 10.410	1.26	0.91
F	3.132	939.6	G: -17.340, H: 3.340	0.52	0.37

Continues on next page...

Spin	δ (ppm)	ω_{rot} (Hz)	J (Hz)	T_1 (s)	T_2 (s)
G	2.670	801.0	F : -17.340, H : 8.470	0.55	0.39
H	4.288	1286.4	F : 3.340, G : 8.470, I : 3.300	0.89	0.64
I	1.276	382.8	E : 10.410, H : 3.300, J : 3.290	0.97	0.70
J	3.150	945.0	I : 3.290, K : 4.110, L : 1.960, R : 1.610, T : 0.470	1.07	0.77
K	2.360	708.0	J : 4.110, L : -14.350, M : 4.330	0.42	0.30
L	1.462	438.6	J : 1.960, K : -14.350, M : 2.420	0.41	0.30
M	3.963	1188.9	K : 4.330, L : 2.420	1.15	0.83
N	1.890	567.0	O : -13.900, P : 5.500, Q : 7.200	0.48	0.35
O	1.890	567.0	N : -13.900, P : 3.200, Q : 10.700	0.47	0.34
P	3.219	965.7	N : 5.500, O : 3.200, Q : -13.900	0.49	0.36
Q	2.878	863.4	N : 7.200, O : 10.700, P : -13.900	0.42	0.30
R	3.716	1114.8	J : 1.610, S : -14.800, T : 1.790	0.47	0.34
S	2.745	823.5	R : -14.800	0.45	0.33
T	5.915	1774.5	J : 0.470, R : 1.790, U : 7.000, V : 6.100	1.48	1.06
U	4.148	1244.4	T : 7.000, V : -13.800	0.48	0.34
V	4.066	1219.8	T : 6.100, U : -13.800	0.55	0.40

Parameter	Four Multiplets	Sucrose
$f_{\text{bf}}^{(1)}$ (MHz)	500	300
$f_{\text{off}}^{(2)}$ (Hz)	0	1000
$f_{\text{off}}^{(2)}$ (ppm)	0	3.333
$f_{\text{sw}}^{(1)}$ (Hz)	40	30
$f_{\text{sw}}^{(2)}$ (Hz)	1000	2200
$f_{\text{sw}}^{(2)}$ (ppm)	2	7.333
$N^{(1)}$	128	64
$N^{(2)}$	1024	4096

TABLE A2: Experiment parameters for 2DJ simulations run using SPINACH.

**Table of spinach experiment parameters Description of 2DJ and Invrec simulations
(i.e. relaxation model used, approximations to basis used etc.)**

Parameter	Five Multiplets
$f_{\text{bf}}^{(1)}$ (MHz)	500
$f_{\text{off}}^{(1)}$ (Hz)	2500
$f_{\text{off}}^{(1)}$ (ppm)	5
$f_{\text{sw}}^{(1)}$ (Hz)	5000
$f_{\text{sw}}^{(1)}$ (ppm)	10
$N^{(1)}$	16384
K	21
τ_{\max} (s)	4

TABLE A3: Experiment parameters for inversion recovery simulations run using SPINACH. K specifies the number of increments run, and τ_{\max} specifies the largest delay time used. Delays were generated with linear spacings, with the first delay always being 0 s, such that the n^{th} delay was $(n-1)\tau_{\max}/K-1$.

C.2 Experimental datasets

C.2.1 Structures

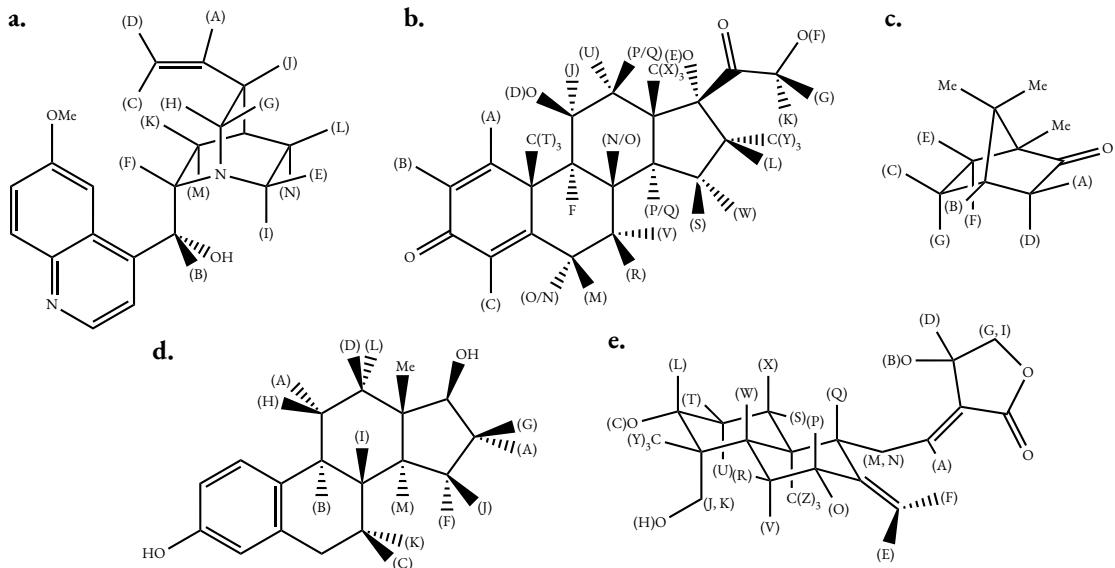


FIGURE A1: The molecular structures of species giving rise to the experimental NMR datasets considered in this work. **a.** Quinine, **b.** Dexamethasone, **c.** Camphor, **d.** 17β -estradiol, **e.** Andrographolide. Proton environments giving rise to signals which are considered in this work are denoted with bracketed alphabetical characters. Non-bracketed alphabetical characters denote chemical symbols. Me denotes the methyl group, equivalent to CH_3 . **TODO: check assignments (esp. estradiol). If more structures need adding, edit the ChemDraw file on Chive called simon_stuff/thesis_structures.cdxml. Use EB-Garamond for atom labels. One a structure is made, scale to 75% of original size, and set font to 7pt. Then in inkscape, rescale this by multiplying by 0.8.**

Spin	Coupling partners	Apparent multiplet structure
(A)	(D) ^l , (M) ^v , (N) ^v	dt
(B)	(D) ^{g,ex}	d
(E)	(F) ^{g,v} , ...	d...
(F)	(E) ^{g,v} , ...	d...
(K)	(J) ^g , (H) ^{ex}	d
(L)	(C) ^{ex} , (T) ¹⁸⁰ , (U) ⁶⁰	dd
(O)	(P) ^g , (R) ⁶⁰ , (V) ⁶⁰	ddd
(P)	(O) ^g , (R) ⁶⁰ , (V) ¹⁸⁰	dt
(Q)	(M) ^v , (N) ^v	dd
(V)	(O) ⁶⁰ , (P) ¹⁸⁰ , (R) ^g , (W) ¹⁸⁰	dq

TABLE A4: Coupling partners associated with spins whose signals are considered in Figure 3.2. Coupling partners are labelled as follows: ^{g,v} geminal coupling between two vinylic protons, ^{g,ex} geminal coupling between two protons, in which one is bonded to an oxygen, leading to exchange decoupling, ^g geminal coupling, ^l long-range coupling, ^v vicinal coupling, ⁶⁰ geminal coupling, with a fixed dihedral angle of 60°, ¹⁸⁰ geminal coupling, with a fixed dihedral angle of 180°.

C.2.2 Datasets related to CUPID

The 2D J-Resolved datasets presented were acquired on various Bruker NMR spectrometers. In each case `jresqf` from Bruker's pulse sequence library was used which comprises: $\pi/2(\Phi_1) \rightarrow t^{(1)}/2 \rightarrow \pi(\Phi_2) \rightarrow t^{(1)}/2 \rightarrow t^{(2)}(\Phi_{rx})$, with the EXORCYCLE phase-cycling scheme[27: Section 11.6]:

$$\begin{aligned}\Phi_1 : & \quad 0^\circ \quad 0^\circ \quad 0^\circ \quad 0^\circ \\ \Phi_2 : & \quad 0^\circ \quad 90^\circ \quad 180^\circ \quad 270^\circ \\ \Phi_{rx} : & \quad 0^\circ \quad 180^\circ \quad 0^\circ \quad 180^\circ\end{aligned}$$

Key experiment parameters are provided in Table A5.

The pulse sequence used for the acquisition of the estradiol PSYCHE spectrum (Figure TODO) is presented in Figure A2, where pulses, gradients, and delays are described in detail. Equivalent parameters were used for the basic setup as the estradiol 2DJ experiment, given in Table A5.

The dexamethasone PSYCHE spectrum (Figure 3.11.a) was generated using the TSE-PSYCHE experiment, which is presented in Figure A3.

C Information on Datasets

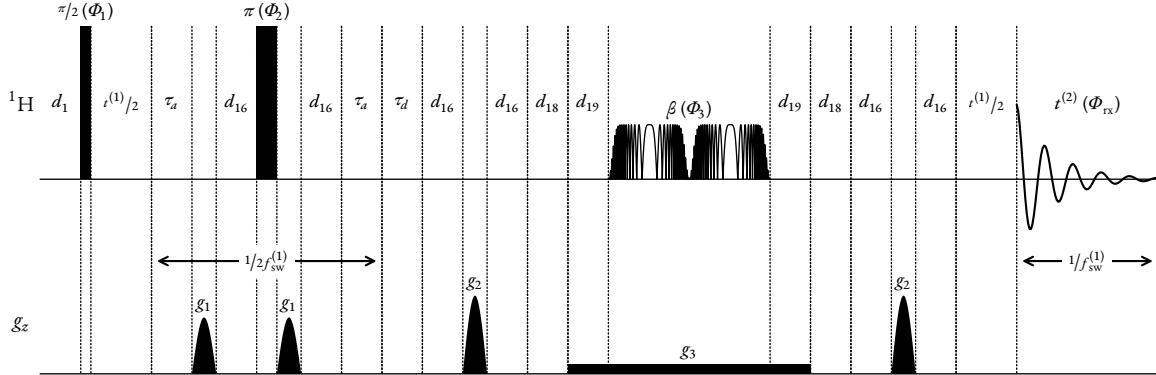


FIGURE A2: PSYCHE pulse sequence used for the acquisition of estradiol data. All delays are included, though they are not to scale. Delays: d_0 (relaxation delay): 1 s, d_{16} : (gradient recovery delay): 200 μ s, d_{18} : 200 μ s, d_{19} : 1 ms, τ_a : 1.3 ms, τ_d : 18.9 ms. The PSYCHE element featured two saltire chirp pulses with a wideband, uniform rate, smooth truncation (WURST)[122] amplitude envelope, with a target flip angle $\beta = 20^\circ$. Each saltire pulse had a bandwidth of 10 kHz, a duration of 25 ms, and a power of 280 μ W. Hard pulses had a power of 31.537 W, with the duration of the $\pi/2$ pulse being 15 μ s. G_1 and G_2 were gradients for coherence order selection. Each comprised a 100-point sine shape profile, and lasted 1 ms. G_3 was a rectangular weak gradient applied during the PSYCHE element, with a duration of 52 ms. The gradeint strengths as a percentage of the maximum permissible z-gradient were, respectively 31%, 47%, 1.6%. The phase cycling scheme used was: $\Phi_1 : 2 \times (0^\circ, 180^\circ)$; $\Phi_2 : 4 \times 0^\circ$; $\Phi_3 : 2 \times 0^\circ, 2 \times 90^\circ$; $\Phi_{rx} : 0^\circ, 180^\circ, 180^\circ, 0^\circ$.

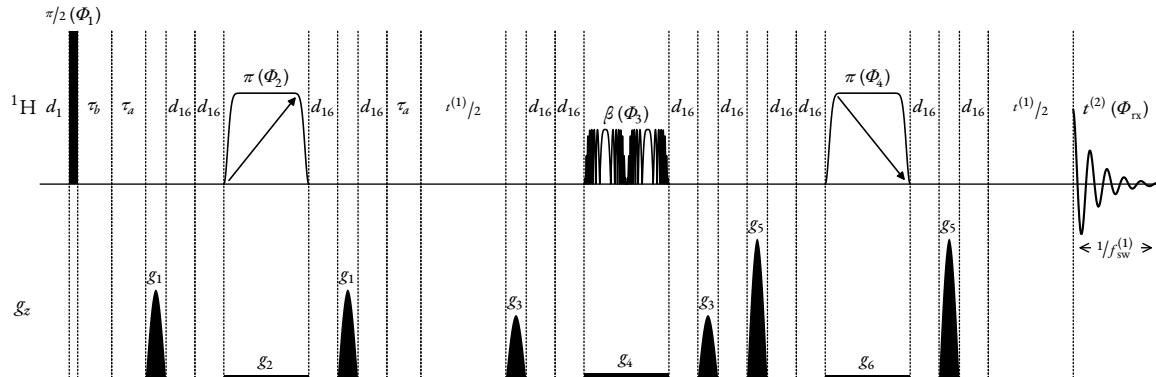


FIGURE A3: TSE-PSYCHE pulse sequence used for the acquisition of dexamethasone data (Figure 3.11.a). Delays: d_1 (relaxation delay): 2 s, d_{16} : (gradient recovery delay): 200 μ s, τ_a : 5 ms ($= 1/4f_{sw}^{(1)}$). The hard $\pi/2$ pulse had a duration of 12 μ s, and a power of 24 W. The two π pulses were unidirectional frequency-swept (chirped) pulses, with the first pulse sweeping from low to high frequencies, and the second pulse sweeping from high to low. These each had a WURST amplitude envelope, lasted a duration of 40 ms, and had a power of 11.05 mW. The PSYCHE element had a target flip angle $\beta = 15^\circ$, and featured two saltire chirp pulses. Both saltire pulses had a WURST amplitude envelope, a duration of 15 ms, and a power of 1.28 mW. g_1 , g_3 and g_5 were gradients for coherence order selection. Each comprised a 100-point sine profile, and lasted 1 ms. g_2 , g_4 and g_6 were weak rectangular gradients which were applied at the same time as the chirped pulses. The magnitudes of gradients g_1 to g_6 as a percentage of the maximum permitted gradient were, respectively: 49%, 2%, 35%, 3%, 77%, 2%. The phase cycling scheme used was: $\Phi_1 : 8 \times 0^\circ$; $\Phi_2 : 2 \times (2 \times 0^\circ, 2 \times 180^\circ)$; $\Phi_3 : 2 \times (0^\circ, 90^\circ), 2 \times (180^\circ, 270^\circ)$; $\Phi_4 : 8 \times 0^\circ$; $\Phi_{rx} : 4 \times (0^\circ, 180^\circ)$.

	Quinine	Dexamethasone	Camphor	Estradiol
f_{bf} (MHz)	500.13	600.18	500.13	500.3
$f_{\text{off}}^{(2)}$ (Hz)	2500	2815.4	1000	2501.5
$f_{\text{sw}}^{(1)}$ (Hz)	50	50	50	100
$f_{\text{sw}}^{(2)}$ (Hz)	7500	7211.5	5000	5000
$f_{\text{sw}}^{(2)}$ (ppm)	14.996	12.016	9.9974	9.994
$N^{(1)}$	128	64	128	128
$N^{(2)}$	16384	8192	16384	16384
NS	4	2	4	4
DS	4	8	4	2
PLW1 (W)	20.893	24	20.893	31.537
P1 (μs)	10	12	10	15
D1 (s)	2	1.5	2	1

TABLE A5: Noteworthy experiment parameters for the 2D J-Reolved and PSYCHE experiments run. NS: Number of scans, DS: Number of dummy scans, PLW1: Hard pulse power (W), P1: Duration of $\pi/2$ pulse, D1: Duration of relaxation delay.

MISCELLANEOUS TOPICS OF INTEREST

D

D.0.1 The Nyquist Frequency

The Nyquist frequency defines the highest frequency an oscillator can possess such that it is sampled at least twice per period:

$$f_N = \frac{1}{2\Delta t} \quad (\text{D.1})$$

This quantity is at the heart of the Nyquist-Shannon Sampling Theorem, which - in the words of Claude Shannon[123] - states:

If a function $y(t)$ contains no frequencies higher than f Hz, it is completely determined by giving its ordinates at a series of points spaced $1/2f$ seconds apart.

The implication of this is that the continuous FID $y(t)$ may be completely described by its digitisation with sampling interval Δt so long as the frequencies f_j that make it up satisfy:

$$-\frac{1}{2\Delta t} \leq f_j \leq \frac{1}{2\Delta t} \quad \forall j \quad (\text{D.2})$$

If any frequency f_j does not satisfy (??), the signal will be insufficiently sampled, and will be spuriously represented as having a frequency f_a that satisfies

$$f_a = 2mf_N - f_j \quad (f_j > f_N) \quad (\text{D.3})$$

$$f_a = 2mf_N + f_j \quad (f_j < -f_N), \quad (\text{D.4})$$

where $m \in \mathbb{Z}_+$. This phenomenon is called *aliasing*.

D.0.2 Hessian Validation

Talk about checking that my Hessian code is correct by using the finite difference method

D.0.3 Statistical Definitions

Definitions of standard error, likelihood function, ...