Mathematics of Biomolecular Structure Experiments

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium”

der Georg-August-Universität Göttingen

im Promotionsprogramm

“Mathematical Sciences (SMS)”

der Georg-August University School of Science (GAUSS)

vorgelegt von

Henrik Wiechers

aus Holzminden

Göttingen, 2023

Thesis Committee:

Prof. Dr. Stephan Huckemann
Institute for Mathematical Stochastics, University of Göttingen

PD Dr. Benjamin Eltzner
Research Group Computational Biomolecular Dynamics, Max Planck Institute for Multidisciplinary Sciences

Dr. Yvo Pokern
Department of Statistical Science, University College London

Members of the Examination Board:

Reviewer:
Prof. Dr. Stephan Huckemann
Institute for Mathematical Stochastics, University of Göttingen

Second Reviewer:
PD Dr. Benjamin Eltzner
Research Group Computational Biomolecular Dynamics, Max Planck Institute for Multidisciplinary Sciences

Further Members of the Examination Board:

Dr. Yvo Pokern
Department of Statistical Science, University College London

Prof. Dr. Dominic Schuhmacher
Institute for Mathematical Stochastics, University of Göttingen

Jun.-Prof. Dr. Anne Wald
Institute for Numerical and Applied Mathematics, University of Göttingen

Prof. Dr. Damaris Schindler
Mathematical Institute, University of Göttingen

Date of the Oral Examination: September 4, 2023

Acknowledgments

I would like to express my deepest gratitude to all those who have contributed to the completion of this work and supported me over the past years.

First, I would like to thank my supervisor **Stephan Huckemann** for his supervision since my time in the Master's program and for giving me the opportunity to become his PhD student. I would like to thank you not only for the intensive scientific supervision, but also for the opportunity to participate in various conferences that allowed me to develop and meet many interesting people. Furthermore, I would like to thank you for many pleasant conversations, for example in the regular tea sessions with the working group, which contributed to a positive working atmosphere.

I would like to thank my second supervisor **Benjamin Eltzner**, who supervised me since my master thesis and always had an open ear for all my questions. The cooperation during the master thesis was the main reason to start my doctoral studies. I want to thank you for the many conversations, both scientific and personal, which significantly increased the quality of my work and contributed to a very pleasant and productive working atmosphere.

I would like to thank my third supervisor **Yvo Pokern** for supervising me since the beginning of my PhD studies. I would like to thank you for all the discussions about the research field of ENDOR spectroscopy and the related mathematics, which allowed me to quickly get into the research field and increased the quality of my work throughout my PhD studies. In particular, I highly appreciated the collaboration during your research visits in Göttingen.

Furthermore, I would like to thank **Kanti Mardia** for the opportunity to work with him. Your motivation and energy in conducting research are truly inspiring to me and have greatly improved the quality of my work.

I would also like to thank the 'Electron Paramagnetic Resonance' research group. Especially, I would like to express my gratitude to **Marina Bennati**, **Annemarie Kehl**, **Markus Hiller**, **Igor Tkach**, **Andreas Meyer** and **Laura Remmel** for the pleasant and effective interdisciplinary collaboration. In particular, I would like to thank you for presenting our research together at the Collaborative Research Center 1456 meetings, working on the papers together, and giving me a tour of the very impressive research laboratory.

I am grateful for the support and work of the **Collaborative Research Center 1456**. The regular retreats and financial support for conferences diversified my research interests.

I would like to thank **Markus Zobel**, who collaborated with me as part of his bachelor's thesis and subsequently worked with me on a joint paper. I would like to thank **Franziska Hoppe**, who is currently working with me on her master's thesis. In addition, I would like to thank **Weishi Chen** for the conversations about the ENDOR project and **Rajan Alexander** for discussing our joint research.

I would like to thank the Richardson Laboratory (consisting of **Jane Richardson**, **David Richardson**, **Michael Prisant**, **Vincent Chen**, and **Christopher Williams**) and **Ezra Miller** for the recent collaborations.

I would like to thank all my amazing colleagues who have always contributed to a positive working atmosphere in the office. Lunching together in the canteen, playing soccer together on Tuesday evenings, eating cake together and much more have greatly loosened up the working atmosphere. In particular, I would like to thank **Christian Böhm**, who always took care of my technical problems and saved my laptop when it was broken.

I would like to thank my best friends **Jonas** & **Julia**, **Stefan** & **Linda**, **Max**, **Mariko**, **Lina** & **Tayfun** and **Arvid** & **Yvonne** for the wonderful time together. The moments of shared game nights, vacations, meals in the canteen and walks guided by *Wolfgang Dahms' hiking books* are always something very special to me.

I would like to give a special thanks to my partner **Victoria** for the wonderful time we spend together. Your emotional support, your understanding of the often unusual working hours, your trust in me, your encouraging words and the time we spend together on excursions and with our friends (and much more) are very special to me.

Last but not least, I would like to thank my family for their constant support in all circumstances. I would like to thank my parents **Ute** and **Andreas** for the permanent advice, support, the joint hiking vacations, and much more. I would like to thank my sisters **Lisa** and **Anna** and my brother-in-law **Sören** for an always trustful and warm relationship. I would like to thank my grandparents **Ingrid**, **Heinz** and **Thea** who were always there for me. I would also like to thank the rest of my family for the always cordial relationship.

Contents

CHAPTER 1

Introduction

This dissertation comprises and summarizes four articles, which are appended as Sections A, B, C and D in the Addenda and explain the results of this dissertation in full detail.

A large part of this work is based on a collaboration with the 'Electron Paramagnetic Resonance' research group of Marina Bennati at the Max Planck Institute for Multidisciplinary Sciences in the framework of the *Collaborative Research Center* (CRC) 1456 with the name *Mathematics of Experiment*. In this project, new statistical methods for *electron-nuclear double resonance* (ENDOR) spectroscopy experiments have been developed and published in Paper C and Paper D. In addition, a presentation of my master thesis, which dealt with statistical modeling of ribonucleic acid (RNA) molecules, led to a collaboration with the famous statistician and major long-term protagonist in directional statistics Kanti Mardia (Senior Research Professor at University of Leeds and Visiting Professor at Oxford University). The results of the collaboration have been published in Paper A and Paper B and more recently led to a joint project with the Richardson Laboratory from the Duke Department of Biochemistry (consisting of Jane Richardson, David Richardson, Michael Prisant, Vincent Chen, and Christopher Williams) and Ezra Miller from the Mathematics Department of the Duke University.

The guiding methodological principle that recurs throughout the various projects is the concept of *generalized Fréchet means*, which we briefly introduce below. The expected value $\mu := \mathbb{E}(X)$ of an n-dimensional real random vector X with existing second moment, i.e. $\mathbb{E}(X^T X) < \infty$, can be defined as the minimizer of the expected squared Euclidean distance

$$\arg\min_{y \in \mathbb{R}^n} \mathbb{E}\left((X-y)^T(X-y)\right) = \arg\min_{y \in \mathbb{R}^n} \left(y^T y - 2y^T \mu\right) = \{\mu\}\,.$$

Fréchet (1948) generalized this geometrical property as a definition for a *mean location* (see Hendriks and Landsman (1998)) on a metric space, which was soon called the *Fréchet mean* in his honor. The Fréchet mean of a random variable X taking values in a metric space $(\mathfrak{Q}, d)$ is an element of the set

$$E^{(d^2)} := \arg\min_{q \in \mathfrak{Q}} \mathbb{E}(d^2(X, q))$$

where $E^{(d^2)}$ is called the *set of population Fréchet means*. For *Hadamard spaces* (complete metric spaces with global non-positive curvature), such as Euclidean spaces, it was proven by Sturm (2003) that the Fréchet mean is unique.

However, in general, $E^{(d^2)}$ may consist not only of a single element, but may also be empty, or it may consist of several elements. For instance, for $(\mathfrak{Q}, d) = (\mathbb{S}^2, d_{\mathbb{S}^2})$ with spherical distance $d_{\mathbb{S}^2}(x, y) = \arccos(x^T y)$, the set of population Fréchet means $E^{(d^2_{\mathbb{S}^2})}$ for a random variable X that only takes values on the north and south poles with the same probability of $1/2$ correspond to the entire equator, which follows directly from the spherical coordinate representation where $q(\theta, \phi) \coloneqq (\sin(\theta)\cos(\phi),\ \sin(\theta)\sin(\phi),\ \cos(\theta))^T \in \mathfrak{Q}$:

$$\underset{\theta\in[0,\pi),\phi\in[0,2\pi)}{\arg\min}\ \mathbb{E}(d^2_{\mathbb{S}^2}(X, q(\theta,\phi))) = \underset{\theta\in[0,\pi),\phi\in[0,2\pi)}{\arg\min}\left(\frac{1}{2}\theta^2 + \frac{1}{2}(\pi-\theta)^2\right) = \left\{(\theta,\phi) : \theta = \frac{\pi}{2},\ \phi \in [0, 2\pi)\right\}.$$

In contrast, for $(\mathfrak{Q}, d) = (\mathbb{R} \setminus \{0\}, d_{\mathbb{E}})$ with metric function $d(x, y) = |x - y|$, the set of Fréchet means $E^{(d^2_{\mathbb{E}})}$ for a random variable taking only the values $+1$ and -1 with the same probability of $1/2$ corresponds to the empty set. Huckemann (2011b) extended this concept to generalized Fréchet means (defined in Definition 1.1) as follows. Let X be a random variable mapping into the *data space* $\mathfrak{Q}$, which is a general topological space equipped with the Borel σ-algebra. Moreover, the *parameter space* $(\mathfrak{P}, d)$ is defined as a metric space with the metric function $d : \mathfrak{P} \times \mathfrak{P} \mapsto [0, \infty)$ and the topology induced by the metric. Then for $\rho : \mathfrak{Q} \times \mathfrak{P} \mapsto \mathbb{R}$ which is continuous in $\mathfrak{P}$ for all fixed $q \in \mathfrak{Q}$ and measurable in $\mathfrak{Q}$ for all fixed $p \in \mathfrak{P}$, the set of generalized population Fréchet means is defined as

$$E^{(\rho)} \coloneqq \underset{p\in\mathfrak{P}}{\arg\min}\, \mathbb{E}\left(\rho(X, p)\right).$$

For an independent and identically distributed (i.i.d.) random sample $X_1, \ldots X_n \sim X$ one then defines the *generalized sample Fréchet mean* as an estimator of the population by the following

$$E_n^{(\rho)}(\omega) \coloneqq \underset{p\in\mathfrak{P}}{\arg\min} \sum_{i=1}^{n} \rho(X_i(\omega), p).$$

Generalized Fréchet means are a very versatile modeling tool that comprises geometric objects like Fréchet means, *L^p-Fréchet means* using $\rho(X, p) = d^p(X, p)$ (see Afsari (2009, 2011)), *extrinsic means* (see Hendriks and Landsman (1996, 1998)), geodesics (see Fletcher and Joshi (2004); Huckemann and Ziezold (2006); Huckemann et al. (2010)) and *submanifolds*, especially *backward nested subspaces* (see Jung et al. (2012); Huckemann and Eltzner (2018)). The formulation given here is identical to the earlier concept of *M-estimators* (see for example van der Vaart (2000)) which include *Maximum Likelihood* (ML) estimators and a wide variety of other estimators of interest, especially *robust estimators* like the minimizer of the *Huber loss* (see Huber (1964)). Since we will mostly be concerned with geometrical objects, we will use the term 'generalized Fréchet mean' throughout this thesis. It is highlighted whenever generalized Fréchet means and the associated methodology and theory are used.

This work has been motivated by biomolecular structure reconstruction. There is a wide range of different methods to determine the structure of biomolecules, which are applicable in different cases (see Section 2 for an overview). This naturally leads to data at different resolutions (Figure 1.1 shows exemplarily two different RNA structures measured at two different resolutions) and to the question of how to model data at different scales and develop learning algorithms. To tackle this issue, Paper B introduces a novel approach to model RNA strands at two scales, the *microscopic* (atomic level) and the *mesoscopic* (intermediate scale between the microscopic scale and *macroscopic* scale (e.g. the whole RNA strand)). At the microscopic scale, we work with suites (see Figure 2.1) which can be represented on the seven-dimensional *torus* $\mathbb{T}^7$ (introduced in Section 1.1.1). At the *mesoscopic* scale, we work with *mesoscopic shapes*, which are modeled in the *size-and-shape space* $S\Sigma_3^6$ (introduced in Section 1.1.2). In order to learn clash-free corrections for both scales, we developed a new clustering method in Paper A, which can be applied to data in general metric spaces. It consists of an *iterative pre-clustering* and a post-clustering which separates clusters with statistical significance, based on the dimension reduction with *Principal Nested Spheres* (PNS) of Jung et al. (2012) (introduced in Section 1.2), which gains power due to the statistically advantageous geometry of spheres. The correction on both scales is based on the concept of Fréchet means from the classes learned with this clustering. Note that PNS can be integrated into the framework of generalized Fréchet means, for which Huckemann and Eltzner (2018) prove strong consistency and a central limit theorem.

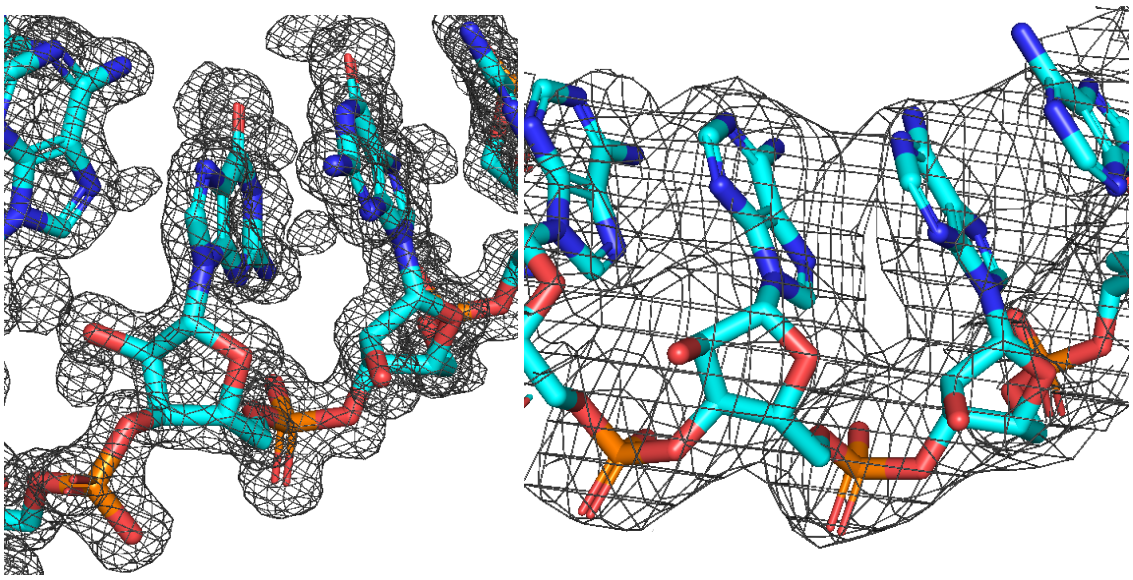


Figure 1.1: Reconstructed RNA structure and electron density contour surface created with PyMOL at level of one σ (see see Schrödinger, LLC (2015)) at resolution 1.6 Å (left, from Ippolito and Steitz (2000)) and at resolution 3 Å (right, from benchmark file 1f8v). The Figure is taken from Paper B.

ENDOR spectroscopy (introduced in Section 2.2) can be used to determine intramolecular distances (see right panel of Figure 1.2). For this purpose, two different challenges have been worked on. The first challenge is to *denoise* the data and is addressed in Paper C: during an ENDOR experiment, the

spectrometer stores a data matrix $Y \in \mathbb{C}^{B\times N}$, where B is the number of *batches* and N is the number of *frequencies* (see left panel of Figure 1.2), from which a real-valued spectrum $\hat{I} \in \mathbb{R}^N$ has to be estimated. In Pokern et al. (2021), the *homoscedastic drift model* for experiments at a microwave frequency of 263 GHz was presented, which is a parametric model that takes into account various thermal drifts occurring in experiments and achieves a very good model fit in practice, see Pokern et al. (2021); Hiller et al. (2022) and Paper D. The homoscedastic drift model allows the application of the *parametric bootstrap*, which in turn enables *hypothesis testing* and *confidence intervals* for the spectra, see denoising step of Figure 1.2.

In Paper C, we first develop asymptotic theory for the estimation of the spectrum with the homoscedastic drift model. In the homoscedastic drift model, the complex-valued maximum likelihood estimator $\hat{\kappa}$, which is required to satisfy $\sum_\nu |\hat{\kappa}_\nu|^2 = 1$ for identifiability reasons, of the parameter κ is computed. Subsequently, κ is rotated with a suitable complex rotation by $\lambda \in [0, 2\pi)$ radians so that the estimated spectrum $\hat{I} = \Re(e^{i\lambda}\kappa)$ is the real part. Consequently, we naturally obtain the *complex projective space* (defined in Section 1.1.3) as *parameter space* for κ (see Section 2.2 for a more detailed explanation). To prove strong consistency for the estimation of κ, we extend the theory of strong consistency for generalized Fréchet means from Huckemann (2011b) (see Section 1.4). Subsequently, a central limit theorem for the estimation of both κ and I is proven. Secondly, we extend the homoscedastic drift model to cover experiments at a microwave frequency of 94 GHz, for which ENDOR spectrometers are more widely available.

In Paper D we address the issue of how physical parameters describing the conformation of the biomolecule are estimated (see optimize step from Figure 1.2), including statistically rigorous error propagation from the spectral uncertainties made accessible by the homoscedastic drift model. To this end, we drastically accelerated a *spectrum simulation code* to enable optimizations (a simulated spectrum is depicted in red in the optimize step of Figure 1.2). Building on this, a Bayesian optimization-based pipeline was implemented and successfully applied to ENDOR data (see Section 2.2 or Paper D for more details).

The thesis is structured as follows. This section describes the mathematical methods used and extended in this work. Section 1.1 introduces the torus, *landmark-based shape spaces*, and the complex projective space. Then, we discuss prior work on generalizations of *principal component analysis* (PCA) for non-Euclidean data in Section 1.2. Section 1.3 gives an overview of the topic of clustering and motivates our clustering algorithm developed in Paper A and used in Paper B. Subsequently, in Section 1.4, we review strong consistency for generalized Fréchet means and point out how we generalize two results in this field. In Section 2, an overview of different methods, all aiming to obtain structural information of biomolecules, is presented. As part of this, the necessary fundamentals for both RNA structure analysis and ENDOR experiments are introduced and subsequently the individual papers are summarized. In addition, for all the papers, my own contributions are highlighted. Finally, in Section 3 an outlook is given, including some developments which are currently in progress.

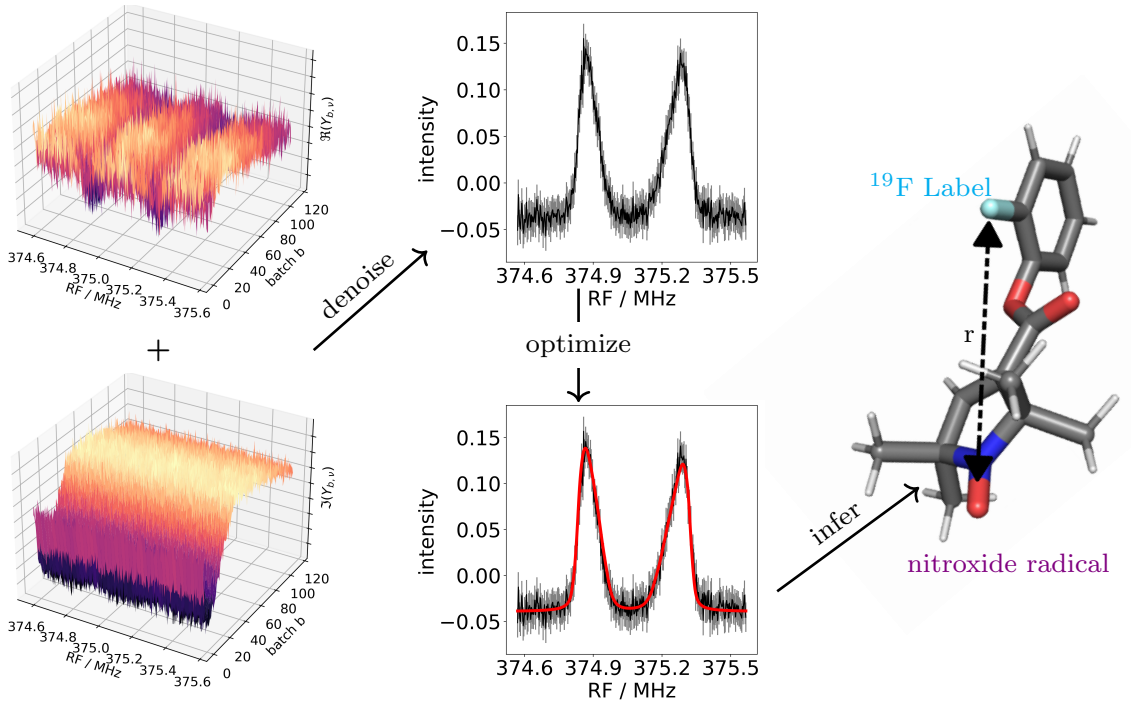


Figure 1.2: The ENDOR experiment for the fluorine-nitroxide compound (analyzed in Paper D). Left: the real part (top) and the imaginary part (bottom) of the raw data matrix Y of orientation g_x. Center: the data matrix is denoised with the homoscedastic drift model from Pokern et al. (2021) (top) to obtain the estimated spectrum $\hat{I}$ (including pointwise confidence bands obtained by parametric bootstrap). In a second step (center bottom), a spectrum simulation software is used to search for the physical parameters such that the spectrum simulated from those (plotted in red) are closest to the estimated spectrum. Right: the corresponding energy-minimized structure of Meyer et al. (2020) predicted by DFT calculations. The *interspin distance* r between the nitroxide radical electron (mean location at the bond between the nitrogen atom (blue) and the oxygen atom (red) that are part of the nitroxide radical) and the ^{19}F nucleus can be inferred from the estimated physical parameters.

1.1 Non-Euclidean data and parameter spaces

This section serves as an introduction to the manifolds used in the papers. We represent the shape of biomolecules in Paper A and Paper B either using the torus (Section 1.1.1) or using landmark-based shape spaces (Section 1.1.2). In Paper C and Paper D we work with the complex projective space (Section 1.1.3), which is the well-known planar similarity shape space.

1.1.1 Sphere and Torus

The following serves as an introduction to Paper A and Paper B and is thus close in content to parts of the respective papers.

We define the k-dimensional real unit sphere $\mathbb{S}^k := \{y \in \mathbb{R}^{k+1} : \|y\| = 1\}$. It is a metric space, which is usually equipped with the spherical distance

$$d_{\mathbb{S}^k}(x, y) := \arccos(x^T y), \quad \text{for} \quad x, y \in \mathbb{S}^k.$$

To define the m-dimensional (flat) torus we first start with the one-dimensional case. The one-dimensional torus is defined by

$$\mathbb{T} := [0, 2\pi]/\sim,$$

where $\sim$ denotes the identification of 0 with 2π. It is a metric space with canonical distance

$$d_{\mathbb{T}}(\phi, \psi) = \min\{|\phi - \psi|, 2\pi - |\phi - \psi|\}, \quad \text{for} \quad \phi, \psi \in \mathbb{T}.$$

It follows directly that $\mathbb{T} \cong \mathbb{S}^1$, however, we use $\mathbb{T}$ for consistent notation. The m-dimensional torus $\mathbb{T}^m$ is a metric space defined by the canonical product of m one-dimensional tori. It is a metric space with canonical distance

$$d_{\mathbb{T}^m}(\phi, \psi) = \sqrt{\sum_{j=1}^{m} d_{\mathbb{T}}(\phi_j, \psi_j)^2}, \quad \text{for} \quad \phi = (\phi_1, \ldots, \phi_m), \psi = (\psi_1, \ldots, \psi_m) \in \mathbb{T}^m.$$

For the torus case, the Fréchet mean is also called *torus mean*. Specialized methods for the torus have been developed by several authors in particular to represent biomolecules, see e.g. Altis et al. (2008); Kent and Mardia (2009); Sargsyan et al. (2012); Eltzner et al. (2018); Zoubouloglou et al. (2022). A detailed discussion about dimension reduction on the torus is given in Section 1.2.

1.1.2 Landmark-based shape spaces

The following content serves as an introduction to landmark-based shape spaces and is summarized from Dryden and Mardia (2016).

Intuitively speaking, the *shape* is the geometric information that remains after *location*, *scale* and *rotation* effects have been filtered out. We are often interested in keeping the information about the *size*, i.e. the scale information is not filtered out, which leads to the concept of *size-and-shape*.

In landmark-based shape spaces, k different *landmarks* $x_1, \ldots, x_k$ (usually in $\mathbb{R}^m$) are considered, which form a *configuration matrix* $X = (x_1, \ldots, x_k)^T \in M$, where M is typically a subset of $\mathbb{R}^{k \times m}$. The landmarks are considered modulo a group action of a group G on the set of landmarks M to compare objects with each other. This leads to a variety of different *shape spaces* M/G.

Probably the best known shape space is the *similarity shape space* Σ_m^k introduced by Kendall (1977), where we have k landmarks in $\mathbb{R}^m$ and G is the group of *proper (i.e. orientation preserving) similarity transformations*, which include rotation, translation and scale, $T^{(\Sigma)} = (r, R, v) \in \mathbb{R}_+ \times \mathrm{SO}(m) \times \mathbb{R}^m$

and act on X via

$$T^{(\Sigma)}.X := (rRx_1 + v, \ldots, rRx_k + v)^T.$$

Then the similarity shape space is defined by performing each group action for convenience in the following sequence. The *translational shape space* $R^{k\times m}/R^m$ resulting from the group of translations is isometric to $R^{(k-1)\times m}$. Each *orthogonal complement* $H \in R^{(k-1)\times k}$ of $e_0 := \frac{1}{\sqrt{k}}(1, \ldots, 1)^T$ (i.e. $(H^T \mid e_0) \in \mathbb{R}^{k\times k}$ is an orthogonal matrix) defines an isometry, as for all $X \in [X], Y \in [Y] \in \mathbb{R}^{k\times m}/\mathbb{R}^m$ it holds

$$d_{R^{k\times m}/R^m}([X],[Y]) := \min_{v\in\mathbb{R}^m} \|X - e_0 v^T - Y\| = \|HX - HY\|.$$

One possible orthogonal complement is the *Helmert sub-matrix* where the row vectors are the *Helmert orthonormal basis vectors*

$$h_j := \frac{1}{\sqrt{j(j+1)}}\left(\left(\sum_{m=1}^{j} e_m\right) - j e_{j+1}\right), \quad j = 1, \ldots, k-1.$$

Landmarks that are transformed with the Helmert sub-matrix are called *Helmertized landmarks*. By multiplying the Helmertized landmarks from the left with H^T, an isometry is defined between $\mathbb{R}^{(k-1)\times m}$ and the *centered landmarks* $\{X \in \mathbb{R}^{k\times m} : e_0^T X = 0\} \subset R^{k\times m}$. Landmarks transformed with $H^T H = \mathrm{Id}_k - \frac{1}{k} e_0 e_0^T$ are called centered landmarks. Removing the configurations, $\{e_0 v^T : v \in \mathbb{R}^m\} \subset \mathbb{R}^{k\times m}$ where all landmarks are the same, corresponds to removing the origin in $\mathbb{R}^{(k-1)\times m}$. Then

$$\left(\mathbb{R}^{(k-1)\times m} \setminus \{0\}\right)/R_+$$

has the structure of a unit sphere $\mathbb{S}^{(k-1)-1\times m}$, which is called *pre-shape sphere*. The group of rotations acts isometrically on the pre-shape sphere and the quotient

$$\begin{aligned}\Sigma_m^k := \mathbb{S}^{(k-1)-1\times m}/\mathrm{SO}(m) &\cong \left(\mathbb{R}^{k\times m} \setminus \{e_0 v^T : v \in \mathbb{R}^m\}\right) / (\mathbb{R}_+ \times \mathrm{SO}(m) \times \mathbb{R}^m) \\ &\cong \left(\{X \in R^{k\times m} : e_0^T X = 0\} \setminus \{e_0 v^T : v \in \mathbb{R}^m\}\right) / (\mathbb{R}_+ \times \mathrm{SO}(m))\end{aligned}$$

is called similarity shape space. The similarity shape space is for example equipped with the *full Procrustes distance* d_F, the *Procrustes distance* ρ or the *partial Procrustes distance* d_P, which are defined respectively by

$$\begin{aligned} d_F([X],[Y]) &:= \min_{r\in\mathbb{R}, R\in\mathrm{SO}(m)} \|X - rYR\| \\ \rho([X],[Y]) &:= \min_{R\in\mathrm{SO}(m)} d_{\mathbb{S}^{m\times(k-1)-1}}(X, YR) \\ d_P([X],[Y]) &:= \min_{R\in\mathrm{SO}(m)} \|X - YR\| \end{aligned}$$

where $X \in [X], Y \in [Y]$ are pre-shapes. We call the corresponding Fréchet means *full Procrustes mean*, *Procrustes mean* and *partial Procrustes mean*, respectively.

In the applications in Paper B we are working with the size-and-shape space: let $M := \mathbb{R}^{k\times m}$ and G correspond to the *proper (i.e. orientation preserving) Euclidean transformations* involving rotations and translations, $T^{(S\Sigma)} := (R, v) \in \mathrm{SO}(m) \times \mathbb{R}^m$ which act on X via

$$T^{(S\Sigma)}.X := (Rx_1 + v, \ldots, Rx_k + v)^T .$$

Then the *size-and-shape space* is defined by

$$S\Sigma_m^k := \{[X] : X \in \mathbb{R}^{k\times m}\} \quad \text{where} \quad [X] := \{T^{(S\Sigma)}.X : T^{(S\Sigma)} \in \mathrm{SO}(m) \times \mathbb{R}^m\}. \tag{1.1}$$

Analogous to the procedure for the similarity shape space follows

$$S\Sigma_m^k \cong \mathbb{R}^{(k-1)\times m}/\mathrm{SO}(m) \cong \{X \in R^{k\times m} : e_0^T X = 0\}/\mathrm{SO}(m).$$

In our applications in Paper A and Paper B we work with centered landmarks and equip them with the partial Procrustes distance

$$d_P([X], [Y]) \quad := \quad \min_{R\in\mathrm{SO}(m)} \|X - YR\| \tag{1.2}$$

where $X \in [X], Y \in [Y]$. On $S\Sigma_m^k$ we call the Fréchet mean defined by the partial Procrustes distance *partial Procrustes mean*.

1.1.3 Planar shape spaces and complex projective space

In this section, we introduce the complex projective space and highlight its relation to the planar shape space. For $k \in \mathbb{N}_0$ the *complex-valued unit sphere* is defined by

$$\mathcal{S}^{2k+1} := \{z \in \mathbb{C}^{k+1} : \|z\| = 1\} \subset \mathbb{C}^{k+1}.$$

It is a manifold of real dimension $2k + 1$. In particular, we obtain the *complex-valued unit circle*

$$\mathcal{S}^1 := \{z \in \mathbb{C}^1 : |z| = 1\} = \{e^{i\lambda} : \lambda \in [0, 2\pi)\}.$$

of real dimension 1. With complex multiplication, it is a one-dimensional *Lie group* (i.e., a group that is also a differentiable manifold). The complex projective space $\mathbb{C}P^k$ is the set of *one-dimensional complex linear subspaces* of $\mathbb{C}^{k+1}$. Each one-dimensional complex linear subspace is determined by a single $v \in \mathbb{C}^{k+1} \setminus \{0\}$ as

$$\{r\exp(i\phi)v : r \geq 0,\ \phi \in [0, 2\pi)\}$$

and intersects the unit sphere in an $\mathcal{S}^1$ orbit. Consequently, the complex projective space can be defined as

$$\mathbb{C}P^k = \mathcal{S}^{2k+1}/\mathcal{S}^1.$$

It is a Riemannian manifold of real dimension $2k$ and complex dimension k. In the following, we illustrate that planar shape spaces are isomorphic to complex projective spaces of corresponding dimension, see for example Huckemann and Hotz (2009).

For k planar landmarks $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots \begin{pmatrix} x_k \\ y_k \end{pmatrix} \in \mathbb{R}^2$, the similarity shape space (see Section 1.1.2) is defined as

$$\Sigma_2^k := \mathbb{S}^{(k-1)-1\times 2}/\mathrm{SO}(2),$$

where

$$\mathrm{SO}(2) := \left\{ \begin{pmatrix} \cos(\lambda) & -\sin(\lambda) \\ \sin(\lambda) & \cos(\lambda) \end{pmatrix} : \lambda \in [0, 2\pi) \right\}.$$

For all $z \in \mathbb{C}$ holds

$$\begin{pmatrix} \cos(\lambda) & -\sin(\lambda) \\ \sin(\lambda) & \cos(\lambda) \end{pmatrix} \begin{pmatrix} \Re(z) \\ \Im(z) \end{pmatrix} = \begin{pmatrix} \Re(e^{i\lambda} z) \\ \Im(e^{i\lambda} z) \end{pmatrix}$$

and consequently SO(2) acts equivalently on the planar landmarks $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots \begin{pmatrix} x_k \\ y_k \end{pmatrix} \in \mathbb{R}^2$ as $\mathcal{S}^1$ does on the corresponding complex representation $z_1 = x_1 + iy_1, \dots, z_k = x_k + iy_k \in \mathbb{C}$. Thus

$$\mathbb{C}P^{k-2} \cong \Sigma_2^k.$$

1.2 PCA for non-Euclidean data

The following serves as an introduction to Paper A and Paper B and is thus close in content to parts of the respective papers.

PCA is one of the most important exploratory methods in Euclidean multivariate statistics and is used for dimensionality reduction of data. It was developed for Euclidean data by statisticians Karl Pearson in 1901 (Pearson (1901)) and Harold Hotelling in the 1930s (Hotelling (1936)). Consider a Euclidean data matrix $X \in \mathbb{R}^{n\times p}$, where $p \in \mathbb{N}$ corresponds to the number of variables and $n \in \mathbb{N}$ corresponds to the number of objects. Usually, the transposed rows $x_1, \dots, x_n \in \mathbb{R}^p$ of X form a random sample. We search for a *few* linear combinations of the variables that summarize the data while losing as little variance as possible. For this, we determine the *empirical covariance matrix*

$$S_X := \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{n} X^T X - \bar{x}\bar{x}^T,$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean. From the spectral decomposition theorem, it follows that S_X can be written as

$$S_X = GLG^T$$

where $L = \text{diag}(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix with $\lambda_1 \geq \lambda_2, \ldots, \geq \lambda_p \geq 0$ and G is an orthogonal matrix. We define the matrix W by the *principal component transformation*

$$W = \left(X - 1_n \bar{x}^T\right) G, \quad \text{where} \quad 1_n = (1, \ldots, 1)^T \in \mathbb{R}^n.$$

The columns of W are called *principal components* and represent an uncorrelated linear combination of the variables, which means that the covariance matrix of W is diagonal

$$S_W = \frac{1}{n} W^T W = \frac{1}{n} G^T \left(X^T - \bar{x} 1_n^T\right)\left(X - 1_n \bar{x}^T\right) G = G^T S_x G = L.$$

The columns of G are called *principal directions*. In practice, dimension reduction is performed by taking only the principal components that belong to the highest variance, see e.g. Mardia et al. (1979). Extending PCA to non-Euclidean data, where data points are assumed to lie on some manifold rather than in some Euclidean space, is an active area of research and several concepts have been developed already, some of which are outlined below.

There are several concepts on how to extend PCA to data on non-Euclidean spaces embedded in Euclidean spaces, for instance *tangent space PCA* and *geodesic PCA*. Tangent space PCA attempts to exploit the locally Euclidean property of manifolds by, in principle, mapping the data to a suitable Euclidean tangent space of the underlying space and then subjecting it to Euclidean PCA. This can be accomplished using the tangent space at a suitable Fréchet mean. The next step is to project the data onto the tangent space of the respective mean. In *generalized Procrustes analysis*, the data are orthogonally projected with respect to a suitable ambient space onto the tangent space of the full Procrustes mean (see Section 1.1.2), see Gower (1975). In contrast, in *Principal Geodesic Analysis* (PGA), the *inverse Riemannian exponential map* at the Fréchet mean is used, see Fletcher and Joshi (2004). Tangent space PCA methods can be useful when the data are close to the Fréchet mean. However, note that tangent space PCA may not be the canonical choice due to its dependence on the chosen tangent space, i.e. the random base point, and that in the presence of curvature, in general, no tangent space can accurately represent the mutual distance between all data points. As exemplarily illustrated in Figure 1.3, tangent space PCA methods are limited if the Fréchet mean is far from the data.

In contrast to tangent space PCA, *Geodesic PCA* (G-PCA) for Riemannian manifolds is based on geodesics with respect to the intrinsic metric (e.g. Huckemann and Ziezold (2006); Huckemann et al. (2010)). In Euclidean space, the best fitting geodesic, i.e. the straight line with the lowest sum of square distances to the data, always runs through the mean. This is not necessarily the case in non-Euclidean spaces, therefore flexibility is gained by allowing arbitrary geodesics, in contrast to exclusively considering geodesics passing through the Fréchet mean. Once the principal geodesic is identified, the higher order principal geodesics are determined iteratively in a manner that ensures orthogonality to the previously determined principal geodesics. Alternatively, in *Horizontal*

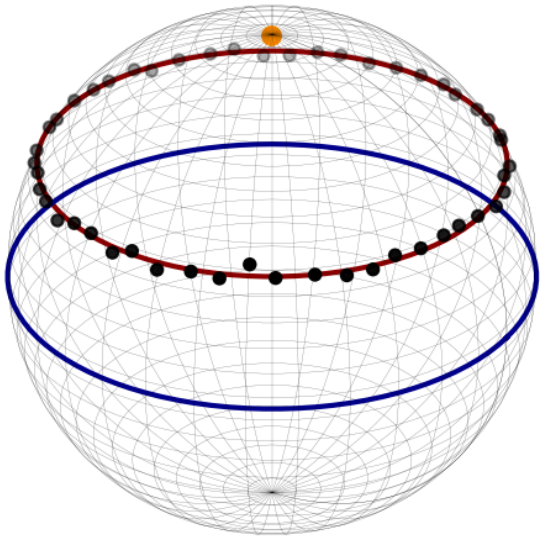


Figure 1.3: A sample of points (black) on a 2-dimensional sphere. In *Principal Geodesic Analysis* from Fletcher and Joshi (2004) only great subspheres passing through the Fréchet mean (orange) are allowed. In *Geodesic PCA* (e.g. Huckemann and Ziezold (2006); Huckemann et al. (2010)) this is relaxed to arbitrary great circles, resulting in the blue best fitting great subsphere. In *Principal Nested Spheres* (from Jung et al. (2012)), small subspheres, i.e. small circles in this 2-dimensional example, are additionally allowed, leading to the best fitting small circle (red).

Component Analysis from Sommer (2013), the higher order components are determined using parallel transport. Since one starts with a one-dimensional approximation of the data and then adds more dimensions, these methods are classified as mainly *forward methods*.

Two manifolds of particular interest for biological data are the sphere and the torus, see for example Mardia and Jupp (2000); Dryden and Mardia (2016). For spheres, a major advance was made with *Principal Nested Spheres* (PNS) analysis, introduced by Jung et al. (2012). In the forward methods introduced above, arbitrary geodesics were allowed, which do not pass through the Fréchet mean. In generalization of this concept, PNS proposed by Jung et al. (2012) is a *backward method* (see Huckemann and Eltzner (2018)), in which a sequence of nested subspheres is determined by intersecting the sphere with an affine hyperplane in its Euclidean ambient space. As a result, not only great subspheres but also small subspheres are available, leading to increased flexibility, depicted in Figure 1.3. For data on an n-dimensional Euclidean space, the family of first principal components (i.e. of straight lines) has dimension $2(n-1)$ (a line is defined by 2 points and both points can vary on the line) while for data that are on an n-dimensional sphere, the family of main principal nested components (i.e. of small circles) has dimension $3(n-1)$ (a circle is defined by 3 points and all three points can vary on the circle), see Huckemann and Eltzner (2018). Consequently, PNS is particularly advantageous for dimension reduction based clustering methods, as it allows in general the separation of three clusters by using only the main nested principal circle, for which two Euclidean principal components would be required for separation, as demonstrated in Figure 1.4. A more general approach is *barycentric subspace analysis on manifolds* by Pennec (2018). For $k+1$ given points, the barycentric subspace for a Riemannian manifold is defined by all Fréchet means

obtained by assigning different weights to each of these points, giving in principle a d-dimensional space. Consequently, adding or removing points can create nested sequences of subspaces, allowing the method to be used as a forward or backward technique.

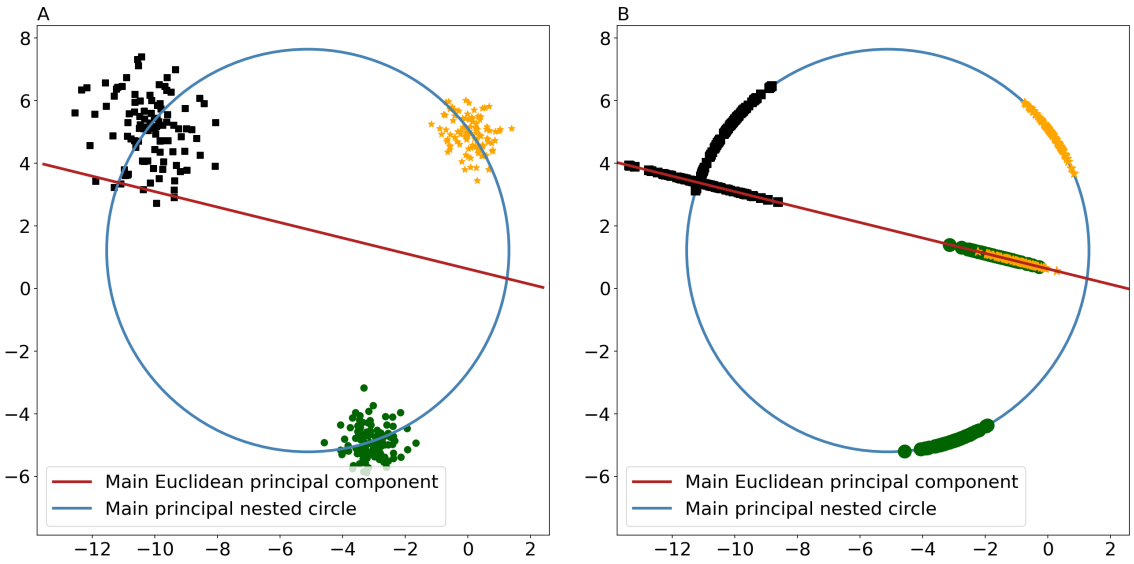


Figure 1.4: Left: a sample of points in a 2-dimensional plane which comprises three clusters (black, yellow, green), as well as the main principal nested circle of the sample (blue) and the main Euclidean principal component (red). Right: the projections of the points onto the main principal nested circle and the main Euclidean principal component, respectively.

For the case of the torus, which is the product space of two or more spheres (see Section 1.1.1), the situation is more complicated, as neither tangent space PCA nor geodesic PCA are well applicable, since in the former the periodicity of the torus is not taken into account and in the latter all irrational slope geodesics approximate any data arbitrarily well, which is a dead end for meaningful statistics. Kent and Mardia (2009) proposed an approach based on wrapped normals to avoid the problem of winding geodesics and in Kent and Mardia (2015), an approach is discussed in which the *winding* number of geodesics is restricted. In Sargsyan et al. (2012), an approach is presented that maps from the torus to the sphere and then uses the G-PCA from Huckemann and Ziezold (2006) for dimension reduction. A major advance is *Torus PCA* (T-PCA) of Eltzner et al. (2018), which inherits the positive properties of PNS. It uses a data driven function to map the data from a torus to a *stratified sphere* (i.e. a sphere with self-gluing to account for the periodicity of the torus) and then reduces the dimension using PNS extended to stratified spheres. Recently, *scaled torus principal component analysis* was developed by Zouboulоglou et al. (2022). Here, the data are mapped from the torus to the sphere using spherical multidimensional scaling and then passed to PNS to find the sequence of best fitting subspheres.

1.3 Adaptive iterative clustering for metric data

The following is used to introduce the clustering algorithm from Paper A, which is also used in Paper B and is thus similar in content to parts of the respective papers.

Clustering is an umbrella term for a broad class of unsupervised data segmentation learning methods widely used in many fields, most prominently pattern recognition, machine learning and it has particularly many applications in biology. Clustering aims to group n distinct data points $X^{(1)}, \dots, X^{(n)}$ that are assumed to be *heterogeneous* into m distinct *homogeneous* groups, where m is usually unknown. Homogeneous in this context means that the individual members of each group are close to each other, but further away from the members of the other groups, see Mardia et al. (1979). This rather broad definition of clustering leads to numerous different clustering algorithms, see Estivill-Castro (2002) for a detailed review on this topic. Some prominent classes of clustering algorithms are *centroid* models, such as *k-means clustering*, see Lloyd (1982), *density models*, such as *density-based spatial clustering of applications with noise* (DBSCAN), see Ester et al. (1996), *distribution models*, such as *Gaussian mixture* models based on the *expectation-maximization* algorithm, see Dempster et al. (1977), or *hierarchical clustering*. All different clustering algorithms have different weaknesses and strengths and are useful for different situations.

We use and refine two particular hierarchical clustering methods for metric spaces, namely *average linkage* clustering, also known as *unweighted pair group method with arithmetic mean*, first developed by Sokal and Michener (1958) and *single linkage* clustering, also known as nearest neighbor clustering, developed by Florek et al. (1951). For points $X^{(1)}, \dots, X^{(n)}$ in an arbitrary metric space with distance d, a distance matrix $D = (d_{i,j})_{i,j=1}^{n}$, where $d_{i,j} \coloneqq d(X^{(i)}, X^{(j)})$ for all $i, j \in \{1, \dots n\}$ is used instead of the points themselves. A rooted tree is created by first making each data point its own cluster. Iteratively, the clusters with the smallest distance are merged to form a new cluster. The distance between two clusters A and B is defined in average linkage clustering, respectively single linkage clustering by

$$d_a(A, B) \coloneqq \frac{1}{|A| \cdot |B|} \cdot \sum_{X \in A} \sum_{Y \in B} d(X, Y), \qquad d_s(A, B) \coloneqq \min_{X \in A, Y \in B} d(X, Y),$$

where $|A|$ is the cardinality of the set A. The cluster tree is extended by a parent node above the merged clusters, which is tagged with the distance between the two merged clusters. We refer to the top node of the tree, which represents the set of all data points, as the *root* of the cluster tree. For both methods, the node values increase: if the clusters A and B are merged into the cluster $A \cup B$ and C is another cluster, then the distance between A and B is smaller than between A and C or B and C, respectively, and therefore

$$\begin{aligned} d_a(A \cup B, C) &= \frac{1}{|A \cup B| \cdot |C|} \left(|A| \cdot |C| \cdot d_a(A, C) + |B| \cdot |C| \cdot d_a(B, C)\right) \;\geq\; d_a(A, B)\,, \\ d_s(A \cup B, C) &= \min_{X \in A \cup B, Y \in C} d(X, Y) = \min\left\{d_s(A, C), d_s(B, C)\right\} \;\geq\; d_s(A, B). \end{aligned}$$

Consequently, any choice of a distance value $r > 0$ leads to a clustering in which all elements are in the same cluster that are in the cluster tree below a joint node carrying a distance of at most r, see Mardia et al. (1979).

Single linkage clustering tends to return long elongated clusters, an effect called *chaining*. Two different clusters can be clustered together because a chain of outliers links them. In addition, both methods fail when two closely neighboring clusters have a higher density than a third cluster, see the example in Figure 1.5. To tackle such problems, Langfelder et al. (2007) and Obulkasim et al. (2015)

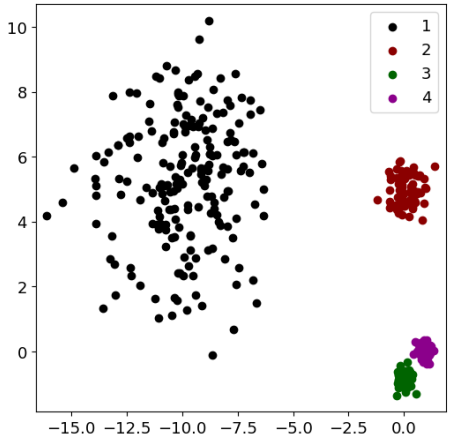


Figure 1.5: Toy data set featuring four clusters, all of which cannot be separated by single or average linkage clustering as the first features a large spread, the other three are very dense, two of them nearby. The Figure is adapted from Paper A.

developed data-adaptive cutting procedures. However, the question remains how to detect with statistical guarantees whether a found cluster can be further decomposed into several sub-clusters. For this reason, a clustering was developed in Paper A, which consists of an iterative pre-clustering (which allows clusters of different densities) and a subsequent post-clustering based on PNS (introduced in Section 1.2). In Paper A, the clustering method is demonstrated on both Euclidean and non-Euclidean data. In Paper B this clustering method is a crucial part of a classification based learning algorithm.

1.4 Strong consistency for generalized Fréchet means

The following section serves as an introduction to the research area of strong consistency theorems for generalized Fréchet means, to which we contribute in Paper C. The following content is a condensed version of Section 3 of Paper C.

For real-valued random vectors, the Fréchet mean is unique if it exists. Uniqueness is not necessarily given in non-Euclidean spaces, as demonstrated in the examples in Section 1. In consequence, a general formulation in terms of set-valued Fréchet means is required for data on non-Euclidean spaces. Two versions of the set-valued strong consistency were introduced in Ziezold (1977);

Bhattacharya and Patrangenaru (2003) and shown under rather broad conditions. In the following we introduce both versions of consistency, show how they have been generalized and explain how we extended them in Paper C.

Consider i.i.d. random elements $X_1, X_2, \cdots \sim X$ mapping from a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ into the *data space* $\mathfrak{Q}$, a topological space equipped with the Borel σ-algebra. Also, define the *parameter space* $(\mathfrak{P}, d)$, as a *separable* (i.e., there exists a dense countable subset) metric space with metric function $d : \mathfrak{P} \times \mathfrak{P} \mapsto [0, \infty)$ and the topology induced by the metric.

Definition 1.1 (Generalized sample and population Fréchet mean)**.** *For $\rho : \mathfrak{Q} \times \mathfrak{P} \mapsto \mathbb{R}$, which is continuous in $\mathfrak{P}$ for all fixed $q \in \mathfrak{Q}$ and measurable in $\mathfrak{Q}$ for all fixed $p \in \mathfrak{P}$, we define*

$$E_n^{(\rho)}(\omega) := \arg\min_{p \in \mathfrak{P}} \sum_{i=1}^{n} \rho(X_i(\omega), p), \qquad E^{(\rho)} := \arg\min_{p \in \mathfrak{P}} \mathbb{E}\left(\rho(X, p)\right).$$

$E^{(\rho)}$ and $E_n^{(\rho)}(\omega)$ are called the sets of generalized sample and population Fréchet means*, respectively.*

The sample Fréchet mean sets are closed random sets (studied by Choquet (1954); Kendall (1974); Matheron (1974), among others) since ρ is continuous. In the following, we introduce two different definitions of strong consistency, which are commonly named after the authors who proposed them.

Definition 1.2 (Two versions of set strong consistency)**.** *We say that the estimator $E_n^{(\rho)}(\omega)$ for $E^{(\rho)}$ is*

ZC: Ziezold strongly consistent *if*

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \subseteq E^{(\rho)} \text{ for all } \omega \in \Omega \text{ almost surely},$$

BPC: Bhattacharya and Patrangenaru strongly consistent *if $E^{(\rho)} \neq \emptyset$ and if for every $\epsilon > 0$ and almost surely for all $\omega \in \Omega$ there is a number $n = n(\epsilon, \omega) > 0$ such that*

$$\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega) \subseteq \{p \in \mathfrak{P} : d(E^{(\rho)}, p) \leq \epsilon\}.$$

Ziezold (1977) proved strong consistency in the sense of (ZC) for separable spaces $\mathfrak{P} = \mathfrak{Q}$ and $\rho = d^2$ where d is a *quasi-metric* (i.e. it holds $d(x, y) \geq 0$, $d(x, x) = 0$, $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$). For metric spaces $(\mathfrak{Q}, d)$, with the *Heine Borel property* (all closed and bounded sets are compact, see Williamson and Janos (1987)), (BPC) was proven by Bhattacharya and Patrangenaru (2003) for $\mathfrak{Q} = \mathfrak{P}$ and $\rho = d^2$. In Huckemann (2011b), (ZC) and (BPC) were extended for generalized Fréchet means. To show (ZC), a continuity condition in the second argument, uniformly over the first argument, was required for a non-negative ρ, and for (BPC), the Heine-Borel property, a nonempty $E^{(\rho)}$, and a coercivity property in the second argument

were additionally required. Evans and Jaffe (2020) generalized strong consistency for so-called *Fréchet p-means*, a special case of Fréchet means, where $\mathfrak{P} = \mathfrak{Q}$ and $\rho = d^p$. In Schötz (2022), the continuity assumptions for ρ required in Huckemann (2011b) were relaxed to lower semi-continuous assumption and the non-negative assumption on ρ was relaxed to $\mathbb{E}(inf_{q\in Q}\rho(Y,q)) > -\infty$. However, to the best of our knowledge, the available strong consistency results for generalized Fréchet means (Schötz (2022); Huckemann (2011b)) cannot cover the simple case of maximum likelihood parameters of a univariate or multivariate Gaussian, because the log-likelihood is unbounded from below.

In Paper C we show (ZC) and (BPC) under even weaker assumptions, namely a modulus of continuity along with its prefactor for (ZC) and using non-emptiness of E^ρ and a weakened coercivity assumption for (BPC). Unlike the cases treated in the above articles, the ρ function does not have to be bounded from below and therefore we are able to show that our framework encompasses consistency in terms of (ZC) and (BPC) for joint estimation of μ and σ in the Gaussian MLE setting. Such a generalization is typically necessary to cover cases where a generalized Fréchet mean is estimated along with a (co-)variance-like quantity. This is for example the case for diffusion means with simultaneously estimated variance, see Eltzner et al. (2022). Moreover, in Paper C we apply the developed theory to prove strong consistency for the homoscedastic drift model (see Section 2.2.1).

CHAPTER 2

Contributions to the research publications

This section aims to briefly summarize the main results of each paper and to highlight my contribution to each paper. For this purpose, we first give an overview of different methods to determine the structure of biomolecules. Then, in Section 2.1, a more detailed introduction to the structure of RNA molecules is given and the contents of papers A and B are summarized. Subsequently, in Section 2.2, an introduction to ENDOR spectroscopy is given and the contents of Paper C and Paper D are presented.

One of the main objectives of structural biology is to understand the complicated three-dimensional structure of biomolecules, and thus provide meaningful links between structure and functionality. In particular, this information can be used in the field of structure-based drug design, see for example Schlick and Pyle (2017) or Anderson (2003a). The first breakthroughs in this field were made in the early 1990s, and thanks to the evolution of new experimental methods and the availability of increasingly effective computational clusters, it has become an essential part of drug design. There is a wide range of different methods to determine the structure, some of which are listed in the following.

A popular method is *X-ray crystallography* (X-ray), in which, using a suitable substrate, molecules are crystallized and subjected to X-ray imaging. A further emerging method is *cryogenic electron microscopy* (cryo-EM), which uses electron microscopes to study the structure of biomolecules in a frozen state. Although the resolution of cryo-EM has improved greatly in recent years due to technical advancement, the resolution usually does not approach the atomic resolution of X-ray. Therefore, both methods are often used complementary, due to the difficulty of crystallizing flexible and large molecules, cryo-EM is used to understand the overall shape and X-ray is used to better understand the atomic structure of individual smaller subcomponents, see for example Wang and Wang (2017). This naturally leads to data at different resolutions and to the question of how to model data at different scales and develop learning algorithms.

Another approach to obtain structural information of molecules is the use of spectroscopic methods. There are a variety of different spectroscopic methods. Possibly the best known is *nuclear magnetic resonance* (NMR) spectroscopy, which studies the interactions between the atoms of a molecule

using radio frequency (RF) pulses. On the other hand, *electron paramagnetic resonance* (EPR), utilizes microwave (MW) pulses to study the local environment and different kind of interactions of the spins of unpaired electrons. Since EPR targets only the tiny minority of unpaired electrons among the large number of electrons in a biomolecule, it can be more selective than NMR. A large part of this thesis work is concerned with the development of methods for ENDOR spectroscopy in a collaboration with the 'Electron Paramagnetic Resonance' research group of Marina Bennati at the Max Planck Institute for Multidisciplinary Sciences. In a two-step experiment, ENDOR spectroscopy combines the advantages of EPR spectroscopy and NMR spectroscopy by applying both techniques to the same sample (see Feher (1956); Gemperle and Schweiger (1991)). Using an unpaired electron that interacts with both an external magnetic field and the atomic magnetic nuclei in its vicinity, the ENDOR experiment first irradiates the sample with a *microwave* (MW) to target the unpaired electron. In the second step, the sample is irradiated with a *radio frequency* (RF) that matches the resonance condition of a specific nucleus. Roughly speaking, the double resonance approach restricts the unpaired electron to interact with specific magnetic nuclei of a chosen kind individually at different frequencies, leading to distinct peaks in a RF spectrum. Artificially inserting *labels*, i.e., magnetic nuclei that do not occur naturally in biomolecules (for example *fluorine labels*), allows determination of certain features, such as distances between interacting spins, see Meyer et al. (2020). Once the experiment is completed in the laboratory, the following steps are performed to obtain structural information from the data matrix recorded by the spectrometer during the experiment, illustrated in Figure 1.2. The first step is to denoise the spectrum, for which we develop in Paper C asymptotic theory for a drift model developed by Pokern et al. (2021) and develop a new drift model for 94 GHz data (illustrated in the denoising step in Figure 1.2). In a second step, an analysis of the spectra was developed in Paper D to determine parameters describing the conformation of the biomolecules from the spectra (exemplified in the optimize step in Figure 1.2).

2.1 Learning torus PCA-based classification for multiscale RNA correction

The following is a compilation of the biological foundations and the multiscale approach in Paper B. The biological foundations are summarized from Watson et al. (2004); Murray et al. (2003).

Ribonucleic acid (RNA) strands are formed of repeating elements that are called *nucleotides*. A nucleotide consists of three components: a sugar ring with 5 carbon atoms called *ribose*, a *phosphate group* bonded to the ribose ring at the O5' atom and one of four *nucleobases* attached to the ribose by a bond between the C1' atom of the ribose sugar ring and a nitrogen atom (which is called N1 or N9 depending on the corresponding nucleobase). RNA chains are formed by single nucleotides linking to the next phosphate group through their O3' atoms, see Figure 2.1. To derive the molecular structures, usually X-ray and cryo-EM methods are used. In the resulting data, physically and chemically impossible molecular configurations frequently occur, so-called clashes, in which two atoms are reconstructed closer to each other than is chemically possible. Clashes between two backbone atoms are the most relevant and most difficult to correct; for a detailed discussion, see

Murray et al. (2003). Most clashes of the RNA backbone occur within suites, which is the part from one sugar ring to the next, see e.g. Murray et al. (2003) and Figure 2.1. In Paper A and

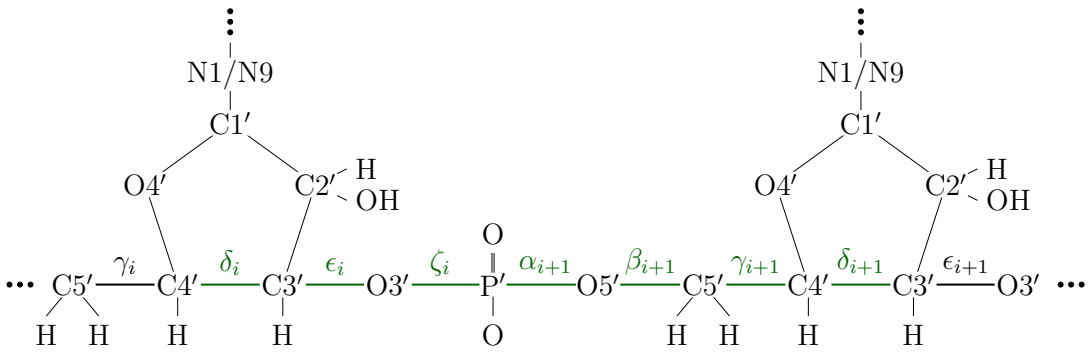


Figure 2.1: 2D scheme of backbone suite number i with 7 dihedral angles $\delta_i, \epsilon_i, \zeta_i, \alpha_{i+1}, \beta_{i+1}, \gamma_{i+1}, \delta_{i+1}$ describing the suite's 3D structure, adapted from Paper A.

Paper B we work with two different scales: at microscopic scale we work with suites, which can be represented by a tuple of 7 dihedral angles (each dihedral angle in the RNA backbone chain, defined by four consecutive atom positions, see Figure 2.2), giving a data point on the seven dimensional torus $\mathbb{T}^7$ for each suite, see Section 1.1.1. Description of suites with dihedral angles is possible because the bond lengths (distances between two consecutive atoms) and the bond angles (angles between three consecutive atoms) are approximately constant due to chemical laws. The shape is therefore determined exclusively by dihedral angles. At *mesoscopic* scale, we model the mesoscopic

angle	atom bonds
α	O3' - P' - O5' - C5'
β	P' - O5' - C5' - C4'
γ	O5' - C5' - C4' - C3'
δ	C5' - C4' - C3' - O3'
ϵ	C4' - C3' - O3' - P'
ζ	C3' - O3' - P' - O5'

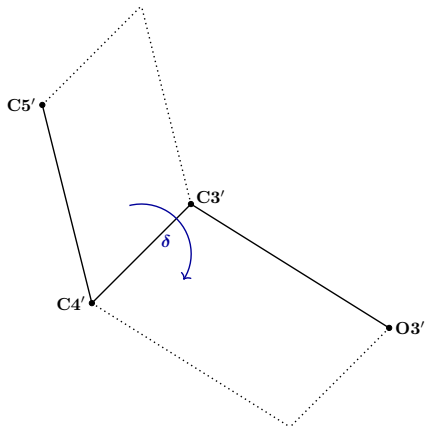


Figure 2.2: Left: names (first column) of dihedral angles along the two central atoms of the four atoms involved (second column), see Figure 2.1. Right: the dihedral angle δ of the bond between the atoms C4' and C3' is the directed angle between the plane spanned by the atoms C5', C4', C3' and the plane spanned by C4', C3', O3'. More precisely, it is the angle determined by turning the vector normal to the plane spanned by C3', C4', C5' to the vector normal to the plane spanned by O3', C3', C4' (with fixed orientation of normals determined by the order of spanning points). Taken from Paper B.

shape belonging to a suite as follows: the *mesoscopic strand* corresponds to the configuration matrix obtained by the centers of the sugar rings (*pseudo landmarks*) belonging to the 2 suites before and after the central suite, see Figure 2.3. The size-and-shape of the mesoscopic strands

is not completely defined by the dihedral angles of four consecutive sugar rings alone, as the distances between two consecutive sugar rings and the angles between three consecutive sugar rings also vary, leading to data in the size-and-shape space $S\Sigma_3^6$, see Section 1.1.2. To understand the interdependence of the different scales and to classify the suites, the following clustering was introduced in Paper A. Based on this, in Paper B we developed our two scale correction algorithm.

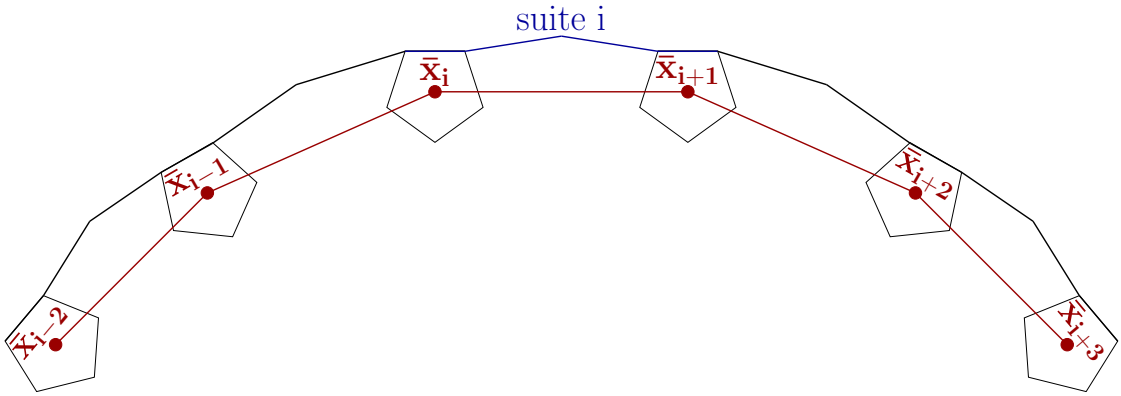


Figure 2.3: The mesoscopic shape (red lines) centered at the i-th suite is determined by the six centers of the sugar rings $\bar{\mathbf{x}}_{\mathbf{i}-2}, \ldots, \bar{\mathbf{x}}_{\mathbf{i}+3}$. Their connecting backbones (blue and black lines) comprise 5 suites, two before and two after suite i. Taken from Paper B.

2.1.1 Paper A: Principal component analysis and clustering on manifolds

The paper *Principal component analysis and clustering on manifolds* is included in Section A and is published in the *Journal of Multivariate Analysis*, see https://doi.org/10.1016/j.jmva.2021.104862. The paper is a collaboration with Kanti V. Mardia, Benjamin Eltzner, and Stephan F. Huckemann.

In the paper, we developed a clustering called *Mode huntIng on the main principle Nested small Circle of prE-clusters* (MINCE) post *Adaptive linkaGe clustEring* (AGE) for manifold data, which consists of three main steps, see Section 4 of Paper A. The first step is to cluster with the AGE pre-clustering to obtain a list of pre clusters, see Algorithm 1 in Paper A. It is a hierarchical clustering method illustrated in Figure 2.4 and Figure 2.5. First, a rooted tree is created as described in Section 1.3. Then, roughly speaking, branches with decreasing cluster size are iteratively removed from the cluster tree (see Figure 2.5). This procedure allows the separation of clusters with different densities (see right panel of Figure 2.4).

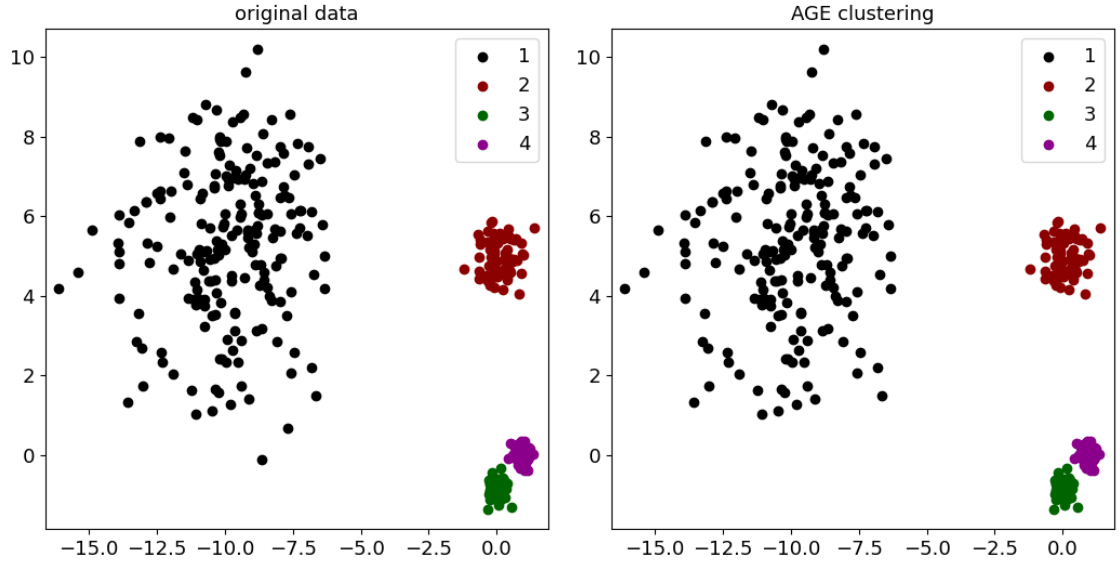


Figure 2.4: Left: original data set featuring four clusters, all of which cannot be separated by single or average linkage clustering as the first features a large spread, the other three are very dense, two of them nearby. Right: all of the clusters are retrieved by AGE except for two outliers from Cluster 1 (not shown in the right panel). The figure is taken from Paper A and the left panel is taken from Figure 1.5.

The second step is to map each pre-cluster in a suitable way to a sphere or a stratified sphere. In the third step, PNS is used to determine the *main principal small circle* (i.e. the last (one-dimensional) element in a sequence of nested spheres in the PNS, introduced in Section 1.2) for each pre-cluster. Then, the data from the pre-cluster are projected onto the main principal small circle and *circular mode hunting* (adapted from the non-circular linear case Dümbgen and Walther (2008)) is applied to assign each mode found with statistical significance a cluster. MINCE post AGE gains strength from the statistically advantageous geometry of spheres. In principle, with the main principal small circle three clusters can be separated reliably, while with the principal component of Euclidean

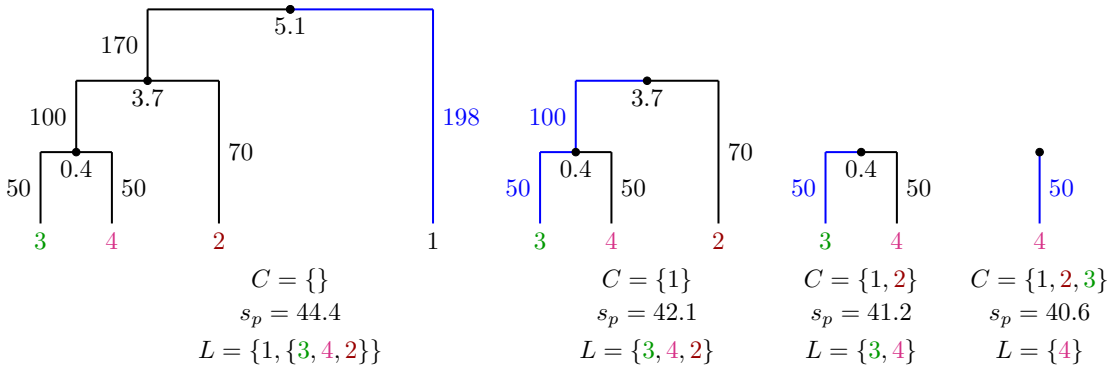


Figure 2.5: Illustrating four iterations of the AGE Algorithm applied to the data of Fig. 2.4 (with cluster colors taken from there) with node labels indicating distance of clusters joined and branch labels indicating the number of data points represented. For each iterate the initial cluster list C, the current minimal cluster size s_P and the resulting auxiliary list L (before Step 8) are shown below. Further details are given in Algorithm A6 of Paper A. The cluster that is removed from the cluster tree in each iteration is marked in blue. The figure is taken from Paper A.

PCA only two can be separated reliably, see Figure 1.4. We apply our clustering on three different datasets: data on a torus, data on a shape space, and Euclidean "worms" data. The first two examples are SARS-CoV-2 RNA backbone relevant data from the Protein Data Bank (PDB) (compiled from the Coronavirus Structural Task Force (CSTF)). In the third example, we obtain our benchmark data from the simulation software *worms*, see Sieranoja and Fränti (2019). Exemplarily, Figure 2.6 plots the cluster result for the suites (each suite is represented by 7 dihedral angles) from the SARS-CoV-2 RNA backbone data set, using a scatter plot relating pairs of dihedral angles with one another.

Structure of the paper:

- Section 1: Introduction.
- Section 2: Notation and terminology are introduced.
- Section 3: The torus PCA from Eltzner et al. (2018) is extensively introduced and explained. Moreover, asymptotic consistency and asymptotic normality are shown for the principal nested spheres from the Torus PCA using the framework of backward nested subspaces, which are an extension of generalized Fréchet means.
- Section 4: The new clustering method MINCE post AGE is introduced, and applied to three different data sets: data on a torus, data on a shape space, and Euclidean "worms" data.
- Section 5: A model based approach for PCA on the torus, based on wrapped multivariate normal distributions (see Kent and Mardia (2009)) is briefly presented.

Own contribution: In the collaboration, I mainly worked (guided by the co-authors) on the contents of Section 4, which is concerned with the following challenges: we developed a clustering method specifically for processing biological data from the PDB. The data are usually not Euclidean, but are

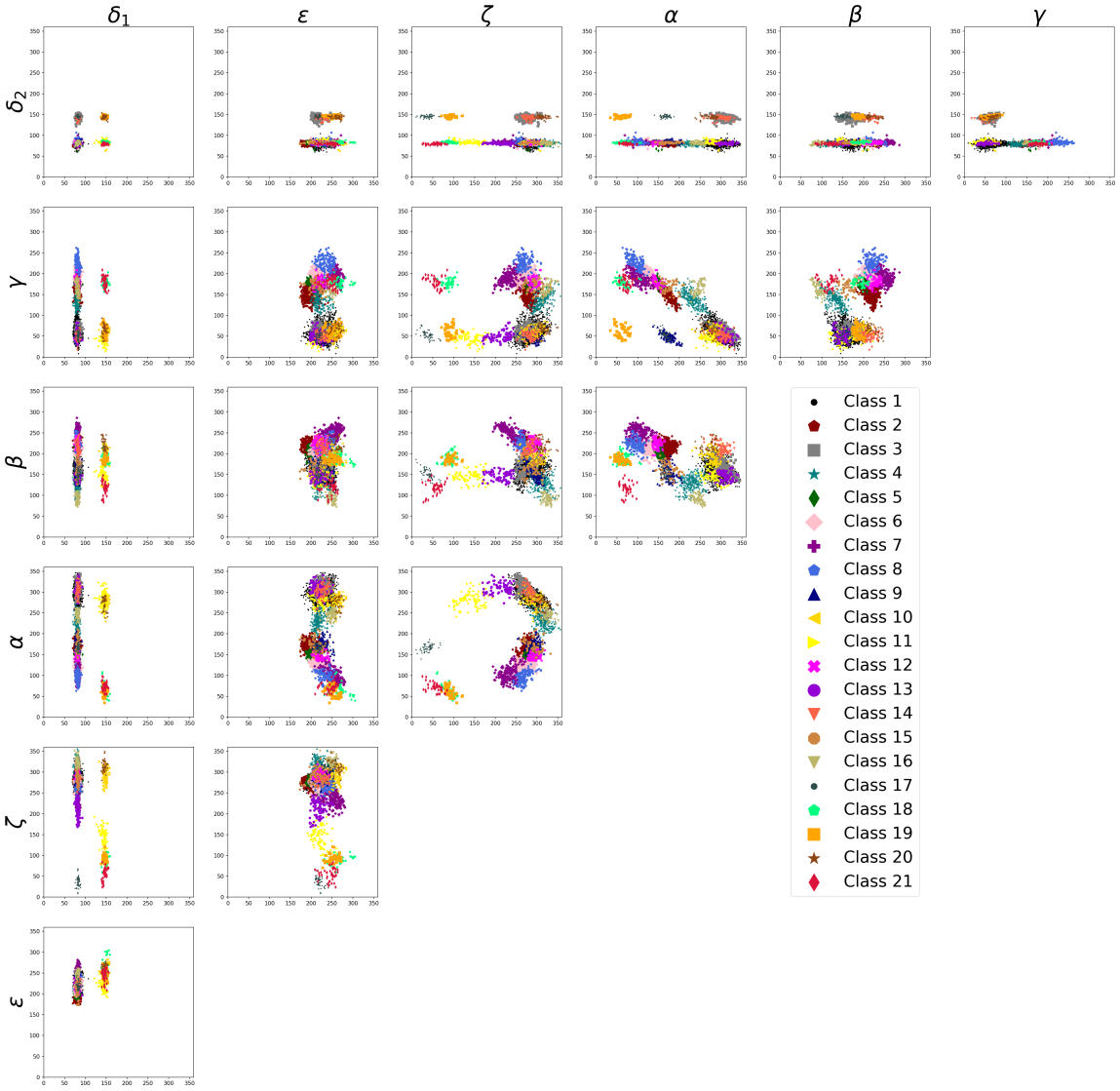


Figure 2.6: Scatterplots of the clusters found by the cluster algorithm MINCE post AGE. The original data can be found in Figure A8 of Paper A. Table A1 of Paper A gives the cluster sizes of each cluster. The figure is taken from Paper A.

usually on manifolds, such as the torus, sphere, shape space, or size-and-shape space. In addition, there are a lot of outliers, clusters with widely varying densities and numbers of elements. Further challenges included the compilation of the data, the data analysis and the implementation of the cluster algorithm for the different data sets. I presented the contents of the paper in the Pizer (Shape Stats Discussion Group) seminar in two talks in 2021 and at the JMVA Jubilee Edition conference in 2022.

2.1.2 Paper B: Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2

The paper *Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2* is included in Section B and can be found in the *Journal of the Royal Statistical Society Series C: Applied Statistics*, see https://doi.org/10.1093/jrsssc/qlad004. The paper is a collaboration with Kanti V. Mardia, Benjamin Eltzner, and Stephan F. Huckemann and has the following content.

We developed a fast, data-driven reconstruction algorithm called CLEAN MINT-AGE (Classification based on muLtiscale structurE enhAncemeNt using Mode huntINg after Torus pca on Adaptive cutting averaGe linkage trEes). Usually, methods based on molecular dynamics are used to correct clashes, see for example Chou et al. (2013). However, due to the large variability of RNA shapes, these simulations are very computationally intensive and often molecules corrected in this way may still contain clashes, see Richardson et al. (2018). In contrast, CLEAN MINT-AGE is based on a two-scale shape analysis: at the microscopic scale we work with suites (see Figure 2.1) which can be represented on the seven-dimensional torus $\mathbb{T}^7$ (see Section 1.1.1). At the *mesoscopic* scale, we model the mesoscopic shapes (see Figure 2.3) in the size-and-shape space $S\Sigma_3^6$, see Section 1.1.2.

On a benchmark data set, we empirically investigate the interdependence between the mesoscopic and microscopic scales, which results in the conclusion that for clash-free data, concentrated clusters on the mesoscopic scale belong to concentrated clusters on the microscopic scale (see left two panels of Figure 2.7). This is the justification of our CLEAN-MINT-AGE correction algorithm: to

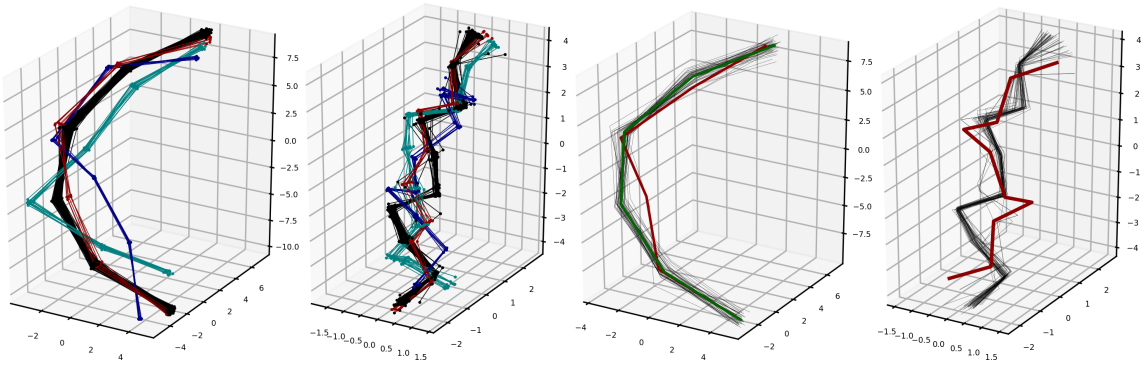


Figure 2.7: Left: four exemplary mesoscopic clusters at mesoscopic scale. Center left: their central suites at microscopic scale. Right two panels: a clash suite (red) (from benchmark file 1f8v, Tang et al. (2001), see Table B in the supplement of Paper B) with its clash free neighbors (black) and proposed clash free correction (green) at mesoscopic scale (center right) and microscopic scale (right). The figure is composed of different figures from Paper B.

correct clashes at the microscopic scale, classes of clash free suites are learned using the MINCE post AGE clustering from Paper A (adapted to torus data). The second step is to determine the class that will be used for structure correction. For this, we determine a set of clash-free mesoscopic shapes that are closest to the mesoscopic shape corresponding to a clash suite (see center right panel of Figure 2.7). As a third step, we consider the suites corresponding to this set of closest

mesoscopic shapes and determine the class that dominates this set (see right panel of Figure 2.7). At the microscopic scale, the correction is determined by the torus mean (introduced in Section 1.1.1) of the corresponding class. At mesoscopic scale, our correction is defined by a novel orthogonal projection (with respect to the partial Procrustes distance defined in 1.1.2) of the partial Procrustes mean of the closest mesoscopic shapes from the dominant class onto a subset of the size-and-shape space defined by special distance constraints inherited from mesoscopic and microscopic scale. The projection is in general unique and can be computed explicitly (see Theorem 2.3 in Paper B). The corrections are designed to be clash free and, on the benchmark data set, the proposed corrections are in general well below the order of resolution of the respective measurement.

To show the potential of our method, we apply our method to two suites (one of them is shown in Figure 2.8 center and right panel) of the *frameshift stimulation element* (which enables decoding of more than one protein from a single RNA molecule) of SARS-CoV-2 (depicted in left panel of Figure 2.8). It is a promising target for drug development and to our knowledge, no consistent 3D structure has been proposed to date. Based on the cryo-EM map, ten different three-dimensional structure models are proposed in Zhang et al. (2021). The central and right panel of Figure 2.8 illustrates the correction suite 28/29 (marked with a red arrow in the left panel). In all 10 different proposed models of Zhang et al. (2021) the suite is a clash suite. In contrast, CLEAN MINTAGE consistently classifies the clash suites for each of the 10 models into the same class, which is clash-free by design.

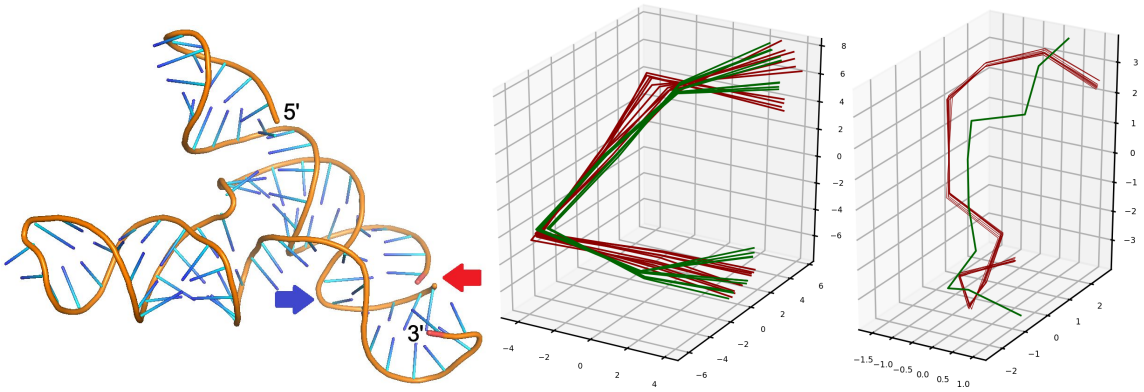


Figure 2.8: Left: one out of 10 proposed 3D RNA structures of the SARS-CoV-2 frameshift stimulation element by Zhang et al. (2021), graphically reproduced with PyMOL (Schrödinger, LLC, 2015) with backbone (orange) and nucleobases (blue), yielding helical structures whenever the latter point to each other. Center and right: ten proposed reconstructions (red) by Zhang et al. (2021), which are all clashing, for Suite 28/29 (marked with a red arrow in the left panel) connecting two helical segments in the frameshift stimulation element of SARS-CoV-2 and our ten clash free corrections (green) at mesoscopic scale (center) and at microscopic scale (right). The figure is composed of different figures from Paper B.

Structure of the paper:

- Section 1: Introduction.

- Section 2: Tools from shape analysis are introduced: the torus, the size-and-shape space, the Fréchet mean for both scales and the novel orthogonal projection onto a constrained-size-and-shape space.
- Section 3: Biological concepts and tools that are required to understand our data analysis and correction algorithm are discussed in detail, namely RNA backbone geometry, clash detection and the concept of multiscale modeling.
- Section 4: A detailed data analysis on a benchmark data set:
  - First we learn the classes using the clustering method presented in Paper A.
  - We statistically investigate the interdependence between the mesoscopic and microscopic scales.
  - We introduce the CLEAN-MINT-AGE correction algorithm, which is subsequently applied to the benchmark data set.
- Section 5: We apply our method to two suites of the frameshift stimulation element of SARS-CoV-2.
- Section 6: Discussion.

Own contribution: For the paper, I worked on all sections. This includes (guided by the co-authors) the development of the statistical methods, statistical analysis and modeling and working with various biological programs, such as PHENIX (see Liebschner et al. (2019)) and ERRASER (see Chou et al. (2013)). In addition, I presented the contents of the paper at three conferences: GSI in Paris 2021, GPSD Mannheim 2021 and ADISTA in Santiago de Compostela 2022.

2.2 The ENDOR experiment and Drift Models on Complex Projective Space

This section is intended to provide an introductory exposition of ENDOR spectroscopy necessary for the contents of Paper C and Paper D and is close in content to the respective papers. A detailed introduction to the physical theory of ENDOR can be found, for example, in Gemperle and Schweiger (1991).

During the ENDOR experiment, a chemical sample (for example the model fluorine-nitroxide compound depicted in the right panel of Figure 1.2) is located in an external temporally constant and spatially homogeneous magnetic field with magnetic field strength B_0. In the first part of the experiment, the sample is irradiated with a sequence of microwave (MW) pulses. The combination of the magnetic field strength B_0 and the microwave frequency ν_{MW} selects a set of orientations of the molecule relative to the external magnetic field that lead to a certain resonance condition being satisfied. Those molecules in the chemical sample whose orientation is in this set take part in the resonance experiment. Usually, the experiment is performed at five orientations, denoted as g_x, g_{xy}, g_y, g_{yz} and g_z. A receiver records the microwave echo signal emitted by the molecules for which the resonance conditions are satisfied in two different components. In a technique called *quadrature detection*, a reference microwave signal is used to obtain a measurement in two different components: a first component, which forms the real part and is in phase with the reference microwave signal and a second component, called the *quadrature component*, whose phase is shifted by 90 degrees, which forms the imaginary part. For an orientation, that is, a fixed magnetic field strength and a fixed MW frequency, the sample is irradiated with each of the RF frequencies from $\{f_\nu : \nu \in 1, \dots, N\}$ in a pseudo-random order, each of which affects the echo signal. The set of measurements on all frequencies is called a *scan*. The sum of the echo signals of several (typically 50) scans is stored by the spectrometer as a single *batch* in $\mathbb{C}^N$ and B different batches resulting in a data matrix $Y \in \mathbb{C}^{B\times N}$. Two different pulse sequences are commonly used (see Mims (1965) and Davies (1974)) in order to synchronize a large number of molecules participating in the resonance experiment so that combined microscopic signals can be measured macroscopically. Different pulse sequences result in different characteristics of the experimental noise.

The common practice for extracting ENDOR spectra from a data matrix Y prior to the work in our CRC project consisted in applying the *averaging model*. In the averaging model, the batches are first averaged $Z_\nu = \frac{1}{B}\sum_{b=1}^{B} Y_{b,\nu}$, then phase correction is performed to obtain a real-valued spectrum $\tilde{I} = \Re(e^{i\lambda}Z)$ where $\lambda \in [0, 2\pi)$, is determined for example with the *maximum method*, see (2.5). Subsequently, a normalization is performed to obtain the spectrum I. However, since ENDOR experiments often run for hours and at low temperatures, significant thermal drifts occur in practice. To model these, the *homoscedastic drift model* for spectrometers using a microwave frequency of $\nu_{MW} = 263\,\text{GHz}$ was developed in Pokern et al. (2021). Here, the data matrix is decomposed as follows

$$Y_{b,\nu} = \psi_b + \phi_b\kappa_\nu + \epsilon_{b,\nu}. \tag{2.1}$$

One can interpret $\psi \in \mathbb{C}^B$ as the EPR signal as well as a possible offset of the spectrometer, $\phi \in \mathbb{C}^B$ as magnitude and phase of the ENDOR effect and $\kappa \in \mathbb{C}^N$ as the ENDOR effect consisting of the ENDOR spectrum I and an orthogonal component ω that contains a resonance artifact of the spectrometer. The experimental noise is represented by $\epsilon \in \mathbb{C}^{B\times N}$, with the property $\operatorname{vec}(\epsilon_{b,\nu}) := \begin{pmatrix} \Re(\epsilon_{b,\nu}) \\ \Im(\epsilon_{b,\nu}) \end{pmatrix} \sim \mathcal{N}(0, \Sigma)$ where the covariance matrix $\Sigma \in \mathrm{SPD}(2)$ is positive definite symmetric. The following conditions are introduced for the identifiability of κ

$$\sum_{\nu} \kappa_\nu \overset{!}{=} 0 \quad \text{to avoid non-identifiability from} \quad \tilde{\kappa} = \kappa + c, \quad \tilde{\psi} = \psi - c\phi \tag{2.2}$$

$$\sum_{\nu} |\kappa_\nu|^2 \overset{!}{=} 1 \quad \text{to avoid non-identifiability from} \quad \tilde{\kappa} = r\kappa, \quad \tilde{\phi} = r^{-1}\phi, \quad r \in \mathbb{R}_{>0}. \tag{2.3}$$

In an iterative procedure, the maximum likelihood estimators $\hat{\psi}, \hat{\phi}, \hat{\kappa}, \hat{\Sigma}$ are computed and then the estimated spectrum $\hat{I}$ and its orthogonal component $\hat{\omega}$ are extracted from $\hat{\kappa}$ using a complex rotation

$$\hat{I} = \Re(e^{i\lambda}\hat{\kappa}), \qquad \hat{\omega} = \Im(e^{i\lambda}\hat{\kappa}), \tag{2.4}$$

where λ is selected according to an optimality criterion (for example with the maximum method, see (2.5)). Exemplarily, for orientation g_x from a chemical sample of the fluorine-nitroxide compound depicted in the right panel of Figure 1.2, maximum likelihood estimates (including point-wise confidence bands obtained by parametric bootstrap using 10,000 bootstrap samples) are depicted in Figure 2.9. For different chemical compounds, the Kolmogorov-Smirnov goodness of fit tests of

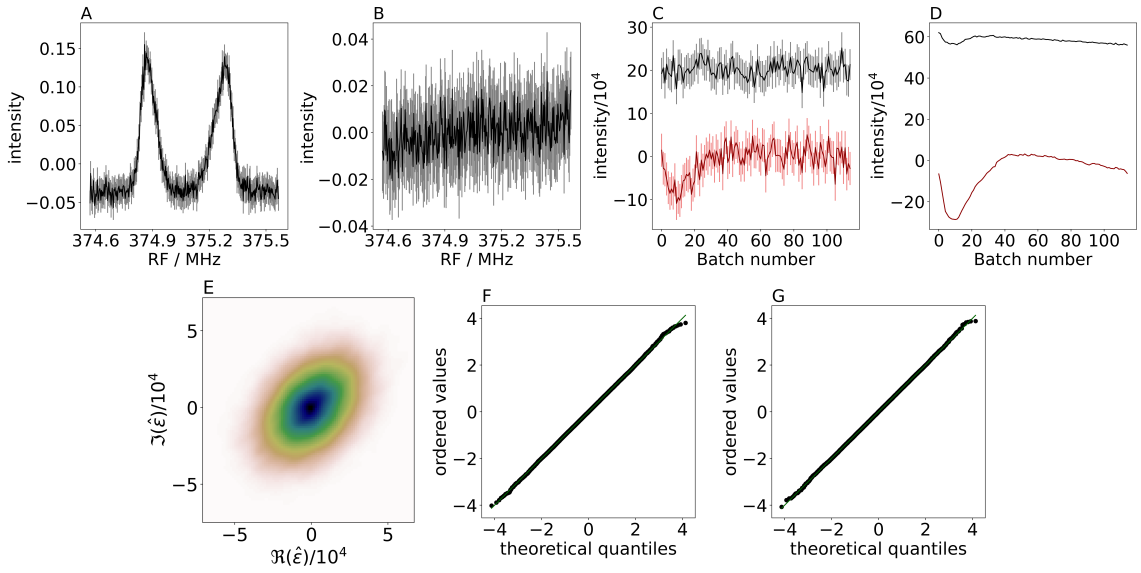


Figure 2.9: Applying the homoscedastic drift model to a typical ENDOR measurement. Panel A displays the estimated spectrum $\hat{I}$, while panel B displays the component $\hat{\omega}$ that is orthogonal to the estimated spectrum $\hat{I}$. Panel C and D show the real (black) and imaginary (red) components of $\hat{\phi}$ and $\hat{\psi}$, respectively. Panel E displays a kernel-density-estimation of the complex residuals $\hat{\epsilon}_{b,\nu}$, while panels F and G depict q-q-plots for the real and imaginary components of the standardized residuals, respectively. The Figure is adapted from Paper D.

the real and imaginary parts of the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu$ (exemplarily shown in panels E, F, and G of Figure 2.9) usually cannot be rejected, see Pokern et al. (2021); Hiller et al. (2022); Wiechers et al. (2023).

2.2.1 Paper C: Drift Models on Complex Projective Space for Electron-Nuclear Double Resonance

The Paper *Drift Models on Complex Projective Space for Electron-Nuclear Double Resonance* is included in Section C and the paper is submitted to arXiv (https://doi.org/10.48550/arXiv.2307.12414). The paper has been developed in an interdisciplinary collaboration with the 'Electron Paramagnetic Resonance' research group headed by Marina Bennati. From the spectroscopy side, Igor Tkach and Marina Bennati contributed to the paper. From the statistics group, Markus Zobel, Benjamin Eltzner, Stephan F. Huckemann, Yvo Pokern and myself contributed to the paper.

The homoscedastic drift model (2.1) is the first of its kind in the field of ENDOR spectroscopy, achieving surprisingly good model fits for experiments at a microwave frequency of $\nu_{MW} = 263\,\mathrm{GHz}$. In order to justify statistical methods such as the parametric bootstrap and statistical tests, there is great interest in asymptotic results for the homoscedastic drift model. For the special case $\Sigma = \mathrm{diag}(c, c)$ for $c > 0$, the maximum likelihood estimators $\hat{\phi}$ and $\hat{\kappa}$ are determined by

$$(\hat{\phi}, \hat{\kappa}) \in \operatorname*{arg\,min}_{\phi\in\mathbb{C}^B, \kappa\in M} \left\|\tilde{Y} - \phi\kappa^T\right\|^2, \quad \text{where} \quad M \coloneqq \left\{\kappa \in \mathbb{C}^N : \ \sum_\nu \kappa_\nu = 0, \ \sum_\nu |\kappa_\nu|^2 = 1\right\}$$

and $\tilde{Y}_{b,\nu} \coloneqq Y_{b,\nu} - \hat{\psi}_b = Y_{b,\nu} - \frac{1}{B}\sum_{\nu=0}^N Y_{b,\nu}$ is the row-centered data matrix. In this case, the estimate is given by the *singular value decomposition* of $\tilde{Y}$, where $\sigma_1^{-1}\hat{\phi}$ and $\overline{\hat{\kappa}}$ are the left and right singular vectors, respectively, belonging to the largest singular value σ_1. Note that in contrast to PCA, the row centered matrix is used rather than the column centered matrix (cf. Section 1.2). Thus, the κ of central tendency that gives the largest variability across frequencies is determined rather than the κ that corresponds to the largest variability across repeated measures. However, in the more general real-world case where Σ has two distinct eigenvalues and is not diagonal, this analogy and the asymptotic results for PCA (for example in Anderson (2003b)) are not applicable. The maximum likelihood estimators $\hat{\phi}$ and $\hat{\kappa}$ are determined by

$$(\hat{\phi}, \hat{\kappa}) \in \operatorname*{arg\,min}_{\phi\in\mathbb{C}^B, \kappa\in M} \sum_{b=1}^B \sum_{\nu=1}^N \mathrm{vec}\left(\tilde{Y}_{b,\nu} - \phi_b\kappa_\nu\right)^T \Sigma^{-1} \mathrm{vec}\left(\tilde{Y}_{b,\nu} - \phi_b\kappa_\nu\right)$$

Strong consistency for $B \to \infty$ can be proved, proceeding as follows: first, we insert the *conditional maximum likelihood estimator*

$$\hat{\phi}_b(\kappa, \Sigma, \tilde{Y}) = \left(\sum_{\nu=1}^N M(\kappa_\nu)^T\Sigma^{-1}M(\kappa_\nu)\right)^{-1}\left(\sum_{\nu=1}^N M(\kappa_\nu)^T\Sigma^{-1}\mathrm{vec}\left(\tilde{Y}_{b,\nu}\right)\right).$$

where $M(c) \coloneqq \begin{pmatrix} \Re(c) & -\Im(c) \\ \Im(c) & \Re(c) \end{pmatrix}$ for $c \in \mathbb{C}$. Then we obtain the following minimization problem

$$\hat{\kappa} \in \operatorname*{arg\,min}_{\kappa \in M} \sum_{b=1}^{B} \sum_{\nu=1}^{N} \operatorname{vec}\left(\tilde{Y}_{b,\nu} - \hat{\phi}_b(\kappa, \Sigma, \tilde{Y})\kappa_\nu\right)^T \Sigma^{-1} \operatorname{vec}\left(\tilde{Y}_{b,\nu} - \hat{\phi}_b(\kappa, \Sigma, \tilde{Y})\kappa_\nu\right) = \operatorname*{arg\,min}_{\kappa \in M} \sum_{b=1}^{B} \rho(\tilde{Y}_b, \kappa),$$

where

$$\rho(\tilde{Y}_b, \kappa) \coloneqq \sum_{\nu=1}^{N} \operatorname{vec}\left(\tilde{Y}_{b,\nu} - \hat{\phi}_b(\kappa, \Sigma, \tilde{Y})\kappa_\nu\right)^T \Sigma^{-1} \operatorname{vec}\left(\tilde{Y}_{b,\nu} - \hat{\phi}_b(\kappa, \Sigma, \tilde{Y})\kappa_\nu\right).$$

This can be identified as a generalized Fréchet mean, introduced in Section 1.4. A difficulty is that due to matrix multiplication with Σ^{-1} matrices of the form $\sum_{\nu=1}^{N} M(\kappa_\nu)^T \Sigma^{-1} M(\kappa_\nu) \in \mathbb{R}^{2\times 2}$ appear, which usually cannot be interpreted as complex numbers. We proceed as follows. The constraint (2.2) can be eliminated using for example the Helmert sub-matrix (defined in Section 1.1.2), resulting in a projected $\kappa^H \in \mathbb{C}^{N-1}$ and $\tilde{Y}^H \in \mathbb{C}^{B\times N-1}$. We treat the projected row-centered data matrix $\tilde{Y}^H$ as a sequence of B identically and independently distributed $\mathbb{C}^{N-1}$-valued random variables $\tilde{Y}_1^H, \ldots, Y_B^H$. Thus, our data space is $\mathfrak{Q} = \mathbb{C}^{N-1}$. Adding the constraint (2.3), the projected κ is an element on a complex sphere. For any phase $\lambda \in [0, 2\pi)$, it holds $(\exp(i\lambda)\phi)((\exp(-i\lambda)\kappa)^T = \phi\kappa^T$. Removing this additional phase factor gives us the complex projective space $\mathbb{C}P^{N-2}$ as parameter space (introduced in Section 1.1.3). It is a Riemannian manifold of real dimension $2N - 4$.

To show strong consistency for the homoscedastic drift model, we first extend strong consistency in the sense of (ZC) and of (BPC) as described in Section 1.4. Then (under some technical conditions), the generalized theory is used to prove strong consistency for κ. For this purpose, in particular, a local Lipschitz constant is determined (see Theorem 4.3 of Paper C) and the uniqueness of the population Fréchet ρ-means is proved (see Theorem 4.5 of Paper C). Then the properties required for Theorem 6 from Huckemann (2011a) are shown to prove a CLT in a suitable chart β around the true but unknown $[\kappa^{(0)}]$, which has the property $\beta([\kappa^{(0)}]) = 0$. We denote the gradient of $x \mapsto \rho(Y, \beta^{-1}(x))$ by $\operatorname{grad}_2\rho(Y, [\kappa])$ and by $H_2\rho(Y, [\kappa])$ the corresponding Hessian matrix. The regularity conditions (10) and (11) of Huckemann (2011a) are shown to hold and it is verified that the population Hessian matrix $\mathbb{E}[H_2\rho(Y, [\kappa^{(0)}])]$ is invertible. We obtain the following CLT.

Theorem 2.1 (CLT)**.** *For the centered homoscedastic drift model from Definition 4.1 and Assumption 4.7, let $[\hat{\kappa}^{(B)}(\omega)] \in E_B^{(\rho)}(\omega)$ be a measurable selection for all $\omega \in \Omega$, then, omitting ω,*

$$\sqrt{B}\beta([\hat{\kappa}^{(B)}]) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\mathbb{E}\left[H_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1} \left(\operatorname{cov}\left[\operatorname{grad}_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right) \left(\mathbb{E}\left[H_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1}\right)$$

holds for the chart β defined in Definition 4.8 of Paper C.

In the ENDOR experiment one is interested in the spectrum I, which is extracted from κ using a complex multiplication, see (2.4): one possible approach is the maximum method (see for example Paper D) in which

$$\lambda_{\text{opt}} \in \underset{\lambda \in [0,\pi)}{\arg\max} \left\| \Re(\exp(i\lambda)\kappa) \right\| \tag{2.5}$$

is chosen such that the norm of $I = \Re(\exp(i\lambda_{\text{opt}})\kappa)$ is maximal. The function that determines the corresponding I for a given κ is explicitly determined and the corresponding Jacobi matrix is computed. Using the delta method (see, e.g., van der Vaart (2000)), we show a central limit theorem (similar to the central limit theorem from Theorem 2.1) for the ENDOR spectrum $\hat{I}$ extracted by the maximum method.

Subsequently, we address the more complicated case of joint estimation of κ and Σ in the profile likelihood model. It is demonstrated that the joint estimation of κ and Σ is not consistent. To achieve a consistent estimator for κ and Σ, one would heuristically expect that a proper treatment of the randomness in both ϵ and ϕ is required, which the profile likelihood does not provide for ϕ. This calls for further research (see Section 3.3).

The homoscedastic drift model was developed for ENDOR data recorded at $\nu_{MW} = 263\,\text{GHz}$. Only a handful of groups worldwide have the equipment to perform experiments at such high MW frequencies. Far more groups have equipment for experiments at $\nu_{MW} = 94\,\text{GHz}$, the so-called *W-band*. However, the homoscedastic drift model introduced by Pokern et al. (2021) does not fit the spectral data recorded at $\nu_{MW} = 94\,\text{GHz}$ because heteroscedastic noise is observed. The heteroscedasticity is attributed to phase noise of the EPR echo, leading to a new parametric *heteroscedastic drift model*, which is presented and tested on ENDOR data in Paper C. It achieves a fair model fit and a remarkable improvement in the signal-to-noise ratio, compared to the averaging model.

Structure of the paper:

- Section 1: Introduction.
- Section 2: The homoscedastic drift model is introduced.
- Section 3: We introduce the research area of strong consistency for generalized Fréchet means and prove two new theorems for a general framework.
- Section 4: The theory developed in Section 3 is applied to prove strong consistency for the homoscedastic drift model under certain assumptions. Subsequently, a central limit theorem is proved for the maximum likelihood estimator of κ and, building on it, for I.
- Section 5: In Section 5, we demonstrate with detailed calculations that the joint estimate of κ and Σ is not consistent.

- Section 6: In Section 6, we introduce, discuss, and apply the heteroscedastic drift model to data.

- Section 7: Outlook.

Own contribution: With guidance from my supervisors, I developed and formulated all the proofs and calculations from Sections 2, 3, 4, and 5. In addition, I worked on early versions of the heteroscedastic drift model. First, I worked on a nonparametric approach that used a penalized likelihood. This approach achieved a reasonable model fit and improved signal-to-noise, but suffered from *shrinkage*. For this reason, I subsequently worked on a first parametric model. This model was significantly extended and improved in the bachelor thesis of Markus Zobel, who mainly worked on the corresponding parts of the paper (Section 6 and corresponding parts of the Supplement).

2.2.2 Paper D: Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra

The paper *Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra* is published in *Journal of Magnetic Resonance* (see https://doi.org/10.1016/j.jmr.2023.107491) and is included in Section D of this thesis. The paper has been developed in an interdisciplinary collaboration with the 'Electron Paramagnetic Resonance' research group headed by Marina Bennati. From the spectroscopy side, Annemarie Kehl, Markus Hiller, Andreas Meyer, Igor Tkach and Marina Bennati contributed to the paper. From the statistics group, Benjamin Eltzner, Stephan F. Huckemann, Yvo Pokern and myself contributed to the paper.

In this paper, physical parameters describing the conformation of two different fluoronitroxide compounds (one of them is shown in the right panel of Figure 1.2) are estimated, including rigorous statistical error propagation. Previous spectroscopic work (for example Kehl et al. (2021)) proceeds by laborious manual adjustment of parameters starting from *density functional theory* (DFT) derived values. Uncertainties were specified by varying one physical parameter at a time, thereby obscuring large correlations of uncertainties in the parameter vector. In contrast, in Paper D we replace manual parameter tuning by an optimization consisting of *Bayesian optimization* and a local optimization that uses an accelerated simulation code to determine the parameters that are closest to the estimated spectrum. In addition to a better fit between measured and simulated spectra, the approach also provides the stochastic error of the obtained parameter estimates, by propagating the spectral uncertainties, made available by the homoscedastic drift model, to the parameters. For this purpose, we developed the following pipeline for ^{19}F ENDOR spectra (simplified for a single orientation). Notably, it can also be adapted to general ENDOR spectra.

1. Use the *drift model* to obtain the estimated spectrum $\hat{I} \in \mathbb{R}^N$ and the parametric bootstrap to obtain the corresponding covariance matrix $\chi \in \mathbb{R}^{N \times N}$ (denoising step in the Figure 1.2).

2. Solve the optimization problem

$$\hat{\theta} \in \arg\min_{\theta \in \Theta} \sum_{\nu=1}^{N} |\hat{I}_\nu - I_\nu(\theta)|^2,$$

where θ is a parameter vector and $I(\theta)$ is the output of our spectrum simulation software for a given $\theta \in \Theta$. We employ *Bayesian optimization* to perform a global search, followed by a local optimization by a gradient-based method (optimization step in the Figure 1.2).

3. Calculate the Jacobi matrix $J \in \mathbb{R}^{\dim(\Theta)\times N}$ of the simulated spectrum $I(\theta)$ with respect to the parameter vector θ using finite difference approximation. Approximate the covariance matrix $\mathrm{Cov}(\theta)$ of the parameters by linear propagation of Gaussian errors

$$\mathrm{Cov}(\theta) \approx (J^T)^+ \chi J^+ \in \mathbb{R}^{\dim(\Theta)\times\dim(\Theta)},$$

where J^+ is the *Moore–Penrose inverse* of J.

We successfully apply this pipeline to two different fluoronitroxide compounds to obtain estimates for parameter values and the covariance matrices of their errors. For the compound shown in the right panel of Figure 1.2, the following parameters are estimated (a more detailed explanation of the individual parameters can be found in the Paper D):

- The hyperfine interaction tensor A, which depends on four parameters, all of which provide information about the structure of the sample. In particular, the tensor A depends on the so-called *dipolar coupling strength* T, from which the interspin distance r between the nitroxide radical and the fluorine nucleus (see infer step of Figure 1.2) can be calculated directly.
- A chemical shielding tensor σ, a symmetric 3×3 matrix (with 6 parameters), from which structural information about the phenyl ring can be derived.
- For each orientation, two experimental parameters (namely magnetic field strength (denoted with Field in Figure 2.10) and a line broadening parameter (denoted with lw in Figure 2.10) are estimated. A statistical analysis has shown that these are not as precisely known as originally expected and the additional estimation of these parameters leads to a significantly better fit.

Panel A of Figure 2.10 shows the ^{19}F ENDOR spectra estimated with the homoscedastic drift model and the spectra simulated with the estimated parameters. The spectral residuals are very close to white noise, which indicates that the optimization procedure results in a very good fit. In panel B, the correlation matrix of the corresponding parameters is shown. For example, a strong positive correlation between the magnetic field strengths and the chemical shielding tensor σ can be inferred.

Structure of the paper:

- Section 1: Introduction.

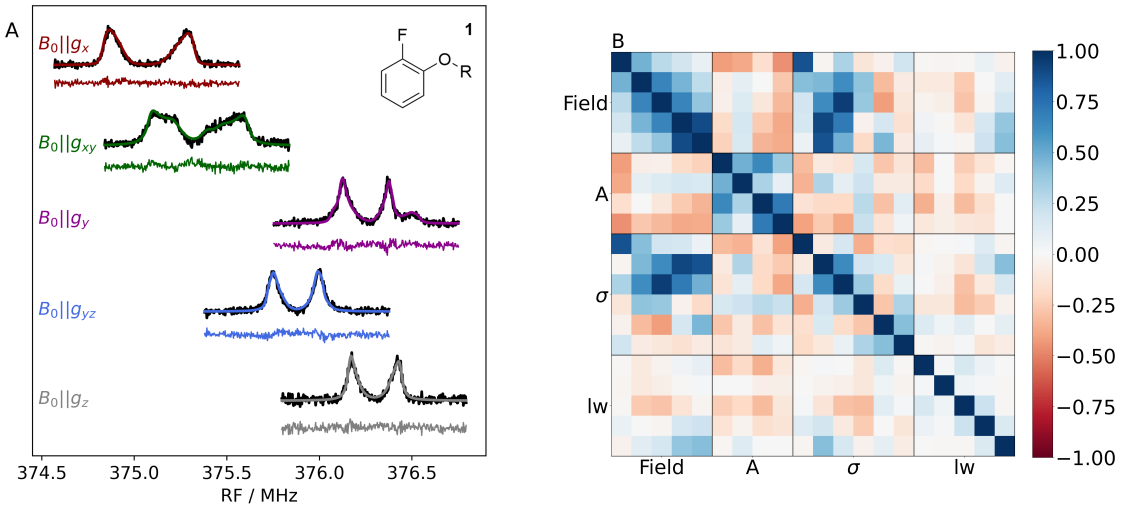


Figure 2.10: A: the ^{19}F ENDOR spectra $\hat{I}_{o,\nu}$ (black) extracted with the homoscedastic drift model, where o stands for the orientations. In different colors: the corresponding spectra $I_{o,\nu}(\hat{\theta})$ simulated with the values that emerged from the optimization from Paper D. The spectral residuals are plotted below each of the spectra using the same color. B: correlation matrix of the corresponding parameters (from top to bottom/left to right): the magnetic field strength parameters for each of the five orientations, the four parameters of the hyperfine matrix A, the six parameters of the chemical shielding tensor σ and broadening parameter lw for each of the orientations. The figure is taken from Paper D, where the individual parameters are explained in more detail.

- Section 2: The experiment, the data (EPR and ENDOR), and data processing with the homoscedastic drift model from Pokern et al. (2021) are presented.
- Section 3: The parameters included in the optimization process are discussed and the spectrum simulation algorithm is presented and compared with the version of Kehl et al. (2021).
- Section 4: The methodology for estimating the parameters and quantifying the stochastic part of the error is presented.
- Section 5: The exact procedure for the optimization and the optimization results are presented.
- Section 6: Materials and Methods

Own contribution: Since ENDOR spectra are recomputed each time the loss function is called, both speed and accuracy are crucial for the success of the optimization. When I started working on the project, it took several minutes to compute a single spectrum with sufficient accuracy, therefore one of my first challenges was to speed up the code to allow for optimization. By *tensorizing* the code to take advantage of powerful numerical linear algebra subroutines, *parallelizing* blocks of code that are independent of each other, and *precomputing* expressions that do not change in the optimization pipeline, significant speed improvements were achieved (see Table S1 in Paper D and Section C in the Supplementary Information of Paper D). Furthermore, I developed (guided by my supervisors) the code for estimating the parameters and quantifying the stochastic error (Section 4 in Paper D and Section E in the Supplementary Information of Paper D). In regular meetings I

presented intermediate results, ideas or problems to the spectroscopy group, which often resulted in adjustments of methods or new data analysis both on the spectroscopy and on our side (Section 3 and 5 in Paper D and section A and F in the Supplementary Information of Paper D). I also presented the project at two different CRC retreats. Together with Markus Hiller in Hofgeismar in 2021 and together with Annemarie Kehl in Goslar in 2022.

CHAPTER 3

Outlook

In this section, an overview of ongoing work and some research questions that need to be addressed in the future are presented.

3.1 Impact of the Paper B on the biological community and regression between stratified spaces

Recently, the methods from Paper B received the attention from the Richardson group from the Duke Department of Biochemistry (see Section 1), who are experts in the field of three-dimensional structure analysis of RNA molecules. The aim of the joint work is to understand the interdependence of two different scales, namely between the so-called *low-resolution scale* and the *high-resolution scale* corresponding to the suites we have studied in Paper A and Paper B. A suite comprises the RNA region between two sugar rings. The corresponding low-resolution scale is characterized by 5 atoms: atom N1 or N9 (depending on the base) and atom C1' at both sugar rings and additionally atom P' between the two sugar rings (see Figure 2.1). These atoms were chosen because their coordinates can be identified even in experiments with low resolution, while the atomic positions in between can only be determined with less accuracy. In the nascent collaboration, we obtained filtered PDB data from the Richardson group, which we split into four different sub-data sets depending on the folding of the ribose sugar ring.

At the moment, we are working empirically on how to parameterize the low-resolution scale suitably (e.g. in a landmark-based shape space (see Section 1.1.2), with *Bookstein coordinates* (see for example Dryden and Mardia (2016)) or using angles, dihedral angles and distances between the consecutive atoms). One of the objectives is to identify small clusters in the sub-data sets. Therefore, we are currently developing a new mode hunting method together with the master student Franziska Hoppe, which is specifically suited to identify small clusters.

A first approach to understand the interdependence of the two scales is to model the low-resolution scale in the similarity shape space (see Section 1.1.2), opening the possibility to perform PNS (introduced in Section 1.2) from Jung et al. (2012) since the pre-shapes are represented on a high-

dimensional sphere, similar to Dryden et al. (2019). Analogously, Torus PCA (see Section 1.2) can be performed on the high-resolution suites. The respective Euclidean matrices of the residuals can be used, for example, for regression between the two scales.

A more general approach to understand the interdependence of the two scales is to develop non-linear regression methods between stratified spaces. The low-resolution scale can be modeled on a stratified (see Huckemann and Hotz (2014)) landmark-based shape space, and data on the torus can be mapped to a stratified sphere in the same way as in torus PCA (see Section 1.2). A challenge is to develop regressions between such spaces (see e.g. Marzio et al. (2014, 2019) for regressions between spheres).

3.2 Nonlinear Regression for DFT-Calculations

In Pokern et al. (2021) and Hiller et al. (2022), ^{1}H Davies ENDOR spectra of E. coli ribonucleotide reductase $Y_{122}^{\bullet}$ were measured, revealing broad features that cannot be explained by a single conformation of the molecule, but by a distribution of molecular conformations. The distribution of molecular conformations depends on the dihedral angle α determined by the atoms C^{α}, C^{β}, C_1 and C_2 (see left panel of Figure 3.1). Using a computationally intensive DFT calculation, hyperfine tensors which are elements of the symmetric 3×3 matrices Sym(3) were computed for various values of α in 5 degree steps (depicted in the right panel of Figure 3.1). DFT can

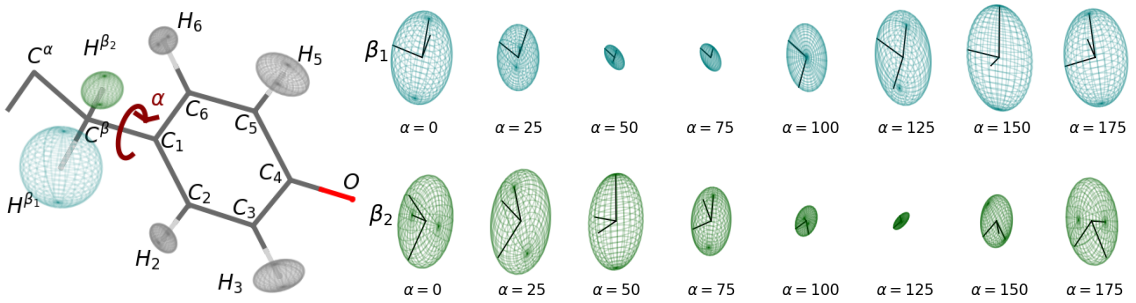


Figure 3.1: Left: the atomic structure of $Y_{122}^{\bullet}$ for the dihedral angle $\alpha = 150°$. The corresponding hyperfine tensors determined with the DFT calculation are displayed as ellipsoids at the centers of the atomic positions. Right: the β_1 and β_2 hyperfine tensors determined with DFT calculation for different dihedral angle values α.

predict both the distributions of conformations and the map from the conformations to the hyperfine tensors. However, the computations are very time-consuming and are only approximations. In a first manuscript, we developed a parametric domain knowledge-driven model which, for a given dihedral angle α, predicts hyperfine tensors. At the moment, Rajan Alexander (a PhD student supervised by Yvo Pokern) is working on Gaussian process based modeling on manifolds to learn the map from the conformations to the hyperfine tensors from DFT data (which may provide an alternative approach to tackle the regression scenarios from Section 3.1).

In particular such approaches may tackle more complicated molecules, such as the ^{19}F-labeled nitrosyl radical (studied by the Bennati group), where, on the one hand, domain knowledge-driven models are hardly feasible and, on the other hand, the distribution is no longer a simple Gaussian but will be a mixture of different Gaussians. They may even be informed by molecular dynamics simulations with appropriate adaptive force field refinement.

3.3 New drift models and asymptotic theory

So far, the homoscedastic drift model exists for ENDOR data recorded at $\nu_{MW} = 263\,\text{GHz}$ and we also developed a heteroscedastic drift model for $\nu_{MW} = 94\,\text{GHz}$ Mims (see Mims (1965)) data in Paper C. For the homoscedastic drift model, asymptotic theory was developed for the case where Σ is known. Since the joint estimate of κ and Σ is not consistent in the profile likelihood model (see Section 2.2.1), it would be desirable to obtain a consistent estimate by including the randomness of ϕ in the statistical model. Possible approaches are to model the ϕ as i.i.d. Gaussian or, to reflect the likelihood being invariant under permutations of batches, as exchangeable random variables or, perhaps most realistically, as a Gaussian process. For the latter two approaches, one would need to generalize the theory about generalized strong consistency of Fréchet means (see Section 1.4) for random variables that are not i.i.d. Furthermore, an asymptotic analysis for the heteroscedastic model is future work. Especially challenging is that the mean ψ_b and the variance Σ_b are dependent on each other.

Note that only a handful of groups worldwide have the equipment to perform experiments at $\nu_{MW} = 263\,\text{GHz}$. Far more groups have equipment for experiments at $\nu_{MW} = 94\,\text{GHz}$ and even more groups have equipment at $\nu_{MW} = 34\,\text{GHz}$ or $\nu_{MW} = 9\,\text{GHz}$. Therefore, it is challenging to develop drift models for all microwave frequencies and pulse sequences to make them applicable to a large audience. The group of Yvo Pokern at UCL has worked on extending the heteroscedastic drift model to $\nu_{MW} = 94\,\text{GHz}$ Davies (a specific pulse sequence, see Davies (1974)) data. In addition, the Bennati group has collected experimental data at $\nu_{MW} = 34\,\text{GHz}$ and $\nu_{MW} = 9\,\text{GHz}$, for which drift models are not yet available.

3.4 Impact of the accelerated spectra simulation code

The accelerated spectral simulation code (developed in Paper D) is currently used by Laura Remmel (from the Bennati group) to work on the so-called *fluoride riboswitch* (see Ren et al. (2012)). Here, spin labels are present in the frozen sample in different conformations. The ENDOR spectrum is a sum of these conformations, and to properly simulate the spectrum, a sum of all individual spectra has to be simulated. Therefore, the accelerated spectral simulation code is crucial for the realization of this project. In general, these or similar labelling methods can be applied to many other biomolecules (e.g. proteins, RNA, DNA).

Bibliography

Afsari, B. (2009). *Means and averaging on Riemannian manifolds*. University of Maryland.

Afsari, B. (2011). Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139:655–773.

Altis, A., Otten, M., Nguyen, P. H., Rainer, H., and Stock, G. (2008). Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *The Journal of Chemical Physics*, 128(24):245102.

Anderson, A. C. (2003a). The process of structure-based drug design. *Chemistry & Biology*, 10(9):787–797.

Anderson, T. W. (2003b). *An Introduction to Multivariate Statistical Analysis*. Wiley Interscience.

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29.

Choquet, G. (1954). Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295.

Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., and Das, R. (2013). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods*, 10(1):74–76.

Davies, E. R. (1974). A new pulse endor technique. *Physics Letters A*, 47(1):1–2.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38. With discussion.

Dryden, I. L., Kim, K.-R., Laughton, C. A., and Le, H. (2019). Principal nested shape space analysis of molecular dynamics data.

Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis*. Wiley, Chichester, 2nd edition.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785.

Eltzner, B., Hansen, P., Huckemann, S. F., and Sommer, S. (2022). Diffusion means in geometric spaces.

Eltzner, B., Huckemann, S., and Mardia, K. V. (2018). Torus principal component analysis with applications to RNA structure. *Ann. Appl. Stat.*, 12(2):1332–1359.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press.

Estivill-Castro, V. (2002). Why so many clustering algorithms: A position paper. *SIGKDD Explor. Newsl.*, 4(1):65–75.

Evans, S. N. and Jaffe, A. Q. (2020). Strong laws of large numbers for fr\'echet means. *arXiv preprint arXiv:2012.12859*.

Feher, G. (1956). Observation of nuclear magnetic resonances via the electron spin resonance line. *Phys. Rev.*, 103:834–835.

Fletcher, P. T. and Joshi, S. C. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. *ECCV Workshops CVAMIA and MMBIA*, pages 87–98.

Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, Hugo, and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, 2(3-4):282–285.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. 10(4):215–310.

Gemperle, C. and Schweiger, A. (1991). Pulsed electron-nuclear double resonance methodology. *Chemical Reviews*, 91(7):1481–1505.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40(1):33–51.

Hendriks, H. and Landsman, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statistics & Probability Letters*, 26(2):169–178.

Hendriks, H. and Landsman, Z. (1998). Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions. *Journal of Multivariate Analysis*, 67(2):227–243.

Hiller, M., Tkach, I., Wiechers, H., Eltzner, B., Huckemann, S., Pokern, Y., and Bennati, M. (2022). Distribution of h-β hyperfine couplings in a tyrosyl radical revealed by 263 ghz endor spectroscopy. *Applied Magnetic Resonance*, 53:1015–1030.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321–377.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

Huckemann, S. (2011a). Inference on 3d procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics*, 38(3):424–446.

Huckemann, S. and Hotz, T. (2009). Principal component geodesics for planar shape spaces. *Journal of Multivariate Analysis*, 100(4):699–714.

Huckemann, S. and Hotz, T. (2014). On means and their asymptotics: circles and shape spaces. *Journal of mathematical imaging and vision*, 50(1-2):98–106.

Huckemann, S., Hotz, T., and Munk, A. (2010). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica*, 20(1):1–100.

Huckemann, S. and Ziezold, H. (2006). Principal component analysis for riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319.

Huckemann, S. F. (2011b). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics*, 39(2):1098 – 1124.

Huckemann, S. F. and Eltzner, B. (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *The Annals of Statistics*, 46(5):1994 – 2019.

Ippolito, J. A. and Steitz, T. A. (2000). The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution. *Journal of Molecular Biology*, 295(4):711–717.

Jung, S., Dryden, I. L., and Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, 99(3):551–568.

Kehl, A., Hiller, M., Hecker, F., Tkach, I., Dechert, S., Bennati, M., and Meyer, A. (2021). Resolution of chemical shift anisotropy in ^{19}F ENDOR spectroscopy at 263 GHz/9.4 T. *Journal of Magnetic Resonance*, 333:107091.

Kendall, D. G. (1974). Foundations of a theory of random sets, stochastic geometry (harding ef and kendall dg, eds.).

Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability*, 9(3):428–430.

Kent, J. and Mardia, K. (2015). The winding number for circular data. In *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop*.

Kent, J. T. and Mardia, K. V. (2009). Principal component analysis for the wrapped normal torus model. *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop 2009*.

Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5):719–720.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D*, 75(10):861–877.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, New York.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic press.

Marzio, M. D., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *Journal of the American Statistical Association*, 109(506):748–763.

Marzio, M. D., Panzera, A., and Taylor, C. C. (2019). Nonparametric rotations for sphere-sphere regression. *Journal of the American Statistical Association*, 114(525):466–476.

Matheron, G. (1974). *Random sets and integral geometry*. John Wiley & Sons.

Meyer, A., Dechert, S., Dey, S., Höbartner, C., and Bennati, M. (2020). Measurement of Angstrom to Nanometer Molecular Distances with ^{19}F Nuclear Spins by EPR/ENDOR Spectroscopy. *Angewandte Chemie International Edition*, 59(1):373–379.

Mims, W. B. (1965). Pulsed Endor Experiments. *Proceedings of the Royal Society of London Series A — Mathematical and Physical Sciences*, 283(1395):452–457.

Murray, L. J. W., Arendall, W. B., Richardson, D. C., and Richardson, J. S. (2003). RNA backbone is rotameric. *Proceedings of the National Academy of Sciences*, 100(24):13904–13909.

Obulkasim, A., Meijer, G. A., and van de Wiel, M. A. (2015). Semi-supervised adaptive-height snipping of the hierarchical clustering tree. *BMC Bioinformatics*, 16(1):15.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Pennec, X. (2018). Barycentric subspace analysis on manifolds. 46(6A):2711–2746.

Pokern, Y., Eltzner, B., Huckemann, S. F., Beeken, C., Stubbe, J., Tkach, I., Bennati, M., and Hiller, M. (2021). Statistical analysis of ENDOR spectra. *Proceedings of the National Academy of Sciences*, 118(27):e2023615118.

Ren, A., Rajashankar, K. R., and Patel, D. J. (2012). Fluoride ion encapsulation by mg2+ ions and phosphates in a fluoride riboswitch. *Nature*, 486(7401):85–89.

Richardson, J. S., Williams, C. J., Hintze, B. J., Chen, V. B., Prisant, M. G., Videau, L. L., and Richardson, D. C. (2018). Model validation: local diagnosis, correction and when to quit. *Acta Crystallographica Section D*, 74(2):132–142.

Sargsyan, K., Wright, J., and Lim, C. (2012). Geopca: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic acids research*, 40(3):e25–e25.

Schlick, T. and Pyle, A. M. (2017). Opportunities and challenges in rna structural modeling and design. *Biophysical journal*, 113(2):225–234. 28162235[pmid].

Schötz, C. (2022). Strong laws of large numbers for generalizations of fréchet mean sets. *Statistics*, 56(1):34–52.

Schrödinger, LLC (2015). The PyMOL molecular graphics system, version 1.8.

Sieranoja, S. and Fränti, P. (2019). Fast and general density peaks clustering. *Pattern Recognition Letters*, 128:551–558.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38:1409–1438.

Sommer, S. (2013). Horizontal dimensionality reduction and iterated frame bundle development. In *Geometric Science of Information*, pages 76–83. Springer.

Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces: Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs: April 16-July 13, 2002, Emile Borel Centre of the Henri Poincaré Institute, Paris, France*, 338:357.

Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M., and Johnson, J. E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex rna. *Nature Structural Biology*, 8(1):77–83.

van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ. Press.

Wang, H.-W. and Wang, J.-W. (2017). How cryo-electron microscopy and x-ray crystallography complement each other. *Protein Sci.*, 26(1):32–39.

Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., and Losick, R. (2004). *Molecular Biology of the Gene*. Pearson Education, fifth edition.

Wiechers, H., Kehl, A., Hiller, M., Eltzner, B., Huckemann, S., Meyer, A., Tkach, I., Bennati, M., and Pokern, Y. (2023). Bayesian optimization to estimate hyperfine couplings from ^{19}f endor spectra. *Journal of Magnetic Resonance*.

Williamson, R. and Janos, L. (1987). Constructing metrics with the heine-borel property. *Proceedings of the American Mathematical Society*, 100(3):567–573.

Zhang, K., Zheludev, I. N., Hagey, R. J., Haslecker, R., Hou, Y. J., Kretsch, R., Pintilie, G. D., Rangan, R., Kladwang, W., Li, S., Wu, M. T.-P., Pham, E. A., Bernardin-Souibgui, C., Baric, R. S., Sheahan, T. P., D'Souza, V., Glenn, J. S., Chiu, W., and Das, R. (2021). Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nature Structural & Molecular Biology*, 28(9):747–754.

Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces.

Zoubouloglou, P., García-Portugués, E., and Marron, J. (2022). Scaled torus principal component analysis. *Journal of Computational and Graphical Statistics*, pages 1–12.

Addenda

The addenda below contains Articles A, B, C, and D, which form the basis for this thesis. In addition, an introduction is included that contains the sources and abstracts of each article.

Principal component analysis and clustering on manifolds

Kanti V. Mardia, Henrik Wiechers, Benjamin Eltzner, and Stephan F. Huckemann

Published in Journal of Multivariate Analysis: https://doi.org/10.1016/j.jmva.2021.104862

Abstract

Big data, high dimensional data, sparse data, large scale data, and imaging data are all becoming new frontiers of statistics. Changing technologies have created this flood and have led to a real hunger for new modelling strategies and data analysis by scientists. In many cases data are not Euclidean; for example, in molecular biology, the data sit on manifolds. Even in a simple non-Euclidean manifold (circle), to summarize angles by the arithmetic average cannot make sense and so more care is needed. Thus non-Euclidean settings throw up many major challenges, both mathematical and statistical. This paper will focus on the PCA and clustering methods for some manifolds. Of course, the PCA and clustering methods in multivariate analysis are one of the core topics. We basically deal with two key manifolds from a practical point of view, namely spheres and tori. It is well known that dimension reduction on non-Euclidean manifolds with PCA-like methods has been a challenging task for quite some time but recently there has been some breakthrough. One of them is the idea of nested spheres and another is transforming a torus into a sphere effectively and subsequently use the technology of nested spheres PCA. We also provide a new method of clustering for multivariate analysis which has a fundamental property required for molecular biology that penalizes wrong assignments to avoid chemically no go areas. We give various examples to illustrate these methods. One of the important examples includes dealing with COVID-19 data.

Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2

Henrik Wiechers, Benjamin Eltzner, Kanti V. Mardia and Stephan F. Huckemann

Published in Journal of the Royal Statistical Society Series C: Applied Statistics: https://doi.org/10.1093/jrsssc/qlad004

Abstract

Three-dimensional RNA structures frequently contain atomic clashes. Usually, corrections approx-

imate the biophysical chemistry, which is computationally intensive and often does not correct all clashes. We propose fast, data-driven reconstructions from clash free benchmark data with two-scale shape analysis: microscopic (suites) dihedral backbone angles, mesoscopic sugar ring centre landmarks. Our analysis relates concentrated mesoscopic scale neighbourhoods to microscopic scale clusters, correcting within-suite-backbone-to-backbone clashes exploiting angular shape and size-and-shape Fréchet means. Validation shows that learned classes highly correspond with literature clusters and reconstructions are well within physical resolution. We illustrate the power of our method using cutting-edge SARS-CoV-2 RNA.

Drift Models on Complex Projective Space for Electron-Nuclear Double Resonance

Henrik Wiechers, Markus Zobel, Marina Bennati, Igor Tkach, Benjamin Eltzner, Stephan F. Huckemann, Yvo Pokern

Submitted to arXiv: https://doi.org/10.48550/arXiv.2307.12414

Abstract

ENDOR spectroscopy is an important tool to determine the complicated three-dimensional structure of biomolecules and in particular enables measurements of intramolecular distances. Usually, spectra are determined by averaging the data matrix, which does not take into account the significant thermal drifts that occur in the measurement process. In contrast, we present an asymptotic analysis for the homoscedastic drift model, a pioneering parametric model that achieves striking model fits in practice and allows both hypothesis testing and confidence intervals for spectra. The ENDOR spectrum and an orthogonal component are modeled as an element of complex projective space, and formulated in the framework of generalized Fréchet means. To this end, two general formulations of strong consistency for set-valued Fréchet means are extended and subsequently applied to the homoscedastic drift model to prove strong consistency. Building on this, central limit theorems for the ENDOR spectrum are shown. Furthermore, we extend applicability by taking into account a phase noise contribution leading to the heteroscedastic drift model. Both drift models offer improved signal-to-noise ratio over pre-existing models.

Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra

H. Wiechers, A. Kehl, M. Hiller, B. Eltzner, S. F. Huckemann, A. Meyer, I. Tkach, M. Bennati, Y. Pokern

Published in Journal of Magnetic Resonance: https://doi.org/10.1016/j.jmr.2023.107491

Abstract

ENDOR spectroscopy is a fundamental method to detect nuclear spins in the vicinity of paramagnetic centers and their mutual hyperfine interaction. Recently, site-selective introduction of ^{19}F as nuclear labels has been proposed as a tool for ENDOR-based distance determination in biomolecules, complementing pulsed dipolar spectroscopy in the range of angstrom to nanometer. Nevertheless, one main challenge of ENDOR still consists of its spectral analysis, which is aggravated by a large parameter space and broad resonances from hyperfine interactions. Additionally, at high EPR frequencies and fields ($\geq$ 94 GHz/3.4 Tesla), chemical shift anisotropy might contribute

to broadening and asymmetry in the spectra. Here, we use two nitroxide-fluorine model systems to examine a statistical approach to finding the best parameter fit to experimental 263 GHz ^{19}F ENDOR spectra. We propose Bayesian optimization for a rapid, global parameter search with little prior knowledge, followed by a refinement by more standard gradient-based fitting procedures. Indeed, the latter suffer from finding local rather than global minima of a suitably defined loss function. Using a new and accelerated simulation procedure, results for the semi-rigid nitroxide-fluorine two and three spin systems lead to physically reasonable solutions, if minima of similar loss can be distinguished by DFT predictions. The approach also delivers the stochastic error of the obtained parameter estimates. Future developments and perspectives are discussed.

CHAPTER A

Principal component analysis and clustering on manifolds

Principal component analysis and clustering on manifolds

Kanti V. Mardia [3,4], Henrik Wiechers [1,*], Benjamin Eltzner [2] and Stephan F. Huckemann [1,*]

[1]Department of Statistics, School of Mathematics, University of Leeds, LS2 9JT and Department of Statistics, University of Oxford, OX1 3LB, UK,
[2]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, Georgia-Augusta-University, Göttingen, 37077, Germany,
[3]Max Planck Institute for Biophysical Chemistry, Göttingen, 37077, Germany,
[4]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, Georgia-Augusta-University, Göttingen, 37077, Germany.

May 25, 2023

Abstract

Big data, high dimensional data, sparse data, large scale data, and imaging data are all becoming new frontiers of statistics. Changing technologies have created this flood and have led to a real hunger for new modelling strategies and data analysis by scientists. In many cases data are not Euclidean; for example, in molecular biology, the data sit on manifolds. Even in a simple non-Euclidean manifold (circle), to summarize angles by the arithmetic average cannot make sense and so more care is needed. Thus non-Euclidean settings throw up many major challenges, both mathematical and statistical. This paper will focus on the PCA and clustering methods for some manifolds. Of course, the PCA and clustering methods in multivariate analysis are one of the core topics.

We basically deal with two key manifolds from a practical point of view, namely spheres and tori. It is well known that dimension reduction on non-Euclidean manifolds with PCA-like methods has been a challenging task for quite some time but recently there has been some breakthrough. One of them is the idea of nested spheres and another is transforming a torus into a sphere effectively and subsequently use the technology of nested spheres PCA. We also provide a new method of clustering for multivariate analysis which has a fundamental property required for molecular biology that penalizes wrong assignments to avoid chemically no go areas. We give various examples to illustrate these methods. One of the important examples includes dealing with COVID-19 data.

Keywords: adaptive linkage clustering, circular mode hunting, dimension reduction, multivariate wrapped normal, SARS-CoV-2 geometry, stratified spheres, torus PCA
MSC 2020 Primary 62H11, 62H15, Secondary 62P10, 62H3

1 Introduction

PCA and clustering analysis are well established topics in multivariate analysis. There are more challenging data that have appeared on non-Euclidean manifolds such as tori, spheres and shape spaces where these new subjects have evolved significantly in the last two decades (see e.g. Mardia and Jupp (2000); Dryden and Mardia (2016)). However, the progress with PCA and clustering for these manifolds has just taken a major step change,

bringing new mathematical tools. Namely the nested sphere methods for PCA on a sphere by Jung et al. (2012) and torus PCA by Eltzner et al. (2018). In shape space PCA methods are now well established (see Gower (1975); Dryden and Mardia (2016)). Using the main principal component of torus PCA we show how a new clustering method for Euclidean and non-Euclidean data can be developed.

For non-Euclidean data, extrinsic as well as tangent space PCA methods have been developed, e.g. by Gower (1975), Fletcher and Joshi (2004), Boisvert et al. (2006), Arsigny et al. (2006), and more recently, intrinsic (geodesic) methods by e.g. Huckemann and Ziezold (2006); Huckemann et al. (2010); Sommer (2013), both of which have been usually successful. Also, methods based on geodesic flow have been proposed, e.g. Panaretos et al. (2014); Yao and Zhang (2020). For the special case of data on spheres, intrinsic and extrinsic methods come close as tangent space PCs naturally map to great circles. Mimicking the forward and backward nested nature of Euclidean PCA, Jung et al. (2012) introduced principal nested spheres (PNS) analysis. A generalization to arbitrary manifolds are barycentric subspaces by Pennec (2018). It is a subtle point that PNS approximates data not only by great subspheres but also by small subspheres, thereby adding increased flexibility. For instance, for data on a m-dimensional sphere, the family of main principal nested circle components (i.e. of small circles) has dimension $3(m-1)$ while for data on a m-dimensional Euclidean space the family of first PCs (i.e. of straight lines) has dimension $2(m-1)$. For PCA-based clustering, this above property is very desirable because clusters that would require two Euclidean PCs to represent can often be separated along the main principal nested circle. We show how they can be separated with statistical guarantees by circular mode hunting.

However, for the torus case (a direct product space of two or more angles), the above methods are inadequate as tangent space PCA fails to take into account the periodicity of the torus and, even worse, geodesic PCA is completely inapplicable because almost all geodesics densely wind around. To circumvent this dead end, one could attempt at mapping a torus to a sphere. Indeed, it is well known that any well behaved manifold can be mapped canonically to a sphere: If Q is simplicial complex with single largest dimensional cell Q^* of dimension $m \in \mathbb{N}$, then the Alexandroff compactification $Q/(Q \setminus Q^*)$ of Q^* (identfying all boundary cells of Q^* with a single point) carries canonically the topological structure of $\mathbb{S}^m$, see for example (Hatcher, 2005, Proposition 2.22, phrased in more general terms). For the torus $\mathbb{T}^m$ viewed as a cube $[0, 2\pi]^m$ with opposite faces identified ($\sim$) with one another, see Definition 1, this would mean to identify all opposite faces with a single point. In contrast, our torus-to-stratified-sphere (TOSS) map introduced in Section 3.1 preserves torus angles nearly unchanged, thus reducing dimensions of faces less drastically and preserving much more of the cyclic structure, at the price, however, of arriving at a properly stratified sphere, as depicted in Fig. A2.

Also data on other spaces can be directly mapped to spheres, for instance data on a shape space can be horizontally lifted to the pre-shape sphere, e.g. Preston and Wood (2010); Dryden et al. (2019); Tran et al. (2021), which then allows for PNS, e.g. Dryden et al. (2019); Yang and Vemuri (2021). Another example is data on a Euclidean space that can be mapped to a best fitting hypersphere. For these three cases: tori, shape spaces and Euclidean spaces we illustrate the new clustering method Mode huntIng on the main principle Nested small Circle of prE-clusters (MINCE) post Adaptive linkaGe clustEring (AGE). The MINCE method is specifically designed for analysis of biomolecular structure: its strength lies in detecting clusters correctly and avoiding assignment to wrong clusters, rather it assigns such data to the outlier set. While no previous knowledge of cluster numbers is required, clusters with too few members are also assigned to the outlier set.

We give here advances in one of the important topics of PCA and clustering on manifolds. However, statistics on manifolds is a fast growing area which has been motivated by challenging and cutting edge applications, see e.g., Huckemann (2021); Mardia (2021); Pewsey and García-Portugués (2021). It should be noted that this subject is geometry driven so somewhat more complicated than the traditional multivariate analysis. In this paper, after introducing our notation and terminology in Section 2, we give a comprehensive introduction to torus PCA in Section 3 followed by a description of the MINCE post AGE method in Section 4 where we cover three typical applications. The first two address SARS-CoV-2 structure clustering on the microscopic level of backbone suites using torus PCA and on a mesoscopic level using PNS on lifts to the pre-shape sphere of shape spaces. Using simulated data from a recent benchmark algorithm, the last application illustrates how to apply the method to Euclidean data. One of the simplest model based methods for torus PCA is to use the wrapped multivariate normal distribution following indirectly a multivariate normal formulation for Euclidean PCA. However, this method is not yet well explored. This is briefly presented in Section 5. We conclude with an overview addressing some important research questions.

2 Notation and Terminology

Let $x := (x_1, \ldots, x_k)^T \in \mathbb{R}^k$ denote a column vector with Euclidean norm $\|x\| := \sqrt{x^T x}$ and $\mathbb{S}^k := \{y \in \mathbb{R}^{k+1} : \|y\| = 1\}$ the unit sphere with spherical distance

$$d_{\mathbb{S}^k}(x, y) := \arccos(x^T y), \quad x, y \in \mathbb{S}^k .$$

For a space Q with equivalence relation $\sim$ we set

$$Q/\sim := \{[q] : q \in Q\}, \quad [q] := \{q' \in Q : q' \sim q\} \text{ for } q \in Q .$$

We will be dealing with many such quotient spaces, most fundamental is the torus.

Definition 1. For $k \in \{1, 2, \ldots\}$, $\mathbb{T}^k := [0, 2\pi]^k/\sim$ is the k-dimensional torus where

$$[0, 2\pi]^k \ni x := (x_1, \ldots, x_k) \sim (x'_1, \ldots, x'_k) =: x' \in [0, 2\pi]^k$$

either if $x = x'$ or if there are different indices $i_1, \ldots, i_r, j_1, \ldots, j_s \in \{1, \ldots, k\}$, $r, s \in \{1, \ldots, k\}$ with $r + s \leq k$ such that

$$x_{i_1} = \ldots = x_{i_r} = 0, \quad x_{j_1} = \ldots = x_{j_s} = 2\pi, \quad x'_{i_1} = \ldots = x'_{i_r} = 2\pi, \quad x'_{j_1} = \ldots = x'_{j_s} = 0$$

and $x_\ell = x'_\ell$ for all $l \in \{1, \ldots, k\} \backslash \{i_1, \ldots, i_r, j_1, \ldots, j_s\}$. The torus distance for $[x], [x'] \in \mathbb{T}^k$ is defined as

$$d_{\mathbb{T}^k}([x], [x']) := \sqrt{\sum_{j=1}^{k} \Big(\min\{|x_j - x'_j|, 2\pi - |x_j - x'_j|\}\Big)^2} .$$

Usually, subscripts denote coordinate indices. When subscripts denote repeated measurements, coordinate indices, if necessary, will be moved to superscripts in parentheses. Subscripts in parentheses will denote ordered data.

3 Torus PCA

Torus PCA consists of four steps: the torus-to-stratified-sphere (TOSS) mapping, choosing datadriven torus angles, principal nested subsphere analysis for the torus and statistical tests against overfitting. In Eltzner et al. (2018) torus PCA has been introduced from a differential geometric perspective illustrating the deformation undergone by a Riemannian line element. Here we describe the TOSS map more explicitly, give implementation details and some asymptotic theory.

3.1 The Torus-To-Stratified-Sphere (TOSS) Mapping

We start with a description of the torus-to-stratified-sphere (TOSS) mapping in two dimensions and then generalize to higher dimensions. In particular, the stratified sphere is defined below.

3.1.1 TOSS in Two Dimensions

Topologically, a two-dimensional torus can be viewed as a two-dimensional sphere with north and south pole identified, maintaining the torus' periodicity. Geometrically this can be achieved by the price of creating a singularity, preferably far away from data in practice. The following is illustrated in Fig. A1.

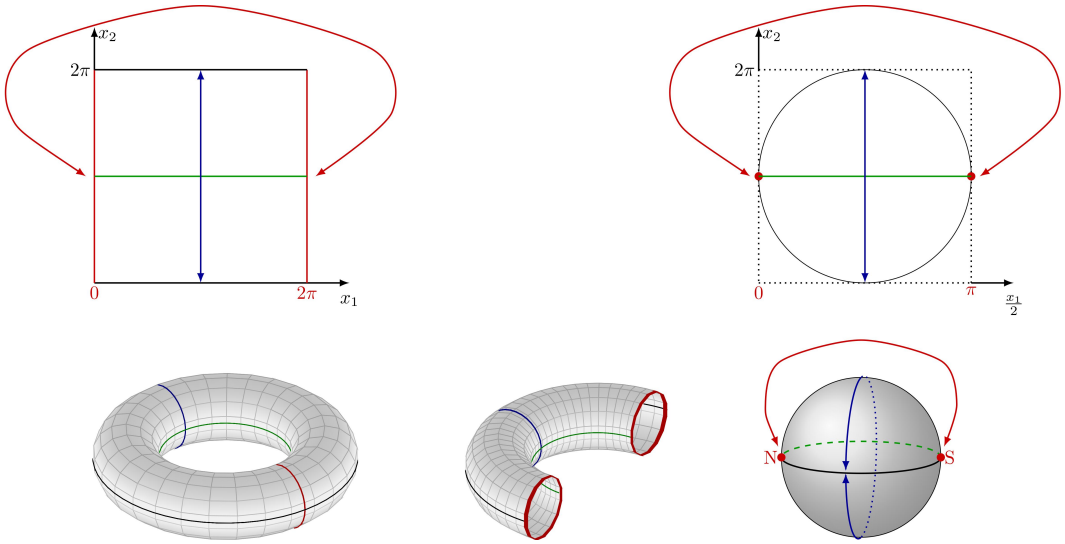


Fig. A1: Top left: fundamental region of a 2D torus with boundary identification given by arrows. Bottom left: a 2D torus (non-isometrically) embedded in 3D Euclidean space. Bottom middle: cutting the torus open at red circles (the lines $x_1 = 0$ and $x_1 = 2\pi$ in top left) one arrives at a topological cylinder. Top and bottom right: collapsing the two red circles to points one arrives at a topological sphere with north (N) and south (S). If the first angle is halved, torus angles naturally become polar angles. Identification of the red parts (formerly $x_1 = 0$ and $x_1 = 2\pi$, now north and south pole) results in a stratification of the sphere (conveyed by red arrows). The black lines (formerly $x_2 = 0$ and $x_2 = 2\pi$, now the front part of the meridian) remain identified in the sphere's topology (conveyed by blue arrows). In order to depict a standard torus in the bottom left panel, on the sphere (bottom right panel), north and south pole are left and right of the sphere.

If the torus is suitably parametrized by angles, and the sphere suitably by polar coor-

dinates, the corresponding TOSS transformation is particularly simple. To this end recall the 2D torus and the 2D sphere

$$\begin{aligned}\mathbb{T}^2 &:= \{(x_1, x_2) \in [0, 2\pi]^2\}/\sim,\\ \mathbb{S}^2 &:= \{(\cos\theta_1,\ \sin\theta_1\cos\theta_2,\ \sin\theta_1\sin\theta_2) : \theta_1 \in [0,\pi], \theta_2 \in [0, 2\pi)\},\end{aligned}$$

respectively. Here $[0, 2\pi]^2 \ni (x_1, x_2) = x \sim x' = (x'_1, x'_2) \in [0, 2\pi]^2$ if either $x = x'$ or if

$$x_i = x'_i \text{ and either } x_j = 0, x'_j = 2\pi \text{ or } x_j = 2\pi, x'_j = 0, \quad \{i, j\} = \{1, 2\}.$$

Obviously the second torus coordinate x_2 can be mapped directly to the longitude θ_2 while, conveniently, the trigonometric functions preserve periodicity. Then the other coordinate x_1 can be halved to be mapped directly to the colatitude θ_1. Since the trigonometric functions are not periodic along a half period only, in order to restore periodicity, north and south pole, corresponding to $\theta_1 = 0$ and $\theta_1 = \pi$, respectively, are identified, giving a stratified sphere $\mathbb{S}^2/\sim$ as follows: for a pair of points $y, y' \in \mathbb{S}^2$ the identification is given by $(y_1, y_2, y_3) = y \sim y' = (y'_1, y'_2, y'_3)$ if either $y = y'$ or $y = (1, 0, 0), y' = (-1, 0, 0)$ or $y = (-1, 0, 0), y' = (1, 0, 0)$ (north and south pole).

Indeed, the sphere with north and south pole removed (the manifold stratum), topologically a cylinder, is diffeomorphic to a torus with a circle removed, where this circle is determined by the first coordinate being zero (or equivalently 2π). However, near the north and south pole (the singular stratum), $\mathbb{S}^2/\sim$ has no manifold structure.

This leads to the following TOSS mapping in two dimensions:

$$\Phi : \mathbb{T}^2 \ \to\ \mathbb{S}^2/\sim, \quad [x_1, x_2] \ \mapsto\ \left[\cos\frac{x_1}{2},\ \sin\frac{x_1}{2}\cos x_2,\ \sin\frac{x_1}{2}\sin x_2\right]. \tag{1}$$

Note that the entire line $x_1 = 0$ is mapped to the north pole $(1, 0, 0)$ of the sphere, irrespective of x_2, and the line $x_1 = 2\pi$ is mapped to the south pole $(-1, 0, 0)$ on the sphere, again, irrespective of x_2. Since both lines are identical on the torus, north and south pole of the sphere are identified (preserving the torus' periodicity), making this space a stratified sphere as it is no longer a manifold at the identified poles. Fig. A1 illustrates the TOSS mapping step by step.

3.1.2 TOSS in Higher Dimensions

Recall the k-dimensional torus $\mathbb{T}^k := [0, 2\pi]^k/\sim$ from Section 2 and the polar representation for the k-dimensional sphere,

$$\mathbb{S}^k := \{y \in \mathbb{R}^{k+1} : \|y\| = 1\},$$

see, for example, (Mardia et al., 1979, p. 35), given by

$$\begin{aligned} y_1 &:= \cos\theta_1, & y_i &:= \left(\prod_{j=1}^{i-1}\sin\theta_j\right)\cos\theta_i,\\ y_k &:= \left(\prod_{j=1}^{k-1}\sin\theta_j\right)\cos\theta_k, & y_{k+1} &:= \left(\prod_{j=1}^{k-1}\sin\theta_j\right)\sin\theta_k \end{aligned} \tag{2}$$

where $\theta_j \in [0, \pi], j \in \{1, \ldots, k-1\}$ and $\theta_k \in [0, 2\pi)$.

As before, the last torus angle x_k can be mapped directly to last polar angle θ_k, and the trigonometric functions preserve periodicity. All of the other torus angles $(x_1, \ldots, x_{k-1})$

can be halved and mapped directly to the other polar coordinates $(\theta_1, \ldots, \theta_{k-1})$. Since the trigonometric functions are not periodic along a half period only, periodicity is restored by identifying

$$\mathbb{S}^k \ni y := (y_1, \ldots, y_{k+1}) \sim (y'_1, \ldots, y'_{k+1}) =: y' \in \mathbb{S}^k\,,$$

if $y = y'$ or if there is an index $1 \leq \ell \leq k-1$ such that

$$y_j = \begin{cases} y'_j\,, & \text{if} \quad 1 \leq j < \ell\,, \\ \pm y'_j\,, & \text{if} \quad j = \ell\,, \\ y'_j = 0\,, & \text{if} \quad \ell < j \leq k+1\,. \end{cases}$$

This leads to the following stratified sphere $\mathbb{S}^k/\sim$ and the TOSS mapping

$$\begin{aligned} &\Phi : \mathbb{T}^k \to \mathbb{S}^k/\sim \qquad (3) \\ &[x_1, \ldots, x_k] \mapsto \left[\cos\frac{x_1}{2}, \ldots, \left(\prod_{j=1}^{i-1} \sin\frac{x_j}{2}\right) \cos\frac{x_i}{2}, \ldots, \right. \\ &\qquad \left. \left(\prod_{j=1}^{k-1} \sin\frac{x_j}{2}\right) \cos x_k, \left(\prod_{j=1}^{k-1} \sin\frac{x_j}{2}\right) \sin x_k \right], \end{aligned}$$

In particular, $\mathbb{S}^k/\sim$ has the manifold part

$$H_k := \{y \in \mathbb{S}^k : y_k^2 + y_{k+1}^2 > 0\}$$

and its singular part is itself stratified comprising the disjoint stratified sets

$$H_\ell/\sim,\ H_\ell := \{(z, 0) \in \mathbb{R}^{k+1} : z \in \mathbb{S}^{\ell+1}, z_{\ell+1} \neq 0\}$$

of dimensions $\ell \in \{0, 1, \ldots, k-2\}$, where $(z, 0) \sim (z', 0)$ if $z = z'$ or $z_j = z'_j$ for $1 \leq j \leq l$ and $z_{\ell+1} = -z'_{\ell+1}$. The singular part is the image of $k-1$ $(k-1)$-dimensional subtori (each collapsing, the first subtorus losing one dimension, the last subtorus losing $k-1$ dimensions sequentially) of $\mathbb{T}^k$ as illustrated in Fig. A2 for dimension $k = 3$.

3.2 Datadriven Torus Angles

Let $X_1, \ldots, X_n$ be given data on $(\mathbb{T}^k, d_{\mathbb{T}^k})$. As detailed above, the TOSS map from $\mathbb{T}^k$ to $\mathbb{S}^k/\sim$ map is singular on $k-1$ subtori of dimension $k-1$. Datadriven torus angles place this singularity set as far away as possible from the data, and they order the angles according to their marginal variances, either increasing or decreasing. Also for the first goal, there are two options.

3.2.1 Mean and gap centering for dislodging data from the singularity set

First, we assume that the sample $X_1, \ldots, X_n$ features a unique sample Fréchet mean $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_k)$. This is the minimizer of the sum of squared torus distances $d_{\mathbb{T}^k}$, it is not the mean direction on every circle. For some concepts of different means see e.g. (Dryden and Mardia, 2016, p. 114-115). Choosing (by translation) torus angles $(x_1, \ldots, x_k)$ such that $\hat{\mu}_j = \pi$, $j \in \{1, \ldots, k\}$, we speak of mean centered (MC) torus angles. From Arnaudon and Miclo (2014) it is known that every sample $X_1, \ldots, X_n$ of a random variable X on $(\mathbb{T}^k, d_{\mathbb{T}^k})$ that has a density with respect to the uniform measure of $\mathbb{T}^k$, has an almost surely unique sample Fréchet mean. Alternatively, choosing torus angles such that every component

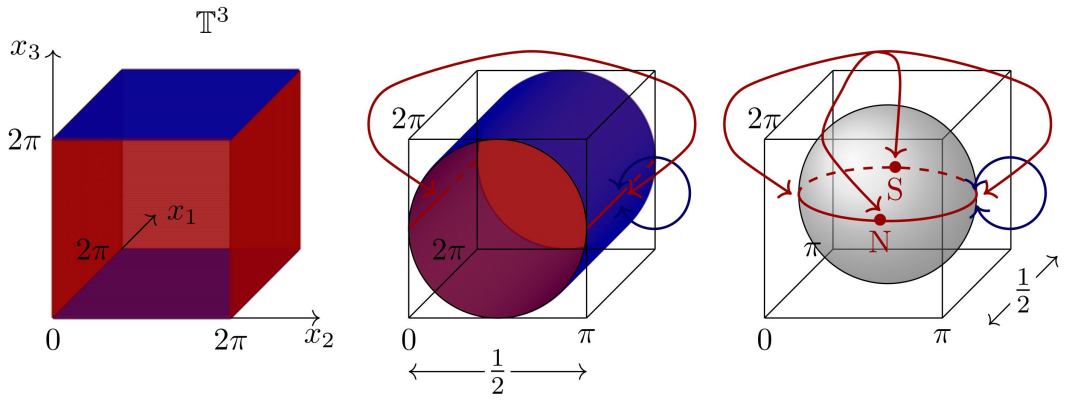


Fig. A2: Illustrating the TOSS map for $k = 3$. Left: fundamental region of a 3D torus which is a solid cube with opposite faces identified. Middle: after collapsing the faces $x_2 = 0$ and $x_2 = 2\pi$ to lines and halving x_2, the torus becomes a solid cylinder with the two bounding disks identified and further boundary identifications depicted by arrows. Right: after collapsing the faces $x_1 = 0$ and $x_1 = 2\pi$ (now to the points N and S) and halving x_1 one obtains a solid ball with upper and lower hemisphere identified (blue arrows, canonically conveyed by $x_3 \in [0, 2\pi]$) yielding a topological 3D sphere. Its stratification is illustrated by red arrows: two opposite open meridians (H_1) are identified (making $H_1/\sim$) and so are north (N) and south pole (S) of the 3D sphere (making $H_0/\sim$).

is gap centered (as detailed below) with respect to each of the angular components of $X_1, \ldots, X_n$, we speak of gap centered (GC) torus angles.

In order to gap center a sample of angles $\phi_1, \ldots, \phi_n \in [0, 2\pi)$ having one single angular component, for $\psi \in [-\pi, \pi)$ let $\phi_1^\psi, \ldots, \phi_n^\psi \in [0, 2\pi)$ be representatives of

$$\phi_1 + \psi \mod 2\pi, \quad \ldots, \quad \phi_n + \psi \mod 2\pi\,.$$

Setting $\phi_{(1)}^\psi := \min\{\phi_1^\psi, \ldots, \phi_n^\psi\}$, $\phi_{(n)}^\psi := \max\{\phi_1^\psi, \ldots, \phi_n^\psi\}$ and $\Phi^\psi := 2\pi - \phi_{(n)}^\psi + \phi_{(1)}^\psi$ (this is the length of the circular gap between $\phi_{(n)}^\psi$ and $\phi_{(1)}^\psi$), consider

$$\psi^* \in \operatorname*{arg\,max}_{\psi \in [-\pi, \pi)} \Phi^\psi\,,$$

which is the length of the largest circular gap of the angles $\phi_1, \ldots, \phi_n$. Then, $\phi_j^{(g)} := \phi_j^{\psi^*} - \phi_{(1)}^{\psi^*} + \Phi^{\psi^*}/2 \in [0, 2\pi)$, $j \in \{1, \ldots, n\}$, are called largest gap centered representatives of the original $\phi_1, \ldots, \phi_n$. Indeed, the smallest of the gap centered representatives is $\phi_{(1)}^{(g)} = \Phi^{\psi^*}/2$ and the largest is $\phi_{(n)}^{(g)} = 2\pi - \Phi^{\psi^*}/2$.

3.2.2 Spread inside and spread outside ordering of torus angles

Recall that $X_1, \ldots, X_n$ are the torus data, the j-th coordinate of X_i will be denoted by $X_i^{(j)}$, $1 \le j \le k$, $1 \le i \le n$. For mean centered data we define $\overline{X} := \hat{\mu}$ whereas for gap centered data we define $\overline{X}^{(j)} := \frac{1}{2}\left((X^{(j)})_{(n)}^\psi + (X^{(j)})_{(1)}^\psi\right)$. Then, for $1 \le j \le k$,

$$\sigma_j := \frac{1}{n}\sum_{i=1}^n \left(X_i^{(j)} - \overline{X}^{(j)}\right)^2$$

is called the *data spread* of the j-th torus angle. We consider two permutations τ_{max} and τ_{min} of the $1, \ldots, k$ determined by

$$\sigma^2_{\tau_{max}(1)} \geq \sigma^2_{\tau_{max}(2)} \geq \cdots \geq \sigma^2_{\tau_{max}(n)}\,, \quad \sigma^2_{\tau_{min}(1)} \leq \sigma^2_{\tau_{min}(2)} \leq \cdots \leq \sigma^2_{\tau_{min}(n)}\,.$$

The case of setting $\tau = \tau_{max}$ will be denoted with spread outside (SO) and of setting $\tau = \tau_{min}$ with spread insided (SI). As there is no ambiguity in the following, the permuted torus angles

$$[x_{\tau(1)}, \ldots, x_{\tau(k)}]$$

will be again denoted by

$$[x_1, \ldots, x_k]\,.$$

Eventually we come up with four possible choices of datadriven torus angles

(MC, SO) or (MC, SI) or (GC, SO) or (GC, SI).

These choices will be important for the MINCE algorithm in Section 4.1.

3.3 Principal Nested Subspheres for the Torus

Before describing in detail, we first give an overview of this section. As principal components on the torus we will use principal nested small spheres under the TOSS map as described below. Principal nested small spheres have been first introduced by Jung et al. (2012). In our setting, these small spheres will live on the manifold part of a stratified sphere. In fact, this will be the case if the first principal small subsphere does not pass through the singular stratum. We will see below that the singular stratum is of codimension 2 on the stratified sphere, which will happen with probability one under realistic conditions. In fact, the space of principal nested subspheres can be equipped with a manifold structure. This manifold structure from Huckemann and Eltzner (2018) is first reviewed. We require that the data are approximated by subspheres with respect to the torus distance. Hence, we use the distance on the stratified sphere as the distance between the data under the TOSS map and a subsphere. Details are given in the second part of this section. Those readers who are not familiar with directional data analysis we want to remind that small subsphere and great subsphere have the same connotation as small circle and great circle.

3.3.1 Nested small subspheres

Again, we first give an overview of the details following: Every linear subspace of $\mathbb{R}^k$ of dimension ℓ is determined by $k-\ell$ linearly independent vectors orthogonal to it and these vectors can be assumed to be orthonormal. In particular, every full rank linear recombination of these vectors yields the same linear space, so the space of ℓ-dimensional linear subspaces of $\mathbb{R}^k$ is a quotient space modulo $(k-\ell)\times(k-\ell)$ orthogonal transformations. Since an affine subspace of dimension ℓ is obtained by translating a linear ℓ-dimensional space along each of the vectors orthogonal to it, the corresponding affine space is determined by an additional vector $\alpha \in \mathbb{R}^{k-\ell}$ of signed distances from the origin along the orthogonal vectors. Finally, every pair of $k-\ell$ orthogonal vectors and $k-\ell$ signed distances yields the same affine subspace if the orthogonal vectors and the distance vector are obtained by suitably applying a common orthogonal transformation to both of them. Since ℓ-dimensional small subspheres can be viewed as intersections of k-dimensional spheres embedded in $\mathbb{R}^{k+1}$ with suitable $(\ell+1)$-dimensional affine subspaces, in the following definitions, some dimensions are increased by one.

Definition 2. The space

$$O(r,s) := \{V = (v_1, \dots, v_r) \in \mathbb{R}^{s\times r} : v_i^T v_j = \delta_{ij}, 1 \le i, j, \le r\}, \quad 1 \le r \le s\,,$$

of matrices with r orthonormal columns with s components is called a *Stiefel manifold*. For $1 \le \ell \le k-1$ denote by

$$A_{V,\alpha} := \{x \in \mathbb{R}^{k+1} : v_j^T x = \alpha_j, 1 \le j \le k-\ell\}$$

the affine subspace of dimension $\ell+1$ in $\mathbb{R}^{k+1}$ determined by $V = (v_1, \dots, v_{k-\ell}) \in O(k-\ell, k+1)$ and the vector $\alpha = (\alpha_1, \dots, \alpha_{k-\ell})^T \in \mathbb{R}^{k-\ell}$. Further, denote by

$$S_{V,\alpha} := A_{V,\alpha} \cap \mathbb{S}^k \tag{4}$$

the ℓ-dimensional small subsphere of $\mathbb{S}^k$ determined by $V \in O(k-\ell, k+1)$ and

$$\alpha \in \mathbb{B}^{k-\ell} := \{x \in \mathbb{R}^{k-\ell} : \|x\| < 1\}\,.$$

Finally, let

$$P_\ell := \big(O(k-\ell, k+1) \times \mathbb{B}^{k-\ell}\big) / \sim$$

where $(V,\alpha) \sim (V',\alpha')$, for $V, V' \in O(k-\ell, k+1)$ and $\alpha, \alpha' \in \mathbb{B}^{k-1}$ if $V' = VR$ and $\alpha' = R^T\alpha$ for some

$$R \in O(k-\ell) := O(k-l, k-\ell)\,.$$

Note that we restrict $\|\alpha\| < 1$ above to ensure that the intersection in (4) is not void or just a point. Moreover, we have

$$S_{V,\alpha} = S_{VR,R^T\alpha}$$

for any $R \in O(k-\ell)$, so that P_ℓ parametrizes all small ℓ-dimensional subspheres of $\mathbb{S}^k$. In the following theorem, we collect more results from (Huckemann and Eltzner, 2018, Appendix A).

Theorem 1. *With the above notation, for $1 \le \ell \le k-1$,*

(i) *P_ℓ is a smooth manifold of dimension $(l+2)(k-\ell)$,*

(ii) *and for $(V,\alpha) \in O(k-\ell, k+1) \times \mathbb{B}^{k-\ell}$, the spherical projection of from $\mathbb{S}^k$ to $S_{V,\alpha}$ is given by*

$$\pi_{S_{V,\alpha}} : \mathbb{S}^k \;\to\; S_{V,\alpha}\,, \quad y \;\mapsto\; V\alpha + \sqrt{1-\|\alpha\|^2}\,\frac{(I_{k+1} - VV^T)y}{\|I_{k+1} - VV^T)y\|}\,,$$

which is the same as first projecting to $S_{V',\alpha'} \supset S_{V,\alpha}$ and then to $S_{V,\alpha}$, where V' comprises some of the columns of V and α' the corresponding elements of α. Here I_{k+1} is the $(k+1)\times(k+1)$ unit matrix.

Remark 1. With the above notation, for $1 \le \ell \le k-1$, $V = (v_1, \dots, v_\ell) \in O(k-\ell, k+1)$ and $\alpha = (\alpha_1, \dots, \alpha_{k-l})^T \in \mathbb{B}^{k-l}$ we have a sequence of nested small subspheres from dimension k to dimension l:

$$\mathbb{S}^k \supset S_{v_1,\alpha_1} \supset S_{(v_1,v_2),(\alpha_1,\alpha_2)^T} \supset S_{(v_1,v_2,v_3),(\alpha_1,\alpha_2,\alpha_3)^T} \supset \cdots \supset S_{V,\alpha}\,. \tag{5}$$

Furthermore, on P_ℓ we have the canonical *extrinsic quotient distance*

$$d_{P_\ell}([V,\alpha],[V',\alpha']) := \min_{R\in O(k-l)} \sqrt{\|RV - RV'\|^2 + \|\alpha - R^T\alpha'\|^2}\,,$$

which is due to restricting the Euclidean distance of $\mathbb{R}^{(k-\ell)\times(k+1)} \times \mathbb{R}^{k-\ell}$ to $O(k-\ell, k+1) \times \mathbb{B}^{k-\ell}$.

The following definition gives the analog of the first Euclidean principal component in principal nested small spheres analysis.

Definition 3 (Main Principal Nested Circle)**.** The *main principal nested circle* is the last element in a sequence of nested nested spheres as in (5), down to dimension $\ell = 1$.

3.3.2 Estimating nested small subspheres with respect to the stratified sphere's distance

For a given combination of torus coordinates from $\{MC, GC\}$ and $\{SO, SI\}$ from Section 3.2, we have the following *distance* $\rho_\ell(S_{V,\alpha}, x)$ between a torus element $x \in \mathbb{T}^k$ and a ℓ-dimensional subsphere $S_{V,\alpha}$, $(\ell \in \{1, \ldots, k-1\}, V \in O(k-\ell, k+1), \alpha \in \mathbb{D}^{k-l})$ of $\mathbb{S}^k$ given by

$$\rho_\ell(S_{V,\alpha}, x) := \min \Bigg\{ \arccos\left(y^T \pi_{S_{V,\alpha}}(y)\right), \qquad (6)$$
$$\min_{d=0,\ldots,k-1} \left(\arccos\left(\sqrt{\sum_{j=1}^{d+1} y_j^2} \right) + \arccos\left((\widetilde{y}^{(d)})^T \pi_{S_{V,\alpha}}(\widetilde{y}^{(d)}) \right) \right) \Bigg\},$$

where $y = \Phi(x)$ with the TOSS map Φ from (3), and

$$y^{(d)} = \frac{(y_1, \ldots, y_{d+1}, 0, \ldots, 0)}{\sqrt{y_1^1 + \cdots + y_{d+1}^2}},$$

is its projection to the stratum $H_d/\sim$, $d \in \{0, \ldots, k-1\}$, as in Theorem 1 (ii), and

$$\widetilde{y}^{(d)} := \frac{(y_1, \ldots, y_\ell, -y_{d+1}, 0, \ldots, 0)}{\sqrt{y_1^1 + \cdots + y_{d+1}^2}}$$

is identified with $y^{(d)}$. Hence the distance is either spherical distance, or the spherical distance to a suitable stratum $H_d/\sim$ and the shorter spherical distance from that stratum.

We now use the concept of generalized Fréchet ρ-means in their sample and population version from Huckemann (2014). Let $1 \le \ell < k$ and define with respect to the above distances the sequence of sample stratified principal nested small spheres for $X_1, \ldots, X_n$ by

$$\left.\begin{array}{rcl}
(\hat{v}_1, \hat{\alpha}_1) & \in & \underset{\substack{v_1 \in \mathbb{S}^k \\ \alpha_1 \in (-1,1)}}{\arg\min} \; \frac{1}{n} \sum_{i=1}^n \rho_{k-1}(S_{v_1,\alpha_1}, X_i), \\
(\hat{v}_2, \hat{\alpha}_2) & \in & \underset{\substack{\hat{v}_1 \perp v_2 \in \mathbb{S}^k \\ \alpha_2 \in \left(-\sqrt{1-\hat{\alpha}_1^2}, \sqrt{1-\hat{\alpha}_1^2}\right)}}{\arg\min} \; \frac{1}{n} \sum_{i=1}^n \rho_{k-2}(S_{(\hat{v}_1,v_2),(\hat{\alpha}_1,\alpha_2)^T}, X_i), \\
 & \vdots & \\
(\hat{v}_{k-\ell}, \hat{\alpha}_{k-\ell}) & \in & \underset{\substack{\hat{v}_1, \ldots, \hat{v}_{k-\ell-1} \perp v_{k-\ell} \in \mathbb{S}^k \\ \alpha_{k-\ell} \in (-\sqrt{1-\|\hat{\alpha}'\|^2}, \sqrt{1-\|\hat{\alpha}'\|^2})}}{\arg\min} \; \frac{1}{n} \sum_{i=1}^n \rho_{k-\ell}(S_{(\hat{V}', v_{k-\ell}),((\hat{\alpha}')^T, \alpha_{k-\ell})^T}, X_i),
\end{array}\right\} \qquad (7)$$

where $\hat{V}' = (\hat{v}_1, \ldots, \hat{v}_{k-\ell-1})$ and $\hat{\alpha}' = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{k-\ell-1})^T$ in the last expression. Similarly

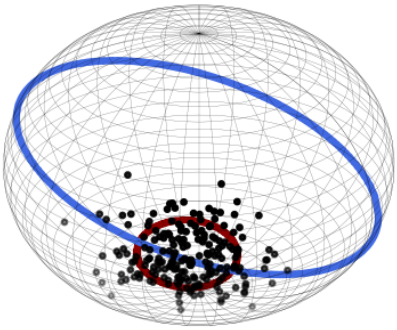


Fig. A3: Depicting a spherical sample represented in black dots from a distribution with best fitting population great subsphere. It is best fitted among small subspheres in red of a smaller size. Its best sample great subsphere fit in blue which is not rejected by the likelihood ratio test given by (12). This prevents overfitting.

define the sequence of population stratified principal nested small spheres for X by

$$\left.\begin{aligned}
(v_1^*, \alpha_1^*) &\in \underset{\substack{v_1 \in \mathbb{S}^k \\ \alpha_1 \in (-1,1)}}{\arg\min}\, \mathbb{E}\left[\rho_{k-1}(S_{v_1,\alpha_1}, X)\right] , \\
(v_2^*, \alpha_2^*) &\in \underset{\substack{v_1^* \perp v_2 \in \mathbb{S}^k \\ \alpha_2 \in (-\sqrt{1-(\alpha_1^*)^2}, \sqrt{1-(\alpha_1^*)^2})}}{\arg\min}\, \mathbb{E}\left[\rho_{k-2}(S_{(v_1^*,v_2),(\alpha_1^*,\alpha_2)^T}, X)\right] , \\
&\vdots \\
(v_{k-\ell}^*, \alpha_{k-\ell}^*) &\in \underset{\substack{v_1^*,\ldots,v_{k-\ell-1}^* \perp v_{k-\ell} \in \mathbb{S}^k \\ \alpha_{k-\ell} \in (-\sqrt{1-\|\alpha'\|^2}, \sqrt{1-\|\alpha'^*\|^2})}}{\arg\min}\, \mathbb{E}\left[\rho_{k-\ell}(S_{(V^{*\prime},v_{k-\ell}),((\alpha^{*\prime})^T,\alpha_{k-\ell})^T}, X)\right] ,
\end{aligned}\right\} \tag{8}$$

where $V^{*\prime} = (v_1^*, \ldots, v_{k-\ell-1}^*)$ and $\alpha^{*\prime} = (\alpha_1^*, \ldots, \alpha_{k-\ell-1}^*)^T$ in the last expression.

If the argminima above are unique up to the action of the corresponding orthogonal groups we speak of unique sequences. Indeed, then the above subspheres are uniquely determined.

3.4 Preventing Overfitting

Suppose that for a random variable X on $\mathbb{S}^k$ a great subsphere yields its best approximating population great and small subsphere. For a sample of X, however, usually a proper (not a great subsphere) small subsphere will yield its best approximating subsphere. In fact, in realstic scenarios such a best sample small subsphere may have a rather small size, see Fig. A3. Clearly, this is overfitting and in order to avoid this, in every step, every sample small subsphere fit is compared to a sample great subsphere fit, both with respect to the sample fitted to the preceding subsphere.

Without loss of generality we may assume that the preceding "best" subspere sample fit is a great sphere $\mathbb{S}^\ell \subseteq \mathbb{S}^k$, $2 \leq \ell \leq k$ and that the data is already projected to it, i.e., $X_1, \ldots, X_n \in \mathbb{S}^\ell$. If it was a proper small subsphere, move its center to the origin and rescale it to unit radius. Further assume that S_1 and S_2 are the best fitting great subsphere and small subsphere of dimension $\ell - 1$, respectively, to the sample in $\mathbb{S}^\ell$. Since S_2 is a proper small subsphere, it has a unique center denoted by $p \in \mathbb{S}^\ell$ and it is of positive radius less than π. We approximate the marginal distribution of the data radii r from $p \in \mathbb{S}^\ell$ by

the model

$$q \mapsto r := \arccos(p^T q) \mapsto C(\rho,\sigma)\, f(r;\rho,\sigma)\, \sin^{\ell-1}(r),\ q \in \mathbb{S}^l,\ r \in [0,\pi],\ \rho,\sigma > 0\,, \quad (9)$$

with parameters ρ and σ, a suitable integration constant $C(\rho,\sigma)$ making the above a density on $r \in [0,\pi]$ and the folded Gaussian

$$\begin{aligned} f(r;\rho,\sigma) &:= \frac{1}{\sqrt{2\pi}\sigma}\left(\exp\left(-\frac{(r-\rho\sigma)^2}{2\sigma^2}\right) + \exp\left(-\frac{(r+\rho\sigma)^2}{2\sigma^2}\right)\right) \\ &= \frac{2}{\sqrt{2\pi}\sigma}\exp\left(-\frac{r^2}{2\sigma^2} - \frac{\rho^2}{2}\right)\cosh\left(\frac{r\rho}{\sigma}\right). \end{aligned}$$

Let us briefly explain the role of $\sin^{\ell-1} r$ in (9). It is the size of the $(\ell-1)$-dimensional small sphere at distance r from the center p divided by the size of the ℓ-dimensional unit sphere. As detailed in Eltzner et al. (2018), thus taking the change of volume into account improves the test from Jung et al. (2012). Numerical experiments by Tsagris et al. (2014) indicate that the function $r \mapsto f(r;\rho,\sigma)$ has a single node located at $r = 0$ for $0 \leq \rho \leq 1$ and located at some $r > 0$ for $\rho > 1$. Hence, in order to test the two hypothesis

$$H_0 : \rho = 1 \text{ (great subsphere) vs. } H_1 : \rho > 1 \text{ (small subsphere)}\,, \quad (10)$$

after estimation of p by $\hat{p}$, resulting in $\hat{r}_i := \arccos \hat{p}^T X_i$, with the log likelihoods, up to a constant, given by

$$\begin{aligned} \ell(\rho,\sigma|\{\hat{r}_i\}_{i=1}^n) = &- n\ln C(\rho,\sigma) + (d-1)\sum_{i=1}^n \ln\sin(\hat{r}_i) - \frac{n\rho^2}{2} - n\ln(\sigma) \\ &+ \sum_{i=1}^n\left(-\frac{\hat{r}_i^2}{2\sigma^2} + \ln\cosh\left(\frac{\hat{r}_i\rho}{\sigma}\right)\right), \end{aligned}$$

we have the minus two times log likelihood ratio

$$\lambda = 2\sup\{\ell(\rho,\sigma|\{r_i\}_{i=1}^n) : \rho \in (1,\infty),\, \sigma \in \mathbb{R}^+\} - 2\sup\{\ell(\rho,\sigma|\{r_i\}_{i=1}^n) : \rho = 1,\, \sigma \in \mathbb{R}^+\}\,. \quad (11)$$

This is, under H_0, due to Wilks' theorem (e.g. (Mardia et al., 1979, Theorem 5.2.1)), asymptotically distributed as χ^2_1. We usually use a 5% significance level for our test, which means that

$$H_0 \text{ is rejected if } \lambda > \chi^2_{1,0.95} \approx 3.84 \quad (12)$$

and we reject the great subsphere fit and we keep the small subsphere fit; otherwise, we keep the great subsphere fit. The normalizing constant $C(\rho,\sigma)$ and the MLEs for ρ and σ in (11) are obtained using standard numerical optimization.

3.5 Asymptotics of Torus PCA

In Huckemann and Eltzner (2018) asymptotic results for principal nested spheres on spheres have been derived. The following similar results hold also for torus PCA, as we prove below. Taking $\ell = 1$ below yields the corresponding assertions on the asymptotics of the main principal circle.

Theorem 2 (Asymptotic Consistency)**.** *Let* $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X$ *be random variables on* $\mathbb{T}^k$ *and* $1 \leq \ell < k$ *such that the population sequence from (8),*

$$S_{v_1^*,\alpha_1^*} \supset S_{(v_1^*,v_2^*),(\alpha_1^*,\alpha_2^*)^T} \supset S_{(v_1^*,v_2^*,v_3^*),(\alpha_1^*,\alpha_2^*,\alpha_3^*)^T} \supset \cdots \supset S_{V^*,\alpha^*}$$

determined by $V^* = (v_1^*, \ldots, v_{k-\ell}^*) \in O(k-\ell, k+1), \alpha^* = (\alpha_1^*, \ldots, \alpha_{k-l}^*)^T \in \mathbb{B}^{k-\ell}$ *is unique. Then, considering any measurable selection of sample sequences from (7)*

$$S_{\hat{v}_1(n),\hat{\alpha}_1(n)} \supset S_{(\hat{v}_1(n),\hat{v}_2(n)),(\hat{\alpha}_1(n),\hat{\alpha}_2(n))^T} \supset \cdots \supset S_{\hat{V}(n),\hat{\alpha}(n)}$$

determined by

$$\hat{V}(n) = \big(\hat{v}_1(n), \ldots, \hat{v}_{k-\ell}(n)\big) \in O(k-\ell, k+1) \quad \textit{and} \quad \hat{\alpha}(n) = \big(\hat{\alpha}_1(n), \ldots, \hat{\alpha}_{k-\ell}(n)\big)^T \in \mathbb{B}^{k-\ell}$$

we have that

$$S_{\hat{V}(n),\hat{\alpha}(n)} \textit{ converges almost surely to } S_{V^*,\alpha^*} \textit{ as } n \to \infty\,.$$

Proof. Since $\mathbb{T}^k$ is compact, and so is every P_ℓ, and since every ϕ_ℓ is continuous, the assumptions for (Huckemann and Eltzner, 2018, Theorem 4.1) hold, yielding the assertion. □

Remark 2 (Asymptotic Normality)**.** With the notation and assumptions of the above Theorem 1, under additional technical assumptions from (Huckemann and Eltzner, 2018, Assumption 3.10), with a local chart $f : U \to \mathbb{R}^{(\ell+1)(k-\ell)}$ near $[V^*, \alpha^*] \in U$ open $\subseteq P_\ell$ with $f([V^*, \alpha^*]) = 0$ and a suitable symmetric positive definite matrix Σ, as $n \to \infty$,

$$\sqrt{n} f([\hat{V}(n), \hat{\alpha}(n)]) \quad \overset{\mathcal{D}}{\to} \quad \mathcal{N}(0, \Sigma)\,.$$

If the above additional assumptions are not met, slower rates may be possible as discussed in Eltzner and Huckemann (2019).

4 Mode Hunting on the Main Principal Nested Circle Post Adaptive Clustering (MINCE post AGE)

We give a new method of clustering for manifold data which involves pre-clustering by an adaptive clustering procedure followed by a refinement which uses the main nested principal circle (component). Indeed, this provides a new powerful tool to cluster Euclidean data. As detailed in Section 1, the main principal nested circle offers more degrees of freedom than the main Euclidean principal component. Consequently, PNS allows for a simple and rather powerful dimension reduction method, not only on spheres and tori but also on general manifolds including Euclidean space. Recall from Section 1, that PNS is parsimonious as it also allows for curved main principal components (circles) that in Euclidean PCA require two principal components. We illustrate our method for three data sets: data on a torus, data on a shape space and Euclidean "worms" data. The first two examples are taken from the SARS-CoV-2 RNA backbone data from the protein data bank and for the last example we obtain our benchmark data from the *worms* simulation software (Sieranoja and Fränti (2019)). First, let us detail the MINCE post AGE method Mode huntIng on the main principle Nested small Circle of prE-clusters post Adaptive linkaGe clustEring).

4.1 Overview of the MINCE post AGE method

The MINCE post AGE method clusters a data set on an original Euclidean or non-Euclidean space by proceeding along the following steps:

1. Pre cluster the data set with a given method of choice, for instance with AGE as detailed in Section 4.2 below.

2. Map every pre cluster to a sphere or a stratified sphere, depending on the topology or geometry of the original non-Euclidean space. For torus angles, use a combination of $\{MC, GC\} \times \{SI, SO\}$ from Section 3.2. If the original space is

   - a torus, Section 3.1 above details the torus-to-stratified-sphere (TOSS) map,
   - a shape space, Section 4.4 below details how to map the data to a horizontal sphere,
   - a Euclidean space, Section 4.5 below details how to map onto a sphere.

3. Subject each pre-cluster mapped to a sphere to principal nested spheres (Section 3.3) and compute its main principal small circle (the last principal nested sphere)

4. Subject the pre-cluster data projected to the main principal small circle to circular mode hunting as detailed below.

5. Assign every mode found with statistical significance as detailed below a post cluster and return all these post clusters as the final clustering result. For original torus data, if only one mode has been found for this pre-cluster and other combination of $\{MC, GC\} \times \{SI, SO\}$ for torus angles in Step 2 have not been tried, go to Step 2 and try another combination.

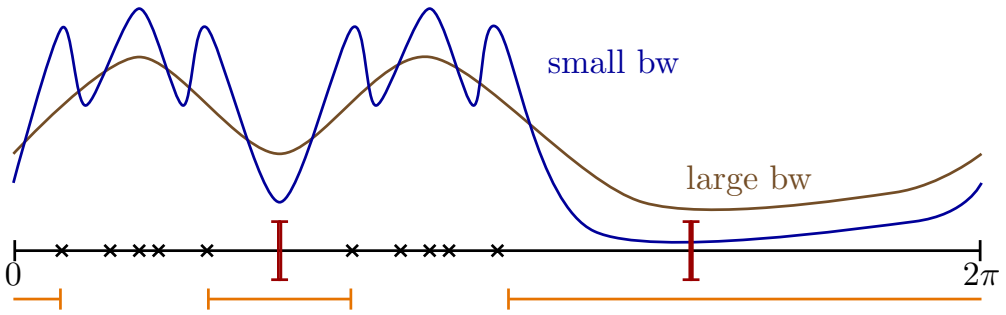


Fig. A4: Adapted from Wiechers et al. (2021): As the circular version of the test from Dümbgen and Walther (2008) for the number of modes for the data (black asterisks) yields a significant number of 2 (intervals containing an antimode with statistical significance are shown in orange), density smoothing with a wrapped normal is performed with varying bandwidths. While the blue density (small bandwidth) features 6 modes, the brown density (low bandwidth) features the right number of 2 modes. For the two modes' case the minima serve as cluster boundaries (red verticals).

4.1.1 Post clustering based on circular mode hunting

Suppose that $0 \leq X_{(1)} < \cdots < X_{(n)} < 2\pi$ are the ordered data projected to the last (the one-dimensional) nested principal nested small sphere component, which is, after

re-scaling, the circle $\mathbb{S}^1 = [0, 2\pi)/\sim$. Recall that this is the main component. Let

$$X_{i,j,k} := \begin{cases} \frac{X_{(i)}-X_{(j)}}{X_{(k)}-X_{(j)}}, & \text{if } 1 \le j \le i \le k \le n\,, \\ \frac{X_{(i)}-X_{(j)}}{2\pi+X_{(k)}-X_{(j)}}, & \text{if } k < j \le i \le n\,, \\ \frac{2\pi+X_{(i)}-X_{(j)}}{2\pi+X_{(k)}-X_{(j)}}, & \text{if } 1 \le i \le k < j\,, \end{cases}$$

be slopes normalized to lie in $[0, 1]$. Then, for any pair $(j, k) \in \{1, \ldots, n\}^2$ with $j \neq k$, define

$$T_{j,k} := \begin{cases} \sum_{i=j+1}^{k-1} (2X_{i,j,k} - 1)\ , \text{ if } j < k\,, \\ \sum_{i=j+1}^{n} (2X_{i,j,k} - 1) + \sum_{i=1}^{k-1} (2X_{i,j,k} - 1)\ , \text{ if } k < j\,, \end{cases}$$

where undefined sums are zero.

We can reasonably assume that the projected angular data has come from a circular distribution with a density. In that case the sign of $T_{j,k}$ indicates whether this density is increasing or decreasing between (in the circular sense) $X_{(j)}$ and $X_{(k)}$. An increase followed by a decrease indicates a mode of the density and a decrease followed by an increase an antimode. This gives rise to a test simultaneously considering all distinct pairs (j, k) as above, inferring an increasing or decreasing density between $X_{(j)}$ and $X_{(k)}$, if $T_{j,k} > c_{j,k}(\alpha)$ or $T_{j,k} < -c_{j,k}(\alpha)$, respectively, with statistical significance $\alpha \in [0, 1]$. For the non-circular linear case, i.e. inferring about a density on $\mathbb{R}$, defining $T_{j,k}$ in the obvious non-circular way, optimal $c_{j,k}(\alpha)$ have been determined and simulated by Dümbgen and Walther (2008). In MINCE we adapt their method and estimate the statistically significant number of modes (each corresponding to a cluster) based on Monte Carlo simulations for the circular uniform distribution. Then, antimodes (their number equals the number of modes on the circle) serve as cluster boundaries and these are located using the *WiZer* software from Huckemann et al. (2016) which is the circular analog of the *SiZer* from Chaudhuri and Marron (1999, 2000). Due to causality of kernel smoothing (the number of modes is nonincreasing with increasing bandwidth) of data with the normal distribution on the line (see Lindeberg (2011)) and the wrapped normal on the circle (see Huckemann et al. (2016)), respectively, the number of modes and antimodes is non-increasing with bandwidth, so that the smoothed density will have the statistically significant number of modes found above for some bandwidth interval. Then, post cluster boundaries will be set to local minimal loci of the wrapped normal (see Section 5) smoothed densities with the middle bandwidth as illustrated in Fig. A4.

4.2 Overview of the AGE Method

From the abundance of clustering methods for data $X_1, \ldots, X_n$ in a metric space (Q, d), for clustering directional data see e.g. Pewsey and García-Portugués (2021), we discuss here briefly the frequent issue of detecting clusters with varying densities as depicted in left panel of Fig. A5 from a hierarchical linkage tree clustering viewpoint.

Recall that in hierarchical linkage tree clustering, see e.g. (Mardia et al., 1979, pp. 369–375), initially each data point is its own cluster, defining the n leaves of the cluster tree, eventually obtained. As long as there are at least two elements in the running cluster list, the clusters with the smallest distance are removed from the running cluster list and the union of these two is added to the running cluster list. The cluster tree is expanded by creating a parent node over the two nodes that correspond to the merged clusters and adding branches that connect them to the parent node. The value indicating the distance between the two merged clusters is assigned to the node. The iteration terminates when

the running list contains only one cluster that includes all the data; this forms the root of the tree.

For the distance of two clusters A and B typically average linkage, also known as unweighted pair group method with arithmetic mean, developed by Sokal and Michener (1958) and single linkage, developed by Florek et al. (1951), are often used:

$$d_a(A,B) := \frac{1}{|A|\cdot|B|}\cdot\sum_{X\in A}\sum_{Y\in B} d(X,Y)\,, \tag{13}$$

$$d_s(A,B) := \min_{X\in A, Y\in B} d(X,Y)\,, \tag{14}$$

respectively. Importantly, the node values increase monotonically for both linkage methods, for if clusters A and B are merged to form the cluster $A\cup B$ and C is another cluster, the distance between A and B is smaller than between A and C, and B and C, respectively, and hence we have

$$d_a(A\cup B,C) = \frac{1}{|A\cup B|\cdot|C|}\left(|A|\cdot|C|\cdot d_a(A,C) + |B|\cdot|C|\cdot d(B,C)\right) \geq d_a(A,B)\,,$$

$$d_s(A\cup B,C) = \min_{X\in A\cup B, Y\in C} d(X,Y) = \min\left\{d_s(A,C), d_s(B,C)\right\} \geq d_s(A,B).$$

Cutting the tree at a certain height (distance value) c and taking the running cluster list where the last node with distance value smaller than c was added to the cluster tree, a specific clustering is obtained. If clusters feature different densities, however, as in the left panel of Fig. A5, less dense clusters may be found at the price of joining nearby denser clusters or the latter are discriminated at the price of missing less dense clusters. This can be avoided by allowing varying densities in iterative clustering as proposed by Wiechers et al. (2021), see also Langfelder et al. (2007) and Obulkasim et al. (2015).

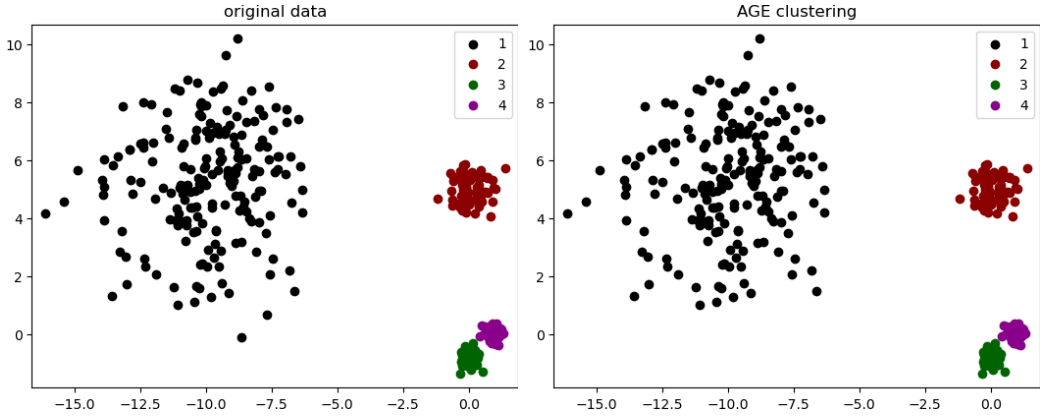


Fig. A5: Left: original data set featuring four clusters, all of which cannot be separated by single or average linkage clustering as the first features a large spread, the other three are very dense, two of them nearby. Right: all of the clusters are retrieved by AGE except for two outliers from Cluster 1 (not shown in the right panel).

Algorithm 1 (Adaptive Linkage Clustering (AGE) by Wiechers et al. (2021))**.** Inputs are

- the three tuning parameters $(d_{\max}, \kappa, q)$:

  $d_{\max}$ = *maximal outlier distance*, controlling cluster density,

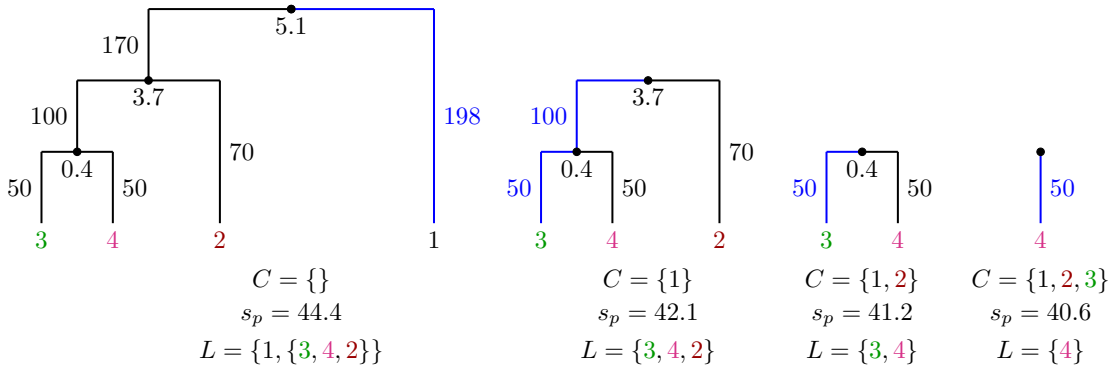


Fig. A6: Illustrating four iterations of the AGE Algorithm 1 applied to the data of Fig. A5 (with cluster colors taken from there) with node labels indicating distance of clusters joined and branch labels indicating the number of data points represented. For each iterate the initial cluster list C, the current minimal cluster size s_P and the resulting auxiliary list L (before Step 8) are shown below. Further details are given in the text.

> κ = *minimal cluster size*, ensuring that mode hunting from Section 4.1, later applied, has sufficient power to separate one-dimensional pre-clusters (for all of the applications below we use $\kappa = 40$),
>
> q = *relative branching distance*, ensuring that two clusters are only split if their parent node's distance value is significant in relation to the greatest distance value of its child nodes (for the applications below it is chosen by the following rule of thumb: In case of single linkage, it is chosen smaller than all $d/r - 1$ where r is within cluster neighboring point distance and d is distance to neighboring clusters) in case of average linkage, r is cluster radius;

- and n data points $P = \{X_1, \ldots, X_n\}$.

Let R be the outlier list and C be the cluster list, each of which are initially empty. They are filled iteratively as follows.

1. Compute the (single, average or other) linkage clustering tree from P.
2. Perform a tree cut at distance $d_{\max}$ to obtain a clustering, move from P to R all data points that are in clusters with less than κ data points.
3. Compute the linkage cluster tree for the new P as in Step 1.
4. Set $s_P = \sqrt{|P| + \kappa^2}$ (inspired by the square root rule of thumb used in histogram binning).
5. Create an empty list L of clusters.
6. Begin at the root and always follow the branch with more points at each node. From each node add the child node corresponding to the smaller subcluster to L
   (a) if it contains more than s_P data points,
   (b) and if the q-fold of its parent node's distance value is greater than the two children nodes' distance values.
7. At the last node, where the smaller subcluster is added to L, also add the larger subcluster to L

8. Consider L:
   - if L is empty, move the union of all data points from P to C; these correspond then to one single cluster,
   - else, add the largest cluster in L to C and remove its points from P.
9. If $|P| > 0$, go to Step 1.
10. Return the clusters list C and the outlier list R.

The four panels of Fig. A6, from left to right, illustrate the iteration in AGE with single linkage and the tuning parameters $q = 0$ (reflecting that the distance of neighbors within the small clusters is of the same order as the distance between the two small clusters) and $d_{\max}$ chosen such that at most 1% of the suites in the single linkage tree are in a branch with less than $\kappa + 1$ data points. Indeed, AGE correctly retrieves all clusters at the price of 2 outliers from 370 original data points, see right panel of Fig. A5.

Iteration 1: In the first iteration, C is initially empty and 2 points from the true Cluster 1 are removed from P as outliers. At the top node the smaller branch comprising true Clusters 2, 3 and 4 is added to L as a single cluster and following the larger branch, the iterate terminates and adds the remains of Cluster 1 to L, which, as being largest is hence moved to C.

Iteration 2: Thus, in the second iterate, C contains Cluster 1 which has been removed from P. At its top node, now the branch comprising Clusters 3 and 4 is largest. Hence, the smaller branch, comprising only Cluster 1 is added to L. Then, in the larger branch, one of the equally large Clusters 3 and 4 is considered smaller, this is added to L and finally, as the iterate terminates, also the other cluster is added to L. Finally the largest cluster in L, which is Cluster 2, is moved to C.

Iteration 3: In the third iteration, only the equally large Clusters 3 and 4 are left in P, first one is added to L, then the other one, is then moved to C.

Last iteration: In the last iteration, only Cluster 4 is left, no nodes are left, so Cluster 4 as the only cluster left in P is then moved to L and then to C. Then, P is empty and AGE terminates.

4.3 Example: MINCE post AGE on a Torus

Since the exact geometric configuration of a biomolecule is decisive for its function, it is of high interest to correctly reconstruct this structure, see Schlick and Pyle (2017). This is usually done with elaborate methods assessing electron density from which with inverse methods the locations of atom nuclei are inferred Jain et al. (2015). As this method is expensive and time consuming, other methods learn how to predict geometry – also called second order and higher order structure – from the sequences of nucleic bases (RNA structure) or amino acids (protein structure) – also called primary structure, see, e.g., Schlick and Pyle (2017).

Typically such learning methods rely on classification and clustering of geometric structure Frellsen et al. (2009); Olsson et al. (2011) and in the following application we use representations on the torus $\mathbb{T}^7$ comprising the seven dihedral angles of RNA suite structure, detailed in Fig. A7.

From the protein data base (PDB), Berman et al. (2000), we have selected (dated 4/21/2021) 13,439 SARS-CoV-2 and related structures containing atom positions and

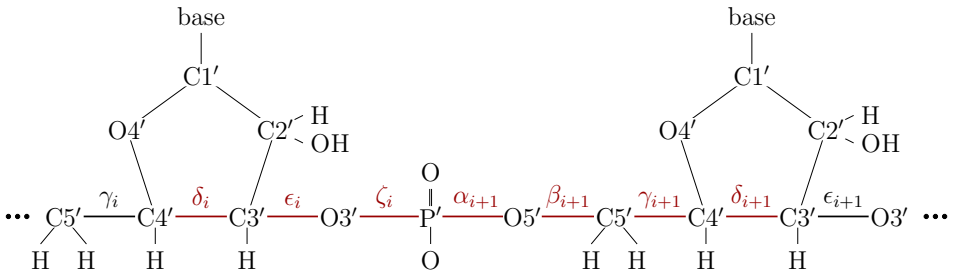


Fig. A7: Adapted from Wiechers et al. (2021): Schematic structure of suite (from one sugar ring which is attached to a nucleic base to the next sugar ring) number i within an RNA backbone with seven dihedral angles (red) yielding the suite's geometric shape.

mesoscopics (detailed in Section 4.4) compiled in a git repository by the Coronavirus Structural Task Force (CSTF) headed by Andrea Thorn. Fig. A8 illustrates this data set by a scatterplot relating pairs of dihedral angles with one another (for a discussion on dihedral angles in biominformatics see e.g. Mardia (2013)). Notably, such data may contain clashes, in particular reconstructed hydrogen nuclei, which are not found by electron density as their electrons usually move to bonded neighboring nuclei, may collide with other atoms' nuclei reconstructions. Since clashes of atom positions are physically not possible, clash correction, i.e. clash free reconstruction of molecular geometry is one major challenge, see e.g. Murray et al. (2003).

We have applied the MINCE post AGE method of Section 4.1. For AGE, Algorithm 1, we choose tuning parameters $q = 3/20$ (taking into account that, due to chemical constraints, cluster centers are often highly concentrated and should thus not be separated from their less dense neighborhood) and $d_{\max}$ such that at most 25% of the suites in the average linkage tree are in a branch with less than $\kappa + 1$ data points (for this and all applications here, we use $\kappa = 40$). This leads to 21 clusters. Their sizes and the size of the outlier set are listed in Table A1. The clusters' pairwise dihedral angle scatterplots are depicted in Fig. A9.

In Wiechers et al. (2021) we have analyzed a similar data set, with higher resolution and removed clashes using this methods. However, the current data set is of lower quality and contains an unknown number of clashing suites (suite configurations that are chemically not possible and thus cannot be assigned to clusters). For this reason, the choice of $d_{\max}$ anticipates a high but realistic number of outliers. Still most of the clusters from Table A1 roughly correspond to non clashing suite clusters found in Wiechers et al. (2021), this is in particular the case for the first two clusters.

Table A1: Clusters and outliers (numbers and sizes, the relative amount of outliers 24.4 % is governed by the choice of $d_{\max}$) found by MINCE post AGE, as depicted in Figre A9, in SARS-CoV-2 and related suite structures from the *protein data base* (PDB), Berman et al. (2000).

Cluster	1	2	3	4	5	6	7	8	9	10	11	12
Size	7206	687	245	234	222	201	193	157	115	109	98	91
Cluster	13	14	15	16	17	18	19	20	21	R	$\sum$	
Size	81	77	64	64	58	54	50	48	44	3281	13439	

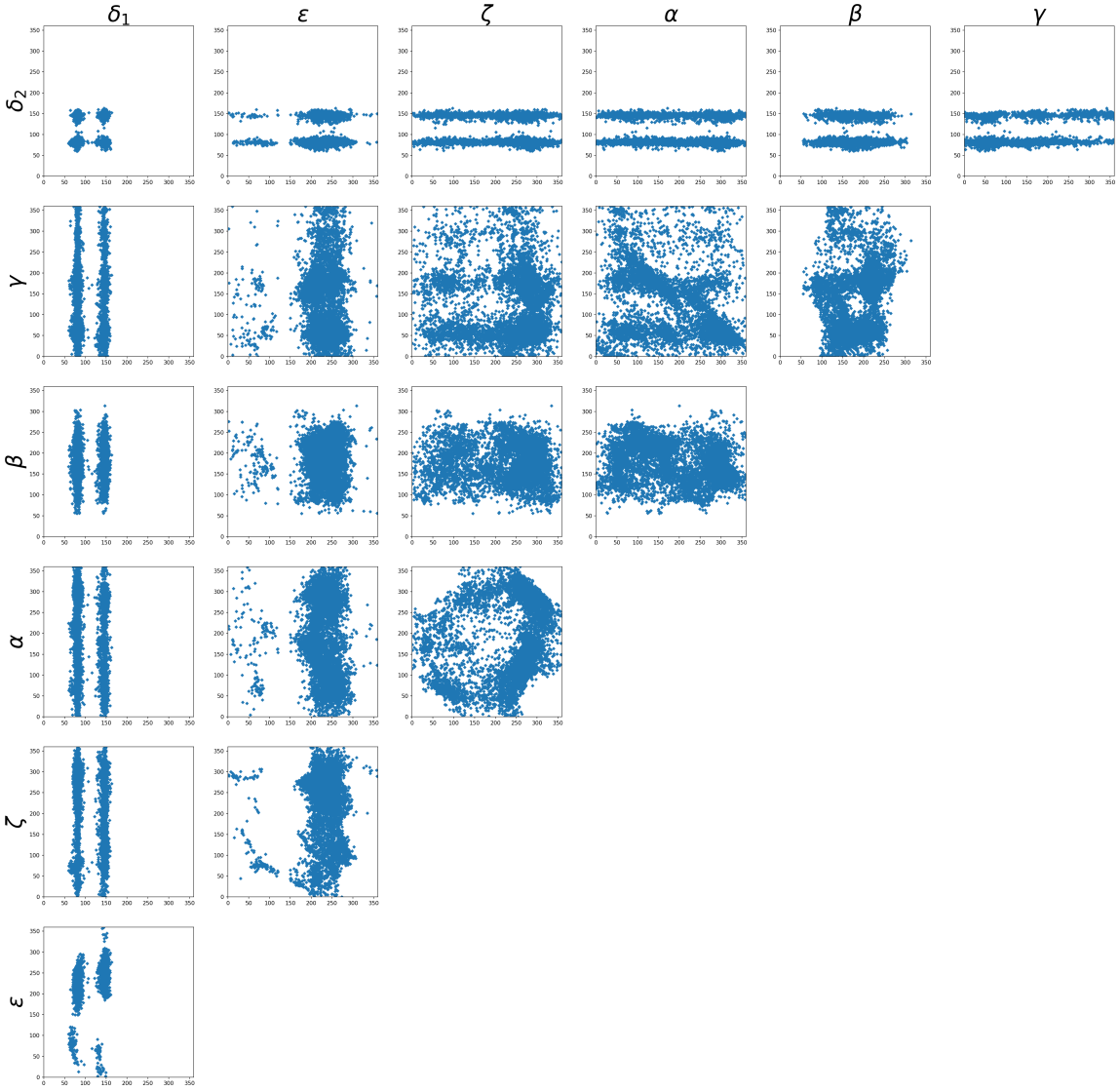


Fig. A8: Scatterplots pairwise relating all seven dihedral suite angles of 13,439 SARS-CoV-2 and related structures from the *protein data base* (PDB), Berman et al. (2000).

4.4 Example: MINCE post AGE on a Shape Space

Usually, biomolecules come as long strands, along a backbone of repetitive structures, see Fig. A7. In RNA such a repetitive structure is a sugar ring (called ribose) forming a suite's boundary and mesoscopics are shapes of landmark configurations where landmarks are placed at the centers of k subsequent sugar rings, cf. Fig. A10. Here, as in Wiechers et al. (2021) we consider $k = 6$, roughly corresponding to a half helix turn in helical configurations. Combining suite clustering with mesoscopic clustering is a powerful tool in RNA structure classification based on statistical learning as detailed in Wiechers et al. (2021). Here we cluster the mesoscopics of the SARS-CoV-2 data set described in the previous Section 4.3. To this end, each of the 13,439 mesoscopics is represented in the shape space Σ_3^6 of three-dimensional six-landmark configuration, see Dryden and Mardia (2016), which, as a metric space, allows for pre-clustering with AGE, Algorithm 1, with tuning parameters $q = 0.0005$ (reflecting that some within cluster distances are almost of the same order as their distances to neighboring clusters) and $d_{\max}$ such that 40% of

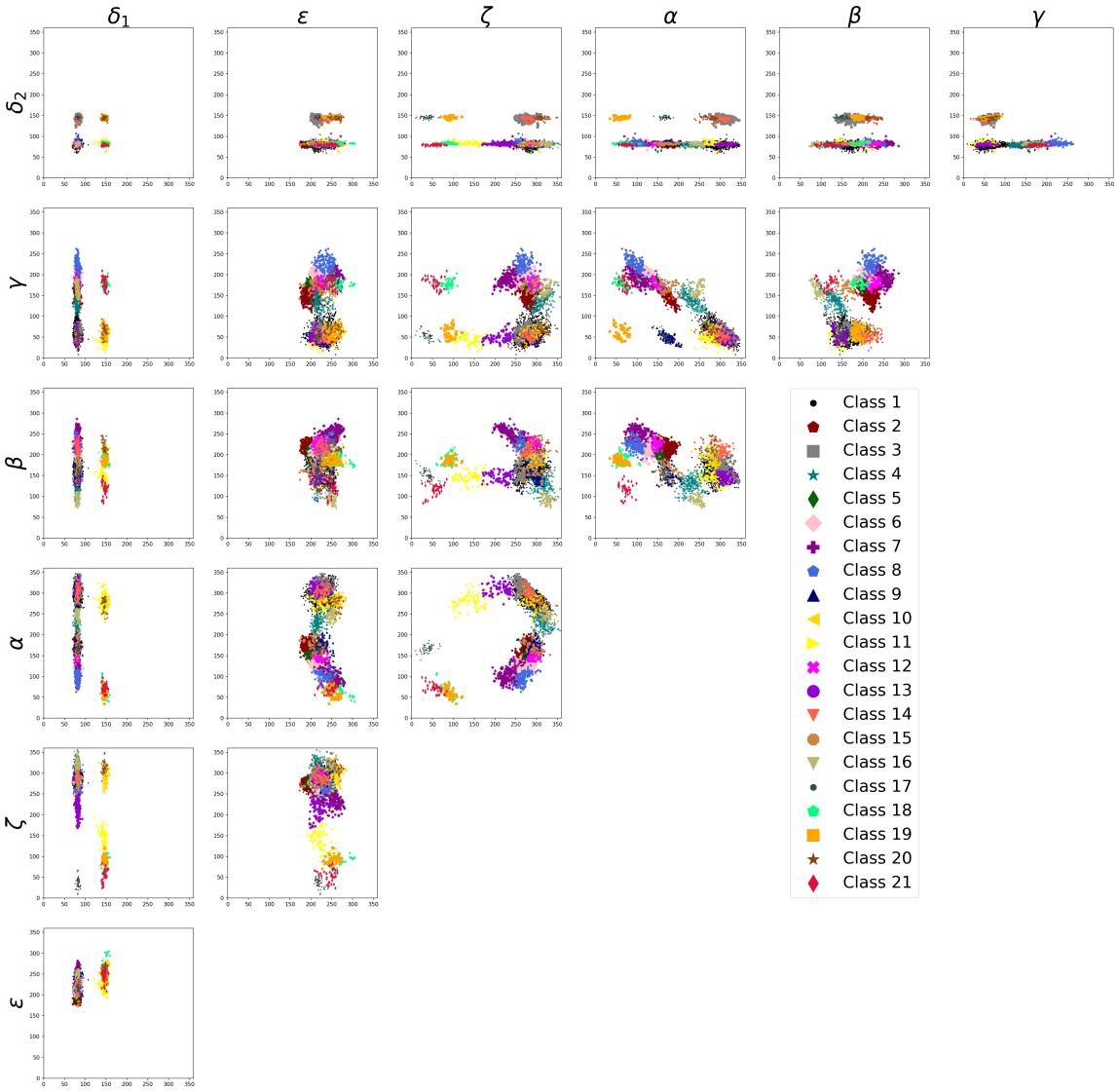


Fig. A9: Scatterplots of data in Fig. A8, now depicting the 21 clusters (Table A1) found by MINCE post AGE.

the mesosocopic shapes in the average linkage tree are in a branch with less than $\kappa + 1$ data points (for this and all applications here, we use $\kappa = 40$). In addition to the remarks on data quality in the previous section, the mesoscopic shape data is geometrically very variable and contains a high number of small dispersed clusters. Discriminating those from outliers requires higher sample sizes than used here, see e.g. Jain et al. (2015), so in order not to misclassify parts of larger disperse clusters, we allow here for 40 % of outliers.

For the MINCE method of Section 4.1, every pre-cluster is mapped to the pre-shape-sphere in optimal position to its Procrustes mean (see e.g. Dryden and Mardia (2016)) and PNS using spherical distance (see e.g. Dryden et al. (2019)) is conducted. Typical shapes of clusters are depicted in the left panel of Fig. A11. In total, 15 clusters have been found. Table A2 lists their sizes as well as the number of outliers. Notably, Clusters 5 and 6 depicted in the right panel of Fig. A11 have been separated by circular mode hunting. More details can be found in Wiechers et al. (2021).

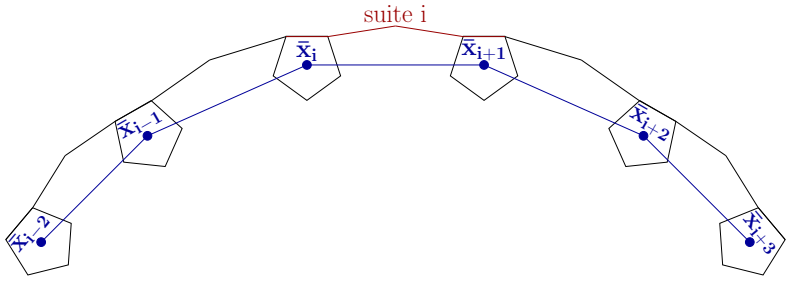


Fig. A10: Adapted from Wiechers et al. (2021): Centers of the sugar rings from Fig. A7, their connecting backbones (red and black lines) giving 5 suites (two before and two after suite i) yielding the mesoscopic shape (blue lines).

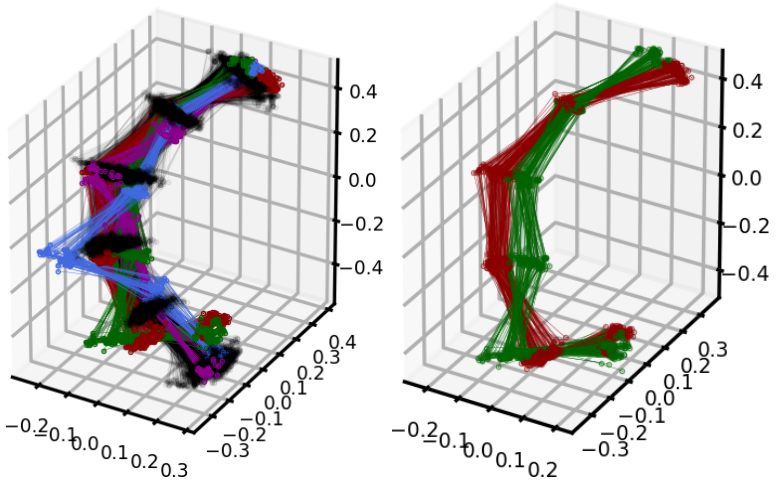


Fig. A11: Left: Five mesoscopic clusters found by MINCE post AGE, that can be well displayed together with cluster numbers in parentheses from Table A1: black (1), red (5), green (6), magenta (9) and blue (10). Right: A pre-cluster divided into Cluster 5 (green) and Cluster 6 (green) by circular mode hunting.

Table A2: Clusters and outliers (numbers and sizes, the relative amount of outliers 43.9 % is governed by the choice of $d_{\max}$) found by applying MINCE post AGE to the mesoscopic shape data detailed in Section 4.3. Asterisks indicate clusters displayed in Fig. A11.

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	R	$\sum$
Size	5788	385	295	180	145	115	100	80	76	69	68	61	60	57	56	5904	13409

4.5 A New Clustering Method: MINCE post AGE on a Euclidean Space

We apply MINCE post AGE on a Euclidean space, thus mapping every pre-cluster found by AGE to a sphere, performing PNS and, as key ingredient, on the main PNS component, which is a circle, conducting circular mode hunting to obtain post clusters. This leads to a new clustering method which we now describe.

First of all, we need to define a mapping from a data pre-cluster $x_1, \ldots, x_n \in \mathbb{R}^d$ to a suitable sphere. While several options come to mind, we use the following method. Estimate a hypersphere of dimension $d-1$ in $\mathbb{R}^d$ center $\mu \in \mathbb{R}^d$ and a radius $r > 0$ by

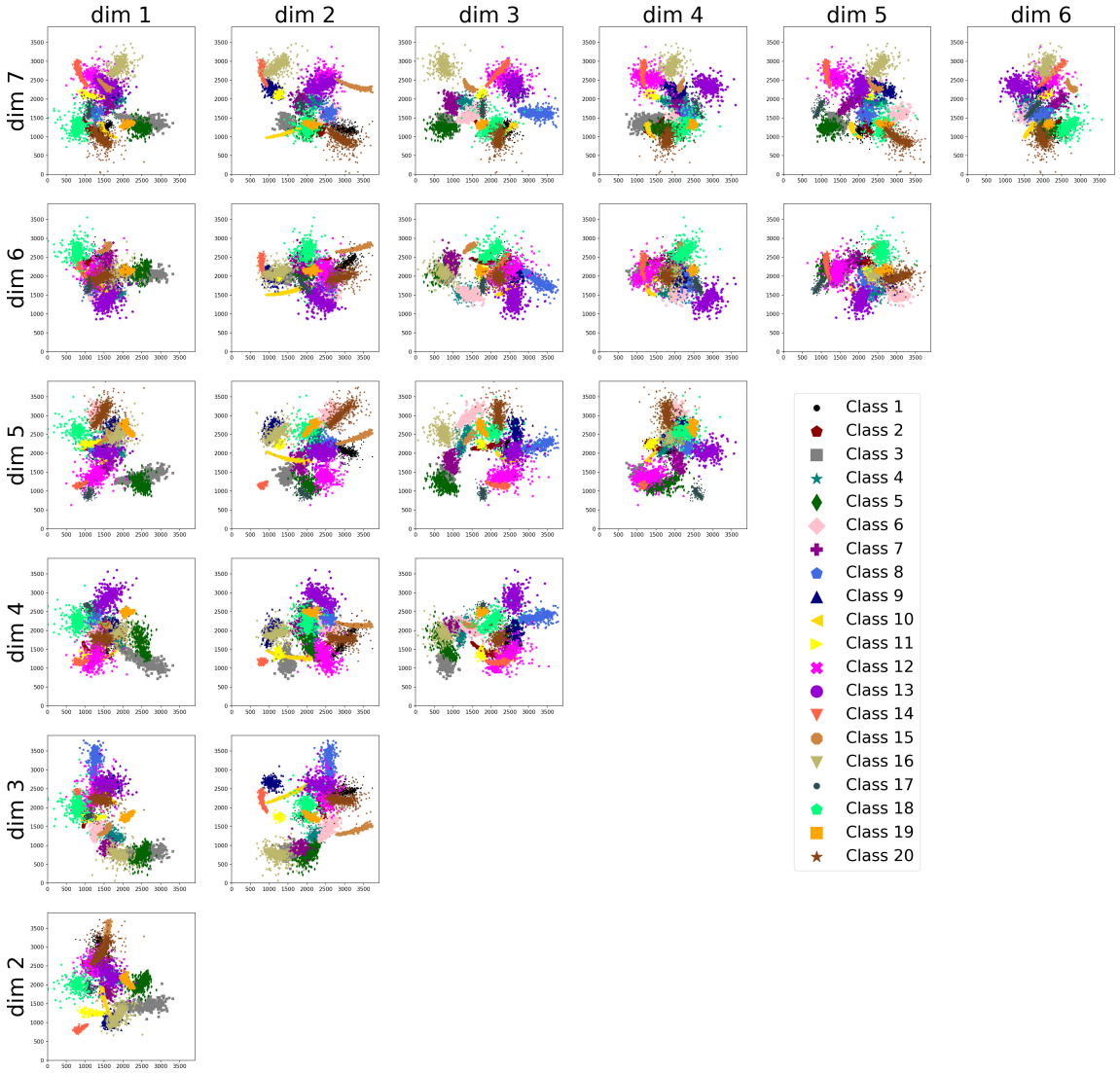


Fig. A12: Coordinate scatter plots of the challenging worms data set of 20 worm-like original clusters in $\mathbb{R}^7$ produced with an adaption of the Matlab script from Sieranoja and Fränti (2019).

minimizing the sum of squared residuals

$$\sum_{j=1}^{n} (r - \|\mu - x_j\|)^2 \ .$$

Here, we test against overfitting by comparing with a best fitting affine hyperspace $A_{v,\alpha}$ from (4) defined by $v \in \mathbb{S}^d$ and $\alpha > 0$, minimizing the sum of squared residuals

$$\sum_{j=1}^{n} \left(\alpha - v^T x_j\right)^2 \ .$$

For an illustration a *worms* data set has been created using the Matlab script from Sieranoja and Fränti (2019) for dimension $m = 7$. In total, $12,711$ data points in $\mathbb{R}^7$ have been produced (for computational feasibility we have changed some of the script's default flags to reduce the number of points produced by a factor of approximately 5) in 20 clusters

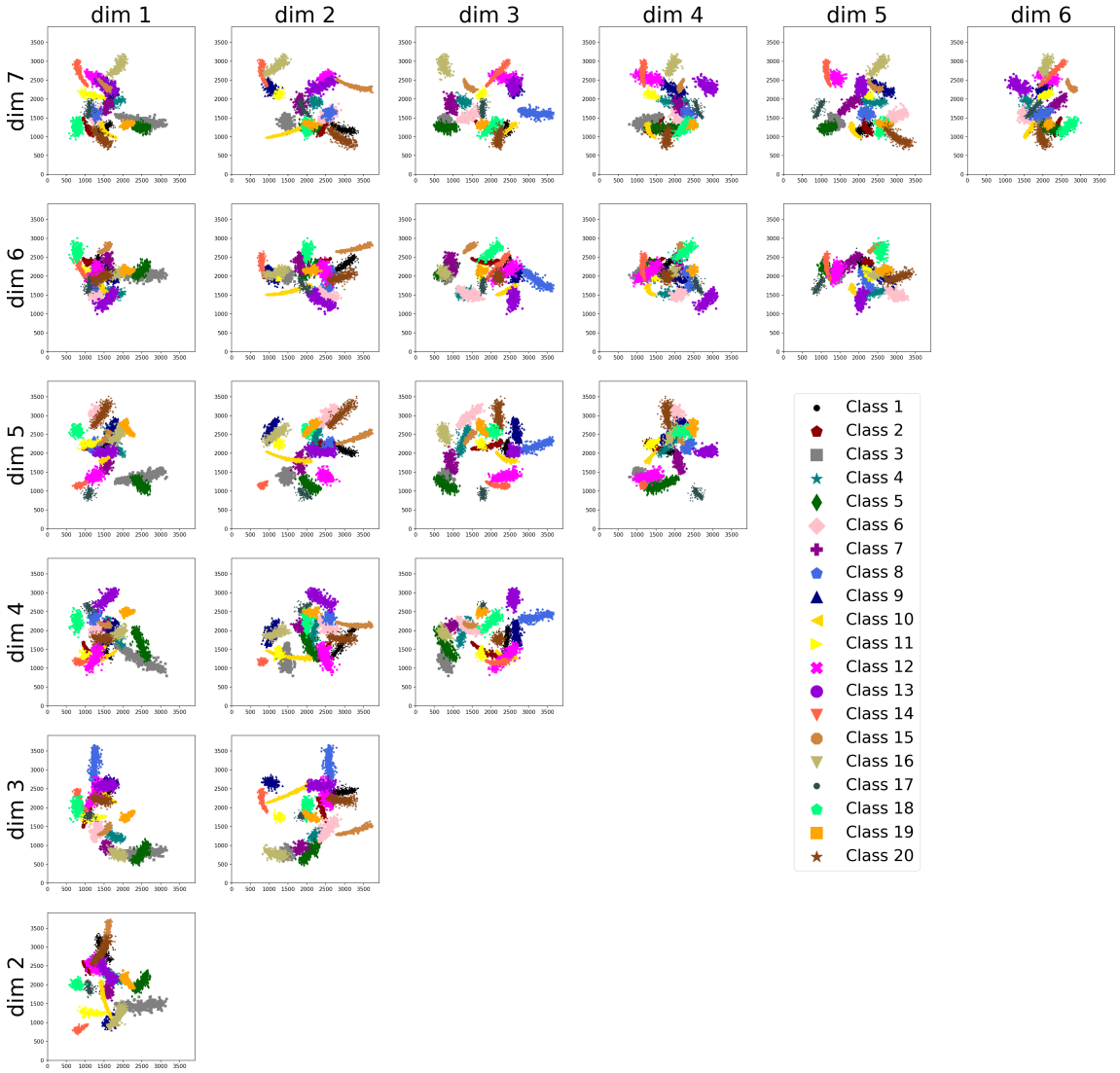


Fig. A13: Coordinate scatter plots of clusters found by applying MINCE post AGE to the challenging worms data set in Fig. A12.

consisting of Gaussian samples centered at initially random positions that drifted in each step into a random direction with increased variance, thus producing worm-like samples similar to those depicted in Fig. A12. For AGE, Algorithm 1, choosing tuning parameters $q = 0$ (reflecting that within cluster neighboring distances may be of the same order as between cluster distances) and $d_{\max}$ such that 10% of the elements in the single linkage tree are in a branch with less than $\kappa + 1$ data points (for this and all applications here, we use $\kappa = 40$), the MINCE post AGE method from Section 4 detailed above has been able to correctly retrieve all 20 clusters, only assigning approximately 10 % to the outlier set.

4.5.1 The challenging worms data set

For the second experiment, we produce more challenging clusters of varying members and density. To this end the script of Sieranoja and Fränti (2019) has been modified:

- The step-length parameter *stepl* has been increased from the default value of 2, to the value of 5, so that the individual clusters become longer.

Table A3: MINCE post AGE cluster sizes (2nd colomun) ground truth cluster (sizes in 3rd column) numbers (1st column) and number of data points classified as outliers (4th column) of the challenging *worms* data set depicted in Fig. A12.

Cluster number	MINCE post AGE size	ground truth size	unassigned by MINCE post AGE
1	666	768	102
2	666	666	0
3	842	891	49
4	619	720	101
5	604	708	104
6	558	588	30
7	503	540	37
8	589	657	68
9	602	648	46
10	900	900	0
11	486	486	0
12	493	666	173
13	456	585	129
14	714	714	0
15	678	678	0
16	407	543	136
17	405	426	21
18	353	480	127
19	369	369	0
20	530	678	148
Outliers	1271	0	0
Total	12711	12711	1271

- Initially the variance range parameter had value $var\_range = [10, 50]$, which meant that for every cluster the variance moved from 10 at first sampling to 50 at the end of sampling. Now for each cluster a random variable X uniform in $[0, 1]$ is drawn and setting $var\_range = [X * 20, X * 100]$ individually, yields clusters of different densities.

- To obtain clusters with different numbers of elements, the number of steps, called *numsteps* is now randomly sampled for each cluster from a normal distribution with expected value $\mu = 200$ and standard deviation $\sigma = 40$. The number is rounded to the closest integer. As in the original script, in each step 3 points are sampled.

The challenging worms data set's ground truth clusters are displayed in Fig. A12. Not changing our tuning parameters for AGE in Algorithm 1, from the challenging worms data set MINCE post AGE has correctly identified 19 out of the 20 clusters, two clusters could not be separated with statistical significance. Setting $q = 1/5$, however, MINCE post AGE has been able to retrieve all 20 clusters, no data point has been assigned to a wrong cluster, 7 clusters have been retrieved perfectly and 1,217 data points have been classified as outliers. For comparison the retrieved clusters are illustrated in Fig. A13 and their true and retrieved cluster sizes as well as the number of assigned outliers are listed in Table A3. Notably, this has been achieved at the price of classifying 10 % as outliers.

5 PCA on the Torus through Wrapped Normal Distributions

So far we have given PCA methods for a torus which are intuitive but one simple way is to follow classical PCA and use the wrapped normal distribution in place of the multivariate normal distribution. In fact, the underlying covariance matrix of the multivariate wrapped normal distribution has simple explicit expressions for its trigonometric moments. We describe the procedures suggested by Kent and Mardia (2009) and propose a new method. Suppose that θ is a vector of angles following a wrapped normal torus distribution; that is, $\theta_j = X_j \mod 2\pi, j \in \{1, \ldots, p\}$, where $X \sim N_p(0, \Sigma)$. Then it can be shown that

$$\text{var}(\cos\theta) = DAD - cc^T, \quad \text{var}(\sin\theta) = DBD, \quad \text{cov}(\cos\theta, \sin\theta) = 0. \tag{15}$$

where the elements of the vector c and matrices A, B are given by

$$c_j = \exp\{-\frac{1}{2}\sigma_{jj}\}, \quad a_{jk} = c_j c_k \cosh(\sigma_{jk}), \quad b_{jk} = c_j c_k \sinh(\sigma_{jk})$$

and $D = \text{diag}(c)$. Thus Σ can be recovered from the trigonometric moments through the equation

$$\Sigma = \sinh^{-1}(D^{-1}\text{var}(\sin\theta)D^{-1}). \tag{16}$$

Here the notation $\sinh^{-1}(\cdot)$ applied to a matrix means that the inverse sinh function, $\sinh^{-1}(u) = \log(u + \sqrt{u^2+1})$, is applied to each element of the matrix.

These results suggest a method to estimate Σ from an $n \times p$ matrix of torus data:

(a) Calculate the sample first order trigonometric moments for the p angles, and rotate each angle so that the resultant vector points towards the positive horizontal axis, leading to an estimate of D.

(b) Calculate the sample second trigonometric moments to get an estimate of corresponding to $\text{var}(\sin\theta)$.

(c) Now use (16) to produce an estimate of Σ.

Similarly, we can also estimate Σ from $\text{var}(\cos\theta)$ using $\Sigma = \cosh^{-1}(D^{-1}\text{var}(\cos\theta)D^{-1}) + 11^T$ where 1 is a p-dimensional vector of ones; let these estimates be denoted by $\hat{\Sigma_1}$ and $\hat{\Sigma_2}$ respectively. Now we can obtain a pooled estimate of Σ as given in Kent and Mardia (2009). Another approach is to get the PCA from each of these estimates and since from (15) $\text{cov}(\cos\theta, \sin\theta) = 0$, we can get back into the original space as follows. For example, if y_1 and z_1 are the first principal components from $\hat{\Sigma_1}$ and $\hat{\Sigma_2}$ respectively then the first angular component ϕ_1 say can be obtained by setting $y_1 = r\sin\phi_1, z_1 = r\cos\phi_1$ and so on. This and other proposals of Kent and Mardia (2009) need further investigation.

6 Discussion

We have used the word manifold broadly which includes shape spaces Σ_m^k that are no longer manifolds but stratified spaces, if objects of dimension $m = 3$ or higher are considered. Future work aims at extending MINCE to more general manifolds and stratified spaces.

With the AGE algorithm presented, the MINCE method is specifically designed to deliver perfect cluster assignment, thus avoiding wrong cluster assignment at the price of a potentially larger outlier set. In particular, this is desirable in view of application in

biomolecular structure analysis. The AGE algorithm comes with three tuning parameters that, in future applications, can be learned over larger databanks in molecular biology. Notably, MINCE post AGE requires no previous knowledge about the numbers of clusters to be found.

At the core of MINCE is a data adaptive transformation from a Euclidean, manifold or stratified space to a sphere, possibly a stratified sphere in order to preserve as much topology and geometry of the original space. For the Euclidean data from Section 4.5 a different method chooses suitable points $p, q \in \mathbb{R}^d$ which become north pole $(0, \ldots, 0, 1)$ and south pole $(0, \ldots, 0, -1)$, respectively, of $\mathbb{S}^d$ to which data are mapped via inverse stereographic projection

$$\mathbb{R}^d \ni x \mapsto \frac{x-p}{\|x-q\|} =: y \mapsto \left(\frac{2y}{1+\|y\|^2}, \frac{1-\|y\|^2}{1+\|y\|^2} \right) \in \mathbb{S}^d \subset \mathbb{R}^{d+1}.$$

Alternatively $q = \infty$ can be chosen (yielding the one-point Alexandroff compactification of $\mathbb{R}^d$ mentioned in Section 1) such that more simply, $y = x - p$. Additionally choosing $p = \frac{1}{n}\sum_{j=1} x_j$ seems canonical.

As another approach to turn from a metric space into spherical data, one may use multidimensional scaling, not with a Euclidean space but with an underlying sphere. For data on a torus, the ST-PCA approach of Zouboulоglou et al. (2021) thus refines our torus PCA by reducing distortions, followed by PNS.

Finally, let us point to statistical testing, which relies asymptotically on central limit theorems. Due to curvature, in particular on positive curvature manifolds, limiting rates for quite large sample sizes can considerably deviate from Euclidean analogs (underlying Remark 1) for a large number of reasonable data models, requiring specifically designed bootstrap methods, see Hundrieser et al. (2020). This new effect has been called finite sample smeariness by Hundrieser et al. (2020) and it is an open problem how this manifests in the asymptotics of the main principal nested circle.

Funding Information

H. Wiechers and B. Eltzner gratefully acknowledge funding by DFG SFB 1456. Additionally, S. F. Huckemann acknowledges funding by the Felix-Bernstein-Institute at the University of Göttingen, DFG HU 1575/7 and the Niedersachsen Vorab of the Volkswagen Foundation. K. V. Mardia acknowledges the Leverhulme Trust for the Emeritus Fellowship.

References

M. Arnaudon, L. Miclo, Means in complete manifolds: uniqueness and approximation, ESAIM: Probability and Statistics 18 (2014) 185–206.

V. Arsigny, O. Commowick, X. Pennec, N. Ayache, A log-euclidean framework for statistics on diffeomorphisms, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, Springer, 2006, pp. 924–931.

H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, Nucleic Acids Research 28 (2000) 235–242.

J. Boisvert, X. Pennec, H. Labelle, F. Cheriet, N. Ayache, Principal spine shape deformation modes using Riemannian geometry and articulated models, in: Articulated Motion and Deformable Objects, Springer, 2006, pp. 346–355.

P. Chaudhuri, J. Marron, SiZer for exploration of structures in curves, Journal of the American Statistical Association 94 (1999) 807–823.

P. Chaudhuri, J. Marron, Scale space view of curve estimation, The Annals of Statistics 28 (2000) 408–428.

I. L. Dryden, K.-R. Kim, C. A. Laughton, H. Le, Principal nested shape space analysis of molecular dynamics data, arXiv preprint arXiv:1903.09445 (2019).

I. L. Dryden, K. V. Mardia, Statistical Shape Analysis, Wiley, Chichester, 2nd edition, 2016.

L. Dümbgen, G. Walther, Multiscale inference about a density, Ann. Statist. 36 (2008) 1758–1785.

B. Eltzner, S. Huckemann, K. V. Mardia, Torus principal component analysis with applications to RNA structure, Ann. Appl. Statist. 12 (2018) 1332–1359.

B. Eltzner, S. F. Huckemann, A smeary central limit theorem for manifolds with application to high-dimensional spheres, Ann. Statist. 47 (2019) 3360–3381.

P. T. Fletcher, S. C. Joshi, Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors, ECCV Workshops CVAMIA and MMBIA (2004) 87–98.

K. Florek, J. Lukaszewicz, J. Perkal, Steinhaus, Hugo, S. Zubrzycki, Sur la liaison et la division des points d'un ensemble fini, Colloquium Mathematicum 2 (1951) 282–285.

J. Frellsen, I. Moltke, M. Thiim, K. V. Mardia, J. Ferkinghoff-Borg, T. Hamelryck, A Probabilistic Model of RNA Conformational Space, PLoS Comput Biol 5 (2009) e1000406.

J. C. Gower, Generalized Procrustes analysis, Psychometrika 40 (1975) 33–51.

A. Hatcher, Algebraic topology, Cambridge Univ. Press, 2005.

S. Huckemann, (Semi-)intrinsic statistical analysis on non-Euclidean spaces, in: Advances in Complex Data Modeling and Computational Methods in Statistics, Springer, 2014, pp. 103–118.

S. Huckemann, T. Hotz, A. Munk, Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion)., Statistica Sinica 20 (2010) 1–100.

S. Huckemann, K.-R. Kim, A. Munk, F. Rehfeldt, M. Sommerfeld, J. Weickert, C. Wollnik, The circular sizer, inferred persistence of shape parameters and application to early stem cell differentiation, Bernoulli 22 (2016) 2113–2142.

S. Huckemann, H. Ziezold, Principal component analysis for Riemannian manifolds with an application to triangular shape spaces, Advances of Applied Probability (SGSA) 38 (2006) 299–319.

S. F. Huckemann, Comments on: Recent advances in directional statistics, TEST (2021) 1–5.

S. F. Huckemann, B. Eltzner, Backward nested descriptors asymptotics with inference on stem cell differentiation, The Annals of Statistics (2018) 1994 – 2019.

S. Hundrieser, B. Eltzner, S. F. Huckemann, Finite sample smeariness of Fréchet means and application to climate, 2020. ArXiv preprint arXiv:2005.02321.

S. Jain, D. C. Richardson, J. S. Richardson, Chapter Seven - Computational Methods for RNA Structure Validation and Improvement, in: S. A. Woodson, F. H. Allain (Eds.), Structures of Large RNA Molecules and Their Complexes, volume 558 of Methods in Enzymology, Academic Press, 2015, pp. 181 – 212.

S. Jung, I. L. Dryden, J. S. Marron, Analysis of principal nested spheres, Biometrika 99 (2012) 551–568.

P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, Bioinformatics 24 (2007) 719–720.

T. Lindeberg, Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space, Journal of Mathematical Imaging and Vision 40 (2011) 36–81.

K. V. Mardia, Statistical approaches to three key challenges in protein structural bioinformatics, Journal of the Royal Statistical Society: SERIES C: Applied Statistics (2013) 487–514.

K. V. Mardia, Comments on: Recent advances in directional statistics, TEST (2021) 1–5.

K. V. Mardia, P. E. Jupp, Directional Statistics, Wiley, New York, 2000.

K. V. Mardia, J. T. Kent, J. M. Bibby, Multivariate Analysis, Academic press, 1979.

L. J. W. Murray, W. B. Arendall, D. C. Richardson, J. S. Richardson, RNA backbone is rotameric, Proceedings of the National Academy of Sciences 100 (2003) 13904–13909.

A. Obulkasim, G. A. Meijer, M. A. van de Wiel, Semi-supervised adaptive-height snipping of the hierarchical clustering tree, BMC Bioinformatics 16 (2015) 15.

S. Olsson, W. Boomsma, J. Frellsen, S. Bottaro, T. Harder, J. Ferkinghoff-Borg, T. Hamelryck, Generative probabilistic models extend the scope of inferential structure determination, Journal of Magnetic Resonance 213 (2011) 182–186.

V. M. Panaretos, T. Pham, Z. Yao, Principal flows, Journal of the American Statistical Association 109 (2014) 424–436.

X. Pennec, Barycentric subspace analysis on manifolds, The Annals of Statistics 46 (2018) 2711–2746.

A. Pewsey, E. García-Portugués, Recent advances in directional statistics, TEST (2021) 1–58.

S. P. Preston, A. T. Wood, Two-sample bootstrap hypothesis tests for three-dimensional labelled landmark data, Scandinavian journal of statistics 37 (2010) 568–587.

T. Schlick, A. M. Pyle, Opportunities and challenges in rna structural modeling and design, Biophysical journal 113 (2017) 225–234. 28162235[pmid].

S. Sieranoja, P. Fränti, Fast and general density peaks clustering, Pattern Recognition Letters 128 (2019) 551–558.

R. R. Sokal, C. D. Michener, A statistical method for evaluating systematic relationships, University of Kansas Science Bulletin 38 (1958) 1409–1438.

S. Sommer, Horizontal dimensionality reduction and iterated frame bundle development, in: Geometric Science of Information, Springer, 2013, pp. 76–83.

D. Tran, B. Eltzner, S. F. Huckemann, Improved two-sample tests on manifolds and non-smooth quotient spaces, 2021. Soon available on arXiv.

M. Tsagris, C. Beneki, H. Hassani, On the folded normal distribution, Mathematics 2 (2014) 12–28.

H. Wiechers, B. Eltzner, K. V. Mardia, S. F. Huckemann, Learning torus PCA based classification for multiscale RNA backbone structure correction with application to SARS-CoV-2, 2021. BioRxiv, https://doi.org/10.1101/2021.08.06.455406, submitted.

C.-H. Yang, B. C. Vemuri, Nested Grassmanns for dimensionality reduction with applications to shape analysis, in: International Conference on Information Processing in Medical Imaging, Springer, pp. 136–149.

Z. Yao, Z. Zhang, Principal boundary on Riemannian manifolds, Journal of the American Statistical Association 115 (2020) 1435–1448.

P. Zoubouloglou, E. García-Portugués, J. S. Marron, Scaled torus principal component analysis, 2021. Soon available on arXiv.

CHAPTER B

Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2

Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2

Henrik Wiechers [1,*], Benjamin Eltzner [2], Kanti V. Mardia [3,4] and Stephan F. Huckemann [1,*]

[1]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, Georgia-Augusta-University, Göttingen, 37077, Germany,
[2]Max Planck Institute for Biophysical Chemistry, Göttingen, 37077, Germany,
[3]Department of Statistics, School of Mathematics, University of Leeds, LS2 9JT, England,
[4]Department of Statistics, University of Oxford, 24-29 St Giles', Oxford OX1 3LB.

May 25, 2023

Abstract

Three-dimensional RNA structures frequently contain atomic clashes. Usually, corrections approximate the biophysical chemistry, which is computationally intensive and often does not correct all clashes. We propose fast, data-driven reconstructions from clash free benchmark data with two-scale shape analysis: microscopic (suites) dihedral backbone angles, mesoscopic sugar ring centre landmarks. Our analysis relates concentrated mesoscopic scale neighbourhoods to microscopic scale clusters, correcting within-suite-backbone-to-backbone clashes exploiting angular shape and size-and-shape Fréchet means. Validation shows that learned classes highly correspond with literature clusters and reconstructions are well within physical resolution. We illustrate the power of our method using cutting-edge SARS-CoV-2 RNA.

Keywords: angular shape analysis, clash correction, frameshift stimulation element, Fréchet and Procrustes means, geodesic projection, mesoscopic shape and microscopic shape, size-and-shape space

1 Introduction

Understanding the structure of active biomolecules is ever more important for maintaining and improving human health, as has been summarized by Schlick and Pyle (2017). In particular, this pertains to RNA molecules in designing drugs which target specific structures (see Batool *et al.* (2019)), as recently impressively demonstrated by the worldwide effort confronting the SARS-CoV-2 (severe acute respiratory syndrome) virus responsible for the COVID-19 (corona virus disease) pandemic (see Croll *et al.* (2021)).

Extracting RNA primary structure (sequencing) is nowadays fairly well feasible using currently available gene sequencing technology (e.g. Wang *et al.* (2009)). Predicting the 3D structure (helices, etc.) from that, however, is a still unsolved fundamental problem (e.g. Schlick and Pyle (2017)). Although elaborate methods such as X-ray crystallography and cryo-EM (cryogenic electron microscopy) are used that determine spatial electron densities – and from these densities individual atom positions can be inferred – frequently, the inferred molecular structures contain so-called clashes as detailed by Murray *et al.* (2003); Chen *et al.* (2010) and others.

Definition 1.1. *A* **clash** *is a forbidden molecular configuration, where two atoms are reconstructed closer to each other than is chemically possible.*

In case of RNA, clashes most relevant and most difficult to correct are between atoms along the backbone (main chain), in particular when single hydrogen atoms not contributing to electron densities are added to inferred structures (see Figure B1); a detailed discussion is given in Murray *et al.* (2003).

In order to correct such clashes, methods from *molecular dynamics* are usually employed: Simulated atoms are allowed to fluctuate into positions of minimal energy, following approximations of the laws of biophysical chemistry (e.g. Chou *et al.* (2013a)). For RNA molecules, these simulations are highly computation intensive due to the large variability of RNA shape. If local and not global energy minima are achieved, thus corrected molecules may still feature clashes and their geometries may be outliers in comparison to clash free geometries (e.g. Richardson *et al.* (2018)). Supplement D briefly sketches the state of the art correction method ERRASER by Chou *et al.* (2013a) and details this observation.

As most RNA backbone clashes appear within ***suites*** (the section from one sugar ring to the next, e.g. Murray *et al.* (2003), see Figure B1 and Notation 3.1, Section 3), we therefore apply our method to ***within-suite-backbone-to-backbone*** clashes here; although it can be more generally applied. For the scope of this article, we call here suites ***clash free*** if they are free of within-suite-backbone-to-backbone clashes. We analyze the RNA backbone simultaneously at two scales exploiting their interdependence as follows.

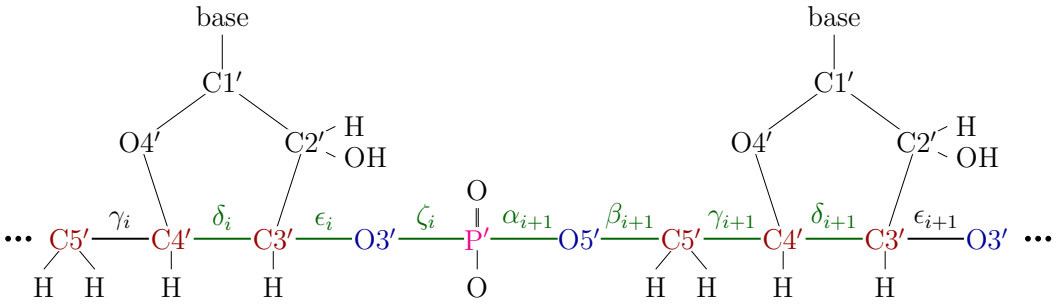


Figure B1: *2D scheme of backbone suite number i with 7 dihedral angles (see Figure B7) $\delta_i, \epsilon_i, \zeta_i, \alpha_{i+1}, \beta_{i+1}, \gamma_{i+1}, \delta_{i+1}$ describing the suite's 3D structure.*

We work on two levels, the microscopic (atomic level) and the mesoscopic (level of objects). At the ***microscopic*** scale we model the backbone of suites by tuples of 7 dihedral angles, each between 0 and 2π from the backbone atoms, giving a data point on the seven dimensional torus $\mathbb{T}^7$. We are thus working on a form of shape analysis from angles (angular shape analysis). At the ***mesoscopic*** scale we model k suites before and k suites after a central suite of concern represented by $2k+2$ *pseudo-landmarks*, the centers of sugar rings, see Figure B2. Our interest will be the ***size-and-shape*** (see Dryden and Mardia (2016)) of these landmarks. Setting $k = 2$. i.e. six landmarks in total (which depicts roughly a half helix turn), our data analysis leads to the conclusion that for clash free data, concentrated clusters at mesoscopic scale correspond to clusters at microscopic scale. This correspondence is at the heart of our two-scale correction method: Since we aim to correct potential errors at the microscopic scale, we first learn classes of clash free microscopic shapes by clustering a benchmark data set of clash free data at the microscopic scale. As illustrated in the left two panels of Figure B3 we provide a data driven correction (green) for a clash suite (red) by a Fréchet mean on the torus at the microscopic scale (left

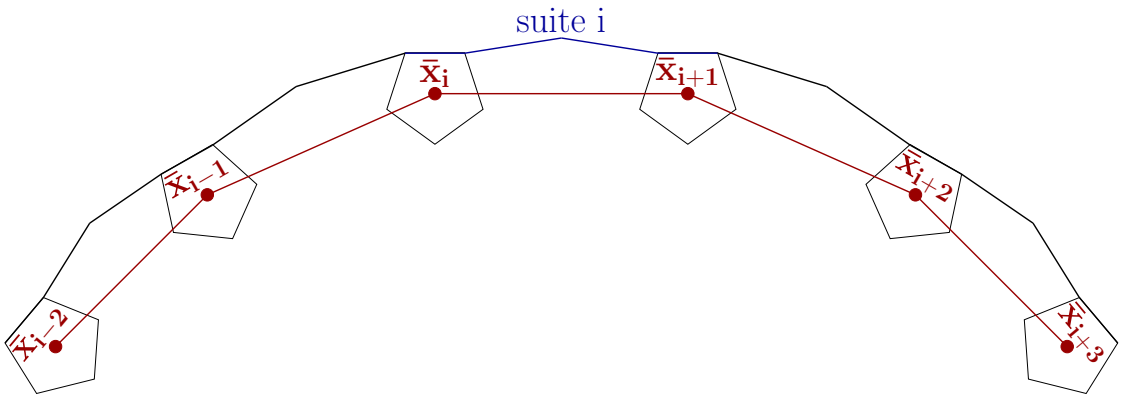


Figure B2: *The mesoscopic shape (red lines) for $k = 2$ centered at the i-th suite is determined by the six centers of the sugar rings $\bar{\mathbf{x}}_{\mathbf{i-2}}, \ldots, \bar{\mathbf{x}}_{\mathbf{i+3}}$. Their connecting backbones (blue and black lines) give 5 suites, two before and two after suite i.*

panel) within a specific class of clash free suites (grey) from Tang *et al.* (2001) (file 1f8v, see Table B16). To determine the class which is used for microscopic structure correction, we leverage the corresponding mesoscopic shape describing the geometry of the RNA strand in proximity to the clash suites by determining a set of closest mesoscopic shapes to the mesoscopic shape containing the clash suite. We then consider the microscopic suite shapes corresponding to these nearby mesoscopic shapes and determine the class which dominates this set (center left panel, same colors). At the mesoscopic scale, our correction (green) is the geodesic projection of the corresponding Procrustes mean to the mesoscopic shape featuring the same endpoints and the length of the corrected suite. Typically, our correction at mesoscopic scale requires only a few moderate shifts of sugar centers (left center, see also Figure B13, right panel).

We validate our correction method based on the interdependence of clash free RNA backbone shape at the two scales (microscopic and mesoscopic) by showing that the corrections proposed stay well below resolution level on the benchmark data. We also validate our classification by comparison with a suite clustering proposed by Richardson *et al.* (2008) who investigated a larger data set (comprising about twice as many suites than our benchmark data set): The classes we propose correspond well to their clusters, where some of our classes comprise several of their clusters.

In application, we propose clash free corrections for ten structure proposals from Zhang *et al.* (2021) for two suites of the *frameshift stimulation element* (which facilitates decoding more than one protein from a single RNA strand) of SARS-CoV-2 which are difficult to reconstruct, and for which, to the best knowledge of the authors, there are no consistent 3D structures known to date. Our method proposes structure which are strikingly consistent, and by design, are clash free. For one of the two suites, the situation is exemplified below in the two right panels of Figure B3: For each of the ten clashing proposals (red), at mesoscopic scale (right panel) we propose clash free corrections (green) and at microscopic scale (center right panel, same colors) our corrections agree nearly unambiguously.

Our paper is structured as follows. First, we introduce the two shape spaces: the torus (angular shape space) describing the RNA backbone uniquely at microscopic (atomic) scale and the size-and-shape space describing the RNA backbone at mesoscopic scale. Then follows the concept of Fréchet means used at both scales for clash correction. At mesoscopic scale (here Fréchet means are Procrustes means), we provide a novel projection (preserving constraints from the original mesoscopic shape and its microscopic correction)

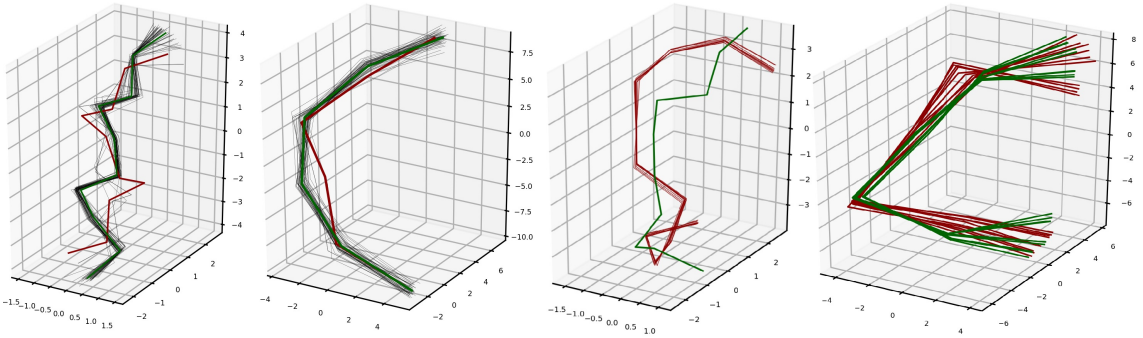


Figure B3: *Left two panels: A clashing suite (red) (from benchmark file 1f8v, Tang* et al. *(2001), see Table B16 in the supplement) with its clash free neighbors (black) and proposed clash free correction (green) at microscopic scale (left) and mesoscopic scale (left center). Right two panels: Ten proposed reconstructions (red) by Zhang* et al. *(2021), which are all clashing, for Suite 28/29 (cf. Figure B5) connecting two helical segments in the frameshift stimulation element of SARS-CoV-2 and our ten clash free corrections (green) at microscopic scale (center right) and at mesoscopic scale (right).*

for the Procrustes mean. In Section 3 we link the 3D RNA backbone structure at two scales to our two shape spaces, overview clash detection and provide our benchmark data. Section 4 proposes our multiscale RNA backbone correction method, first introducing learned classes from the clash free benchmark data and validating them. We then present the interdependence of clash free RNA backbone shape at the two scales (microscopic and mesoscopic) and detail how we exploit this for the new method proposed and validate it. Finally we apply our method to the correction of the RNA backbone of SARS-CoV-2. In Section 6 we discuss further potentials of our method, in particular how ***multiscale shape analysis*** can be more fully developed and how it could be used to complement existing reconstruction methods for long stranded biomolecules based on molecular dynamics.

While we measure angles in radians, for instant comparison with other research in this area, some of the Figures report results in degrees.

Finally we list the content of our supplementary material, containing all code and all data, as well as further data analysis and an overview of the MINT-AGE algorithm from Mardia *et al.* (2022).

2 Tools from Shape Analysis

For Fréchet means defined in Section 2.3 below we will need appropriate distances for the microscopic and mesoscopic scale which we now give in Sections 2.1 and 2.2, respectively. For the mesoscopic scale we develop in Section 2.4 a geodesic projection since we have to impose suitable geometric constraints.

2.1 The Torus for Microscopic Scale

The one-dimensional *torus* is

$$\mathbb{T} := [0, 2\pi]/\sim$$

where "$\sim$" denotes that 0 and 2π are identified. It is a metric space with canonical distance

$$d_{\mathbb{T}}(\phi, \psi) \quad = \quad \min\{|\phi - \psi|, 2\pi - |\phi - \psi|\}, \quad \phi, \psi \in \mathbb{T}\,. \tag{1}$$

The canonical product of m one-dimensional tori is the m-dimensional torus $\mathbb{T}^m$ with the canonical product distance given by

$$d_{\mathbb{T}^m}(\phi,\psi) \quad = \quad \sqrt{\sum_{j=1}^{m} d(\phi_j,\psi_j)^2}\,, \tag{2}$$

for $\phi = (\phi_1,\ldots,\phi_m), \psi = (\psi_1,\ldots,\psi_m) \in \mathbb{T}^m$. Several authors have studied data on the torus, especially representing large biomolecules, and developed specialized methods, including Parsons *et al.* (2005); Altis *et al.* (2008); Kent and Mardia (2009); Sargsyan *et al.* (2012); Eltzner *et al.* (2018); AlQuraishi (2019); Zoubouloglou *et al.* (2021).

2.2 Size-and-Shape for Mesoscopic Scale

We describe a landmark configuration matrix $X = (\mathbf{x}_1,\ldots,\mathbf{x}_m) \in \mathbb{R}^{3\times m}$ encoding $m \in \mathbb{N}$, three-dimensional landmark positions $\mathbf{x}_i \in \mathbb{R}^3$, $i = 1,\ldots,m$ by its *size-and-shape* as follows, see Dryden and Mardia (2016): Proper (i.e. orientation preserving) Euclidean transformations comprising rotations and translations $T = (R,v) \in \mathrm{SO}(3) \times \mathbb{R}^3$ act on X columnwise via

$$T.X := (R\mathbf{x}_1 + v,\ldots,R\mathbf{x}_m + v)\,.$$

Then

$$S\Sigma_3^m := \{[X] : X \in \mathbb{R}^{3\times m}\} \text{ where } [X] := \{T.X : T \in \mathrm{SO}(3)\times\mathbb{R}^3\} \tag{3}$$

is the *size-and-shape space* which is equipped with the quotient distance, also called *Procrustes distance*

$$d_\Sigma([X],[Y]) \quad := \quad \min_{T\in\mathrm{SO}(3)\times\mathbb{R}^3} \|X - T.Y\| \tag{4}$$

with the standard Frobenius norm on $\mathbb{R}^{3\times m}$. We say that X and Y are in *optimal position* if

$$d_\Sigma([X],[Y]) = \|X - Y\|\,.$$

Taking derivatives and using a singular value decomposition (SVD) it follows at once that configurations X, Y in optimal position have coinciding mean landmarks with symmetric YX^T (e.g. Dryden and Mardia (2016, Result 7.1)). For this reason, we assume that all landmark configurations are *centered*, i.e. their landmarks vectors add up to zero. Optimal positioning is then conveyed by rotations $R \in \mathrm{SO}(3)$ only, i.e. RY is in optimal position to X if $R = VSU^T$ with a suitable diagonal matrix S with entries in $\{-1,1\}$ and a SVD $YX^T = UDV^T$ (here U, V are orthogonal, D is diagonal with nonnegative entries).

2.3 Fréchet Means for Both Scales

Definition 2.1. *For data $X_1,\ldots,X_n \in M$ on an arbitrary metric space (M,d), define their* Fréchet means *by*

$$\operatorname*{argmin}_{X\in M} \sum_{j=1}^{n} d(X,X_i)^2\,.$$

The Fréchet mean is a generalization of the classical Euclidean mean. On complete spaces, Fréchet means exist, and on manifolds, if samples are drawn from continuous distributions, they are almost surely unique (see Arnaudon and Miclo (2014)). On stratified

quotient spaces, such as size-and-shape space for 3D configurations, they lie on the manifold part (the top-dimensional dense stratum) if the manifold part is assumed with positive probability (see Huckemann (2012)).

On $S\Sigma_3^m$, Fréchet means defined by Procrustes distance are also called *Procrustes means*. On $\mathbb{T}^m$ we call them *torus means*.

2.4 Geodesic Projection to Constrained Size-and-Shape

Our CLEAN MINT-AGE Algorithm in Section 4.3.1 corrects clashes not only at atomic suite (microscopic) scale but also at mesoscopic scale. The corrected mesoscopic shape m_{τ_c} in Section 4.3.1 features two constraints. The first one sets the distance between its first and last landmark to the corresponding distance of the original mesoscopic shape, thus assuring its fit into a larger RNA strand. The second one sets the distance between its two central landmarks to the length of the corrected suite, assuring the fit of the latter into the former.

With more general future applications in mind, assume that the distances between $r \in \mathbb{N}$, $(2 \leq 2r \leq m)$ landmark pairs are constants $a_1, \ldots, a_r > 0$. With a permutation σ of $(1, \ldots, m)$ we may assume that landmark $\sigma(j)$ is paired with landmark $\sigma(j+r)$ for $j = 1, \ldots, r$ while landmarks $\sigma(j)$ for $2r < j \leq m$ (if $2r < m$) are unconstrained.

Definition 2.2. *Let $r \in \mathbb{N}$ with $2r \leq m$, $a := (a_1, \ldots, a_r)$ with $a_1, \ldots, a_r > 0$ and σ be a permutation of $(1, \ldots, m)$. Then the* constrained-size-and-shape space *is given by*

$$S\Sigma_3^m(\sigma, a) := \{[Y] \in S\Sigma_3^m : Y = (y_1, \ldots, y_m) \in \mathbb{R}^{3\times m}, \\ \|y_{\sigma(j)} - y_{\sigma(j+r)}\| = a_j \text{ for } j = 1, \ldots, r\}\,.$$

An orthogonal projection from Σ_3^m to $\Sigma_3^m(\sigma, a)$ can be given explicitly as the following theorem teaches.

Theorem 2.3. *Let $r \in \mathbb{N}$ with $2r \leq m$, $a = (a_1, \ldots, a_r)$ with $a_1, \ldots, a_r > 0$, $[Z] \in S\Sigma_3^m$ with centered $Z \in (z_1, \ldots, z_m)$, i.e. $z_1 + \ldots + z_m = 0$ and σ be a permutation of $(1, \ldots, m)$. Then $Y^* = (y_1^*, \ldots, y_m^*)$ with*

$$\begin{aligned} y^*_{\sigma(j)} &= \beta_{\sigma(j)} z'_{\sigma(j)} + (1 - \beta_{\sigma(j)}) z'_{\sigma(j+r)}\,, \\ y^*_{\sigma(j+r)} &= (1 - \beta_{\sigma(j)}) z'_{\sigma(j)} + \beta_{\sigma(j)} z'_{\sigma(j+r)}\,, \text{ with} \\ \beta_{\sigma(j)} &= \frac{1}{2}\left(1 + \frac{a_j}{\|z'_{\sigma(j)} + z'_{\sigma(j+r)}\|}\right), \end{aligned}$$

for $j = 1, \ldots, r$ where we set $z'_{\sigma(j)} := z_{\sigma(j)}$, $z'_{\sigma(j+r)} := z_{\sigma(j+r)}$, if $z_{\sigma(j)} \neq z_{\sigma(j+r)}$ and $z'_{\sigma(j)} := z_{\sigma(j)} + v_j$, $z'_{\sigma(j+r)} := z_{\sigma(j)} - v_j$ if $z_{\sigma(j)} = z_{\sigma(j+r)}$ with an arbitrary nonzero vector $v_j \in \mathbb{R}^{3\times m}$, and, furthermore

$$y^*_{\sigma(j)} = z_{\sigma(j)} \text{ for } j = 2r+1, \ldots, m\,,$$

gives an orthogonal projection

$$[Y^*] \in \operatorname*{argmin}_{[Y] \in S\Sigma_3^m(\sigma,a)} d_{S\Sigma_3^m}([Z], [Y])\,.$$

The orthogonal projection is unique if $z_{\sigma(j)} \neq z_{\sigma(j+r)}$ for all $j = 1, \ldots, r$.

Proof. W.l.o.g. assume that σ is the identity. Furthermore, note that by construction Y^* is centered as Z is centered.

Every orthogonal projection is a minimizer of the Lagrange function

$$\mathcal{L}(Y, \lambda_1, \ldots, \lambda_r) = \|Y - Z\|^2 + \sum_{j=1}^{r} \lambda_j \left(\|y_{j+r} - y_j\|^2 - a_j^2 \right)$$

incorporating proximity of $Y = (y_1, \ldots, y_m)$ to Z and the constraining conditions. All of its critical points Y^* are determined by the equations

$$\begin{aligned} y_j^* - z_j &= \lambda_j (y_{j+r}^* - y_j^*) \text{ for } j = 1, \ldots, r & (5) \\ y_{j+r}^* - z_{j+r} &= -\lambda_j (y_{j+r}^* - y_j^*) \text{ for } j = 1, \ldots, r & (6) \\ y_j^* &= z_j \text{ for } j \in \{2r+1, \ldots, m\}\,. \end{aligned}$$

Notably, the last equations yield the unique minimizers of the non-constrained landmarks. Now fix $j \in \{1, \ldots, r\}$ and subtract (6) from (5) to obtain

$$(y_j^* - y_{j+r}^*)(1 + 2\lambda_j) = z_j - z_{j+r}\,. \tag{7}$$

If $z_j \neq z_{j+r}$ then (5) yields

$$y_j^* = z_j - \frac{\lambda_j}{1 + 2\lambda_j}(z_j - z_{j+r})\,,$$

i.e. with $\beta_j = \frac{1+\lambda_j}{1+2\lambda_j}$

$$y_j^* = \beta_j z_j + (1 - \beta_j) z_{j+r}\,, \tag{8}$$

and similarly, (6) yields

$$y_{j+r}^* = z_{j+r} + \frac{\lambda_j}{1 + 2\lambda_j}(z_j - z_{j+r})\,,$$

i.e.

$$y_{j+r}^* = (1 - \beta_j) z_j + \beta_j z_{j+r}\,. \tag{9}$$

This implies at once that

$$\begin{aligned} \|y_j^* - z_j\|^2 + \|y_{j+r}^* - z_{j+r}\|^2 &= 2(1 - \beta_j)^2 \|z_j - z_{j+r}\|^2 \\ &= \frac{2\lambda_j^2}{(1 + 2\lambda_j)^2} \|z_j - z_{j+r}\|^2\,. \end{aligned} \tag{10}$$

In order to determine λ_j we exploit the constraining condition to obtain from (7) that $|1 + 2\lambda_j| = \frac{\|z_{j+r} - z_j\|}{a_j}$. The cases of $1 + 2\lambda_j > 0$ and $1 + 2\lambda_j < 0$ correspond to

$$\lambda_j = \frac{1}{2}\left(\frac{\|z_{j+r} - z_j\|}{a_j} - 1 \right) \text{ and } \lambda_j = -\frac{1}{2}\left(\frac{\|z_{j+r} - z_j\|}{a_j} + 1 \right)$$

respectively, so that, taking into account (10), $\mathcal{L}$ assumes the minimal value for the positive branch yielding

$$\beta_j = \frac{1}{2}\left(1 + \frac{a_j}{\|z_{j+r} - z_j\|} \right)\,,$$

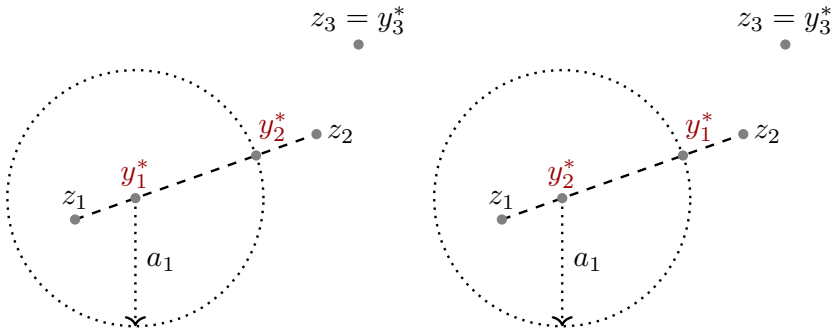


Figure B4: *Planar representation of the o.g. projection of* $Z = (z_1, z_2, z_3)$ *to the constraint* $\|y_1 - y_2\| = a_1$. *Left: The global minimum determined by* $1 + 2\lambda_1 > 0$ *is attained for* y_1^* *and* y_2^* *balanced between* z_1 *and* z_2. *Notably, fixing* $y_1 = y_1^*$ *the constrained* y_2 *is confined to a sphere of radius* a_1, *centered at* y_1^*. *Right: Swapping the minimal* y_1^* *and* y_2^* *from the left side corresponds to* $1 + 2\lambda_1 < 0$. *Fixing* $y_2 = y_2^*$, *the constrained* y_1 *lies on a sphere of radius* a_1, *centered at* y_2^*, *for which* $y_1 = y_1^*$ *produces a local maximum.*

as asserted. Moreover, then the above equations (8) and (9) yield the asserted landmarks for y_j^* and y_{j+r}^* in case of $z_j \neq z_{j+r}$.

If $z_j = z_{j+r}$ adding (6) to (5) yields

$$y_j^* + y_{j+r}^* \quad = \quad \frac{z_j}{2}\,,$$

which, taking into account the constraining condition, is solved by

$$y_j^* = z_j + a_j \frac{v_j}{2\|v_j\|}, \quad y_{j+r}^* = z_j - a_j \frac{v_j}{2\|v_j\|}$$

with an arbitrary nonzero vector $v_j \in \mathbb{R}^{3\times m}$. Then the above argument, after replacing z_j with $z_j' := z_j + v_j$ and $z_{j+r}' := z_j - v_j$ above in (5) and (6), yields the asserted equations.

We note that we have indeed found a minimum, for we can reparametrize the matrix Y by arbitrary $Y' := (y_1, \ldots, y_r, y_{2r+1}, \ldots, y_m) \in \mathbb{R}^{3\times(m-r)}$, and by $(w_1, \ldots w_r)$, each w_j arbitrary on the compact sphere $\{w \in \mathbb{R}^3 : \|w\| = 1\}$ which model the constraining conditions via $y_{j+r} = y_j + a_j w_j$ for $j = 1, \ldots, r$. Along the columns of Y' there is a unique minimum and along each of the w_j $(j = 1, 2)$ there is a maximum and a minimum given by the two choices of λ_j as detailed above and illustrated in Figure B4, and each such minimum is unique if $z_j \neq z_{j+r}$.

Finally, we claim that Y^* is already in optimal position to Z. In fact it suffices to see this for two landmarks only z_j, z_{j+r} $(1 \leq j \leq r)$ and y_j^*, y_{j+r}^* from (8) and (9). Indeed, in case of $z_j \neq z_{j+r}$, with the 3×3 unit matrix I, minimizing

$$\|z_j - Ry_j^*\|^2 + \|z_{j+r} - Ry_{j+r}^*\|^2 = \|(I - \beta R)z_j - (1-\beta)Rz_{j+r}\|^2 + \|(I - \beta R)z_{j+r} - (1-\beta)Rz_j\|^2$$

over $R \in SO(m)$ can be cast into the two dimensional complex problem with $z = z_j, w = z_{j+r} \in \mathbb{C}$, $\beta = \beta_j > 1/2$ minimizing

$$\begin{aligned} &|(1 - \beta e^{i\alpha})z - (1-\beta)e^{i\alpha}w|^2 + |(1 - \beta e^{i\alpha})w - (1-\beta)e^{i\alpha}z|^2 \\ &\quad = \quad (|z|^2 + |w|^2)\left(1 + \beta^2 - 2\beta\cos\alpha + (1-\beta)^2\right) - 4(1-\beta)\mathrm{Re}\,(\overline{z}\,w)\,(\cos\alpha - \beta) \end{aligned}$$

over $\alpha \in [0, 2\pi)$. Due to $0 \leq |z \pm w|^2 = |z|^2 + |w|^2 \pm 2\mathrm{Re}\,(\overline{z}\,w)$ and $\beta > 1/2$ this is minimized for $\alpha = 0$, corresponding to $R = I$ above.

In case of $z_j = z = z_{j+r}$, with arbitrary but fixed $v_j \in \mathbb{R}^3, \|v_j\| = 1$ such that $y_j^* = z + a_j v_j/2$ and $y_{j+r}^* = z - a_j v_j/2$, as above, we have similarly for $R \in SO(3)$ that

$$\begin{aligned} \|z_j - Ry_j^*\|^2 + \|z_{j+r} - Ry_{j+r}^*\|^2 &= \|z - R(z + a_j v_j/2)\|^2 + \|z - R(z - a_j v_j/2)\|^2 \\ &= 2\|z - Rz\|^2 + \frac{a_j^2}{2}\,, \end{aligned}$$

which is minimized by $R = I$. □

Remark 2.4. *The case $z_j = z_{j+r}$ has been discussed for exhaustive mathematical treatment. In the application in Section 4.3.1, this only happens if the neighborhoods in the classes learned feature degenerate Procrustes means, a clear sign that the learning algorithm failed. In this case we suggest to reevaluate learned classes, rather than choosing any v_j of suitable length.*

3 Multiscale Modeling of RNA Backbone Geometry, Clash Detection and Data Sets

Ribonucleic acid (RNA) molecules are composed of repeating elements called *nucleotides* and each nucleotide is composed of three building blocks, see Watson *et al.* (2004) and Figure B1: A sugar ring called *ribose* comprising 5 carbon atoms, one of 4 possible *nucleobases* which is attached to the ribose at the C1' position and a *phosphate group* connected to the ribose ring at the O5' atom. The single nucleotides are connected by their O3' atoms to the next phosphate group to form long RNA chains.

3.1 RNA Folding

In contrast to DNA which usually forms a double helix of complementary strands, in principle, RNA is single stranded and the form of its ribose (which is not "desoxy" as in DNA, i.e. it has an additional hydroxyl group) allows for complex folding structures. Figure B5 shows helical structures followed by mismatching sites: a *hairpin* in a 2D schematic and the 3D structure of the *frameshift stimulation element* of the SARS-CoV-2 genome proposed by Zhang *et al.* (2021). Its 2D schematic is depicted in the first panel of Figure B15.

3.2 Multiscale Modeling

In this section we describe the two scales modeled. Their surprising interaction which has lead to the two Hypotheses 4.1 and 4.2 underlying our method is detailed in Section 4.2.

On a *microscopic* scale, nucleotides are either studied as *suites*, i.e. from one sugar to the next, or as *residues*, i.e. from one phosphate to the next, e.g. Murray *et al.* (2003); Jain *et al.* (2015). As clashes often occur between neighboring residues but within same suites, cf. Murray *et al.* (2003), for our analysis, we use suites. Indexing, however, is usually done on residue level, so that within a single suite, atom indices change, cf. Figure B1. For the dihedral angles of concern, Figure B7 lists the 4 consecutive atoms, defining the respective dihedral angle of the bond between the two central atoms.

On the *mesoscopic* scale we additionally take the coordinates of the k preceding and k succeeding sugar rings into account. This can be seen as an intermediate scale between the microscopic suite scale and the macroscopic scale of a whole RNA strand.

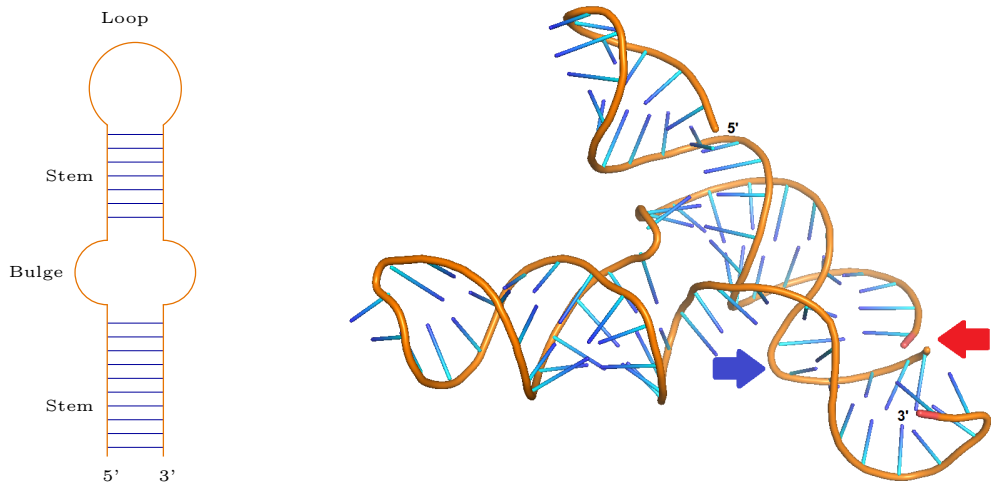


Figure B5: *Left: 2D schematic of the common* hairpin *structure: Double helices* (stems) *formed by bindings between matching nucleobases (blue) are followed by* mismatching *nucleobases* (bulges)*, not depicted, and a terminating mismatching site* (loop)*. Orientation is conveyed by the 5' and 3' ends. Right: One out of 10 proposed 3D RNA structures of the SARS-CoV-2 frameshift stimulation element by Zhang* et al. *(2021), graphically reproduced with PyMOL (Schrödinger, LLC, 2015) with backbone (orange) and nucleobases (blue), yielding helical structures whenever the latter point to each other. Arrows indicate suites with problematic (blue arrow, Suite 2 determined by Residues 33/34) and non-connected backbone (red arrow, Suite 1 determined by Residues 28/29) proposals discussed in Figures B3 and B15.*

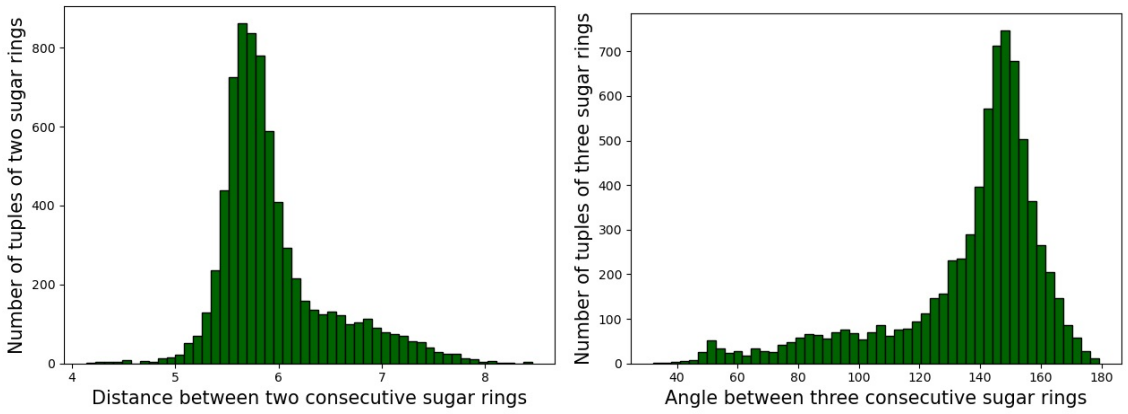


Figure B6: *Histograms of the distribution of distances between two successive sugar ring centers in Å(left) and of the distribution of angles in degrees spanned by three successive sugar ring centers (right).*

Notation 3.1. *We consider a connected RNA strand with $N \in \mathbb{N}$ consecutive nucleotides indexed by $i \in \{1, \dots, N\}$.*

Microscopic scale: *The i-th* suite *comprises the RNA region between a $C5'_i$ atom and the second next $O3'$ atom labeled $O3'_{i+1}$ and the* backbone shape *of the suite is described by the seven dihedral angles $(\delta_i, \epsilon_i, \zeta_i, \alpha_{i+1}, \beta_{i+1}, \gamma_{i+1}, \delta_{i+1}) \in \mathbb{T}^7$ for $i = 1, \dots, N-1$, cf. Figure B1.*

Mesoscopic scale: *As each nucleotide comes with a sugar ring formed by the atoms $C1'_i$,*

$C2'_i$, $C3'_i$, $C4'_i$ and $O4'_i$ (see Figure B1), denoting their centers of gravity (i.e. average location) with $\bar{\mathbf{x}}_\mathbf{i}$, for all $i = k+1, \ldots, N-k-1$, the mesoscopic strand *corresponding to the i-th suite is the* configuration matrix $X^{(i)} = (\bar{\mathbf{x}}_{\mathbf{i-k}}, \bar{\mathbf{x}}_{\mathbf{i-k+1}}, \ldots, \bar{\mathbf{x}}_{\mathbf{i+k+1}}) \in \mathbb{R}^{3\times(2k+2)}$. *Its size-and-shape in $S\Sigma_3^{2k+2}$ is called its* mesoscopic shape.

Indeed, geometric suite variability is solely governed by the dihedral angles, since bond lengths (distances between two consecutive atoms) and bond angles (angles between three consecutive atoms) are approximately constant due to the laws of chemistry, see e.g. Watson *et al.* (2004). In consequence, the geometry of the i-th suite is described, up to a proper Euclidean transformation (translation and rotation), by an element of the seven-dimensional torus $\mathbb{T}^7$ given by its seven dihedral angles.

angle	atom bonds
α	O3' - P - O5' - C5'
β	P - O5' - C5' - C4'
γ	O5' - C5' - C4' - C3'
δ	C5' - C4' - C3' - O3'
ϵ	C4' - C3' - O3' - P
ζ	C3' - O3' - P - O5'

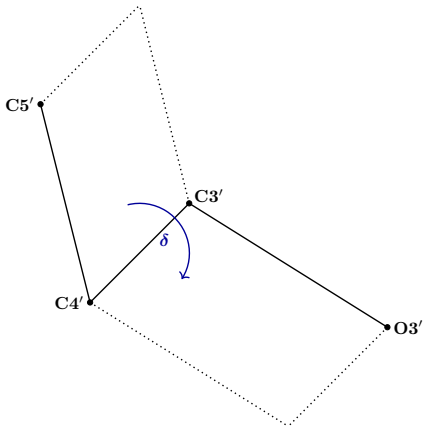


Figure B7: Left: *Names (first column) of dihedral angles along the two central atoms of the four atoms involved (second column), see Figure B1. Right: The dihedral angle δ of the bond between the atoms C4' and C3' is the directed angle between the plane spanned by the atoms C5', C4', C3' and the plane spanned by C4', C3', O3'. More precisely, it is the angle determined by turning the vector normal to the plane spanned by C3', C4', C5' to the vector normal to the plane spanned by O3', C3', C4' (with fixed orientation of normals determined by the order of spanning points).*

Since distances between two neighboring sugar rings and angles between three consecutive sugar rings vary due to folding at microscopic scale, see Figure B6, dihedral angles defined by four consecutive sugar rings are not sufficient to completely define the geometry of mesoscopic strands up to proper Euclidean transformations. The size-and-shape representation, modeling geometric landmark configurations determined by central positions of sugar rings modulo translation and rotation, however, suffices.

Remark 3.2. *For the mesoscopic strands we include the sugar ring centers of the $k = 2$ suites preceding and the $k = 2$ suites following the suite of concern, cf. Figure B2. This choice of k presents a trade-off, since a small k emphasizes the central, potentially faulty, suite and a large k leads to a great variety of shapes at transitions between secondary structure elements. For a given mesoscopic shape, this reduces the number of potentially similar mesoscopic shapes. Empirically, $k = 2$ yields a good balance between these two effects by modelling the local geometry at an intermediate (mesoscopic) scale. On the side of biochemistry, the $5 + 1 = 6$ bases from the $2k + 1 = 5$ suites correspond roughly to the number of bases involved in a half helix turn, see e.g. Watson* et al. *(2004). For future applications we anticipate that involving more scales by suitably choosing larger k will prove useful.*

We only work with suites that have a corresponding mesoscopic strand, i.e. we exclude the two suites at the end of an RNA strand.

Definition 3.3. *For an RNA strand of length $N \geq 2k+2$ the suites numbered $i = k+1, \ldots, N-k-1$ are called* admissible*, so that every admissible suite $\mathfrak{a}$ has a mesoscopic shape $m_{\mathfrak{a}} \in S\Sigma_3^{2k+2}$ and vice versa.*

Definition 3.4. *We call a suite a* **clash suite** *if two of its backbone atoms (including associated hydrogen atoms and oxygen atoms associated with the phosphate) clash with each other. All other suites that have $2k = 4$ neighboring non-clash suites (i.e. their mesoscopic strands have no within-suite-backbone-to-backbone clashes) are called* **clash free***.*

3.3 Cryo-EM, X-Ray Crystallography and Clash Detection

Cryo-EM (cryogenic electron microscopy) and X-ray crystallography are popular methods to determine atomic positions in RNA, protein and similar biomolecular structures, cf. Jain *et al.* (2015). For the former, molecules are shock frosted and subjected to electron microscopy. For the latter, using a suitable substrate, molecules are crystallized and subjected to X-ray imaging. The resolution of X-ray crystallography is defined as the smallest distance of two objects such that their diffraction patterns can be separated. In cryo-EM, resolution has been defined in various ways, usually via properties of the Fourier transformed electron density, in order to be comparable to the resolution values given for X-ray crystallography measurements. For a review, see Liao and Frank (2010). From different angles, via inverse Fourier transforms, the electron density can be reconstructed and, in principle, density peaks correspond to atom positions. Figure B8 shows exemplary level surfaces of electron densities with estimated atom positions.

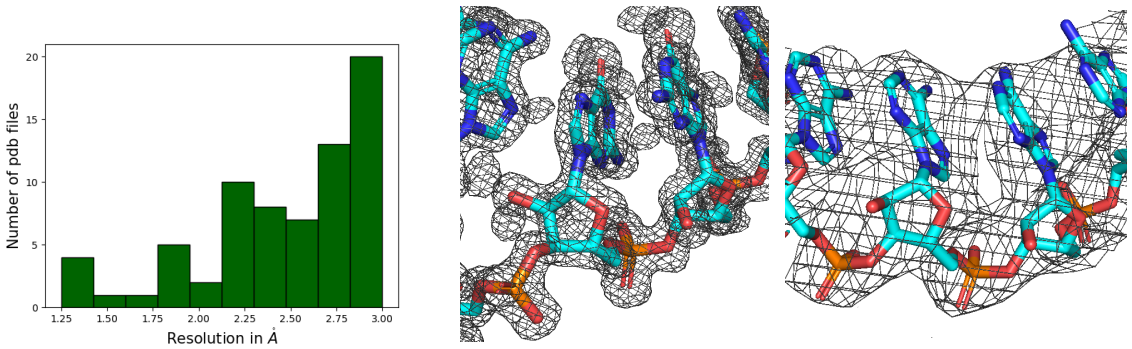


Figure B8: *Left: Histogram of X-ray crystallography resolutions in the benchmark data set from Section 3.4 below. Middle and right: reconstructed RNA structure and electron density contour surface created with PyMOL at level of one σ, see Schrödinger, LLC (2015), at resolution 1.6 Å (middle, from benchmark file 1csl, Ippolito and Steitz (2000), see Table B16 in the supplement) and at resolution 3 Å (right, from benchmark file 1f8v, Tang* et al. *(2001), see Table B16 in the supplement).*

At a resolution of 2.5 to 4$\mathring{A}$, which is typical for large RNA strands, base pairings can be predicted well and phosphates are well identified by strong peaks of density Jain *et al.* (2015). It is, however, more challenging to precisely estimate single atom positions along the backbone, see for example Murray *et al.* (2003). In addition, structural disorder due to crystallization and thermal oscillation contribute to uncertainties.

Since it is computationally not feasible to include the positions of all atoms and a full quantum chemical treatment into the fitting, the ambiguities in the measured density occasionally result in incompatible reconstructed atom positions. Indeed, our benchmark data set contains approximately 2.5 % clash suites.

The PHENIX (Python-based Hierarchical ENvironment for Integrated Xtallography) software by Liebschner *et al.* (2019) provides validation tools that detect such errors. Since hydrogen atoms are not visible in the electron density measurements (H-atoms contain only one electron which is shifted to the covalent-bond partner atom), first, the PHENIX tool `phenix.reduce` adds the hydrogen atoms. Then, `phenix.probe` performs an all-atom contact analysis (Word *et al.*, 1999), which declares atoms that are not bonded to each other as a *clash* if they are closer together than is physically possible (i.e. if van der Waals shells overlap by more than 0.4 Å). For each PDB file, `phenix.clashscore` generates a list of all clashes. From all of the different types of clashes detected, in this work we are only concerned with within-suite-backbone-to-backbone clashes as in Definition 3.4.

3.4 The Benchmark, Training and Test Data Sets

In our applications, we analyze a subset of a classical RNA data set. The classical RNA data set comprises 8665 suites, carefully selected for high experimental X-ray precision (of 3 Å= 0.3 nanometers) by Duarte and Pyle (1998); Wadley *et al.* (2007) and analyzed by them and by others, for example Murray *et al.* (2003); Richardson *et al.* (2008) and Eltzner *et al.* (2018). The data originate from 71 different measurements and the atomic positions of each measurement have been stored in the *PDB* format of a *protein data bank* file, online at the Protein Data Bank, see Berman *et al.* (2000). More details on the PDB files can be found in Table B16 of Supplement A.

From this classical data set, we consider the 7648 admissible suites (which have an associated mesoscopic strand, see Definition 3.3) and call this data set the *benchmark data set*.

Applying PHENIX as detailed in Section 3.3 to the benchmark data set, we obtain 5957 clash free suites that also have clash free mesoscopic strands (see Definition 3.4) and these form the *benchmark training data set* $\mathfrak{T}$. Figure B17 in the supplement gives a scatterplot at microscopic scale for all pairs of the seven dihedral angles.

From the remaining suites we chose those suites that feature within-suite-backbone-to-backbone clashes, forming the *benchmark test data set* $\mathfrak{C}$, containing 198 suites.

As our purpose lies in demonstrating our methods rather than correcting all clashes, all other suites (e.g. those not themselves clashing but featuring clashes in their mesoscopic strands) are disregarded in our analysis.

4 CLEAN-MINT-AGE

After classifying clash free suites by the MINT-AGE algorithm (Mardia *et al.*, 2022) from the benchmark training data set, we validate the classes obtained by comparing with the outcome of the clustering method by Richardson *et al.* (2008). Motivating our multiscale approach by analyzing clusters at two scales, then we propose and validate the CLEAN method classifying suites exploiting the observed relationship between the two scales.

4.1 Microscopic Classification and its Validation

We apply the non-supervised cluster learning method from Mardia *et al.* (2022) to the microscopic suite representations on the torus $\mathbb{T}^7$, of the benchmark training data set.

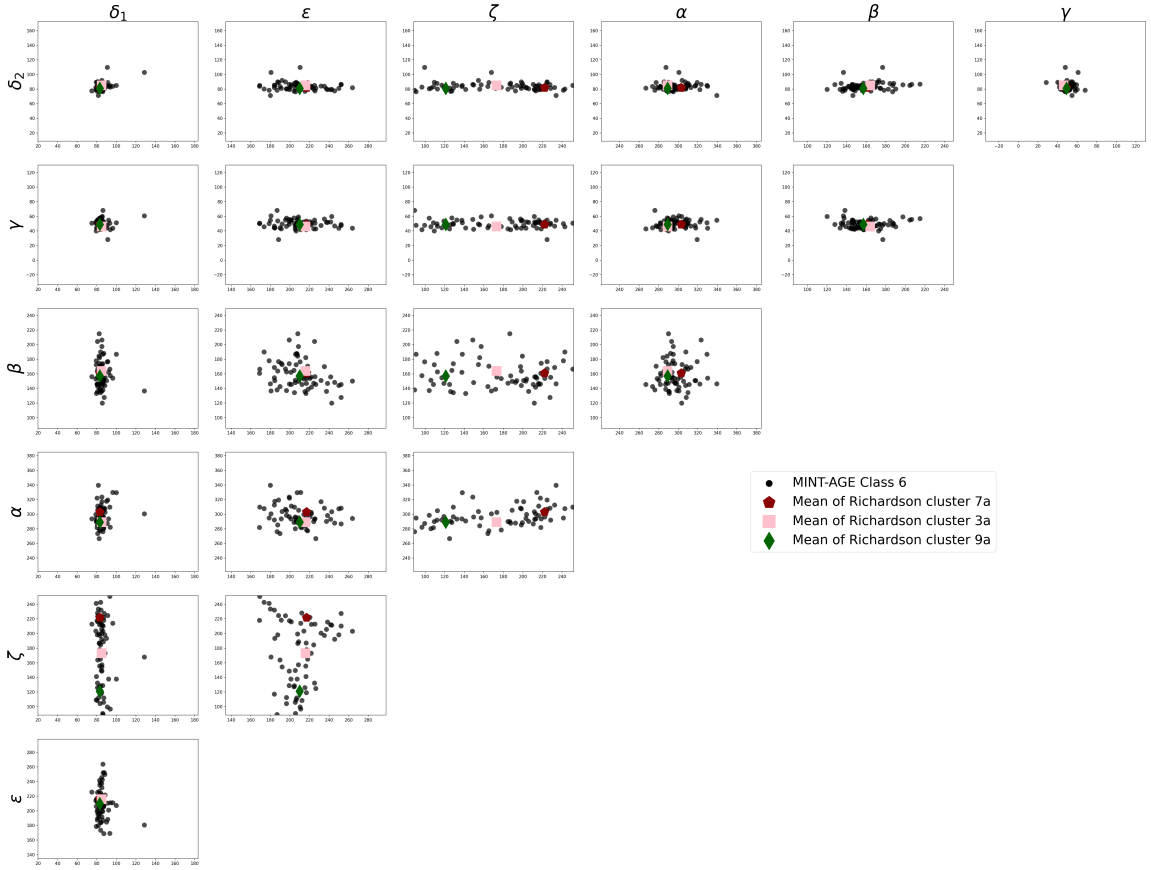


Figure B9: *Scatterplots of all two dimensional dihedral angle pairs (in degrees) of MINT-AGE Class 6 (black) and the reported means of Clusters 7a, 3a and 9a from Richardson* et al. *(2008).*

In brief, in a first step (AGE) it proposes preclusters based on an iterative, adaptive, average linkage clustering method for general metric spaces, that allows to detect clusters of different densities and sizes. In a second step (MINT), each precluster is subjected to torus PCA (see Eltzner *et al.* (2018)) and its projection to its main one-dimensional component is subjected to circular mode hunting, so that each statistically significant antimode corresponds to a post-cluster boundary. For convenience the MINT-AGE (Mode huntINg on Torus pca post iterative Adaptive linkaGe clustEring) algorithm is reproduced in supplement Section C including a discussion of parameters and our choices. Its general version is described in Mardia *et al.* (2022).

As discussed in detail in Eltzner *et al.* (2018), performing PCA analogs on non-Euclidean manifolds may be challenging, in particular on a torus: tangent space PCA (e.g. Fletcher *et al.* (2004)) misses data periodicity, intrinsic PCA (see Huckemann and Ziezold (2006)) produces geodesics winding infinitely often around, each of which approximating all possible data perfectly, and restricting winding numbers (e.g. Altis *et al.* (2008); Kent and Mardia (2009, 2015)) greatly reduces flexibility. In contrast on spheres, *principle nested spheres* (PNS, by Jung *et al.* (2012)) is a PCA analog that is even more flexible and this flexibility persists on suitably stratified spheres which represent the torus in *torus PCA* (see also Mardia *et al.* (2022)): On the m-dimensional sphere, the dimension of the family of main principal nested circle components is $3(m-1)$, while the dimension of the family of first PCs for data on an m-dimensional Euclidean space is dimension $2(m-1)$.

This feature is advantageous for PCA-based clustering, since clusters that would require two Euclidean PCs to be separated can often be separated along the main principal nesting circle.

MINT-AGE	Size	Richardson *et al.* (2008)
1*	3933	1a, (1m), (1L), (&a)
2*	294	1c
3	203	1b, 1[
4*	93	1g
5	91	2a
6	67	7a, 3a, 9a
7	64	0a, (4a)
8	50	
9	46	1e
10*	40	5z
11	37	6p
12	31	2[
13	29	0i, 6n
14	29	4b, (0b)
15	23	
16*	23	6g
17	23	4g
Outliers	881	
Total	5957	

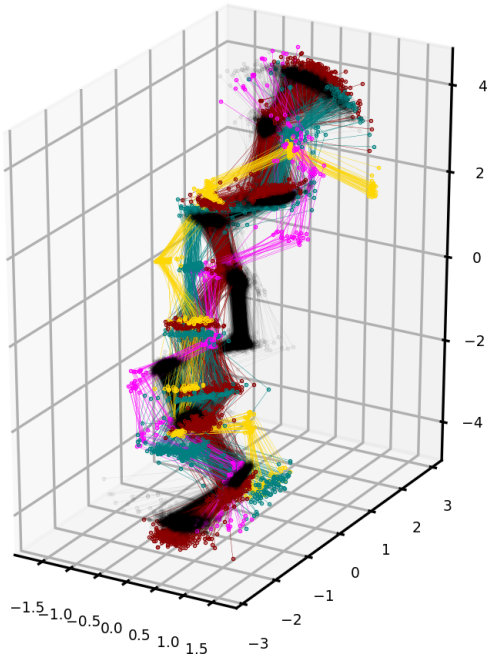


Figure B10: *Left: MINT-AGE class numbers and outliers (left column) with size (middle column) from the benchmark training data set with corresponding two-character cluster names (a number for the first character and a letter or "[" for the second character) from Richardson* et al. *(2008). Asterisks mark MINT-AGE classes displayed in the right panel. Right: Five exemplary classes that can be well displayed together at microscopic scale: Class 1 (black), class 2 (red), class 4 (turquoise), class 10 (yellow), class 16 (magenta). Parentheses indicate that Richardson* et al. *(2008) cluster means are at boundaries of MINT-AGE classes.*

Application of MINT-AGE to the benchmark training data set yields 17 classes. The largest corresponding to the A helix shape contains 3933 elements and is highly dominant. All classes are rather dense and even the smallest has a credible size of 21 elements. The number of outliers (881), however, is quite large. We conjecture that a considerable number of these are due to incorrect structure reconstructions, which have not been detected because they have not led to clashes. Figure B18 in the supplement displays all classes in dihedral angle representation.

The table in Figure B10 compares our MINT-AGE classes with clusters found by Richardson *et al.* (2008, Table 2) in a larger set encompassing the benchmark training data set. As they report every cluster only by its mean dihedral angles, we have manually assigned these means to MINT-AGE classes. Typically, Figure B9 illustrates how three Richardson *et al.* (2008) cluster means are assigned to MINT-AGE Class 6. This larger data set and allowing some clusters with less than 10 elements has lead to a larger number of 46 Richardson *et al.* (2008) clusters. Remarkably, more than half (24) of them can be assigned to MINT-AGE clusters and among the ones that could not be assigned, only two have more than 20 elements (7p with 27 elements and 8d with 24).

4.2 Motivation for a Multiscale Ansatz

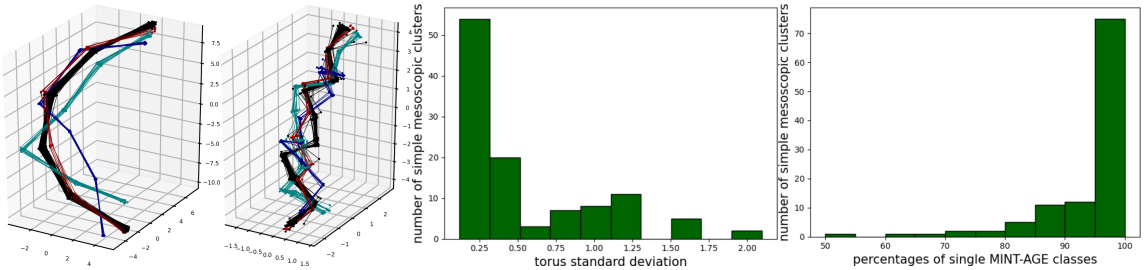


Figure B11: *Left: Four exemplary* simple *mesoscopic clusters at mesoscopic scale. Center left: Their central suites at microscopic scale. Simple mesoscopic Cluster 1 (black) of size 77 contains 73 suites from MINT-AGE Class 1, all of the others clusters are in 1-to-1 correspondence to MINT-AGE classes: Cluster 30 (turquoise, size 13) to Class 4, Cluster 55 (blue, size 8) to 7 and Cluster 92 (red, size 6) to Class 2. Center right: Binned torus (angular) standard deviation of the suites belonging to simple mesoscopic clusters. For instance, the suites of Cluster 1 from the two left panels have a standard deviation of* 0.83*, so that Cluster 1 is counted in the 4th green bar from the left. Right: Percentages of the largest MINT-AGE class in each cluster. For instance the rightmost bar indicates that for 75 out of the 110 clusters at least* 95 % *of their suites belong to a single MINT-AGE class.*

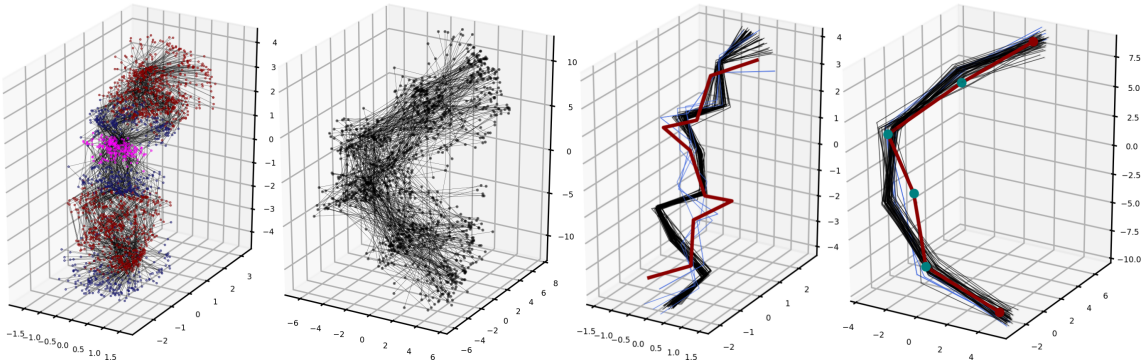


Figure B12: *Left: The 198 clash suites from the benchmark data set in Section 3.4 with carbon (dark red), oxygen (dark blue) and phosphorus atoms (pink), cf. Figure B1 at microscopic scale. Left center: Same at mesoscopic scale. Right center: At microscopic scale a typical clash suite* $\mathfrak{c}$ *(red), the* 46 *suites (black) from the dominant MINT-AGE class and the other 4 suites (blue) in the neighborhood* $U_{\mathfrak{c}}$ *with respect to mesoscopic shape distance, see also Figure B3. Right: Same at mesoscopic scale where shapes are highly concentrated. The landmarks (teal) of the clash suite at mesoscopic scale (red), except for the middle one, require only very moderate correction.*

In a first fundamental study, we establish a relationship between suites that have similar mesoscopic shapes, see Figure B2 and Notation 3.1. To this end, we cluster the mesoscopic shapes of the suites of the benchmark training data set (Section 3.4) using the *simple version* of AGE from Mardia *et al.* (2022) (Algorithm C.1 from Supplement C.1, performing only Steps 1 and 2 with $\kappa = 5$ and $d_{\max}$ such that 50% of the mesoscopic strands are outliers) yielding the *simple mesoscopic clusters*. By design, we obtain many (110) clusters that are rather concentrated. It turns out that

1. the suites corresponding to each simple mesoscopic cluster also form rather concen-

trated suite clusters: for most, the standard deviation of angles (between 0 and 2π) of their suites is less than 0.6 and only very few clusters with low cardinality (close to the minimum of $\kappa + 1 = 6$) have higher suite standard deviation (Figure B11, third panel);

2. simple mesoscopic clusters are in high correspondence with the 17 MINT-AGE classes from Section 4.1 above as clearly visible in in the rightmost panel of Figure B11 and detailed for exemplary clusters in the caption of Figure B11.

This leads to the following hypothesis.

Hypothesis 4.1. *Correctly reconstructed suites with similar mesoscopic shapes have also similar suite shape. In particular, concentrated mesoscopic clusters relate to suite classes.*

In a second fundamental study, we consider the 198 clash suites in the benchmark data set forming the test data set, see Section 3.4. Their suite shapes as well as their mesoscopic shapes feature a rather larger spread, see Figure B12 (first two panels). As before, we consider training suites from concentrated neighborhoods in the mesoscopic shape space, of size $\rho \in \mathbb{N}$. For a given clash suite $\mathfrak{c}$ such a neighborhood is

$$U_{\mathfrak{c}} \;:=\; \Big\{\mathfrak{t} \in \mathfrak{T} : \#\big\{\mathfrak{t}' \in \mathfrak{T} : d_\Sigma(m_{\mathfrak{t}'}, m_{\mathfrak{c}}) \leq d_\Sigma(m_{\mathfrak{t}}, m_{\mathfrak{c}})\big\} \leq \rho\Big\}. \tag{11}$$

The neighborhood $U_{\mathfrak{c}}$ is the set of the ρ suites of the training data, whose mesoscopic shapes are most similar to $m_{\mathfrak{c}}$ with respect to mesoscopic shape space distance. Recall from Section 3.4 that $\mathfrak{T}$ is the set of training suites (the clash free suites in the benchmark data set) and that $m_{\mathfrak{t}}$ denotes the mesoscopic shape of $\mathfrak{t} \in \mathfrak{T}$. On close inspection of the 198 $U_{\mathfrak{c}}$'s we find a situation typically illustrated in the last two panels of Figure B12, which leads to the following hypothesis.

Hypothesis 4.2. *While at microscopic scale, clash suite shapes are rather irregular among the suite shapes of their clash free neighbors, at mesoscopic scale, their mesoscopic shapes differ only mildly from nearby clash free mesoscopic shapes.*

The theoretical argument underlying this hypothesis is that even drastic errors on the atomic suite scale can still be compatible with electron density measurement results due to finite resolution, while drastic errors on the mesoscopic scale are excluded since they would strongly contradict the measured electron density. Indeed, we find empirically at mesoscopic scale that only one of the four teal landmarks in the middle (FigureB12, right panel) differs more strongly from the neighboring clash free mesoscopic shapes in $U_{\mathfrak{c}}$. For all 198 clash shapes the histogram in Figure B13 shows that for the vast majority of clash suites $\mathfrak{c} \in \mathfrak{C}$, the distance (detailed in Section 4.4) of its mesoscopic shape to its clash free correction is only rarely barely above and mostly well below the resolution order.

Remark 4.3. *There are databases that store different RNA motifs and their interaction: In* RNA Bricks *(Chojnowski* et al.*, 2013), the elements of simple mesoscopic Clusters 1 and 92 are often found in a* stem cluster *(corresponding to helical backbone shapes) and the elements of simple mesoscopic Cluster 30 are found in a* loop cluster*. Similarly, in Petrov* et al. *(2013), the elements of simple mesoscopic Cluster 30 are classified in the* hairpin loop *with the name HL_43074.14.* Stems *and* loops *are depicted in the* hairpin *structure scheme in the left panel of Figure B5.*

4.3 The Multiscale RNA Backbone Structure Correction Prodecure

Exploiting the above Hypotheses 4.1 and 4.2, the following multiscale backbone correction procedure simultaneously corrects clashing suites at microscopic and at mesoscopic scale, working with concentrated neighborhoods as in (11), defined by mesoscopic shape distance. In these concentrated neighborhoods, dominating classes from MINT-AGE of the training data set provide guidance for correction. Recall from the two left panels of Figure B3, with more detail in the two right panels of Figure B12, that even a minor correction of one of the sugar ring centers at mesoscopic scale can have great impact on the shape of the suite of interest, which is positioned between the third and fourth sugar ring at mesoscopic scale.

4.3.1 Multiscale Correction (CLEAN)

Input:

- a training data set $\mathfrak{T}$ comprising only clash free admissible suites (suites that feature a mesoscopic shape, see Definition 3.3),
- a list of classes $C_1, \dots, C_r$ and an outlier set R for $\mathfrak{T}$ obtained from applying the MINT-AGE algorithm (see Section 4.1 and Algorithm C.3 from the supplement),
- a clash suite $\mathfrak{c}$ and its corresponding mesoscopic shape $m_{\mathfrak{c}}$.
- the size $\rho \in \mathbb{N}$ of the neighborhood $U_{\mathfrak{c}}$ from (11), we choose $\rho = 50$ as roughly twice the size of the smallest class, and
- the flag `DOMINATING` set to `ABSOLUTE` or `RELATIVE` which will return either the absolutely dominating cluster in $U_{\mathfrak{c}}$ or the relatively dominating cluster with at least $\rho/10$ elements, taking into account cluster size (in Step (b) below).

Implementation steps:

1. Calculate
   (a) the neighborhood $U_{\mathfrak{c}}$ as defined in (11) of the ρ suites of the training data, whose mesoscopic shapes are most similar to $m_{\mathfrak{c}}$ with respect to mesoscopic shape space distance;
   (b) according to flag `DOMINATING`, the number
   $j_{\mathfrak{c}} \in \operatorname{arg\,max}_{j=1,\dots,m} \#\big(C_j \cap U_{\mathfrak{c}}\big)$, (`ABSOLUTE`), or
   $j_{\mathfrak{c}} \in \operatorname{arg\,max}_{j=1,\dots,m} \mathbf{1}_{\{C_j \mid \#(C_j \cap U_{\mathfrak{c}}) \geq \rho/10\}} \#\big(C_j \cap U_{\mathfrak{c}}\big)/\#C_j$,
   (`RELATIVE`), respectively, of the dominant MINT-AGE class in $U_{\mathfrak{c}}$;
   (c) a Fréchet mean $\tau_{\mathfrak{c}} \in \operatorname{argmin}_{\mathfrak{t} \in \mathbb{T}^7} \sum_{\mathfrak{t}' \in C_{j_{\mathfrak{c}}} \cap U_{\mathfrak{c}}} d_{\mathbb{T}^7}(\mathfrak{t}, \mathfrak{t}')^2$, of the dominant class' suites in the neighborhood;
   (d) the approximate length $\ell_{\tau_{\mathfrak{c}}}$ of the suite by the mean distance of the two central sugar rings $k+1$ and $k+2$ of the mesoscopic shapes corresponding to the suites of $C_{j_{\mathfrak{c}}} \cap U_{\mathfrak{c}}$;
   (e) a Procrustes mean

$$\mu_{\mathfrak{c}} \in \operatorname*{argmin}_{m \in S\Sigma_3^{2k+2}} \sum_{\mathfrak{t} \in C_{j_{\mathfrak{c}}} \cap U_{\mathfrak{c}}} d_{\Sigma}(m, m_{\mathfrak{t}})^2 ,$$

   of the corresponding mesoscopic shapes.

2. With a mesoscopic shape $m_{\mathfrak{c}} = [x_1, \ldots, x_{2k+2}]$ defined as in Equation (3) by a landmark configuration matrix $(x_1, \ldots, x_{2k+2})$, determine the corrected mesoscopic shape $m_{\tau_{\mathfrak{c}}}$ as the orthogonal projection of the size-and-shape Y^* of the Procrustes mean $\mu_{\mathfrak{c}} = [z_1, \ldots, z_{2k+2}]$ to the set

$$\left\{ m = [y_1, \ldots, y_{2k+2}] \in S\Sigma_3^{2k+2} : \|y_1 - y_{2k+2}\| = a_1,\ \|y_{k+2} - y_{k+1}\| = a_2 \right\} \quad (12)$$

of mesoscopic shapes whose configurations have distance $a_1 = \|x_1 - x_{2k+2}\|$ between the first and the last landmark given by that of any configuration of $m_{\mathfrak{c}}$ and whose distance a_2 between the central landmarks is the length $\ell_{\tau_{\mathfrak{c}}}$ which is chosen so that the Fréchet mean suite $\tau_{\mathfrak{c}}$ fits between them. By Theorem 2.3, with $m = 2k+2$, $r = 2$, $\sigma(1) = k+2$, $\sigma(k+1) = k+2$ and $\sigma(j) = j$ for $j \in \{2, \ldots, k, k+3, \ldots, 2k+1\}$, the (in practice there will no ties between the landmarks) desired orthogonal projection to $S\Sigma_3^{2k+2}(\sigma, a_1, a_2)$ which is the space determined by (12) is given by

$$\begin{aligned} y_1^* &= \alpha z_1 + (1-\alpha) z_{2k+2}\,, & y_{2k+2}^* &= \alpha z_{2k+2} + (1-\alpha) z_1 \\ y_{k+1}^* &= \beta z_{k+1} + (1-\beta) z_{k+2}\,, & y_{k+2}^* &= \beta z_{k+2} + (1-\beta) z_{k+1} \end{aligned}$$

where

$$\alpha = \frac{1}{2}\left(1 + \frac{\|x_{2k+2} - x_1\|}{\|z_{2k+2} - z_1\|}\right), \qquad \beta = \frac{1}{2}\left(1 + \frac{\ell_{\tau_{\mathfrak{c}}}}{\|z_{k+2} - z_{k+1}\|}\right)$$

and $y_j^* = z_j$ for $j \in \{1, \ldots 2k+2\} \backslash \{1, k+1, k+2, 2k+2\}$.

Output:

- the corrected suite shape $\tau_{\mathfrak{c}}$ and its corrected mesoscopic shape $m_{\tau_{\mathfrak{c}}} := [Y^*]$.

As mentioned above, we suggest to choose $\rho = 50$ as roughly twice the size of the smallest class. A larger value for ρ would make it very unlikely that the plurality of neighboring suites for a clash suite are from the smallest cluster, because any other nearby clusters will outnumber it. A smaller value for ρ would lead to less reliable results and, in some cases, to a majority of outliers in the set.

For many applications of CLEAN, setting DOMINATING = ABSOLUTE can be used as we do for analyzing two suites of SARS-CoV-2 RNA in the following Section 5. If considerably differing class sizes are of concern, setting DOMINATING = RELATIVE ensures assignment to smaller classes that dominate neighborhoods at mesoscopic scale only relatively to their total size. This results in greater diversity as illustrated in Figure B19 in the supplement, applying CLEAN to the entire benchmark test set from Section 3.4.

4.4 Validation of CLEAN

We apply the CLEAN method from Section 4.3.1 to the 198 clash suites which form the test data set $\mathfrak{C}$ from Section 3.4. For validation we confirm that backbone correction is realistic and neither arbitrary nor ambiguous. For the former, we verify that corrections happen on a scale not larger than the underlying X-ray crystallography resolution, see Section 3.3, and for the latter we verify that the largest MINT-AGE classes in neighborhoods $U_{\mathfrak{c}}$ from (11) are indeed strongly dominating in most cases.

In order to relate the amount of correction to resolution, consider the normalized Procrustes distance between the mesoscopic shape $m_{\mathfrak{c}}$ of a clash suite $\mathfrak{c} \in \mathfrak{C}$ and the mesoscopic shape $m_{\tau_{\mathfrak{c}}}$ of its correction by CLEAN,

$$\widetilde{d}_{\mathfrak{c}}^2 \;:=\; \frac{1}{\text{resolution}^2}\,\frac{3}{\text{degrees of freedom}}\,d_\Sigma(m_{\mathfrak{c}}, m_{\tau_{\mathfrak{c}}})^2\,. \tag{13}$$

Recalling that the group of 3D Euclidean transformations is of dimension 6, the *degrees of freedom* are given by $3(2k+2)-6 = 3\cdot 2k$, so that the inverse of the second quotient above gives the number $2k$ of free landmarks in Σ_3^{2k+2} taking into account that the resolution incorporates the spatial dimension 3.

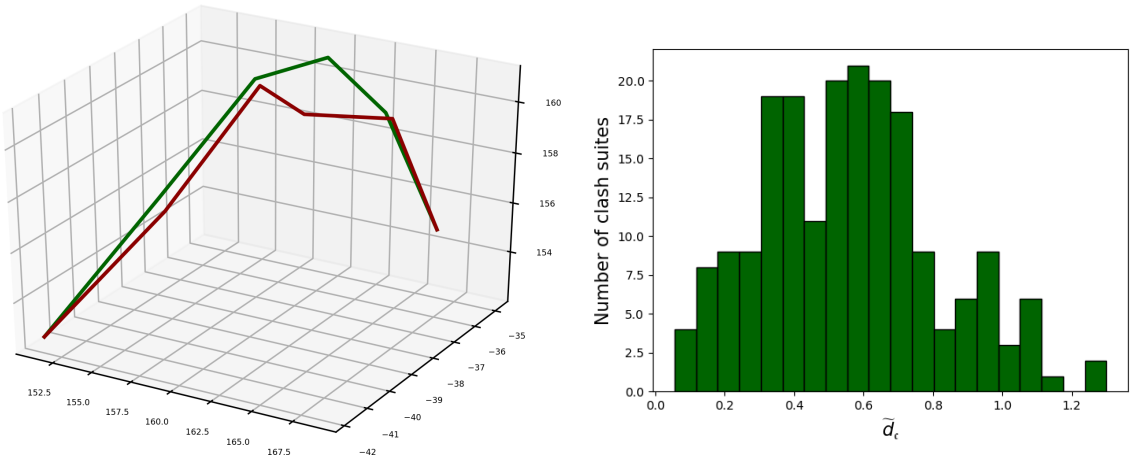


Figure B13: *Left: At mesoscopic scale a clash suite (red) and its mesoscopic shape (green) corrected by CLEAN. Right: Histogram of relative distances $\widetilde{d}_{\mathfrak{c}}$ between corrected mesoscopic shapes and original mesoscopic shapes from (13) over all $\mathfrak{c} \in \mathfrak{C}$.*

The histogram in Figure B13 shows that for the vast majority of clash suites $\mathfrak{c} \in \mathfrak{C}$, $\widetilde{d}_{\mathfrak{c}}$ is smaller than 1. Thus, corrections are only rarely slightly above and mostly well below the order of resolution.

In order to assess how dominating torus MINT-AGE classes are in neighborhoods $U_{\mathfrak{c}}$ ($\mathfrak{c} \in \mathfrak{C}$), the histogram in Figure B14 shows the number of suites in the dominating classes $C_{j_{\mathfrak{c}}}$. Indeed, for considerably more than half of the neighborhoods, the dominating cluster contains more than half of the neighboring suites. Remarkably, the negative correlation visible in the scatter plot in Figure B14 (right) shows that a smaller amount of correction tends to correlate with more elements being in the dominating cluster.

5 Application to SARS-CoV-2 Suites

With the recent worldwide pandemic of the *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2), the virus' RNA structure reconstruction and backbone correction has become ever more relevant. Indeed, effective drug and vaccine development necessitates good understanding of the three-dimensional RNA structure, see Croll *et al.* (2021). Recently, a large number of measurements has been added to the Protein Data Bank, see Berman *et al.* (2000), and as part of the *Coronavirus Structural Task Force* (CSTF) headed by Andrea Thorn, a large number of data sets of SARS-CoV-2 and related structures are compiled in a git repository, see Thorn *et al.* (2021). While X-ray crystallography can achieve very high resolution in principle, the large viral genome, comprising $\sim 20,000$ bases, is very

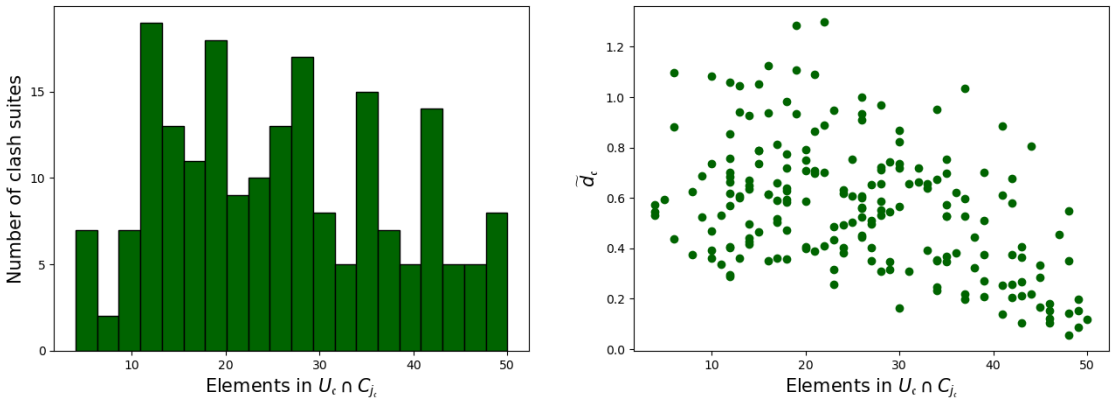


Figure B14: *Left: Histogram of the number of suites in* $U_{\mathfrak{c}}$ *from (11), over all (198) clash suites* $\mathfrak{c} \in \mathfrak{C}$ *(test data set), of the dominating MINT-AGE class. Right: Scatter plot relating the number of suites in the dominating class of* $U_{\mathfrak{c}}$ *and the normalized distance* $\widetilde{d}_{\mathfrak{c}}$ *from (14), over* $\mathfrak{c} \in \mathfrak{C}$.

difficult to crystallize. Therefore, many structures are determined by cryogenic electron microscopy (cryo-EM).

5.1 The Frameshift Stimulation Element

In Zhang *et al.* (2021), the frameshift stimulation element of the SARS-CoV-2 genome was studied (see Figure B5, right panel), which, due to its *slippery site* encodes different proteins simultaneously (this method of information compression is shared with other viruses such as HIV-1). As their balanced expression is required for virus replication, this element is believed to be fairly resistant against mutations. Hence it is a promising target for antiviral drug design. Its three-dimensional structure has been assessed by cryo-EM with a resolution of 6.9Å using the ribosolve pipeline from Kappel *et al.* (2020), see also Section 3.3. Using a consensus secondary structure of the molecule and the cryo-EM map, Zhang *et al.* (2021) proposed 10 possible three-dimensional structure models (based on a measurement with mean pairwise root mean squared deviation of 5.68Å) and stored them to the Protein Data Bank. Notably, it was not possible to reliably assign individual atom positions, but the secondary arrangement of helical segments and the non-helical linking segments could be reconstructed, see Zhang *et al.* (2021) and first panel of Figure B15. In particular, the suites linking different helical segments have been difficult to reconstruct. Here we focus on the suite determined by Residues 28/29 which we call *Suite 1* and on the suite determined by Residues 33/34 which we call *Suite 2* (referring to enumeration in the PDB file).

5.2 Reconstructing Suite 1

Suite 1 (red arrow in Figure B5, right panel, and the left red dot in Figure B15, left panel) is a clash suite in all 10 models proposed by Zhang *et al.* (2021), as determined by PHENIX, see Section 3.3. Notably, the P'-O3' bonds are unphysically long (red verticals in Figure B3, center right panel), hinting towards a bad structure fit of all 10 proposals. Figure B15 (3rd panel) shows $\mathfrak{c}_1$, the first clashing proposal for Suite 1, at mesoscopic scale and its highly concentrated neighborhood $U_{\mathfrak{c}_1}$ from (11), in which 43 out of the 50 suites belong to MINT-AGE Class 4. Its torus mean and $\mathfrak{c}_1$ at microscopic scale are shown in Figure B15

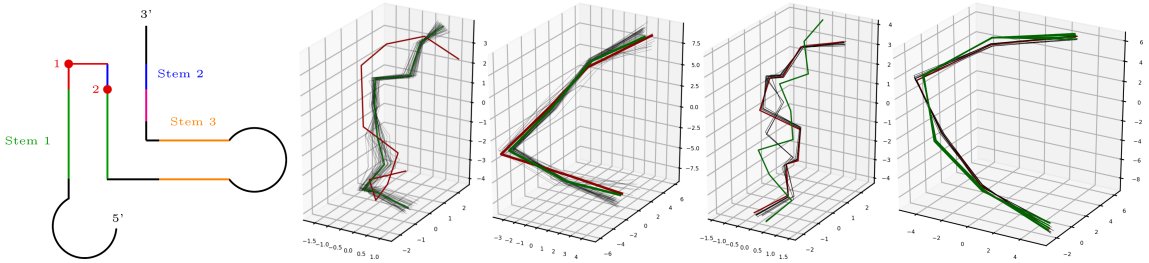


Figure B15: *First: 2D scheme of the SARS-CoV-2 frameshift stimulation element, adapted from (Zhang* et al.*, 2020, Figure 8), see also Figure B5 (right panel), with double-stranded helical stems (green, yellow and blue) and connecting Suites 1 and 2 (red dots, only one nucleotide is on the red branch) Second: Model 1 ($\mathfrak{c}_1$) of Suite 1 (red, clashing) proposed by Zhang* et al. *(2021), the* 43 *suites from MINT-AGE Class 4 dominating in neighborhood $U_{\mathfrak{c}_1}$ (black) and their torus mean (green) at microscopic scale. Third: The corresponding mesoscopic shape $m_{\mathfrak{c}_1}$ (red), the* 43 *mesoscopic shapes of suites from MINT-AGE Class 4 in neighborhood $U_{\mathfrak{c}_1}$ (black) and their Procrustes mean geodesically projected to a mesoscopic shape featuring length constraints from $m_{\mathfrak{c}_1}$ and the microscopic correction of $\mathfrak{c}_1$ (green). Fourth: The ten different model proposals (one in red clashes, the others in black form two clusters) by Zhang* et al. *(2021) of Suite 2 and their highly consistent correction from MINT-AGE Class 1 (green) at microscopic scale. Right: Using same coloring, at mesoscopic scale all 10 models from Zhang* et al. *(2021) of Suite 2 form one cluster for which CLEAN-MINT-AGE provides a moderate correction only.*

(2nd panel). The situation is very similar for the other clashing proposals $\mathfrak{c}_2, \ldots, \mathfrak{c}_{10}$ for Suite 1: MINT-AGE Class 4 dominates strongly in their concentrated neighborhoods, each warranting only minor corrections at mesoscopic scale (Figure B3, 4th panel) and all of their torus means at microscopic scale are nearly indistinguishable (Figure B3, 3rd panel). Notably, MINT-AGE Class 4 corresponds to one Richardson *et al.* (2008) cluster only (namely 2a, see Figure B10) which has been characterized there as *GNRA 1-2; U-turn.*

5.3 Reconstructing Suite 2

Suite 2 (blue arrow in Figure B5, right panel, and the right red dot in Figure B15, left panel) is a clash suite only in one out of the 10 models proposed by Zhang *et al.* (2021), as determined by PHENIX, see Section 3.3. At microscopic scale (Figure B15, fourth panel, red and black) these models are inconclusive as they feature two different clusters and one of the models (red) from the larger cluster has a clash score 0.401Å, slightly above the threshold of 0.4Å. As before, at mesoscopic scale (Figure B15, fifth panel, red and black), the shapes of all 10 models proposed are very similar and consistent and there is a single MINT-AGE class that strongly dominates every neighborhood (11), namely Class 1. Figure B15 (fifth panel, green) shows its Procrustes means projected to the mesoscopic shapes featuring length constraints from the corresponding mesoscopic shapes $m_{\mathfrak{c}_1}, \ldots, m_{\mathfrak{c}_{10}}$ of the 10 models and the suite lengths of the corrections from $\mathfrak{c}_1, \ldots, \mathfrak{c}_{10}$ as detailed in Section 4.3.1. In consequence, the CLEAN-MINT-AGE corrections are the torus means of the suites of Class 1 in the respective neighborhoods. Again these are nearly indistinguishable, giving one consistent correction for Suite 2 in Figure B15 (fourth panel, green).

6 Discussion

The CLEAN-MINT-AGE procedure presented here, yielding

1. hierarchical (different shape spaces for multiscale interrelationships),
2. probabilistic (Fréchet means in iterative adaptive torus clusters obtained after circular mode hunting, projected to a shape space featuring data driven constraints),
3. clash free, and
4. fast,

RNA backbone correction which is an important and challenging contribution warranting further research in various directions, of which we sketch three.

In particular, we have discovered, described and exploited a relationship of RNA 3D structure between a microscopic and a mesoscopic scale. Further research, building on larger datasets, beyond the scope of this paper, will investigate this relationship more closely and identify relationships between other scales as well and exploit these similarly. As we have found that shape at different scales is best described by fundamentally different shape spaces, this involves statistically linking different geometrical models of shape.

At this point, the two-scale correction method CLEAN we propose corrects a central suite at microscopic scale only. More realistic, again beyond the scope of this paper, are simultaneous corrections of all suites involved at the mesoscopic scale (notably, adjacent suites overlap at four atoms), and correction of suites linked by nucleobase bindings, potentially far away along the backbone. Such corrections can, after elaborate extension, also address backbone-to-backbone-extra-suites clashes and even the more rare nucleobase clashes. Obviously these methods extend to various other biomolecules and in particular to protein structure correction, see Hamelryck *et al.* (2010).

As mentioned in the introduction, there are elaborate correction methods, for example ERRASER from Chou *et al.* (2013b), building on approximations of highly complex molecular dynamics simulations yielding 3D structures following the laws of biophysical chemistry. This aims not only at correcting all clashes (i.e. within-suite and between-suites, as well as backbone or base to backbone or base), it also aims at various other structure improvements. While for the test data set this entire process took several days on the ROSIE servers Chou *et al.* (2013a), frequently not removing all clashes, our CLEAN method, removing all within-suite-backbone-to-backbone clashes, ran within minutes. Since in contrast to corrections based on molecular dynamics, as demonstrated in Figures B3 and B15, our proposed corrections can be quite different from original clash suite shapes, they may serve as additional initial states for subsequent molecular dynamics, and thus provide a powerful tool.

Supplement Overview

In the supplementary material we

A. list the PDB files making the benchmark data set;
B. give scatterplots of suites (microscopic scale) in pairwise dihedral angle representation: first of the training data (the clash free suites of the benchmark data set), secondly of its classification by MINT-AGE and thirdly the corrections of the test data (the clashing suites of the benchmark data) by CLEAN;

C. give, for convenience and to make this paper self contained, an overview of the MINT-AGE algorithm from Mardia *et al.* (2022) including the parameter choices leading to the MINT-AGE classes of the training data set and, specifically, the classes found by circular mode hunting in the MINT-step;

D. briefly sketches the state of the art method ERRASER from Chou *et al.* (2013a) and provides some comparison with CLEAN MINT-AGE;

E. provide urls to access code and data (e.g. the PDB files and all the code used to generate the analyses and plots presented in this paper.).

Acknowledgements

The authors are grateful for fruitful discussions with Markus Hiller, Thomas Hamelryck and Carina Wollnik. The authors also gratefully acknowledge the very valuable comments provided by the three anonymous referees, improving this paper.

Funding

All authors except K. V. Mardia acknowledge DFG CRC 1456, K. V. Mardia acknowledges the Leverhulme Trust for the Emeritus Fellowship.

Supplement Overview

In the supplementary material we

A. list the PDB files making the benchmark data set;

B. give scatterplots of suites (microscopic scale) in pairwise dihedral angle representation: first of the training data (the clash free suites of the benchmark data set), secondly of its classification by MINT-AGE and thirdly the corrections of the test data (the clashing suites of the benchmark data) by CLEAN;

C. give, for convenience and to make this paper self contained, an overview of the MINT-AGE algorithm from Mardia *et al.* (2022) including the parameter choices leading to the MINT-AGE classes of the training data set and, specifically, the classes found by circular mode hunting in the MINT-step;

D. briefly sketches the state of the art method ERRASER from Chou *et al.* (2013a) and provides some comparison with CLEAN MINT-AGE;

E. provide urls to access code and data (e.g. the PDB files and all the code used to generate the analyses and plots presented in this paper.).

References

AlQuraishi, M. (2019). Parallelized natural extension reference frame: parallelized conversion from internal to cartesian coordinates. *Journal of computational chemistry*, **40**(7), 885–892.

Altis, A., Otten, M., Nguyen, P. H., Rainer, H., and Stock, G. (2008). Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *The Journal of Chemical Physics*, **128**(24), 245102.

Arnaudon, M. and Miclo, L. (2014). Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics*, **18**, 185–206.

Batool, M., Ahmad, B., and Choi, S. (2019). A structure-based drug discovery paradigm. *International Journal of Molecular Sciences*, **20**(11).

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, **28**(1), 235–242.

Chen, V. B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, David C., E.-m. d. (2010). Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D: Biological Crystallography*, **66**(Pt 1).

Chojnowski, G., Waleń, T., and Bujnicki, J. M. (2013). RNA Bricks—a database of RNA 3D motifs and their interactions. *Nucleic Acids Research*, **42**(D1), D123–D131.

Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., and Das, R. (2013a). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods*, **10**(1), 74–76.

Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., and Das, R. (2013b). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature methods*, **10**(1), 74—76.

Croll, T. I., Williams, C. J., Chen, V. B., Richardson, D. C., and Richardson, J. S. (2021). Improving SARS-CoV-2 structures: Peer review by early coordinate release. *Biophysical journal*, **120**(6), 1085–1096. 33460600[pmid].

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, **36**(4), 1758–1785.

Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R. Second Edition.* John Wiley and Sons, Chichester.

Duarte, C. M. and Pyle, A. M. (1998). Stepping through an RNA structure: A novel approach to conformational analysis 11. Edited by D. Draper. *Journal of Molecular Biology*, **284**(5), 1465 – 1478.

Eltzner, B., Huckemann, S., and Mardia, K. V. (2018). Torus principal component analysis with applications to RNA structure. *Ann. Appl. Stat.*, **12**(2), 1332–1359.

Everitt, B. (1993). *Cluster Analysis*. Edward Arnold, third edition.

Fletcher, P., Lu, C., Pizer, S., and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, **23**(8), 995–1005.

Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, Hugo, and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, **2**(3-4), 282–285.

Hamelryck, T., Borg, M., Paluszewski, M., Paulsen, J., Frellsen, J., Andreetta, C., Boomsma, W., Bottaro, S., and Ferkinghoff-Borg, J. (2010). Potentials of mean force for protein structure prediction vindicated, formalized and generalized. *PLOS ONE*, **5**(11), e13714.

Huckemann, S. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics*, **64**(6), 1227–1259.

Huckemann, S. and Ziezold, H. (2006). Principal component analysis for riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, **38**(2), 299–319.

Huckemann, S., Kim, K.-R., Munk, A., Rehfeldt, F., Sommerfeld, M., Weickert, J., and Wollnik, C. (2016). The circular SiZer, inferred persistence of shape parameters and application to early stem cell differentiation. *Bernoulli*, **22**(4), 2113 – 2142.

Huckemann, S. F. and Eltzner, B. (2015). Polysphere PCA with applications. In *Proceedings of the 33th LASR Workshop*, pages 51–55. Leeds University Press. http://www1.maths.leeds.ac.uk/statistics/workshop/lasr2015/Proceedings15.pdf.

Ippolito, J. A. and Steitz, T. A. (2000). The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution. *Journal of Molecular Biology*, **295**(4), 711–717.

Jain, S., Richardson, D. C., and Richardson, J. S. (2015). Chapter Seven - Computational Methods for RNA Structure Validation and Improvement. In S. A. Woodson and F. H. Allain, editors, *Structures of Large RNA Molecules and Their Complexes*, volume 558 of *Methods in Enzymology*, pages 181 – 212. Academic Press.

Jung, S., Dryden, I. L., and Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, **99**(3), 551–568.

Kappel, K., Zhang, K., Su, Z., Watkins, A. M., Kladwang, W., Li, S., Pintilie, G., Topkar, V. V., Rangan, R., Zheludev, I. N., Yesselman, J. D., Chiu, W., and Das, R. (2020). Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods*, **17**(7), 699–707.

Kent, J. and Mardia, K. (2015). The winding number for circular data. In *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop*.

Kent, J. T. and Mardia, K. V. (2009). Principal component analysis for the wrapped normal torus model. *Proceedings of the Leeds Annual Statistical Research (LASR) Workshop 2009*.

Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, **24**(5), 719–720.

Liao, H. Y. and Frank, J. (2010). Definition and estimation of resolution in single-particle reconstructions. *Structure (London, England : 1993)*, **18**(7), 768–775. 20637413[pmid].

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D*, **75**(10), 861–877.

Mardia, K. V., Wiechers, H., Eltzner, B., and Huckemann, S. F. (2022). Principal component analysis and clustering on manifolds. *Journal of Multivariate Analysis*, **188**, 104862. 50th Anniversary Jubilee Edition.

Murray, L. J. W., Arendall, W. B., Richardson, D. C., and Richardson, J. S. (2003). RNA backbone is rotameric. *Proceedings of the National Academy of Sciences*, **100**(24), 13904–13909.

Obulkasim, A., Meijer, G. A., and van de Wiel, M. A. (2015). Semi-supervised adaptive-height snipping of the hierarchical clustering tree. *BMC Bioinformatics*, **16**(1), 15.

Parsons, J., Holmes, J. B., Rojas, J. M., Tsai, J., and Strauss, C. E. (2005). Practical conversion from torsion space to cartesian space for in silico protein synthesis. *Journal of computational chemistry*, **26**(10), 1063–1068.

Petrov, A. I., Zirbel, C. L., and Leontis, N. B. (2013). Automated classification of rna 3d motifs and the rna 3d motif atlas. *RNA (New York, N.Y.)*, **19**(10), 1327–1340. 23970545[pmid].

Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J., Richardson, D. C., Ham, D., Hershkovits, E., Williams, L. D., Keating, K. S., Pyle, A. M., Micallef, D., Westbrook, J., Berman, H. M., and Consortium, R. O. (2008). Rna backbone: consensus all-angle conformers and modular string nomenclature (an rna ontology consortium contribution). *RNA (New York, N.Y.)*, **14**(3), 465–481. 18192612[pmid].

Richardson, J. S., Williams, C. J., Hintze, B. J., Chen, V. B., Prisant, M. G., Videau, L. L., and Richardson, D. C. (2018). Model validation: local diagnosis, correction and when to quit. *Acta Crystallographica Section D*, **74**(2), 132–142.

Sargsyan, K., Wright, J., and Lim, C. (2012). GeoPCA: a new tool for multivariate analysis of dihedral angles based on principal component geodesics. *Nucleic Acids Research*, **40**(3), e25.

Schlick, T. and Pyle, A. M. (2017). Opportunities and challenges in rna structural modeling and design. *Biophysical journal*, **113**(2), 225–234. 28162235[pmid].

Schrödinger, LLC (2015). The PyMOL molecular graphics system, version 1.8.

Sokal, R. R. and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, **38**, 1409–1438.

Tang, L., Johnson, K. N., Ball, L. A., Lin, T., Yeager, M., and Johnson, J. E. (2001). The structure of pariacoto virus reveals a dodecahedral cage of duplex rna. *Nature Structural Biology*, **8**(1), 77–83.

Thorn, A., Gao, Y., Nolte, K., Kirsten, F., and Stäb, S. (2021). Coronavirus structural task force. https://github.com/thorn-lab/coronavirus_structural_task_force.

Wadley, L. M., Keating, K. S., Duarte, C. M., and Pyle, A. M. (2007). Evaluating and Learning from RNA Pseudotorsional Space: Quantitative Validation of a Reduced Representation for RNA Structure. *Journal of Molecular Biology*, **372**(4), 942 – 957.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**(1), 57–63.

Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., and Losick, R. (2004). *Molecular Biology of the Gene*. Pearson Education, fifth edition.

Word, J., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999). Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. Edited by J. Thornton. *Journal of Molecular Biology*, **285**(4), 1711 – 1733.

Zhang, K., Zheludev, I. N., Hagey, R. J., Wu, M. T.-P., Haslecker, R., Hou, Y. J., Kretsch, R., Pintilie, G. D., Rangan, R., Kladwang, W., Li, S., Pham, E. A., Bernardin-Souibgui, C., Baric, R. S., Sheahan, T. P., D'Souza, V., Glenn, J. S., Chiu, W., and Das, R. (2020). Cryo-electron microscopy and exploratory antisense targeting of the 28-kda frameshift stimulation element from the SARS-CoV-2 RNA genome. *bioRxiv*.

Zhang, K., Zheludev, I. N., Hagey, R. J., Haslecker, R., Hou, Y. J., Kretsch, R., Pintilie, G. D., Rangan, R., Kladwang, W., Li, S., Wu, M. T.-P., Pham, E. A., Bernardin-Souibgui, C., Baric, R. S., Sheahan, T. P., D'Souza, V., Glenn, J. S., Chiu, W., and Das, R. (2021). Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nature Structural & Molecular Biology*, **28**(9), 747–754.

Zouboulouglou, P., García-Portugués, E., and Marron, J. S. (2021). Scaled torus principal component analysis. arXiv stat.ME 2110.04758.

Supplementary Information

A The PDB files

Table B16 lists the PDB IDs and the corresponding resolution of the benchmark data set, see Section 3.4.

Figure B16: *PDB IDs and resolutions (see Section 3.3) from the 71 different measurements from the benchmark data set (see Section 3.4).*

PDB ID	Resolution	PDB ID	Resolution
1cvj	2.60	1ooa	2.45
1ddl	2.70	1q2r	2.90
1duh	2.70	1qf6	2.90
1e7k	2.90	1qtq	2.25
1ec6	2.40	1r3e	2.10
1ehz	1.93	1r3o	1.90
1et4	2.30	1rlg	2.70
1f1t	2.80	1s03	2.70
1f7u	2.20	1s72	2.40
1f7y	2.80	1sds	1.80
1f8v	3.00	1u9s	2.90
1f27	1.30	1vfg	2.80
1ffy	2.20	1wpu	1.48
1h3e	2.90	1xjr	2.70
1h4s	2.85	1xmq	3.00
1hmh	2.60	1xok	3.00
1hr2	2.25	1y3s	2.25
1i6u	2.60	1yfg	3.00
1ivs	2.90	1yls	3.00
1jbr	2.15	1z43	2.60
1k8w	1.85	2a8v	2.40
1kh6	2.90	2a43	1.34
1kq2	2.71	2atw	2.25
1kxk	3.00	2bh2	2.15
1l2x	1.25	2bte	2.90
1l9a	2.90	2bu1	2.20
1lng	2.30	2bx2	2.85
1m5k	2.40	2csx	2.70
1m8v	2.60	2fmt	2.80
1m8x	2.20	7msf	2.80
1mzp	2.65	361d	3.00
1n78	2.10	397d	1.30
1ntb	2.90		

B Scatterplots of all pairwise dihedral angle pairs

Figure B17 illustrates all two dimensional dihedral angle pairs of the suites of the training data set $\mathfrak{T}$, introduced in Section 3.4. Their MINT-AGE classes (see Algorithm C.3) with the parameters described in Supplement C.1 are shown in Figure B18. The corresponding cluster sizes are summarized in Figure B10 in the main text. The scatterplots of the 198 clash suites $\mathfrak{C}$ from the benchmark test data set are depicted in Figures B19 as black diamonds. Figure B19 shows the CLEAN (Section 4.3.1) corrections as green circles and blue crosses, respectively.

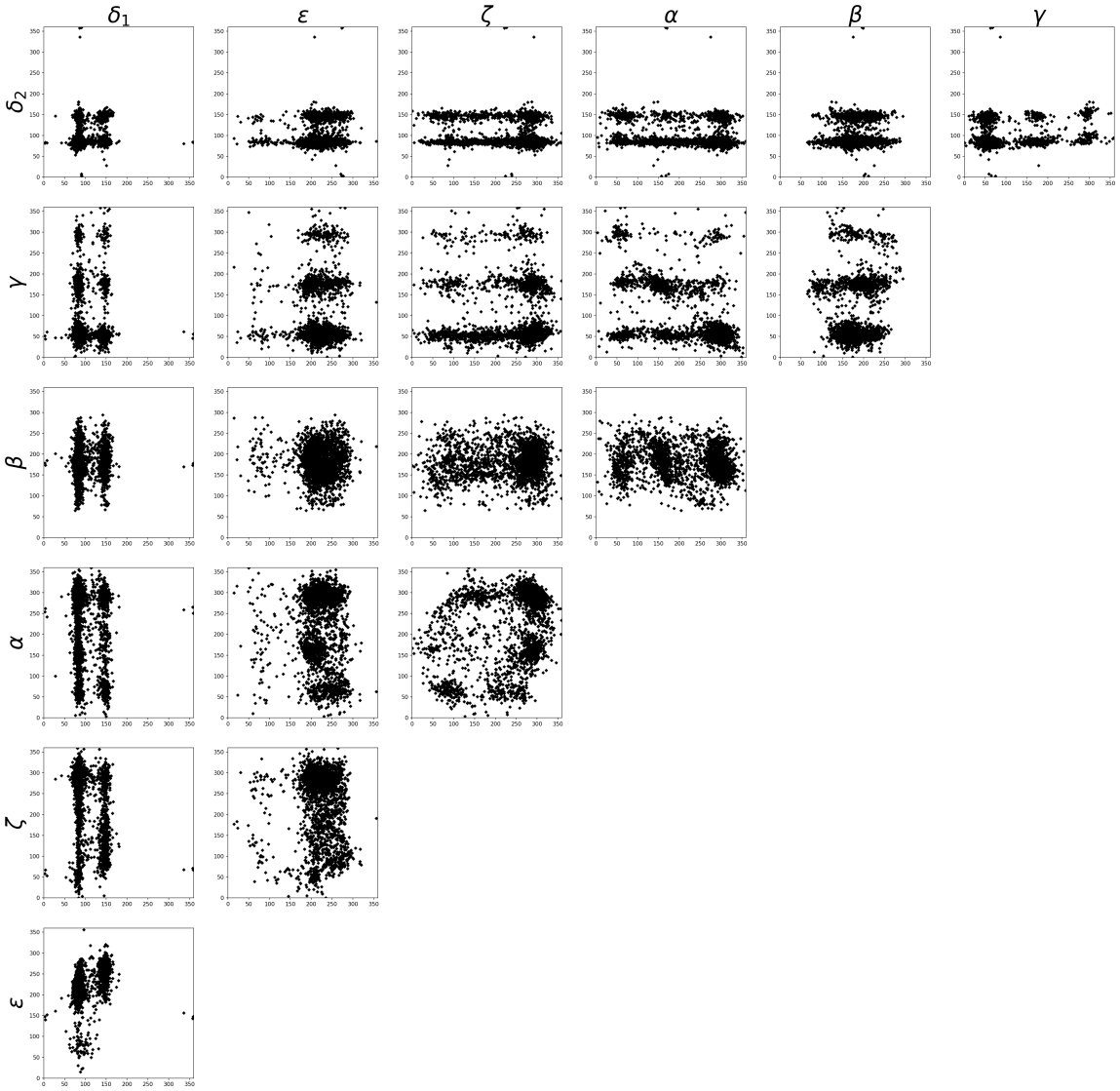


Figure B17: *Scatterplots of all two dimensional dihedral angle pairs (in degrees) of the suites of the training data set $\mathfrak{T}$, see Section 3.4.*

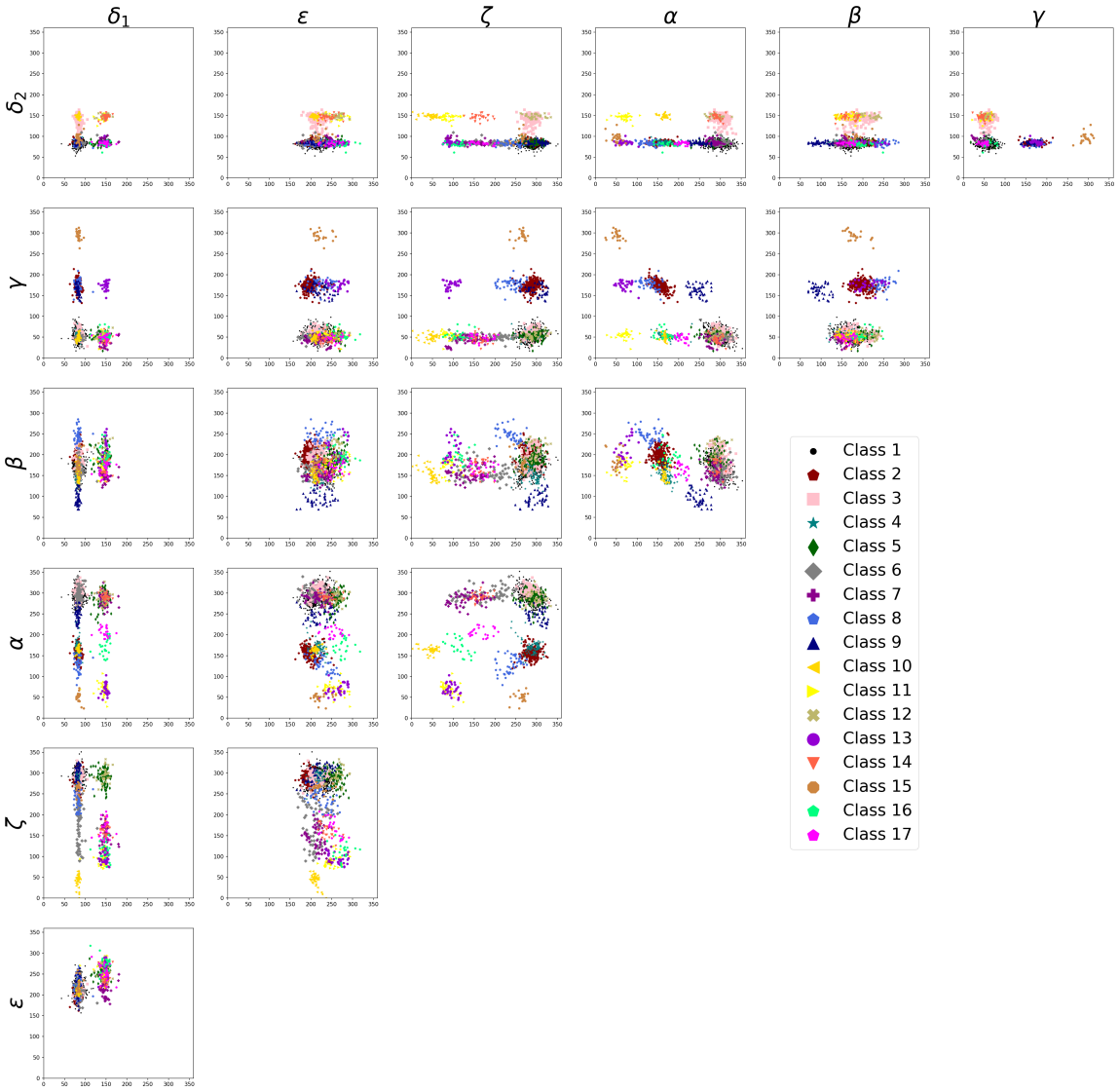


Figure B18: *Scatterplots of all two dimensional dihedral angle pairs (in degrees) for the 17 MINT-AGE benchmark classes described in Supplement C.*

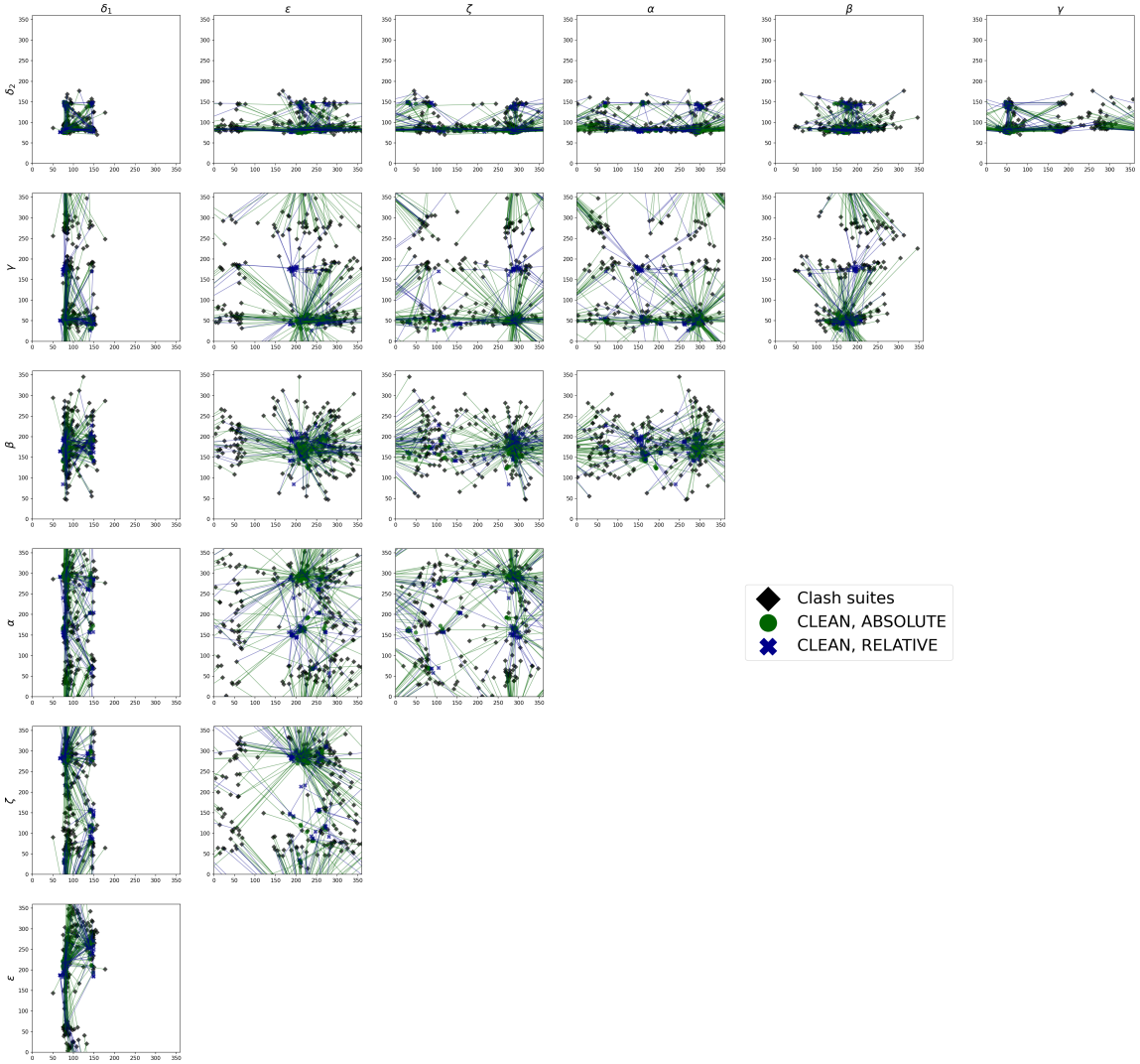


Figure B19: *Scatterplots of all two dimensional dihedral angle pairs (in degrees) of the 198 clash suites* $\mathfrak{C}$ *(black) from the benchmark test data set with their corrections by CLEAN (green circles correspond to flag DOMINATING = ABSOLUTE in Section 4.3.1 and blue crosses to flag DOMINATING = RELATIVE). Each correction belongs to a clash free class.*

C MINT-AGE

The following two algorithm (Algorithm C.1 and Algorithm C.3) have been proposed in Mardia *et al.* (2022). For convenience, we reproduce them here including an introduction.

The *Mode huntINg after Torus pca on Adaptive cutting averaGe linkage trEes* (MINT-AGE) algorithm builds on three components. Initially torus data is pre-clustered by adaptively cutting an average linkage clustering tree, as detailed in Supplement C.1. Then, each cluster is reduced to a one dimensional torus representation using *torus PCA*, recently developed by Eltzner *et al.* (2018), with varying flag parameters, see Supplement C.2. While the torus is rather inconvenient for PCA based dimension reduction methods (almost all geodesics are dense and tangent space methods lose periodicity), torus PCA deforms a torus into a stratified sphere, opening up the toolbox of *principal nested spheres* from Jung *et al.* (2012). This makes the torus even more attractive for PCA based dimension

reduction methods than Euclidean space, cf. Huckemann and Eltzner (2015). Finally, if the sum of squared residual distances to these one dimensional torus representations of the pre-clusters is less than one fourth of the cluster's total Fréchet variance, these one dimensional representations are subjected to circular mode hunting that identifies subclusters with statistical significance, as detailed in Supplement C.3.

C.1 AGE: Iterative Adaptive Cutting Average Linkage Tree Clustering on a Metric Space

The first building block is AGE pre clustering. It builds on *average linkage clustering*, also known as the *unweighted pair group method with arithmetic mean* or simply as *UPGMA*, which was first developed by Sokal and Michener (1958). It is a hierarchical clustering method that creates a rooted tree where each node stands for a cluster comprising all leaves below that node.

Starting with data $X^{(1)}, \ldots, X^{(n)}$ in a metric space with distance d, initially, each $X^{(i)}$ $(i = 1, \ldots, n)$ is assigned its own cluster yielding the initial *running cluster list*. The tree constructed starts with a graph comprising n leaves labeled from 1 to n, representing each of these initial clusters. Then, iteratively, if there is more than one cluster in the running cluster list, the two clusters with the smallest average distance are merged to form a new cluster which is added to the running list, from which the two merged clusters are deleted. Here, the *average distance* between two clusters A and B is given by

$$d(A, B) = \frac{1}{|A| \cdot |B|} \cdot \sum_{X \in A} \sum_{Y \in B} d(X, Y). \tag{14}$$

The graph is extended by adding a parent node above the two nodes corresponding to the merged clusters and adding branches joining them to the parent node. The iteration terminates if the running list contains only one cluster, comprising all data, which forms the root of the tree, called the *average linkage clustering tree*. Notably, node values increase when approaching the root. Indeed, if clusters A and B are merged to cluster $A \cup B$ and C is any other cluster, then $d(A, B) \leq d(A, C), d(B, C)$ and hence

$$d(A \cup B, C) = \frac{1}{|A \cup B| \cdot |C|} \left(|A| \cdot |C| \cdot d(A, C) + |B| \cdot |C| \cdot d(B, C) \right) \geq d(A, B)\,.$$

Every choice of a distance value $d_c > 0$ yields a clustering by a *tree cut* at distance d_c, i.e. by taking that running list when the last node with distance value $\leq d_c$ has been added.

We note that instead of the average distance function in (14), one may consider the minimal cluster distance yielding *single linkage* or *nearest neighbor clustering*, developed by Florek *et al.* (1951), which is currently also highly popular. It tends to return long elongated clusters, an effect called *chaining*, see Everitt (1993). For our purpose of structure correction relying on Fréchet means, which are generalizations of averages to metric spaces, clustering based on average distance ensures that Fréchet means are within or close to their clusters with a tendency towards isotropic spread.

Obviously, this simple approach fails to separate frequent cluster configurations, for instance if two closely neighboring clusters have a higher density than a third one. Such configurations are separated, however, using a data-adaptive cutting procedure, see for example Langfelder *et al.* (2007) and Obulkasim *et al.* (2015). We choose the following tuning parameters:

- *maximal outlier distance* $d_{\max}$ controlling cluster density,

- *minimal cluster size* κ controlling, if not too small, that our mode hunting from Supplement C.3 can significantly separate one-dimensional clusters with several modes,
- *relative branching distance* q ensuring that two clusters are only split if their parent node's distance value is significant in relation to the greatest distance value of its child nodes.

Algorithm C.1 (AGE)**.** Consider the set $P = \{X^{(1)}, \ldots, X^{(n)}\}$ of n data points. Let R be the outlier list and $\mathcal{C}$ be the cluster list, each of which are initially empty. They are filled iteratively as follows:

1. Compute the average linkage clustering tree from P
2. Perform a tree cut at distance $d_{\max}$ to obtain a clustering, move from P to R all data points that are in clusters with less than κ data points.
3. Compute the average linkage cluster tree for the new P as in Step 1.
4. Set $s_P = \sqrt{|P| + \kappa^2}$ (inspired by the square root rule of thumb used in histogram binning), create an empty list $\mathcal{L}$ of clusters.
5. Begin at the root and always follow the branch with more points at each node. From each node add the child node corresponding to the smaller subcluster to $\mathcal{L}$
   (a) if it contains more than s_P data points,
   (b) and if the q-fold of its parent node's distance value is greater than the two children nodes' distance values.
6. At the last node, where the smaller subcluster is added to $\mathcal{L}$, also add the larger subcluster to $\mathcal{L}$
7. Consider $\mathcal{L}$:
   - if $\mathcal{L}$ is empty, move the union of all data points from P to $\mathcal{C}$; these correspond then to one single cluster,
   - else, add the largest cluster in $\mathcal{L}$ to $\mathcal{C}$ and remove its points from P.
8. If $|P| > 0$, go to Step 1.
9. Return the clusters list $\mathcal{C}$ and the outlier list R.

Some of the pre-clusters resulting from the benchmark training set contain obvious subclusters that have not be identified by Algorithm C.1, for a similar data set this is detailed in Mardia *et al.* (2022). It turns out they can be well separated by mode hunting (see Section C.3) applied to their one dimensional torus PCA representation. To this end, for the pre-clustering Algorithm C.1 we choose the *minimal cluster size* $\kappa = 20$, the *relative branching distance* $q = 0.15$ and the *maximal outlier distance* $d_{\max}$ such that 15% of the suites in the average linkage tree are in a branch with less than κ data points.

Remark C.2. *Since we use the clustering results in Section 4.3.1 in the main text to suggest corrections for clash suites, we aim at larger and concentrated clusters, possibly at the price of a larger number of outliers which cannot be allocated to any of the clusters. However, other choices of tuning parameters are conceivable:*

- *Increasing d_{max} would result in reducing the number of outliers. If d_{max} is chosen too large, outliers will be added to set of clusters, too small d_{max} will cause elements that actually belong to clusters to be assigned to the set of outliers.*
- *Increasing of the minimal cluster size κ would cause some smaller clusters to no longer be detected. Decreasing κ would result in reduced ability to separate smaller clusters in the mode hunting step, see Supplement C.3, with statistical significance.*
- *Cluster centers are often concentrated due to chemical constraints. To ensure that they are not separated from the less dense neighborhood, $q = 0.15$ was chosen.*

C.2 Torus PCA Based Clustering

Algorithm C.3 (MINT-AGE)**.**

AGE-step: From input suite data $X^{(1)}, \ldots, X^{(n)}$ obtain a list of pre-clusters using Algorithm C.1 from the supplement and store it as the *remaining cluster list* $\mathcal{R}$. Create the initially empty *final cluster list* $\mathcal{F}$.

MINT-step: While $\mathcal{R}$ is non-empty:

1. Take a cluster C from $\mathcal{R}$ and set $m = 1$:
2. For C perform torus PCA from Eltzner *et al.* (2018) with the flags GC (gap centered), MC (mean centered), SI (spread inside), SO (spread outside) as detailed there
   - if m = 1: GC, SI
   - if m = 2: GC, SO
   - if m = 3: MC, SI
   - if m = 4: MC, SO
   - if m = 5: Remove C from the remaining cluster list $\mathcal{R}$, add it to $\mathcal{F}$ and go to Step 1.
3. For the suite $X^{(j)} \in C$ let $X^{(j,1D)}$ denote their one-dimensional torus PCA projections and let μ denote the torus PCA nested mean from Eltzner *et al.* (2018). Whenever $4\left(\sum_{j=1}^{n} d_{\mathbb{T}^7}\left(X^{(j)}, X^{(j,1D)}\right)^2\right) \leq \sum_{j=1}^{n} d_{\mathbb{T}^7}\left(X^{(j)}, \mu\right)^2$, perform mode hunting from Section C.3:
   - if subclusters were found, add them to the remaining cluster list $\mathcal{R}$ and remove C from it.
   - else: Set $m = m + 1$ and go to Step 2.

Return: $\mathcal{F}$.

C.3 Circular Mode Hunting

We cluster the one-dimensional projections obtained by torus PCA in Step 2(c) of Algorithm C.3 using the multiscale method described by Dümbgen and Walther (2008). Although this method was originally defined for the real line, its numerical implementation for circular data is even simpler. Since modes are separated by minima, we use this method to identify regions in which minima are located with a certain confidence level. (Throughout the applications, we use a fixed confidence level of 95%). In each of these

regions, we estimate the minima by estimating the density of the one-dimensional projections using a wrapped Gaussian kernel. As the bandwidth increases, the number of minima of the density estimate in each of the regions decreases. Whenever there is only one minimum left in a region – this will inevitably happen due to the causality of the wrapped normal kernel, see Huckemann *et al.* (2016) – we take this as a cluster boundary, see Figure B20 for an illustration of our method.

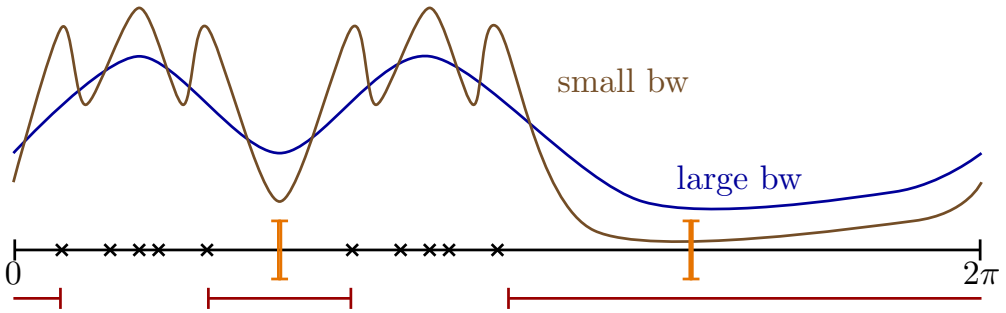


Figure B20: *Circular mode hunting for data (black asterisks). Intervals containing minima with statistical significance (red) and wrapped Gaussian kernel smoothed densities with varying bandwidths having too many minima (brown) and having the statistically significant number of minima (blue). The latter minima are taken as cluster boundaries (orange), adapted from Mardia* et al. *(2022).*

From the training data set, the AGE-step of Algorithm C.3 produced 13 pre-clusters, three of which containing sub-clusters identified by the MINT-step, see Table B1.

Pre-cluster number	MINT-AGE class numbers
7	9, 12
4	6, 7, 14
8	8, 15

Table B1: *MINT-AGE classes (2nd column) found in single AGE pre-clusters (1st column).*

D ERRASER

A popular and well established (see for example Richardson *et al.* (2018), Jain *et al.* (2015)) tool for structure correction is ERRASER (Enumerative Real-space Refinement ASsisted by Electron-density under Rosetta) Chou *et al.* (2013a). It automatically corrects complete RNA structures as well as individual residues and usually executes the following three steps three times in succession, see Chou *et al.* (2013a):

1. The high-resolution *Rosetta energy function*, extended by *electron density correlation evaluation*, subjects, among others, all dihedral angles to minimization in order to obtain a new reconstruction.

2. Residues in this new reconstruction are labeled if the PHENIX validation tools (see Section 3.3 in the main text) detect errors, if the backbone's configuration is not recognized, or if other geometric errors occur (e.g. in the structure of the sugar ring).

3. Labeled residues are reconstructed one after the other by Single Nucleotide StepWise Assembly (SWA) which samples all nucleotide atoms from an exhaustive grid search.

D.1 Using ERRASER

To correct a PDB file with ERRASER, one needs a 2mFo-DFc density map in CCP4 format in addition to the raw PDB file. We created the 2mFo-DFc electron density maps with the PHENIX map tool, see Liebschner *et al.* (2019). However, some older PDB files have not published the associated experimental files necessary to create a 2mFo-DFc density map in CCP4 format and were therefore left out. We used the offered online server ROSIE (Rosetta Online Server that Includes Everyone) see Chou *et al.* (2013a) to obtain the ERRASER corrections. ERRASER returns a statistic for each corrected PDB file that identifies a *clashscore* (the number of clashes per 1000 atoms) in the raw data set and the clashscore in the PDB file corrected by ERRASER.

D.2 ERRASER data set

As explained above, the ERRASER method on the ROSIE server can only correct PDB files that come with an associated 2mFo-DFc density map in CCP4 format and do not exceed a specific maximum size. This is the case for only 49 PDB files of the 71 PDB files from our benchmark data set, comprising 2325 suites. We denote this set by $\mathfrak{R}$ and it is a subset of our benchmark data set from Section 3.4. Whether clashing or not, all of the suites from $\mathfrak{R}$ are corrected by ERRASER. We obtain the *ERRASER test data set* $\mathfrak{C}' = \mathfrak{C} \cap \mathfrak{R}$ of clash suites which has size 73 (recall that $\mathfrak{C}$ is the test data set introduced in Section 3.4 in the main text).

D.3 Clash Reduction by ERRASER

We apply the validation method `phenix.clashscore`, see Section 3.3, which returns the number of clashes per 1000 atoms found in a given PDB file. Note that this *clash score* includes all clashes between two atoms in a measurement, not just the clashes between two backbone atoms.

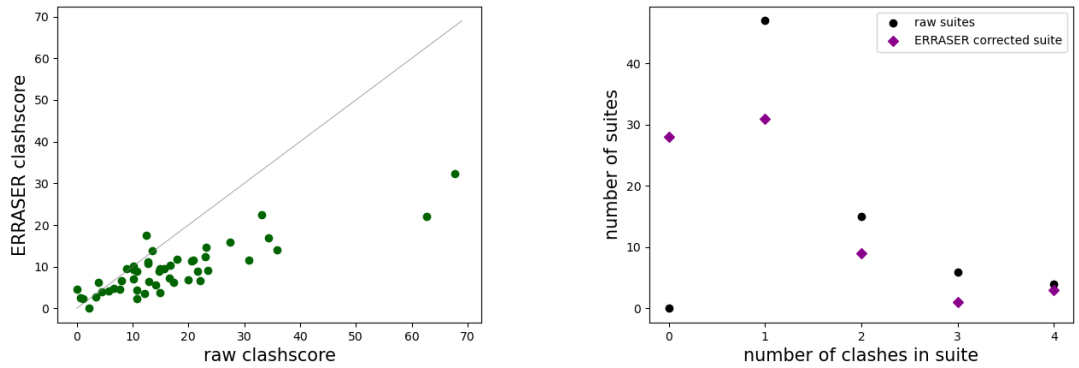


Figure B21: *Left: Each of the 73 green points compares the clash score of a single PDB files before (horizontal) and after (vertical) correction by ERRASER. Right: Histogram of numbers of atom clashes (horizontal) in clash suites before (black) and after correction by ERRASER (magenta).*

The left panel of Figure B21 shows that ERRASER effectively reduces clash scores, as an overwhelming part of the green points (each corresponding to a single PDB file) lies substantially below the diagonal. The right panel of Figure B21 shows the amount of clash reduction on suite level: after correction by ERRASER, from $\mathfrak{C}'$, approx. 40 % (29) of the suites have been made clash free and almost equally many (30) feature only one clash; correcting the few suites with a higher number of clashes was only partially possible.

D.4 Comparing CLEAN MIN-TAGE with ERRASER

By design, our CLEAN MIN-TAGE algorithm corrects clashes by assigning clash suites to one of the clash free classes of the underlying training data set. In contrast, after correction by ERRASER, many clash suites remain outliers. Figure B22 displays the clash suites $\mathfrak{C}'$ before and after correction by ERRASER and CLEAN MIN-TAGE and in Figure B23 displays all two dimensional dihedral angle pairs of the the clash suites $\mathfrak{C}'$ before and after correction by ERRASER. Through ERRASER correction, no distinct structures become visible. In stark contrast, CLEAN MIN-TAGE assigns most clash suites to the dominating (first cluster, A helix) and some to smaller clusters, see Figure B22.

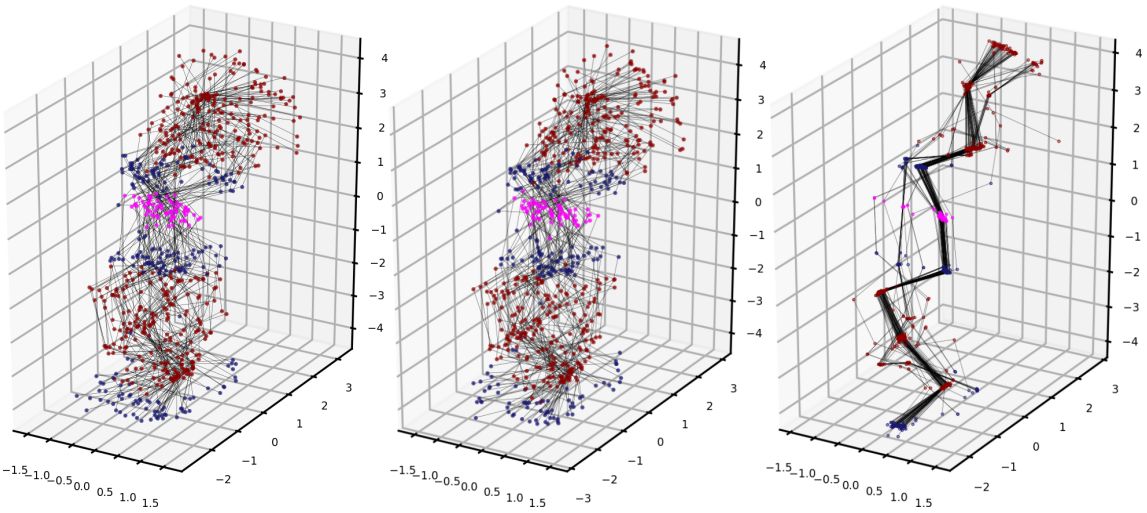


Figure B22: *RNA backbone suites with carbon (dark red), oxygen (dark blue) and phosphorus atoms (pink), see Figure B1 in the main text. Left: 73 clash suites which form our ERRASER test set $\mathfrak{C}'$. Center: Their corrections by ERRASER. More than half of them still feature clashes. Right: Their corrections by CLEAN MIN-TAGE. Each correction belongs to a clash free class. See Figures B23 and B19 for both corrections in dihedral angle representation.*

All two dimensional dihedral angle pairs of the suites of the 73 clash suites $\mathfrak{C}'$ from the ERRASER test set are depicted in Figure B23 together with the corresponding ERRASER corrections as magenta circles.

Figure B23: *Scatterplots of all two dimensional dihedral angle pairs (in degrees) of the 73 clash suites* $\mathfrak{C}'$ *(black) from the ERRASER test set with their corrections (magenta) by ERRASER from Chou* et al. *(2013a). More than half of them still feature clashes, cf. Supplement D.3 and Figure B22.*

E Data and Code

The PDB files and all the code used to generate the analyses and plots presented in this paper can be found https://gitlab.gwdg.de/henrik.wiechers1/clean-mintage-code.

CHAPTER C

Drift Models on Complex Projective Space for Electron-Nuclear Double Resonance

Drift Models on Complex Projective Space for Electron-Nuclear Double Resonance

Henrik Wiechers[1], Markus Zobel[1], Marina Bennati[2,3], Igor Tkach[2], Benjamin Eltzner[2], Stephan Huckemann[1], Yvo Pokern[4]

[1]Felix-Bernstein-Institute for Mathematical Statistics, Georg-August-University, Göttingen, 37077 Göttingen, Germany.
[2]Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany.
[3]Department of Chemistry, Georg-August University of Göttingen, Tammannstr. 2, Göttingen, Germany
[4]Department of Statistical Science, University College London, London WC1E 6BT, United Kingdom.

Abstract

ENDOR spectroscopy is an important tool to determine the complicated three-dimensional structure of biomolecules and in particular enables measurements of intramolecular distances. Usually, spectra are determined by averaging the data matrix, which does not take into account the significant thermal drifts that occur in the measurement process. In contrast, we present an asymptotic analysis for the homoscedastic drift model, a pioneering parametric model that achieves striking model fits in practice and allows both hypothesis testing and confidence intervals for spectra. The ENDOR spectrum and an orthogonal component are modeled as an element of complex projective space, and formulated in the framework of generalized Fréchet means. To this end, two general formulations of strong consistency for set-valued Fréchet means are extended and subsequently applied to the homoscedastic drift model to prove strong consistency. Building on this, central limit theorems for the ENDOR spectrum are shown. Furthermore, we extend applicability by taking into account a phase noise contribution leading to the heteroscedastic drift model. Both drift models offer improved signal-to-noise ratio over pre-existing models.

1 Introduction

One of the main objectives of structural biology is to understand the complicated three-dimensional structure of biomolecules, and thus provide meaningful links between structure and functionality. In particular, this information can be used in the field of structure-based drug design, see for example [And03a]. There is a wide range of different methods to determine the structure, such as *X-ray crystallography* (X-ray), *cryogenic electron microscopy* (cryo-EM) and *spectroscopic* methods. *Nuclear magnetic resonance* (NMR) spectroscopy is possibly the most widely used spectroscopic method: it studies the interactions between the nuclei of a molecule using radio frequency (RF) pulses. *Electron paramagnetic resonance* (EPR), on the other hand, studies the local environment and different kinds of interactions of the spins of unpaired electrons using microwave (MW) pulses. It can be more selective than NMR in that it targets only the tiny minority of unpaired electrons among the large number of electrons present in a biomolecule. Additionally, the larger gyromagnetic ratio of the electron compared to any magnetic nucleus usually leads to higher detection sensitivity and thus to better signal-to-noise ratio (SNR). Electron Nuclear Double Resonance (ENDOR) spectroscopy [Feh56, GS91, Har16] seeks to combine the advantages of EPR and NMR by interacting with both, nuclei and radical electrons, using both MW and RF pulses in a single experiment (see Section 2 for an accessible exposition of how this works). It should be emphasized that NMR, EPR and ENDOR differ in their domain of applicability, in particular in the range of distances between interacting spins, rather than one method being generally superior to another. Roughly, ENDOR's double resonance approach yields information on how the unpaired electron interacts with magnetic nuclei of a chosen kind (e.g. protons, deuterium nuclei or fluorine nuclei) and explores their environment. Artificially inserting *labels*, i.e. magnetic nuclei rarely present in biomolecules such as fluorine or deuterium, as well as radicals containing unpaired electrons that do not naturally occur in the biomolecule under study such as nitroxide radicals, allows highly specific measurements of intramolecular distances and orientations between selectable parts of the biomolecule, see [MDD+20].

Prior to [PEH+21], the standard approach [EABG03, RB14] for extracting ENDOR spectra from the recorded echo signals was equivalent to the *averaging model* [PEH+21] whereby echo responses are simply averaged across a large number of replications of the ENDOR experiment and only the average response is processed further. However, as ENDOR experiments typically run for several hours and at low temperatures, significant thermal drifts over time occur in practice. In [PEH+21], the *homoscedastic drift model* was introduced for ENDOR experiments at a microwave frequency of 263 GHz, which uses the echo signals at each of the $N+1$ (with $N \in \mathbb{N}$) RF frequencies recorded in $B \in \mathbb{N}$ batches over time in a data matrix $Y \in \mathbb{C}^{B\times(N+1)}$. This model accounts for thermal drift by decomposing the data matrix accounting separately for signal drift and spectrum. It is the first of its kind in the field of ENDOR spectroscopy and, relative to common practice in applied statistics, achieves surprisingly good model fit that is maintained across a

number of chemical compounds in follow-up studies, cf. [PEH$^+$21, HTW$^+$22, WKH$^+$23], yielding improved SNRs relative to the averaging model. The homoscedastic drift model enables the application of the parametric bootstrap, which in turn enables hypothesis testing and confidence intervals for the spectra: In [PEH$^+$21], a flatness and a difference test were introduced and performed, which together confirmed unequivocally the presence of broad features that were suspected on visual inspection. [WKH$^+$23] utilizes the spectral uncertainties provided by the drift model to determine stochastic errors in the estimation of physical parameters from which intramolecular distances can be determined. The parameter of greatest applied interest in the homoscedastic drift model, κ, is complex-valued and contains both the ENDOR spectrum as well as an orthogonal component containing a resonance artefact. It is standardized so that $\sum_{\nu=0}^{N} \kappa_\nu = 0$ and $\sum_{\nu=0}^{N} |\kappa_\nu|^2 = 1$. Additionally, the spectrum I is extracted in a step following MLE estimation of κ by selecting a direction in the complex plane that contains the spectrum rather than the resonance artefact based on application-driven criteria so that $I = \mathrm{Re}\{\exp\left(i\lambda^{\mathrm{opt}}\right)\kappa\}$ holds for some $\lambda^{\mathrm{opt}} \in [0, 2\pi]$ which is determined from κ alone. Indeed, we will show that rotation of κ in the complex plane leaves the spectrum I invariant and, thus, it is the application that drives us to consider the complex projective space $\mathbb{C}P^{N-1}$ as the appropriate parameter space in this estimation problem. This paper addresses two main challenges:

Firstly, in order to justify the use of the above methods, we will address the asymptotic theory of ENDOR spectra in this paper. More precisely, both strong consistency and a central limit theorem (CLT) for the parameter κ are proved in the limit of large numbers of batches B. To this end, the theory of strong consistency of generalized Fréchet means is extended in Section 3 and applied in Section 4. Fréchet means (introduced by [Fré48]) take the notion of arithmetic mean to the non-Euclidean setting, and generalized Fréchet means are non-Euclidean data descriptors that do not necessarily live in the data space, that arises naturally in our application and create challenges arising from their implicit definition and potentially set-valued nature. We furthermore establish a CLT for the ENDOR spectrum I justifying the construction of confidence intervals for the ENDOR spectra at least in the case of known noise covariance and comment on the case of unknown noise covariance in Section 5.

Secondly, in Section 6, we extend the homoscedastic drift model to cover other microwave frequencies such as 94 GHz for which EPR spectrometers with an ENDOR capability are more widely available. This necessitates generalizing the drift model to the heteroscedastic case. Given the presence of boundary maxima and the unsatisfactory performance of penalized methods, a carefully devised parametric extension of the homoscedastic drift model is found to work best yielding fairly good fit to the data and notable improvements in SNR.

1.1 Merging Complex and Real Notation and Complex Projective Space

Switching conveniently between complex-valued and real-valued matrices, vectors and scalars, the following notation is used throughout the paper.

For a complex number $z = x + iy \in \mathbb{C}$ and a complex vector $(z_1, \ldots, z_N)^T \in \mathbb{C}^N$ define

$$\mathrm{vec}\,(z) \coloneqq \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2\,, \quad M(z) \coloneqq \begin{pmatrix} x & -y \\ y & x \end{pmatrix}\,, \quad \mathrm{vec} \begin{pmatrix} z_1 \\ \vdots \\ z_N \end{pmatrix} \coloneqq \begin{pmatrix} \mathrm{vec}\,(w_1) \\ \vdots \\ \mathrm{vec}\,(w_N) \end{pmatrix} \in \mathbb{R}^{2N}$$

Conversely for real vectors $(x, y)^T \in \mathbb{R}$ and $(r_1, \ldots, r_{2N})^T \in \mathbb{R}^{2N}$, define

$$\mathfrak{c} \begin{pmatrix} x \\ y \end{pmatrix} \coloneqq x + iy \in \mathbb{C}\,, \quad \tilde{\mathfrak{c}} \begin{pmatrix} r_1 \\ r_{2N} \end{pmatrix} \coloneqq \begin{pmatrix} r_1 + ir_2 \\ \vdots \\ r_{2N-1} + ir_{2N} \end{pmatrix} \in \mathbb{C}^N$$

The following Lemma summarizes basic rules, verified at once.

Lemma 1.1. *For $z, w \in \mathbb{C}$ we have*

1. $M(z)^T = M(\bar{z})$
2. $\mathrm{vec}(zw) = M(z)\mathrm{vec}(w) = M(w)\mathrm{vec}(z)$

Further, for $z, w \in \mathbb{C}^N$ and $A \in \mathrm{SPD}(2)$ we define

$$z \diamond_A w \coloneqq \sum_{i=1}^{N} M(z_i)^T A\, M(w_i) \in \mathbb{R}^{2\times 2}, \qquad z \bullet_A w \coloneqq \sum_{n=1}^{N} M(z_n)^T A\, \mathrm{vec}(w_n) \in \mathbb{R}^2,$$

as well as a Mahalanobis inner product, norm and distance,

$$\langle z, w \rangle_A \coloneqq \sum_{i=1}^{N} \mathrm{vec}(z_i)^T A\, \mathrm{vec}(w_i) \in \mathbb{R}, \qquad ||z||_A \coloneqq \sqrt{\langle z, z \rangle_A} \in \mathbb{R}$$
$$d_A(z, w) \coloneqq ||z - w||_A \in \mathbb{R}.$$

Next, we introduce complex projective space. It is the space of *complex directions* in $\mathbb{C}^N$ that can be viewed as the space of real directions modulo the *phase*

$$\lambda = \mathrm{Arg}(re^{i\lambda}) \in [0, 2\pi)$$

of a complex number $z = re^{i\lambda} \in \mathbb{C}$.

For a complex column vector $z \in \mathbb{C}^N$, $z^T = (z_1, \ldots, z_N)$, its Hermitian conjugate is the row vector

$$z^* \coloneqq (\overline{z_1}, \ldots, \overline{z_N})\,.$$

With the unit sphere

$$\mathcal{S}^{2N-1} := \{\kappa \in \mathbb{C}^N : \kappa^*\kappa = 1\}$$

of real dimension $2N-1$, the complex projective space of complex dimension $N-1$ and real dimension $2N-2$ is

$$\mathbb{C}P^{N-1} := \mathcal{S}^{2N-1}/\sim,$$

where "$\sim$" denotes the equivalence relation

$$\kappa \sim \tilde{\kappa} \quad \Leftrightarrow \quad \exists \lambda \in \mathbb{R}, \quad e^{i\lambda}\tilde{\kappa} = \kappa.$$

Furthermore we define the equivalence class of κ by $[\kappa]$. The distance between $[\kappa], [\tilde{\kappa}] \in \mathbb{C}P^{N-1}$ is defined by

$$d([\kappa], [\tilde{\kappa}]) = \min_{\lambda \in \mathbb{R}} ||\kappa - e^{i\lambda}\tilde{\kappa}||$$

where $\kappa \in [\kappa], \tilde{\kappa} \in [\tilde{\kappa}]$ are arbitrary representatives.

We say that $\kappa, \tilde{\kappa} \in \mathcal{S}^{2N-1}$ are in optimal position if

$$d\left([\kappa], [\tilde{\kappa}]\right)) = \|\kappa - \tilde{\kappa}\|.$$

Lemma 1.2. *For arbitrary $\kappa, \tilde{\kappa} \in \mathcal{S}^{2N-1}$ we have that they are in optimal position if $\tilde{\kappa}^*\kappa = 0$, or else,*

$$\kappa, \quad \frac{\tilde{\kappa}^*\kappa}{|\tilde{\kappa}^*\kappa|}\,\tilde{\kappa}$$

are in optimal position.

Proof. The assertion follows at once from

$$d\left([\kappa], [\tilde{\kappa}]\right)) = \min_{\lambda \in \mathbb{R}} \left(\kappa - e^{i\lambda}\tilde{\kappa}\right)^* \left(\kappa - e^{i\lambda}\tilde{\kappa}\right) = \min_{\lambda \in \mathbb{R}} \left(2 - 2\,\mathrm{Re}\left(e^{i\lambda}\kappa^*\tilde{\kappa}\right)\right).$$

□

2 Homoscedastic Drift Model

In this section, we selectively review those aspects of the ENDOR experiment that are necessary for the present work with more background available in [GS91] and full experimental details in [PEH+21]. We then introduce the setting for the homoscedastic drift model from [PEH+21] in preparation for its asymptotic analysis.

In the ENDOR experiment, a sequence of MW and RF pulses is sent into a chemical sample that is placed in an external magnetic field with field strength B_0. The magnetic field strength B_0, as well as the MW frequency ν_{MW} and MW pulse lengths together determine the set of orientations relative to the external magnetic field of those molecules in the chemical sample that participate in the resonance experiment. Typically, five dif-

ferent field strengths B_0 are used to select five different sets of orientations denoted as g_x, g_{xy}, g_y, g_{yz} and g_z. The microwave echo signal returned by the participating molecules in the chemical sample is recorded in two separate components: a component that is in phase with a reference MW signal constitutes the real part and a component whose phase is shifted by 90 degrees, known as 'in quadrature', constitutes the imaginary part. This echo signal is influenced by a RF pulse that is part of the pulse sequence. While the MW frequency is constant throughout the ENDOR experiment (we report measurements for $\nu_{MW} = 263\,\mathrm{GHz}$ and, in Section 6, $\nu_{MW} = 94\,\mathrm{GHz}$), the RF frequency is varied in a pseudo-random sequence covering each of the RF frequencies $\{f_\nu : \nu \in \{0, \dots, N\}\}$, $N \in \mathbb{N}$ once. This is known as a *scan*. Since the SNR in a single scan is very low, a number $S \in \mathbb{N}$ of scans are performed in succession which constitute a batch of measurements. The batches are enumerated by $b \in \{1, \dots, B\}$. The resulting echo signals $X_{s,b,\nu} \in \mathbb{C}$ are summed up to form $Y_{b,\nu} = \sum_{s=1}^{S} X_{s,b,\nu}$. Here, S is chosen large enough to yield a SNR sufficient to allow adjustment of experimental parameters based on a single batch $Y_{b,:} := (Y_{b,0}, \dots, Y_{b,N})^T$ but small enough for the thermal drift that affects phase and amplitude of the echo signal to be negligible. Thus, we obtain the data matrix $Y \in \mathbb{C}^{B\times(N+1)}$, and a sample data matrix is illustrated in Figure C6 of the Supplementary Information (SI). Prior to [PEH+21], the standard approach [RB14, EABG03] to extract ENDOR spectra from the echo signal Y was the *averaging model*:

Definition 2.1 (Averaging Model)**.** *In the averaging model, the batches are averaged according to*

$$Z_\nu = \frac{1}{B}\sum_{b=1}^{B} Y_{b,\nu}. \tag{1}$$

In a second step, a phase correction*, i.e. a complex multiplication by $e^{i\lambda}$ with a manually tuned $\lambda \in [0, 2\pi)$ to obtain a real valued non-normalized spectrum $\tilde{I} = \mathrm{Re}(e^{i\lambda}Z)$ is applied followed by normalization to obtain the spectrum*

$$I_\nu = \frac{\tilde{I}_\nu - \min_{\nu'\in\{0,\dots,N\}} \tilde{I}_{\nu'}}{\max_{\nu'\in\{0,\dots,N\}} \tilde{I}_{\nu'} - \min_{\nu'\in\{0,\dots,N\}} \tilde{I}_{\nu'}}. \tag{2}$$

In [PEH+21], the statistical flaws of this approach were addressed. Firstly, normalization via $Z_\nu = \psi + \phi\kappa_\nu$ with $\psi \in \mathbb{C}$, $\phi \in \mathbb{R}_{\geq 0}, \kappa_\nu \in \mathbb{C}$ and imposing

$$\sum_{\nu=0}^{N} \kappa_\nu \overset{!}{=} 0 \tag{3}$$

$$\sum_{\nu=0}^{N} |\kappa_\nu|^2 \overset{!}{=} 1 \tag{4}$$

is less sensitive to outliers. Note that the condition 3 removes a complex degree of freedom, motivating our choice of $N+1$ rather than N RF frequencies. Secondly, various algorithms

for phase correction without potentially biased operator intervention were studied to obtain the spectrum that is now given by $I_\nu = \mathrm{Re}(\exp(i\lambda_{\mathrm{opt}})\kappa_\nu)$. In this paper, we exclusively utilize the maximum method [WKH+23], in which $\lambda_{\mathrm{opt}} \in \arg\max_{\lambda\in[0,\pi)} \|\mathrm{Re}(\exp(i\lambda)\kappa)\|$ is chosen so that the norm of I is maximal. In measurements where the spectrum $\hat{I}$ consists of little else than the central peak which carries no conformational information, the minimum method minimizing deviation of $\hat{\omega}$ from a parametric model of the wave has proven to be very effective in [PEH+21, HTW+22]. In both methods, additionally, a sign flip is performed when required to ensure that the spectrum's central peak points in the positive direction, effectively optimizing λ over $[0, 2\pi]$.

However, as ENDOR experiments often run for hours, in practice the aforementioned thermal drift can be substantial, see Figure C6 of the SI. Thus, in [PEH+21] the drift model was introduced, which allows for thermal drift of ψ and ϕ, decomposing the data matrix according to the homoscedastic drift model:

Definition 2.2 (Homoscedastic Drift Model)**.** *The homoscedastic drift model is given by*

$$Y_{b,\nu} = \psi_b + \phi_b\kappa_\nu + \epsilon_{b,\nu}, \quad \mathrm{vec}\,(\epsilon_{b,\nu}) \overset{i.i.d.}{\sim} \mathcal{N}(0,\Sigma),\ 1 \le \nu \le N,\ 1 \le b \le B. \tag{5}$$

By way of interpretation, $\psi \in \mathbb{C}^B$ represents the signal from electron paramagnetic resonance (i.e. what the echo signal would be if the RF pulse were absent) as well as a possible offset of the measurement apparatus, $\phi \in \mathbb{C}^B$ represents the magnitude and phase of the ENDOR effect, $\kappa \in \mathbb{C}^{N+1}$ comprises the ENDOR spectrum I as well as an orthogonal component ω which we call the wave (see panels A and B in Figure C1) and $\epsilon_{b,\nu}$ represents the experimental noise. We use the notation $\mathrm{vec}\,(\epsilon_{b,\nu}) = \begin{bmatrix}\mathrm{Re}\{\epsilon_{b,\nu}\}\\ \mathrm{Im}\{\epsilon_{b,\nu}\}\end{bmatrix}$ so that the noise components follow a bivariate normal distribution with positive definite symmetric covariance matrix $\Sigma \in \mathrm{SPD}(2)$.

The condition 3 serves to eliminate non-identifiability due to $\tilde{\kappa} = \kappa + c$, $\tilde{\psi} = \psi - c\phi$ with $\tilde{\psi}, \phi, \tilde{\kappa}, \Sigma$ yielding the same $Y_{\nu,b}$ as $\psi, \phi, \kappa, \Sigma$ for any $c \in \mathbb{C}$. Similarly, the condition 4 eliminates non-identifiability due to $\tilde{\kappa} = r\kappa$, $\tilde{\phi} = r^{-1}\phi$ with $\psi, \tilde{\phi}, \tilde{\kappa}, \Sigma$ yielding the same $Y_{\nu,b}$ as $\psi, \phi, \kappa, \Sigma$ for any $r \in \mathbb{R}_{>0}$.

Maximum likelihood estimators $\hat{\kappa}, \hat{\psi}, \hat{\phi}, \hat{\Sigma}$ are calculated (see [PEH+21] and Section 2.1 for details) and in a second step, the estimated spectrum $\hat{I} = \mathrm{Re}(e^{i\lambda_{\mathrm{opt}}}\hat{\kappa})$ and the orthogonal component $\hat{\omega} = \mathrm{Im}(e^{i\lambda_{\mathrm{opt}}}\hat{\kappa})$ are extracted from $\hat{\kappa}$ using the maximum method. Additionally to the above mentioned size non-identifiability of κ, the maximum (minimum) method and optional sign-flip eliminate the phase non-identifiability due to $\tilde{\kappa} = \alpha\kappa$, $\tilde{\phi} = \alpha^{-1}\phi$ yielding the same data distribution for $\psi, \phi, \kappa, \Sigma$ and $\psi, \tilde{\phi}, \tilde{\kappa}, \Sigma$ for all $\alpha \in \mathbb{C}$ with $|\alpha| = 1$.

The following data example illustrates that the homoscedastic drift model, Definition 2.2, exhibits unusually good fit to experimental data at $\nu_{\mathrm{MW}} = 263\,\mathrm{GHz}$, and yields improved SNR compared to the averaging model, Definition 2.1. Confidence regions are computed and will be justified via the $B \to \infty$ asymptotics developed in Section 4.2. It also

prepares for extension to the heteroscedastic drift model, Definition 6.1, for $\nu_{\mathrm{MW}} = 94\,\mathrm{GHz}$.

Data Example 2.3 (Homoscedastic Drift Model for 263 GHz dataset)**.** *The maximum likelihood estimates obtained using Algorithm 1 for the orientation g_y from a chemical sample of the D2-$Y_{122}^{\bullet}$ E. coli ribonucleotide reductase using the Davies pulse sequence, see [Dav74], studied in [PEH+21] are presented in Figure C1. This also includes pointwise confidence bands obtained via parametric bootstrap using* 10000 *bootstrap samples. In simulating data for the bootstrap, an additive bias correction for $\hat{\Sigma}$ and a multiplicative bias correction for $\hat{\phi}$ were used owing to substantial bias in these estimators. This bias, which does not disappear with increasing batch number B, likely arises from omitting the randomness in ϕ from the model as will be set out in detail in Section 5. A detailed analysis*

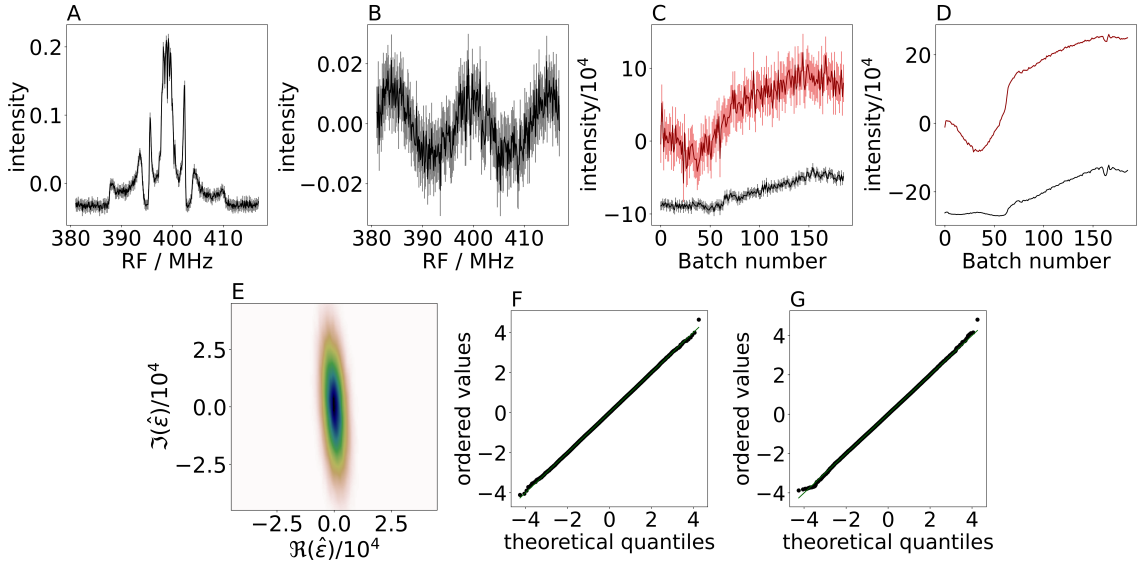


Figure C1: Applying the homoscedastic drift model (5) to an ENDOR example measurement. Panel A displays the estimated spectrum $\hat{I}$, while panel B displays the component $\hat{\omega}$ that is orthogonal to the estimated spectrum $\hat{I}$. C and D show the real (black) and imaginary (red) components of $\hat{\phi}$ and $\hat{\psi}$, respectively. Panel E displays the kernel-density-estimation of the complex residuals $\hat{\epsilon}_{b,\nu}$, while panels F and G depict q-q-plots for the real and imaginary components of the standardized residuals, respectively. The corresponding raw data are plotted in Figure C6 in the SI.

of all D2-$Y_{122}^{\bullet}$ measurements is presented in Appendix A. The real and imaginary parts of the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu$ (shown in panels E, F, and G of Figure C1 for orientation g_y) for all orientations are examined for goodness of fit using a Kolmogorov-Smirnov test, see [For07]. The resulting p-values are provided in Table C2 of the SI and are all clearly above the Bonferroni-corrected critical value of 0.05/10=0.005. Additionally, a comparison of the SNR of the averaging and drift models is performed. The drift model exhibits a better SNR than the averaging model in 4 out of 5 measurements, as shown in Table C1 and Figure C7. This is attributable to partial cancellation in (1) of ENDOR signal components when ϕ_b changes significantly over batches. As an extreme example, $\phi_b = \exp(ib/B)$ would lead to the extracted spectrum from the averaging model being nothing

but noise. Indeed, the larger the drift of $\arg\phi_b$ *shown in Figure C8, the more pronounced the SNR advantage of the drift model over the averaging model in Table C1. This drift in* ϕ *is positively correlated with the drift observed in* ψ*, see panels C and D of Figure C1. While the correlation is not perfect, which argues against including it as a fixed component of the model, it points to the dominant source of drift for both* ψ *and* ϕ *originating from phase and amplitude changes due to thermal drift of the MW coupling to the ENDOR resonator containing the chemical sample.*

2.1 Maximum Likelihood Estimation and Parameter Space

Based on the statistical model 5, the log likelihood is easily found to be

$$\ell_Y(\psi, \phi, \kappa, \Sigma) = -\frac{B(N+1)}{2} \log\left((2\pi)^2 \det(\Sigma)\right) - \frac{1}{2} \sum_{b=1}^{B} \left\| \tilde{Y}_{b,:} - \phi_b \kappa \right\|_P^2, \tag{6}$$

where the precision matrix $P \coloneqq \Sigma^{-1}$ and the centered data matrix $\tilde{Y}_{b,\nu} \coloneqq Y_{b,\nu} - \hat{\psi}_b$ have been used. Note that, contrary to rank one principal component analysis (PCA) where interest is in the direction κ of greatest variability *across repeated measurements*, our interest is in a measure of central tendency for κ that shows greatest variability *across frequencies*. Hence, we centre the data to achieve zero empirical row mean (removing $\hat{\psi}_b = \frac{1}{N}\sum_{\nu=0}^{N} Y_{b,\nu}$) rather than zero empirical column mean. The interest in the direction of greatest variability across frequencies manifests itself in the use of the maximum method to estimate λ_{opt} which, given a direction $\hat{\kappa} \in \mathbb{C}^{N+1}$, selects the phase in the complex plane in which the variability across frequencies is greatest. However, we will see shortly that the proposed model is not equivalent to PCA of the transpose of the data matrix.

For each parameter, the MLE when assuming all other parameters known is available in closed form (see [PEH+21]). Note that $\phi \diamond_P \phi$ and $\kappa \diamond_P \kappa$ are invertible due to Lemma D.4 assuming $\|\phi\| > 0$.

$$\hat{\kappa}_\nu(\phi, P, \tilde{Y}) = \mathfrak{c}\left((\phi \diamond_P \phi)^{-1} \left(\phi \bullet_P \tilde{Y}_{:,\nu} \right) \right) \tag{7}$$

$$\hat{\phi}_b(\kappa, P, \tilde{Y}) = \mathfrak{c}\left((\kappa \diamond_P \kappa)^{-1} \left(\kappa \bullet_P \tilde{Y}_{b,:} \right) \right) \tag{8}$$

$$\hat{\Sigma}(\phi, \kappa, \tilde{Y}) = \frac{1}{B(N+1)} \sum_{b=1}^{B} \sum_{\nu=0}^{N} \left(\mathrm{vec}(\tilde{Y}_{b,\nu}) - M(\phi_b)\mathrm{vec}(\kappa_\nu)\right)\left(\mathrm{vec}(\tilde{Y}_{b,\nu}) - M(\phi_b)\mathrm{vec}(\kappa_\nu)\right)^T. \tag{9}$$

In the special case when $\Sigma = r\mathrm{Id}_2$ for $r \in \mathbb{R}_{>0}$ is known, this reduces to a rank one singular value decomposition (SVD) of the centered data matrix $\tilde{Y}$ with $\frac{\hat{\phi}}{\|\hat{\phi}\|}$ and $\bar{\hat{\kappa}}$ being left and right singular vectors and $\left\|\hat{\phi}\right\|$ the leading singular value, respectively. Therefore, an iterative method imitating the standard power iteration method [TI97] would be a natural algorithm to solve this problem. Indeed, [PEH+21] iterate the formulae 7, 8, 9 to numerically compute the MLE even though they solve the more general and practically relevant case involving given correlated, non-isotropic noise ϵ. In this more general case,

there is no simple analogy to the SVD and its well-established asymptotic theory, see e.g. [And03b], is not applicable.

While the entries of $\tilde{Y}$ are complex numbers, the metric implied by the presence of the $\|\cdot\|_P$-norm in the log likelihood and the fact that the 2×2 matrix $\kappa \diamond_P \kappa$ occurring in the conditional MLE 8 cannot generally be written as the matrix representation $M(c)$ of any complex number $c \in \mathbb{C}$, suggest a different approach. It is possible to conceive of the entries of $\tilde{Y}$ as 2×2 matrices $M(\tilde{Y}_{b,\nu})$ so that computation takes place on the ring R of 2×2 real matrices. κ is then an element of the Hilbert module $(R^{N+1}, \langle \cdot, \cdot \rangle_P)$. A Cauchy-Schwartz type inequality is available on this Hilbert module [Bul82] which would facilitate some of our analysis but we ultimately perceive this algebraic sophistication as a hindrance rather than as a simplification.

Instead, we initially view $\kappa \in \mathbb{C}^{N+1}$ subject to the constraints 3 and 4 as an element of an $N+1$ dimensional complex sphere intersected with the hyperplane defined by 3. Removing an additional phase factor (since κ and $\alpha\kappa$ lead to equivalent models for $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ as previously noted), we are naturally lead to identifying those κ that differ only by a root of unity and hence arrive at the complex projective space $\mathfrak{P} = \mathbb{C}P^{N-1}$, a Riemannian manifold of real dimension $2(N-1)$ as the relevant parameter space for κ, where the Riemannian metric tensor is implied by the natural quotient embedding in $\mathbb{C}^N$.

We additionally choose a new basis that deals with the constraint 3 by re-writing the noise ϵ according to $\tilde{Y}_{b,\nu} = \phi_b \kappa_\nu + \tilde{\epsilon}_{b,\nu}$ where

$$\tilde{\epsilon}_{b,\nu} \coloneqq \frac{N}{N+1}\epsilon_{b,\nu} - \frac{1}{N+1}\sum_{\tilde{\nu}=0,\tilde{\nu}\neq\nu}^{N} \epsilon_{b,\tilde{\nu}}.$$

Now, we transition from the standard basis vectors $e_k \in \mathbb{R}^{N+1}$ to the Helmert orthonormal basis vectors

$$h_j \coloneqq \frac{1}{\sqrt{j(j+1)}}\left(\left(\sum_{k=1}^{j} e_k\right) - j e_{j+1}\right), \quad j = 1, \ldots, N$$

to form the Helmert sub-matrix $H = (h_1, \ldots, h_N)^T \in \mathbb{R}^{(N+1)\times N}$ which is in turn used to Helmertize the data matrix $\tilde{Y}^H \coloneqq H\tilde{Y}$, error $\tilde{\epsilon}^H \coloneqq H\tilde{\epsilon}$ and spectral parameter $\kappa^H \coloneqq H\kappa$, see [DM98] for details on this standard approach. While the covariance structure of $\tilde{\epsilon}$ is slightly cumbersome, that of the Helmertized error is simply $\mathrm{vec}\left(\tilde{\epsilon}^H_{b,\nu}\right) \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma)$ as shown in Lemma B.1 in the SI.

3 Extending strong consistency for generalized Fréchet means

For our purpose in Section 4 to infer geometric parameters of the drift model, in this section we extend the strong law of large numbers for *generalized Fréchet means*, which is usually called *strong consistency* in this context. Let us first introduce this underlying concept.

Noting that expected values of a random variable in a linear space are equivalently described as minimizers of expected squared distance, Fréchet [Fré48] used this latter geometric property as a definition for a *mean location* (cf. [HL98]) on a metric space, which was soon called the *Fréchet mean* in his honor [Kar14]. Further, medians, as minimizers of expected distance lead to *Fréchet medians* on metric spaces [FVJ08], and more generally, any L^p mean can thus be generalized [Afs11]. For Fréchet means, two version of set-valued strong consistency under rather broad conditions have been shown [Zie77, BP03], followed by more versions of strong consistency for Fréchet L^p means by [EJ20, Sch22]. The general formulation in terms of set valued Fréchet means is necessary for data on non-Euclidean spaces, since for example for a sphere with equal point masses on the north and south pole the Fréchet mean set is the whole equator, cf. [Huc12]. The generalized consistency results also apply to *generalized Fréchet means* introduced by [Huc11b] to extend and model geometric data descriptors beyond location, such as principal components of the covariance. For instance in a geodesic space, the first principal component can be generalized to a best approximating geodesic. Notably, then minimization has to be conducted no longer over the data space, but over a descriptor space, in case of geodesics, this is the space of geodesics. Likewise, parameters of a parametric model can be viewed as generalized Fréchet means.

Curiously, to the best knowledge of the authors, available strong consistency results for generalized Fréchet means ([Sch22, Huc11b]) always assume a loss function which is bounded from below and thus do not cover the simple case of maximum likelihood parameters of a univariate or multivariate Gaussian. Such a generalization is typically necessary to cover cases where a generalized Fréchet mean is estimated along with a (co-)variance-like quantity. This is for example the case for diffusion means with simultaneously estimated variance, see [EHHS22], and for a generalization of the asymptotic theory of the drift model to include the covariance of the noise, as discussed in Section 5. As it turns out such a generalization of strong consistency results is possible with moderate effort.

For all of the following, let $X_1, X_2, \cdots \sim X$ be i.i.d. be random elements mapping from a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ into a topological space $\mathfrak{Q}$ equipped with its Borel σ-algebra, called the *data space*. Moreover, let $(\mathfrak{P}, d)$ be a separable metric space, called the *parameter space*.

Definition 3.1 (Sample and Population and Fréchet ρ-mean)**.** *With a function $\rho : \mathfrak{Q} \times \mathfrak{P} \mapsto \mathbb{R}$ which is continuous in $\mathfrak{P}$ for all fixed $q \in \mathfrak{Q}$ and measurable in $\mathfrak{Q}$ for all fixed*

$p \in \mathfrak{P}$, *define, if existent,*

$$\mathcal{F}_n^{(\rho)}(\omega, p) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \rho(X_i(\omega), p), \qquad \mathcal{F}^{(\rho)}(p) \coloneqq \mathbb{E}\left(\rho(X, p)\right),$$
$$\ell_n(\omega) \coloneqq \inf_{p\in\mathfrak{P}} \mathcal{F}_n(\omega, p), \qquad \ell \coloneqq \inf_{p\in\mathfrak{P}} \mathcal{F}(p),$$
$$E_n^{(\rho)}(\omega) \coloneqq \{p \in \mathfrak{P} \mid \mathcal{F}_n(p) = \ell_n(\omega)\}, \qquad E^{(\rho)} \coloneqq \{p \in \mathfrak{P} \mid \mathcal{F}^{(\rho)}(p) = \ell\}.$$

The functions $\mathcal{F}^{(\rho)}$ and $\mathcal{F}_n^{(\rho)}$ are called the population and sample Fréchet ρ-functions*, respectively, and $E^{(\rho)}$ and $E_n^{(\rho)}$ are the sets of* sample and population Fréchet ρ-means*, respectively.*

Definition 3.1 is a generalization of a mean originally introduced for the case $\mathfrak{P} = \mathfrak{Q}$ and $\rho = d^2$ by [Fré48] which is called the *Fréchet mean*, see above. Due to continuity of ρ, $E^{(\rho)}$ is a closed set and $E_n^{(\rho)}(\omega)$ is a random closed set, introduced and studied by [Cho54, Ken74, Mat74], see also [Mol05].

Definition 3.2 (Two versions of set strong consistency)**.** *We say that the estimator $E_n^{(\rho)}(\omega)$ for $E^{(\rho)}$ is*

ZC: Ziezold strongly consistent *if*

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \subseteq E^{(\rho)} \text{ for all } \omega \in \Omega \text{ almost surely,}$$

BPC: Bhattacharya and Patrangenaru strongly consistent *if $E^{(\rho)} \neq \emptyset$ and if for every $\epsilon > 0$ and almost surely for all $\omega \in \Omega$ there is a number $n = n(\epsilon, \omega) > 0$ such that*

$$\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega) \subseteq \{p \in \mathfrak{P} : d(E^{(\rho)}, p) \leq \epsilon\}.$$

Remark 3.3. *ZC was originally introduced by [Zie77] and established in case of $\mathfrak{P} = \mathfrak{Q}$ and ρ a squared quasi-metric. BP was originally introduced by [BP03] and established for Fréchet means on Heine-Borel spaces under the additional condition that $E^{(\rho)}$ be not empty. More generally, [EJ20] put the two concepts of strong consistency into the more general context of Kuratowski limits, see also [Sch22] (ZC corresponds to outer limits there and BPC to limits in one-sided Hausdorff distance).*

As noted above, Fréchet ρ-means for nonnegative ρ have been introduced by [Huc11b], studying both versions of consistency under a uniform continuity and a coercivity assumption on ρ. [Sch22] relaxed these assumptions, among others to lower semicontinuity and some assumtpions on bounds. We show ZC and BPC under even weaker assumptions, namely a modulus of continuity along with its prefactor for ZC and using non-emptiness of E^{ρ} for BPC.

Assumption 3.4. *In the setup of Definition 3.1 there are*

1. $\dot\rho : \mathfrak{Q} \times \mathfrak{P} \mapsto [0,\infty)$ *which is continuous in* $\mathfrak{P}$ *for all fixed* $q \in \mathfrak{Q}$ *and measurable in* $\mathfrak{Q}$ *for all fixed* $p \in \mathfrak{P}$*, with* $\mathbb{E}\left[\dot\rho(X,p)\right] < \infty$ *for all* $p \in \mathfrak{P}$*,*

2. $h : [0,\infty) \to [0,\infty)$ *continuous with* $h(0) = 0$*, and*

3. $\delta > 0$*, such that for every* $p, p' \in \mathfrak{P}$ *with* $d(p,p') < \delta$

$$\left|\rho(q,p) - \rho(q,p')\right| \le \dot\rho(q,p)\; h\big(d(p,p')\big)\,. \tag{10}$$

Further, assume that $\mathbb{E}\left(\rho(X,p)\right)$ *exists for all* $p \in \mathfrak{P}$*.*

Definition 3.5. *For* $\omega \in \Omega, p \in \mathfrak{P}$*, under Assumption 3.4, define*

$$\dot{\mathcal{F}}_n(\omega,p) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \dot\rho(X_i(\omega),p), \qquad \dot{\mathcal{F}}(p) \coloneqq \mathbb{E}\left(\dot\rho(X,p)\right).$$

Lemma 3.6. *Under Assumption 3.4 there is a dense countable subset* $\tilde{\mathfrak{P}} \subset \mathfrak{P}$ *and measureable* $A \subset \Omega$ *with* $\mathcal{P}(A) = 1$ *such that for all* $\tilde p \in \tilde{\mathfrak{P}}$ *and all* $\omega \in A$ *the following hold:*

(i) $\mathcal{F}_n(\omega,\tilde p) \overset{n\to\infty}{\to} \mathcal{F}(\tilde p)$ *and* $\dot{\mathcal{F}}_n(\omega,\tilde p) \overset{n\to\infty}{\to} \dot{\mathcal{F}}(\tilde p)$*,*

(ii) for all $p \in \mathfrak{P}$ *with* $d(p,\tilde p) < \delta/2$ *and* $(p_n)_{n=1}^{\infty} \subset \mathfrak{P}$ *with* $p_n \to p$*,*

$$\begin{aligned}\mathcal{F}(\tilde p) - h\big(d(\tilde p,p)\big)\,\dot{\mathcal{F}}(\tilde p) &\le \liminf_{n\to\infty} \mathcal{F}_n(\omega,p_n)\\ &\le \limsup_{n\to\infty} \mathcal{F}_n(\omega,p_n) \le \mathcal{F}(\tilde p) + h\big(d(\tilde p,p)\big)\,\dot{\mathcal{F}}(\tilde p).\end{aligned}$$

Proof. Since $\mathfrak{P}$ is a separable space, there is a countable subset $\tilde{\mathfrak{P}} = \{\tilde p_i\}_{i=1}^{\infty} \subset \mathfrak{P}$ that is dense in $\mathfrak{P}$. For every $\tilde p_i \in \tilde{\mathfrak{P}}$ there is, due to the classical strong law of large numbers, a measurable set $A_i \in \mathcal{A}$ with $\mathcal{P}(A_i) = 1$ such that

$$\mathcal{F}_n(\omega,\tilde p_i) \overset{n\to\infty}{\to} \mathcal{F}(\tilde p_i) \quad \text{and} \quad \dot{\mathcal{F}}_n(\omega,\tilde p_i) \overset{n\to\infty}{\to} \dot{\mathcal{F}}(\tilde p_i) \quad \text{for every} \quad i = 1,2,\ldots \quad \text{and} \quad \omega \in A_i.$$

Thus, for $A \coloneqq \bigcap_{i=1}^{\infty} A_i$ we have Assertion (i).

In order to see Assertion (ii), consider $\omega \in A$, $p, p_n \in \mathfrak{P}$ with $p_n \to p$ and $\tilde p \in \tilde{\mathfrak{P}}$ with $d(p,\tilde p) < \delta/2$ and $\delta > 0$ from Assumption 3.4. Then, there is $n_0 \in \mathbb{N}$ with $d(p,p_n) < \delta/2$ for all $n \ge n_0$, and hence $d(p_n,\tilde p) < \delta$ for all $n \ge n_0$ (illustrated in the left panel of Figure C2). Thus

$$\mathcal{F}_n(\omega,\tilde p) - \left|\mathcal{F}_n(\omega,\tilde p) - \mathcal{F}_n(\omega,p_n)\right| \le \mathcal{F}_n(\omega,p_n) \le \mathcal{F}_n(\omega,\tilde p) + \left|\mathcal{F}_n(\omega,\tilde p) - \mathcal{F}_n(\omega,p_n)\right|, \tag{11}$$

and from Assumption 3.4 we have for all $n \geq n_0$

$$|\mathcal{F}_n(\omega, \tilde{p}) - \mathcal{F}_n(\omega, p_n)| \leq \frac{1}{n} \sum_{i=1}^{n} |\rho(X_i(\omega), \tilde{p}) - \rho(X_i(\omega), p_n)|$$
$$\leq h\big(d(p_n, \tilde{p})\big) \frac{1}{n} \sum_{i=1}^{n} \dot{\rho}(X_i(\omega), \tilde{p}) = h\big(d(p_n, \tilde{p})\big)\, \dot{\mathcal{F}}_n(\tilde{p}). \tag{12}$$

Letting $n \to \infty$ in (11), exploiting (12), continuity of d, continuity of h, $h(0) = 0$, $h \geq 0$ and Assertion (i) yield at once Assertion (ii). □

Theorem 3.7. *Under Assumption 3.4, ZC holds for the set of Fréchet ρ-means on $\mathfrak{P}$.*

Proof. We follow the steps originally introduced by [Zie77] and adopted by [Huc11b]. With $A \subset \Omega$ of full measure and the dense countable subset $\tilde{\mathfrak{P}}$ of $\mathfrak{P}$, both from Lemma 3.6, fix $p \in \mathfrak{P}$ and $(p_n)_{n=1}^{\infty} \subset \mathfrak{P}$ with $p_n \to p$. We first show that

$$\mathcal{F}_n(\omega, p_n) \overset{n\to\infty}{\to} \mathcal{F}(p)\,, \tag{13}$$

for all $\omega \in A$.

To this end, with $\delta > 0$ from Assumption 3.4, let $\tilde{p} \in \tilde{\mathfrak{P}}$ with $d(p, \tilde{p}) < \delta/2$. Then, due to Assertion (ii) from Lemma 3.6,

$$\mathcal{F}(\tilde{p}) - h\big(d(\tilde{p}, p)\big)\, \dot{\mathcal{F}}(\tilde{p}) \leq \liminf_{n\to\infty} \mathcal{F}_n(\omega, p_n) \leq \limsup_{n\to\infty} \mathcal{F}_n(\omega, p_n) \leq \mathcal{F}(\tilde{p}) + h\big(d(\tilde{p}, p)\big)\, \dot{\mathcal{F}}(\tilde{p}), \tag{14}$$

for all $\omega \in A$.

Letting $(\tilde{p}_k)_{k=1}^{\infty} \subset \tilde{\mathfrak{P}}$ with $\tilde{p}_k \overset{k\to\infty}{\to} p$, there is $k_0 \in \mathbb{N}$ with $d(p, \tilde{p}_k) < \delta/2$ for all $k \geq k_0$ (illustrated in the right panel of Figure C2). Plugging these in, into (14) we obtain for all $\omega \in A$,

$$\begin{aligned} \lim_{k\to\infty} \Big(\mathcal{F}(\tilde{p}_k) - h\big(d(\tilde{p}_k, p)\big)\, \dot{\mathcal{F}}(\tilde{p}_k)\Big) &\leq \liminf_{n\to\infty} \mathcal{F}_n(\omega, p_n) \\ &\leq \limsup_{n\to\infty} \mathcal{F}_n(\omega, p_n) \\ &\leq \lim_{k\to\infty} \Big(\mathcal{F}(\tilde{p}_k) + h\big(d(\tilde{p}_k, p)\big)\, \dot{\mathcal{F}}(\tilde{p}_k)\Big)\,. \end{aligned} \tag{15}$$

This yields (13), as, due to continuity of $\mathcal{F}, \dot{\mathcal{F}}$ and h, as well as $h(0) = 0$,

$$\lim_{k\to\infty} \Big(\mathcal{F}(\tilde{p}_k) - h\big(d(\tilde{p}_k, p)\big)\, \dot{\mathcal{F}}(\tilde{p}_k)\Big) = \mathcal{F}(p) = \lim_{k\to\infty} \Big(\mathcal{F}(\tilde{p}_k) + h\big(d(\tilde{p}_k, p)\big)\dot{\mathcal{F}}(\tilde{p}_k)\Big)\,.$$

Next we show the assertion of the theorem. Since it is trivial in case of

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} = \emptyset,$$

it is sufficient to show that

$$\text{if } \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \neq \emptyset \quad \text{then} \quad \ell_n(\omega) \to \ell \quad \text{for} \quad \omega \in A.$$

To see this, we show the following two inequalities for all $\omega \in A$

$$\liminf_{n\to\infty} \ell_n(\omega) \geq \ell, \tag{16}$$

$$\limsup_{n\to\infty} \ell_n(\omega) \leq \ell\,. \tag{17}$$

Noting that

$$\text{if} \quad p \in \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \quad \text{then} \quad p \in \overline{\bigcup_{k=j}^{\infty} E_{n_k}^{(\rho)}(\omega)} \quad \text{for all} \quad j \in \mathbb{N}\,,$$

where $n_j \to \infty$ is a sequence with

$$\lim_{j\to\infty} \ell_{n_j}(\omega) = \liminf_{n\to\infty} \ell_n(\omega)\,,$$

and recalling that the closure of a set in a metric space is given by all cluster points of sequences in it, there is a sequence $\{p_i\}_{i=1}^{\infty}$ with $p_i \to p$ and $p_i \in E_{k_i}^{(\rho)}(\omega)$ for a subsequence $\{k_i\}_{i=1}^{\infty}$ of n_j. Using (13) we obtain

$$\liminf_{n\to\infty} \mathcal{F}_{k_n}(\omega, p_n) = \lim_{i\to\infty} \ell_{k_i}(\omega) = \mathcal{F}(p) \geq \ell.$$

for all $\omega \in A$, yielding (16).

To see (17), set $p_n := p$ for some $p \in E^{(\rho)}$, so that with (13) there is a nonnegative random sequence $\{\epsilon_n(\omega)\}_{n=1}^{\infty}$, converging to zero for all $\omega \in A$, with

$$\ell = \mathcal{F}(p) \geq \mathcal{F}_n(\omega, p) - \epsilon_n(\omega) \geq \ell_n(\omega) - \epsilon_n(\omega)\,,$$

for all $\omega \in A$, yielding at once (17). This completes the proof. □

Assumption 3.8. *The population Fréchet ρ-mean is not empty: $E^{(\rho)} \neq \emptyset$ and for all random sequences $\{p_n\}_{n\in\mathbb{N}}$ without accumulation points in $\mathfrak{P}$, there is a constant $\ell < C \leq \infty$ such that a.s.*

$$\liminf_{n\to\infty} \rho(X, p_n) \geq C. \tag{18}$$

Theorem 3.9. *Under Assumptions 3.4, and 3.8 BPC holds for the set of Fréchet ρ-means on $\mathfrak{P}$.*

Proof. It suffices to show that for random sequence $p_n(\omega) \in E_n^{(\rho)}(\omega)$ with underlying

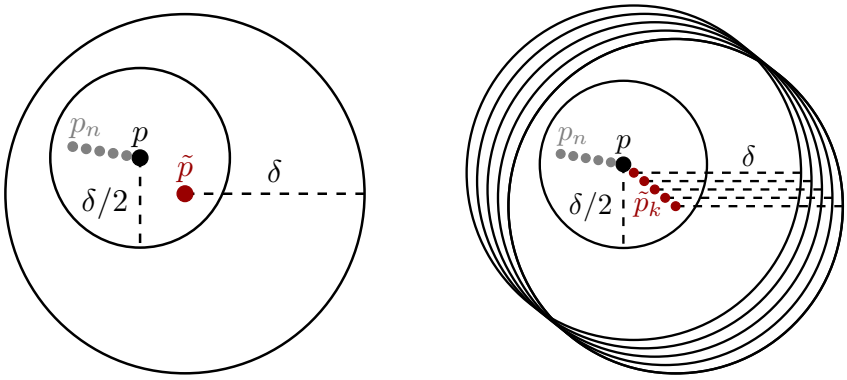


Figure C2: Left: Illustration of Lemma 3.6. For all $\tilde{p} \in \tilde{\mathfrak{P}}$ (red point) with distance $d(p,\tilde{p}) < \delta/2$ to $p \in \mathfrak{P}$ (black point), the inequality (11) holds for $(p_n)_{n=1}^{\infty} \subset \mathfrak{P}$ with $p_n \to p$ (gray points). Right: Illustration of (15) in the proof of Theorem 3.7. Compared to the left panel, where $\tilde{p} \in \mathfrak{P}$ is fixed, the sequence $(\tilde{p}_k)_{k=1}^{\infty} \subset \tilde{\mathfrak{P}}$ (red dots) converges also to p (black dot) for $k \to \infty$.

random one-sided Hausdorff distance $r_n(\omega)$, i.e.

$$p_n(\omega) \in \operatorname*{arg\,max}_{p \in E_n^{(\rho)}(\omega)} \left(\min_{p' \in E^{(\rho)}} d(p,p') \right), \qquad r_n(\omega) = \max_{p \in E_n^{(\rho)}(\omega)} \left(\min_{p' \in E^{(\rho)}} d(p,p') \right), n \in \mathbb{N}$$

$r_n(\omega) \to 0$ for all $\omega \in \Omega$ a.s.

If this was not the case, then there would be $\tilde{A} \subset \Omega$ with $\mathcal{P}(\tilde{A}) > 0$ such that for all $\omega \in \tilde{A}$, there is a subsequence $n_k(\omega)$ with $r_{n_k(\omega)}(\omega) \geq r_0(\omega) > 0$ and $r_0(\omega) > 0$. We now derive a contradiction.

Due to Theorem 3.7, we have ZC, so that with a null set B, all cluster points of $p_{n_k(\omega)}(\omega)$ lie in $E^{(\rho)}$ for all $\omega \in \tilde{A} \setminus B$. In consequence, $p_{n_k(\omega)}(\omega)$ has no cluster points for all $\omega \in \tilde{A} \setminus B$. Fixing $\omega_0 \in \tilde{A} \setminus B$, set

$$\tilde{p}_k(\omega) := \begin{cases} p_{n_k(\omega)}(\omega) & \text{if} \quad \omega \in \tilde{A} \setminus B \\ p_{n_k(\omega_0)}(\omega_0) & \text{if} \quad \omega \in \Omega \setminus (\tilde{A} \setminus B) \end{cases}, k \in \mathbb{N},$$

to obtain a sequence $\tilde{p}_k(\omega)$ without any cluster points for all $\omega \in \Omega$, so that in consequence of Assumption 3.8, there is $C > \ell$ and a.s. $m(\omega) \in \mathbb{N}$ such that

$$\rho\left(X(\omega), \tilde{p}_k(\omega)\right) \geq C$$

for all $k \geq m(\omega)$, almost surely. Hence, by construction, with a null set $\tilde{B} \subset \Omega$,

$$\begin{aligned} \mathcal{F}_{n_k(\omega)}\left(\omega, p_{n_k(\omega)}(\omega)\right) &= \frac{1}{n_k(\omega)} \sum_{j=1}^{n_k(\omega)} \rho\left(X_j(\omega), p_{n_k(\omega)}(\omega)\right) \\ &\geq C \text{ for all } \omega \in \tilde{A} \setminus \tilde{B}, \text{ if } n_k(\omega) > m(\omega). \end{aligned} \tag{19}$$

By hypothesis there is $p \in E^{(\rho)}$ with, due to the strong law of large numbers,

$$\mathcal{F}_{n_k(\omega)}(\omega, p) \overset{a.s.}{\to} \mathcal{F}(p) = \ell < C,$$

by construction, i.e. there is $n(\omega) \in \mathbb{N}$ such that $\mathcal{F}_{n_k(\omega)}(\omega, p) < C$ for all $n_k(\omega) > n(\omega)$, a.s. In conjunction, with (19), letting $n_k(\omega) > \max\{n(\omega), m(\omega)\}$ we have thus

$$\mathcal{F}_{n_k(\omega)}(\omega, p) < C \leq \mathcal{F}_{n_k(\omega)}\big(\omega, p_{n_k(\omega)}(\omega)\big)$$

on a set of positive measure, a contradiction to $p_{n_k(\omega)}(\omega) \in E^{\rho}_{n_k(\omega)}(\omega)$, completing the proof.

□

Remark 3.10. *The BPC version of the strong law in the literature usually requires that a Heine-Borel property, e.g. [BP03,EJ20,Sch22]. If $(\mathfrak{P}, d)$ satisfies the Heine-Borel property, all sequences without accumulation points diverge, so Assumption 3.8 holds for all $\omega \in \Omega$. If $\mathfrak{P}$ is compact, then Assumption 3.8 holds and BPC follows immediately from ZC.*

4 Strong consistencies and CLTs for the homoscedastic drift model

In this section, the theory developed in Section 3 is applied to prove strong consistency for the homoscedastic drift model as the number B of batches tends to infinity. For this purpose, we reformulate in Definition 4.1 the homoscedastic drift model from Definition 2.2 making ϕ explicitly stochastic and assume it to be i.i.d. Without loss of generality, we consider Y centered, i.e. ψ has been subtracted and the basis has been transformed with the Helmert sub-matrix as described in Section 2.1. For ease of notation, we assume that the original random variable is $N+1$ dimensional, so that Y is N dimensional, and we omit writing the tilde and the superscript ^{H} symbol from Section 2.1.

Then, below in Section 4.1 the modulus of continuity h with its prefactor from Assumption 3.4 is explicitly calculated and ZC is proven. With a little more effort it is shown that the population Fréchet mean $E^\rho = \{[\kappa^{(0)}]\}$ is unique, in order to establish BPC. Finally, a central limit theorem for $\hat{\kappa}$ is shown in Section 4.2 and one for $\hat{I}$ in Section 4.3.

Definition 4.1 (Centered Homoscedastic Drift Model)**.** *The complex N-dimensional random vector Y is given by*

$$Y = \phi\kappa^{(0)} + \epsilon \tag{20}$$

where $\kappa^{(0)} \in \mathcal{S}^{2N-1} \coloneqq \{\kappa \in \mathbb{C}^N : ||\kappa|| = 1\}$ comprises the true but unknown ENDOR spectrum, ϕ is a complex random value and $\epsilon = (\epsilon_1, \ldots, \epsilon_N)$ is a complex N-dimensional random vector independent of ϕ with

$$0 < \mathbb{E}[\mathrm{vec}\,(\phi)^T\,\mathrm{vec}\,(\phi)] = c_\phi < \infty, \qquad \mathrm{vec}\,(\epsilon_\nu) \overset{i.i.d.}{\sim} \mathcal{N}(0,\Sigma), 1 \leq \nu \leq N.$$

Moreover, we assume that the precision matrix $P = \Sigma^{-1}$ has two positive eigenvalues $\lambda_1 > \lambda_2 > 0$.

Thus, $\mathfrak{Q} = \mathbb{C}^N$ is the data space, as descriptor space we choose $\mathfrak{P} = \mathbb{C}P^{N-1}$ and for the loss we choose

$$\rho : \mathfrak{Q} \times \mathfrak{P} \to \mathbb{R}, \quad (Y,[\kappa]) \mapsto d_P\left(Y, \hat{\phi}(\kappa,P,Y)\kappa\right)^2, \tag{21}$$

with

$$\hat{\phi}(\kappa,P,Y) = \mathfrak{c}\left((\kappa \diamond_P \kappa)^{-1}(\kappa \bullet_P Y)\right) \in \mathbb{C}. \tag{22}$$

Thus, for a sample $Y(1), Y(2), \ldots \overset{i.i.d.}{\sim} Y$ from (20) we have the following sample and population Fréchet functions

$$\mathcal{F}_B(\omega,[\kappa]) = \frac{1}{B}\sum_{b=1}^{B}\rho(Y(b),[\kappa]), \qquad \mathcal{F}([\kappa]) = \int \rho(Y,[\kappa])\,\mathrm{d}\mathbb{P}\,(\phi,\epsilon)\,.$$

Remark 4.2. *Note that (22) is the MLE (8) for a single $b \in \{1, \dots, B\}$ of the homoscedastic drift model (5).*

Further, note that (21) is well defined, since for $\kappa, \tilde{\kappa} \in [\kappa]$ there is $\lambda \in \mathbb{R}$ such that $\tilde{\kappa} = e^{i\lambda}\kappa$, whence

$$\hat{\phi}(\tilde{\kappa}, P, Y) = e^{-i\lambda}\mathfrak{c}\left((\kappa \diamond_P \kappa)^{-1}(\kappa \bullet_P Y)\right)$$

yielding

$$\hat{\phi}(\tilde{\kappa}, P, Y)\tilde{\kappa} = \hat{\phi}(\kappa, P, Y)\kappa.$$

4.1 Strong consistencies for the centered homoscedastic drift model

Here, we establish that, as more data is accumulated and hence $B \to \infty$, the estimator $[\hat{\kappa}]$ arising from the centered homoscedastic drift model in Definition 4.1, i.e. the generalized Fréchet mean, is strongly consistent in the sense of ZC and BPC.

Theorem 4.3. *Under Assumption 4.1 ZC holds for the centered homoscedastic drift model. In particular, in Assumption 3.4 the modulus of continuity can be chosen as*

$$h([\kappa], [\kappa']) = d([\kappa], [\kappa'])$$

with prefactor

$$\dot{\rho}(Y, P) := \sqrt{\lambda_1}\left(\frac{\lambda_1^2 + \lambda_2^2}{\lambda_1\lambda_2}\right)\left((\lambda_1 + 2)\sqrt{2N} + 8\sqrt{2}N + \frac{32\sqrt{2}N\left(\lambda_1^2 + \lambda_2^2\right)}{\lambda_1\lambda_2}\right)||Y||^2.$$

Proof. Let $[\kappa], [\kappa'] \in \mathfrak{P}$ and $\kappa \in [\kappa], \kappa' \in [\kappa']$ arbitrary. Recalling

$$\rho(Y, [\kappa]) = \langle Y, Y\rangle_P - \left\langle \hat{\phi}(\kappa, P, Y)\kappa, Y\right\rangle_P.$$

from Lemma D.1 and using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}\left|\rho(Y, [\kappa]) - \rho(Y, [\kappa'])\right| &= \left|\left\langle \hat{\phi}(\kappa', P, Y)\kappa' - \hat{\phi}(\kappa, P, Y)\kappa, Y\right\rangle_P\right| \\ &\leq \sqrt{\langle Y, Y\rangle_P}\, d_P\left(\hat{\phi}(\kappa, P, Y)\kappa, \hat{\phi}(\kappa', P, Y)\kappa'\right).\end{aligned}$$

Since λ_1 is the largest eigenvalue of P, by definition of the Mahalanobis inner product in Section 1.1, the first term of the bottom line above is bounded by

$$\sqrt{\langle Y, Y\rangle_P} = \sqrt{\sum_{\nu=1}^{N} \mathrm{vec}\,(Y_\nu)^T R \begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix} R^T \mathrm{vec}\,(Y_\nu)} \leq \sqrt{\lambda_1}\,\|Y\|. \tag{23}$$

A bound for the second term can be obtained from Lemma D.17 in conjunction with

Lemmata D.5 and D.11,

$$\begin{aligned}
&d_P\Big(\hat{\phi}(\kappa,P,Y)\kappa,\hat{\phi}(\kappa',P,Y)\kappa'\Big)\\
&\leq\left(\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\right)\left(\lambda_1\sqrt{2N}+8\sqrt{2}N+\frac{32\sqrt{2}N\left(\lambda_1^2+\lambda_2^2\right)}{\lambda_1\lambda_2}+2\sqrt{2N}\right)||Y||\,\big|\big|\kappa-\kappa'\big|\big|\,.
\end{aligned}$$

In consequence, since $\kappa\in[\kappa],\kappa'\in[\kappa']$ have been arbitrary,

$$\begin{aligned}
&\big|\rho(Y,[\kappa])-\rho(Y,[\kappa'])\big|\\
&\leq\sqrt{\lambda_1}\left(\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\right)\left((\lambda_1+2)\sqrt{2N}+8\sqrt{2}N+\frac{32\sqrt{2}N\left(\lambda_1^2+\lambda_2^2\right)}{\lambda_1\lambda_2}\right)||Y||^2\,d([\kappa],[\kappa'])\,,
\end{aligned}$$

yielding the assertion on h and $\dot{\rho}$.

Further, since $\mathfrak{P}$ is separable, $\dot{\rho}$ does not depend on κ, and because of Assumption 4.1 the second moment of Y exists, it follows from Theorem 3.7 that ZC holds. □

The stronger BPC hinges on existence and uniqueness of the generalized Fréchet population mean. To this end we first decompose and compute the generalized Fréchet population function.

Lemma 4.4. *For the centered homoscedastic drift model from Definition 4.1 we have*

$$\begin{aligned}
&(i)\qquad \mathcal{F}([\kappa])=\int\rho(\phi\kappa^{(0)},[\kappa])\,\mathrm{d}\mathbb{P}\,(\phi)+\int\rho(\epsilon,[\kappa])\,\mathrm{d}\mathbb{P}\,(\epsilon)\,,\\
&(ii)\qquad \int\rho(\epsilon,[\kappa])\,\mathrm{d}\mathbb{P}\,(\epsilon)=2N-2\,,\\
&(iii)\qquad \int\rho(\phi\kappa^{(0)},[\kappa])\,\mathrm{d}\mathbb{P}\,(\phi)=\left(\tilde{\eta}^2-\frac{\tilde{\eta}^4}{4}\right)\int\mathrm{vec}\,(\phi)^T\,S\mathrm{vec}\,(\phi)\,\mathrm{d}\mathbb{P}\,(\phi)\,,
\end{aligned}$$

where $\tilde{\eta}=d\left([\kappa],\left[\kappa^{(0)}\right]\right)$ *and* S *is a matrix with eigenvalues greater than or equal to* λ_2.

Proof. To see (i), note that by Definition (22),

$$\hat{\phi}(\kappa,P,Y)=\hat{\phi}(\kappa,P,\phi\kappa^{(0)})+\hat{\phi}(\kappa,P,\epsilon)\,,$$

whence in conjunction with (21),

$$\begin{aligned}
\mathcal{F}([\kappa])&=\int\Big|\Big|\phi\kappa^{(0)}+\epsilon-\hat{\phi}(\kappa,P,\phi\kappa^{(0)}+\epsilon)\kappa\Big|\Big|_P^2\,\mathrm{d}\mathbb{P}\,(\phi,\epsilon)\\
&=\int\rho(\phi\kappa^{(0)},[\kappa])+2\left\langle\phi\kappa^{(0)}-\hat{\phi}(\kappa,P,\phi\kappa^{(0)})\kappa,\epsilon-\hat{\phi}(\kappa,P,\epsilon)\kappa\right\rangle_P+\rho(\epsilon,[\kappa])\,\mathrm{d}\mathbb{P}\,(\phi,\epsilon)
\end{aligned}$$

for any $\kappa\in[\kappa],\kappa^{(0)}\in[\kappa^{(0)}]$. Since ϕ and ϵ are independent and $\mathbb{E}[\epsilon]=0$ the integral over the mixed term vanishes yielding the first asserted equation.

To see the (ii), use Lemma D.1 to obtain

$$\int \rho(\epsilon,[\kappa])\,\mathrm{d}\mathbb{P}\,(\epsilon) = \int \left(\langle \epsilon,\epsilon\rangle_P - \left\langle \hat{\phi}(\kappa,P,\epsilon)\kappa,\epsilon\right\rangle_P\right)\mathrm{d}\mathbb{P}\,(\epsilon)$$

for any $\kappa \in [\kappa]$ with the MLE from (8). By definition of the Mahalanobis type inner product, independence of the $\epsilon_\nu \sim \mathcal{N}(0,\Sigma)$ $(\nu = 1,\ldots,N)$ and $P = \Sigma^{-1}$, calculate the first term:

$$\int \langle \epsilon,\epsilon\rangle_P\,\mathrm{d}\mathbb{P}\,(\epsilon) = \sum_{\nu=1}^{N} \mathrm{Tr}\left(P\ \mathrm{vec}\,(\epsilon_\nu)\,\mathrm{vec}\,(\epsilon_\nu)^T\,\mathrm{d}\mathbb{P}\,(\epsilon)\right) = \sum_{\nu=1}^{N}\mathrm{Tr}(P\Sigma) = \sum_{\nu=1}^{N}\mathrm{Tr}(\mathrm{Id}_2) = 2N\,.$$

Similarly, compute the second term:

$$\begin{aligned}
&\int \left\langle \hat{\phi}(\kappa,P,\epsilon)\kappa,\epsilon\right\rangle_P \mathrm{d}\mathbb{P}\,(\epsilon) = \int \sum_{\nu=1}^{N} \mathrm{vec}\,(\epsilon_\nu)^T\,PM(\kappa_\nu)\mathrm{vec}\left(\hat{\phi}(\kappa,P,\epsilon)\right)\mathrm{d}\mathbb{P}\,(\epsilon)\\
&= \int (\kappa \bullet_P \epsilon)^T\,(\kappa\diamond_P\kappa)^{-1}\,(\kappa\bullet_P\epsilon)\,\mathrm{d}\mathbb{P}\,(\epsilon) = \mathrm{Tr}\left((\kappa\diamond_P\kappa)^{-1}\int(\kappa\bullet_P\epsilon)\,(\kappa\bullet_P\epsilon)^T\,\mathrm{d}\mathbb{P}\,(\epsilon)\right)\\
&= \mathrm{Tr}\left((\kappa\diamond_P\kappa)^{-1}\,(\kappa\diamond_P\kappa)\right) = 2\,,
\end{aligned}$$

since

$$\begin{aligned}
&\int (\kappa\bullet_P\epsilon)\,(\kappa\bullet_P\epsilon)^T\,\mathrm{d}\mathbb{P}\,(\epsilon) = \int \left(\sum_{\nu=1}^{N} M(\kappa_\nu)^T P\mathrm{vec}\,(\epsilon_\nu)\right)\left(\sum_{\nu=1}^{N}\mathrm{vec}\,(\epsilon_\nu)^T\,PM(\kappa_\nu)\right)\mathrm{d}\mathbb{P}\,(\epsilon)\\
&= \sum_{\nu=1}^{N} M(\kappa_\nu)^T P\left(\int \mathrm{vec}\,(\epsilon_\nu)\,\mathrm{vec}\,(\epsilon_\nu)^T\,\mathrm{d}\mathbb{P}\,(\epsilon)\right)PM(\kappa_\nu) = \kappa\diamond_P\kappa.
\end{aligned}$$

Subtracting the first term from the second gives the second asserted equation.

Proving (iii) is a more elaborate. We have

$$\begin{aligned}
&\int \rho(\phi\kappa^{(0)},[\kappa])\,\mathrm{d}\mathbb{P}\\
&= \int \mathrm{vec}\,(\phi)^T\left(\left(\kappa^{(0)}\diamond_P\kappa^{(0)}\right) - \left(\kappa^{(0)}\diamond_P\kappa\right)(\kappa\diamond_P\kappa)^{-1}\left(\kappa\diamond_P\kappa^{(0)}\right)\right)\mathrm{vec}\,(\phi)\,\mathrm{d}\mathbb{P}\,(\phi)
\end{aligned}$$

where $\kappa \in [\kappa]$. Without loss of generality, assume that κ and $\kappa^{(0)}$ are in optimal position, i.e. $d\left([\kappa],\left[\kappa^{(0)}\right]\right)) = ||\kappa-\kappa^{(0)}|| =: \widetilde{\eta}$. Therefore, due to Lemma 1.2,

$$\kappa^*\kappa^{(0)} = \mathrm{Re}(\kappa^*\kappa^{(0)}) = \widetilde{\mathrm{vec}}\,(\kappa)^T\,\widetilde{\mathrm{vec}}\left(\kappa^{(0)}\right) = \cos(\eta) \quad \text{where} \quad \eta = 2\arcsin(\widetilde{\eta}/2).$$

We can rewrite the above formula by using matrix notation

$$\kappa\diamond_P\kappa = \begin{pmatrix} M(\kappa_1)^T & \ldots & M(\kappa_N)^T\end{pmatrix}\begin{pmatrix} P & 0\ldots & 0\\ 0 & \ddots & 0\\ 0 & \ldots 0 & P\end{pmatrix}\begin{pmatrix} M(\kappa_1)\\ \vdots\\ M(\kappa_N)\end{pmatrix}.$$

Now one can define a matrix containing only rotations on the diagonal, such that all $M(\kappa_\nu)$ become diagonal, i.e. real:

$$R_1 := \begin{pmatrix} R_{\alpha_1} & 0\dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots 0 & R_{\alpha_N} \end{pmatrix} \in \mathbb{R}^{2N\times 2N}$$

and then define a matrix $R_2 = (\widetilde{R}_2 \otimes \mathrm{Id}_2) \in \mathbb{R}^{2N\times 2N}$ which rotates these real blocks such that we get $R_2 R_1 \widetilde{\mathrm{vec}}\,(\kappa) = (1, 0, \dots, 0)^T$. From the construction of R_1 and R_2 follows directly

$$\tilde{\mathfrak{c}}\left(R_2 R_1 \widetilde{\mathrm{vec}}\,(\kappa)\right)^* \tilde{\mathfrak{c}}\left(R_2 R_1 \widetilde{\mathrm{vec}}\left(\kappa^{(0)}\right)\right) = \tilde{\mathfrak{c}}\left(R_2 \widetilde{\mathrm{vec}}\begin{pmatrix} e^{i\alpha_1}\kappa_1 \\ \vdots \\ e^{i\alpha_N}\kappa_N \end{pmatrix}\right)^* \tilde{\mathfrak{c}}\left(R_2 \widetilde{\mathrm{vec}}\begin{pmatrix} e^{i\alpha_1}\kappa_1^{(0)} \\ \vdots \\ e^{i\alpha_N}\kappa_N^{(0)} \end{pmatrix}\right)$$

$$= \tilde{\mathfrak{c}}\left(\widetilde{\mathrm{vec}}\begin{pmatrix} e^{i\alpha_1}\kappa_1 \\ \vdots \\ e^{i\alpha_N}\kappa_N \end{pmatrix}\right)^* \widetilde{R}_2^* \widetilde{R}_2 \tilde{\mathfrak{c}}\left(\widetilde{\mathrm{vec}}\begin{pmatrix} e^{i\alpha_1}\kappa_1^{(0)} \\ \vdots \\ e^{i\alpha_N}\kappa_N^{(0)} \end{pmatrix}\right) = \kappa^*\kappa^{(0)} = \cos(\eta).$$

Thus, it follows that $\tilde{\mathfrak{c}}\left(R_2 R_1 \widetilde{\mathrm{vec}}\left(\kappa^{(0)}\right)\right)_1 = \cos(\eta)$. Next, we define a Matrix R_3 which rotates all $M(\kappa_\nu^{(0)})$ for $\nu \geq 2$ to real numbers and leaves the component $\nu = 1$ unchanged (thus leaving κ unchanged) and a Matrix $R_4 = (\widetilde{R}_4 \otimes \mathrm{Id}_2)$ which rotates only the components $\nu \geq 2$, such that we get $\kappa_\nu^{(0)} = 0$ for $\nu \geq 2$. As a trade-off for this simplification, the matrix in the center becomes more complicated:

$$(\mathrm{Id}_N \otimes P) \quad \to \quad Q := R_4 R_3 R_2 R_1 (\mathrm{Id}_N \otimes P) R_1^T R_2^T R_3^T R_4^T.$$

This leads to

$$\begin{aligned}
&\int \rho(\phi\kappa^{(0)}, [\kappa])\,\mathrm{d}\mathbb{P} \\
&= \int \mathrm{vec}\,(\phi)^T \Bigg(\cos^2(\eta) Q_{11} + \cos(\eta)\sin(\eta)\,(Q_{12} + Q_{21}) + \sin^2(\eta) Q_{22} \\
&- \Big(\cos(\eta) Q_{11} + \sin(\eta) Q_{21}\Big) Q_{11}^{-1} \Big(\cos(\eta) Q_{11} + \sin(\eta) Q_{12}\Big)\Bigg) \mathrm{vec}\,(\phi)\,\mathrm{d}\mathbb{P}\,(\phi) \\
&= \sin^2(\eta) \int \mathrm{vec}\,(\phi)^T \Big(Q_{22} - Q_{12}^T Q_{11}^{-1} Q_{12}\Big) \mathrm{vec}\,(\phi)\,\mathrm{d}\mathbb{P}\,(\phi) \\
&= \left(\tilde{\eta}^2 - \frac{\tilde{\eta}^4}{4}\right) \int \mathrm{vec}\,(\phi)^T \Big(Q_{22} - Q_{12}^T Q_{11}^{-1} Q_{12}\Big) \mathrm{vec}\,(\phi)\,\mathrm{d}\mathbb{P}\,(\phi)
\end{aligned}$$

where Q_{ij} are 2×2 blocks from Q. Note furthermore that $Q_{ji}^T = Q_{ij}$ since Q is symmetric.

We define $S \coloneqq Q_{22} - Q_{12}^T Q_{11}^{-1} Q_{12}$. The matrix

$$\widetilde{Q} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix} \in \mathbb{R}^{4\times 4}$$

is positive definite with eigenvalues in $[\lambda_2, \lambda_1]$ since it is a leading principal minor of Q. It follows for $v \in \mathbb{R}^2 \setminus \{0\}$

$$\begin{aligned} v^T S v &= v^T Q_{22} v - v^T Q_{12}^T Q_{11}^{-1} Q_{12} v \\ &= v^T Q_{12}^T Q_{11}^{-1} Q_{11} Q_{11}^{-1} Q_{12} v - v^T Q_{12}^T Q_{11}^{-1} Q_{12} v - v^T Q_{12}^T Q_{11}^{-1} Q_{12} v + v^T Q_{22} v \\ &= \begin{pmatrix} -Q_{11}^{-1} Q_{12} v \\ v \end{pmatrix}^T \widetilde{Q} \begin{pmatrix} -Q_{11}^{-1} Q_{12} v \\ v \end{pmatrix} \geq \lambda_2 \begin{pmatrix} -Q_{11}^{-1} Q_{12} v \\ v \end{pmatrix}^T \begin{pmatrix} -Q_{11}^{-1} Q_{12} v \\ v \end{pmatrix} \geq \lambda_2 ||v||^2. \end{aligned}$$

Thus, S has only eigenvalues greater than or equal to λ_2.

□

Theorem 4.5 (Uniqueness)**.** *For the centered homoscedastic drift model from Definition 4.1 we have for every $\epsilon > 0$ that*

$$\inf_{[\kappa]: d([\kappa],[\kappa^{(0)}])>\epsilon} \mathcal{F}([\kappa]) > \mathcal{F}\left(\left[\kappa^{(0)}\right]\right).$$

In particular, the Fréchet population mean is uniquely $\left[\kappa^{(0)}\right]$.

Proof. This follows at once from

$$F([\kappa]) = 2N - 2 + \left(\tilde{\eta}^2 - \frac{\tilde{\eta}^4}{4}\right) \int \operatorname{vec}(\phi)^T S \operatorname{vec}(\phi) \, \mathrm{d}\mathbb{P}(\phi)$$

with $\tilde{\eta} = d\left([\kappa], \left[\kappa^{(0)}\right]\right)$ and S having eigenvalues greater than or equal to λ_2, due to Lemma 4.4. □

Corollary 4.6. *Under Assumption 4.1 BPC holds for the centered homoscedastic drift model.*

Proof. Since $E^\rho = \left\{\left[\kappa^{(0)}\right]\right\}$ due to Theorem 4.5, the BPC follows directly from Theorem 3.9 as $\mathfrak{P}$ is compact. □

4.2 The CLT for the centered homoscedastic drift model

To prove a central limit theorem for the centered homoscedastic drift model from Definition 4.1, we apply Theorem 6 of [Huc11a]. For this we need the following additional assumption.

Assumption 4.7. *The random variable ϕ has a finite fourth moment*

$$\mathbb{E}[(\operatorname{vec}(\phi)^T \operatorname{vec}(\phi))^2] < \infty.$$

Definition 4.8. *For $\kappa^{(0)} \in [\kappa^{(0)}] \in \mathfrak{P}$ define a unitary matrix $R \in \mathbb{C}^{N\times N}$ satisfying $R\kappa^{(0)} = e_N$ where e_N is the N-th vector of the standard basis. For any $\kappa \in [\kappa]$ define $\tilde{\kappa} \coloneqq R\kappa$ and $\mathfrak{U} = \{[\kappa] \mid \tilde{\kappa}_N \neq 0\}$ and define the chart*

$$\beta : \mathfrak{U} \to \mathbb{R}^{2(N-1)}, \quad [\kappa] \mapsto \left(\mathrm{Re}\left(\frac{\tilde{\kappa}_1}{\tilde{\kappa}_N}\right), \mathrm{Im}\left(\frac{\tilde{\kappa}_1}{\tilde{\kappa}_N}\right), \ldots, \mathrm{Re}\left(\frac{\tilde{\kappa}_{N-1}}{\tilde{\kappa}_N}\right), \mathrm{Im}\left(\frac{\tilde{\kappa}_{N-1}}{\tilde{\kappa}_N}\right)\right).$$

For $x \in \mathbb{R}^{2(N-1)}$ we define

$$\tilde{x} = \left(x_1 + ix_2, \ldots, x_{2(N-1)-1} + ix_{2(N-1)}, 1\right)$$

and get

$$\beta^{-1} : \mathbb{R}^{2(N-1)} \to \mathfrak{U}, \quad x \mapsto \left[R^* \frac{\tilde{x}}{||\tilde{x}||}\right].$$

Note that $\beta([\kappa])$ is indeed independent of the choice of representative $\kappa \in [\kappa] \in \mathfrak{U}$ and thus is well-defined. In a local chart $(\beta, \mathfrak{U})$ of $\mathfrak{P}$ near $\beta^{-1}(0)$, we denote the gradient of $x \mapsto \rho(Y, \beta^{-1}(x))$ by $\mathrm{grad}_2\rho(Y, [\kappa])$ and by $H_2\rho(Y, [\kappa])$ the corresponding Hesse matrix.

Theorem 4.9 (CLT)**.** *For the centered homoscedastic drift model from Definition 4.1, under Assumption 4.7, let $[\hat{\kappa}^{(B)}(\omega)] \in E_B^{(\rho)}(\omega)$ be a measurable selection for all $\omega \in \Omega$, then, omitting ω,*

$$\begin{aligned}&\sqrt{B}\beta([\hat{\kappa}^{(B)}])\\ &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\mathbb{E}\left[H_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1}\left(\mathrm{cov}\left[\mathrm{grad}_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)\left(\mathbb{E}\left[H_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1}\right)\end{aligned}$$

holds for the chart β defined in Definition 4.8.

Proof. We show in this proof that the following conditions for Theorem 6 of [Huc11a] are satisfied.

1. $x \mapsto \rho(Y, \beta^{-1}(x))$ is smooth for $|x| < \epsilon$,
2. $\mathbb{E}[\mathrm{grad}_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)]$ exists,
3. $\mathbb{E}[H_2\rho(Y, [\kappa])]$ exists for κ near $\kappa^{(0)}$ and is continuous at $\kappa = \kappa^{(0)}$,
4. $\mathrm{cov}[\mathrm{grad}_2\rho\left(Y, \left[\kappa^{(0)}\right]\right)]$ exists,
5. $\mathbb{E}[H_2\rho\left(Y, \left[\kappa^{(0)}\right]\right))]$ is invertible.

First, we rewrite ρ, see (21), to

$$\rho(Y, [\kappa]) = \langle Y, Y\rangle_P - \sum_{\nu=1}^{N}\sum_{\tilde{\nu}=1}^{N} \mathrm{vec}\left(Y_{\tilde{\nu}}\right)^T f_{\tilde{\nu},\nu,P}([\kappa])\mathrm{vec}\left(Y_\nu\right)$$

where

$$f_{\tilde{\nu},\nu,P}([\kappa]) \coloneqq PM(\kappa_{\tilde{\nu}}) \left(\kappa \diamond_P \kappa\right)^{-1} M(\kappa_\nu)^T P$$

for $\kappa \in [\kappa]$. Using Lemma D.3 we get

$$f_{\tilde{\nu},\nu,P}([\kappa]) = \frac{PM(\kappa_{\tilde{\nu}}) \left(\kappa \diamond_{\tilde{P}} \kappa\right) M(\kappa_\nu)^T P}{\det\left(\kappa \diamond_P \kappa\right)}.$$

1.) Both $\kappa \mapsto PM(\kappa_{\tilde{\nu}}) \left(\kappa \diamond_{\tilde{P}} \kappa\right) M(\kappa_\nu)^T P$ and $\kappa \mapsto \det\left(\kappa \diamond_P \kappa\right)$ are fourth degree polynomials and it follows from Lemma D.4 that $\det\left(\kappa \diamond_P \kappa\right) \geq \lambda_2 \lambda_1$ for $\kappa \in \mathfrak{P}$. Using the chain rule, it follows that the function $x \mapsto f_{\tilde{\nu},\nu,P}(\beta^{-1}(x))$ is smooth. Therefore, it follows that the function $x \mapsto \rho(Y, \beta^{-1}(x))$ is also smooth.

2.) We start with

$$\frac{\partial}{\partial x_i} \rho(Y, \beta^{-1}(x)) = -\sum_{\nu=1}^{N} \sum_{\tilde{\nu}=1}^{N} \operatorname{vec}\left(Y_{\tilde{\nu}}\right)^T \left(\frac{\partial}{\partial x_i} f_{\tilde{\nu},\nu,P}(\beta^{-1}(x))\right) \operatorname{vec}\left(Y_\nu\right)$$

Since ρ is smooth we get

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial}{\partial x_i} \rho(Y, \beta^{-1}(x))\right] &= -\mathbb{E}\left[\sum_{\nu=1}^{N} \sum_{\tilde{\nu}=1}^{N} \operatorname{vec}\left(Y_{\tilde{\nu}}\right)^T \left(\frac{\partial}{\partial x_i} f_{\tilde{\nu},\nu,P}(\beta^{-1}(x))\right) \operatorname{vec}\left(Y_\nu\right)\right] \\
&= -\sum_{\nu=1}^{N} \sum_{\tilde{\nu}=1}^{N} \operatorname{Tr}\left(\left(\frac{\partial}{\partial x_i} f_{\tilde{\nu},\nu,P}(\beta^{-1}(x))\right) \mathbb{E}\left[\operatorname{vec}\left(Y_{\tilde{\nu}}\right) \operatorname{vec}\left(Y_\nu\right)^T\right]\right).
\end{aligned}$$

Since ρ is smooth with respect to x and we know from Assumption 4.1 that the second moment of Y exists, it follows that $\mathbb{E}[\operatorname{grad}_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)]$ exist.

3.) Analogously, we conclude that the following function exists for x near $\beta(\kappa^{(0)})$ and is continuous at $x = \beta(\kappa^{(0)})$

$$\mathbb{E}\left[\frac{\partial^2}{\partial x_i \partial x_j} \rho(Y, \beta^{-1}(x))\right] = -\sum_{\nu=1}^{N} \sum_{\tilde{\nu}=1}^{N} \operatorname{Tr}\left(\left(\frac{\partial^2}{\partial x_i \partial x_j} f_{\tilde{\nu},\nu,P}(\beta^{-1}(x))\right) \mathbb{E}\left[\operatorname{vec}\left(Y_{\tilde{\nu}}\right) \operatorname{vec}\left(Y_\nu\right)^T\right]\right).$$

Thus $\mathbb{E}[H_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)]$ exists for κ near $\kappa^{(0)}$ and is continuous at $\kappa = \kappa^{(0)}$.

4.) From the smoothness of ρ with respect to x and the Assumption 4.7, the existence of

$$\mathbb{E}\left[\operatorname{grad}_2 \rho(Y, \left[\kappa^{(0)}\right]) \operatorname{grad}_2 \rho(Y, \left[\kappa^{(0)}\right])^T\right]$$

follows, since

$$\left(\frac{\partial}{\partial x_i} \rho(Y, \beta^{-1}(x))\right) \left(\frac{\partial}{\partial x_j} \rho(Y, \beta^{-1}(x))\right)$$

is a polynomial of fourth degree for all $i, j = 1, \ldots, N$ with respect to Y. Consequently, $\operatorname{cov}[\operatorname{grad}_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)]$ exists.

5.) Inserting the calculations from Lemma D.18 into the results of Lemma 4.4 yields:

$$\begin{aligned}\mathcal{F}(\beta(x)) &= 2N-2+\left(d\left(\beta(x),\left[\kappa^{(0)}\right]\right)^2-\frac{d\left(\beta(x),\left[\kappa^{(0)}\right]\right))^4}{4}\right)\int \mathrm{vec}\left(\phi\right)^T S\mathrm{vec}\left(\phi\right)\mathrm{d}\mathbb{P}\left(\phi\right)\\ &\geq 2N-2+\left(1-\frac{1}{\sqrt{||x||^2+1}}-\frac{\left(1-\frac{1}{\sqrt{||x||^2+1}}\right)^2}{4}\right)\lambda_2 c_\phi\\ &= 2N-2+\frac{||x||^2}{2}\lambda_2 c_\phi+\mathcal{O}(||x||^3).\end{aligned}$$

It follows from the Taylor expansion of $\mathcal{F}(\beta(x))$ at 0:

$$x^T\mathbb{E}[H_2\rho\left(Y,\left[\kappa^{(0)}\right]\right)]x \geq ||x||^2\lambda_2 c_\phi+\mathcal{O}(||x||^3).$$

Consequently $\mathbb{E}[H_2\rho\left(Y,\left[\kappa^{(0)}\right]\right)]$ is invertible. □

4.3 The CLT for the spectrum I

In the ENDOR experiment, one is particularly interested in the spectrum I (see Figure C1, panel A). Different rotation methods are possible to extract the estimated spectrum $\hat{I}$ from the maximum likelihood estimator $\hat{\kappa}$. As discussed in Section 2.1, in this section (as well as throughout the paper) we work with the maximum method and use the notation $\hat{I} = \mathrm{Re}(e^{i\hat{\lambda}}\hat{\kappa})$. Lemma D.19 provides an explicit formula for computing $\hat{\lambda}$ according to the maximum method, from which we derive the function in equation (24) below, which maps κ to an optimally rotated I. This explicit function is used in Corollary 4.11, which provides a central limit theorem for $\hat{I}$.

In fact, in conjunction with the chart β from Definition 4.8 we will construct functions $g, f, f_\pm$ making the diagram below commutative (on the corresponding domains) and smooth outside singularity sets $M_1 \cup M_2$ (defined in (26)) in $\mathfrak{P}$:

$$\begin{array}{rcl}\mathbb{C}P^{N-1}=\mathfrak{P} & \xrightarrow{f} & \mathfrak{I}=\mathbb{R}^N/\sim_\pm\\ \beta\downarrow & & \downarrow f_\pm\\ \mathbb{R}^{2N-2} & \xrightarrow{g} & \mathbb{R}^N\end{array}$$

For this purpose, we define $\mathbb{S}^1 := [0, 2\pi]/\sim$ where "$\sim$" denotes

$$x_1 \sim x_2 \quad \Leftrightarrow \quad x_1 = x_2 \quad \text{or} \quad x_1, x_2 \in \{0, 2\pi\}$$

From Lemma D.19 in the SI, with the definition for the complex argument from Section 1.1 follows

$$\underset{\lambda\in\mathbb{S}^1}{\arg\max}\left|\left|\mathrm{Re}(e^{i\lambda}\kappa)\right|\right|^2=\begin{cases}\left\{\pi-\frac{\mathrm{Arg}(\kappa^T\kappa)}{2},2\pi-\frac{\mathrm{Arg}(\kappa^T\kappa)}{2}\right\}, & \text{if } \kappa^T\kappa\neq 0\\ \mathbb{S}^1, & \text{else}\end{cases}$$

for $\kappa \in \mathcal{S}^{2N-1}$. Note that $\mathrm{Re}(e^{i(\lambda+\pi)}) = -\mathrm{Re}(e^{i\lambda})$, which leads to the metric space $(\mathfrak{I}, d_{\mathfrak{I}})$ given by $\mathfrak{I} \coloneqq \mathbb{R}^N / \sim_{\pm}$, where the equivalence relation "$\sim_{\pm}$" is defined as

$$I \sim_{\pm} I' \quad \Leftrightarrow \quad I = I' \quad \text{or} \quad I = -I'.$$

and

$$d_{\mathfrak{I}}([I]_{\pm}, [I']_{\pm}) = \min_{k\in\{0,1\}} \left|\left| I - (-1)^k I' \right|\right|^2$$

where $I \in [I]_{\pm}, I' \in [I']_{\pm}$. This gives rise to the function

$$\tilde{f} : \mathcal{S}^{2N-1} \to \mathfrak{I}, \quad \kappa \mapsto \begin{cases} \left[\mathrm{Re}\left(e^{\frac{-i}{2}\mathrm{Arg}(\kappa^T\kappa)}\kappa\right)\right]_{\pm} & \text{for } \kappa^T\kappa \neq 0, \\ [(0,\dots,0)]_{\pm} & \text{else.} \end{cases} \tag{24}$$

In Lemma D.20 in the SI, we show that the function $\tilde{f}$ is well-defined for $\mathfrak{P}$, which means that $\tilde{f}(\kappa) = \tilde{f}(\tilde{\kappa})$ for all $\kappa, \tilde{\kappa} \in [\kappa]$. Therefore, we can define a function

$$f : \mathfrak{P} \to \mathfrak{I}, \quad [\kappa] \mapsto \tilde{f}(\kappa).$$

To obtain the spectrum $I \in \mathbb{R}^N$ from $f([\kappa]) = [I]_{\pm} \in \mathfrak{I}$, an additional sign flip is performed, if necessary, as one usually wants the peaks to be in the positive direction. This can be uniquely achieved under the condition

$$\left|\max_{\nu=1,\dots N} I_\nu\right| > \left|\min_{\nu=1,\dots N} I_\nu\right|,$$

using the following sign flip function

$$f_{\pm} : \mathfrak{I} \to \mathbb{R}^N, \quad [I]_{\pm} \mapsto \begin{cases} I, & \text{if } |\max_{\nu=1,\dots N} I_\nu| > |\min_{\nu=1,\dots N} I_\nu| \\ -I, & \text{if } |\max_{\nu=1,\dots N} I_\nu| < |\min_{\nu=1,\dots N} I_\nu| \\ 0, & \text{else.} \end{cases} \tag{25}$$

Using the map β from Definition 4.8 we finally define

$$g : \mathbb{R}^{2N-2} \to \mathbb{R}^N, \quad x \mapsto f_{\pm}(f(\beta^{-1}(x))).$$

Ensuring that the functions $\widetilde{f}$ and $f_{\pm}$ are smooth we excluded the following two singularity sets

$$\begin{aligned} M_1 &\coloneqq \{[\kappa] \in \mathfrak{P} \mid \exists \kappa \in [\kappa] \text{ s.t. } \kappa^T\kappa = 0\}, \\ M_2 &\coloneqq \left\{[\kappa] \in \mathfrak{P} \mid \exists \kappa \in [\kappa] \text{ s.t. } \left|\max_{\nu=1,\dots N} f(\kappa)_\nu\right| = \left|\min_{\nu=1,\dots N} f(\kappa)_\nu\right|\right\}. \end{aligned} \tag{26}$$

Note that the defining relations in the Equation (26) are independent of representative, i.e. "$\exists$" can be replaced with "$\forall$" without loss of generality. With R from Definition 4.8, the

Jacobian matrix of the function g for $[\kappa^{(0)}] \in \mathfrak{P} \setminus (M_1 \cup M_2)$ at location $x = 0$ is computed in Lemma D.21 in the SI and is of the following form

$$J_x g(0) = \pm \operatorname{Re}\left(e^{-\frac{i\alpha}{2}} \left(i\kappa^0 \operatorname{Re}\left(i \frac{\overline{(\kappa^{(0)})^T \kappa^{(0)}} \left(\kappa^{(0)}\right)^T R^* A}{r^2} \right) + R^* A \right) \right) \in \mathbb{R}^{N \times 2(N-1)}.$$

where $\left|(\kappa^{(0)})^T \kappa^{(0)}\right| =: r$, $\alpha \coloneqq \operatorname{Arg}\left((\kappa^{(0)})^T \kappa^{(0)}\right)$ and

$$A \coloneqq \begin{pmatrix} 1 & i & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & i & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & i \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \in \mathbb{C}^{N \times 2(N-1)}. \tag{27}$$

Remark 4.10. *In our applications, we observed a Jacobi-matrix $J_x g(0)$ of full rank (see plots of the singular values for the different orientations from a chemical sample of the D2-$Y_{122}^{\bullet}$ E. coli ribonucleotide reductase in Figure C9 in the SI). However, this may not be the case in general. For example, for $\kappa^{(0)} = e_N$ we have $R = \mathrm{Id}_N$ so that $J_x g(0) = \pm \operatorname{Re}(A)$, which has rank $N - 1$.*

Corollary 4.11. *If $[\kappa^{(0)}] \in \mathfrak{P} \setminus (M_1 \cup M_2)$, in the centered homoscedastic drift model from Definition 4.1, we have the true spectrum $I^{(0)} = g \circ \beta([\kappa^{(0)}])$ with β from Definition 4.8 and under Assumption 4.7 with the estimator $\hat{I}^{(B)} = g \circ \beta([\hat{\kappa}^{(B)}])$ from a measurable selection $[\hat{\kappa}^{(B)}] \in E_B^{(\rho)}$ that*

$$\sqrt{B}\left(\hat{I}^{(B)} - I^{(0)}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(J_x g(0)\right) \mathfrak{G}_\beta \left(J_x g(0)\right)^T\right)$$

where $\mathfrak{G}_\beta = \left(\mathbb{E}\left[H_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1} \left(\operatorname{cov}\left[\operatorname{grad}_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right) \left(\mathbb{E}\left[H_2 \rho\left(Y, \left[\kappa^{(0)}\right]\right)\right]\right)^{-1}$ and β is defined as in Definition 4.8.

Proof. Follows directly from Theorem 4.9 and Lemma D.21 using the delta method (see, for example, Section 3 of [vdV00]). □

5 Inconsistency for joint estimation of κ and Σ in the homoscedastic drift model

In contrast to Section 4, in this section we do not work with the assumption that Σ is known, but we investigate the more complicated case that Σ and κ are estimated simultaneously. To this end, we reformulate in Definition 5.1 the Centered Homoscedastic Drift Model from Definition 4.1 from Section 4 by introducing the true but unknown $\Sigma^{(0)}$. Particularly, this section demonstrates that the joint estimation of κ and Σ is not consistent. For ease of notation, as in Section 4, we assume that the original random variable is $N+1$ dimensional, so that Y below is N dimensional and we omit writing the tilde and the superscript ^{H} symbol from Section 2.1.

Definition 5.1 (Centered Extended Homoscedastic Drift Model)**.** *The complex N-dimensional random vector Y is given by*

$$Y = \phi\kappa^{(0)} + \epsilon \tag{28}$$

where $\kappa^{(0)} \in \mathcal{S}^{2N-1} \coloneqq \{\kappa \in \mathbb{C}^N : ||\kappa|| = 1\}$ comprises the true but unknown ENDOR spectrum, ϕ is a complex random value and $\epsilon = (\epsilon_1, \dots, \epsilon_N)$ is a complex N-dimensional random vector independent of ϕ with

$$0 < \mathbb{E}[\mathrm{vec}\,(\phi)^T \,\mathrm{vec}\,(\phi)] = c_\phi < \infty, \qquad \mathrm{vec}\,(\epsilon_\nu) \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma^{(0)}), 1 \le \nu \le N.$$

Moreover, we assume that the precision matrix $P^{(0)} = \left(\Sigma^{(0)}\right)^{-1}$ has two positive eigenvalues $\lambda_1 > \lambda_2 > 0$.

Thus, $\mathfrak{Q} = \mathbb{C}^N$ is the data space. Since, in contrast to Section 4, we additionally want to estimate the strictly positive definite symmetric matrix Σ, we obtain the following parameter space

$$\mathfrak{P} \coloneqq \mathbb{C}P^{N-1} \times \mathrm{SPD}(2)\,,$$

and for the loss we choose

$$\rho(Y, ([\kappa], P)) = d_P\left(Y, \hat{\phi}(\kappa, P, Y)\kappa\right)^2 - N\log(\det(P)) \tag{29}$$

with

$$\hat{\phi}(\kappa, P, Y) = \mathfrak{c}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \bullet_P Y\right)\right) \in \mathbb{C}.$$

Thus, for a sample $Y(1), Y(2), \dots \overset{i.i.d.}{\sim} Y$ from (28) we have the following sample and

population Fréchet functions

$$\mathcal{F}_B(\omega, ([\kappa], P)) = \frac{1}{B}\sum_{b=1}^{B} \rho(Y(b), [\kappa]), \qquad \mathcal{F}([\kappa], P) = \int \rho(Y, [\kappa])\, \mathrm{d}\mathbb{P}\,(\phi, \epsilon)\,.$$

Analogous to Lemma 4.4, we decompose the first term of $\mathcal{F}([\kappa], P)$ into the ϵ and ϕ parts

$$\begin{aligned}
&\int d_P\left(Y, \hat{\phi}(\kappa, P, Y)\kappa\right)^2 \mathrm{d}\mathbb{P}\,(\phi, \epsilon) \\
&= \int \Bigg(d_P\left(\phi\kappa^{(0)}, \hat{\phi}(\kappa, P, \phi\kappa^{(0)})\kappa\right)^2 + 2\left\langle \phi\kappa^{(0)} - \hat{\phi}(\kappa, P, \phi\kappa^{(0)})\kappa, \epsilon - \hat{\phi}(\kappa, P, \epsilon)\kappa \right\rangle_P \\
&\qquad + d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \Bigg) \mathrm{d}\mathbb{P}\,(\phi, \epsilon)\,,
\end{aligned}$$

for any $\kappa \in [\kappa]$. Since ϕ and ϵ are independent and $\mathbb{E}[\epsilon] = 0$ the integral over the mixed term vanishes and consequently

$$\begin{aligned}
\mathcal{F}([\kappa], P) = &\int d_P\left(\phi\kappa^{(0)}, \hat{\phi}(\kappa, P, \phi\kappa^{(0)})\kappa\right)^2 \mathrm{d}\mathbb{P}\,(\phi) + \int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\,(\epsilon) \\
&- N\log(\det(P)).
\end{aligned}$$

In contrast to Section 4.1, the expression

$$\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\,(\epsilon) = N\,\mathrm{Tr}\left(\Sigma^{(0)}P\right) - \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\,(\kappa \diamond_{P\Sigma^{(0)}P} \kappa)\right)$$

depends on κ, see Lemma E.1 in the SI. In Lemma E.4 in the SI

$$\frac{\partial}{\partial P}\left(N\,\mathrm{Tr}\left(\Sigma^{(0)}P\right) - \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\,(\kappa \diamond_{P\Sigma^{(0)}P} \kappa)\right) - N\log(\det(P))\right)$$

is calculated and from Lemma E.5 in the SI follows

$$\begin{aligned}
\frac{\partial}{\partial P}\mathcal{F}([\kappa], P)\Big|_{[\kappa]=[\kappa^{(0)}], P=P^{(0)}} &= 2\left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)} \diamond_{P^{(0)}} \kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)}\right) \\
&\quad - \mathrm{diag}\left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)} \diamond_{P^{(0)}} \kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)}\right).
\end{aligned}$$

In general, $\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)} \diamond_{P^{(0)}} \kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)}$ is not equal to 0. Thus, in general, it does not hold that $E = \left\{\left([\kappa^{(0)}], P^{(0)}\right)\right\}$. This shows that the matrix Σ, which describes the random vector ϵ, cannot be estimated consistently from the profile likelihood. To achieve a jointly consistent estimator for κ and Σ, one would heuristically expect that a proper treatment of the randomness in both ϵ and ϕ is required, which the profile likelihood does not provide for ϕ. See Section 7 for a fuller discussion.

6 Heteroscedastic Drift Model

The homoscedastic drift model has been found to fit spectroscopic data recorded at MW frequency 263 GHz well, across a range of RF frequencies (20 MHz-400 MHz) and nuclei (^{1}H, ^{2}H, ^{19}F). [PEH$^+$21, HTW$^+$22, WKH$^+$23]. However, application to the lower MW frequency of 94 GHz, more commonly encountered in biochemistry groups, reveals very poor fit arising from the noise containing a *phase noise* component that is affected by phase drift. This necessitated development of a heteroscedastic drift model that we will detail in this section. It exhibits much improved fit and again results in improved SNR compared to the averaging model.

6.1 Modelling of the Heteroscedastic Drift Model

We test the homoscedastic drift model with ENDOR data recorded at a MW frequency of 94 GHz targeting the ^{2}H resonance in the twice deuterated Y_{122} Tyrosyl radical [HTW$^+$22] and using the Mims pulse sequence, see [Mim65]. This pulse sequence is known for yielding strong EPR echos, so we expect $|\psi_b|$ to be large. As in Section 2, the goodness of fit was assessed by applying Kolmogorov-Smirnov tests to the real and imaginary parts of the standardized residuals for each of the five datasets (orientations $g_x, g_{xy}, g_y, g_{yz}, g_z$) yielding the $p-$values reported in Table C3 in Section F.1 in the SI. For all orientations except g_z, at least one of the two p-values falls far below the Bonferroni-corrected significance level of 0.005, with some $p-$values of order 10^{-20} indicating very poor fit. This lack of fit can also be observed from the kernel density estimates and q-q-plots shown in Figure C3. Further examination of the residuals shown in Panel A of Figure C4 hints at an underly-

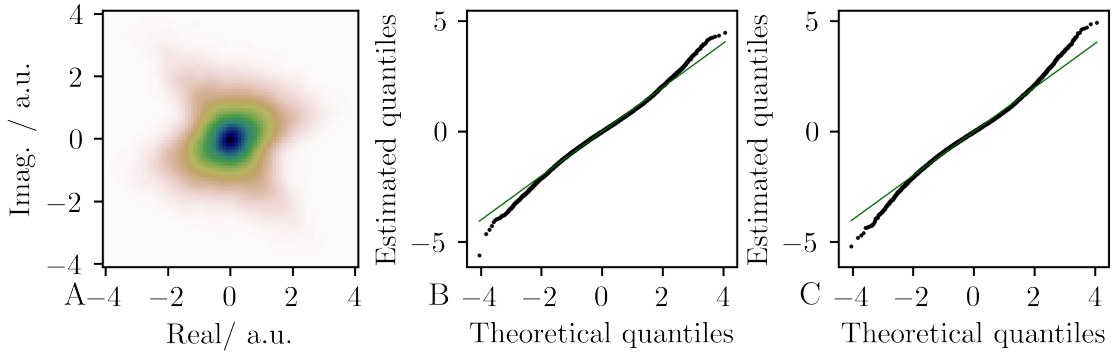


Figure C3: Results of the goodness of fit methods for the homoscedastic drift model applied to 94 GHz Mims data. Panel A displays the kernel density estimator applied to the standardized residuals $\hat{\Sigma}^{-\frac{1}{2}}\hat{\epsilon}_{b,\nu}$ with $\hat{\Sigma}^{-\frac{1}{2}}$ the inverse of the matrix square root of $\hat{\Sigma}_b$. Panels B and C show q-q-plots for the real part (B) and the imaginary part (C) of the standardized residuals against a standard normal (black) and a reference of perfect fit (green).

ing heteroscedastic noise structure w.r.t. the batches. Taking a general batch-dependent covariance matrix for the noise $\epsilon_{b,\nu} \overset{ind}{\sim} \mathcal{N}(0, \Sigma_b), \Sigma_b \in \mathrm{SPD}(2)$, constitutes a very flexible

extension of the homoscedastic drift model which is non-parametric in the sense that the number of parameters $\{\Sigma_b | b \in \{1, \dots, B\}\}$ increases with the amount of data available. However, the likelihood of this model has boundary maxima which can be obtained by choosing the parameters ϕ, κ such as to yield zero residuals $\hat{\epsilon}_{b^*,\nu} = 0$ for one particular batch b^* and letting the covariance matrix Σ_{b^*} tend to zero resulting in $-\frac{1}{2}\log\det\Sigma_{b^*}$ tending to infinity. Such an approach therefore needs additional penalization for the Σ_b resulting in shrinkage and a parametric extension was pursued instead. Based on the empirical observation presented in Panels B and C of Figure C4 that the batch-wise principal component of the homoscedastic residuals $\hat{\epsilon}_{b,:}$ is rotated by 90° compared to the spectrum mean $\mathrm{vec}\left(\hat{\psi}_b\right)$, the noise was modelled as a sum of a homoscedastic noise source and one whose covariance is given as a function of $\mathrm{vec}\,(\psi)$. This batch-dependent noise is attributed to the phase noise of the EPR echo ψ_b as phase noise is known to be orthogonal in phase and proportional in amplitude to the carrier signal it arises from [Hag09]. Both these properties of phase noise are empirically found to apply to the residuals of the homoscedastic drift model arising from our data: see panels A and B of Figure C4 for orthogonality and panel C of that figure for amplitude.

The expansion of a phase noise term modeled as a wrapped Gaussian

$$\begin{aligned}\tilde{\psi}_{b,\nu} &= \psi_b \exp\{i\tilde{\sigma}\varphi_{b,\nu}\} \\ \varphi_{b,\nu} &\overset{i.i.d}{\sim} \mathcal{N}(0,1)\end{aligned}$$

in small $\tilde{\sigma} > 0$ up to linear order gives

$$\mathrm{vec}\left(\tilde{\psi}_{b,\nu}\right) = \mathrm{vec}\,(\psi_b) + \tilde{\sigma}\varphi_{b,\nu}\mathrm{vec}\,(i\psi_b) + \mathcal{O}_P(\tilde{\sigma}^2). \tag{30}$$

This random variable has mean $\mathrm{vec}\,(\psi_b) + \mathcal{O}_P(\tilde{\sigma}^2)$ and covariance matrix

$$\tilde{\sigma}^2\mathrm{vec}\,(i\psi_b)\,\mathrm{vec}\,(i\psi_b)^T + \mathcal{O}_P(\tilde{\sigma}^4).$$

So, this expansion reproduces the homoscedastic drift model mean ψ_b and the empirical orthogonality structure of the residuals to linear and quadratic order in $\tilde{\sigma}$, respectively. It adds a further dependency of the moments on ψ_b and therefore $\hat{\psi}_b = \frac{1}{N+1}\sum_{\nu=0}^{N} Y_{b,\nu}$ is not the MLE estimate anymore. In the homoscedastic case, we needed the condition in Equation 3 of standardized mean 0 spectra for identifiability. To retain this standardization for κ, we introduce an additional parameter, the spectrum mean $c \in \mathbb{C}$. In total, the heteroscedastic drift model hence decomposes the data matrix as follows:

Definition 6.1 (Heteroscedastic Drift Model)**.**

$$\begin{aligned}Y_{b,\nu} &= \psi_b + \phi_b(\kappa_\nu + c) + \epsilon_{b,\nu} \\ \epsilon_{b,\nu} &\overset{ind.}{\sim} \mathcal{N}(0, \Sigma_b) \\ \Sigma_b &= \Sigma_0 + \tilde{\sigma}^2\mathrm{vec}\,(i\psi_b)\,\mathrm{vec}\,(i\psi_b)^T.\end{aligned}$$

Even though this model also has boundary maxima in the limit when Σ_0 is rank deficient, see Lemma F.1 in the SI, this is easily overcome by specifying lower bounds on the eigenvalues of Σ_0 that arise from reasonable estimates of minimal MW receiver noise. No penalization is needed in practice to enforce these bounds when starting optimizers from parameter estimates derived from the homoscedastic drift model, see Section F.2 in the SI for details.

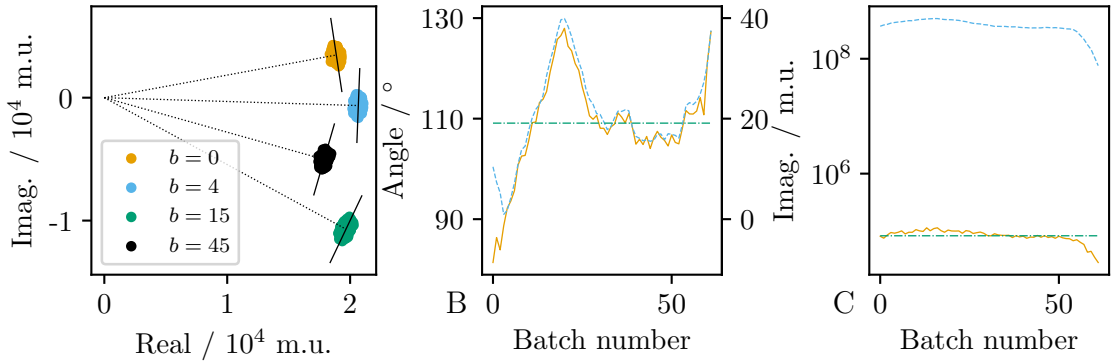


Figure C4: Examination of the 94 GHz data using the results of the homoscedastic drift model. Panel A shows the comparison of the principal component (direction of black line segment) of the data from one batch $Y_{b,:}$. The batch mean $\hat{\psi}_b$ is visualized as the dotted line segment $\overline{0\hat{\psi}_b}$. Panel B shows the angle with the real axis of the principal component of the residuals per batch (orange) and the eigenvector corresponding to the maximal eigenvalue of $\hat{\Sigma}$ (constant, green dot-dashed) both quantified on the left y-axis and the angle of $\hat{\psi}_b$ (light blue dashed) quantified on the right y axis which is offset by 90°. In Panel C, the magnitude of the largest singular value of the residuals per batch (orange), $\left|\hat{\psi}_b\right|^2$ (light blue dashed) and the largest eigenvalue of $\hat{\Sigma}$ (constant green dot-dashed) are depicted.

In the heteroscedastic drift model, we truncate the expansion of the phase noise term in $\tilde{\sigma}$ at the linear order for the mean and quadratic order for the variance, respectively. Careful comparison of higher order terms with empirically observed values of Σ_0, see Section F.3 in the SI, reveals that the former are at least two orders of magnitude smaller than the latter which justifies our chosen truncation.

6.2 Results of the Heteroscedastic Drift Model

The algorithmic implementation of heteroscedastic drift model estimates a local MLE by iteratively updating the parameters by their conditional MLE. In contrast to the homoscedastic drift model, the conditional MLE's have to be approximated for $\Sigma_0, \tilde{\sigma}$ and ψ. In order to improve convergence properties of the algorithm, we included an additional step wherein we calculate the conditional MLE for the parameter Δ_c where $\tilde{\psi} = \psi - \Delta_c\phi$, $\tilde{c} = c + \Delta_c$, see Appendix F of the SI for the algorithm and further details.

Finding the optimal rotation and flip of $\hat{\kappa}$ to obtain the final spectrum $\hat{I}$ is done as in the homoscedastic drift model. The algorithm will report if one eigenvalue of Σ_0 is lower

than the empirical cut-off $\delta = 1 \times 10^{-20}$. The results of applying this algorithm to the 94 GHz data can be found in Data Example 6.2.

Data Example 6.2 (Heteroscedastic Drift Model for 94 GHz Dataset)**.** *The result of applying Algorithm 2 to the* 94 GHz *dataset are shown in Figure C5. The goodness of fit methods are applied to the real and imaginary part of the standardized residuals* $\tilde{\epsilon}_{b,\nu} = \hat{\Sigma}_b^{-\frac{1}{2}}\left(Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu\right)$*. The results of the kernel-density estimation are in Panel F and the q-q plots in Panel G and H of Figure C5. The Kolmogorov-Smirnov test is carried out at the Bonferroni-corrected level of* $0.005 = 0.05/10$*. The p-values can be found in Table C5. The model is narrowly rejected by the K-S-test as the p value in orientation y is significant. Still, the graphical goodness of fit results are improved over the results from applying the homoscedastic drift model. The SNR estimated from the heteroscedastic drift model is visibly larger than the one from the averaging model in all orientations as can be seen in Figure C10 and Table C4 in the SI. Thus, while there is potential improvement to be gained by further modelling, the heteroscedastic drift model is already a successful extension of the homoscedastic drift model.*

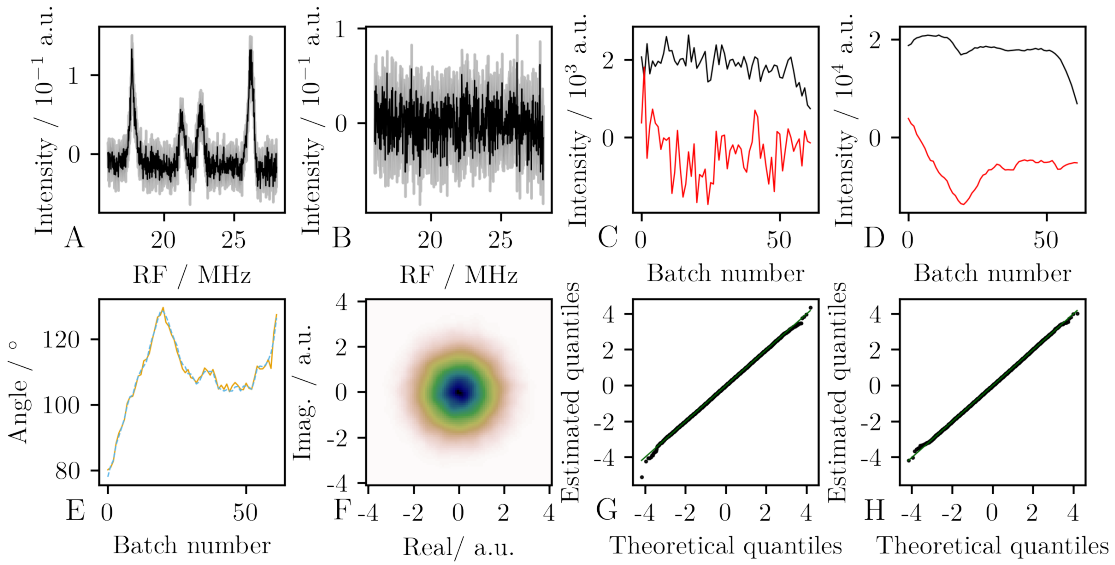


Figure C5: The results of applying the heteroscedastic drift model to the 94 GHz data. Panel A displays the estimated spectrum $\hat{I}$, while panel B displays the component $\hat{\omega}$ that is orthogonal to the estimated spectrum $\hat{I}$. Both panels show Bootstraped confidence intervalls based on 50 Bootstrap samples with out bias reduction. C and D show the real (black) and imaginary (red) components of $\exp\{i\alpha_{opt}\}\hat{\phi}$ and $\hat{\psi}$, respectively, where the α_{opt} was chosen to maximize the correlation between the rotated $\hat{\phi}$ and $\hat{\psi}$. Panel E shows the angle of the principal component of the residuals per batch (orange) and ψ_b (light blue dashed) with the real axis. Panel F displays the kernel-density-estimation of the complex residuals standardized residuals $\tilde{\epsilon}_{b,\nu}$, while panels G and H depict q-q-plots for the real and imaginary components of the residuals, respectively in black with the identity shown in green.

7 Outlook

For the homoscedastic drift model, asymptotic theory was developed for the case where Σ is known (see Section 4). Since the joint estimate of κ and Σ is not consistent in the profile likelihood model (see Section 5), it would be desirable to obtain a consistent estimate by including the randomness of ϕ in the statistical model. Possible approaches are to model the $\{\phi_b\}_{b=1}^B$ as i.i.d. Gaussian or, to reflect the likelihood being invariant under permutations of batches, as exchangeable random variables or, perhaps most realistically, as a Gaussian process. For the latter two approaches, one would need to generalize the theory about generalized strong consistency of generalized Fréchet means (see Section 3) for random variables that are not i.i.d.. Furthermore, an asymptotic analysis for the heteroscedastic model is future work. Here, particular challenges arise as the mean ψ_b and the variance Σ_b are dependent on each other. In addition, it is challenging to develop drift models for all microwave frequencies and pulse sequences to make them usable for a large audience. Initial work on Davies $\nu_{MW} = 94\,\mathrm{GHz}$ data (a special pulse sequence, see [Dav74]) shows the heteroscedastic drift model not to fit well in this case, likely due to cancellation of the main echo signal leading to small and noisy $\hat{\psi}_b$. In addition, there are other experiments at $\nu_{MW} = 34\,\mathrm{GHz}$ and $\nu_{MW} = 9\,\mathrm{GHz}$ for which drift models are not yet available. A possible avenue may be separate modelling of mean and variance via $\Sigma_b = \Sigma_0 + \alpha_b \alpha_b^T$ with $\alpha_b \in \mathbb{R}^2$ to be estimated which may subsume homoscedastic and heteroscedastic noise models and would also apply to pulse sequences where $\hat{\psi}$ is afflicted by noise and cancellation effects.

8 Acknowledgements

H.W., B.E., S.H., M.B. and Y.P. thank the DFG — project-ID 432680300 — CRC 1456 for financial support. M.B. acknowledges the ERC Advanced Grant 101020262 BIO-enMR. We thank the Max Planck Society for financial support. S.H. acknowledges the Niedersachsen Vorab of the Volkswagen foundation, DFG-HU 1575/7 and the IMSI workshop on Object Oriented Data Analysis in Health Sciences 2023. Y.P. gratefully acknowledges Royal Society International Exchanges grant IE150666.

References

[Afs11] B. Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139:655–773, 2011.

[And03a] Amy C. Anderson. The process of structure-based drug design. *Chemistry & Biology*, 10(9):787–797, 2003.

[And03b] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley Interscience, 2003.

[BP03] Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29, 2003.

[Bul82] Adhemar Bultheel. Inequalities in Hilbert modules of matrix-valued functions. *Proceedings of the American Mathematical Society*, 85(3):369–372, 1982.

[Cho54] Gustave Choquet. Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295, 1954.

[Dav74] E. R. Davies. A new pulse endor technique. *Physics Letters A*, 47(1):1–2, 1974.

[DM98] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis.* Wiley Series in Probability and Statistics. Wiley, 1998.

[EABG03] B. Epel, D Arieli, D Baute, and D. Goldfarb. Improving w-band pulsed endor sensitivity—random acquisition and pulsed special triple. 164:78–83, 2003.

[EHHS22] Benjamin Eltzner, Pernille Hansen, Stephan F. Huckemann, and Stefan Sommer. Diffusion means in geometric spaces. 2022.

[EJ20] Steven N. Evans and Adam Q. Jaffe. Strong laws of large numbers for Fréchet means. *arXiv preprint arXiv:2012.12859*, 2020.

[Feh56] G. Feher. Observation of nuclear magnetic resonances via the electron spin resonance. *Phys. Rev.*, 103, 1956.

[For07] Kevin Ford. *From Kolmogorov's theorem on empirical distribution to number theory.* Springer, 2007.

[Fré48] Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310, 1948.

[FVJ08] P.T. Fletcher, S. Venkatasubramanian, and S.C. Joshi. Robust statistics on Riemannian manifolds via the geometric median. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[GS91] Claudius. Gemperle and Arthur. Schweiger. Pulsed electron-nuclear double resonance methodology. *Chemical Reviews*, 91(7):1481–1505, 1991.

[Hag09] Jon B. Hagen. *Radio-Frequency Electronics: Circuits and Applications*. Cambridge University Press, 2nd edition, 2009.

[Har16] Jeffrey R. Harmer. Hyperfine spectroscopy – endor. *eMagRes*, 5, 2016.

[HL98] H. Hendriks and Z. Landsman. Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *Journal of Multivariate Analysis*, 67:227–243, 1998.

[HTW+22] Markus Hiller, Igor Tkach, Henrik Wiechers, Benjamin Eltzner, Stephan Huckemann, Yvo Pokern, and Marina Bennati. Distribution of H-β hyperfine couplings in a tyrosyl radical revealed by 263 Ghz endor spectroscopy. *Applied Magnetic Resonance*, 53:1015–1030, 2022.

[Huc11a] Stephan Huckemann. Inference on 3d Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics*, 38(3):424–446, 2011.

[Huc11b] Stephan F. Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics*, 39(2):1098 – 1124, 2011.

[Huc12] Stephan Huckemann. On the meaning of mean shape: Manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics*, 64(6):1227–1259, 2012.

[Kar14] Hermann Karcher. Riemannian center of mass and so called Karcher mean. *arXiv preprint arXiv:1407.2087*, 2014.

[Ken74] David G Kendall. Foundations of a theory of random sets, stochastic geometry (Harding E.F. and Kendall D.G., eds.), 1974.

[Mat74] Georges Matheron. *Random sets and integral geometry*. John Wiley & Sons, 1974.

[MDD+20] Andreas Meyer, Sebastian Dechert, Surjendu Dey, Claudia Höbartner, and Marina Bennati. Measurement of Angstrom to Nanometer Molecular Distances with ^{19}F Nuclear Spins by EPR/ENDOR Spectroscopy. *Angewandte Chemie International Edition*, 59(1):373–379, 2020.

[Mim65] W. B. Mims. Pulsed Endor Experiments. *Proceedings of the Royal Society of London Series A — Mathematical and Physical Sciences*, 283(1395):452–457, 1965.

[Mol05] Ilya Molchanov. *Theory of random sets.* Probability and Its Applications, Springer, 2005.

[PEH+21] Yvo Pokern, Benjamin Eltzner, Stephan F. Huckemann, Clemens Beeken, JoAnne Stubbe, Igor Tkach, Marina Bennati, and Markus Hiller. Statistical analysis of ENDOR spectra. *Proceedings of the National Academy of Sciences*, 118(27), 2021.

[RB14] Roberto Rizzato and Marina Bennati. Enhanced sensitivity of electron-nuclear double resonance (ENDOR) by cross polarisation and relaxation. *Phys. Chem. Chem. Phys.*, 16:7681–7685, 2014.

[Sch22] Christof Schötz. Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics*, 56(1):34–52, 2022.

[TI97] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra.* SIAM, 1997.

[vdV00] A.W. van der Vaart. *Asymptotic statistics.* Cambridge Univ. Press, 2000.

[WKH+23] Henrik Wiechers, Annemarie Kehl, Markus Hiller, Benjamin Eltzner, Stephan Huckemann, Andreas Meyer, Igor Tkach, Marina Bennati, and Yvo Pokern. Bayesian optimization to estimate hyperfine couplings from ^{19}F ENDOR spectra. *Journal of Magnetic Resonance*, 2023.

[Zie77] Herbert Ziezold. *On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces*, pages 591–602. Springer Netherlands, Dordrecht, 1977.

Supplementary Information

A Homoscedastic drift model

The real part and imaginary part of the raw data matrix Y for orientation g_y from a chemical sample of D2-$Y^{\bullet}_{122}$ are presented in Figure C6.

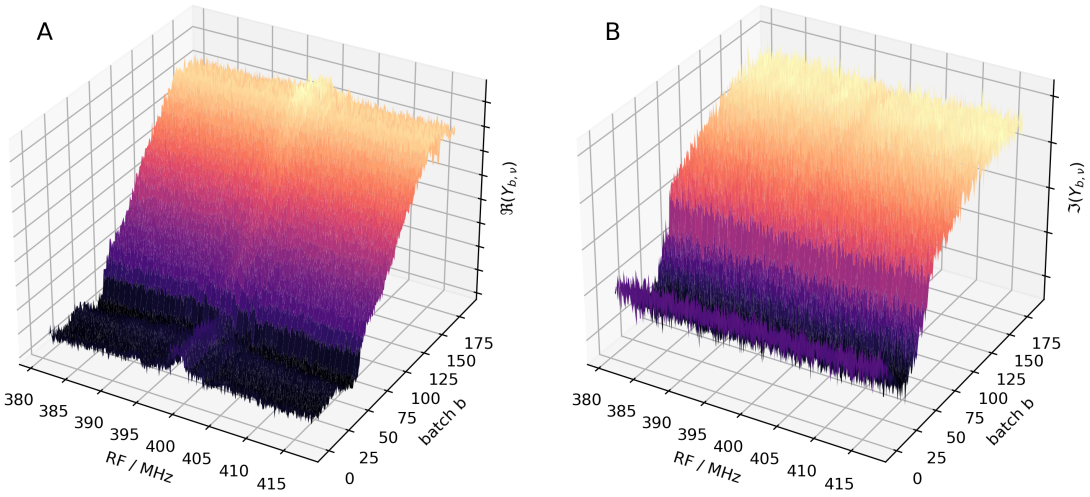


Figure C6: Panel A shows the real component of the raw data matrix Y for orientation g_y from a chemical sample of D2-$Y^{\bullet}_{122}$, while panel B displays the imaginary component of the same data matrix.

The algorithm used to fit the homoscedastic drift model is given in Algorithm 1.

In Figure C7 we compare the SNR of the averaging model with SNR of the homoscedastic drift model. For this purpose, the spectrum is extracted from the data matrices of the different orientations of the measurements of the chemical sample D2-$Y^{\bullet}_{122}$ using both the averaging model (plotted in green) and the homoscedastic drift model (plotted in black). In both models, the maximum method is used for phase correction. The regions of RF frequencies where the true ENDOR spectrum is judged to be constant, referred to as *flat frequency regions* defined in [PEH$^+$21], are plotted in the right panel and the standard deviations of the spectrum in the flat frequency regions are listed in Table C1. In four out of five orientations, the homoscedastic drift model provides an improved SNR. Only at orientation g_x is the SNR of the averaging model slightly better than that of the homoscedastic drift model, which can be explained by the fact that at orientation g_x the least phase drift of $\hat{\phi}$ is observed (see Figure C8).

Algorithm 1: Homoscedastic drift model MLE

Load $\boldsymbol{y}$
$\widehat{\boldsymbol{\psi}} \leftarrow \widehat{\boldsymbol{\psi}}_{hom}(\boldsymbol{y})$
$\tilde{\boldsymbol{y}} \leftarrow \boldsymbol{y} - \widehat{\boldsymbol{\psi}}$
$\boldsymbol{u}, \eta, \overline{\boldsymbol{v}} \leftarrow$ SVD($\tilde{\boldsymbol{y}}$,1st component)
$\widehat{\boldsymbol{\phi}}^{(0)} \leftarrow \boldsymbol{u}\eta$
$\widehat{\boldsymbol{\kappa}}^{(0)} \leftarrow \overline{\boldsymbol{v}}$
$k \leftarrow 0$
while $k \leq maxiter = 200$ **do**
  $\widehat{\boldsymbol{\Sigma}}^{(k)} \leftarrow \widehat{\boldsymbol{\Sigma}}_{hom}(\widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\boldsymbol{\kappa}}^{(k)}, \tilde{\boldsymbol{y}})$
  $\widehat{\boldsymbol{\phi}}^{(k+1)} \leftarrow \widehat{\boldsymbol{\phi}}_{hom}(\widehat{\boldsymbol{\kappa}}^{(k)}, (\widehat{\boldsymbol{\Sigma}}^{(k)})^{-1}, \tilde{\boldsymbol{y}})$
  $\widehat{\boldsymbol{\kappa}}^{(k+1)} \leftarrow \widehat{\boldsymbol{\kappa}}_{hom}(\widehat{\boldsymbol{\phi}}^{(k+1)}, (\widehat{\boldsymbol{\Sigma}}^{(k)})^{-1}, \tilde{\boldsymbol{y}})$
  $\widehat{\boldsymbol{\kappa}}^{(k+1)}, \widehat{\boldsymbol{\phi}}^{(k+1)} \leftarrow \frac{\widehat{\boldsymbol{\kappa}}^{(k+1)}}{\left\|\widehat{\boldsymbol{\kappa}}^{(k+1)}\right\|}, \left\|\widehat{\boldsymbol{\kappa}}^{(k+1)}\right\| \widehat{\boldsymbol{\phi}}^{(k+1)}$
  $\ell^{(k)} \leftarrow \ell(\tilde{\boldsymbol{y}}, \widehat{\boldsymbol{\phi}}^{(k+1)}, \widehat{\boldsymbol{\kappa}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}^{(k)})$
  if $k > 0$ **then**
    if $\ell^{(k)} - \ell^{(k-1)} < min\_delta\_loglik = 10^{-4}$ **then**
      break
    end if
  end if
  $k \leftarrow k + 1$
end while
return $\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\boldsymbol{\kappa}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k-1)}$

Orientation	Averaging model	Homoscedastic drift model
g_x	0.0074	0.0107
g_{xy}	0.0072	0.0061
g_y	0.0085	0.0033
g_{yz}	0.0045	0.0035
g_z	0.0111	0.0042

Table C1: The standard deviation of the spectrum across the flat frequency regions shown in Panel B of Figure C7, computed for both the averaging model and the homoscedastic drift model.

Orientation	Real	Imag
g_x	0.098	0.220
g_{xy}	0.023	0.237
g_y	0.736	0.938
g_{yz}	0.373	0.271
g_z	0.022	0.374

Table C2: Results of Kolmogorov–Smirnov tests for Gaussianity applied to the real and imaginary parts of the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b \hat{\kappa}_\nu$, pooled over b and ν, obtained from the homoscedastic drift model applied to all measurements.

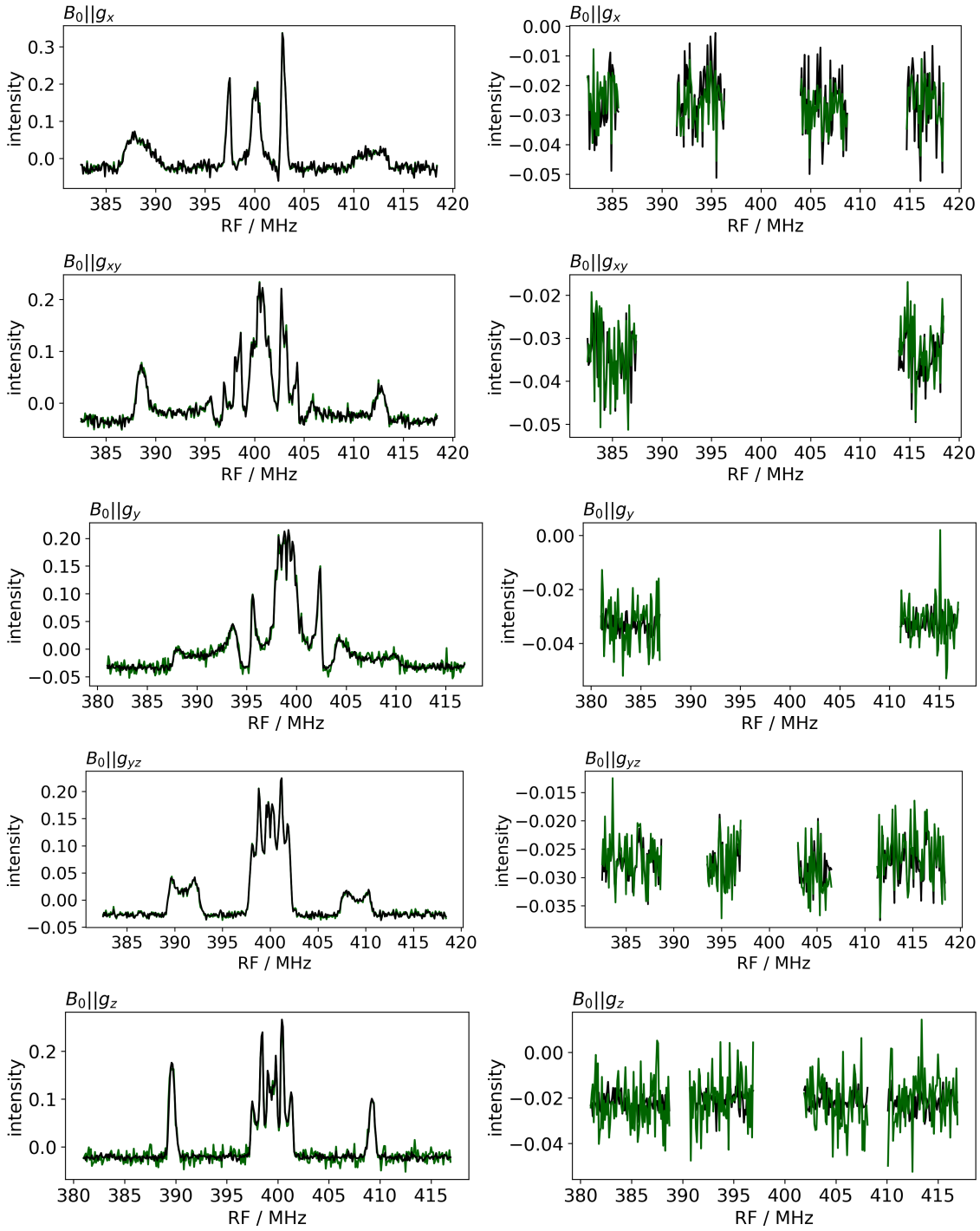


Figure C7: The estimated spectra $\hat{I}$ for all different orientations, using the homoscedastic drift model (black) and the averaging model (green) for all frequencies (left) and only for the flat frequency regions (right). The standard deviations of the spectra across the flat frequency regions from the right panel is given in Table C1.

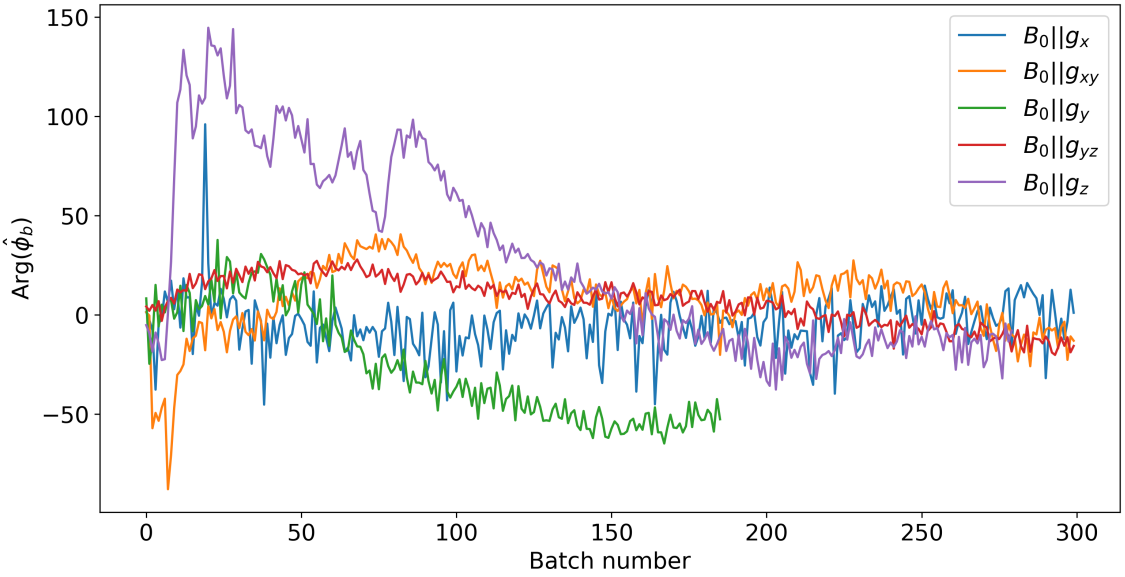


Figure C8: The angle of $\hat{\phi}$ for all five orientations of the chemical sample D2-$Y^{\bullet}_{122}$ at 263 GHz.

B Lemma for the Helmert Matrix

We prove in the following a lemma related to Section 2.1 from the main text and consequently have the same notation. In particular, the vectors $h_1, \ldots, h_N$, the $\tilde{\epsilon}^H_{b,\nu}$ and $\text{vec}\left(\tilde{\epsilon}^H_{b,\nu}\right)$ are defined as in Section 2.1 in the main text.

Lemma B.1. *In the new basis $h_1, \ldots, h_N$, the $\tilde{\epsilon}^H_{b,\nu}$ are i.i.d distributed and $\text{vec}\left(\tilde{\epsilon}^H_{b,\nu}\right) \sim \mathcal{N}(0, \Sigma)$ for all $b = 1 \ldots, B$ and $\nu = 1, \ldots, N$.*

Proof. It holds $\mathbb{E}\left[\text{vec}\left(\tilde{\epsilon}_{b,\nu}\right)\right] = 0$ for all $b = 1, \ldots, B$ and $\nu = 1, \ldots, N$. For all $b = 1, \ldots, B$ and $\nu = 1, \ldots, N$ we get

$$
\begin{aligned}
&\mathbb{E}\left[\text{vec}\left(\tilde{\epsilon}^H_{b,\nu}\right)\text{vec}\left(\tilde{\epsilon}^H_{b,\nu}\right)^T\right] = \mathbb{E}\left[\text{vec}\left(h_\nu^T\tilde{\epsilon}_b\right)\text{vec}\left(h_\nu^T\tilde{\epsilon}_b\right)^T\right] \\
&= \frac{1}{\nu(\nu+1)}\mathbb{E}\left[\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right) - \nu\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)\right)\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right) - \nu\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)\right)^T\right] \\
&= \frac{1}{\nu(\nu+1)}\mathbb{E}\Bigg[\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)^T - \nu\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)^T \\
&\qquad\qquad - \nu\left(\sum_{k=1}^{\nu}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)^T + \nu\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)\nu\text{vec}\left(\tilde{\epsilon}_{b,\nu+1}\right)^T\Bigg] \\
&= \frac{1}{\nu(\nu+1)}\left(\nu\left(-\frac{\nu}{N}\Sigma + \Sigma\right) + \frac{2\nu^2}{N}\Sigma + \nu^2\left(1 - \frac{1}{N}\right)\Sigma\right) = \frac{1}{\nu+1}\left(\nu + \nu^2\right)\Sigma = \Sigma.
\end{aligned}
$$

For $\nu_1 < \nu_2$ we get

$$
\begin{aligned}
&\mathbb{E}\left[\text{vec}\left(\tilde{\epsilon}^H_{b,\nu_1}\right)\text{vec}\left(\tilde{\epsilon}^H_{b,\nu_2}\right)^T\right] = \mathbb{E}\left[\text{vec}\left(h_{\nu_1}^T\tilde{\epsilon}_b\right)\text{vec}\left(h_{\nu_2}^T\tilde{\epsilon}_b\right)^T\right] \\
&= \frac{1}{\sqrt{\nu_1(\nu_1+1)\nu_2(\nu_2+1)}} \\
&\qquad \cdot \mathbb{E}\left[\left(\sum_{k=1}^{\nu_1}\text{vec}\left(\tilde{\epsilon}_{b,k}\right) - \nu_1\text{vec}\left(\tilde{\epsilon}_{b,\nu_1+1}\right)\right)\left(\sum_{k=1}^{\nu_2}\text{vec}\left(\tilde{\epsilon}_{b,k}\right) - \nu_2\text{vec}\left(\tilde{\epsilon}_{b,\nu_2+1}\right)\right)^T\right] \\
&= \frac{1}{\sqrt{\nu_1(\nu_1+1)\nu_2(\nu_2+1)}} \\
&\qquad \cdot \mathbb{E}\left[\left(\sum_{k=1}^{\nu_1}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)\left(\sum_{k=1}^{\nu_2}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)^T - \nu_1\text{vec}\left(\tilde{\epsilon}_{b,\nu_1+1}\right)\left(\sum_{k=1}^{\nu_2}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)^T\right] \\
&+ \frac{1}{\sqrt{\nu_1(\nu_1+1)\nu_2(\nu_2+1)}} \\
&\qquad \cdot \mathbb{E}\left[-\nu_2\text{vec}\left(\tilde{\epsilon}_{b,\nu_2+1}\right)\left(\sum_{k=1}^{\nu_1}\text{vec}\left(\tilde{\epsilon}_{b,k}\right)\right)^T + \nu_1\text{vec}\left(\tilde{\epsilon}_{b,\nu_1+1}\right)\nu_2\text{vec}\left(\tilde{\epsilon}_{b,\nu_2+1}\right)^T\right] \\
&= \frac{1}{\sqrt{\nu_1(\nu_1+1)\nu_2(\nu_2+1)}}\left(\nu_1\left(1 - \frac{\nu_2}{N}\right)\Sigma - \nu_1\left(1 - \frac{\nu_2}{N}\right)\Sigma + \frac{\nu_1\nu_2}{N}\Sigma - \frac{\nu_1\nu_2}{N}\Sigma\right) = 0.
\end{aligned}
$$

The independence of $\tilde{\epsilon}^H_{b_1,\nu_1}$ from $\tilde{\epsilon}^H_{b_2,\nu_2}$ for $b_1 \neq b_2$ and $\nu_1, \nu_2 = 1, \ldots, N$ follows directly from the independence of $\tilde{\epsilon}_{b_1,\nu_1}$ from $\tilde{\epsilon}_{b_2,\nu_2}$. □

C Example Strong Consistency

We show that using the theory developed in Section 3 in the main text we can prove strong consistency for the simultaneous estimation of μ and σ in the univariate normal distribution.

Assumption C.1. *The random variable X has distribution $X \sim \mathcal{N}\left(\mu^{(0)}, \left(\sigma^{(0)}\right)^2\right)$ where $\mu^{(0)}$ and $\left(\sigma^{(0)}\right)^2$ are the true but unknown parameters of the normal distribution.*

For observations $x_1, \dots, x_n$ we get the following log-likelihood function

$$\ell_x(\mu, \sigma) = -\frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}\left(-\ln(\sigma) - \frac{1}{2\sigma^2}(x_i - \mu)^2\right).$$

Since our theory was developed for minimization we have to change the sign and get

$$\rho(x, (\mu, \sigma)) = \ln(\sigma) + \frac{1}{2\sigma^2}(x - \mu)^2,$$

the data space $\mathfrak{Q} = \mathbb{R}$ and the parameter space $\mathfrak{P} \coloneqq \mathbb{R} \times \mathbb{R}_{>0}$ with the metric

$$d((\mu, \sigma), (\tilde{\mu}, \tilde{\sigma})) = |\mu - \tilde{\mu}| + |\ln(\sigma) - \ln(\tilde{\sigma})| + \left|\frac{1}{\sigma^2} - \frac{1}{\tilde{\sigma}^2}\right|.$$

Remark C.2. *We cannot use [EJ20], because there $\mathfrak{Q} = \mathfrak{P}$ is required and we cannot use [Huc11b], because there $\rho \geq 0$ is required. [Sch22] requires*

$$\mathbb{E}\left(\inf_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+}\rho(x, (\mu, \sigma))\right) > -\infty.$$

However, one sees directly, if one inserts $x = \mu$

$$\rho(x, (x, \sigma)) = \ln(\sigma) \overset{\sigma\to 0}{\longrightarrow} -\infty.$$

Theorem C.3. *Under Assumption C.1 ZC holds for the normal distribution.*

Proof. Let $(\mu, \sigma), (\tilde{\mu}, \tilde{\sigma}) \in \mathfrak{P}$ with $d((\mu, \sigma), (\tilde{\mu}, \tilde{\sigma})) < 1$

$$|\rho(x, (\mu, \sigma)) - \rho(x, (\tilde{\mu}, \tilde{\sigma}))| \leq |\ln(\sigma) - \ln(\tilde{\sigma})| + \frac{1}{2}\left|\frac{1}{\sigma^2}(x - \mu)^2 - \frac{1}{\tilde{\sigma}^2}(x - \tilde{\mu})^2\right| \tag{31}$$

$$\leq |\ln(\sigma) - \ln(\tilde{\sigma})| + \frac{1}{4}\left|\frac{1}{\sigma^2} - \frac{1}{\tilde{\sigma}^2}\right|\left|(x - \mu)^2 + (x - \tilde{\mu})^2\right| \tag{32}$$

$$+ \frac{1}{4}\left|\frac{1}{\sigma^2} + \frac{1}{\tilde{\sigma}^2}\right|\left|(x - \mu)^2 - (x - \tilde{\mu})^2\right|.$$

It follows directly

$$|\ln(\sigma) - \ln(\tilde{\sigma})| \leq d((\mu, \sigma), (\tilde{\mu}, \tilde{\sigma})), \qquad \left|\frac{1}{\sigma^2} - \frac{1}{\tilde{\sigma}^2}\right| \leq d((\mu, \sigma), (\tilde{\mu}, \tilde{\sigma})).$$

We get

$$\begin{aligned}\left|(x-\mu)^2+(x-\tilde{\mu})^2\right| &\leq 2x^2+|2x|\,|\mu+\tilde{\mu}|+\left|\mu^2+\tilde{\mu}^2\right| \\ &\leq 2x^2+2\,|x|\,(2\,|\mu|+1)+2\mu^2+2\,|\mu|+1.\end{aligned}$$

We get for the last part of (31)

$$\left|\frac{1}{\sigma^2}+\frac{1}{\tilde{\sigma}^2}\right| \leq 2\left|\frac{1}{\sigma^2}\right|+\left|\frac{1}{\sigma^2}-\frac{1}{\tilde{\sigma}^2}\right| \leq 2\left|\frac{1}{\sigma^2}\right|+1$$

and

$$\begin{aligned}\left|(x-\mu)^2-(x-\tilde{\mu})^2\right| &\leq 2\,|x|\,|\mu-\tilde{\mu}|+2\,|\mu|\,d((\mu,\sigma),(\tilde{\mu},\tilde{\sigma}))+d((\mu,\sigma),(\tilde{\mu},\tilde{\sigma}))^2 \\ &\leq d((\mu,\sigma),(\tilde{\mu},\tilde{\sigma}))\Big(2\,|x|+2\,|\mu|+1\Big).\end{aligned}$$

Substituting the inequalities into (31) results in

$$|\rho(x,(\mu,\sigma))-\rho(x,(\tilde{\mu},\tilde{\sigma}))| \leq d((\mu,\sigma),(\tilde{\mu},\tilde{\sigma}))\dot{\rho}(x,(\mu,\sigma))$$

where $\dot{\rho}(x,(\mu,\sigma))$ is defined as

$$\begin{aligned}\dot{\rho}(x,(\mu,\sigma)) \coloneqq 1&+\frac{1}{4}\left(2x^2+2\,|x|\,(2\,|\mu|+1)+2\mu^2+2\,|\mu|+1\right) \\ &+\frac{1}{4}\left(\frac{2}{\sigma^2}+1\right)(2\,|x|+2\,|\mu|+1)\,.\end{aligned}$$

Therefore we get

$$\begin{aligned}\dot{\mathcal{F}}(\mu,\sigma) = 1&+\frac{1}{4}\left(2\mathbb{E}\left(x^2\right)+2\mathbb{E}\left(|x|\right)(2\,|\mu|+1)+2\mu^2+2\,|\mu|+1\right) \\ &+\frac{1}{4}\left(\frac{2}{\sigma^2}+1\right)(2\mathbb{E}(|x|)+2\,|\mu|+1)\,,\end{aligned}$$

which is smaller than infinity and continuous. □

Theorem C.4. *Under Assumption C.1 BPC holds for the normal distribution.*

Proof. It holds

$$\left\{\left(\mu^{(0)},\sigma^{(0)}\right)\right\} = E^{(\rho)}.$$

If $(\mu_n,\sigma_n^2)_n^\infty \subset \mathfrak{P}$ is without accumulation points then a.s. $\liminf \rho(X,(\mu_n,\sigma_n)) \to \infty$. Thus BPC follows immediately. □

D Technical Theorems and Lemmas for the homoscedastic drift model

In this section, we prove technical theorems and lemmas needed for the strong consistency and central limit theorem in Section 4 in the main text. In particular, we use the definitions of $\rho, \mathfrak{Q}$ and $\mathfrak{P}$ from Section 4.

Lemma D.1. *For $\rho : \mathfrak{Q} \times \mathfrak{P} \mapsto \mathbb{R}$ that is defined, as in (21) holds*

$$\rho(Y, [\kappa]) = \langle Y, Y \rangle_P - \left\langle \hat{\phi}(\kappa, P, Y)\kappa, Y \right\rangle_P$$

for all $\kappa \in [\kappa]$.

Proof. We start with the Definition of ρ

$$\begin{aligned}\rho(Y, [\kappa]) &= \sum_{\nu=1}^{N} \left\| \text{vec}\,(Y_\nu) - M(\kappa_\nu)\text{vec}\left(\hat{\phi}(\kappa, P, Y)\right) \right\|_P^2 \\ &= \langle Y, Y \rangle_P - 2\left\langle \hat{\phi}(\kappa, P, Y)\kappa, Y \right\rangle_P + \left\langle \hat{\phi}(\kappa, P, Y)\kappa, \hat{\phi}(\kappa, P, Y)\kappa \right\rangle_P .\end{aligned}$$

We get for the last part of the equation

$$\begin{aligned}&\left\langle \hat{\phi}(\kappa, P, Y)\kappa, \hat{\phi}(\kappa, P, Y)\kappa \right\rangle_P = \sum_{\nu=1}^{N} \text{vec}\left(\hat{\phi}(\kappa, P, Y)\right)^T M(\kappa_\nu)^T P M(\kappa_\nu)\text{vec}\left(\hat{\phi}(\kappa, P, Y)\right) \\ &= \text{vec}\left(\hat{\phi}(\kappa, P, Y)\right)^T (\kappa \diamond_P \kappa)\,\text{vec}\left(\hat{\phi}(\kappa, P, Y)\right) \\ &= \text{vec}\left(\hat{\phi}(\kappa, P, Y)\right)^T (\kappa \diamond_P \kappa)\,(\kappa \diamond_P \kappa)^{-1}\,(\kappa \bullet_P Y) \\ &= \text{vec}\left(\hat{\phi}(\kappa, P, Y)\right)^T (\kappa \bullet_P Y) .\end{aligned}$$

It follows directly $\text{vec}\left(\hat{\phi}(\kappa, P, Y)\right)^T (\kappa \bullet_P Y) = \left\langle \hat{\phi}(\kappa, P, Y)\kappa, Y \right\rangle_P$ and therefore

$$\rho(Y, [\kappa]) = \langle Y, Y \rangle_P - \left\langle \hat{\phi}(\kappa, P, Y)\kappa, Y \right\rangle_P .$$

□

Definition D.2. *For $P = R\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}R^T \in \text{SPD}(2)$ we define $\tilde{P} = R\begin{pmatrix}\lambda_2 & 0\\ 0 & \lambda_1\end{pmatrix}R^T$ where $\lambda_1 \geq \lambda_2 > 0$ and R is a rotation matrix.*

For the rest of this section, we simplify the notation of sums. Every sum symbol without bounds, $\sum$, is to be understood as a sum $\sum_{\nu=1}^{N}$.

Lemma D.3. *For $\kappa \in \mathcal{S}^{2N-1}$ and $P, \tilde{P} \in \text{SPD}(2)$ as defined in Definition D.2 holds*

$$(\kappa \diamond_P \kappa)^{-1} = \frac{1}{\det(\kappa \diamond_P \kappa)}\left(\kappa \diamond_{\tilde{P}} \kappa\right).$$

Proof. First, we define $M(\tilde{\kappa}_\nu) \coloneqq RM(\kappa_\nu)$ for all $\nu = 1, \dots, N$. Therefore,

$$\begin{aligned}\kappa \diamond_P \kappa &= \tilde{\kappa} \diamond_{\mathrm{diag}(\lambda_1,\lambda_2)} \tilde{\kappa} \\ &= \begin{pmatrix} \lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 & (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) \\ (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) & \lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 \end{pmatrix}.\end{aligned}$$

Using the standard rule for calculating a 2×2 inverse matrix, we get the desired result

$$\begin{aligned}&(\kappa \diamond_P \kappa)^{-1} \\ &= \frac{1}{\det(\kappa \diamond_P \kappa)} \begin{pmatrix} \lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 & (\lambda_1 - \lambda_2) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) \\ (\lambda_1 - \lambda_2) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) & \lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 \end{pmatrix} \\ &= \frac{1}{\det(\kappa \diamond_P \kappa)} \left(\kappa \diamond_{\tilde{P}} \kappa\right).\end{aligned}$$

□

Lemma D.4. *For all $\kappa \in \mathcal{S}^{2N-1}$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\det(\kappa \diamond_P \kappa) \geq \lambda_1 \lambda_2.$$

Proof. We define $\tilde{\kappa}_\nu$ in the same way as in the proof of Lemma D.3. We get using the Cauchy-Schwarz inequality

$$\begin{aligned}&\det(\kappa \diamond_P \kappa) = \det \begin{pmatrix} \lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 & (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) \\ (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) & \lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 \end{pmatrix} \\ &= \left(\lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \left(\lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \\ &- \left((\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu)\right)^2 \\ &\geq \left(\lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \left(\lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \\ &- (\lambda_2 - \lambda_1)^2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2\right) \left(\sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \\ &= \lambda_1 \lambda_2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2\right)^2 + \lambda_1 \lambda_2 \left(\sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 + 2\lambda_1 \lambda_2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2\right) \left(\sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \\ &= \lambda_1 \lambda_2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 = \lambda_1 \lambda_2.\end{aligned}$$

□

D.1 Calculating the modulus of continuity along with its prefactor

This section contains all the calculations needed for the Theorem 4.3 from the main text. 4.3

Lemma D.5. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}, Y \in \mathbb{C}^N$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2*

holds

$$\frac{1}{2}d_P\left(\left(\hat{\phi}(\kappa,P,Y)+\hat{\phi}(\kappa',P,Y)\right)(\kappa-\kappa'),0\right)\leq\lambda_1\sqrt{2N}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\,||Y||\,||\kappa-\kappa'||.$$

Proof. Using first Lemma D.6 and then Lemma D.10 yields

$$\begin{aligned}&\frac{1}{2}d_P\left(\left(\hat{\phi}(\kappa,P,Y)+\hat{\phi}(\kappa',P,Y)\right)(\kappa-\kappa'),0\right)\\&\leq\frac{\lambda_1}{2}\left|\hat{\phi}(\kappa,P,Y)+\hat{\phi}(\kappa',P,Y)\right|\,||\kappa-\kappa'||\leq\lambda_1\sqrt{2N}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\,||Y||\,||\kappa-\kappa'||.\end{aligned}$$

□

Lemma D.6. *For $x\in\mathbb{C}, a\in\mathbb{C}^N$ and $P\in\mathrm{SPD}(2)$ as defined in Definition D.2 with $\lambda_1\geq\lambda_2$ holds*

$$d_P(xa,0)\leq\sqrt{\lambda_1}|x|\cdot||a||.$$

Proof. Since $P\in\mathrm{SPD}(2)$ we can write $P=R^T\mathrm{diag}(\lambda_1,\lambda_2)R$ where

$$R=\begin{pmatrix}\cos(\alpha)&-\sin(\alpha)\\\sin(\alpha)&\cos(\alpha)\end{pmatrix}$$

is a rotation matrix. Therefore

$$\begin{aligned}d_P(xa,0)&=\sqrt{\langle xa,xa\rangle_P}=\sqrt{\langle e^{i\alpha}xa,e^{i\alpha}xa\rangle_{\mathrm{diag}(\lambda_1,\lambda_2)}}\leq\sqrt{\lambda_1\,\langle e^{i\alpha}xa,e^{i\alpha}xa\rangle_{\mathrm{Id}_2}}\\&=\sqrt{\lambda_1}|x|\cdot||a||.\end{aligned}$$

□

Lemma D.7. *For $\kappa\in\mathcal{S}^{2N-1}$ and $P\in\mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$||\kappa\diamond_P\kappa||\leq\sqrt{\lambda_1^2+\lambda_2^2}.$$

Proof. We define $\tilde{\kappa}_\nu$ in the same way as in the proof of Lemma D.3. Using Cauchy–Schwarz

we get

$$\begin{aligned}
||\kappa \diamond_P \kappa||^2 &= \left|\left| \begin{pmatrix} \lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 & (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) \\ (\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu) & \lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2 \end{pmatrix} \right|\right|^2 \\
&= \left(\lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 + 2\left((\lambda_2 - \lambda_1) \sum \mathrm{Re}(\tilde{\kappa}_\nu)\,\mathrm{Im}(\tilde{\kappa}_\nu)\right)^2 \\
&+ \left(\lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 \\
&\leq \left(\lambda_1 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_2 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 \\
&+ 2(\lambda_2 - \lambda_1)^2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2\right)\left(\sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right) \\
&+ \left(\lambda_2 \sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \lambda_1 \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 \\
&= \lambda_1^2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 + \lambda_2^2 \left(\sum \mathrm{Re}(\tilde{\kappa}_\nu)^2 + \sum \mathrm{Im}(\tilde{\kappa}_\nu)^2\right)^2 \\
&= \lambda_1^2 + \lambda_2^2.
\end{aligned}$$

□

Lemma D.8. *For $\kappa \in \mathcal{S}^{2N-1}$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\left|\left|(\kappa \diamond_P \kappa)^{-1}\right|\right| \leq \frac{\sqrt{\lambda_1^2 + \lambda_2^2}}{\lambda_1 \lambda_2}.$$

Proof. We get by using Lemma D.3, Lemma D.4 and Lemma D.7

$$\left|\left|(\kappa \diamond_P \kappa)^{-1}\right|\right| = \left|\left|\frac{1}{\det(\kappa \diamond_P \kappa)} (\kappa \diamond_{\tilde{P}} \kappa)\right|\right| \leq \frac{\sqrt{\lambda_1^2 + \lambda_2^2}}{\lambda_1 \lambda_2}.$$

□

Lemma D.9. *For $\kappa \in \mathcal{S}^{2N-1}, Y \in \mathbb{C}^N$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$||\kappa \bullet_P Y|| \leq \sqrt{2N}\sqrt{\lambda_1^2 + \lambda_2^2}\,||Y||\,.$$

Proof. We define $\tilde{\kappa}_\nu$ in the same way as in the proof of Lemma D.3. We calculate

$$\begin{aligned}
||\kappa \bullet_P Y|| &= \left|\left| \begin{pmatrix} M(\kappa_1)^T & \dots & M(\kappa_N)^T \end{pmatrix} (\mathrm{Id}_N \otimes P) \begin{pmatrix} \mathrm{vec}\,(Y_1) \\ \vdots \\ \mathrm{vec}\,(Y_N) \end{pmatrix} \right|\right| \\
&\leq \left|\left| \begin{pmatrix} M(\kappa_1)^T & \dots & M(\kappa_N)^T \end{pmatrix} \right|\right| ||(\mathrm{Id}_N \otimes P)||\,||Y|| = \sqrt{2N}\sqrt{\lambda_1^2 + \lambda_2^2}\,||Y||\,.
\end{aligned}$$

□

Lemma D.10. *For $\kappa \in \mathcal{S}^{2N-1}, Y \in \mathbb{C}^N$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\left|\hat{\phi}(\kappa, P, Y)\right| \leq \sqrt{2N}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\left|\left|Y\right|\right|.$$

Proof. Using Lemma D.8 and Lemma D.9 we get

$$\left|\hat{\phi}(\kappa, P, Y)\right| = \left|\left|\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \bullet_P Y\right)\right|\right| \leq \left|\left|\left(\kappa \diamond_P \kappa\right)^{-1}\right|\right| \left|\left|\left(\kappa \bullet_P Y\right)\right|\right| \leq \sqrt{2N}\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\left|\left|Y\right|\right|.$$

□

Lemma D.11. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}$ and $Y \in \mathbb{C}^N$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\begin{aligned}
&\frac{1}{2}d_P\left(\left(\hat{\phi}(\kappa, P, Y) - \hat{\phi}(\kappa', P, Y)\right)(\kappa+\kappa'), 0\right) \\
&\leq \left(\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\right)\left(8\sqrt{2}N + \frac{32\sqrt{2}N\left(\lambda_1^2+\lambda_2^2\right)}{\lambda_1\lambda_2} + 2\sqrt{2N}\right)\left|\left|Y\right|\right|\left|\left|\kappa-\kappa'\right|\right|.
\end{aligned}$$

Proof. First, we use Lemma D.6

$$\begin{aligned}
&\frac{1}{2}d_P\left(\left(\hat{\phi}(\kappa, P, Y) - \hat{\phi}(\kappa', P, Y)\right)(\kappa+\kappa'), 0\right) \leq \frac{\sqrt{\lambda_1}}{2}\left|\hat{\phi}(\kappa, P, Y) - \hat{\phi}(\kappa', P, Y)\right| \cdot \left|\left|\kappa+\kappa'\right|\right| \\
&\leq \sqrt{\lambda_1}\left|\hat{\phi}(\kappa, P, Y) - \hat{\phi}(\kappa', P, Y)\right| = \sqrt{\lambda_1}\left|\left|\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \bullet_P Y\right) - \left(\kappa' \diamond_P \kappa'\right)^{-1}\left(\kappa' \bullet_P Y\right)\right|\right| \\
&\leq \frac{\sqrt{\lambda_1}}{2}\left|\left|\left(\kappa \diamond_P \kappa\right)^{-1} - \left(\kappa' \diamond_P \kappa'\right)^{-1}\right|\right|\left|\left|\left(\kappa \bullet_P Y\right) + \left(\kappa' \bullet_P Y\right)\right|\right| \\
&+ \frac{\sqrt{\lambda_1}}{2}\left|\left|\left(\kappa \diamond_P \kappa\right)^{-1} + \left(\kappa' \diamond_P \kappa'\right)^{-1}\right|\right|\left|\left|\left(\kappa \bullet_P Y\right) - \left(\kappa' \bullet_P Y\right)\right|\right|.
\end{aligned}$$

From Lemma D.8, Lemma D.9, Lemma D.12 and Lemma D.13 it follows that

$$\begin{aligned}
&\frac{1}{2}d_P\left(\left(\hat{\phi}(\kappa, P, Y) - \hat{\phi}(\kappa', P, Y)\right)(\kappa+\kappa'), 0\right) \\
&\leq \left(\left(\frac{4\sqrt{N}\sqrt{\lambda_1^2+\lambda_2^2}}{\lambda_1\lambda_2} + \frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)^{3/2}}{(\lambda_1\lambda_2)^2}\right)\left|\left|\kappa-\kappa'\right|\right|\right)\left(2\sqrt{2N}\sqrt{\lambda_1^2+\lambda_2^2}\left|\left|Y\right|\right|\right) \\
&+ \left(2\frac{\sqrt{\lambda_1^2+\lambda_2^2}}{\lambda_1\lambda_2}\right)\left(\sqrt{2N}\sqrt{\lambda_1^2+\lambda_2^2}\left|\left|Y\right|\right|\left|\left|\kappa-\kappa'\right|\right|\right) \\
&= \left(\frac{\lambda_1^2+\lambda_2^2}{\lambda_1\lambda_2}\right)\left(8\sqrt{2}N + \frac{32\sqrt{2}N\left(\lambda_1^2+\lambda_2^2\right)}{\lambda_1\lambda_2} + 2\sqrt{2N}\right)\left|\left|Y\right|\right|\left|\left|\kappa-\kappa'\right|\right|.
\end{aligned}$$

□

Lemma D.12. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}, Y \in \mathbb{C}^N$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2*

holds

$$\left|\left|(\kappa \bullet_P Y) - \left(\kappa' \bullet_P Y\right)\right|\right| \leq \sqrt{2N}\sqrt{\lambda_1^2+\lambda_2^2}\,||Y||\,\left|\left|\kappa-\kappa'\right|\right|.$$

Proof. Analogous to the proof from Lemma D.9 it follows directly

$$\begin{aligned}
&\left|\left|(\kappa \bullet_P Y) - \left(\kappa' \bullet_P Y\right)\right|\right| \\
&= \left|\left| \begin{pmatrix} M(\kappa_1-\kappa_1')^T & \dots & M(\kappa_N-\kappa_N')^T \end{pmatrix} (\mathrm{Id}_N \otimes P) \begin{pmatrix} \mathrm{vec}\,(Y_1) \\ \vdots \\ \mathrm{vec}\,(Y_N) \end{pmatrix} \right|\right| \\
&\leq \left|\left| \begin{pmatrix} M(\kappa_1-\kappa 1')^T & \dots & M(\kappa_N-\kappa_N')^T \end{pmatrix} \right|\right| ||(\mathrm{Id}_N \otimes P)||\,||Y|| \\
&= \sqrt{2N}\sqrt{\lambda_1^2+\lambda_2^2}\,||Y||\,\left|\left|\kappa-\kappa'\right|\right|.
\end{aligned}$$

□

Lemma D.13. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\begin{aligned}
&\left|\left|(\kappa \diamond_P \kappa)^{-1} - \left(\kappa' \diamond_P \kappa'\right)^{-1}\right|\right| \\
&\leq \left(\frac{4\sqrt{N}\sqrt{\lambda_1^2+\lambda_2^2}}{\lambda_1\lambda_2} + \frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)\sqrt{\lambda_1^2+\lambda_2^2}}{(\lambda_1\lambda_2)^2}\right)\left|\left|\kappa-\kappa'\right|\right|.
\end{aligned}$$

Proof. We get by using Lemma D.3

$$\begin{aligned}
&\left|\left|(\kappa \diamond_P \kappa)^{-1} - \left(\kappa' \diamond_P \kappa'\right)^{-1}\right|\right| = \left|\left|\frac{1}{\det(\kappa \diamond_P \kappa)}\left(\kappa \diamond_{\tilde{P}} \kappa\right) - \frac{1}{\det(\kappa' \diamond_P \kappa')}\left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right| \\
&\leq \frac{1}{2}\left|\frac{1}{\det(\kappa \diamond_P \kappa)} + \frac{1}{\det(\kappa' \diamond_P \kappa')}\right| \left|\left|\left(\kappa \diamond_{\tilde{P}} \kappa\right) - \left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right| \\
&+ \frac{1}{2}\left|\frac{1}{\det(\kappa \diamond_P \kappa)} - \frac{1}{\det(\kappa' \diamond_P \kappa')}\right| \left|\left|\left(\kappa \diamond_{\tilde{P}} \kappa\right) + \left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right|.
\end{aligned}$$

From Lemma D.4, Lemma D.7, Lemma D.14 and Lemma D.16 it follows that

$$\begin{aligned}
&\left|\left|(\kappa \diamond_P \kappa)^{-1} - \left(\kappa' \diamond_P \kappa'\right)^{-1}\right|\right| \\
&\leq \frac{1}{2}\left(\frac{2}{\lambda_1\lambda_2}\right)\left(4\sqrt{N}\sqrt{\lambda_1^2+\lambda_2^2}\left|\left|\kappa-\kappa'\right|\right|\right) \\
&+ \frac{1}{2}\left(\frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)}{(\lambda_1\lambda_2)^2}\left|\left|\kappa-\kappa'\right|\right|\right)\left(2\sqrt{\lambda_1^2+\lambda_2^2}\right) \\
&\leq \left(\frac{4\sqrt{N}\sqrt{\lambda_1^2+\lambda_2^2}}{\lambda_1\lambda_2} + \frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)\sqrt{\lambda_1^2+\lambda_2^2}}{(\lambda_1\lambda_2)^2}\right)\left|\left|\kappa-\kappa'\right|\right|.
\end{aligned}$$

□

Lemma D.14. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\left|\left|\left(\kappa \diamond_{\tilde{P}} \kappa\right) - \left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right| \leq 4\sqrt{N}\sqrt{\lambda_1^2 + \lambda_2^2}\left|\left|\kappa - \kappa'\right|\right|.$$

Proof. First, we use the triangle inequality and the sub-multiplicative property of the Frobenius norm:

$$\begin{aligned}
&\left|\left|\left(\kappa \diamond_{\tilde{P}} \kappa\right) - \left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right| = \frac{1}{2}\left|\left|\left((\kappa - \kappa') \diamond_{\tilde{P}} (\kappa + \kappa')\right) + \left((\kappa + \kappa') \diamond_{\tilde{P}} (\kappa - \kappa')\right)\right|\right| \\
&\leq \left|\left|\left((\kappa + \kappa') \diamond_{\tilde{P}} (\kappa - \kappa')\right)\right|\right| \\
&\leq \left|\left|\begin{pmatrix} M(\kappa_1 + \kappa_1')^T & \dots & M(\kappa_N + \kappa_N')^T \end{pmatrix}\right|\right| \left|\left|(\mathrm{Id}_N \otimes \tilde{P})\right|\right| \left|\left|\begin{pmatrix} M(\kappa_1 - \kappa_1') \\ \vdots \\ M(\kappa_N - \kappa_N') \end{pmatrix}\right|\right|.
\end{aligned}$$

Analogously to Proof of Lemma D.9 we get

$$\left|\left|\begin{pmatrix} M(\kappa_1 + \kappa_1')^T & \dots & M(\kappa_N + \kappa_N')^T \end{pmatrix}\right|\right| \left|\left|(\mathrm{Id}_N \otimes \tilde{P})\right|\right| \leq 2\sqrt{2N}\sqrt{\lambda_1^2 + \lambda_2^2}.$$

Thus, we get the desired result

$$\left|\left|\left(\kappa \diamond_{\tilde{P}} \kappa\right) - \left(\kappa' \diamond_{\tilde{P}} \kappa'\right)\right|\right| \leq 4\sqrt{N}\sqrt{\lambda_1^2 + \lambda_2^2}\left|\left|\kappa - \kappa'\right|\right|.$$

□

Lemma D.15. *For $A, B \in \mathbb{R}^{2\times 2}$ we have*

$$\Big|\det(A) - \det(B)\Big| \leq 2||A + B||||A - B||.$$

Proof. We directly calculate the determinant of the 2×2 matrix and use the triangle inequality

$$\begin{aligned}
&\Big|\det(A) - \det(B)\Big| = \Big|a_{11}a_{22} - a_{12}a_{21} - (b_{11}b_{22} - b_{12}b_{21})\Big| \\
&\leq \Big|a_{11}a_{22} - b_{11}b_{22}\Big| + \Big|a_{12}a_{21} - b_{12}b_{21}\Big| \\
&= \frac{1}{2}\Big|(a_{11} + b_{11})(a_{22} - b_{22}) + (a_{11} - b_{11})(a_{22} + b_{22})\Big| \\
&+ \frac{1}{2}\Big|(a_{12} + b_{12})(a_{21} - b_{21}) + (a_{12} - b_{12})(a_{21} + b_{21})\Big| \\
&\leq \frac{1}{2}\Big|a_{11} + b_{11}\Big|\Big|a_{22} - b_{22}\Big| + \frac{1}{2}\Big|a_{11} - b_{11}\Big|\Big|a_{22} + b_{22}\Big| \\
&+ \frac{1}{2}\Big|a_{12} + b_{12}\Big|\Big|a_{21} - b_{21}\Big| + \frac{1}{2}\Big|a_{12} - b_{12}\Big|\Big|a_{21} + b_{21})\Big|.
\end{aligned}$$

Using that $\left|a_{ij}+b_{ij}\right| \leq ||A+B||$ and $\left|a_{ij}-b_{ij}\right| \leq ||A-B||$ for all $i,j=1,2$ we get

$$\left|\det(A)-\det(B)\right| \leq 2||A+B||||A-B||.$$

□

Lemma D.16. *For $\kappa, \kappa' \in \mathcal{S}^{2N-1}$ and $P \in \mathrm{SPD}(2)$ as defined in Definition D.2 holds*

$$\left|\frac{1}{\det\left(\kappa \diamond_P \kappa\right)}-\frac{1}{\det\left(\kappa' \diamond_P \kappa'\right)}\right| \leq \frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)}{(\lambda_1\lambda_2)^2}\left|\left|\kappa-\kappa'\right|\right|.$$

Proof. It follows from Lemma D.4 and Lemma D.15.

$$\begin{aligned}\left|\frac{1}{\det\left(\kappa \diamond_P \kappa\right)}-\frac{1}{\det\left(\kappa' \diamond_P \kappa'\right)}\right| &\leq \frac{1}{(\lambda_1\lambda_2)^2}\left|\det\left(\kappa' \diamond_P \kappa'\right)-\det\left(\kappa \diamond_P \kappa\right)\right| \\ &\leq \frac{2}{(\lambda_1\lambda_2)^2}\left|\left|\left(\kappa' \diamond_P \kappa'\right)+\left(\kappa \diamond_P \kappa\right)\right|\right|\left|\left|\left(\kappa' \diamond_P \kappa'\right)-\left(\kappa \diamond_P \kappa\right)\right|\right|.\end{aligned}$$

Consequently, from Lemma D.7 and Lemma D.14 follows the desired result:

$$\begin{aligned}\left|\frac{1}{\det\left(\kappa \diamond_P \kappa\right)}-\frac{1}{\det\left(\kappa' \diamond_P \kappa'\right)}\right| &\leq \frac{2}{(\lambda_1\lambda_2)^2}\left(2\sqrt{\lambda_1^2+\lambda_2^2}\right)\left(4\sqrt{N}\sqrt{\lambda_1^2+\lambda_2^2}\left|\left|\kappa-\kappa'\right|\right|\right) \\ &= \frac{16\sqrt{N}(\lambda_1^2+\lambda_2^2)}{(\lambda_1\lambda_2)^2}\left|\left|\kappa-\kappa'\right|\right|.\end{aligned}$$

□

Lemma D.17. *Let $x,y \in \mathbb{C}$ and $a,b \in \mathbb{C}^N$. It holds*

$$d_P(xa,yb) \leq \frac{1}{2}d_P\Big(\left(x+y\right)\left(a-b\right),0\Big)+\frac{1}{2}d_P\Big(\left(x-y\right)\left(a+b\right),0\Big). \tag{33}$$

Proof. Using the Triangle inequality we get

$$\begin{aligned}d_P(xa,yb) &= d_P\left(\frac{1}{2}(x+y)(a-b)+\frac{1}{2}(x-y)(a+b),0\right) \\ &\leq \frac{1}{2}d_P\Big(\left(x+y\right)\left(a-b\right),0\Big)+\frac{1}{2}d_P\Big(\left(x-y\right)\left(a+b\right),0\Big).\end{aligned}$$

□

D.2 CLT

In this section we have the same notation as in the Section 4.2 in the main text.

Lemma D.18. *For β^{-1} from Definition 4.8 in the main text holds*

$$d(\left[\beta^{-1}(x)\right],\left[\kappa^{(0)}\right])^2 = 1-\frac{1}{\sqrt{||x||^2+1}}.$$

Proof. For $[\kappa], [\tilde{\kappa}] \in \mathfrak{P}$ it holds

$$d\left([\kappa],[\tilde{\kappa}]\right)^2 = \min_{\lambda\in\mathbb{R}} ||\kappa - e^{i\lambda}\tilde{\kappa}||^2 = \min_{\lambda\in\mathbb{R}} 2\left(1 - \mathrm{Re}(e^{i\lambda}\kappa^*\kappa)\right).$$

Consequently, if $\kappa^*\tilde{\kappa} \in \mathbb{R}_{>0}$, then $\kappa \in [\kappa], \tilde{\kappa} \in [\tilde{\kappa}]$ are in optimal position and it holds $d\left([\kappa],[\tilde{\kappa}]\right)^2 = ||\kappa - \tilde{\kappa}||^2$. For $\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right) \in \beta^{-1}(x)$ from Definition 4.8 in the main text,

$$\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)^T \kappa^{(0)} = \frac{1}{||\tilde{x}||} \in \mathbb{R}_{>0}$$

holds and thus

$$d(\beta^{-1}(x),[\kappa^{(0)}])^2 = \left|\left|\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right) - \kappa^{(0)}\right|\right|^2 = 1 - \frac{1}{||\tilde{x}||} = 1 - \frac{1}{\sqrt{||x||^2+1}}.$$

□

D.3 Auxiliary calculations for Section 4.3

In this section we have the same notation as in the Section 4.3 in the main text.

Lemma D.19. *For $\kappa \in \mathcal{S}^{2N-1}$ holds*

$$\arg\max_{\lambda\in\mathbb{S}^1} \left|\left|\mathrm{Re}(e^{i\lambda}\kappa)\right|\right|^2 = \begin{cases} \left\{\pi - \frac{\mathrm{Arg}(\kappa^T\kappa)}{2}, 2\pi - \frac{\mathrm{Arg}(\kappa^T\kappa)}{2}\right\}, & \text{if } \kappa^T\kappa \neq 0 \\ \mathbb{S}^1, & \text{else.} \end{cases}$$

Proof. Euler's formula gives us

$$\begin{aligned} &\arg\max_{\lambda\in\mathbb{S}^1} \left|\left|\mathrm{Re}(e^{i\lambda}\kappa)\right|\right|^2 = \arg\max_{\lambda\in\mathbb{S}^1} \left|\left|e^{i\lambda}\kappa + e^{-i\lambda}\bar{\kappa}\right|\right|^2 \\ &= \arg\max_{\lambda\in\mathbb{S}^1} \left(||\kappa||^2 + e^{2i\lambda}\kappa^T\kappa + \overline{e^{2i\lambda}\kappa^T\kappa}\right). \end{aligned}$$

If $\kappa^T\kappa = 0$, then all $\lambda \in \mathbb{S}^1$ maximize the expression. If $\kappa^T\kappa \neq 0$, then there is exactly one $\alpha \in \mathbb{S}^1$ with $\alpha = \mathrm{Arg}(\kappa^T\kappa)$ and we get $\kappa^T\kappa = re^{i\alpha}$, where $r = |\kappa^T\kappa| > 0$. Substituting $\kappa^T\kappa = re^{i\alpha}$ and using the angle addition and subtraction theorems gives us:

$$\arg\max_{\lambda\in\mathbb{S}^1} \left|\left|\mathrm{Re}(e^{i\lambda}\kappa)\right|\right|^2 = \arg\max_{\lambda\in\mathbb{S}^1} \left(r\cos(2\lambda+\alpha)\right).$$

The expression $\cos(2\lambda+\alpha)$ is maximized exactly when $2\lambda + \alpha = 0 \mod 2\pi$ holds. Therefore

$$\arg\max_{\lambda\in\mathbb{S}^1} \left|\left|\mathrm{Re}(e^{i\lambda}\kappa)\right|\right|^2 = \left\{\pi - \frac{\alpha}{2}, 2\pi - \frac{\alpha}{2}\right\}.$$

□

Lemma D.20. *Let $[\kappa] \in \mathfrak{P}$ then it holds for all $\kappa, \tilde{\kappa} \in [\kappa]$ that $\tilde{f}(\kappa) = \tilde{f}(\tilde{\kappa})$, where $\tilde{f}$ is defined as in Equation (24) in the main text.*

Proof. For $\kappa \in \mathcal{S}^{2N-1}$ with $\kappa^T\kappa = 0$ the proposition is trivially satisfied. Let $\kappa \in \mathcal{S}^{2N-1}$ with $\left|\kappa^T\kappa\right| = r > 0$ and let $\tilde{\kappa} \in [\kappa]$ then there is a $\lambda \in \mathbb{S}^1$ with $\tilde{\kappa} = e^{i\lambda}\kappa$. It follows

$$\begin{aligned}
&\operatorname{Arg}\left(\tilde{\kappa}^T\tilde{\kappa}\right) = \operatorname{Arg}\left(\left(e^{i\lambda}\kappa\right)^T\left(e^{i\lambda}\kappa\right)\right) \\
&= \operatorname{Arg}\left(re^{i\left(2\lambda+\operatorname{Arg}\left(\kappa^T\kappa\right)\right)}\right) = 2\lambda + \operatorname{Arg}\left(\kappa^T\kappa\right) \mod 2\pi.
\end{aligned}$$

Thus it follows

$$\tilde{f}(\tilde{\kappa}) = \left[\operatorname{Re}\left(e^{\frac{-i}{2}\left(2\lambda+\operatorname{Arg}\left(\kappa^T\kappa\right)\right)}\left(e^{i\lambda}\kappa\right)\right)\right]_{\pm} = \tilde{f}(\kappa).$$

□

Lemma D.21. *Let g be defined as in Equation (25) and let $\kappa^{(0)} \in \mathfrak{P} \setminus (M_1 \cup M_2)$ then the Jacobian matrix at $x = 0$ is given by*

$$J_x g(0) = \pm \operatorname{Re}\left(e^{-\frac{i\alpha}{2}}\left(i\kappa^0 \operatorname{Re}\left(i\frac{\overline{(\kappa^{(0)})^T\kappa^{(0)}}\left(\kappa^{(0)}\right)^T R^*A}{r^2}\right) + R^*A\right)\right).$$

where $\left|(\kappa^{(0)})^T\kappa^{(0)}\right| =: r$ and $\alpha := \operatorname{Arg}\left((\kappa^{(0)})^T\kappa^{(0)}\right)$.

Proof. As $\kappa^{(0)} \notin M_1$, it follows $\left|(\kappa^{(0)})^T\kappa^{(0)}\right| =: r > 0$ and we can write $(\kappa^{(0)})^T\kappa^{(0)} := re^{i\alpha}$, where $\alpha := \operatorname{Arg}\left((\kappa^{(0)})^T\kappa^{(0)}\right)$. Since $\kappa \notin M_2$, the outer function $f_\pm$ is the identity or minus the identity, only the sign of the Jacobian matrix is determined by this function. We therefore get

$$g : \mathbb{R}^{2N-2} \to \mathbb{R}^N, \quad x \mapsto \pm f(\beta^{-1}(x)) = \pm \operatorname{Re}\left(e^{\frac{-i}{2}\operatorname{Arg}\left(\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)^T\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)\right)}\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)\right)$$

where $\tilde{x}$ and R are defined as in Definition 4.8 in the main text. We first calculate

$$J_x \operatorname{Arg}\left(\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)^T\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)\right) = J_x \operatorname{Arg}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) \tag{34}$$

$$\begin{aligned}
&= \frac{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) J_x\left(\operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)\right)}{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2 + \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2} \\
&- \frac{\operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) J_x\left(\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)\right)}{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2 + \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2}
\end{aligned} \tag{35}$$

where

$$\begin{aligned}
J_x \operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) &= \frac{1}{2}\left(J_x\left((R^*\tilde{x})^T(R^*\tilde{x})\right) + \overline{J_x(R^*\tilde{x})^T(R^*\tilde{x})}\right), \\
J_x \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) &= \frac{-i}{2}\left(J_x\left((R^*\tilde{x})^T(R^*\tilde{x})\right) - \overline{J_x(R^*\tilde{x})^T(R^*\tilde{x})}\right).
\end{aligned}$$

Substitution of this into (34) results in

$$\begin{aligned}
&J_x \operatorname{Arg}\left(\left(R^*\tilde{x}\right)^T\left(R^*\tilde{x}\right)\right)\\
&= \frac{-\frac{1}{2}\left(\operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) + \operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)i\right)\left(J_x(R^*\tilde{x})^T(R^*\tilde{x})\right)}{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2 + \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2}\\
&- \frac{\frac{1}{2}\left(\operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) - \operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)i\right)\left(\overline{J_x(R^*\tilde{x})^T(R^*\tilde{x})}\right)}{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2 + \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2}\\
&= -\operatorname{Re}\left(\frac{\left(\operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right) + \operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)i\right)J_x\left((R^*\tilde{x})^T(R^*\tilde{x})\right)}{\operatorname{Re}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2 + \operatorname{Im}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)^2}\right)
\end{aligned}$$

where

$$J_x(R^*\tilde{x})^T(R^*\tilde{x}) = 2\tilde{x}^T\left(R^*\right)^T R^* A,$$

where

$$A \coloneqq \begin{pmatrix}
1 & i & 0 & 0 & 0 & \dots & 0 & 0\\
0 & 0 & 1 & i & 0 & \dots & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 0 & 0 & \dots & 1 & i\\
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0
\end{pmatrix} \in \mathbb{C}^{N\times 2(N-1)}.$$

Using that same matrix A we get

$$J_x R^* \frac{\tilde{x}}{||\tilde{x}||} = R^*\left(\tilde{x}\left(J_x \frac{1}{||\tilde{x}||}\right) + \frac{1}{||\tilde{x}||}\left(J_x\tilde{x}\right)\right) = R^*\left(-\frac{1}{||\tilde{x}||^{3/2}}\tilde{x}x^T + \frac{1}{||\tilde{x}||}A\right). \tag{36}$$

Consequently, we get

$$\begin{aligned}
&J_x g(x)\\
&= \pm\operatorname{Re}\left(e^{\frac{-i}{2}\operatorname{Arg}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)}\left(-\frac{i}{2}\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)J_x\left(\operatorname{Arg}\left(\left(R^*\tilde{x}\right)^T\left(R^*\tilde{x}\right)\right)\right) + J_x R^*\frac{\tilde{x}}{||\tilde{x}||}\right)\right).
\end{aligned}$$

Inserting $x = 0$ gives us

$$\begin{aligned}
e^{\frac{-i}{2}\operatorname{Arg}\left((R^*\tilde{x})^T(R^*\tilde{x})\right)}\Big|_{x=0} &= \pm\exp\left(\frac{-i\alpha}{2}\right)\\
J_x R^*\frac{\tilde{x}}{||\tilde{x}||}\Big|_{x=0} &= R^*A\\
-\frac{i}{2}\left(R^*\frac{\tilde{x}}{||\tilde{x}||}\right)J_x\left(\operatorname{Arg}\left(\left(R^*\tilde{x}\right)^T\left(R^*\tilde{x}\right)\right)\right)\Big|_{x=0} &= i\kappa^0\operatorname{Re}\left(i\frac{\overline{(\kappa^{(0)})^T\kappa^{(0)}}\left(\kappa^{(0)}\right)^T R^*A}{r^2}\right).
\end{aligned}$$

By substituting into Equation (36) we get

$$J_x g(0) = \pm \operatorname{Re}\left(e^{-\frac{i\alpha}{2}} \left(i\kappa^0 \operatorname{Re}\left(i \frac{\overline{(\kappa^{(0)})^T \kappa^{(0)}} \left(\kappa^{(0)}\right)^T R^* A}{r^2} \right) + R^* A \right) \right).$$

□

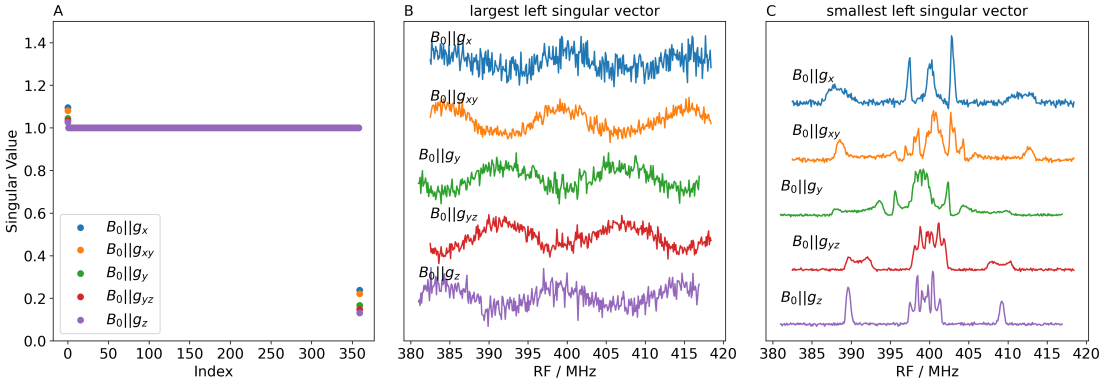


Figure C9: The singular values of $J_x g(0)$ for the different orientations from a chemical sample of the D2-$Y_{122}^{\bullet}$ E. coli ribonucleotide reductase. Remarkably, almost all singular values are equal to 1, with one value being slightly larger and one value being markedly smaller (but clearly separated from zero). The variation of the function g stems from two sources. First, changes in κ are directly translated into changes in the spectrum, which account for the flat eigenvalue spectrum. Second, the complex rotation by λ, which depends on κ, changes the spectrum. The eigenvector to the smallest eigenvalue therefore corresponds very closely to the spectrum itself since we evaluate the Jacobian at this point and thus, when varying κ and hence λ, the change is mostly tangential to that direction. The eigenvector to the largest eigenvalue corresponds closely to the "imaginary part of the spectrum" which is projected out, so when varying κ the corresponding variation in λ, which mixes more or less of the wave into the spectrum, compounds the change in this direction, leading to an increased eigenvalue.

E Technical Theorems and Lemmas for Section 5 in the main text

In this section, we prove technical lemmas for Section 5 in the main text. Consequently, we have the same notation, in particular for ρ, $\mathfrak{Q}$ and $\mathfrak{P}$.

Lemma E.1. *Under Definiton 5.1, it holds for* $([\kappa], P) \in \mathfrak{P}$

$$\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\epsilon\right) = N \operatorname{Tr}\left(\Sigma^{(0)} P\right) - \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right)$$

where $\kappa \in [\kappa]$

Proof. Using Lemma D.1 we get

$$\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\epsilon\right) = \int \langle \epsilon, \epsilon \rangle_P - \left\langle \hat{\phi}(\kappa, P, \epsilon)\kappa, \epsilon \right\rangle_P \mathrm{d}\mathbb{P}\left(\epsilon\right)$$

First, we calculate

$$\int \langle \epsilon, \epsilon \rangle_P \, \mathrm{d}\mathbb{P}\left(\epsilon\right) = \sum_{\nu=1}^{N} \operatorname{Tr}\left(\mathbb{V}\left(\sqrt{P}\mathrm{vec}\left(\epsilon_\nu\right)\right)\right) = \sum_{\nu=1}^{N} \operatorname{Tr}\left(\sqrt{P}\Sigma^{(0)}\sqrt{P}\right) = N \operatorname{Tr}\left(\Sigma^{(0)} P\right).$$

By using the linearity and the cyclic property of the trace operator we get

$$\begin{aligned}
&\int \left\langle \hat{\phi}(\kappa, P, \epsilon)\kappa, \epsilon \right\rangle_P \mathrm{d}\mathbb{P}\left(\epsilon\right) = \int \sum_{\nu=1}^{N} \mathrm{vec}\left(\epsilon_\nu\right) PM(\kappa_\nu) \mathrm{vec}\left(\hat{\phi}(\kappa, P, \epsilon)\right) \mathrm{d}\mathbb{P}\left(\epsilon\right) \\
&= \int \left(\kappa \bullet_P \epsilon\right)^T \left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \bullet_P \epsilon\right) \mathrm{d}\mathbb{P}\left(\epsilon\right) = \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \int \left(\kappa \bullet_P \epsilon\right) \left(\kappa \bullet_P \epsilon\right)^T \mathrm{d}\mathbb{P}\left(\epsilon\right)\right).
\end{aligned}$$

Since $\epsilon_1, \ldots, \epsilon_N \sim \mathcal{N}(0, \Sigma^{(0)})$ are i.i.id random variables, it holds that

$$\begin{aligned}
&\int \left(\kappa \bullet_P \epsilon\right) \left(\kappa \bullet_P \epsilon\right)^T \mathrm{d}\mathbb{P}\left(\epsilon\right) = \int \left(\sum_{\nu=1}^{N} M(\kappa_\nu)^T P \mathrm{vec}\left(\epsilon_\nu\right)\right) \left(\sum_{\nu=1}^{N} \mathrm{vec}\left(\epsilon_\nu\right)^T PM(\kappa_\nu)\right) \mathrm{d}\mathbb{P}\left(\epsilon\right) \\
&= \sum_{\nu=1}^{N} M(\kappa_\nu)^T P \left(\int \mathrm{vec}\left(\epsilon_\nu\right) \mathrm{vec}\left(\epsilon_\nu\right)^T \mathrm{d}\mathbb{P}\left(\epsilon\right)\right) PM(\kappa_\nu) = \kappa \diamond_{P\Sigma^{(0)}P} \kappa.
\end{aligned}$$

Consequently,

$$\int \left\langle \hat{\phi}(\kappa, P, \epsilon)\kappa, \epsilon \right\rangle_P \mathrm{d}\mathbb{P}\left(\epsilon\right) = \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right).$$

and therefore

$$\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\epsilon\right) = N \operatorname{Tr}\left(\Sigma^{(0)} P\right) - \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right).$$

□

Lemma E.2. *For* $([\kappa], P) \in \mathfrak{P}$, $\kappa \in [\kappa]$ *and* $\Sigma^{(0)} \in \mathrm{SPD}(2)$ *we obtain*

$$(i) \quad \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{\left(\frac{\partial P}{\partial p_{ij}}\right)\Sigma^{(0)} P} \kappa\right)\right) = \mathrm{Tr}\left(\Sigma^{(0)} P\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa}\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right),$$

$$(ii) \quad \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right)} \kappa\right)\right) = \mathrm{Tr}\left(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa}\right) P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right)\right),$$

$$(iii) \quad \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{\frac{\partial P}{\partial p_{ij}}} \kappa\right)(\kappa \diamond_P \kappa)^{-1}(\kappa \diamond_{P\Sigma^{(0)} P} \kappa)\right)$$
$$= \mathrm{Tr}\left(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)} P}\kappa\right)(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa}\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right).$$

Proof. We start with (i). From the cyclic property of the trace operator we obtain

$$\begin{aligned}
&\mathrm{Tr}\left(((\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{\left(\frac{\partial P}{\partial p_{ij}}\right)\Sigma^{(0)} P} \kappa\right)\right) \\
&= \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\left(\sum_{\nu=1}^{N} M(\kappa_\nu)^T\left(\frac{\partial P}{\partial p_{ij}}\right)\Sigma^{(0)} P M(\kappa_\nu)\right)\right) \\
&= \mathrm{Tr}\left(\Sigma^{(0)} P\left(\sum_{\nu=1}^{N} M(\kappa_\nu)(\kappa \diamond_P \kappa)^{-1} M(\kappa_\nu)^T\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right) \\
&= \mathrm{Tr}\left(\Sigma^{(0)} P\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa}\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right).
\end{aligned}$$

Analogously, we obtain for (ii)

$$\begin{aligned}
&\mathrm{Tr}\left(((\kappa \diamond_P \kappa)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right)} \kappa\right)\right) \\
&= \mathrm{Tr}\left((\kappa \diamond_P \kappa)^{-1}\left(\sum_{\nu=1}^{N} M(\kappa_\nu)^T P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right) M(\kappa_\nu)\right)\right) \\
&= \mathrm{Tr}\left(\left(\sum_{\nu=1}^{N} M(\kappa_\nu)(\kappa \diamond_P \kappa)^{-1} M(\kappa_\nu)^T\right)\Sigma^{(0)} P\left(\frac{\partial P}{\partial p_{ij}}\right)\right) \\
&= \mathrm{Tr}\left(\left(\sum_{\nu=1}^{N} M(\kappa_\nu)(\kappa \diamond_P \kappa)^{-1} M(\kappa_\nu)^T\right)\Sigma^{(0)} P\left(\frac{\partial P}{\partial p_{ij}}\right)\right) \\
&= \mathrm{Tr}\left(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa}\right)\Sigma^{(0)} P\left(\frac{\partial P}{\partial p_{ij}}\right)\right).
\end{aligned}$$

For (iii) we use the cyclic property of the trace operator

$$\begin{aligned}
&\operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{\frac{\partial P}{\partial p_{ij}}} \kappa\right)\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right)\\
&\operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{\frac{\partial P}{\partial p_{ij}}} \kappa\right)\right)\\
&\operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\left(\kappa \diamond_P \kappa\right)^{-1}\left(\sum_{\nu=1}^{N} M(\kappa_\nu)^T\left(\frac{\partial P}{\partial p_{ij}}\right) M(\kappa_\nu)\right)\right)\\
&\operatorname{Tr}\left(\sum_{\nu=1}^{N}\left(M(\kappa_\nu)\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\left(\kappa \diamond_P \kappa\right)^{-1} M(\kappa_\nu)^T\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right)\\
&= \operatorname{Tr}\left(\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}\left(\kappa\diamond_{P\Sigma^{(0)}P}\kappa\right)(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right)\left(\frac{\partial P}{\partial p_{ij}}\right)\right).
\end{aligned}$$

□

Lemma E.3. *Let $f : \mathrm{SPD}(2) \to \mathbb{R}$ be a differentiable function with the property*

$$\frac{\partial f(P)}{\partial p_{ij}} = \operatorname{Tr}\left(A\left(\frac{\partial P}{\partial p_{ij}}\right)\right)$$

where $P = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix} \in \mathrm{SPD}(2)$ and A is any symmetric matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \in \mathbb{R}^{2\times 2}$. Then holds

$$\frac{\partial f(P)}{\partial P} = 2A - \operatorname{diag}(A).$$

Proof. It holds

$$\operatorname{Tr}\left(A\left(\frac{\partial P}{\partial p_{11}}\right)\right) = a_{11}, \quad \operatorname{Tr}\left(A\left(\frac{\partial P}{\partial p_{22}}\right)\right) = a_{22}, \quad \operatorname{Tr}\left(A\left(\frac{\partial P}{\partial p_{12}}\right)\right) = 2a_{12}$$

and therefore

$$\frac{\partial f(P)}{\partial P} = 2A - \operatorname{diag}(A).$$

□

Lemma E.4. *For* $([\kappa], P) \in \mathfrak{P}$, $\kappa \in [\kappa]$ *and* $\Sigma^{(0)} \in \mathrm{SPD}(2)$ *we obtain*

$$\begin{aligned}
&\frac{\partial}{\partial P}\left(N \operatorname{Tr}\left(\Sigma^{(0)} P\right) - \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right) - N \log(\det(P))\right) \\
&= 2\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right) P\Sigma^{(0)} + \Sigma^{(0)} P\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right) - 2\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}\left(\kappa\diamond_{P\Sigma^{(0)}P}\kappa\right)(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right) \\
&- \operatorname{diag}\Big(\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right) P\Sigma^{(0)} + \Sigma^{(0)} P\left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right) \\
&- \left(\bar{\kappa} \diamond_{(\kappa\diamond_P\kappa)^{-1}\left(\kappa\diamond_{P\Sigma^{(0)}P}\kappa\right)(\kappa\diamond_P\kappa)^{-1}} \bar{\kappa}\right)\Big) \\
&+ N\left(2\Sigma^{(0)} - \operatorname{diag}(\Sigma^{(0)})\right) - N\left(2P^{-1} - \operatorname{diag}\left((2P^{-1})\right)\right).
\end{aligned}$$

Proof. Since $P, \Sigma^{(0)} \in \mathrm{SPD}(2)$ we can write $P = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix}$ and $\Sigma^{(0)} = \begin{pmatrix} \sigma_{11}^{(0)} & \sigma_{12}^{(0)} \\ \sigma_{12}^{(0)} & \sigma_{22}^{(0)} \end{pmatrix}$. For the first term in the sum we get

$$\frac{\partial \operatorname{Tr}\left(\Sigma^{(0)} P\right)}{\partial p_{11}} = \sigma_{11}^{(0)}, \quad \frac{\partial \operatorname{Tr}\left(\Sigma^{(0)} P\right)}{\partial p_{22}} = \sigma_{22}^{(0)}, \quad \frac{\partial \operatorname{Tr}\left(\Sigma^{(0)} P\right)}{\partial p_{12}} = 2\sigma_{12}^{(0)}$$

and for the third term

$$\frac{\partial \log(\det(P))}{\partial p_{11}} = \frac{p_{22}}{\det(P)}, \quad \frac{\partial \log(\det(P))}{\partial p_{22}} = \frac{p_{11}}{\det(P)}, \quad \frac{\partial \log(\det(P))}{\partial p_{12}} = \frac{-2p_{12}}{\det(P)}.$$

Thus we get

$$\begin{aligned}
&\frac{\partial}{\partial P}\left(N \operatorname{Tr}\left(\Sigma^{(0)} P\right) - N \log(\det(P))\right) \\
&= N\left(2\Sigma^{(0)} - \operatorname{diag}(\Sigma^{(0)})\right) - N\left(2P^{-1} - \operatorname{diag}\left((2P^{-1})\right)\right).
\end{aligned} \tag{37}$$

For the second term, we calculate the partial derivatives. For this purpose, we first consider the following auxiliary calculations

$$\begin{aligned}
0 &= \frac{\partial}{\partial p_{ij}}\left(\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_P \kappa\right)\right) \\
&= \left(\frac{\partial}{\partial p_{ij}}\left(\kappa \diamond_P \kappa\right)^{-1}\right)\left(\kappa \diamond_P \kappa\right) + \left(\kappa \diamond_P \kappa\right)^{-1}\left(\frac{\partial}{\partial p_{ij}}\left(\kappa \diamond_P \kappa\right)\right)
\end{aligned}$$

and therefore

$$\frac{\partial \left(\kappa \diamond_P \kappa\right)^{-1}}{\partial p_{ij}} = -\left(\kappa \diamond_P \kappa\right)^{-1}\left(\kappa \diamond_{\frac{\partial P}{\partial p_{ij}}} \kappa\right)\left(\kappa \diamond_P \kappa\right)^{-1}. \tag{38}$$

We also calculate

$$\frac{\partial \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)}{\partial p_{ij}} = \left(\kappa \diamond_{\left(\frac{\partial P}{\partial p_{ij}}\right)\Sigma^{(0)}P} \kappa\right) + \left(\kappa \diamond_{P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right)} \kappa\right). \tag{39}$$

By utilizing equations (38) and (39) and Lemma E.2, we can deduce that

$$\begin{aligned}
&\frac{\partial}{\partial p_{ij}} \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right) \\
&\quad = \operatorname{Tr}\Big(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{\left(\frac{\partial P}{\partial p_{ij}}\right)\Sigma^{(0)}P} \kappa \right) + \left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}\left(\frac{\partial P}{\partial p_{ij}}\right)} \kappa \right) \\
&\qquad - \left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{\frac{\partial P}{\partial p_{ij}}} \kappa \right) \left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) \Big) \\
&\quad = \operatorname{Tr}\Bigg(\Sigma^{(0)} P \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \left(\frac{\partial P}{\partial p_{ij}} \right) + \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) P\Sigma^{(0)} \left(\frac{\partial P}{\partial p_{ij}} \right) \\
&\qquad - \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) (\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \left(\frac{\partial P}{\partial p_{ij}} \right) \Bigg).
\end{aligned} \tag{40}$$

We obtain from (40) and Lemma E.3

$$\begin{aligned}
&\frac{\partial}{\partial P} \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right) \\
&= 2 \left(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) P\Sigma^{(0)} + \Sigma^{(0)} P \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) - \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) (\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \right) \\
&- \operatorname{diag}\Big(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) P\Sigma^{(0)} + \Sigma^{(0)} P \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \Big) \\
&- \operatorname{diag}\Big(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) (\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \Big).
\end{aligned}$$

Using (37) we get the desired result

$$\begin{aligned}
&\frac{\partial}{\partial P} \left(N \operatorname{Tr}\left(\Sigma^{(0)} P\right) - \operatorname{Tr}\left(\left(\kappa \diamond_P \kappa\right)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right)\right) - N \log(\det(P)) \right) \\
&= 2 \left(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) P\Sigma^{(0)} + \Sigma^{(0)} P \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) - \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) (\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \right) \\
&- \operatorname{diag}\Big(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) P\Sigma^{(0)} + \Sigma^{(0)} P \left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \Big) \\
&- \operatorname{diag}\Big(\left(\bar{\kappa} \diamond_{(\kappa \diamond_P \kappa)^{-1} \left(\kappa \diamond_{P\Sigma^{(0)}P} \kappa\right) (\kappa \diamond_P \kappa)^{-1}} \bar{\kappa} \right) \Big) \\
&+ N \left(2\Sigma^{(0)} - \operatorname{diag}(\Sigma^{(0)}) \right) - N \left(2P^{-1} - \operatorname{diag}\left((2P^{-1}) \right) \right).
\end{aligned}$$

□

Lemma E.5. *For* $([\kappa], P) \in \mathfrak{P}$, $\kappa \in [\kappa]$ *and* $P^{(0)}, \Sigma^{(0)} \in \mathrm{SPD}(2)$ *with* $P^{(0)} = \left(\Sigma^{(0)}\right)^{-1}$ *we obtain*

$$\begin{aligned}
\frac{\partial}{\partial P} \mathcal{F}([\kappa], P)\Big|_{[\kappa]=\left[\kappa^{(0)}\right], P=P^{(0)}} &= 2 \left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)} \diamond_{P^{(0)}} \kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)} \right) \\
&\quad - \operatorname{diag}\left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)} \diamond_{P^{(0)}} \kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)} \right).
\end{aligned}$$

Proof. Analogous to Section 5 in the main text, we decompose $\mathcal{F}$ as follows

$$\begin{aligned}
&\frac{\partial}{\partial P}\mathcal{F}([\kappa], P)\Big|_{[\kappa]=[\kappa^{(0)}], P=P^{(0)}} \\
&= \frac{\partial}{\partial P}\int d_P\left(Y, \hat{\phi}(\kappa, P, Y)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\phi\right)\Big|_{\kappa=\kappa^{(0)}, P=P^{(0)}} \\
&+ \frac{\partial}{\partial P}\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\epsilon\right) - N\log(\det(P))\Big|_{\kappa=\kappa^{(0)}, P=P^{(0)}}.
\end{aligned} \tag{41}$$

From Theorem 4.5 in the main text, it follows that all $\kappa \in [\kappa^{(0)}]$ minimize the expression

$$\int d_P\left(Y, \hat{\phi}(\kappa, P, Y)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\phi\right)$$

for any $P \in \mathrm{SPD}(2)$. Consequently,

$$\frac{\partial}{\partial P}\int d_P\left(Y, \hat{\phi}(\kappa, P, Y)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\phi\right)\Big|_{\kappa=\kappa^{(0)}, P=P^{(0)}} = 0.$$

We utilize Lemma E.1 and Lemma E.4 for the second part of (41)

$$\begin{aligned}
&\frac{\partial}{\partial P}\int d_P\left(\epsilon, \hat{\phi}(\kappa, P, \epsilon)\kappa\right)^2 \mathrm{d}\mathbb{P}\left(\epsilon\right) - N\log(\det(P))\Big|_{\kappa=\kappa^{(0)}, P=P^{(0)}} \\
&= 2\left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)}\diamond_{P^{(0)}}\kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)}\right) - \mathrm{diag}\left(\bar{\kappa}^{(0)} \diamond_{\left(\kappa^{(0)}\diamond_{P^{(0)}}\kappa^{(0)}\right)^{-1}} \bar{\kappa}^{(0)}\right).
\end{aligned}$$

□

F Heteroscedastic Drift Model

F.1 Goodness of Fit and Standard Deviations

Orientation	Real	Imaginary
x	1.13×10^{-5}	4.19×10^{-6}
xy	2.08×10^{-8}	0.929
y	2.71×10^{-4}	0.262
yz	2.54×10^{-10}	5.83×10^{-20}
z	0.173	3.53×10^{-2}

Table C3: Results of Kolmogorov–Smirnov tests for Gaussianity applied to the real and imaginary parts of the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu$, pooled over b and ν, obtained from the homoscedastic drift model applied to the 94 GHz data.

Orientation	heteroscedastic drift model	averaging model
x	4.2×10^{-3}	9.2×10^{-3}
xy	2.6×10^{-3}	3.1×10^{-3}
y	2.9×10^{-3}	4.4×10^{-3}
yz	3.9×10^{-3}	9.0×10^{-3}
z	9.3×10^{-3}	1.2×10^{-2}

Table C4: The standard deviation of the flat regions shown in Panel B of Figure C10, computed for both the averaging model and the heteroscedastic drift model.

orientation	$p_{\Re}$	$p_{\Im}$
x	0.827	0.321
xy	1.11×10^{-3}	0.984
y	0.0294	0.587
yz	0.269	0.253
z	0.755	0.889

Table C5: p-values from applying the heteroscedastic drift model to the 94 GHz data.

F.2 Boundary Maxima in the Heteroscedastic Drift Model

The heteroscedastic drift model exhibits boundary maxima as Σ_0 tends to a rank-deficient matrix. A detailed example exhibiting these is given in Lemma F.1. The iterative Algorithm 2 fitting the above model did not find these boundary maxima when initialized from the homoscedastic drift model on the tested datasets. From the log likelihood values resulting from these fits, we looked at the upper bound for the minimal eigenvalue of Σ_0 for which these log likelihood values can be obtained by the parameter sequence constructed in Lemma F.1. These are reported in Table C6 and compared with the minimal eigenvalues of the estimated $\hat{\Sigma_0}$. From the differences, which are about 200 orders of magnitude, we concluded that the algorithm did indeed not find the boundary global maxima but found some local MLE. In practice, we did not actually need to restrict the parameter space for

Σ_0 to impose lower bounds on its eigenvalues, even though this would reasonably represent minimum receiver noise.

Lemma F.1. *The heteroscedastic drift model has boundary maxima.*

Proof. Let $Y_{b,\nu} \in \mathbb{C}$ be arbitrary. WLOG $Y \neq 0, \sum_{\nu=0}^{N} Y_{1,\nu} \neq 0$. And choose $\psi_1 = \frac{1}{N+1}\sum_{\nu=0}^{N} Y_{1,\nu}$. Then we can chose κ_ν such that the residuals $R_{1,\nu} = 0$ for all $\nu = 0, \ldots, N$ by just using the averaging model estimator applied to the first batch

$$\begin{aligned}
\kappa_\nu &= \frac{Y_{1,\nu} - \psi_1}{\sqrt{\sum_{\nu'=0}^{N} \left|Y_{1,\nu'} - \psi_1\right|^2}} \\
c &= 0 \\
\phi_1 &= \sqrt{\sum_{\nu'=0}^{N} \left|Y_{1,\nu'} - \psi_1\right|^2} \\
\Rightarrow R_{1,\nu} &= Y_{1,\nu} - \psi_1 - \phi_1\kappa_\nu = Y_{1,\nu} - \frac{1}{N+1}\sum_{\nu'=0}^{N} Y_{1,\nu'} - \left(Y_{1,\nu} - \frac{1}{N+1}\sum_{\nu'=0}^{N} Y_{1,\nu'}\right) = 0
\end{aligned}$$

We then choose a sequence $\Sigma_0^{(k)}, \tilde{\sigma}$ such that the likelihood diverges to $+\infty$ as $k \to \infty$. For notational convenience, we express all matrices in the basis $\left\{\mathrm{vec}\left(\frac{\psi_1}{|\psi_1|}\right), \mathrm{vec}\left(i\frac{\psi_1}{|\psi_1|}\right)\right\}$.

$$\begin{aligned}
\Sigma_0^{(k)} &= \frac{1}{k}\mathrm{vec}\left(\psi_1\right)\mathrm{vec}\left(\psi_1\right)^T + \mathrm{vec}\left(i\psi_1\right)\mathrm{vec}\left(i\psi_1\right)^T = |\psi_1|^2 \begin{pmatrix} \frac{1}{k} & 0 \\ 0 & 1 \end{pmatrix} \\
\tilde{\sigma} &= 1 \\
\Rightarrow \Sigma_1^{(k)} &= \Sigma_0^{(k)} + \tilde{\sigma}\mathrm{vec}\left(i\psi_1\right)\mathrm{vec}\left(i\psi_1\right)^T = |\psi_1|^2\left(\begin{pmatrix} \frac{1}{k} & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right) = |\psi_1|^2 \begin{pmatrix} \frac{1}{k} & 0 \\ 0 & 2 \end{pmatrix}
\end{aligned}$$

We first focus on the log likelihood contribution $\ell^{(k)}_{Y_{1,:}}$ associated with batch $b = 1$ and consider the remaining contributions later.

$$\begin{aligned}
\ell^{(k)}_{y_{1,:}} &= -\frac{1}{2}\left(\sum_{\nu=0}^{N}\left(\mathrm{vec}\left(R_{1,\nu}\right)^T P_1^{(k)}\mathrm{vec}\left(R_{1,\nu}\right)\right) + (N+1)\left(\log\left(\det\left(\Sigma_1^{(k)}\right)\right) + \log\left((2\pi)^2\right)\right)\right) \\
&= -\frac{N+1}{2}\left(\log\left(\det\left(\Sigma_1^{(k)}\right)\right) + \log\left((2\pi)^2\right)\right) = \frac{N+1}{2}\log\left(\frac{1}{\det\left(\Sigma_1^{(k)}\right)}\right) - (N+1)\log(2\pi) \\
&= \frac{N+1}{2}\log\left(\frac{k}{2|\psi_1|^4}\right) - (N+1)\log(2\pi) \\
&\Rightarrow \lim_{k\to\infty} \ell^{(k)}_{Y_{1,:}} = +\infty
\end{aligned}$$

We now choose $\psi_b^{(k)}$ for $b \neq 1$ such that $\Sigma_b^{(k)}$ is constant in k. So let b not equal to 1

$$\psi_b^{(k)} = -i\sqrt{1-\frac{1}{k}}\psi_1$$

$$\operatorname{vec}\left(i\psi_b^{(k)}\right)\operatorname{vec}\left(i\psi_b^{(k)}\right)^T = \left(1-\frac{1}{k}\right)\operatorname{vec}(\psi_1)\operatorname{vec}(\psi_1)^T = |\psi_1|^2\begin{pmatrix}1-\frac{1}{k} & 0\\ 0 & 0\end{pmatrix}$$

$$\Rightarrow \Sigma_b^{(k)} = \Sigma_0^{(k)} + \tilde{\sigma}\operatorname{vec}\left(i\psi_b^{(k)}\right)\operatorname{vec}\left(i\psi_b^{(k)}\right)^T = |\psi_1|^2\left(\begin{pmatrix}\frac{1}{k} & 0\\ 0 & 1\end{pmatrix} + \begin{pmatrix}1-\frac{1}{k} & 0\\ 0 & 0\end{pmatrix}\right)$$

$$= |\psi_1|^2\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

$$\Rightarrow P_b^{(k)} = |\psi_1|^{-2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

But then the only dependency on k for $b \neq 1$ in the likelihood is in the residuals. (We choose $\phi_b = -i\phi_1$ for convenience so $\phi_b\kappa_\nu = -i\left(Y_{1,\nu} - \frac{1}{N+1}\sum_{\nu'=0}^{N} Y_{1,\nu'}\right) = -i(Y_{1,\nu} - \psi_1)$)

$$\Rightarrow R_{b,\nu}^{(k)} = Y_{b,\nu} - \psi_b^{(k)} - \phi_b\kappa_\nu = Y_{b,\nu} + i\sqrt{1-\frac{1}{n}}\psi_1 + i(Y_{1,\nu} - \psi_1)$$

$$\Rightarrow \lim_{k\to\infty} R_{b,\nu}^{(k)} = Y_{b,\nu} + iY_{1,\nu} =: R_{b,\nu}$$

$$\Rightarrow \ell_{Y_{b,:}}^{(k)} = -\frac{1}{2}\left(\sum_{\nu=0}^{N}\left(\operatorname{vec}\left(R_{b,\nu}^{(k)}\right)^T P_b^{(k)}\operatorname{vec}\left(R_{b,\nu}^{(k)}\right)\right.\right.$$

$$\left.\left. + (N+1)\left(\log\left(\det\left(\mathbf{\Sigma}_b^{(k)}\right)\right) + \log\left((2\pi)^2\right)\right)\right)$$

$$= -\frac{1}{2}\left(\sum_{\nu=0}^{N}\frac{\left|R_{b,\nu}^{(k)}\right|^2}{|\psi_1|^2} + (N+1)\left(\log\left(|\psi_1|^4\right) + \log\left((2\pi)^2\right)\right)\right)$$

$$\Rightarrow \lim_{k\to\infty}\ell_{y_{b,:}}^{(k)} = -\frac{1}{2}\left(\sum_{\nu=0}^{N}\frac{|R_{b,\nu}|^2}{|\psi_1|^2} + (N+1)\left(\log\left(|\psi_1|^4\right) + \log\left((2\pi)^2\right)\right)\right)$$

So the log likelihood of the residuals for $b \neq 1$ does not diverge to $-\infty$ but instead converges to a finite value. But by design the log likelihood of the first batch diverges like $\log(k)$.

orientation	k*	smallest ev $\Sigma_0^{(k*)}$	smallest ev of $\widehat{\Sigma_0}$	$\widehat{\tilde{\sigma}}$
x	10^{271}	10^{-262}	10^{3}	1.73×10^{-4}
xy	10^{339}	10^{-330}	10^{3}	1.69×10^{-4}
y	10^{468}	10^{-460}	10^{3}	1.72×10^{-4}
yz	10^{316}	10^{-307}	10^{-2}	1.51×10^{-4}
z	10^{467}	10^{-459}	10^{4}	1.79×10^{-4}

Table C6: Applying the example of a parameter sequence with divergent log likelihood from Lemma F.1 to the datasets, we can compare the algorithmic fit with these parameters by focusing on the sequence index k^* for which the log likelihood of the fit is first reached by the sequence.

Therefore,

$$\begin{aligned}
&\lim_{k\to\infty} \ell_Y^{(k)} = \lim_{k\to\infty} \left(\ell_{Y_{1,:}}^{(k)} + \sum_{b=2}^{B} \ell_{Y_{b,:}}^{(k)} \right) \\
&= \lim_{k\to\infty} \left(\frac{N+1}{2} \log\left(\frac{k}{2|\psi_1|^4} \right) - (N+1)\log(2\pi) \right. \\
&\left. + \sum_{b=2}^{B} \left(-\frac{1}{2} \left(\sum_{\nu=0}^{N} \frac{\left| R_{b,\nu}^{(k)} \right|^2}{|\psi_1|^2} + (N+1)\Big(\log\Big(|\psi_1|^4 \Big) + \log\big((2\pi)^2 \big) \Big) \right) \right) \right) \qquad (42) \\
&= +\infty - \frac{1}{2} \left(\left(\sum_{b=2}^{B} \sum_{\nu=0}^{N} \frac{|R_{b,\nu}|^2}{|\psi_1|^2} \right) + (N+1)(B-1)\log\Big(|\psi_1|^4 \Big) + (N+1)B\log\big((2\pi)^2 \big) \right) \\
&= +\infty
\end{aligned}$$

□

F.3 Phase Noise Truncation

Looking at the mean and variance of the wrapped Gaussian

$$\begin{aligned}
\mathbb{E}\Big[\tilde{\psi}_{b,\nu} \Big] &= \psi_b \exp\left(-\frac{\tilde{\sigma}^2}{2} \right) = \psi_b \left(1 - \frac{\tilde{\sigma}^2}{2} \right) + \mathcal{O}(\tilde{\sigma}^4) \\
\mathrm{Var}\Big[\tilde{\psi}_{b,\nu} \Big] &:= \mathrm{Cov}\Big[\mathrm{vec}\left(\tilde{\psi}_{b,\nu} \right), \mathrm{vec}\left(\tilde{\psi}_{b,\nu} \right) \Big] = M(\psi_b) \begin{pmatrix} \frac{1+e^{-2\tilde{\sigma}^2}-2e^{-\tilde{\sigma}^2}}{2} & 0 \\ 0 & \frac{1-e^{-2\tilde{\sigma}^2}}{2} \end{pmatrix} M(\psi_b)^T \\
&= M(\psi_b) \begin{pmatrix} 0 + \mathcal{O}(\tilde{\sigma}^4) & 0 \\ 0 & \tilde{\sigma}^2 + \mathcal{O}(\tilde{\sigma}^4) \end{pmatrix} M(\psi_b)^T.
\end{aligned}$$

we see that the expansion of the mean to higher than linear order is not consistent with the mean of Definition 6.1 due to the correction in $\tilde{\sigma}^2$ which comes from the quadratic term $-\frac{\psi_b \tilde{\sigma}^2}{2} \varphi_{b,\nu}^2$. Replacing ψ_b by $\breve{\psi}_b = \psi_b \frac{2-\tilde{\sigma}^2}{2}$ in Definition 6.1 on the other hand, leads

to a different noise scale parameter $\sigma^2 = \left(\frac{2}{2-\tilde{\sigma}^2}\right)^2 \tilde{\sigma}^2$ as

$$\mathrm{vec}\left(\breve{\psi}_b\right) \sigma^2 \mathrm{vec}\left(\breve{\psi}_b\right)^T = \mathrm{vec}\left(\breve{\psi}_b\right) \left(\frac{2}{2-\tilde{\sigma}^2}\right)^2 \tilde{\sigma}^2 \mathrm{vec}\left(\breve{\psi}_b\right)^T =: \mathrm{vec}\left(\psi_b\right) \tilde{\sigma}^2 \mathrm{vec}\left(\psi_b\right)^T .$$

This second parametrization was used in the heteroscedastic drift model. As $\sigma^2 = \tilde{\sigma}^2 + \mathcal{O}(\tilde{\sigma}^4)$, the validity of an expansion to linear order in σ^2 is equivalent to one in $\tilde{\sigma}^2$.

The next term in the expansion of $\tilde{\psi}_{b,\nu} = \psi_b \exp\{i\tilde{\sigma}\varphi_{b,\nu}\}$ not modeled is the quadratic term. The variance contribution of this term is

$$\mathrm{Var}\left[-\mathrm{vec}\left(\psi_b\right) \frac{\tilde{\sigma}^2}{2} \varphi_{b,\nu}^2\right] = \mathrm{vec}\left(\frac{\psi_b}{|\psi_b|}\right) \frac{|\psi_b|^2 \tilde{\sigma}^4}{2} \mathrm{vec}\left(\frac{\psi_b}{|\psi_b|}\right)^T .$$

Given our data, when calculated based on the MLE estimators for ψ_b and $\tilde{\sigma}^2$, this is dominated by the marginal variance of $\hat{\Sigma}_0$ in the subspace spanned by $\mathrm{vec}\left(\hat{\psi}_b\right)$ given by $\left\| \frac{\mathrm{vec}\left(\hat{\psi}_b\right)}{|\hat{\psi}_b|} \right\|_{\hat{\Sigma}_0}$ justifying the truncation. Even when minimizing this comparison over the batch parameter independently, the marginal variance is still larger by 2 orders of magnitude as reported in Table C7. Based on this, explicit modelling of the quadratic term was deemed unnecessary.

orientation	$\min_{b\in B} \mathrm{vec}\left(\frac{\psi_b}{\lvert\psi_b\rvert}\right)^T \Sigma_0 \mathrm{vec}\left(\frac{\psi_b}{\lvert\psi_b\rvert}\right)$	$\max_{b\in B} \frac{\lvert\psi_b\rvert^2\sigma^4}{2}$
x	8.0×10^3	2.6×10^1
xy	3.9×10^3	2.3×10^1
y	2.8×10^3	1.5×10^1
yz	2.4×10^3	8.4
z	1.7×10^3	8.1

Table C7: Comparison of the contribution of Σ_0 in the direction ψ_b (minimized over the batches) with the maximal contribution of $\frac{\sigma^4}{2}|\psi_b|^2$ (maximized over the batches) in the heteroscedastic drift model. Noise contributions from the quadratic order term in the wrapped Gaussian expansion for the phase noise are at least two orders of magnitude smaller than those of Σ_0.

F.4 Algorithm

We included the update step $\tilde{\psi} = \psi - \Delta_c \phi$, $\tilde{c} = c + \Delta_c$ for a numerically optimized value of Δ_c in the optimizer in order to improve convergence properties. It does not change the residuals but only the covariance matrix. Without it, the log likelihood improvements stagnate. Adding this update from the beginning led to unstable trajectories of $\hat{c}$ over the iterations, so the algorithm we used starts this additional update after the 25th iteration.

The initialization of Σ_0 and $\tilde{\sigma}$ is done by regressing the matrices

$$\mathrm{vec}\left(i\hat{\psi}_{bhom}\right) \mathrm{vec}\left(i\hat{\psi}_{bhom}\right)^T ,$$

which are obtained from the residuals arising from fitting the homoscedastic drift model, onto the sample covariance matrix of the homoscedastic drift model. The intercept is taken as an initial value for Σ_0 and the slope initializes $\tilde{\sigma}$.

The full algorithm is given in Algorithm 2.

Algorithm 2: Heteroscedastic drift model MLE

Load $\boldsymbol{y}$

$\widehat{\boldsymbol{\psi}}^{(0)}, \widehat{\boldsymbol{\phi}}^{(0)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(0)}, \widehat{\boldsymbol{\Sigma}}_{hom} \leftarrow Algorithm\ 1(\boldsymbol{y})$

$R \leftarrow \text{vec}\left(y - \hat{\psi}^{(0)}(\mathbf{1}_{N+1})^T - \hat{\phi}^{(0)}(\hat{\tilde{\kappa}}^{(0)})^T\right)$

$\hat{\Psi}, \hat{S} \leftarrow \text{vec}\left(i\hat{\psi}^{(0)}\right)\text{vec}\left(i\hat{\psi}^{(0)}\right)^T, \frac{1}{N+1}\sum_{\nu=0}^{N} R_{:,\nu}R_{:,\nu}^T$

$\widehat{\tilde{\sigma}}^{(0)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(0)} \leftarrow LinReg\left(\hat{\Psi}, \hat{S}\right)$

$k \leftarrow 0$

while $k \leq maxiter = 200$ **do**

  $\ell^{(k)} \leftarrow \ell\left(\boldsymbol{y}, \widehat{\boldsymbol{\psi}}^{(k)}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k)}, \widehat{\tilde{\sigma}}^{(k)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k)}\right)$

  if $k > 0$ **then**

    if $\ell^{(k)} - \ell^{(k-1)} < min\_delta\_loglik = 10^{-4}$ **then**

      break

    end if

  end if

  $\widehat{\tilde{\sigma}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k+1)} \leftarrow L-BFGS-B(x_0 = (\widehat{\tilde{\sigma}}^{(k)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k)}), func = \ell_{\boldsymbol{y}}^{\widehat{\boldsymbol{\psi}}^{(k)}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k)}},$
  $jac = D\ell_{\boldsymbol{y}}^{\widehat{\boldsymbol{\psi}}^{(k)}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k)}})$

  $\widehat{\boldsymbol{\phi}}^{(k+1)} \leftarrow \widehat{\boldsymbol{\phi}}\left(\boldsymbol{y}, \widehat{\boldsymbol{\psi}}^{(k)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k)}, \widehat{\tilde{\sigma}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k+1)}\right)$

  $\widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)} \leftarrow \widehat{\tilde{\boldsymbol{\kappa}}}\left(\boldsymbol{y}, \widehat{\boldsymbol{\psi}}^{(k)}\widehat{\boldsymbol{\phi}}^{(k+1)}, \widehat{\tilde{\sigma}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k+1)}\right)$

  $\hat{r} \leftarrow \left\| \hat{\tilde{\boldsymbol{\kappa}}}^{(k+1)} - \frac{1}{N+1}\sum_{\nu=0}^{N} \widehat{\tilde{\kappa}}_{\nu}^{(k+1)} \right\|$

  $\widehat{\boldsymbol{\phi}}^{(k+1)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)} \leftarrow \hat{r}\widehat{\boldsymbol{\phi}}^{(k+1)}, \frac{\widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)}}{\hat{r}}$

  $\widehat{\boldsymbol{\psi}}^{(k+1)} \leftarrow Nelder-Mead\left(x_0 = \widehat{\boldsymbol{\psi}}^{(k)}, func = \ell_{\boldsymbol{y}}^{\widehat{\boldsymbol{\phi}}^{(k+1)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)}, \widehat{\tilde{\sigma}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}_{\mathbf{0}}}^{(k+1)}}\right)$

  if $k \geq start\_c\_opt = 25$ **then**

    $\Delta_c \leftarrow Nelder-Mead\left(x_0 = 0, func = \ell_{\boldsymbol{y}}^{\widehat{\boldsymbol{\theta}_c}}\right)$

    $\widehat{\boldsymbol{\psi}}^{(k+1)}, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)} \leftarrow \widehat{\boldsymbol{\psi}}^{(k+1)} - \widehat{\boldsymbol{\phi}}^{(k+1)}\Delta_c, \widehat{\tilde{\boldsymbol{\kappa}}}^{(k+1)} + \Delta_c\mathbf{1}_{N+1}$

  end if

  $k \leftarrow k + 1$

end while

$\hat{c}, \widehat{\boldsymbol{\kappa}} \leftarrow \frac{1}{N+1}\sum_{\nu=0}^{N} \widehat{\tilde{\kappa}}_{\nu}^{(k)}, \left(\widehat{\tilde{\boldsymbol{\kappa}}}^{(k)} - \hat{c}\mathbf{1}_{N+1}\right)$

return $\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\boldsymbol{\kappa}}, \hat{c}, \widehat{\tilde{\sigma}}, \widehat{\boldsymbol{\Sigma}_0}^{(k)}$

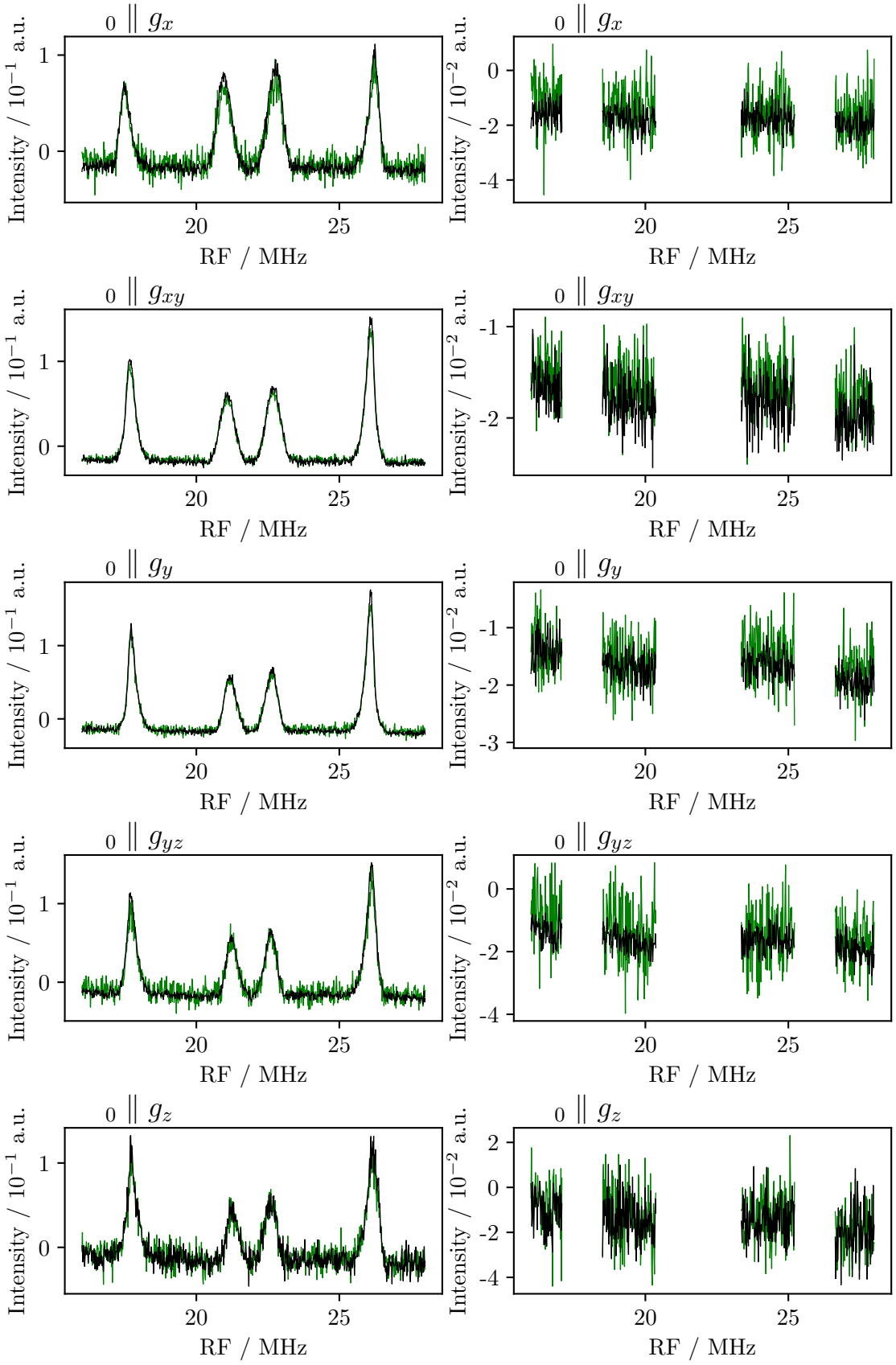


Figure C10: Comparison of the averaging model (green) and the heteroscedastic drift model (black).

CHAPTER D

Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra

Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra

H. Wiechers[1], A. Kehl[2], M. Hiller[2], B. Eltzner[2], S. F. Huckemann[1], A. Meyer[2], I. Tkach[2], M. Bennati[2,3,*] and Y. Pokern[4,*]

[1]Felix-Bernstein-Institute for Mathematical Statistics, Georg-August-University Göttingen, 37077 Göttingen, Germany.
[2]Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany.
[3]Department of Chemistry, Georg-August University of Göttingen, Tammannstr. 2, Göttingen, Germany
[4]Department of Statistical Science, University College London, London WC1E 6BT, United Kingdom.
[*]Corresponding authors

ENDOR spectroscopy is a fundamental method to detect nuclear spins in the vicinity of paramagnetic centers and their mutual hyperfine interaction. Recently, site-selective introduction of ^{19}F as nuclear labels has been proposed as a tool for ENDOR-based distance determination in biomolecules, complementing pulsed dipolar spectroscopy in the range of angstrom to nanometer. Nevertheless, one main challenge of ENDOR still consists of its spectral analysis, which is aggravated by a large parameter space and broad resonances of hyperfine tensors. Additionally, at high EPR frequencies and fields ($\geq$ 94 GHz/3.4 Tesla), chemical shift anisotropy might contribute to broadening and asymmetry in the spectra. Here, we use two nitroxide-fluorine model systems to examine a statistical approach to finding the best parameter fit to experimental 263 GHz ^{19}F ENDOR spectra. We propose Bayesian optimization for a rapid, global parameter search with little prior knowledge, followed by a refinement by more standard gradient-based fitting procedures. Indeed, the latter suffer from finding local rather than global minima of a suitably defined loss function. Using a new and accelerated simulation procedure, results for the semi-rigid nitroxide-fluorine two and three spin systems lead to physically reasonable solutions, if minima of similar loss can be distinguished by DFT predictions. The approach also delivers the stochastic error of the obtained parameter estimates. Future developments and perspectives are discussed.

KEYWORDS: EPR, electron nuclear double resonance, EPR, fluorine labelling, least-squares fitting, Bayesian optimization, spectral simulation

1 Introduction

Electron-nuclear double resonance (ENDOR) spectroscopy measures hyperfine (HF) couplings between a paramagnetic center and nuclear spins. Since the early introduction of the two main pulse sequences, Davies and Mims [1, 2], ENDOR has been extensively used in combination with nuclear isotope labelling to map electron spin density distributions [3–6] and to study the active site of biomolecules and materials [7–15].

Recently, the introduction of fluorine labels has provided an additional opportunity to employ ENDOR for distance measurements in structural biology. The approach exploits some unique properties of the ^{19}F nucleus [16], i.e. its nuclear spin $I = \frac{1}{2}$ in combination with the large gyromagnetic ratio, providing relatively simple ENDOR spectra. The spectra are dominated by dipolar interaction, as long as the ^{19}F nucleus is sufficiently far from the electron spin such that no effective spin density transfer mechanism is operative. Analysis of the spectra reveals the electron spin-fluorine dipolar tensor, from which the interspin distance can be extracted [16]. In the last year, the method has been extended in combination with other paramagnetic labels (trityl, Gd^{3+}) [17, 18] as well as with endogenous tyrosyl radicals [19], and distances in the range between 0.5 and 2 nm have been reported so far. As an attractive feature, ENDOR samples can also be investigated by paramagnetic NMR techniques [20], for instance paramagnetic relaxation enhancements (PRE) [21] or pseudocontact shifts [22], which opens avenues to an integrative approach for structural biology studies. Very recently, this approach has been reported for investigations of proteins in cell [23].

Similarly to NMR, ENDOR spectroscopy benefits from high magnetic fields and frequencies, as nuclear Larmor frequencies become naturally separated, and consequently HF powder patterns arising from different types of nuclei can be better resolved. This is particularly relevant for studies with ^{19}F, which has a gyromagnetic ratio very close to that of protons. For instance, at 34 GHz/1.2 Tesla (Q-band) the ^{19}F and ^{1}H Larmor frequencies are separated by only ca. 3 MHz, which means that ^{1}H/^{19}F overlap will occur in typical nitroxides featuring proton HF coupling constants on the order of 6 MHz [24]. Even 94 GHz/3.4 Tesla (W-band) can present severe complications with proton background subtraction, for example if using tyrosyl radicals as paramagnetic centers [19]. Thus, exploration of ENDOR at even higher frequencies (263 GHz/9.4 Tesla), although instrumentally demanding [25], becomes crucial for future developments.

Alongside their mentioned advantages, ^{19}F ENDOR spectra at high frequency come with complications in the analysis due to two factors: (i) a strong orientation selection that usually prevents immediate read-out of couplings from peak positions, and (ii) the emerging resolution of chemical shielding (CS) anisotropy. Recently, we have reported 263 GHz ^{19}F Mims ENDOR spectra of nitroxide-fluorine model systems and demonstrated an unprecedented, visible asymmetry arising from CS anisotropy [26]. This latter interaction contributes six additional parameters (three tensor eigenvalues plus three Euler angles) per ^{19}F nucleus in ENDOR spectral simulations, rendering standard estimation procedures based on least-square fitting with gradient methods unreliable. In our previous work, spectral simulations were achieved by using a fully DFT-predicted parameter set as input and subsequent, minor manual adjustment. The study provided the motivation to search for more rigorous methods of parameter estimation.

Recently, Stoll *et al.* have presented an example for inferring information about Fermi contact interaction as well as electron-nuclear distances from ENDOR spectra, including estimation of their uncertainties and inference on distributions [27]. The method demonstrates that multiple HF couplings can be extracted from ENDOR spectra if a Bayesian prior distribution based on DFT calculations is considered.

Herein, we follow an alternative route for determining the best parameter set from 263 GHz ENDOR spectra. We employ two representative model fluorine-nitroxide compounds that were investigated in [26], with one and two fluorine nuclei, respectively, see Figure D1. We first neglect

distance distributions, an approximation which turned out acceptable for the investigated semi-rigid model systems, but will be considered in more detail in future work. We combine statistical spectral uncertainties, recently made available through a *statistical drift model* [28, 29] called here SDM, with an accelerated *simulation code* (SimSpec) that explicitly calculates the effect of orientation selection. Since the parameter space associated with the analysis is of moderately high dimension (typically 10 dimensions for a single ^{19}F nucleus) and objective functions typically possess many local minima, we employ Bayesian optimization for determining the set of interaction parameters that provides the best fit to several ENDOR spectra simultaneously. Using SimSpec, we then compute how the simulated spectra change when we vary the interaction parameters. Combining the statistical uncertainty of the spectrum with the dependence of the spectrum on the interaction parameters in turn yields stochastic uncertainties of the interaction parameters. The

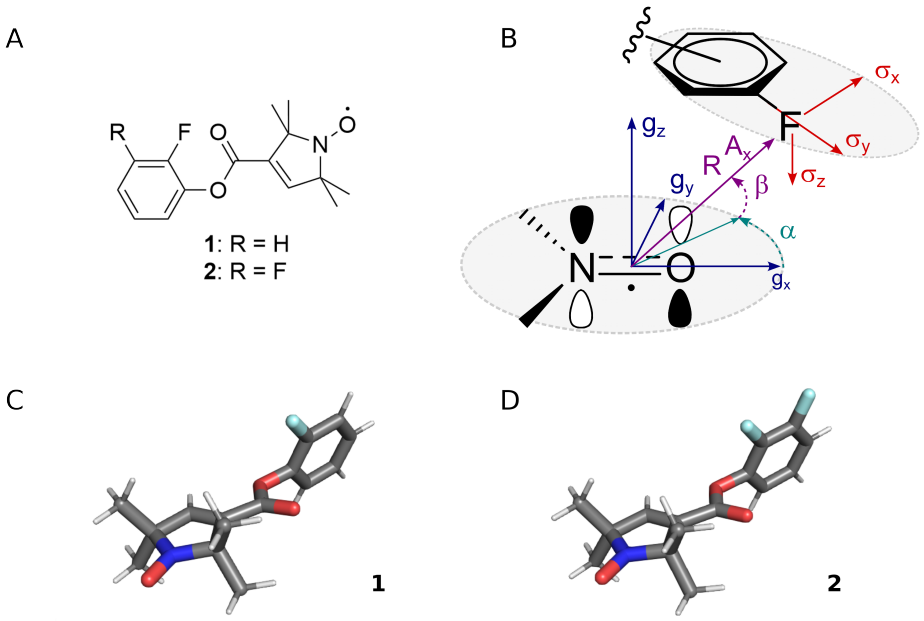


Figure D1: A: Chemical structure of compounds **1** and **2**. B: Visualization of the expected orientation of the g-, HF (A) and CS (σ) tensors with respect to the chemical structure, see [16]. C, D: Energy minimized structure predicted by DFT calculations of compounds **1** (C, [16]) and **2** (D, [26]). Those DFT calculations also predicted that these compounds assume only one predominant conformation.

paper is organized as follows: in Section 2 we experimentally determine the g-values of the two investigated compounds **1** and **2** from EPR spectra and then apply the SDM to ^{19}F Mims ENDOR data to remove baseline and other experimental artefacts. Subsequently, in Section 3, we describe the spin and experimental parameters required in the optimization procedure as well as the accelerated spectral simulation algorithm. In Section 4 we set out the statistical inference methodology. We then report and discuss results obtained using the proposed methodology in Section 5 and provide further details on materials and methods in Section 6.

2 Experiments and Data Processing

Estimation of HF and CS tensors from ENDOR data requires a work-flow that starts with the generation and examination of experimental data. Two types of experimental data are used here: (i) EPR spectra to characterize the nitroxide radical, *i.e.* the g- and the ^{14}N hyperfine coupling tensors, required to simulate orientation selection for ENDOR, and (ii) ENDOR spectra that are free from background signals and other experimental artefacts as the latter considerably affect the results of the optimization procedure. Such ENDOR spectra along with their uncertainties are extracted from recorded ENDOR data using the recently developed SDM [28], here described for ^{19}F-ENDOR.

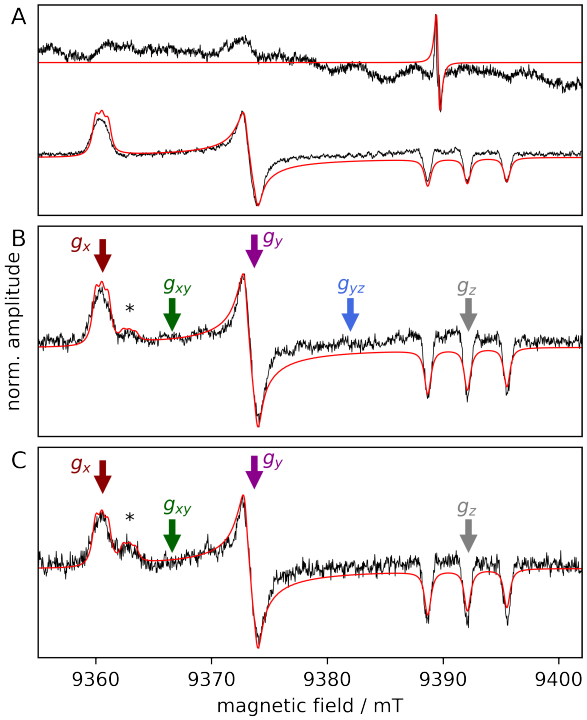


Figure D2: 263 GHz EPR spectra (black) and their simulations (red). A: compound **1** in a frozen solution containing an internal C-fibre standard measured both with CW-EPR (top), and ESE (bottom). The C-fibre spectrum is visible only in CW-EPR, whereas compound **1** can be better detected with ESE. B, C: ESE spectrum of compounds **1** and **2**, respectively, with an asterisk indicating the weaker contribution with a different g_x-value. In the case of the ESE experiments, the first derivatives of the smoothed echo-detected spectra are displayed.

2.1 EPR Spectra

For optimization to succeed, we aimed to reduce the number of parameters that needed to be estimated. To this end, we determined the g-values of the system in a measurement using a carbon (C) fibre ($g = 2.002644$, [30]) as internal reference standard. The 263 GHz continuous wave (CW-) EPR spectrum of the fibre was recorded directly after measuring the echo-detected EPR (ESE) spectrum of compound **1** and is displayed in Figure D2 A. To calibrate the magnetic field strength, a simulated EPR spectrum of the C fibre was used. Based on this calibration, the g-values of compound **1** were obtained from a simulation. The HF and quadrupole interaction parameters for the nitroxide's ^{14}N nucleus were adopted from our previous report [26] with

a minor modification of the HF tensor $\boldsymbol{A}_{^{14}\mathrm{N}}$ eigenvalues to $[15, 11, 95.8]$ MHz to improve the fit. Moreover, we used eigenvalues of $[1.3, 0.5, -1.8]$ MHz for the quadrupolar tensor $\boldsymbol{P}_{^{14}\mathrm{N}}$ and Euler angles $[0, 0, 0]$ relative to the $\boldsymbol{g}$ tensor for both $\boldsymbol{A}_{^{14}\mathrm{N}}$ and $\boldsymbol{P}_{^{14}\mathrm{N}}$. This simulation yielded $g_{x,y,z} = [2.00886, 2.00610, 2.00211]$.
The EPR spectra of compounds **1** and **2** showed a second, weaker contribution with a smaller g_x-value (see Figure D2 B and C, marked with an asterisk). This was found dependent on the freezing conditions and is attributed to a fraction of the sample with a different H-bonding environment of the nitroxide [31]. For the simulation of the second contribution,

$$g_{x,y,z} = [2.00835, 2.00610, 2.00211]$$

and $\boldsymbol{A}_{^{14}\mathrm{N}}$ eigenvalues of $[15, 11, 95.8]$ MHz were used (relative weight 0.15 for compound **1** and 0.25 for compound **2**, inferred from relative g_x peak heights in the measured EPR spectra). The influence of this second contribution on the analysis is discussed in Section 5.

2.2 ENDOR Spectra and Data Processing with the Statistical Drift Model (SDM)

^{19}F 263 GHz Mims ENDOR spectra of compounds **1** and **2** were recorded at orientations of the g tensor as indicated in Figure D2 B and C. We used quadrature detection yielding an in phase and an orthogonal component, referred to as real ($\Re(Y)$) and imaginary ($\Im(Y)$) parts which, as for complex numbers generally, we equivalently write as two-dimensional vectors $\begin{pmatrix}\Re(Y)\\ \Im(Y)\end{pmatrix}$. Spectra of compound **1** were adopted from [26] while spectra of compound **2** were recorded again to obtain a better signal to noise (S/N-) ratio as compared to [26]. Experimental details are given in Section 6.
The spectra were processed with the SDM of [28] whose validity is established here for ^{19}F ENDOR at 263 GHz. The SDM allows quantification of the noise in the spectra as well as compensation of possible phase drifts of the echo signal that maximizes S/N-ratio.
In order to apply the SDM, the ENDOR data are recorded and stored in so-called batches indexed by $b \in \{1, \ldots, B\}$ in a data matrix $Y \in \mathbb{C}^{B\times N}$ with entries $Y_{b,\nu} \in \mathbb{C}$, where $\nu \in \{1, \ldots, N\}$ enumerates the RF frequencies. These data are then modelled as consisting of an offset $\psi_b \in \mathbb{C}$ (mostly the EPR echo but also including a dc offset) which may drift over time, an ENDOR-component $\phi_b\kappa_\nu$ (with $\phi_b, \kappa_\nu \in \mathbb{C}$) and measurement noise $\epsilon_{b,\nu}$ according to

$$Y_{b,\nu} = \psi_b + \phi_b\kappa_\nu + \epsilon_{b,\nu} \qquad \epsilon_{b,\nu} \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma). \tag{1}$$

Here, $i.i.d.$ denotes that the $\epsilon_{b,\nu}$ are independent and all follow the same Gaussian distribution (with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{2\times 2}$), *i.e.* we use additive Gaussian white noise. In Equation (1), ϕ_b models the batch-dependent strength and phase of the ENDOR effect and κ_ν captures the ENDOR spectrum. Applying our procedure delivers maximum likelihood estimates $\hat{\psi}$, $\hat{\phi}$, $\hat{\kappa}$ and $\hat{\Sigma}$ as described in [28], where the hat symbol denotes an estimator. To identify the direction in the complex plane along which $\hat{\kappa}$ contains the spectrum, we select λ_{opt} such that $\sum_\nu \Re(e^{i\lambda_{\mathrm{opt}}}\hat{\kappa}_\nu)^2$ is maximal and then consider $\hat{I}_\nu = \Re(e^{i\lambda_{\mathrm{opt}}}\hat{\kappa}_\nu)$ the measured spectrum, see panel A of Figure D3 and Supplementary Information (SI) A.1. The component orthogonal to the spectrum is shown in panel B of Figure D3. This figure also provides $\hat{\phi}$ and $\hat{\psi}$ in panels C and D.
The real and imaginary parts of the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu$ (see panels E,F and G of Figure D3) are examined for goodness of fit using a Kolmogorov-Smirnov (KS) test [32]. The resulting $p-$values are available in SI A.2 and give no concern over a lack of model fit. Therefore, the SDM is found to also fit ^{19}F 263 GHz ENDOR data.
In order to obtain a confidence region for the estimated spectrum, we employ the bootstrap procedure. This consists of the repeated generation of synthetic data from the estimated spectrum via

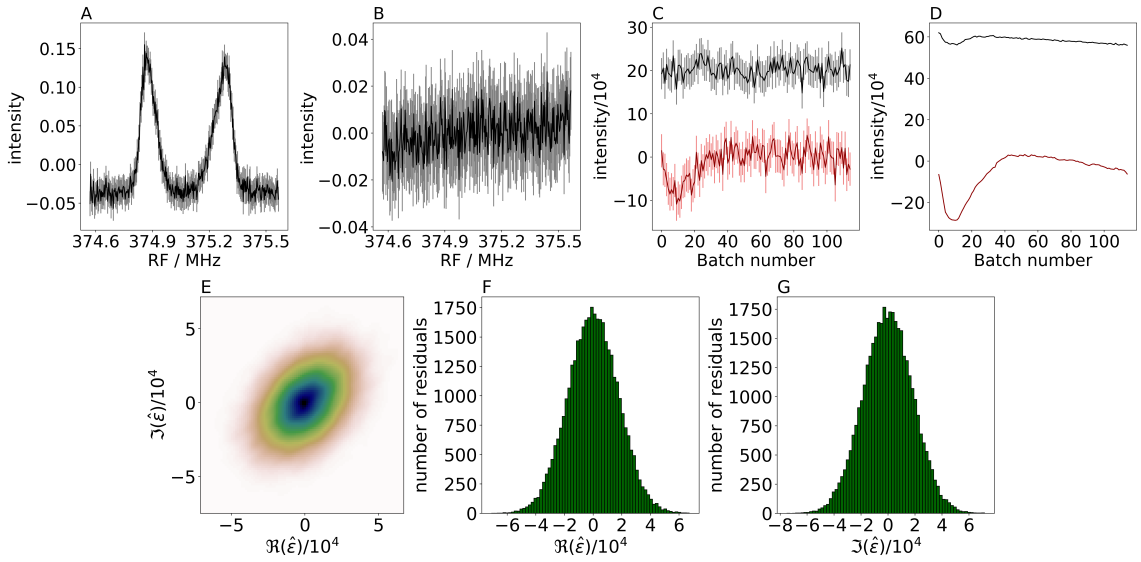


Figure D3: Representative data processing by the SDM for orientation g_x of compound **1**. A: the estimated spectrum $\hat{I}$. B: the component $\hat{\omega}$ that is orthogonal to the estimated spectrum $\hat{I}$ and contains no ENDOR signal, as expected. C, D: the real (black) and imaginary (red) parts of $\hat{\phi}$ and $\hat{\psi}$, respectively. A small phase and baseline drift is visible, particularly in the imaginary component. In A-D, 95% approximate pointwise confidence intervals are indicated as shaded regions; in D, these are so small as to be invisible. E: Kernel-density-estimation of the complex residuals $\hat{\epsilon}_{b,\nu}$. F, G: histograms for the real and imaginary parts of the residuals, respectively.

adding simulated noise, followed by estimation of the spectrum implied by these synthetic data. The variability of the spectra thus obtained indicates the stochastic error of the spectrum. In detail, following [28], bias-corrected estimates $\check{\phi}$ and $\check{\Sigma}$ of ϕ and Σ are used to generate bootstrap samples (denoted by the superscript *) of the error, $\epsilon^*_{b,\nu} \overset{i.i.d.}{\sim} \mathcal{N}(0, \check{\Sigma})$, and hence the synthetic data $Y^*_{b,\nu} = \hat{\psi} + \check{\phi}_b \hat{\kappa}_\nu + \epsilon^*_{b,\nu}$. From J independent samples of these synthetic data, maximum likelihood estimates $\psi^{*,j}, \phi^{*,j}, \kappa^{*,j}_\nu$ and hence $I^{*,j}_\nu$ of ψ, ϕ, κ and I respectively, indexed by $j \in \{1, \ldots, J\}$, are obtained. Their standard deviation is used to obtain approximate 95% confidence intervals displayed as shaded regions in panels A, B, C and D of Figure D3. The resulting ENDOR spectra and their uncertainties for all orientations and compounds are shown in Figure D4. Here, we also compare these spectra with the spectra obtained through the standard averaging method, *i.e.* summing of $Y_{b,\nu}$ over batches followed by normalization and phasing. We note that, in contrast with [28, 29] which tackled ^{1}H ENDOR, there is little difference between the spectra resulting from these two methods because there is very little phase drift in ϕ_b. The advantage of the SDM is that approximate 95% confidence regions naturally arise from the model. If an SDM is not available, it is possible to extract an indication of the stochastic error from spectra obtained by the averaging method. This can proceed via comparing the measured spectrum with a smoothed version, as in the *quasi-bootstrap* method in SI B. We used the quasi-bootstrap method to compute approximate 95% confidence regions for those spectra in Figure D4 that result from the averaging method. However, the distinction between signal and noise is then less reliable and influenced by manual tuning.

In order to prepare uncertainty estimation of tensor parameters, we also estimate the covariance matrix
$\left(\chi_{\nu,\nu'}\right)_{\nu,\nu' \in \{1,\ldots,N\}}$ describing the stochastic error in the spectra. This matrix captures the standard deviation of the stochastic error at each frequency ν as $\sqrt{\chi_{\nu,\nu}}$ as well as the dependency of stochas-

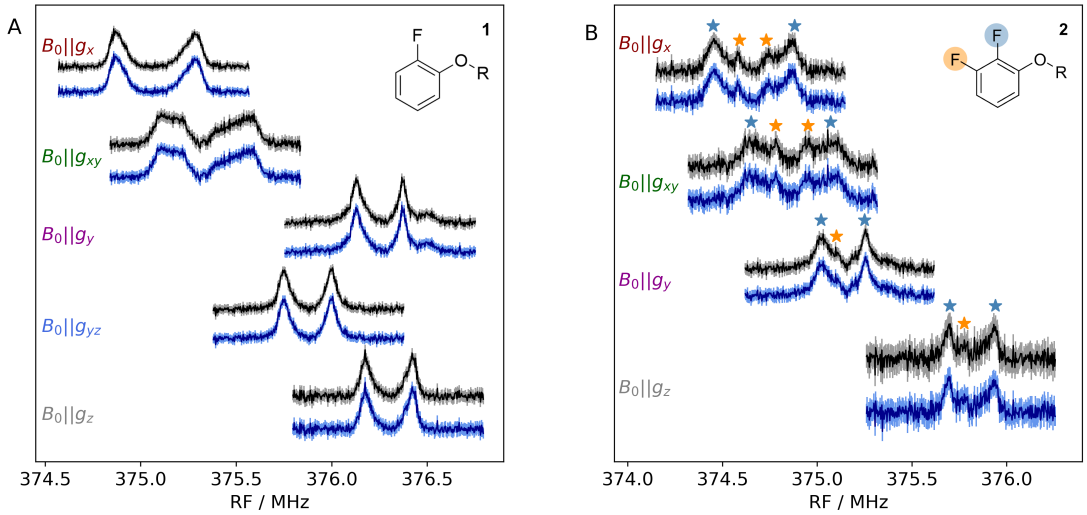


Figure D4: A, B: Comparison of 263 GHz ^{19}F Mims ENDOR spectra of compounds **1** and **2** at different orientations, respectively, estimated via the SDM (black) and the averaging method (blue). The chemical structures are indicated in the inset. Shaded areas correspond to approximate 95% pointwise confidence intervals obtained via bootstrap for the SDM (grey) and quasi-bootstrap for the averaging method (light blue). The dipolar splitting corresponding to the nitroxide-fluorine inter-spin distance is well visible for compound **1**. For compound **2**, only one dipolar splitting is well resolved, whereas the second, smaller splitting is partially suppressed by the central spectral hole of the Mims sequence. The resolved features of the dipolar splittings are indicated by colored asterisks for both F-atoms. The CS tensor leads to an asymmetry in the spectra and it cannot be evaluated visually. The MW frequency used for compound **1**, orientation g_y, differs from the one used for all other spectra resulting in a shift of the ^{19}F resonance.

tic error at different frequencies ν and ν'. Empirically, we found χ to be approximately diagonal, see Figure D10 in SI A.3, which means that the stochastic errors at different RF frequencies are approximately uncorrelated. Finally, we examine the distribution of the $\{I_\nu^{*,1}, \ldots, I_\nu^{*,J}\}$ for some chosen fixed values of ν and find that it is well-approximated by a Gaussian distribution in each case, see Figure D11 in SI A.3. This justifies a Gaussian error model for the stochastic error of the spectra.

3 Spectral Simulation: Parameters and Algorithm

In this section, we discuss the parameters included in the optimization process. These comprise the *spin parameters* in the spin Hamiltonian and the *experimental parameters* magnetic field strength and line broadening.

3.1 Spin Hamiltonian Parameters for the Nitroxide - ^{19}F System

The general spin Hamiltonian for the nitroxide-^{19}F spin system at 9.4 T / 263 GHz was discussed in our previous publication [26]. Briefly, it consists of two parts, $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$:

$$\hat{\mathcal{H}}_1 = \frac{\mu_\mathrm{B}}{\hbar} \boldsymbol{B}_0^T \boldsymbol{g} \hat{\boldsymbol{S}} - \frac{\mu_\mathrm{N} g_{\mathrm{n}(^{14}\mathrm{N})}}{\hbar} \boldsymbol{B}_0^T \hat{\boldsymbol{I}}_{^{14}\mathrm{N}} + \hat{\boldsymbol{S}}^T \boldsymbol{A}_{^{14}\mathrm{N}} \hat{\boldsymbol{I}}_{^{14}\mathrm{N}} + \hat{\boldsymbol{I}}_{^{14}\mathrm{N}}^T \boldsymbol{P}_{^{14}\mathrm{N}} \hat{\boldsymbol{I}}_{^{14}\mathrm{N}} \tag{2}$$

$$\hat{\mathcal{H}}_2 = \sum_{k=1}^{N_{^{19}\mathrm{F}}} \left[-\frac{\mu_\mathrm{N} g_{\mathrm{n}(^{19}\mathrm{F})}}{\hbar} \boldsymbol{B}_0^T (\mathbf{1} - \boldsymbol{\sigma}_{^{19}\mathrm{F}_k}) \hat{\boldsymbol{I}}_{^{19}\mathrm{F}_k} + \hat{\boldsymbol{S}}^T \boldsymbol{A}_{^{19}\mathrm{F}_k} \hat{\boldsymbol{I}}_{^{19}\mathrm{F}_k} \right], \tag{3}$$

where $h = 2\pi\hbar$ is Planck's constant, μ_B and μ_N are the Bohr and nuclear magnetons, respectively, g_n is the nuclear g factor and k enumerates the fluorine nuclei whose total number is $N_{^{19}\mathrm{F}} = 1$ for compound **1** and $N_{^{19}\mathrm{F}} = 2$ for compound **2**. $\boldsymbol{A}$, $\boldsymbol{P}$ and $\boldsymbol{\sigma}$ denote HF, quadrupolar and CS tensors, respectively. Please note that the hat symbol in this section indicates a quantum mechanical operator, not an estimator.

In the high-field approximation for the ^{19}F nuclei, the spin operators $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{I}}_{^{19}\mathrm{F}_k}$ can be replaced by the m_S and $m_{I(^{19}\mathrm{F}_k)}$ quantum numbers, where $\sigma_{zz(^{19}\mathrm{F}_k)}$ and $A_{zz(^{19}\mathrm{F}_k)}$ are the scalar zz-components of the respective $\boldsymbol{\sigma}_{^{19}\mathrm{F}_k}$ and $\boldsymbol{A}_{^{19}\mathrm{F}_k}$ tensors:

$$\hat{\mathcal{H}}_2 \simeq \sum_{k=1}^{N_{^{19}\mathrm{F}}} \left[\frac{\mu_N g_{n(^{19}\mathrm{F})}}{\hbar} B_0 (1 - \sigma_{zz(^{19}\mathrm{F}_k)}) m_{I(^{19}\mathrm{F}_k)} + m_S A_{zz(^{19}\mathrm{F}_k)} m_{I(^{19}\mathrm{F}_k)} \right] \tag{4}$$

The parameters of the spin Hamiltonian $\hat{\mathcal{H}}_1$ are inferred from the simulation of the EPR spectra using full matrix diagonalization. The parameters of the spin Hamiltonian $\hat{\mathcal{H}}_2$ report on the inter-spin distance between the nitroxide and the fluorine and are the subject of optimization. In the case of compounds **1** and **2**, this distance is large so that we assume zero rhombicity [16], whence the HF interaction tensor can be expressed as

$$\boldsymbol{A}_{^{19}\mathrm{F}} = R_A \begin{pmatrix} a_\mathrm{iso} + 2T & & \\ & a_\mathrm{iso} - T & \\ & & a_\mathrm{iso} - T \end{pmatrix} R_A^T, \tag{5}$$

where T represents the dipolar coupling strength, which depends on the nucleus and the inter-spin distance r. The order of the eigenvalues was adopted from [16] as it was assumed that the dipolar axis of the HF tensor would be close to parallel with g_x. Generally, the order is such that the largest tensor component is along the z direction as detailed in [33]. The value a_iso describes the isotropic part arising through the Fermi contact mechanism. Finally, the rotation matrix R_A determines the orientation of $\boldsymbol{A}$ in the nitroxide $\boldsymbol{g}$-tensor frame: we use reduced* Euler angles α_A, β_A. In total, only four parameters are required to describe the $\boldsymbol{A}$ tensor: $a_\mathrm{iso}, T, \alpha_A$ and β_A.

Similarly to the HF tensor, the CS tensor $\boldsymbol{\sigma}_{^{19}\mathrm{F}}$ may be parameterized through its eigenvalues $\tilde{\sigma}_{xx}$, $\tilde{\sigma}_{yy}$ and $\tilde{\sigma}_{zz}$ along with an associated rotation matrix $R_{\boldsymbol{\sigma}_{^{19}\mathrm{F}}}$ (parameterized using the three Euler angles $\alpha_\sigma, \beta_\sigma, \gamma_\sigma$) as given in Equation (6), left. In our code, we adopt an alternative representation via the diagonal and off-diagonal entries of the symmetric 3x3 matrix (Equation (6), right) which avoids singularities of Euler angles as coordinates (for $\beta_\sigma = 0$, any combination of α_σ and γ_σ with the same sum $\alpha_\sigma + \gamma_\sigma$ implies the same orientation, so one degree of freedom is lost) so that we alternatively use parameters $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz}$.

$$\boldsymbol{\sigma}_{^{19}\mathrm{F}} = R_{\boldsymbol{\sigma}_{^{19}\mathrm{F}}} \begin{pmatrix} \tilde{\sigma}_{xx} & 0 & 0 \\ 0 & \tilde{\sigma}_{yy} & 0 \\ 0 & 0 & \tilde{\sigma}_{zz} \end{pmatrix} R_{\boldsymbol{\sigma}_{^{19}\mathrm{F}}}^T = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \tag{6}$$

*The complete rotation sequence is z,y',z" with angles α, β, γ, respectively, with rotation matrix as defined in (B.48) in [34] but the final rotation about z has angle $\gamma = 0$ fixing the orientation ambiguity due to the repeated eigenvalue via $R_{A,yz} \stackrel{!}{=} 0$. The combination of this choice of Euler angles and order of eigenvalues is mathematically cumbersome but it is retained here for comparability with [16, 26].

Finally, we note that a rotation symmetry in the Hamiltonian leads to identical spectra for several distinct sets of interaction parameters. Careful treatment of this symmetry prevents its interfering with interaction parameter estimation. In the $\boldsymbol{g}$-tensor frame, symmetries under rotation of coordinates can be expressed by those rotation matrices M that leave the $\boldsymbol{g}$-tensor invariant. Since $\boldsymbol{g}$ is diagonal with distinct diagonal entries in this frame and these matrices must satisfy $MgM^T = g$, only the rotation matrices M from the set

$$\mathcal{M} := \left\{ \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \begin{pmatrix} -1 & & \\ & 1 & \\ & & -1 \end{pmatrix}, \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \right\} \tag{7}$$

leave $\boldsymbol{g}$ invariant. Just like $\boldsymbol{g}$, the HF and quadrupole interaction tensors $\boldsymbol{A}_{^{14}\mathrm{N}}$ and $\boldsymbol{P}_{^{14}\mathrm{N}}$ are diagonal in the $\boldsymbol{g}$ frame and will therefore also be invariant under transformation by $M \in \mathcal{M}$. Hence, the interaction parameter subsets $(\boldsymbol{\sigma}_{^{19}\mathrm{F}}, \boldsymbol{A}_{^{19}\mathrm{F}})$ and $(M\boldsymbol{\sigma}_{^{19}\mathrm{F}}M^T, M\boldsymbol{A}_{^{19}\mathrm{F}}M^T)$ will yield equivalent Hamiltonians $\hat{\mathcal{H}}_1, \hat{\mathcal{H}}_2$ and hence identical spectra for all $M \in \mathcal{M}$. Geometrically, these symmetry operations correspond to rotations about axes of the $\boldsymbol{g}$ tensor frame by 0 or 180 degrees.

3.2 Experimental Parameters

Several experimental quantities, particularly those connected to the pulse sequence such as MW frequency, MW pulse length and shape, inter-pulse delay τ and RF axis values, were assumed known with sufficient precision as their estimation would add considerable complexity and lead to non-identifiability absent penalization or prior knowledge. Thus, we considered only those experimental parameters that are known to be a major source of error: the static magnetic field strength and the ENDOR line broadening.

Magnetic Field Strength B_0

The magnetic field strength affects orientation selection and has been observed to slowly degrade, likely due to residual resistance losses of the superconducting magnet. This drift amounts to approximately -2.7 G per day. We previously proposed measuring ^{1}H ENDOR resonance frequencies to reference the Larmor frequency of fluorine nuclei [26]. In the present study, we have re-examined this method to calibrate the absolute magnetic field strength and we found it to be subject to stochastic errors of several G (data not shown). Thus, the magnetic field strength was retained as a parameter.

Line Width Convolution Parameter

In ENDOR spectral simulations, a convolution with a line broadening function is usually applied to account for the experimentally observed line width. While a lower bound for the line broadening can be determined from the inverse of the RF pulse length, we previously found that optimal line width typically exceeds this minimum [16] and this depends on the selected orientation. Thus, the line width was retained as an optimization parameter.

3.3 Accelerated ENDOR Simulation Algorithm (SimSpec)

In order to estimate interaction parameters, frequently repeated simulation of ENDOR spectra for slightly different parameter values is required. Therefore, a major step of this work consisted in the acceleration of the spectral simulation code in Matlab, previously reported in [26], to which we refer as Sim. This was achieved by re-writing Sim in Python along with several algorithmic improvements detailed below.

Briefly, the accelerated code SimSpec computes transition energies by exact diagonalization of the nitroxide Hamiltonian $\hat{\mathcal{H}}_1$. These resonance energies are utilized to compute orientation selection and accumulate contributions to the ENDOR spectrum through a histogram weighted by the 'hole'-function proposed by Mehring [35]. Subsequently, the selected orientations are used to compute the fluorine resonance according to $\hat{\mathcal{H}}_2$ in high-field approximation, see Equation (4). The effect of the Mims ENDOR blind spot is treated analytically as proposed by [36]. For more details, see [26].

For the powder pattern and orientation selection, SimSpec offers the choice between a Polar grid and the SOPHE grid [37], whereas Sim uses a grid similar to the Polar grid. Significant speed improvements were obtained by (i) pre-calculation of trigonometric expressions for all positions on the grid so that only changed quantities are re-computed and (ii) tensorification of the code exploiting high performance numerical linear algebra subroutines available through Numerical Python to reduce the number of explicit `for` loops.

A speed comparison between Sim and SimSpec to simulate the 263 GHz ^{19}F-ENDOR spectra of compound **1** yields the execution times in Table D1. All parameters were chosen for comparable computational accuracy between these codes. Repeating diagonalization of the Hamiltonian is only necessary if the magnetic field strength B_0 or the $\boldsymbol{g}$ eigenvalues have been changed and so computational speed-ups are available by selectively updating subsets of the parameters. Additionally, considerable computational savings from pre-calculation of the EPR spectrum and trigonometric expressions are apparent comparing execution times ('w/precalc' vs 'w/o precalc'). Overall, speed-ups by a factor between 10 and 100 or more were achieved. More details are available in SI C.

Code	Sim		SimSpec			
			Polar grid		SOPHE grid	
	total	ENDOR	w/o precalc.	w/ precalc.	w/o precalc.	w/ precalc.
g_x	167	128	3.3	0.31	3.7	0.35
g_y	364	323	3.7	0.33	3.7	0.34
g_{yz}	336	296	3.4	0.32	4.0	0.33
g_z	164	125	3.5	0.31	3.5	0.34

Table D1: Approximate execution times (in seconds) for the simulation of the 263 GHz ENDOR spectra of compound **1**. 'Total' refers to the full execution time including computation of the EPR spectrum and other set-up costs, whereas 'ENDOR' refers to the execution time of the ENDOR spectrum only. Similarly, 'w/ precalc' includes the execution time of setting up the grid, pre-calculating trigonometric expressions and diagonalizing the Hamiltonian whereas 'w/o precalc' excludes these times.

4 Inference Methodology

In this section, we introduce the methodology employed for estimation and quantification of the stochastic part of the error in this estimation. Chiefly, this consists of choosing a reasonable loss function that quantifies the fit between measured and simulated spectrum and an optimization procedure that needs to be carefully designed. We propose Bayesian optimization to perform a global search followed by refinement through a gradient-based method. Bayesian optimization is particularly suited as the loss function exhibits a large number of local minima that many other optimization algorithms tend to get stuck in.

4.1 Loss Function

The full set of spin and experimental parameters can be described by a parameter vector θ (not to be confused with any polar angle) of dimension $2N_o + 10N_{^{19}\mathrm{F}}$, where N_o denotes the number of orientations available (enumerated as $o \in \{1, \ldots, N_o\}$ corresponding to those $g_x, g_{xy}, g_y, g_{yz}, g_z$ for which data are available).
For a given θ, our SimSpec code (Section 3.3) simulates spectra $I_{o,\nu}(\theta)$ which are compared to the measured spectra[†] $\hat{I}_{o,\nu}$ obtained using the SDM from Section 2.2.
We seek a value for θ such as to yield simulated spectra that match the measured spectra for all five orientations as closely as possible. We quantify the deviation of the simulated spectra from the measured spectra through the *loss function* (a re-scaled mean square deviation)

$$\mathcal{L}(\theta) = \sum_{o=1}^{N_o} \sum_{\nu=1}^{N} |\hat{I}_{o,\nu} - I_{o,\nu}(\theta)|^2. \tag{8}$$

The choice of this particular loss function is partly justified by the properties of $\hat{I}$: for large numbers B of batches we expect the estimator $\hat{I}$ to approximately follow a Gaussian distribution and this is also observed approximately in bootstrap experiments (see SI A), hence approximately making $\mathcal{L}(\theta)$ a negative multiple of the log likelihood for θ. The problem of finding those parameter values θ^* in the parameter space Θ that yield the best fit between simulated and measured spectra is hence cast as the problem of minimizing $\mathcal{L}$ over θ.

4.2 Optimization Algorithms

We seek to solve the problem of minimizing the loss function from Equation (8):

$$\theta^* \in \operatorname*{argmin}_{\theta \in \Theta} \mathcal{L}(\theta), \tag{9}$$

where $\mathcal{L} : \mathbb{R}^m \supset \Theta \to \mathbb{R}$, in the context of optimization algorithms, is known as the *objective function*. We observed empirically that $\mathcal{L}$ was reasonably smooth if sufficiently precise spectral simulation was used (see Figure D12 in SI C). However, it possessed a multitude of local minima and its domain is of moderately high dimension (up to 12 in our data depending on the compound under investigation). Hence, we decided to use Bayesian optimization which is an iterative strategy to solve the global optimization problem Equation (9). There are two main ingredients involved in Bayesian optimization: a statistical model for the objective function $\mathcal{L}$ and an acquisition function to determine which $\theta \in \Theta$ should be tried next. *A priori*, the loss function is modelled as a Gaussian process with some mean and covariance kernel ideally chosen to reflect pre-existing understanding of the loss function. Here, we adopted a standard zero mean function and Matérn covariance kernel [38] of order $\nu = 1.5$ with expected improvement as acquisition function. An introduction to this algorithm is provided in SI E, a more detailed exposition can be found in [38]. Instead of specifying a starting value, Bayesian optimization requires boundaries to be specified for all parameters: some parameters have natural boundaries such as α_A and β_A in the HF tensors, whereas other parameter boundaries are chosen to consider only physically reasonable values. One advantage of Bayesian optimization is that it deals well with large regions of parameter space, which enabled us to include the full range of theoretically possible Euler angles.
Once Bayesian optimization has identified a parameter value near the global minimum, refining the estimate using a quasi-Newton method such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) [39] is standard practice because this will converge to θ^* quickly. To enhance the performance of BFGS,

[†]All spectra are normalized by imposing $\sum_{\nu=1}^{N} I_{o,\nu}(\theta) \stackrel{!}{=} \sum_{\nu=1}^{N} \hat{I}_{o,\nu} \stackrel{!}{=} 0$ and $\sum_{\nu=1}^{N} I_{o,\nu}^2(\theta) \stackrel{!}{=} \sum_{\nu=1}^{N} \hat{I}_{o,\nu}^2 \stackrel{!}{=} 1$ for all orientations o.

we elected to supply gradients of $\mathcal{L}$ approximated via a manually tuned finite difference method, see Figure D17 in SI E.

4.3 Approximate Confidence Regions

Having approximately computed the best parameter value θ^*, there is a need to assess its error. Statistical methods allow quantification of the *stochastic error*, *i.e.* the parameter uncertainty implied by the measurement error of the spectrum. We start from the approximate covariance matrix of the measured spectrum, $\chi \in \mathbb{R}^{N\times N}$, which arises from the measurement error and is computed using the bootstrap method, see Section 2.2 for details. We then compute the matrix of partial derivatives of the simulated spectrum with respect to the parameters, $J = \left(\frac{\partial I_{o,\nu}(\theta)}{\partial \theta_i}\right)_{i\in\{1,\ldots,\dim(\Theta)\},\,\nu\in\{1,\ldots,N\}} \in \mathbb{R}^{\dim(\Theta)\times N}$, by finite difference approximation. A linear approximation of the spectral simulation algorithm, $I_{o,\cdot}(\theta) \approx I_{o,\cdot}(\theta^*) + J^T(\theta - \theta^*)$, then enables us to approximately obtain the covariance matrix describing the stochastic uncertainty in θ^* that is implied by the uncertainty in $\hat{I}$ via

$$\chi \approx J^T \mathrm{Cov}(\theta) J. \tag{10}$$

The least-square solution of Equation (10) for $\mathrm{Cov}(\theta)$ is used to construct approximate confidence regions for θ^*. This corresponds to linear propagation of Gaussian errors and delivers good results for small uncertainty ranges but less reliable results for particularly large uncertainty ranges. More detail on how uncertainties of orientations have been handled is available in SI D.

5 Optimization Results

Bayesian optimization requires the specification of boundaries of the parameter space and is sensitive to its dimension. Hence, we proceeded by firstly fixing the experimental parameters (see Subsection 3.2) at reasonable initial values (a line width of 20 kHz from [26] and the magnetic field strengths estimated from ^{1}H Larmor frequencies). Then, we used Bayesian optimization for all spin parameters except for $a_{\mathrm{iso}} = 0$ which was kept fixed. Examples of intermediate steps (iterations 20 and 300) of the Bayesian optimization are plotted in Figure D16 in SI F in the case of compound **1**. These show that after 300 iterations, parameter estimates start to yield reasonably matched spectra and therefore Bayesian Optimization was halted at this point to limit computational time. In a second step, all experimental and spin parameters including a_{iso} were jointly optimized using BFGS.

Choosing boundaries for the parameter space requires prior knowledge. Since Bayesian optimization deals well with large regions of parameter space, the boundaries were chosen to include all plausible parameter values. We considered the visible features in the spectrum (e.g. dipolar peaks in the spectrum, see Figure D4) as well as the spread of DFT-predicted values for CS tensors, reported in [26]. This resulted in the boundaries given in Table D2.

Parameter name	T/kHz	$\alpha_A/^\circ$	$\beta_A/^\circ$	σ_{xx}	σ_{yy}	σ_{zz}	σ_{xy}	σ_{xz}	σ_{yz}
Lower boundary	50	-180	0	191	83	115	-118	-129	-87
Upper boundary	320	180	180	325	558	360	118	129	87

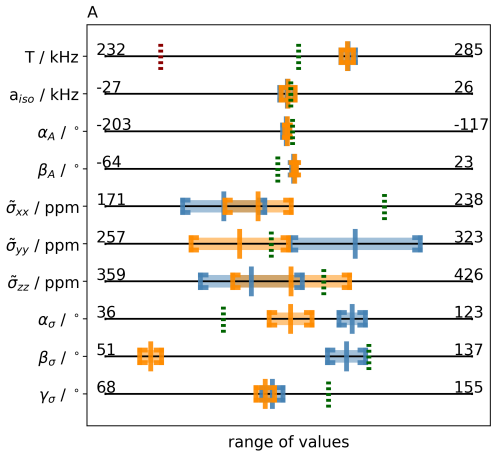
Table D2: Boundaries for the spin parameters used in Bayesian optimization for both compounds **1** and **2**. For CS tensors, we use the parameterization via matrix entries as per Equation (6). Note that the precise values of the boundaries (*e.g.* -129 vs -130) are irrelevant as long as the ranges are large enough to include all plausible values.

Compound 1

For compound **1**, two local minima of the loss function were identified, corresponding to two solutions. These two parameter sets are compared in panel A of Figure D5, where blue and orange bars illustrate the approximate 95% confidence interval of each parameter. Notably, the two minima have indistinguishable HF parameters with narrow confidence intervals. The agreement of the HF parameters between the two minima is an important finding, since we consider the HF parameters as the primary source of information from ^{19}F ENDOR spectroscopy for (biological) structure determination. The two minima differ in their CS parameters and their occurrence is not surprising because of a lack of resolution of the CS tensor and a parameter space of moderately high dimension. Based on the optimization procedure alone, we cannot decide which set of parameters is preferable. We selected the parameter set represented in blue in this panel (values given in panel B) because the CS tensor Euler angle β_σ shows better consistency with DFT whereas the eigenvalues differ less between the two minima considering their confidence regions. In this regard, our approach is similar to that of including DFT-derived information penalizing deviations of estimated parameter values from their anticipated values [27].
The spectral residuals, shown in panel A of Figure D6, demonstrate that the optimization procedure leads to a very close fit, substantially improving over the previously published fit shown in panel A of Figure D7. Close inspection of the spectral residuals reveals some structure, *i.e.* the spectral residuals deviate from the expected pure noise indicating the presence of some systematic error due to imperfect model fit. Our statistical approach also yields correlations between stochastic errors for different parameters (see panel B of Figure D6): for instance, we observe a strong positive correlation between magnetic field strengths and CS tensor values and a weaker negative correlation between magnetic field strengths and HF tensor parameters.
Given the closeness of measured and simulated spectra, the deviation of the estimated T-value from DFT and the previously reported value visible in panels A and B of Figure D5 is striking. This triggered an examination of instrumental parameters, which revealed that nominal RF frequency differed from actual RF frequency, possibly by up to 8 kHz, due to a resolution issue in the commercial RF unit. This lead to a systematic error in the spectra and could partially explain the observed difference (see panel B in Figure D5) in T-values to the ones previously obtained at 94 GHz [16], where the RF issue was absent. This finding underlines that a comparison of stochastic error with observed deviation from theoretical values can trigger a more careful study of instrumental errors.

Compound 2

Compound **2** exhibits a larger number of parameters due to its two non-equivalent ^{19}F nuclei. Therefore, we decided first to use Bayesian optimization to search for the CS parameters, keeping the HF interaction parameters fixed at reasonable values based on the peak positions in the spectra (Figure D4 B). In a second step, BFGS was carried out over all spin parameters of the two nuclei. In a third step, we used BFGS over both spin and experimental parameters. Increasing the dimension of the parameter space via the second step ensures that BFGS stays near the minimum identified by Bayesian optimization rather than being attracted by another local minimum.
A substantial number of local minima of the loss function were identified: a fairly intensive search yielded four local minima (enumerated as minimum #1 to #4) but it is probable that there are further local minima not yet identified. To decide between the four minima, we relied firstly on the qualitative reproduction of the m-fluorine HF coupling. We plotted the HF tensor parameters for both ortho and meta fluorine nuclei for the four minima and compared them with DFT and X-ray-derived values in Figure D8. Panel A of this figure shows that estimated HF parameters for the ortho fluorine nucleus are all indistinguishable taking their stochastic error into account. We note a



B Spin Parameters for Compound **1**

parameter	Minimum #1	Sim. [26]	DFT [26]
T / kHz	267.2 ± 1.1	260	260
a_{iso} / kHz	-0.5 ± 1.1	0	0
α_A / °	-160.5 ± 0.8	-158	-159
β_A / °	-19.2 ± 0.5	-19	-23
$\tilde{\sigma}_{xx}$ / ppm	193 ± 7	222	222
$\tilde{\sigma}_{yy}$ / ppm	302 ± 12	287	287
$\tilde{\sigma}_{zz}$ / ppm	386 ± 9	399	399
α_σ / °	94 ± 3	65	64
β_σ / °	108 ± 5	115	113
γ_σ / °	108 ± 3	120	121

Figure D5: A: Estimated HF interaction and CS parameters and the corresponding approximate 95% confidence regions for the ^{19}F nucleus in compound **1**: minimum #1 in blue, minimum #2 in orange, DFT values as dotted green line and T-value calculated from X-ray structure as dotted red line from [16]. B: Comparison of minimum #1 with manual fitting results and DFT values reported in [26]. g-values were $g_{x,y,z} = [2.00886, 2.00610, 2.00211]$, as reported in Section 2.1. Parameter uncertainties (approximate 95 % confidence regions) consider only stochastic error, not systematic error. Parameter estimates in [26] were based on 94 and 263 GHz data whereas optimized values here are derived solely from 263 GHz data. Uncertainties in [26] were assessed only from the impact of parameter-wise changes on the spectrum and are therefore not reported here. Similarly, DFT uncertainties are difficult to assess in general and are therefore not specified here, either.

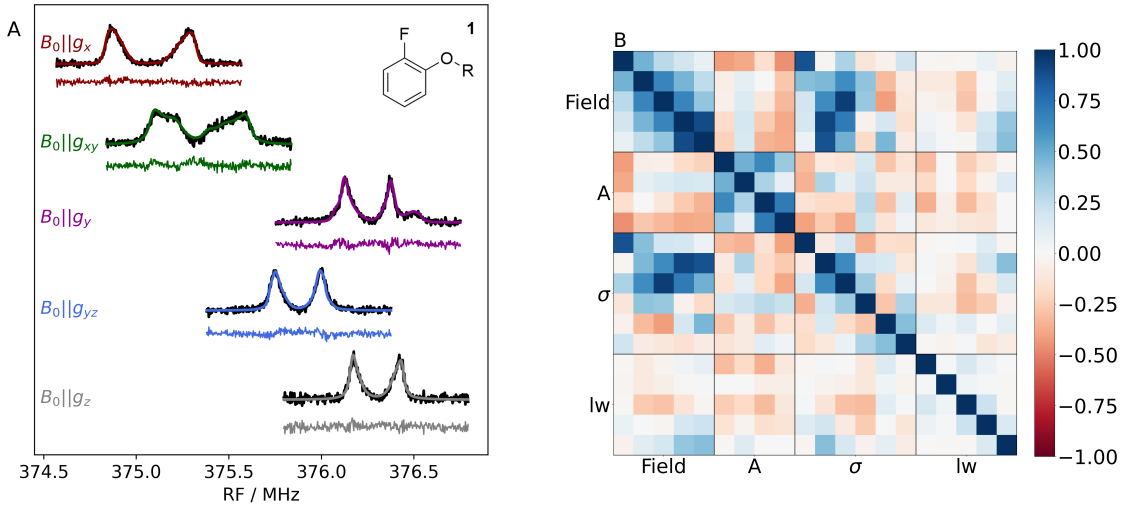


Figure D6: A: The ^{19}F ENDOR spectra $\hat{I}_{o,\nu}$ (black) extracted with the SDM for all orientations. In different colors: the corresponding spectra $I_{o,\nu}(\hat{\theta})$ simulated with the values indicated in the Table D4. The spectral residuals are plotted below each of the spectra using the same color. B: correlation matrix of the corresponding parameters (from top to bottom/left to right: magnetic field strength B_0 at each of the five orientations $g_x, \ldots, g_z$, $a_{\mathrm{iso}}, T, \alpha_A, \beta_A$, $\tilde{\sigma}_{xx}, \tilde{\sigma}_{yy}, \tilde{\sigma}_{zz}, \alpha_\sigma, \beta_\sigma, \gamma_\sigma$ and line width for each of the five orientations $g_x, \ldots, g_z$) calculated as described in Section 4.3.

slight deviation of the T-value from the DFT and previously reported values in the same direction and of similar magnitude to that observed for compound **1**, likely due to similar systematic errors.

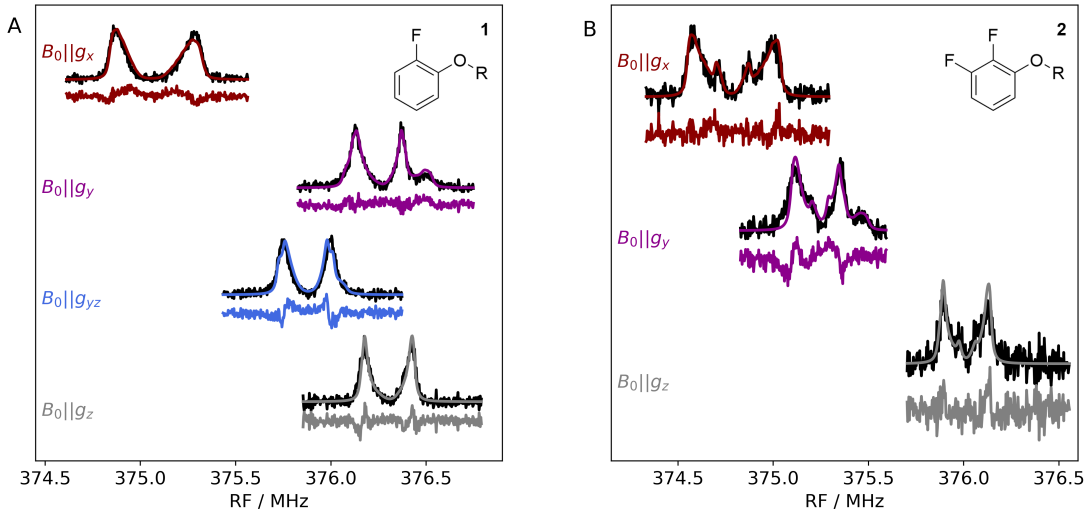


Figure D7: A, B: Spectra, simulation and spectral residuals resulting from the estimated parameters reported in [26] for compounds **1** and **2**, respectively, with the chemical structures indicated in the inset. Parameters are given in panel B of Figure D5 for compound **1** and Table D3 for compound **2**.

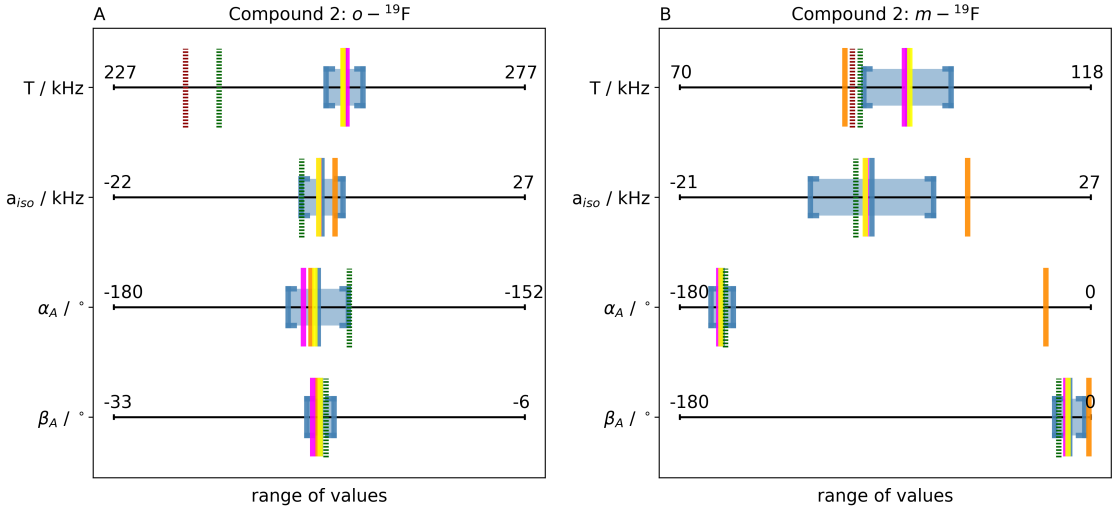


Figure D8: Estimated HF parameters for minima #1, #2, #3 and #4 (blue, orange, pink, yellow) for compound **2**. A: ortho ^{19}F nucleus, B: meta ^{19}F nucleus. For minimum #1, the corresponding approximate 95% confidence regions are shown in light blue. DFT values are shown in green dotted lines and the T-value calculated from X-ray structure in red dotted lines from [26].

Panel B displays the values for the m-fluorine nucleus and it is readily visible that minimum #2 (in orange) exhibits a strong deviation of dipolar tensor orientation from DFT values as well as from those of all other minima. In spectral simulation, this corresponds to the failure to reproduce the $m-$fluorine peak in orientation g_y, see panel A of Figure D20 in SI F.3. Therefore, we discarded minimum #2. In order to decide between minima #1, #3 and #4, we considered the relative orientation of the dominant eigenvectors of the CS tensors (*i.e.* the eigenvector associated with the largest eigenvalue of each CS tensor). These should both be orthogonal to the phenyl ring based on NMR studies [40] and therefore parallel to each other. To visualize this, we computed the angle

between these two dominant eigenvectors for each minimum, taking stochastic error into account. This is represented in Figure D21 in SI F.3 which shows that minimum #1 is best compatible with angle 0°, although minimum #3 is also acceptable. Minimum #4 can be excluded. The results of the optimization procedure including correlations for minimum #1 are summarized in Figure D9 and reported in Table D3. The approximate confidence regions for the CS tensors were determined using error propagation from Section 4.3 and tensor alignment as detailed in SI D. The spectral residuals are plotted in panel A of this figure. We find a very close fit: spectral residuals consist mostly of noise. This represents a main result, namely a clear improvement in the spectral residuals over the results of previous manual parameter tuning [26] as displayed in panel B of Figure D7. Moreover, a great advantage is that our optimization procedure typically has computational time amounting to a few hours on a standard PC as compared to several weeks of manual parameter tuning underlying the results in [26]. Overall, despite limited resolution of CS tensors and moderately high dimension of the parameter space, we conclude that the method was able to identify a physically reasonable set of parameter values.

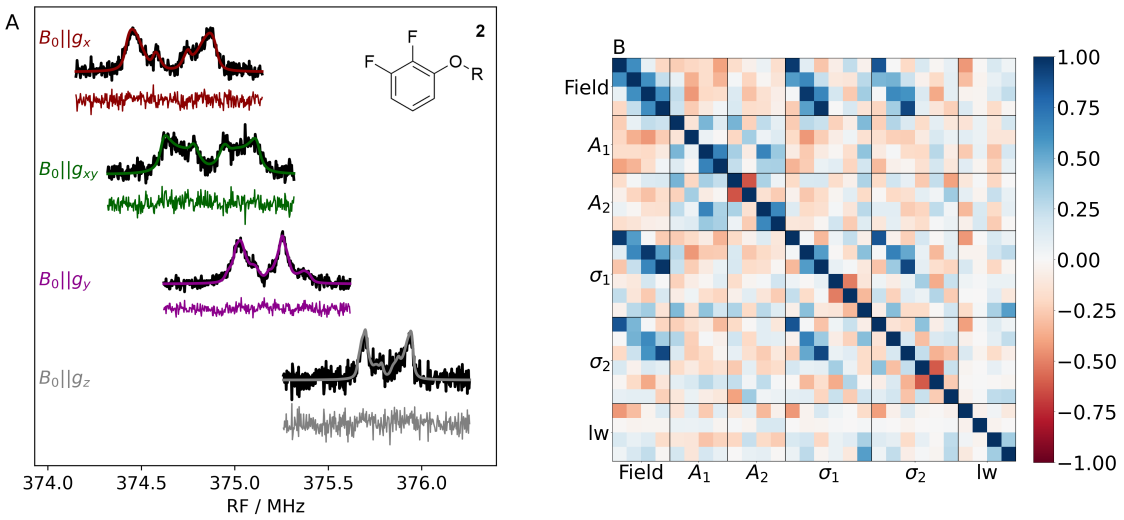


Figure D9: A: the ^{19}F ENDOR spectra $\hat{I}_{o,\nu}$ (black) extracted with the SDM for all orientations of compound **2**. In different colors: The corresponding spectra $I_{o,\nu}(\hat{\theta})$ simulated with HF and CS parameters from minimum #1 and experimental parameters as in Table D4. The spectral residuals $\hat{I}_{o,\nu} - I_{o,\nu}(\hat{\theta})$ are plotted below each of the spectra using the same color. B: The correlation matrix of the corresponding parameters calculated as described in Section 4.3 for compound **2**. Parameters are sorted as in panel B of Figure D6 except for the CS tensors where matrix entries $\sigma_{xx}, \ldots, \sigma_{yz}$ according to Equation (6) are used.

Final Remarks and Conclusion

Finally, we discuss the results regarding estimated magnetic field strengths and line widths which are reported in Table D4. We observe major deviations between magnetic field strengths previously used in spectral simulation and their current estimates, particularly for compound **2**. This strongly suggests that in the previous analysis, a systematic error was present in the combined determination of the magnetic field strengths and g-values. Indeed, only an internal magnetic field strength calibration (Bruker field linearization) was available in that study.
The ENDOR line widths were also optimized and for most orientations the values obtained are reasonably close to the ones used in [26] which are based on RF pulse lengths. However, orientations g_{xy} and g_z in compound **1** as well as orientation g_{xy} in compound **2** showed a markedly

Spin Parameters for Compound **2**

parameter	Minimum #1	Sim.	DFT	Minimum #1	Sim.	DFT
	o-F			*m*-F		
T / kHz	255 ± 2	250	240	97 ± 5	90	91
a_{iso} / kHz	2 ± 3	0	0	2 ± 7	0	0
α_A / °	-166 $\pm$ 2	-160	-164	-161 $\pm$ 5	-168	-160
β_A / °	-19 $\pm$ 1	-15	-19	-9 $\pm$ 7	-16	-14
$\tilde{\sigma}_{xx}$ / ppm	$139 \in [86, 194]$	257	257	$117 \in [56, 174]$	238	238
$\tilde{\sigma}_{yy}$ / ppm	$159 \in [134, 194]$	274	274	$184 \in [129, 235]$	286	286
$\tilde{\sigma}_{zz}$ / ppm	$331 \in [304, 366]$	450	450	$289 \in [257, 337]$	413	413
α_σ / °	$9 \in [-47, 47]$	50*	49*	$32 \in [-40, 68]$	53*	53*
β_σ / °	$116 \in [109, 124]$	100*	102*	$71 \in [53, 87]$	80	83
γ_σ / °	$116 \in [57, 73] \cup [101, 130]$	100*	102*	$-77 \in [-110, -52]$	-60	-59

Table D3: HF interaction and CS tensor parameters for the ^{19}F nuclei in compound **2**. g-values were $g_{x,y,z} = [2.00886, 2.00610, 2.00211]$, as reported in Section 2.1. Parameter uncertainties (approximate 95 % confidence regions) consider only stochastic error, not systematic error. Uncertainties for CS eigenvalues and Euler angles are computed using the method in SI D. Euler angles marked with an asterisk have been changed by symmetry transformations according to Equation (7) to map them closer to the current confidence regions. Previous estimated values from manual parameter tuning (Sim.) and corresponding DFT values (DFT) used as input, both from [26], are included for comparison. Parameter estimates in [26] were based on 94 and 263 GHz data whereas optimized values here are derived solely from 263 GHz data. Uncertainties in [26] were assessed only from the impact of parameter-wise changes on the spectrum and are therefore not reported here. Similarly, DFT uncertainties are difficult to assess in general and are therefore not specified here, either.

increased line width, see Table D4. This may be attributable to the observed magnetic field drift (see Section 3.2) which is not part of the model and results in a mixing of ENDOR spectra across a narrow range of magnetic field strengths. The resulting broadening is then matched by an increased estimated line width. The combined impact of deviations in line widths and magnetic field strengths could explain the systematic error observed in the spectral residuals in Figure D7.

Turning to the specified errors, it is important to point out that the specified approximate confidence regions account solely for stochastic error. While the spectral residuals, *e.g.* in Figure D9, are small, this does not necessarily mean that the implied systematic parameter errors are small. Indeed, including additional parameters in the spectral simulation such as two of the eigenvalues of the $\boldsymbol{g}$ tensor, has been observed to lead to a notable change in estimated parameters, larger than would be expected from stochastic error. This tendency is especially pronounced when the additional parameters are strongly correlated with ones already included, as is the case with eigenvalues of $\boldsymbol{g}$. Out of concern for such larger than expected parameter changes, we have tested an optimization considering the small spectral contribution with a second g_x value for compound **1**, as reported in Section 2.1. The results are displayed in Figure D19 in SI F.2 and show similar spectral residuals as using only a single g_x value, so no unexpectedly large parameter changes were observed in this case.

The main sources of systematic error likely include instrumental error as well as our assumption of a single molecular conformation. Also, the spectral simulation approach does not consider any spin dynamics and relaxation. Modelling of these complex effects requires a different mathematical approach and is left for future work. Our study shows that knowledge of experimental parameters,

	1		**2**	
	this work	Ref [26]	this work	Ref [26]
Loss $\mathcal{L}$ / a.u.	0.272	N/A	0.704	N/A
Experimental Parameters				
B_0 / G : g_x	93628.6 $\pm$ 0.6	93633	93522.0 $\pm$ 1.7	93524
B_0 / G : g_{xy}	93697.1 $\pm$ 0.7	93698†	93573.9 $\pm$ 1.7	93585†
B_0 / G : g_y	93937.5 $\pm$ 0.9	93936	93651.6 $\pm$ 2.5	93660
B_0 / G : g_{yz}	93838.4 $\pm$ 0.8	93838	N/A	N/A
B_0 / G : g_z	93943.5 $\pm$ 1	93942	93810.1 $\pm$ 5.4	93827
lw / kHz : g_x	15.2 $\pm$ 3.2	20	31.7 $\pm$ 6.9	20
lw / kHz : g_{xy}	51.0 $\pm$ 4.9	20†	46.7 $\pm$ 7	20†
lw / kHz : g_y	19.9 $\pm$ 2.4	20	23.9 $\pm$ 5.5	20
lw / kHz : g_{yz}	23.6 $\pm$ 2.8	20	N/A	N/A
lw / kHz : g_z	31.2$\pm$ 3.1	20	16.6 $\pm$ 6.6	20

Table D4: Magnetic field strengths and line widths obtained from the reported optimization procedure. Parameter uncertainties (approximate 95 % confidence intervals) consider only the stochastic error, they do not include systematic error. The values yielding the simulated spectra in our previous report [26] are included for comparison. The values marked with † concern data not presented in [26] and are obtained using the method employed there.

such as magnetic field strengths or actual RF frequency, is crucial for parameter estimation of fluorine nuclei; it is easy to over-estimate the precision with which such experimental parameters are known.

For future work, a Bayesian approach including molecular dynamics as a source of prior information is a possible route towards systems of greater relevance. Similarly, replication of the present results on additional compounds as well as use of ENDOR spectra recorded at other MW frequencies are planned to establish robustness of the method and its ability to deliver HF tensors. Extension to spin systems with a larger number of nuclei is likely challenging and will require the imposition of penalties or the adoption of a Bayesian approach. Nonetheless, the present work provides a first step towards improved and faster parameter estimation through a better fit between measured and simulated spectra.

6 Materials and Methods

6.1 Sample preparation

Compounds **1** and **2** were synthesized as described in [26]. For ESE and ENDOR, solutions contained sample concentrations of about 300 μM in a 1:1.5 mixture of deuterated DMSO and CD_3OD. Solutions were loaded into Suprasil capillary tubes (VitroCom CV2033-S-100; O.D./I.D. = 0.33/0.20 mm), and shock frozen in liquid nitrogen. Then, sample capillaries were inserted into the precooled resonator immersed in a liquid nitrogen bath. Subsequently, the cold resonator was transferred into the cryostat, precooled to about 80 K. Samples contained typical volumes of about 50 nL. A carbon fibre was used for g-value calibration, as proposed in [30]. For this measurement, a solution of compound **1** was loaded into the capillary along with carbon fibre and shock frozen in liquid nitrogen.

6.2 263 GHz EPR/ENDOR spectroscopy

263 GHz pulsed and CW EPR (ESE) as well as pulse ENDOR were recorded on a Bruker ElexSys E780 spectrometer equipped with a Bruker cylindrical TE012-mode EPR/ENDOR resonator (the model is E9501510) [25]. For ENDOR, the RF was produced by a the Bruker DICE II RF synthesizer and pulse forming unit. A 125 W RF-amplifier (Amplifier Research, model 125W1000) was used, and the Mims pulse sequence [1] was applied. All ENDOR spectra were recorded in batches, each containing the sum of between 50 and 100 individual scans. The batches were stored in one data matrix, containing the batches in one dimension and the RF spectra as a second dimension. This matrix was used for data processing with the SDM. All spectra, including CW-EPR, were detected using a quadrature detection scheme, which enabled phasing of the signal prior to acquisition.

Experimental settings and conditions: ESE (compound **1** or **2**): T = 50 K, MW frequency ν = 263 GHz, π/2-pulse = 32-40 ns, delay τ = 300 ns; shot repetition time (SRT) = 3 ms, 256 shots/point, 1 scan, data smoothed with a Savitzky-Golay filter (`window_length` 32 and `polyorder` 2).
ESE (compound 1 along with C-fibre): T = 40 K, MW frequency ν = 263.185 GHz, MW power = 50 mW, π/2 -pulse = 40 ns, π-pulse = 80 ns; delay τ = 1000 ns; SRT = 50 ms, 20 shots/point, 20 scans, data smoothed with a Savitzky-Golay filter (`window_length` 16 and `polyorder` 4).

CW-EPR (compound **1** along with C-fibre): T = 40 K, ν = 263.185 GHz, MW power = 0.5 mW, modulation frequency = 100 kHz, modulation field amplitude = 1 G, number of scans = 20.

Mims-ENDOR (compound **1** or **2**): T = 50 K, ν = 263 GHz, π/2-pulse (MW) = 32-40 ns, τ = 850 ns, π-pulse (RF) = 50 μs, SRT = 3 ms, 1 shot/point in stochastic acquisition, RF resolution: 333 RF points recorded at nominal resolution of 3 kHz. During the writing of the manuscript we became aware of a bug in the Bruker spectrometer software connected with the DICE II unit, which limits the actual RF resolution to likely up to 8 kHz. While this issue is being addressed, the analysis in this paper was performed assuming the nominal 3 kHz resolution to be the actual resolution. Acquisition time between 2 to 21 hours depending on sample and orientation as discussed in [26].

6.3 Data Availability

All raw data files, processing procedures and source codes for figure generation will be made available in an open repository upon publication. Also, the simulation code developed in [26] is freely available at
https://data.goettingen-research-online.de/dataset.xhtml?persistentId=doi:10.25625/FLQKPM.

7 Acknowledgements

C. Beeken worked on early versions of the code. We thank Konstantin Herb (ETH Zürich) for the donation of the C-fibres. H.W., B.E., S.H., A.M., M.B. and Y.P. thank the DFG — project-ID 432680300 — CRC 1456 for financial support. A.K. acknowledges the GGNB program IMPRS-PBCS for a PhD fellowship. M.B. acknowledges the ERC Advanced Grant 101020262 BIOenMR. We thank the Max Planck Society for financial support. S.H. acknowledges the Niedersachsen Vorab of the Volkswagen foundation. Y.P. gratefully acknowledges Royal Society International Exchanges grant IE150666.

References

[1] W. B. Mims, "Pulsed Endor Experiments", *Proceedings of the Royal Society of London Series A — Mathematical and Physical Sciences* **1965**, *283*, 452–457.

[2] E. R. Davies, "A new pulse ENDOR technique", *Physics Letters A* **1974**, *47*, 1–2.

[3] J. Niklas, T. Schulte, S. Prakash, M. van Gastel, E. Hofmann, W. Lubitz, "Spin-density distribution of the carotenoid triplet state in the peridinin-chlorophyll-protein antenna. A Q-band pulse electron-nuclear double resonance and density functional theory study", *Journal of the American Chemical Society* **2007**, *129*, 15442–15443.

[4] C. Teutloff, S. Pudollek, S. Keßen, M. Broser, A. Zouni, R. Bittl, "Electronic structure of the tyrosine D radical and the water-splitting complex from pulsed ENDOR spectroscopy on photosystem II single crystals", *Physical Chemistry Chemical Physics* **2009**, *11*, 6715–6726.

[5] S. Richert, B. Limburg, H. L. Anderson, C. R. Timmel, "On the influence of the bridge on triplet state delocalization in linear porphyrin oligomers", *Journal of the American Chemical Society* **2017**, *139*, 12003–12008.

[6] E. Schleicher, S. Rein, B. Illarionov, A. Lehmann, T. Al Said, S. Kacprzak, R. Bittl, A. Bacher, M. Fischer, S. Weber, "Selective ^{13}C labelling reveals the electronic structure of flavocoenzyme radicals", *Scientific Reports* **2021**, *11*, 1–9.

[7] J. M. Peloquin, K. A. Campbell, D. W. Randall, M. A. Evanchik, V. L. Pecoraro, W. H. Armstrong, R. D. Britt, "^{55}Mn ENDOR of the S_2-state multiline EPR signal of photosystem II: implications on the structure of the tetranuclear Mn cluster", *Journal of the American Chemical Society* **2000**, *122*, 10926–10942.

[8] D. Goldfarb, "High field ENDOR as a characterization tool for functional sites in microporous materials", *Physical Chemistry Chemical Physics* **2006**, *8*, 2325–2343.

[9] G. E. Cutsail III, J. Telser, B. M. Hoffman, "Advanced paramagnetic resonance spectroscopies of iron–sulfur proteins: Electron nuclear double resonance (ENDOR) and electron spin echo envelope modulation (ESEEM)", *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **2015**, *1853*, 1370–1394.

[10] T. U. Nick, W. Lee, S. Koßmann, F. Neese, J. Stubbe, M. Bennati, "Hydrogen Bond Network between Amino Acid Radical Intermediates on the Proton-Coupled Electron Transfer Pathway of *E. coli* α2 Ribonucleotide Reductase", *Journal of the American Chemical Society* **2015**, *137*, 289–298.

[11] E. Morra, S. Maurelli, M. Chiesa, E. Giamello, "Rational design of engineered multifunctional heterogeneous catalysts. The role of advanced EPR techniques", *Topics in Catalysis* **2015**, *58*, 783–795.

[12] S. Van Doorslaer, "Understanding heme proteins with hyperfine spectroscopy", *Journal of Magnetic Resonance* **2017**, *280*, 79–88.

[13] V. Hoeke, L. Tociu, D. A. Case, L. C. Seefeldt, S. Raugei, B. M. Hoffman, "High-resolution ENDOR spectroscopy combined with quantum chemical calculations reveals the structure of nitrogenase Janus intermediate E_4 (4H)", *Journal of the American Chemical Society* **2019**, *141*, 11984–11996.

[14] F. Hecker, J. Stubbe, M. Bennati, "Detection of Water Molecules on the Radical Transfer Pathway of Ribonucleotide Reductase by ^{17}O Electron–Nuclear Double Resonance Spectroscopy", *Journal of the American Chemical Society* **2021**, *143*, 7237–7241.

[15] G. Rao, N. Chen, D. A. Marchiori, L.-P. Wang, R. D. Britt, "Accumulation and Pulse Electron Paramagnetic Resonance Spectroscopic Investigation of the 4-Oxidobenzyl Radical Generated in the Radical S-Adenosyl-l-methionine Enzyme HydG", *Biochemistry* **2022**, *61*, 107–116.

[16] A. Meyer, S. Dechert, S. Dey, C. Höbartner, M. Bennati, "Measurement of Angstrom to Nanometer Molecular Distances with ^{19}F Nuclear Spins by EPR/ENDOR Spectroscopy", *Angewandte Chemie International Edition* **2020**, *59*, 373–379.

[17] N. Asanbaeva, A. Sukhanov, A. Diveikina, O. Rogozhnikova, D. Trukhin, V. Tormyshev, A. Chubarov, A. Maryasov, A. Genaev, A. Shernyukov, G. Salnikov, A. Lomzov, D. Pyshnyi, E. Bagryanskaya, "Application of W-band ^{19}F electron nuclear double resonance (ENDOR) spectroscopy to distance measurement using a trityl spin probe and a fluorine label", *Physical Chemistry Chemical Physics* **2022**, *24*, 5982–6001.

[18] M. Judd, E. H. Abdelkader, M. Qi, J. R. Harmer, T. Huber, A. Godt, A. Savitsky, G. Otting, N. Cox, "Short-range ENDOR distance measurements between Gd (III) and trifluoromethyl labels in proteins", *Physical Chemistry Chemical Physics* **2022**, *24*, 25214–25226.

[19] A. Meyer, A. Kehl, C. Cui, F. A. Reichardt, F. Hecker, L.-M. Funk, M. K. Ghosh, K.-T. Pan, H. Urlaub, K. Tittmann, J. Stubbe, M. Bennati, "^{19}F Electron-Nuclear Double Resonance Reveals Interaction between Redox-Active Tyrosines across the α/β Interface of E. coli Ribonucleotide Reductase", *Journal of the American Chemical Society* **2022**, *144*, 11270–11282.

[20] E. Ravera, L. Gigli, L. Fiorucci, C. Luchinat, G. Parigi, "The evolution of paramagnetic NMR as a tool in structural biology", *Physical Chemistry Chemical Physics* **2022**, *24*, 17397–17416.

[21] E. Matei, A. M. Gronenborn, "^{19}F Paramagnetic Relaxation Enhancement: A Valuable Tool for Distance Measurements in Proteins", *Angewandte Chemie International Edition* **2016**, *55*, 150–154.

[22] T. Müntener, D. Joss, D. Häussinger, S. Hiller, "Pseudocontact shifts in biomolecular NMR spectroscopy", *Chemical Reviews* **2022**, *122*, 9422–9467.

[23] M. Seal, W. Zhu, A. Dalaloyan, A. Feintuch, A. Bogdanov, V. Frydman, X.-C. Su, A. M. Gronenborn, D. Goldfarb, "Gd (III)-^{19}F Distance Measurements of Proteins in Cells by Electron-Nuclear Double Resonance", *Angewandte Chemie* **2023**.

[24] M. Brustolon, A. Maniero, M. F. Ottaviani, M. Romanelli, U. Segre, "Proton hyperfine tensors in nitroxide radicals", *Journal of physical chemistry* **1990**, *94*, 6589–6594.

[25] I. Tkach, I. Bejenke, F. Hecker, A. Kehl, M. Kasanmascheff, I. Gromov, I. Prisecaru, P. Höfer, M. Hiller, M. Bennati, "^{1}H high field electron-nuclear double resonance spectroscopy at 263 GHz/9.4 T", *Journal of Magnetic Resonance* **2019**, *303*, 17–27.

[26] A. Kehl, M. Hiller, F. Hecker, I. Tkach, S. Dechert, M. Bennati, A. Meyer, "Resolution of chemical shift anisotropy in ^{19}F ENDOR spectroscopy at 263 GHz/9.4 T", *Journal of Magnetic Resonance* **2021**, *333*, 107091.

[27] S. Pribitzer, D. Mannikko, S. Stoll, "Determining electron-nucleus distances and Fermi contact couplings from ENDOR spectra", *Physical Chemistry Chemical Physics* **2021**, *23*, 8326–8335.

[28] Y. Pokern, B. Eltzner, S. F. Huckemann, C. Beeken, J. Stubbe, I. Tkach, M. Bennati, M. Hiller, "Statistical analysis of ENDOR spectra", *Proceedings of the National Academy of Sciences* **2021**, *118*, e2023615118.

[29] M. Hiller, I. Tkach, H. Wiechers, B. Eltzner, S. Huckemann, Y. Pokern, M. Bennati, "Distribution of H-β Hyperfine Couplings in a Tyrosyl Radical Revealed by 263 GHz ENDOR Spectroscopy", *Applied Magnetic Resonance* **2022**, *53*, 1015–1030.

[30] K. Herb, R. Tschaggelar, G. Denninger, G. Jeschke, "Double resonance calibration of g factor standards: Carbon fibers as a high precision standard", *Journal of Magnetic Resonance* **2018**, *289*, 100–106.

[31] E. Bordignon in D. Goldfarb, S. Stoll, *EPR Spectroscopy, Fundamentals and Methods*, (Eds.: D. Goldfarb, S. Stoll), John Wiley & Sons, Chichester, **2018**, Chapter 14, pp. 277–301.

[32] K. Ford, *From Kolmogorov's theorem on empirical distribution to number theory*, Springer, **2007**, pp. 97–108.

[33] M. Bennati in D. Goldfarb, S. Stoll, *EPR Spectroscopy, Fundamentals and Methods*, (Eds.: D. Goldfarb, S. Stoll), John Wiley & Sons, Chichester, **2018**, Chapter 5, pp. 81–94.

[34] A. Schweiger, G. Jeschke, *Principles of Pulse Electron Paramagnetic Resonance*, Oxford University Press, **2001**.

[35] A. Grupp, M. Mehring in L. Kevan, M. K. Bowman, *Modern Pulsed and Continuous-Wave Electron Spin Resonance*, John Wiley & Sons, New York, **1990**, pp. 195–229.

[36] C. Gemperle, A. Schweiger, "Pulsed electron-nuclear double resonance methodology", *Chemical Reviews* **1991**, *91*, 1481–1505.

[37] G. Wang, G. R. Hanson, "A New Method for Simulating Randomly Oriented Powder Spectra in Magnetic Resonance: The Sydney Opera House (SOPHE) Method.", *Journal of Magnetic Resonance Series A* **1995**, *117*, 1–8.

[38] F. Archetti, A. Candelieri, *Bayesian Optimization and Data Science*, Springer International Publishing, **2019**.

[39] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, **1999**.

[40] H. Raber, M. Mehring, "19F Chemical shift tensor in fluorobenzene compounds", *Chemical Physics* **1977**, *26*, 123–130.

[41] S. Jung, A. Schwartzman, D. Groisser, "Scaling-Rotation Distance and Interpolation of Symmetric Positive-Definite Matrices", *SIAM Journal on Matrix Analysis and Applications* **2015**, *36*, 1180–1201.

[42] D. Goldfarb, S. Stoll, *EPR Spectroscopy, Fundamentals and Methods*, John Wiley & Sons, Chichester, **2018**.

Bayesian Optimization to Estimate Hyperfine Couplings from ^{19}F ENDOR Spectra

Supplementary Information

H. Wiechers[1], A. Kehl[2], M. Hiller[2], B. Eltzner[2], S. F. Huckemann[1],
A. Meyer[2], I. Tkach[2], M. Bennati[2,3,*] and Y. Pokern[4,*]
[1]Felix-Bernstein-Institute for Mathematical Statistics, Georg-August-University
Göttingen, 37077 Göttingen, Germany.
[2]Max Planck Institute for Multidisciplinary Sciences,
37077 Göttingen, Germany.
[3]Department of Chemistry, Georg-August University of Göttingen,
Tammannstr. 2, Göttingen, Germany
[4]Department of Statistical Science, University College London,
London WC1E 6BT, United Kingdom.
[*]Corresponding authors

A Statistical Drift Model

A.1 Phasing Procedure to Extract Spectrum I from ENDOR Effect κ

The direction in the complex plane in which κ contains the spectrum can be found in several different ways. As explained in the main text, we opted to select λ_{opt} such that $\sum_\nu \Re\left(e^{i\lambda_{\text{opt}}}\kappa_\nu\right)^2$ was maximal. The signal in $\Im\left(e^{i\lambda_{\text{opt}}}\kappa_\nu\right)$ is not necessarily pure white noise but may contain some artefact such as those identified in [28] where a cosine-like perturbation was observed. However, in the ^{19}F data, the range of RF frequencies is much narrower than that considered in [28]. Hence, the cosine-like perturbation is hard to discern because only a fraction of its period is captured: see panel B of Figure D3 in the main text.

A.2 Goodness of Fit

Goodness of fit was checked via examining the residuals $\hat{\epsilon}_{b,\nu} = Y_{b,\nu} - \hat{\psi}_b - \hat{\phi}_b\hat{\kappa}_\nu$. These were pooled over all batches $b \in \{1,\ldots,B\}$ and RF frequencies $\nu \in \{1,\ldots,N\}$ and their real and imaginary parts were submitted to a KS test of Gaussianity whose results are provided in Table D5.

Compound **1**			Compound **2**		
Orientation	Real	Imag	Orientation	Real	Imag
g_x	0.99	1.00	g_x	0.93	0.41
g_{xy}	0.70	0.34	g_{xy}	0.72	0.66
g_y	0.64	0.80	g_y	0.97	0.36
g_{yz}	0.79	0.21			
g_z	0.58	0.96	g_z	0.86	0.71

Table D5: The $p-$values from KS tests of Gaussianity applied to the real and imaginary parts of the residuals $R_{b,\nu}$ (pooled over b and ν) resulting from the SDM applied to all measurements.

A.3 Bootstrap Sample Distribution

The covariance matrix of the bootstrap samples of the spectrum, $\{I^{*,j}\}_{j=1}^J$, is obtained according to

$$\chi_{\nu,\nu'} = \frac{1}{J-1}\sum_{j=1}^{J}\left(I_\nu^{*,j} - \frac{1}{J}\sum_{i=1}^{J} I_\nu^{*,i}\right)\left(I_{\nu'}^{*,j} - \frac{1}{J}\sum_{i=1}^{J} I_{\nu'}^{*,i}\right) \qquad \nu,\nu' \in \{1,\ldots,N\}. \quad (11)$$

It is plotted in panel A of Figure D10 and its eigenvalues are shown in panel B of that figure. The smallest eigenvalue is practically zero reflecting the constraint that the data are centred. The next distinct eigenvalue reflects the constraint of unit size κ, so that spectral deviations are mostly orthogonal to the mean spectrum.

In order to check the distribution of the deviation of the bootstrap samples $I_\nu^{*,j}$ of the spectrum from the estimated spectrum $\hat{I}_\nu$, we plot a histogram of $I_1^{*,j} - \hat{I}_1$ in panel A of Figure D11 and performed KS goodness of fit tests for a Gaussian distribution for all frequencies ν reporting the results in panels B and C of Figure D11 for bootstrap sample size $J = 10^4$. None of the $p-$values obtained are smaller than 10^{-3} so that the null hypothesis of Gaussian distribution cannot be rejected at the Bonferroni-corrected standard significance level of $\frac{0.05}{N} \approx 1.5 \times 10^{-4}$.

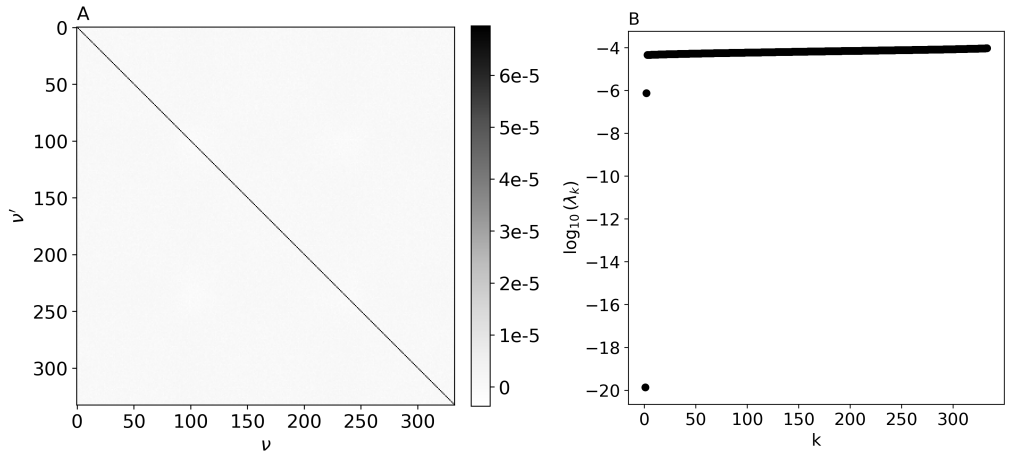


Figure D10: A: the sample covariance matrix $(\chi_{\nu,\nu'})_{\nu,\nu'=1}^{N,N}$ of the bootstrap spectra $\{I^{*,j}\}_{j=1}^{J}$ (see Equation (11)). Note that the shading indicating the value of covariance entries starts at a negative value corresponding to white, so the off-diagonal entries are approximately zero corresponding to light grey. B: the corresponding sorted eigenvalues $\{\lambda_k\}_{k=0}^{332}$ (enumerated by k on the horizontal axis) of the sample covariance matrix plotted on a logarithmic scale.

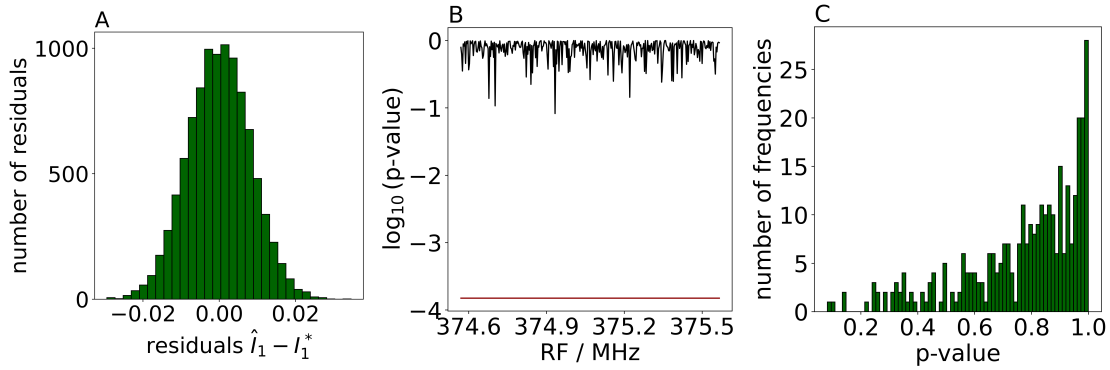


Figure D11: A: checking for Gaussian distribution of the bootstrap residuals $\hat{I}_1 - I_1^*$ for orientation g_x of compound **1** using a histogram. B: p-values of KS tests for Gaussianity of $\hat{I}_\nu - I_\nu^*$ for $\nu \in \{1, \ldots, N\}$. For no frequency is the p-value below the Bonferroni-corrected critical value of $0.05/N$ (indicated by horizontal red line) so that the null hypothesis of Gaussianity cannot be rejected. C: histogram of the p-values calculated in the middle column. Non-uniformity of $p-$values arises from standardization of bootstrap residuals prior to testing.

B Quasi-Bootstrapping for ENDOR Spectra

Here, we briefly describe an alternative method to assess the stochastic error of ENDOR spectra obtained via the averaging method that does not require an SDM. This method allows approximation of the covariance matrix of the ENDOR spectrum needed for stochastic error of the spin and experimental parameters.

This *quasi-bootstrap* approach, summarized in Algorithm 1, uses the difference between the spectrum $\hat{I}$ and its smoothed version $\tilde{I}$ as an indication of the standard deviation s of the stochastic measurement error. The method then samples new independent and identically distributed (i.i.d.) Gaussian noise with this standard deviation and adds it (pointwise) to the smoothed spectrum. Following normalization, we refer to the resulting $I^{*,k}$ as a quasi-bootstrapped spectrum. Equa-

tion (11) can then be used. In choosing a smoothing method (*e.g.* a Savitzky-Golay filter) as well as its parameters, the aim should be to include as much of the true spectrum in $\tilde{I}$ as possible while leaving most of the noise in $\hat{I} - \tilde{I}$.

For Figure D4 in the main text, we chose a Savitzky-Golay filter with parameters window length of 37 and polyorder of 4. This choice was adapted to our spectra and has to be carefully selected based on the properties of the spectra. We then used the standard deviation s as computed in step 3 of Algorithm 1 to compute approximate 95% confidence intervals given by $\hat{I}_\nu \pm 1.96s$ for each RF frequency ν. Since only the approximate confidence intervals for the spectra rather than the covariance matrix from Equation (11) were required for Figure D4, steps 4 to 8 of Algorithm 1 were omitted.

Algorithm 1 Quasi-Bootstrap for the ENDOR Spectrum

1: Input: spectrum $\hat{I}$
2: Smooth $\hat{I}$ to get the smoothed spectrum $\tilde{I}$.
3: Determine the empirical variance $s^2 = \frac{1}{N}\sum_{\nu=1}^{N}\left(\hat{I}_\nu - \tilde{I}_\nu\right)^2$ and perform a quasi-bootstrap:
4: **for** $k \in \{1, \ldots, K\}$ **do**
5: Sample $I^{**,k}$ by $I_\nu^{**,k} = \tilde{I}_\nu + \epsilon_\nu^{*,k}$ where $\epsilon_\nu^{*,k} \overset{i.i.d.}{\sim} \mathcal{N}(0, s^2)$.
6: Normalize to mean zero and sum of squares one: $I_\nu^{*,k} = \frac{I_\nu^{**,k} - \sum_{\nu'} I_{\nu'}^{**,k}}{\sqrt{\sum_{\nu''}\left(I_{\nu''}^{**,k} - \sum_{\nu'} I_{\nu'}^{**,k}\right)^2}}$
7: **end for**
8: **return** The quasi-bootstrap sample $\{I^{*,k}\}_{k=1}^{K}$.

C Accuracy of Simulation for Comparison of Computational Time

To compare (Section 3.3 in the main text) the speed of the accelerated code SimSpec with the code underlying [26], to which we refer as Sim, a similar level of accuracy in the simulated spectrum must be achieved. Additionally, gradient-based optimization is greatly hampered by computing spectra with insufficient accuracy (see Figure D12) due to the ruggedness of the error. Hence, we consider accuracy of spectral simulation in this section.

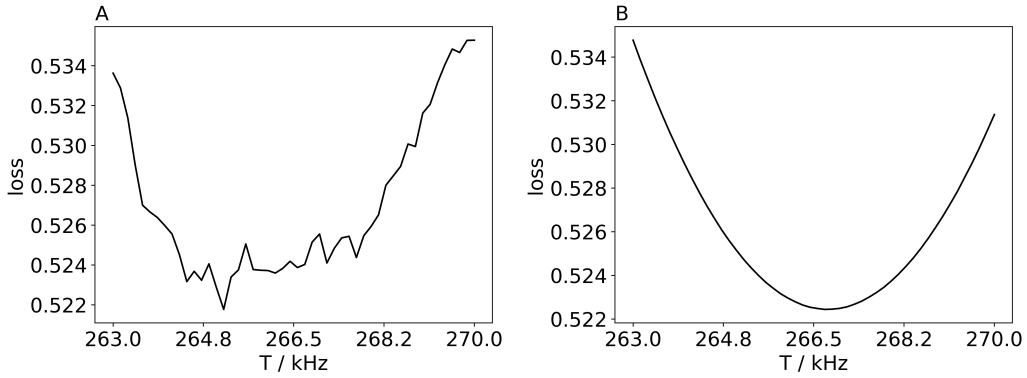


Figure D12: Impact of the choice of the simulation parameters affecting computational time on the loss using SimSpec (summed squared difference of simulated vs measured spectrum) as a function of the HF parameter T. There are two different accuracy settings: A: `rfBinSize_MHz`=0.003 and `Nknots` = 50. B: `rfBinSize_MHz`=$\frac{0.003}{100}$ and `Nknots` = 150. All other parameter values were chosen from [26], Table 1.

In both spectrum simulation functions (SimSpec and Sim), there are two parameters that significantly affect computational time. The first parameter determines (depending on the selected grid) how finely the sphere is partitioned in powder averaging.
The second parameter is `rfBinSize_MHz` which specifies how finely the RF axis is partitioned. Four different grids have been implemented in our simulation software: 'Fibonacci', 'Legacy', 'Polar' and 'SOPHE'. For the SOPHE grid, which was implemented in analogy to Easyspin, only the parameter `Nknots` is required. Increasing this parameter leads to greater accuracy and computational time in the calculation of the spectrum. In a first experiment, we simulate a *reference spectrum* for each of the five orientations from compound **1** with the very high and computationally costly value `Nknots`=400 using SimSpec. Then, various smaller values for `Nknots` are chosen and the loss (summed squared difference of spectra as in Equation (8)) relative to the reference spectrum (see panel A of Figure D13) and the computation time for computing the five spectra are determined (see panel B of Figure D13). The parallelized variant with pre-calculation (displayed in panel B of Figure D13) is by far the fastest variant, since the SOPHE grid is simply loaded and does not have to be computed (the grid does not change when changing the parameters to be optimized). The pre-calculation method gains its speed advantage from the fact that the Hamiltonian only needs to be re-diagonalized when the magnetic field strength is changed.
In a second experiment, we proceeded analogously for the parameter `rfBinSize_MHz`. We fixed the value `Nknots`=150 and simulated a reference spectrum with the `rfBinSize_MHz`=$\frac{0.003}{200}$ using SimSpec. Then, various larger values for `rfBinSize_MHz` were chosen and the loss relative to the reference spectrum (see panel A of Figure D14) and the computation time for computing the five spectra were determined (see panel B of Figure D14).

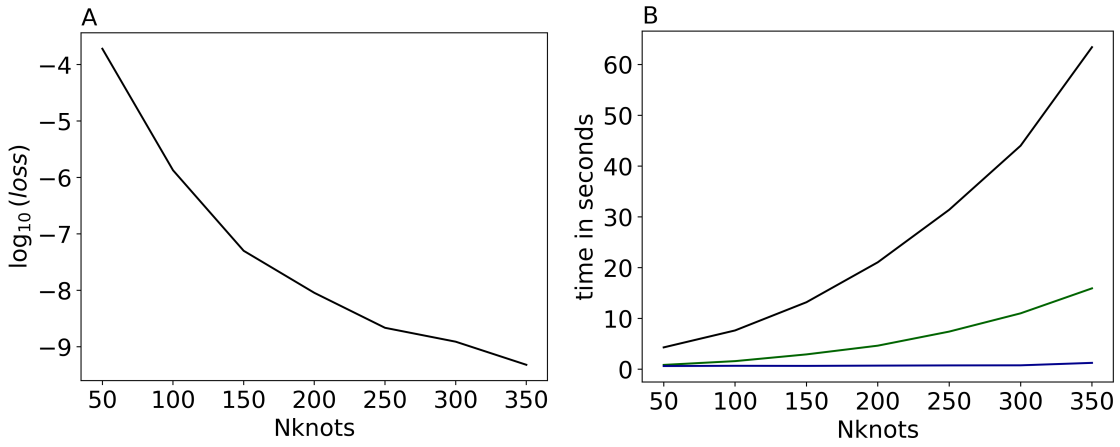


Figure D13: A: the loss (summed squared differences of simulated spectrum with given `Nknots` vs the reference spectra) using SimSpec for different `Nknots` values. The reference spectra were calculated with `Nknots`=400. B: the computation time for calculating the spectra for the five orientations consecutively (black), parallelized based on multiprocessing (green), parallelized based on multiprocessing with additional pre-calculation (blue). For all calculations we used `rfBinSize_MHz` = $\frac{0.003}{100}$.

Finally, for comparison, we simulated spectra using Sim where `Ntheta` (set to equal `Nphimax`) was chosen for the polar grid to yield approximately the same accuracy of the simulated spectrum as the accuracy attained in the SOPHE grid with `Nknots`= 150 and `rfBinSize_MHz`= 0.003, see Figure D15. As usual, the accuracies are considered as the loss relative to the reference spectrum for each grid, respectively. Using these values of `Ntheta` and `Nphimax` makes for a fair comparison of computational time across different grids because similar errors in the computed spectra are achieved.

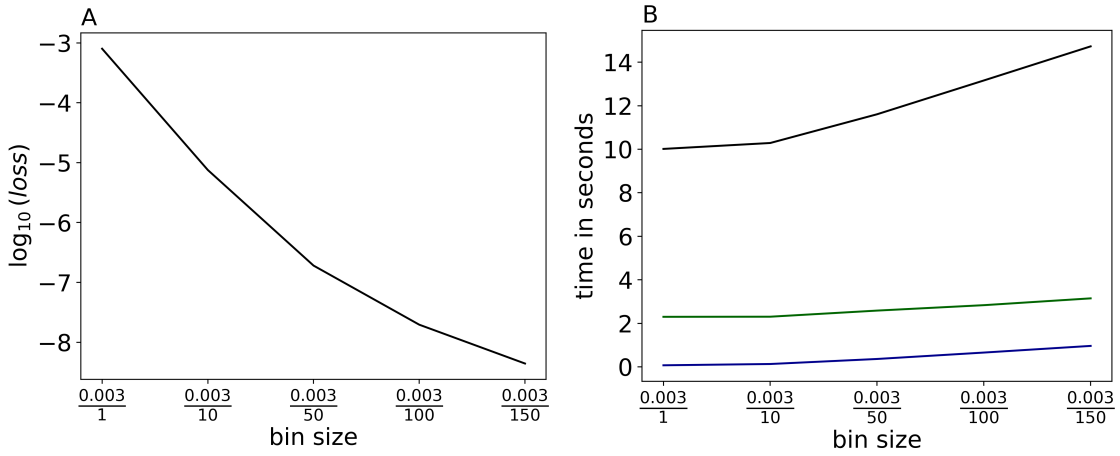


Figure D14: A: the loss (summed squared differences of simulated spectrum with given `bin size` vs the reference spectra) using SimSpec for different `rfBinSize_MHz` values. The reference spectra were calculated with `rfBinSize_MHz`=$\frac{0.003}{200}$. B: the computation time for calculating the five orientations consecutively (black), parallelized based on multiprocessing (green), parallelized based on multiprocessing with additional pre-calculation (blue). For all calculations we used `Nknots` = 150.

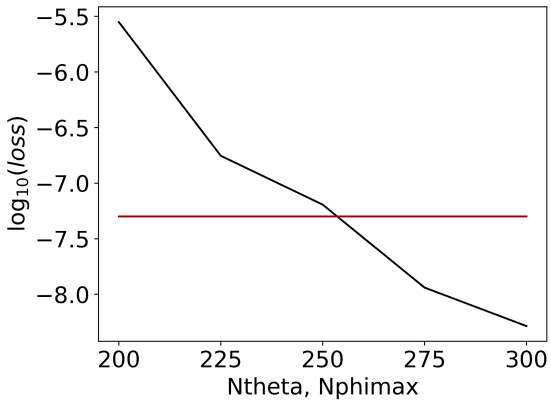


Figure D15: The loss using Sim relative to the reference spectra for different `Ntheta=Nphimax` values. The reference spectra were calculated with `Ntheta=Nphimax`=700. The horizontal line indicates the loss achieved by SimSpec using the SOPHE grid (relative to the SimSpecSOPHE reference spectrum) with the settings used for the computational time evaluation in Table D1 in the main text.

D Symmetries and Uncertainty Quantification

In tensors with considerable uncertainty such as the CS tensors reported in Table D3, specifying confidence regions for Euler angles and eigenvalues in a straightforward manner may lead to unnecessarily large confidence regions. Taking symmetries of the tensors described by those Euler angles and eigenvalues into account and using the joint uncertainty of these quantities instead yields tighter confidence regions. In the following, we show in detail how tensor symmetries are used in an alignment procedure to obtain tight confidence regions while maintaining unimodal uncertainty distributions of the eigenvalues.

For a symmetric 3x3 matrix A with entries $a_{i,j} \in \mathbb{R}$ and distinct eigenvalues, the eigenvalue decomposition $A = U(A)D(A)U(A)^T$ with $U(A) \in SO(3)$ ($SO(3)$ is the set of rotation matrices

in dimension 3) and $D(A) = \begin{pmatrix} \lambda_1(A) & & \\ & \lambda_2(A) & \\ & & \lambda_3(A) \end{pmatrix} \in \mathrm{Diag}(3)$ ($\mathrm{Diag}(3)$ is the set of 3×3 diagonal matrices) is not unique. Even fixing the eigenvalues to be in ascending order so that $\lambda_1(A) < \lambda_2(A) < \lambda_3(A)$ does not lead to uniqueness as the decomposition is invariant under a choice of $U' \in \mathcal{M}$ in $A = UU'D(U')^T U^T$, where $\mathcal{M}$ is given in Equation (7). When aligning bootstrap samples $A^{*,k}$ from the sampling distribution of some estimator $\hat{A}$ with $\hat{A}$ in order to understand the uncertainty of the estimate, one may therefore wish to choose one of these four matrices to make the Euler angles implied by $U(A^{*,k})$ as similar to those implied by $U(\hat{A})$ as possible. Additionally, if some eigenvalues, say $\lambda_1(A^{*,k})$ and $\lambda_2(A^{*,k})$, are close, then it may be preferable to swap these eigenvalues to again achieve a better match of implied Euler angles at the cost of a worse match of eigenvalues. In order carry this out in a systematic fashion, we refer to an adaptation of the distance in [41] to the case of symmetric (rather than positive-definite symmetric) matrices. This strikes a compromise between the angular deviation and the deviation of eigenvalues which is apparent from its definition

$$d(A,B)^2 = \min_{\{U,V \in SO(3),\, D,\Lambda \in \mathrm{Diag}(3):\, UDU^T = A \text{ and } V\Lambda V^T = B\}} \left(k \left\| \log(UV^T) \right\|_F^2 + \|D - \Lambda\|_F^2 \right), \tag{12}$$

where $\|A\|_F^2 = \sum_{i,j=1}^3 a_{i,j}^2$ is the square of the Frobenius norm and $k \in \mathbb{R}_{>0}$ is a positive tuning parameter. Large choices of k correspond to emphasizing small differences of the rotational part and hence the Euler angles whereas small k corresponds to emphasizing a close match of the eigenvalues.

In quantifying the uncertainty for the $\boldsymbol{\sigma}_{^{19}F}$ tensors in compound **2**, we sampled a large number K of realizations $\{\boldsymbol{\sigma}_{^{19}F}^{*,k}\}_{k \in \{1,\ldots,K\}}$ of the Gaussian distribution of the matrix entries in the rightmost part of Equation (6) implied by the linear error propagation in Equation (10). We then aligned these samples with $\hat{\boldsymbol{\sigma}}_{^{19}F}$ (with a fixed choice of $U(\hat{\boldsymbol{\sigma}}_{^{19}F})$ in Equation (12)) by minimizing $d(\hat{\boldsymbol{\sigma}}_{^{19}F}, \boldsymbol{\sigma}_{^{19}F}^{*,k})^2$ from Equation (12) and extracted Euler angles $\alpha^{*,k}, \beta^{*,k}, \gamma^{*,k}$ from V that attained the minimum in Equation (12). We chose a value of k so as to ensure that only the two small eigenvalues of the CS tensors are swapped, maintaining a unimodal distribution of the eigenvalues. Approximate confidence regions for the Euler angles were then chosen as those regions of maximal kernel-density-estimated probability density that accumulate 95% of probability. Using the alignment described here resulted in unimodal distributions for all Euler angles but one, as well as unimodal distributions for all eigenvalues. Omission of the described alignment would have resulted in larger approximate confidence regions for all Euler angles.

E Bayesian Optimization

Here, we provide an overview of Bayesian optimization as well as details on our gradient-based refinement. For more detail and background on the concepts mentioned, see [38]. Bayesian optimization approximately solves the global minimization[‡] problem $\theta^* \in \mathrm{argmin}_{\theta \in \Theta} \mathcal{L}(\theta)$ for an objective function $\mathcal{L} : \mathbb{R}^m \supset \Theta \to \mathbb{R}$. It takes a more global view of the parameter space Θ than gradient-based methods and necessitates neither computation nor approximation of derivatives of $\mathcal{L}$ with respect to θ. This is achieved in an iterative approach using a Gaussian process statistical model $\hat{\mathcal{L}}(\theta)$ for the objective function $\mathcal{L}$ and an acquisition function a to determine which $\theta \in \Theta$ should be tried next. The Gaussian process $\hat{\mathcal{L}}$ is specified through two quantities: its prior expected value and its prior covariance kernel which are used to specify *a priori* beliefs about the objective

[‡]Mimimization of $\mathcal{L}$ is equivalent to maximization of $-\mathcal{L}$, so, for ease of presentation, we consider only minimization here.

function $\mathcal{L}$, *i.e.* beliefs held before any evaluations of $\mathcal{L}$ are available. Its expected value (equal to the average over an infinite number of samples of the Gaussian process) at the point $\theta \in \Theta$ is denoted by $\mu^{(0)}(\theta) = \mathbb{E}^{(0)}[\hat{\mathcal{L}}(\theta)]$ which we chose to be identically zero as this is a standard choice, although it will in general depend on the point θ at which $\hat{\mathcal{L}}$ is evaluated and $\mu^{(0)}$ is therefore known as the prior *mean function*. The covariance of its value at $\theta \in \Theta$ with its value at $\theta' \in \Theta$, denoted as $c^{(0)}(\theta, \theta') = \mathrm{Cov}^{(0)}\left(\hat{\mathcal{L}}(\theta), \hat{\mathcal{L}}(\theta')\right)$ was chosen to be a Matérn kernel of order $\nu = 1.5$ again following a standard choice expressing the prior belief that $\mathcal{L}$ is at least once differentiable with respect to θ. Naturally, this covariance will generally depend on the two points θ and θ' and $c^{(0)}$ is therefore known as the prior *covariance kernel*. In specifying the prior mean function and covariance kernel, the exponent $^{(0)}$ denotes that zero observations of $\mathcal{L}$ have been incorporated so far and expectations and covariances are taken over all possible realizations of the Gaussian process $\hat{\mathcal{L}}$. $\mathcal{L}$ is then computed at a number of parameter values $\theta^{(1)}, \ldots, \theta^{(k)} \in \Theta$ that are selected randomly in the parameter space. Bayes' theorem is used to update the Gaussian process combining the prior mean and covariance with the observations $\mathcal{L}(\theta^{(1)}), \ldots, \mathcal{L}(\theta^{(k)})$ to obtain the posterior Gaussian process which is specified via its mean function $\mu^{(k)}$ (where $\mu^{(k)}(\theta) = \mathbb{E}^{(k)}[\hat{\mathcal{L}}(\theta)]$) and its covariance kernel $c^{(k)}$ (where $c^{(k)}(\theta, \theta') = \mathrm{Cov}^{(k)}\left(\hat{\mathcal{L}}(\theta), \hat{\mathcal{L}}(\theta')\right)$) with the exponent $^{(k)}$ denoting that k observations have been taken into account. Properties of Gaussian processes ensure that, in order to compute the posterior mean $\mu^{(k)}(\theta)$ and the posterior covariance $c^{(k)}(\theta, \theta')$ at fixed points $\theta, \theta' \in \Theta$, only matrix operations involving matrices of size of order $\mathcal{O}(k \times k)$ are required. At points θ close to one of the $\theta^{(1)}, \ldots, \theta^{(k)}$, the variance $c^{(k)}(\theta, \theta)$ of the posterior Gaussian process is small (it is indeed zero at all $\theta^{(1)}, \ldots, \theta^{(k)}$ so that $\mathcal{L}(\theta) = \hat{\mathcal{L}}(\theta)$ when $\theta \in \{\theta^{(1)}, \ldots, \theta^{(k)}\}$), whereas in regions of Θ far away from all the $\theta^{(1)}, \ldots, \theta^{(k)}$ tried so far, the variance is large.
The problem of how to choose $\theta^{(k+1)}$ is at the core of the Bayesian Optimization idea: should we choose it in the vicinity of some of the $\theta^{(1)}, \ldots, \theta^{(k)}$ to minimize $\mathcal{L}(\theta^{(k+1)})$ locally ("exploitation") or should we examine a region of Θ that has received less attention so far ("exploration")? The purpose of the acquisition function is to strike a balance between exploitation and exploration. We use a standard choice which is the expected improvement, *i.e.* our acquisition function $a^{(k)}$ is the expected value

$$a^{(k)}(\theta) = \mathbb{E}^{(k)}\left[\max\left(\min_{\theta' \in \{\theta^{(1)}, \ldots, \theta^{(k)}\}} \hat{\mathcal{L}}(\theta') - \hat{\mathcal{L}}(\theta), 0\right)\right].$$

Based on the current posterior Gaussian process, it assigns a value to each $\theta \in \Theta$ that is thought of as that point's propensity to be the new minimizer of $\mathcal{L}$ among the $\theta^{(1)}, \ldots, \theta^{(k)}, \theta$. An approximate maximizer $\theta \in \Theta$ of the acquisition function is used as $\theta^{(k+1)}$ and the objective function $\mathcal{L}(\theta^{(k+1)})$ is then computed at this newly chosen point. This new pair of values $\theta^{(k+1)}, \mathcal{L}(\theta^{(k+1)})$ is then incorporated to build the updated posterior distribution with mean function $\mu^{(k+1)}$ and covariance kernel $c^{(k+1)}$ and the process continues in an iterative fashion finding $\theta^{(k+2)}$ approximately maximizing $a^{(k+1)}$, see Algorithm 2 for an overview of the algorithm.
The reasons for replacing one optimization problem (that of minimizing the objective function $\mathcal{L}$) by yet another optimization problem (that of maximizing the acquisition function a) are: firstly, only an approximate maximizer of the acquisition function is needed whereas an exact global minimum of the objective function is sought and, secondly, a well-chosen acquisition function, being a simple functional of a Gaussian process, is *much faster* to compute than the objective function. One potential complication is that exploration would proceed indefinitely if the parameter space Θ was unbounded whence bounds need to be specified for all parameters.
Bayesian optimization is used to approximately locate the global minimizer of $\mathcal{L}$ for a set maximum number of observations n to limit computational cost. As can be seen in Figure D16, which shows an illustration of the improvement in ENDOR spectra when going from incorporating $n = 20$ observations to incorporating $n = 300$ observations, the approximate minimizer

Algorithm 2 Bayesian Optimization

1: Define a Gaussian process prior $\hat{\mathcal{L}}$ on $\mathcal{L}$.
2: Sample initial points $\theta^{(1)}, \ldots, \theta^{(k)} \in \Theta$ uniformly distributed on Θ and compute $\mathcal{L}(\theta^{(1)}), \ldots, \mathcal{L}(\theta^{(k)})$.
3: **for** $i \in \{k, \ldots, n-1\}$ **do**
4: Update the Gaussian process to posterior based on $\left(\theta^{(1)}, \mathcal{L}(\theta^{(1)})\right), \ldots, \left(\theta^{(i)}, \mathcal{L}(\theta^{(i)})\right)$.
5: Compute $\theta^{(i+1)} \in \Theta$ by approximately optimizing the acquisition function $a^{(i)}(\theta)$.
6: Compute $\mathcal{L}(\theta^{(i+1)})$.
7: **end for**
8: Return the best solution $\theta^{(i^*)}$ where $i^* = \underset{i=1,\ldots,n}{\operatorname{argmin}} \mathcal{L}(\theta^{(i)})$.

obtained based on $n = 300$ observations is still imperfect. Following standard methodology, we then use a local optimization method to improve on this result. We choose BFGS which is a gradient-based quasi-Newton method that uses the objective function and its gradient to iteratively build an approximation to the Hessian of $\mathcal{L}$ and uses this to generate iterative improvements to the approximate minimizer. When started in the vicinity of a local minimum, this method can exhibit faster convergence to that local minimum than Bayesian optimization because it focuses on the local problem.

When supplying BFGS with gradient information, finite difference approximations to the gradient are usually used: $\frac{\partial \mathcal{L}}{\partial \theta_j} \approx \frac{\mathcal{L}(\theta + \epsilon e_j) - \mathcal{L}(\theta)}{\epsilon}$, with e_j the j'th standard basis vector. Here, the increment ϵ needs to be tuned: if ϵ is too large, the difference quotient deviates from the derivative. If ϵ is too small, the error in computing each ENDOR spectrum, while small in absolute terms, is large compared to the difference in the numerator of the finite difference formula. See Figure D17 for a visualization of this issue. We therefore employed increments that were roughly manually tuned, separately for each parameter.

F Optimization results

F.1 Compound 1: minimum #2

Spin Parameters		Experimental Parameters	
T / kHz	267.1 ± 0.9	B_0 / G : g_x	93628.6 ± 0.5
a_{iso} / kHz	-0.4 ± 1.0	B_0 / G : g_{xy}	93697.0 ± 0.7
$\alpha_A/^\circ$	-160.2 ± 0.6	B_0 / G : g_y	93937.1 ± 0.8
$\beta_A/^\circ$	-19.0 ± 0.4	B_0 / G : g_{yz}	93837.9 ± 0.7
$\tilde{\sigma}_{xx}$ / ppm	199 ± 6	B_0 / G : g_z	93943.1 ± 0.9
$\tilde{\sigma}_{yy}$ / ppm	281 ± 9	lw / kHz : g_x	15.2 ± 3.1
$\tilde{\sigma}_{zz}$ / ppm	393 ± 11	lw / kHz : g_{xy}	50.2 ± 5.0
$\alpha_\sigma/^\circ$	80 ± 5	lw / kHz : g_y	19.7 ± 2.5
$\beta_\sigma/^\circ$	61 ± 3	lw / kHz : g_{yz}	22.8 ± 2.7
$\gamma_\sigma/^\circ$	106 ± 2	lw / kHz : g_z	29.9 ± 3.2

Table D6: Estimated parameter values for minimum #2 in compound **1**

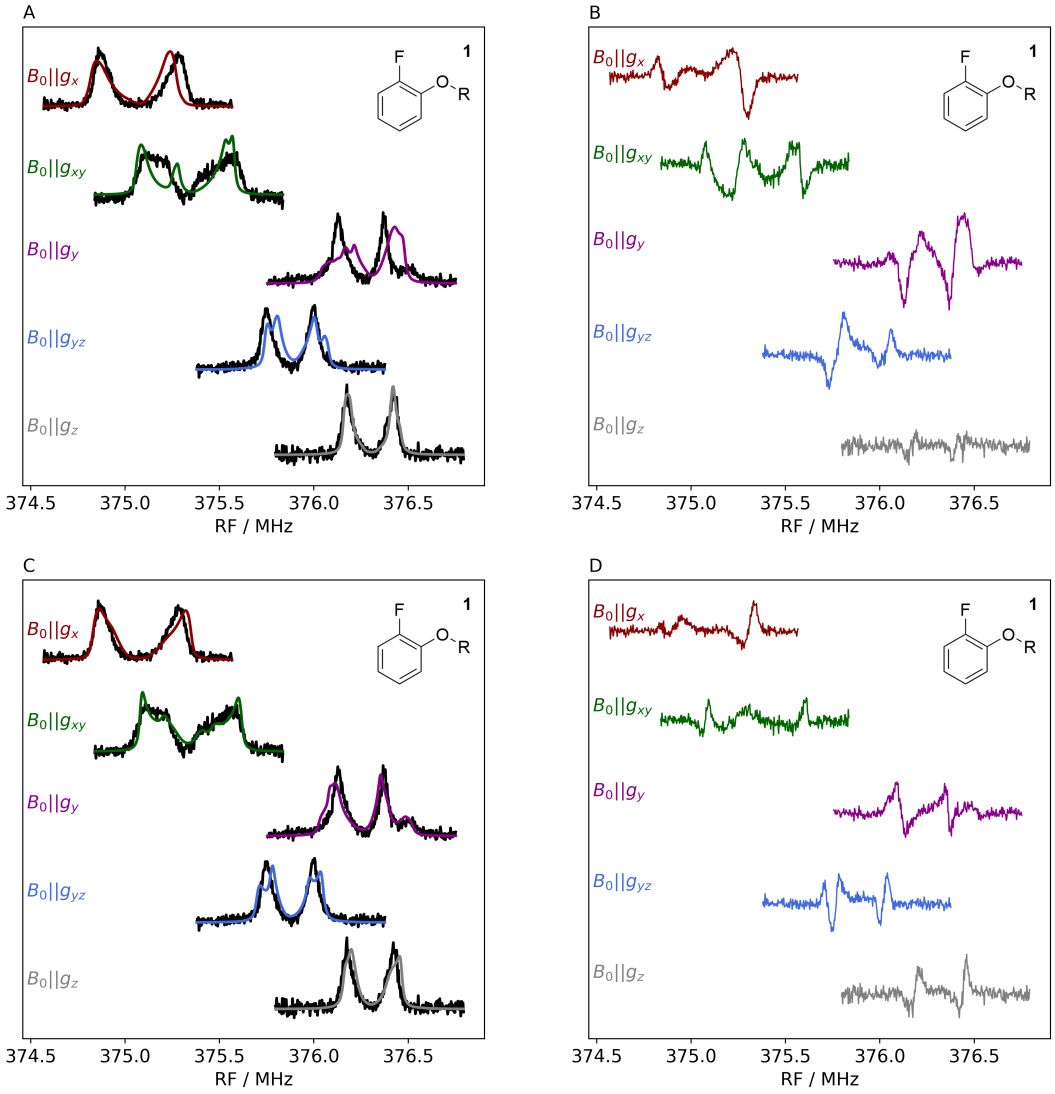


Figure D16: Spectra (left column) and spectral residuals (right column) after different iterations of Bayesian optimization from the minimization described in Section 4.2 in the main text. Top row: After 20 iterations. Bottom row: After 300 iterations.

F.2 Optimization for Compound 1 two g values

F.3 Compound 2

In this subsection, we report on those minima of the loss function $\mathcal{L}$ for compound **2** that have been identified but that have not been reported in detail in the main text. The minimum reported in detail in the main text is referred to as minimum #1 and Table D8 enumerates the additional minima #2, #3 and #4 and provides their spin and experimental parameter values.

In detail, we examined the uncertainty distributions of the CS tensors in compound **2** by creating a bootstrap sample of size K. We sampled $\left\{\left(\boldsymbol{\sigma}^{*,k}_{^{19}F_1}, \boldsymbol{\sigma}^{*,k}_{^{19}F_2}\right)\right\}_{k\in\{1,\ldots,K\}}$ from the Gaussian distribution with mean given by the minimum under investigation and covariance matrix $\mathrm{Cov}(\theta)$ given via Equation (10) for the matrix entries of these tensors using the parametrization given in the rightmost part of Equation (6). Here, K is chosen sufficiently large to yield a representative sample. For each of these samples, we determined the angle $\delta^{*,k}$ subtended by the dominant eigenvectors

Spin Parameters		Experimental Parameters	
T / kHz	266.2 ± 1.0	B_0 / G : g_x	93629.3 ± 0.75
a_{iso} / kHz	-1.3 ± 1.1	B_0 / G : g_{xy}	93697.6 ± 0.75
$\alpha_A/^\circ$	-160.9 ± 1.0	B_0 / G : g_y	93938.0 ± 0.91
$\beta_A/^\circ$	-19.6 ± 0.5	B_0 / G : g_{yz}	93838.7 ± 0.81
$\tilde{\sigma}_{xx}$ / ppm	201.3 ± 8.6	B_0 / G : g_z	93943.8 ± 1.1
$\tilde{\sigma}_{yy}$ / ppm	304.2 ± 12.4	lw / kHz : g_x	13.0 ± 3.3
$\tilde{\sigma}_{zz}$ / ppm	391.4 ± 9.5	lw / kHz : g_{xy}	51.6 ± 5.0
$\alpha_\sigma/^\circ$	93.8 ± 3.4	lw / kHz : g_y	19.2 ± 2.5
$\beta_\sigma/^\circ$	107.7 ± 4.5	lw / kHz : g_{yz}	23.4 ± 2.8
$\gamma_\sigma/^\circ$	107.7 ± 2.6	lw / kHz : g_z	31.2 ± 3.1

Table D7: Parameter values for the minimum shown in Figure D19 assuming contributions from two distinct g_x values $g_{x,1} = 2.00889$ and $g_{x,2} = 2.00835$, weighted 90% and 10%, respectively, in compound **1**. The loss value is 0.273.

$\sigma^{*,k}_{z,^{19}F_1}$ and $\sigma^{*,k}_{z,^{19}F_2}$ of $\boldsymbol{\sigma}^{*,k}_{^{19}F_1}$ and $\boldsymbol{\sigma}^{*,k}_{^{19}F_2}$, respectively, taking into account that $\sigma^{*,k}_{z,^{19}F_2}$ and $-\sigma^{*,k}_{z,^{19}F_2}$ are equivalent eigenvectors and hence restricting the range of values of $\delta^{*,k}$ to between 0 and 90 degrees. To assess whether this uncertainty distribution represented by the sample $\{\delta^{*,k}\}_{k\in\{1,\ldots,K\}}$ was compatible with the dominant eigenvectors being parallel, we computed the quotient of its implied probability density (estimated by kernel density estimation) and the density expected in the case of a uniformly random distribution of the eigenvectors on the sphere. The resulting relative densities are plotted in Figure D21 and show that for minimum #1, the maximum of the relative density is assumed at an angle of $\delta = 0$ with a strong preference for angles close to this value. For minimum #3, we found that the maximum was instead assumed near $\delta \approx 10^\circ$. While the dominant eigenvectors being parallel could not be ruled out, we thus preferred minimum #1. For minimum #4, we found a relative density that essentially ruled out the possibility that the dominant eigenvectors were parallel and therefore we again preferred minimum #1.

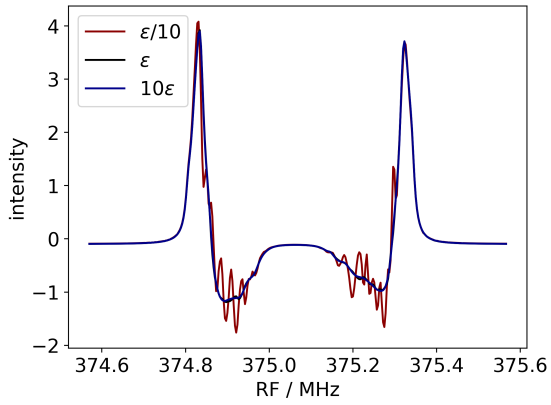


Figure D17: The derivative of the simulated ENDOR spectrum of compound **1** with respect to the parameter T for the orientation g_x approximated by finite difference approximation with $\epsilon = 10^{-3}$ (blue), $\epsilon = 10^{-4}$ (black) and $\epsilon = 10^{-5}$ (red).

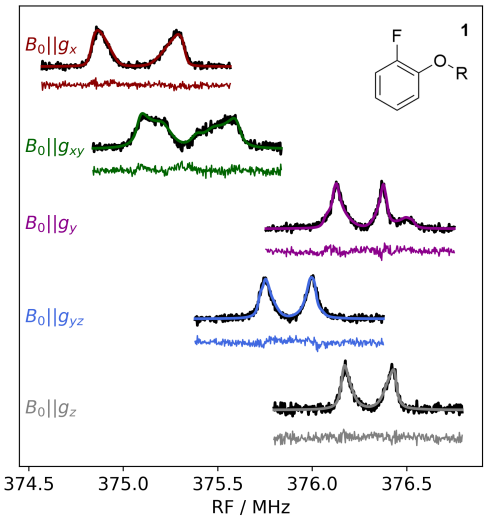


Figure D18: Measured (black) and fitted (orientation colour coded) ENDOR spectra for compound **1** using minimum #2. The spectral residuals are plotted below each of the spectra using the same color. Parameter values of the second minimum are given in Table D6.

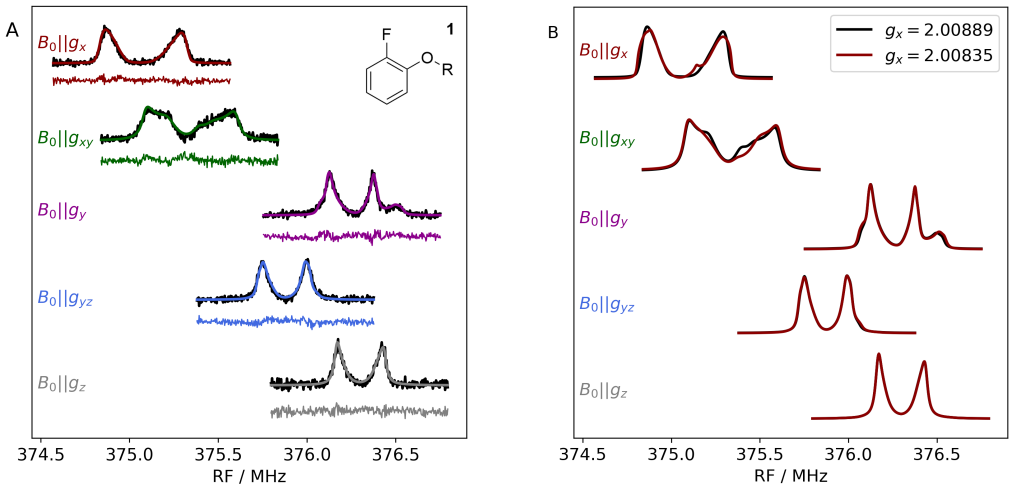


Figure D19: A: measured (black) and simulated (orientation colour coded) ENDOR spectra for compound **1** using the values listed in Table D7 and two contributions weighted 90% and 10% for the g_x-values, namely $g_{x,1} = 2.00889$ and $g_{x,2} = 2.00835$, respectively. The spectral residuals are plotted below each of the spectra using the same color. B: The two component spectra plotted individually (spectrum for $g_{x,1}$ in black, spectrum for $g_{x,2}$ in red), each normalized.

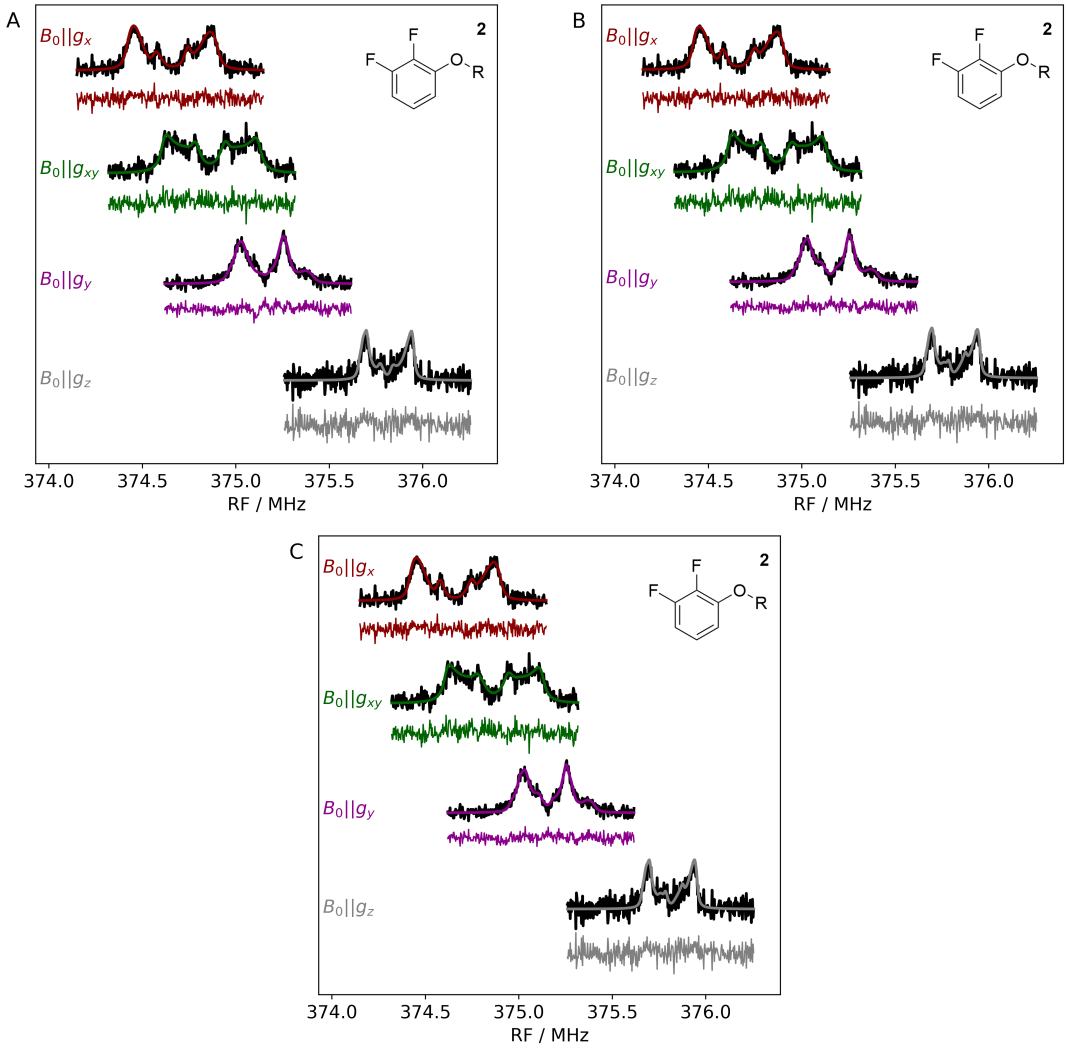


Figure D20: Measured (black) and simulated (orientation colour coded) ENDOR spectra for compound **2** using the parameter values listed in Table D8: minimum #2 (A), minimum #3 (B) and minimum #4 (C). The spectral residuals are plotted below each of the simulated spectra using the same color.

Spin Parameters			
o-F			
	Minimum #2	Minimum #3	Minimum #4
T / kHz	254.8	255.3	254.8
a_{iso} / kHz	4.0	2.1	2.1
$\alpha_A/^\circ$	-166.5	-167.0	-166.28
$\beta_A/^\circ$	-19.6	-19.9	-19.39
$\tilde{\sigma}_{xx}$ / ppm	136.5	169.5	152.1
$\tilde{\sigma}_{yy}$ / ppm	168.6	182.1	167.2
$\tilde{\sigma}_{zz}$ / ppm	317.9	335.1	348.9
$\alpha_\sigma/^\circ$	-14.8	5.0	67.0
$\beta_\sigma/^\circ$	120.3	58.8	119.9
$\gamma_\sigma/^\circ$	-68.2	109.8	115.8
m-F			
	Minimum #2	Minimum #3	Minimum #4
T / kHz	89.2	96.17	96.8
a_{iso} / kHz	13.2	1.2	1.1
$\alpha_A/^\circ$	-19.7	-163.0	-162.0
$\beta_A/^\circ$	-1.0	-10.9	-10.0
$\tilde{\sigma}_{xx}$ / ppm	137.0	169.5	152.1
$\tilde{\sigma}_{yy}$ / ppm	169.0	182.1	167.2
$\tilde{\sigma}_{zz}$ / ppm	318.0	335.1	348.9
$\alpha_\sigma/^\circ$	231.9	312.0	239.0
$\beta_\sigma/^\circ$	53.5	130.1	27.3
$\gamma_\sigma/^\circ$	52.1	172.0	153.0

Experimental Parameters			
	Minimum #2	Minimum #3	Minimum #4
B_0 / G : g_x	93521.2	93522.6	93522.0
B_0 / G : g_{xy}	93572.4	93574.6	93573.8
B_0 / G : g_y	93651.6	93653.0	93652.4
B_0 / G : g_z	93810.7	93813.1	93812.8
lw / kHz : g_x	36.9	31.9	31.4
lw / kHz : g_{xy}	48.3	48.5	47.8
lw / kHz : g_y	35.0	26.5	23.4
lw / kHz : g_z	17.8	17.1	15.8

Table D8: Spin and experimental parameters for the minima not reported in detail in the main text. Minimum #2 has slightly higher loss ($\mathcal{L} = 0.715$) than the other minima which arises from the shoulders of the $m-$F HF coupling (see panel A of Figure D20 in SI F) not being reproduced. Minimum #3 has loss $\mathcal{L} = 0.705$. Minimum #4 has loss $\mathcal{L} = 0.704$.

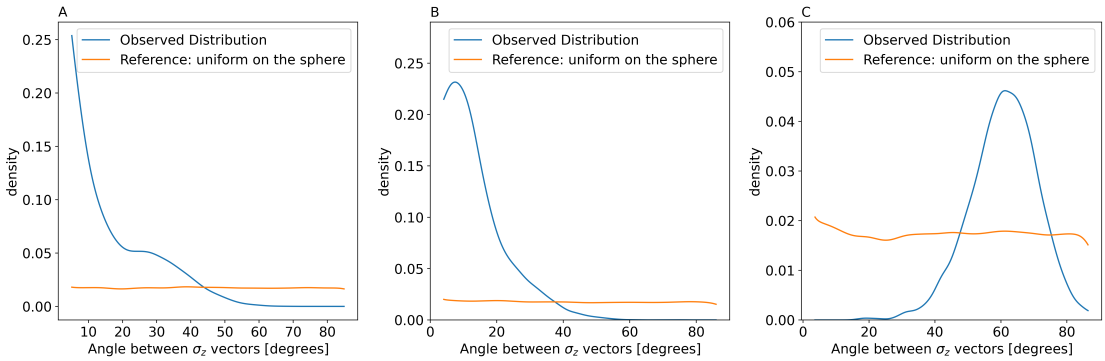


Figure D21: Relative approximate density of uncertainty distribution of angles between the dominant eigenvectors of ortho and meta CS tensors. A: minimum #1 (reported in main text, angle distribution suggests that the eigenvectors are approximately parallel), B: minimum #3 (angle distribution is compatible with the eigenvectors being parallel but with preference for a slight deviation), C: minimum #4 (angle distribution inconsistent with dominant eigenvectors being parallel).