

Masters Thesis: Computational Methods Applied to Rotational Spectroscopy

Marcus Schwarting^{1*}

^{1*}Department of Computer Science, University of Chicago, 5801 S Ellis
Ave, Chicago, 60637, Illinois, United States.

Corresponding author(s). E-mail(s): meschw04@uchicago.edu;

Abstract

Rotational spectroscopy is an incredibly powerful technique for measuring the quantized rotational states of molecules in the range of microwave radiation. While the technique can present a tremendous amount of molecular information that few spectroscopies can emulate, rotational spectroscopy is not commonly utilized by analytical chemists. But the greatest barriers to wider utilization are not experimental; rather, they are computational. We present our ongoing efforts towards the development of RotoCalc Monte Carlo, a method for the inverse mapping of rotational spectra by stochastically sampling from a robust forward model. RotoCalc Monte Carlo seeks to derive important spectroscopic quantities from a rotational spectrum without molecule foreknowledge or spectral annotation, thereby building up rotational spectroscopy as a more viable tool for general-purpose analytical chemistry.

Keywords: Rotational spectroscopy, chirped-pulse Fourier transform microwave spectroscopy, Monte Carlo methods

1 Introduction

Rotational spectroscopy uses microwave radiation to induce quantized rotational transitions on a polar molecule in the gas phase. The technique began in the 1930s to determine the electronic and conformational structures of such molecules. Innovations in instrumentation, especially with the recent advent of the chirped-pulse Fourier transform microwave (CP-FTMW) spectrometer by the Pate group [1], have permitted the structure of many molecules (as well as complexes) to be closely analyzed. Improvements in electronic components are constantly enhancing signal-to-noise (SNR) ratios

of spectra, while simultaneously making rotational instruments more price-competitive with more common spectroscopic techniques. Enhanced SNR enables the identification of conformational species of a molecule, and even allows for the identification of isotopomeric species (molecules with one or more isotopic substitutions of an atom) in natural abundance. Even more exciting is the ability to carry out gas-phase reactions within a CP-FTMW instrument and observe or infer transitional states, thereby experimentally enhancing our understanding of chemical reaction mechanisms.

Despite these experimental advantages and recent instrumentation innovations, rotational spectroscopy remains a boutique method. Two main factors separate rotational spectroscopy from commonly used analytical methods such as nuclear magnetic resonance (NMR) and Fourier transform infrared (FTIR) spectroscopy. First, common spectroscopic techniques must produce spectra which are easy to interpret by humans or automated software. NMR benefits from spectral peaks that match to specific carbon or hydrogen atoms, allowing experts and automated systems to garner important structural information, even for complex mixtures. FTIR lacks this peaks-to-atoms mapping, but benefits from functional groups exhibiting modes which localize to specific "fingerprint regions" of an infrared spectrum, which are readily interpreted at the undergraduate level. Rotational spectroscopy lacks any such characteristics, and in many cases experts can discern little from a spectrum divorced from a molecule-specific context. In practice, rotational spectra are experimentally collected for known molecules and corroborated alongside ab-initio calculations. Second, common spectroscopic techniques do not strictly require quantum number assignments for peak transitions in order to infer structural information. Rotational spectroscopy requires the correct manual assignment of six quantum numbers per peak transition, and while tools and heuristics exist to aid in spectral assignment, the process is often time-intensive and error prone. For these reasons, this so-called "inverse mapping" problem for identifying unknown molecules or mixtures using rotational spectroscopy has remained out of reach.

Our paper is ordered as follows. First, we provide a suitable background of rotational spectroscopy, derive the constituent rotational and dipole constants that comprise a rotational spectrum, and highlight relevant domain-specific computational software. Second, we introduce our own "forward mapping" tool and the Monte Carlo sampling procedures (for both single and multiple species) used to perform an inverse mapping. Third, we describe our results for a single molecule species across varying Monte Carlo sampling procedures. Finally, we highlight future directions for improving the robustness, speed, and scope of our methods.

2 Background

We first provide a background on calculating rotational and dipole constants from molecular geometries, then summarize the calculations required to generate a rotational spectrum from these constants.

2.1 Rotational and Dipole Constants

Consider an arbitrary molecule in the gas phase with a fixed internal geometry (also called a rigid rotor), where $\{(x_1, y_1, z_1, m_1), \dots, (x_n, y_n, z_n, m_n)\}$ describes the Cartesian locations and masses of all atoms. Such a molecule is free to rotate about three mutually orthogonal axes oriented about the center of mass. We can calculate nine rotational inertia constants according to $I_{x,x} = \sum_{i=1}^n m_i(y_i^2 + z_i^2)$ and $I_{p,q} = I_{q,p} = -\sum_{i=1}^n m_i p_i q_i$ where $p \neq q$. Constructing an inertial matrix decomposing, we are left with three inertial magnitudes $I_A \leq I_B \leq I_C$ along our three mutually orthogonal axes (corresponding to the inertial matrix eigenvectors). By convention, we write our rotational constants as $\chi = \frac{h}{8\pi^2 I_\chi}$ (where h is Planck's constant) for $\chi \in \{A, B, C\}$, resulting in $A \geq B \geq C$. While symmetric molecules exist where $A = B$ (oblate) or $B = C$ (prolate) or even $A = B = C$ (spherical), we generally assume the asymmetric case of $A > B > C$, which spans nearly all molecules. The degree of asymmetry is helpfully measured by Ray's asymmetry parameter, defined as $\kappa = \frac{2B-A-C}{A-C}$. These three rotational constants (by convention measured in MHz) will solely define the location of all spectral peaks for a molecular species.

After calculating rotational constants, the dipole constants of the rigid must also be determined. The overall dipole vector in \mathbb{R}^3 of a molecule is automatically calculated using most ab-initio software packages, but can also be computed purely classically using Gastieger charges (albeit with dubious accuracy). Using the previously computed rotational eigenvectors, we can perform a change of basis on the overall dipole vector to determine dipole constants $\mu_A, \mu_B, \mu_C \in \mathbb{R}$ which correspond to their respective rotational constants. These three dipole constants (by convention measured in Debye) will, along with the rotational constants, solely define the intensities of spectral peaks for a molecular species.

We have defined six variables $[A, B, C, \mu_A, \mu_B, \mu_C]$ which, with respect to the forward mapping, solely define a full rotational spectrum for a molecular species. From the forward mapping perspective, we can consider the overall mapping as, approximately, $\mathbb{R}^{4n} \rightarrow \mathbb{R}^6 \rightarrow \mathbb{R}^{2P}$ where P is the number of spectral peaks (considering frequencies and intensities). Note that other variables, such as centrifugal distortion, can be included and have an effect at higher frequencies, however we omit these for now. Six variables is far fewer than the number required for other spectroscopic techniques, making the bottleneck in the forward mapping readily apparent.

2.2 Spectral Calculation

Assuming that the rotational and dipole constants of a molecular species are identified via ab-initio or classical means, we now calculate the frequencies and intensities composing a rotational spectrum. We begin with the Schrödinger equation $\mathcal{H}\Psi = E\Psi$, where \mathcal{H} is a Hamiltonian matrix, E is an energy vector, and Ψ is the wave function matrix. We introduce three quantum numbers for a rotational state: $J \in [0, \dots, J_{\max}]$ represents overall inertial quantization, $K_A \in [-J, \dots, J]$ represents quantized rotational inertia towards an oblate extreme, and $K_C \in [-J, \dots, J]$ represents quantized rotational inertia towards a prolate extreme. The overall Hamiltonian matrix is linearly separable as $\mathcal{H} = \sum_{\chi \in \{A, B, C\}} \chi \mathcal{H}_\chi$, where each \mathcal{H}_χ is a $[(J_{\max} + 1)^2 \times (J_{\max} + 1)^2]$ band

matrix. The eigenvalues $\{\lambda_1, \dots, \lambda_P\}$ of \mathcal{H} represent the resulting unique (J, K_A, K_C) quantum states. A dipole matrix $\boldsymbol{\mu}^2$ can be constructed in a manner similar to the Hamiltonian, where $\boldsymbol{\mu}^2 = \Psi^{-1} [\sum_{\chi \in \{A, B, C\}} \mu_\chi^2 \mathcal{M}_{\mu_\chi^2}] \Psi$, where $\mathcal{M}_{\mu_\chi^2}$ matrices are also of size $[(J_{\max} + 1)^2 \times (J_{\max} + 1)^2]$.

Considering quantum states $\lambda_{lo} > \lambda_{hi}$ satisfying the selection rules (that is, having a nonzero corresponding dipole matrix element), the final peak transition frequencies (in MHz) are simply $\omega_{hi \rightarrow lo} = \lambda_{hi} - \lambda_{lo}$. The corresponding peak intensities (in units of $\text{MHz} * \text{Debye}^2$) are determined as

$$\rho_{hi \rightarrow lo} = \omega_{hi \rightarrow lo} \times \boldsymbol{\mu}^2_{hi \rightarrow lo} \times \left[1 - \exp\left(\frac{-\lambda_{lo}}{k_\beta T}\right) \right] \times \exp\left(\frac{-\omega_{hi \rightarrow lo}}{k_\beta T}\right)$$

where k_β is Boltzmann’s constant and T is a fixed effective temperature (slightly above absolute zero for most CP-FTMW instruments). For more information on the derivation of rotational spectra, see textbooks by Gordy and Cook [2] and Kroto [3].

2.3 Existing Forward and Inverse Mapping Tools

Rotational spectra can be accurately calculated from rotational and dipole parameters using software tools like Pickett’s SPCAT (open-source C) [4], Kisiel’s ASROT (open-source Fortran) [5], and Western’s PGOPHER (Pascal) [6]. While these tools are mostly in agreement on transition frequencies, they may occasionally differ during state assignment. The tools exhibit larger discrepancies when calculating spectral intensities, as theoretical frameworks for calculating intensities differ across tools to some extent. While these tools effectively address the forward problem of mapping rotational and dipole constants to spectra, the methods were never intended to scale beyond individual runs.

The inverse problem of mapping from rotational spectra back to rotational and dipole constants, and eventually to molecule identities, has been far more difficult. All methods rely on the accurate manual identification of some number of transition states in a spectrum, after which a fitting can be carried out. Some tools and heuristics exist to facilitate manual labelling of transition states, such as Loomis-Wood plotting [7] and RAARR scaffolding [8]. Once states have been manually assigned, tools such as SPFIT [4] and ASFIT [9] perform a least-squares fitting on the transition state assignments to determine underlying rotational constants. Unfortunately, fitting in SPFIT or ASFIT also requires a strong set of initial guesses of rotational constants, and poor initial guesses often lead to numerical instabilities or incorrect fittings. Seifert’s AUTOFIT tool [10] takes an initial step towards eliminating this guesswork and automating the fitting process by first running ab-initio calculations to estimate rotational constants from an optimized geometry, then adjusting alignments to experimental spectra in a brute-force grid fashion. An important limitation with all of these established techniques is the experimenter’s assumed prior knowledge of the contents of their sample. Without knowledge of the contents of a sample (as is the norm in analytical chemistry) ab-initio simulations can provide little assistance, and

there can be no clear strategy for manually labelling transition states. For these reasons, mapping the experimental spectrum of a sample of unknown origin to separate molecular species is practically impossible.

3 Methods

We introduce RotoCalc, our tool for spectral calculation, and RotoFit, our tool for fitting spectra based on labelled transitions. We also introduce a Monte Carlo method for inverse spectral mapping.

3.1 Forward Calculation With RotoCalc

First we present RotoCalc, our tool for calculating spectra from rotational and dipole constants. When computing spectra from rotational and dipole constants, RotoCalc follows essentially the same steps as described previously, but with several changes to increase calculation speed even further. First, recall the linear equation for the overall Hamiltonian \mathcal{H} for some J_{\max} where for $\chi \in \{A, B, C\}$:

$$\mathcal{H} = A\mathcal{H}_A + B\mathcal{H}_B + C\mathcal{H}_C = \sum_{\chi \in \{A, B, C\}} \chi \mathcal{H}_\chi.$$

Rather than computing the matrices $\mathcal{H}_\chi; \chi \in \{A, B, C\}$ at every step, we simply load these three $(J_{\max} + 1)^2 \times (J_{\max} + 1)^2$ sparse matrices into computer memory before beginning our calculations. The only intensive operation regarding the Hamiltonian is an eigendecomposition of a block diagonal matrix with eigenvalue reordering sped up using the Hungarian algorithm. We perform a similar memory loading method with the dipole matrix, but with several changes for additional speed improvements. Recall that

$$\boldsymbol{\mu}^2 = \mu_A^2 \mathcal{M}_{\mu_A^2} + \mu_B^2 \mathcal{M}_{\mu_B^2} + \mu_C^2 \mathcal{M}_{\mu_C^2} = \sum_{\chi \in \{A, B, C\}} \mu_\chi^2 \mathcal{M}_{\mu_\chi^2}.$$

Recall the required diagonalization by the eigenvector matrix \mathbf{V}' , a matrix which is dependent only on κ . The separate dipole-axis matrices are independent of the molecule-specific dipoles or principal directions, and are therefore constants. So the overall $\mathcal{M}_{\mu_\chi^2}$ matrices are dependent only on κ . Instead of recovering the dipole component matrix via a diagonalization on a computed and reordered eigenvector matrix, it can be shown that each location of the dipole component matrices can be more simply expressed as a fourth-order polynomial expression as a function of κ . That is,

$$\mathcal{M}_{\mu_\chi^2}(\kappa) = \sum_{p=0}^4 \kappa^p M_{\mu_\chi^2, p}.$$

To compute the overall dipole matrix, we therefore need only load fifteen sparse $(J_{\max} + 1)^2 \times (J_{\max} + 1)^2$ matrices into memory and carry out a handful of matrix additions and scalar multiplications. Peak intensity calculations can also be carried out quickly using simple matrix manipulation. Our RotoCalc program has now optimized the full

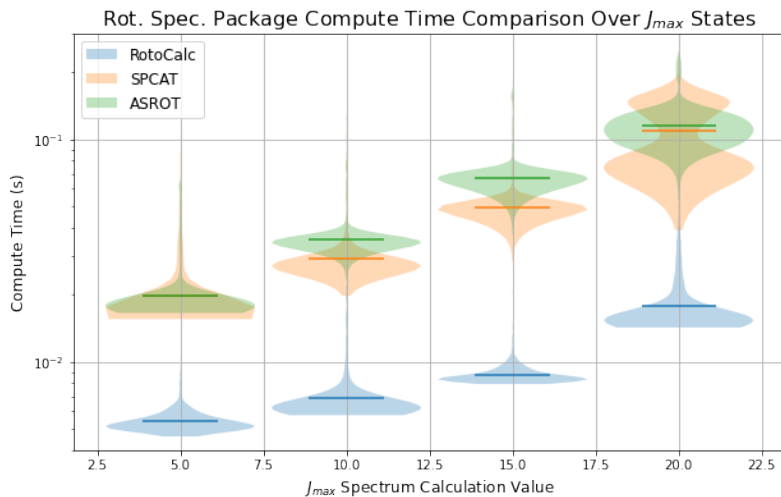


Fig. 1 Violin plot (semi-log) for time comparison of RotoCalc (blue) versus SPCAT (orange) and ASROT (green) for values of J_{\max} .

calculation of rotational spectra to a point where the most compute-intensive steps remaining are essentially unavoidable. These are the eigendecomposition of the block diagonal Hamiltonian matrix, the eigenvector matrix reordering using the Hungarian algorithm, and the final calculation of peak intensities.

Our tool is written in Python, performs calculations at a much faster rate, is easily scaled in parallel across cores, and generates spectra that match closely with existing tools. Figure 1 shows the wall time of RotoCalc versus competing tools SPCAT and ASROT across increasing J_{\max} .

We also introduce RotoFit, which acts as a substitute for SPFIT or ASFIT. For a set of labelled transitions and initial guesses for rotational constants, RotoFit uses a BFGS optimizer (as implemented in Scipy) to perform a least-squares fitting for the labelled transitions. Preliminary testing shows that RotoFit both faster and far less sensitive to initial guesses than SPFIT or ASFIT. However, as our work aims to completely eliminate expert manual annotation of spectra prior to analysis, we will not analyze RotoFit in further depth in this paper.

3.2 Monte Carlo Sampling Methods

RotoCalc provides a fast forward model to generate rotational spectra. For an inverse mapping of a spectrum with no manual annotation or experimental context, our fast forward model naturally begets a sampling approach. A sampling approach also comes with two notable benefits. First, sampling can be more computationally efficient than a grid approach over a large, high-dimensional search space. Second, sampling can provide a distribution over constants, which can lead to uncovering spectral degeneracy as well as allow for multiple conformer identification. One of the simplest Monte Carlo schemes is random walk Metropolis Hastings (RWMH), which is described in the algorithm below.

Algorithm 1 Random Walk Metropolis Hastings (Using RotoCalc)

Inputs A ground truth spectrum $\vec{\omega}^*, \vec{\rho}^*$

- 1: Draw $\Phi_0 = [A_0, B_0, C_0, \mu_{A,0}, \mu_{B,0}, \mu_{C,0}] \sim \mathcal{P}(\cdot)$.
- 2: **for** i from 1, ..., N : **do**
- 3: Propose $A' = A_i + \varepsilon$; $\varepsilon \sim N(0, \sigma_{rot}^2)$, repeat for B', C' .
- 4: Propose $\mu'_A = \mu_{A,i} + \varepsilon$; $\varepsilon \sim N(0, \sigma_{dip}^2)$, repeat for μ'_B, μ'_C .
- 5: Calculate $\vec{\omega}', \vec{\rho}' = \text{RotoCalc}(\Phi' = [A', B', C', \mu'_A, \mu'_B, \mu'_C])$
- 6: Calculate acceptance probability as $a = \frac{\mathcal{P}(\Phi')\mathcal{L}(\vec{\omega}', \vec{\rho}'|\Phi')}{\mathcal{P}(\Phi_i)\mathcal{L}(\vec{\omega}_i, \vec{\rho}_i|\Phi_i)}$
- 7: With probability $\min(1, a)$, let $\Phi_{i+1} = \Phi'$ [accepted]. Else:
- 8: Let $\Phi_{i+1} = \Phi_i$ [rejected].
- 9: **end for**
- 10: **return** $\Phi_{MAP} = \underset{i \in \{0, \dots, N\}}{\operatorname{argmax}} \mathcal{P}(\Phi_i)\mathcal{L}(\vec{\omega}_i, \vec{\rho}_i|\Phi_i)$

In accordance with Bayes theorem, we must define a prior distribution $\mathcal{P}(\cdot)$ and likelihood distribution $\mathcal{L}(\cdot|\cdot)$. Together these define a posterior $\pi(\cdot|\cdot)$, in the context of our rotational spectra written as:

$$\pi(\Phi|\vec{\omega}, \vec{\rho}) \propto \mathcal{L}(\vec{\omega}, \vec{\rho}|\Phi)\mathcal{P}(\Phi)$$

Or, since our implementation primarily takes place in log-space, we can write:

$$\log(\pi(\Phi|\vec{\omega}, \vec{\rho})) = \log(\mathcal{L}(\vec{\omega}, \vec{\rho}|\Phi)) + \log(\mathcal{P}(\Phi)) + \text{constant}.$$

Our choice of prior and likelihood functions can drastically change the behavior of the posterior distribution, and therefore affect the Monte Carlo scheme. We select a prior which assumes the molecule in question is a member of GDB9 (that is, containing fewer than nine heavy atoms). Since we have already computed rotational and dipole constants for DFT-relaxed geometries of conformers of all molecules in GDB9, we can simply parameterize each constant independently using an F-distribution. Figure 2 shows the histogram distribution of rotational and squared dipole constants, all fit to F-distributions. Our log-prior can then be expressed as

$$\log(\mathcal{P}(\Phi' = [A', B', C', \mu'_A, \mu'_B, \mu'_C])) = \sum_{\chi \in \{A, B, C\}} [F_{\chi}(\chi') + F_{\mu_{\chi}^2}(\mu_{\chi}^{\prime 2})]$$

Note that this prior fails to account for relationships between rotational constants (eg. $A \geq B \geq C$ is not enforced), and also fails to consider trends in κ (molecules in GDB9 roughly follow a chi-squared distribution with respect to $\kappa + 1 \in [0, 2]$).

Next, we define our likelihood function, which may be constructed in a number of fashions. First, we construct a likelihood with inspiration from Bretthorst’s Bayesian methods in nuclear magnetic resonance (NMR) quadrature [11, 12]. Consider the comparison of a single ground truth peak (ω^*, ρ^*) with another proposed single peak (ω', ρ'), both of which are in the frequency domain. Presuming the peaks are modeled

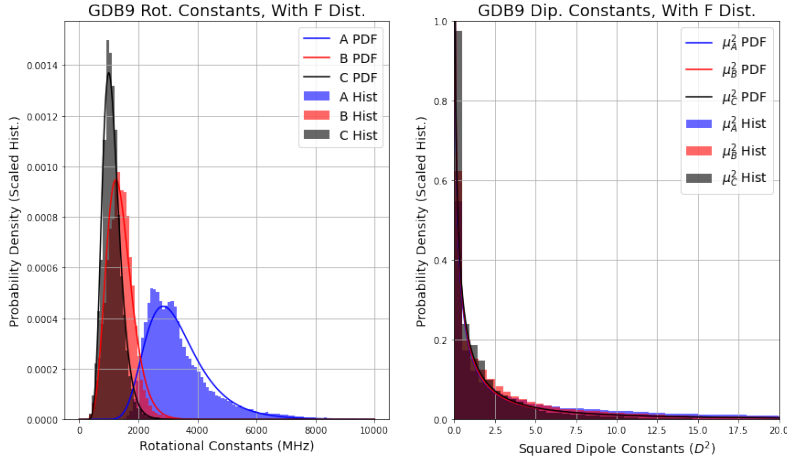


Fig. 2 Left: Histogram of rotational constants A (blue), B (red), and C (black) overlaid with fitted F-distributions. Right: Histogram of squared dipole constants μ_A^2 (blue), μ_B^2 (red), and μ_C^2 (black) overlaid with fitted F-distributions.

as Dirac delta functions in the frequency domain, recall the continuous inverse Fourier transform identity $\rho^* \delta(\omega - \omega^*) \xrightarrow{\mathcal{F}^{-1}} \frac{\rho^*}{2\pi} e^{i\omega^* t} \propto \rho^* \cos(\omega^* t) + \rho^* i \sin(\omega^* t)$. Since rotational spectroscopy is phase-locked and does not measure in quadrature (mainly due to limitations in instrumentation), the imaginary term may be ignored. This leaves $\rho^* \delta(\omega - \omega^*) \xrightarrow{\mathcal{F}^{-1}} \rho^* \cos(\omega^* t)$, which is equivalent to a noise-free free-induction decay measured experimentally (ignoring the exponential decay term). Free-induction decay duration for rotational spectroscopy is typically measured to around 20 microseconds, although the cutoff time can be varied by experimenters based on a number of factors. Of course, free-induction decay measurements are not noise-free, and are often averaged over longer time horizons to improve signal-to-noise ratio. For our likelihood function, we introduce a noise term $\epsilon_T \sim N(0, \sigma_T^2)$, which is applied to a proposed peak in the time domain. To assess the likelihood of the occurrence of one peak versus another, we consider a Gaussian distribution on Euclidean L2 metric across some duration t_{\max} in the time domain, written as

$$\mathcal{L}(\omega^*, \rho^* | \omega', \rho') \propto \frac{1}{\sigma_{\mathcal{L}}^2} \exp \left\{ - \frac{\left(\sum_{t=0}^{t_{\max}} \rho^* \cos(\omega^* t) - \sum_{t=0}^{t_{\max}} [\rho' \cos(\omega' t) + \epsilon_T] \right)^2}{\sigma_{\mathcal{L}}^2} \right\}$$

where $\sigma_{\mathcal{L}}^2$ is the variance of the Gaussian likelihood function.

Generalizing to more than one peak, we consider a ground truth spectrum $\vec{\omega}^*, \vec{\rho}^*$ and a proposed spectrum $\vec{\omega}', \vec{\rho}'$. Let us also consider a vector \vec{t} of arbitrary length Q which evenly spans the time domain within $[0, t_{\max}]$. Then our log-likelihood function may be written as

$$\log(\mathcal{L}(\vec{\omega}^*, \vec{\rho}^* | \vec{\omega}', \vec{\rho}')) = \frac{1}{\sigma_{\mathcal{L}}^2} \cdot \left\{ (\vec{\rho}^*)^T \cos((\vec{\omega}^*)^T \vec{t}) - \right.$$

$$\left. (\vec{\rho}')^T \cos((\vec{\omega}')^T \vec{t}) + \vec{e}_{\mathcal{L}} \right\}^2 - \log(Q\sigma_{\mathcal{L}}^2) + C_{\mathcal{L}}$$

where $C_{\mathcal{L}}$ is an arbitrary constant and the squared operator is applied element-wise to the Q -length matrix. As a matter of notation, we will write our likelihood as $\mathcal{L}(\vec{\omega}^*, \vec{\rho}^* | \Phi')$, since we can generate $\vec{\omega}'$, $\vec{\rho}'$ from our forward model using Φ' as the input.

A second likelihood function to consider is a measure-theoretic approach used in Boyd's DeepLoco model [13]. Considering a spectrum represented as a sum of Dirac deltas, it is natural to consider a convolution kernel ϕ such that, for a spectrum $\vec{\omega}'$, $\vec{\rho}'$, we have a mapping

$$\mathcal{R}_{\phi}(\vec{\omega}', \vec{\rho}') = x \mapsto \sum_i \rho'_i \phi(\omega - \omega'_i).$$

Considering a log-likelihood function with respect to the convolution kernel, we may write

$$\log(\mathcal{L}_{\phi}(\vec{\omega}^*, \vec{\rho}^* | \vec{\omega}', \vec{\rho}')) = \|\mathcal{R}_{\phi}(\vec{\omega}^*, \vec{\rho}^*) - \mathcal{R}_{\phi}(\vec{\omega}', \vec{\rho}')\|_2^2$$

which, in an expanded form, can be written as

$$\begin{aligned} \log(\mathcal{L}_{\phi}(\vec{\omega}^*, \vec{\rho}^* | \vec{\omega}', \vec{\rho}')) &= \sum_{i=1}^R \sum_{l=1}^R \rho_i^* \rho_l^* K_{\phi}(\omega_i^*, \omega_l^*) \\ &- 2 \sum_{i=1}^R \sum_{j=1}^Q \rho_i^* \rho'_j K_{\phi}(\omega_i^*, \omega'_j) + \sum_{j=1}^Q \sum_{k=1}^Q \rho'_j \rho'_k K_{\phi}(\omega'_j, \omega'_k) \end{aligned}$$

where K_{ϕ} is the kernel selected. In our case, two suitable contenders are a Gaussian kernel

$$K_{\phi}(\omega_i, \omega_j) = \frac{1}{2\pi\sigma_{\omega, \phi}} e^{-\frac{(\omega_i - \omega_j)^2}{2\sigma_{\omega, \phi}^2}}$$

and a Student's t kernel (which for a single degree of freedom is equivalent to Cauchy's distribution):

$$K_{\phi}(\omega_i, \omega_j) = \frac{1}{\pi\sigma_{\omega, \phi}} \left(1 + \frac{(\omega_i - \omega_j)^2}{\sigma_{\omega, \phi}^2}\right)^{-1}.$$

We can further extend the DeepLoco likelihood such that the kernel spans both frequencies and intensities. We simply rewrite our previous log-likelihood as

$$\begin{aligned} \log(\mathcal{L}_{\phi}(\vec{\omega}^*, \vec{\rho}^* | \vec{\omega}', \vec{\rho}')) &= \sum_{i=1}^R \sum_{l=1}^R K_{\phi}(\omega_i^*, \rho_i^*, \omega_l^*, \rho_l^*) \\ &- 2 \sum_{i=1}^R \sum_{j=1}^Q \rho_i^* \rho'_j K_{\phi}(\omega_i^*, \rho_i^*, \omega'_j, \rho'_j) + \sum_{j=1}^Q \sum_{k=1}^Q \rho'_j \rho'_k K_{\phi}(\omega'_j, \rho'_j, \omega'_k, \rho'_k) \end{aligned}$$

with two-dimensional Gaussian and Student's t kernel written respectively as

$$K_{\phi}(\omega_i, \rho_i, \omega_j, \rho_j) = \frac{1}{2\pi\sigma_{\omega, \phi}\sigma_{\rho, \phi}} \exp\left\{-\frac{1}{2}\left[\frac{(\omega_i - \omega_j)^2}{\sigma_{\omega, \phi}^2} + \frac{(\rho_i - \rho_j)^2}{\sigma_{\rho, \phi}^2}\right]\right\}$$

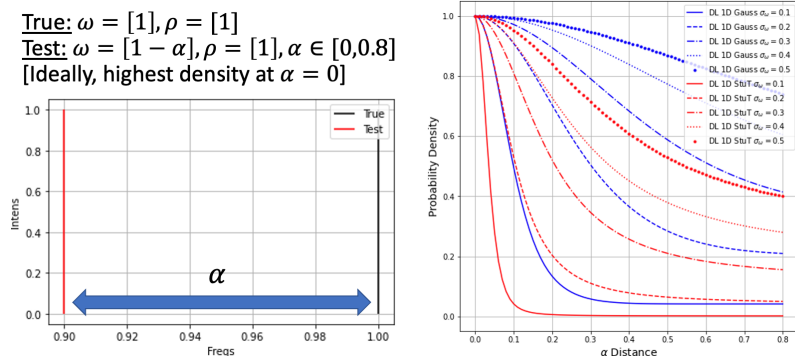


Fig. 3 A simple likelihood calculation test, where a ground truth frequency is fixed and a test frequency is varied along α . Left: a diagram of the simple likelihood test. Right: Varying kernel and σ_ω values using the DeepLoco likelihood function (with a Euclidean norm).

and

$$K_\phi(\omega_i, \rho_i, \omega_j, \rho_j) = \frac{1}{2\sqrt{\pi}\sigma_{\omega,\phi}\sigma_{\rho,\phi}} \left[1 + \frac{(\omega_i - \omega_j)^2}{\sigma_{\omega,\phi}^2} + \frac{(\rho_i - \rho_j)^2}{\sigma_{\rho,\phi}^2} \right]^{-3/2}.$$

Note that other kernels can certainly be used, however Gaussian is the most common choice, with Student's t often useful as well due to, informally, its heavier tails. Figure 3 shows a simple example of adjusting likelihood parameters to change the overall probability density with a Euclidean metric.

Note that a kernel simplification is only possible for L_1 (Manhattan) and L_2 (Euclidean) norms, however we often want to consider Minkowski L_P norms where $P < 1$. Selecting $P < 1$ can be advantageous when we seek to mute the effect of intensity variations, which often play a far smaller role in spectral matching than the underlying frequency alignment. When a likelihood is placed in a grid across variable rotational constants, we can observe a region of high probability density around the true rotational constants. Figure 4 shows one such example where $P = 0.2, \sigma_\omega = 25$. Using this likelihood and prior, we can now begin sampling via a RWMH approach.

4 Conclusion

This preliminary work demonstrates our capacity to identifying rotational constants from spectra without manual annotation or prior knowledge, which is an important step towards accepting rotational spectroscopy as a viable tool for analytical chemistry. In the future, we hope to continue advancing the state of the art for this fascinating inverse mapping problem. We also hope to successfully apply our methods to real experimental measurements supplied by our collaborators.

Acknowledgments. Thanks to advisors Ian Foster and Eric Jonas. Thanks to Argonne mentors, including Ben Blaiszik and Logan Ward. Thanks to experimental collaborators, including Nathan Seifert, Kirill Prozument, and Michael Davis.

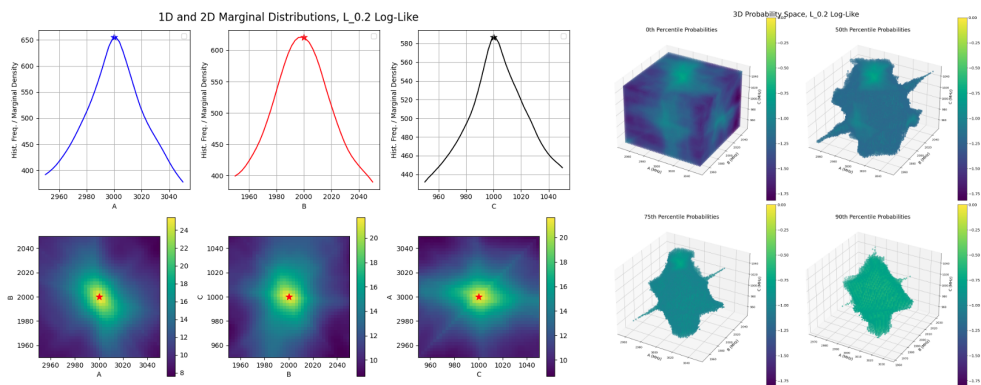


Fig. 4 A spectral likelihood function as a grid over rotational constants, first displaying one-dimensional and two-dimensional marginals (left), then showing a full three-dimensional display of log-likelihoods (across varying percentiles).

References

- [1] Brown, G.G., Dian, B.C., Douglass, K.O., Geyer, S.M., Shipman, S.T., Pate, B.H.: A broadband fourier transform microwave spectrometer based on chirped pulse excitation. *Review of Scientific Instruments* **79**(5), 053103 (2008)
- [2] Gordy, W., Cook, R.L., Weissberger, A.: *Microwave Molecular Spectra* vol. 18. Wiley New York, ??? (1984)
- [3] Kroto, H.W.: *Molecular Rotation Spectra* [by] HW Kroto, (1975)
- [4] Pickett, H.M.: The fitting and prediction of vibration-rotation spectra with spin interactions. *Journal of Molecular Spectroscopy* **148**(2), 371–377 (1991)
- [5] Demaison, J., Sarka, K., Cohen, E.A.: *Spectroscopy from Space* vol. 20. Springer, ??? (2001)
- [6] Western, C.M.: Pgoopher: A program for simulating rotational, vibrational and electronic spectra. *Journal of Quantitative Spectroscopy and Radiative Transfer* **186**, 221–242 (2017)
- [7] Bonah, L., Zingsheim, O., Müller, H.S., Guillemin, J.-C., Lewen, F., Schlemmer, S.: Llwp—a new loomis-wood software at the example of acetone-13c1. *Journal of Molecular Spectroscopy* **388**, 111674 (2022)
- [8] Yeh, L., Satterthwaite, L., Patterson, D.: Automated, context-free assignment of asymmetric rotor microwave spectra. *The Journal of chemical physics* **150**(20), 204122 (2019)
- [9] Kisiel, Z., Pszczółkowski, L., Medvedev, I.R., Winnewisser, M., De Lucia, F.C., Herbst, E.: Rotational spectrum of trans-trans diethyl ether in the ground and

three excited vibrational states. *Journal of Molecular Spectroscopy* **233**(2), 231–243 (2005)

- [10] Seifert, N.A., Finneran, I.A., Perez, C., Zaleski, D.P., Neill, J.L., Steber, A.L., Suenram, R.D., Lesarri, A., Shipman, S.T., Pate, B.H.: Autofit, an automated fitting tool for broadband rotational spectra, and applications to 1-hexanal. *Journal of Molecular Spectroscopy* **312**, 13–21 (2015)
- [11] Bretthorst, G.L.: Bayesian analysis. i. parameter estimation using quadrature nmr models. *Journal of Magnetic Resonance* (1969) **88**(3), 533–551 (1990)
- [12] Bretthorst, G.L.: Bayesian analysis. iii. applications to nmr signal detection, model selection, and parameter estimation. *Journal of Magnetic Resonance* (1969) **88**(3), 571–595 (1990)
- [13] Boyd, N., Jonas, E., Babcock, H., Recht, B.: Deeploco: fast 3d localization microscopy using neural networks. *BioRxiv*, 267096 (2018)