

Python による NMR データフォーマットの変換

熊木康裕

北海道大学大学院理学研究院

1. はじめに

NMR で測定直後に得られる生データを解析可能なスペクトルに変換するためには適切な数学的処理が必要となる。NMR には数多くの測定手法が存在し、それぞれのデータが必要とする処理は異なり、その中には極めて高度で複雑なものもある。また生データのファイルフォーマットは装置メーカーや機種に特有である。このためデータ処理は測定装置のメーカーが提供する NMR データの処理に特化したソフトウェアを利用するのが一般的である。ただしソフトウェアによって使用できる機能や対応しているデータのフォーマットには違いがある。例えば単純な一次元データであればどのソフトウェアも大部分のファイルフォーマットに対応しており、また適用される高速フーリエ変換などの処理はソフトウェア間での相違はほとんどなく、得られるスペクトルはソフトウェアに依存しない。一方拡散 (DOSY) データを例に挙げると、複数のファイルフォーマットに対応しているソフトウェアはほとんどなく、また施される処理手法は多岐にわたっているが、どの処理手法を提供しているかはそれぞれのソフトウェアで異なっている。そのため手持ちのデータに対応していないソフトウェアの機能を使用するためには、自力でファイルフォーマットの変換を行う必要がある。

ここでは JEOL 製 NMR 装置で得られた DOSY データを Bruker 製専用ソフトウェア TopSpin で処理可能なフォーマットに変換するための Python プログラムを作成した事例を報告する。これは 2021 年に NMR Club (全国の NMR に携わる技術職員のコミュニティ) のとある会員より寄せられた要望を受けて行ったものである。

2. データ構造とファイルフォーマットの解析

生データはデータ領域とパラメータ領域から構成されているが、両者の区分が明確である方がフォーマットの解析が容易になる。データ領域は一般にバイナリ形式であり、個々のデータ点に関する属性及びデータ点の配列様式によってそのフォーマットが決定される。一方パラメータ領域は特定の規則的なパターンを持っていないため、アスキー形式で記載されていることが必須となる。また測定パラメータの総数が膨大であるものの DOSY 処理で参照されるパラメータ (以降必須パラメータ群と称する) はその一部である。そこで Delta 及び TopSpin の両システムで必須パラメータ群を選別した後、両者の間で対応付けを行っている。なお DOSY スペクトルは二次元スペクトルの一種であり、横軸が通常の一次元データと同じ化学シフト、縦軸が自己拡散係数を示す。よって必須パラメータ群は一次元データと共通するものに加えて、自己拡散係数を算出するために必要な磁場勾配パルスに関するパラメータを含む。またこの磁場勾配関連のパラメータは使用するパルスシーケンスによっても定義が異なるため、両システム間で相当するパルスシーケンスファイルの対応付けも必要となる。

3. Python による Delta 形式から TopSpin 形式へのフォーマット変換

Python による処理には Jupyter Notebook (Ver. 5.7.4) を使用した。バイナリデータの変換では、NumPy モジュールを利用してデータ点に関する属性を変換している。ただしこの変換は倍精度浮動小数点型から整数型への型変換であるため少数部分の切り捨てが生じる。これによる丸め誤差の影響を最小限に抑えるため、変換元のデータ全域に適切な定数を乗じ、整数型の上下限を超えない範囲で最大限に増幅する操作を事前に行っている。一方測定パラメータの変換については、まず既存の TopSpin 形式の DOSY データのパラメータ領域から必須パラメータ群を全て消

去したものをテンプレートとして用意し、変換元のパラメータ領域から正規表現モジュール **re** を利用して必須パラメータの値を取得し、**TopSpin** 形式の表現でテンプレートに追記することで実現している。

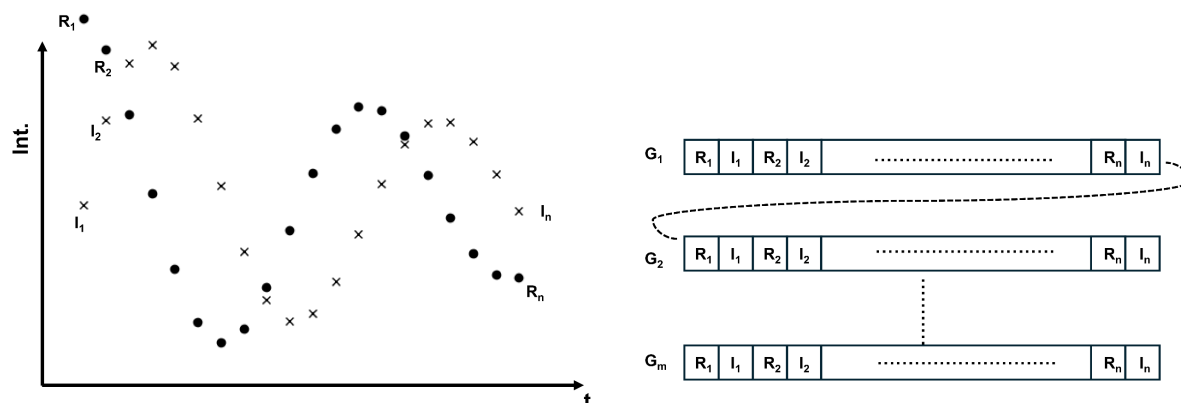


図 1 (左) (一次元)生データは磁化の強度の時間変化を離散的にサンプリングしたものであり、例えば時間 k において実数部 R_k と虚数部 I_k の二つのデータ点が存在する。(右) サンプリングポイント数 n の一次元データはそれぞれのデータ点が連なった一つの長方形に相当する。この例では実数部と虚数部は交互に並んでいる。**DOSY** データの場合、更に異なる磁場勾配条件 ($G_1, G_2 \dots G_m$) の下で得られた m 個の一次元データが連なった構造になっている。

4. TopSpin ソフトウェアによる検証

JEOL 製 ECA600II で得られた生データを上記の変換後に **TopSpin** (Ver4.0.9) 上の動的解析モジュール **Dynamics Center** (Ver 2.6.2)を用いて処理された合成高分子の **DOSY** スペクトルを図 2 に示す。スペクトルの外観は正常である一方、自己拡散係数を算出する際に用いられるパラメータの数値が変換によって約 **0.1%**減少していることが判明した。これは測定に使用された **Delta** システムと処理が行われる **TopSpin** システムの間で、パルスシーケンスのコード内容に微妙に違いがあることに起因すると考えられる。このようなずれを解消するためには、両者のパルスシーケンスが完全に一致するようにソースコードを修正するか、あるいは変換後のデータの拡散軸をスケールアップするといった作業が必要になる。

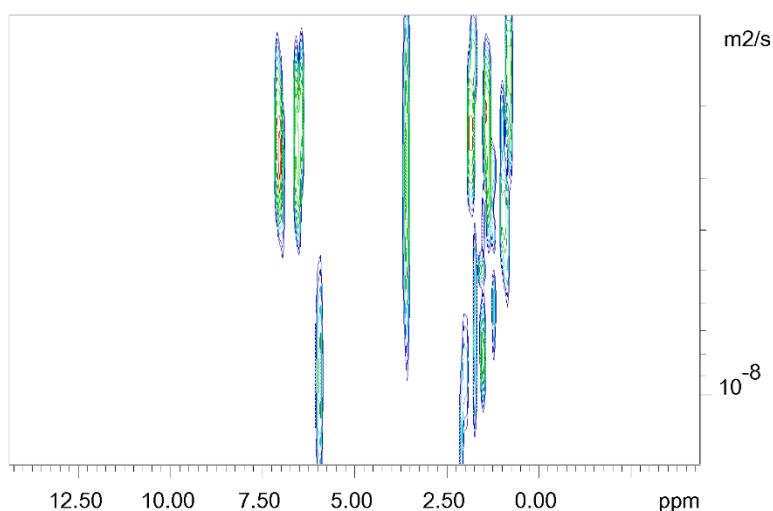


図 2: 合成高分子の **DOSY** スペクトル
自己拡散係数の算出に必要なパラメータである拡散時間(測定パラメータに基づいて計算される)はファイルフォーマット変換後では **49.95ms** であった(正しくは **50ms**)。その結果算出される自己拡散係数は約 **0.1%**大きく見積もられる。