9. Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is a method that can be used to find structures of proteins. NMR spectroscopy is the observation of spins of atoms and electrons in a molecule that is placed in a magnetic field. The spins precess at a frequency in the radio frequency range and the frequency can be detected by the electrical signal that it generates.

We briefly discuss the physics of NMR, and then describe distance geometry, a mathematical theory that can be used to find the protein structure from some types of NMR data.

9.1. Larmor frequency. Spins placed in a magnetic field precess; they wobble like a spinning top. Only certain isotopes of molecules found in organic compounds have spins that react to the magnetic field; the most common ones used in proteins are $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$. The isotopes $^{13}\text{C}$ and $^{15}\text{N}$ are not in common abundance, so specially prepared protein samples must be used.

\begin{figure}[h]
\centering
\includegraphics[width=0.3\textwidth]{nmr_diagram.png}
\caption{An NMR experiment is the observation of the precession of nuclear spins in the presence of a magnetic field. The large arrow represents the magnetic field of the magnet, the small arrow represents magnetic field of the nucleus which is precessing like a spinning top as the tip of the arrow moves on the indicated circle.}
\end{figure}

The frequency of precession is called the Larmor frequency and it is determined mainly by the type of atom and the strength of the magnetic field. The basic NMR
equation is,

\[ \omega = \gamma B_0 \]

where

- \( \omega \) = the Larmor frequency, i.e., the angular frequency of the precession in radians per second
- \( \gamma \) = gyromagnetic ratio, a constant depending on the type of the atom
- \( B_0 \) = the intensity of the magnetic field

9.2. Splitting and chemical shift. Equation (1) assumes that the detected frequency depends only on the atom and the intensity of the magnetic field. There is another factor, however. The magnetic field intensity \( B_0 \) is not the same everywhere in the molecule; it is affected by neighboring atoms and electrons. Neighboring spins have their own magnetic field and this perturbs slightly the field of the magnet and slightly changes the frequency of precession.

This is illustrated by the NMR spectrum of the hydrogen atoms in the molecule Toluene (figure 2). The spectrum is a Fourier transform of the electrical signal showing the intensities (vertical axis) of certain frequencies (horizontal axis). It shows that the peak frequencies cluster around two values. Also a reference signal is shown for hydrogens which are not part of any molecule. The frequencies of the hydrogens in the molecule are different (shifted) from the reference signal. The observed difference is divided by the frequency of the reference signal and the change in frequency is reported in parts per million or ppm. The change is called a chemical shift. The spectrum shows that the 3 methyl hydrogens (to the right) are shifted less that the other 5 hydrogens. From the symmetry of the molecule it is easy to see that the peak on the right comes from the methyl hydrogens.

The spectrum of Toluene shows that we can infer facts about the shape of a molecule by looking at the spectrum. This suggests that we can find chemical structures of larger molecules by NMR. However the spectrum of a protein is much more difficult to interpret. Below is the NMR spectrum of all the hydrogens in the protein thioredoxin indicating which part of the molecule the hydrogen signals are coming from. Although some aspects of the structure can be deduced from the spectrum, it would be difficult to find coordinates of the atoms from this spectrum.

There is another type of spectrum (figure 4) called a 2D NOESY spectrum. This experiment observes two frequencies from an atom, so the intensity is a function of two variables. The figure shows level curves for high intensity, which looks like a set of points. These can be used to estimate distances between atoms. Such estimates are called distance constraints. Distance constraints can be used to find atomic coordinates using techniques of distance geometry. Similar spectra can be used to find orientational constraints which measure angles rather than distances.
Figure 2. The spectrum of Toluene. In a magnetic field, the hydrogens spins in the molecule precess at different frequencies depending on the surrounding atoms and electrons. This is because the neighboring atoms and electrons have spins which contribute to the magnetic field. In Toluene the frequencies are seen in two separate ranges shifted from the reference signal.

9.3. Distance geometry. Determination of the structure of a protein from distance constraints caused renewed interest in an old branch of mathematics called distance geometry. Consider a sequence of points in 3D space. Information about distances between a sequence of points can be put into a matrix called a distance matrix. The atoms in a structure are numbered 1 to \( n \). The distance matrix is an \( n \times n \) matrix. The entry in row \( i \) and column \( j \) is the square of the distance between points \( i \) and \( j \). The object of distance geometry is to find coordinates of the points from the distance matrix.

There is no unique list of coordinates since a rotated and translated set of coordinates gives the same distance matrix, however, coordinates can be found which are unique up to a Euclidean motion. The situation is analogous to viewing a pdb file. Protein viewers change all the coordinates in the structure through rotations and translations, but the shape of the protein and the distances between atoms stays fixed.

9.3.1. Distance matrix example. Label the rows and columns below by the letters a through h. The corresponding entry gives the square of the distance between the
Figure 3. The hydrogen NMR spectrum of the protein thioredoxin. Indicated on the spectrum are the parts of the protein responsible for each part of the signal. The signal is shifted according to the unique chemical environment of each part of the protein.

Two corresponding points on the cube with sides of length 1 in figure 5:

\[
\begin{pmatrix}
0 & 1 & 2 & 1 & 1 & 2 & 3 & 2 \\
1 & 0 & 1 & 2 & 2 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 3 & 2 & 1 & 2 \\
1 & 2 & 1 & 0 & 2 & 3 & 2 & 1 \\
1 & 2 & 3 & 2 & 0 & 1 & 2 & 1 \\
2 & 1 & 2 & 3 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 2 & 2 & 1 & 0 & 1 \\
2 & 3 & 2 & 1 & 1 & 2 & 1 & 0
\end{pmatrix}
\]

9.4. Obtaining coordinates from a distance matrix. The central problem in distance geometry is to recover the coordinates of a sequence of points, up to a Euclidean motion of space, from the distance matrix. The solution is presented in this section.
Figure 4. NOESY spectrum of thioredoxin. A labeled point in the spectrum indicates that the signal comes from the interaction between a certain pair of hydrogen atoms. That the two hydrogen atoms interact indicates that they are not far apart. The distance between the atoms can be estimated by the intensity of the signal.

Figure 5. Cube (octahedron) with vertices labelled.

Suppose there are \( n + 1 \) vectors \( \mathbf{v}_0 \ldots \mathbf{v}_n \). For simplicity, translate the points so that one vector \( \mathbf{v}_0 = 0 \). Consider the vectors as columns of a \( 3 \times n \) matrix

\[
\mathbf{M} = (\mathbf{v}_1, \ldots, \mathbf{v}_n).
\]

Finding the vectors \( \mathbf{v}_j \) from the distance matrix follows in two steps
(1) find the gram matrix \( G = M' M \) from the distance matrix. The matrix \( G \) can be thought of as the square of \( M \).

(2) find \( M \) from the gram matrix \( G \). This can thought of as taking the square root of the gram matrix.

Note that the distance matrix is \((n + 1) \times (n + 1)\) and the gram matrix is \(n \times n\).

9.4.1. **Gram matrix from distance matrix.** The entries in the distance matrix \( D \) are

\[
|v_j - v_k|^2, \quad j, k = 0, 1, \ldots n,
\]

\[
(2) \quad D = \begin{pmatrix}
0 & |v_1|^2 & |v_2|^2 & \cdots & |v_n|^2 \\
|v_1|^2 & 0 & |v_1 - v_2|^2 & \cdots & |v_1 - v_n|^2 \\
|v_2|^2 & |v_2 - v_1|^2 & 0 & \cdots & |v_2 - v_n|^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
|v_n|^2 & |v_n - v_1|^2 & |v_n - v_2|^2 & \cdots & 0
\end{pmatrix}
\]

The gram matrix \( G \) is given by

\[
(3) \quad G = \begin{pmatrix}
v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\
v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & v_2 \cdot v_n \\
\vdots & \vdots & \ddots & \vdots \\
v_n \cdot v_1 & v_n \cdot v_2 & \cdots & v_n \cdot v_n
\end{pmatrix}
\]

The gram matrix can be found from the distance matrix using the vector identity

\[
(4) \quad -2 \, v_i \cdot v_j = |v_i - v_j|^2 - |v_i|^2 - |v_j|^2.
\]

All the information on the right hand side of this equation is in the distance matrix.

Also using (4) we can get the gram matrix from the distance matrix by row and column operations

1. subtract the first row of \( D \) from each row
2. subtract the first column from each column.
3. delete the first row and column.

The result is \(-2G\) with a zero first row and first column added.

9.5. **Coordinates from the gram matrix.** Given \( G \) the problem is to find a \(3 \times n\) matrix \( M \) such that

\[
(5) \quad G = M' M.
\]

This can be done by finding eigenvectors and eigenvalues. Write the eigenvectors of \( G \) in a matrix \( V \) and the eigenvalues as diagonal entries in a diagonal matrix \( E \). Then

\[
(6) \quad GV = VE.
\]

Since the gram matrix is symmetric, it has real eigenvalues with real eigenvectors. In fact, because it came from a distance matrix in 3D space, it can be shown
that it has three non-negative eigenvalues and the rest 0, and we can arrange the eigenvalues so that

$$E = \begin{pmatrix} E_1 & 0 \\ 0 & 0 \end{pmatrix}$$

where $E_1$ is a diagonal matrix with 3 non-negative entries. Eigenvectors can always be found which are an orthonormal set,

$$V'V = I.$$  

Now construct a diagonal matrix whose entries on the diagonal are the square roots of the entries of $E$. Call this matrix $\sqrt{E}$. From (6) and (8) it follows that

$$G = V\sqrt{E}V' = M'M.$$  

where $M = \sqrt{E}V'$. If

$$V' = \begin{pmatrix} A \\ B \end{pmatrix}$$

where $A$ is a $3 \times n$ matrix, then

$$M = \sqrt{E}V' = \begin{pmatrix} \sqrt{E_1}A \\ 0 \end{pmatrix}$$

and first three rows of the matrix $M$ give coordinates of points with the desired gram and distance matrices.

This procedure for finding coordinates is best illustrated using 4 points ($n = 4$).

9.5.1. Example. Find coordinate of points giving the distance matrix

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 2 & 0 & 3 & 1 \\ 1 & 3 & 0 & 2 \\ 1 & 1 & 2 & 0 \end{bmatrix}$$

[Maple demo]


- large experimental error in determination of distance NOESY cross-peak intensities (NOEs) are classified into three categories with distances 1.8 to 2.7 Angstroms, 1.8 to 3.3 Angstroms, and 1.8 to 5.0 Angstroms for strong, medium and weak NOE’s.
- not many distances measured Only atoms within 5.0 Angstroms give rise to NOE’s
It is often easier than solving distance matrices to modify crystallographic refinement packages such as X-plor/CNS to include distance constraints. The program then minimizes the sum of energy and experimental error by a combination of conjugate gradient and simulated annealing minimization. The experimental error in this case is the sum of the square of the calculated distance between two atoms and the distance observed by the NOE resonance,
\[ \sum_j (d_{o,j} - d_{c,j})^2. \]

9.7. Types of NMR. **Solution NMR** This is also called liquid state NMR. The observed molecules are tumbling rapidly in a solution. Although the NMR signal depends on how the molecule is oriented with respect to the magnetic field direction, \( \mathbf{B} \), this orientation is changing constantly and the observed signal is the average over the time of the experiment.

**Solid state NMR** The observed molecules are motionless and fixed with respect to \( \mathbf{B} \). The NMR experiment can give information on the orientation of the molecule or parts of the molecule with respect to the magnetic field direction.

9.8. Quadratic functions in NMR. Very often, observables in physics can be expressed as quadratic functions. For example, the height of a dropped weight affected by only gravity is \( \frac{1}{2}gt^2 \), a quadratic function of time \( t \). This is only an approximation, but a very good one.

The situation is similar in NMR. Observables such as the chemical shift and the dipolar splitting are closely approximated by quadratic functions of coordinates of the magnetic field direction in some frame \( \mathbf{F} \) fixed with respect to the molecule.

The unit vector in the direction of the applied magnetic field is denoted \( \mathbf{B}_0 \). If \( \mathbf{F} = (f_1, f_2, f_3) \) is an orthonormal frame and
\[ \mathbf{B}_0 = x f_1 + y f_2 + z f_3 \]
then the equation chemical for shift \( \sigma \), for example, is of the form
\[(11) \quad \sigma = Ax^2 + By^2 + Cz^2 + 2Dxy + 2Eyz + 2Fxz, \]
a quadratic function of the coordinates of \( \mathbf{B}_0 \). The constants \( A, B, C, D, \) and \( E \) are determined by theory and experiment. The value of \( \sigma \) then gives information on the coordinates \( x, y, z \) of \( \mathbf{B}_0 \) in the frame \( \mathbf{F} \).

9.8.1. Quadratic Functions and symmetric matrices. A quadratic function in three variables can be written in terms of a symmetric matrix and a vector.
Let
\begin{equation}
A = \begin{pmatrix}
A & D & F \\
D & B & E \\
F & E & C \\
\end{pmatrix}
\end{equation}
be a symmetric $3 \times 3$ matrix and
\[X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}\]
a vector. Then $X'AX$ gives the same value as (11). The quadratic function is represented by a symmetric matrix.

9.8.2. **Parameters for a symmetric matrix.** A symmetric $3 \times 3$ matrix is given by just six entries, the ones on the diagonal and above the diagonal. The matrix has an alternate description in terms of its three eigenvalues and an orthonormal set of eigenvectors. Using the eigenvalues and eigenvectors the matrix $A$ can be diagonalized,
\[A = FDF^t\]
with $F$ orthogonal
\[D = \begin{pmatrix} d_1 & 0 & 0 \\
0 & d_2 & 0 \\
0 & 0 & d_3 \end{pmatrix}\]
The eigenvalues $d_1$, $d_2$ and $d_3$ are called the *principal values* and the matrix $F$ with eigenvectors as columns is called *principal axis frame*. This alternate way of describing a symmetric matrix in terms of principal values and the principal axis frame is often used in NMR and in many other applications.