Math 127 - Fall 2016 - Class Project

Nisha Janyaprayot Shangxing(Tony) Jiang Tanmay Ghai
Fraser Holmes-Mackay CJ Kim

December 9, 2016

Contents

1 Electron-Microscopy (Basic Deliverables)

1.1 Introduction to the Biological Problem

Proteins are biological molecules that carry out many important specialized functions in organisms. Antibodies - which help our immunity, enzymes - which facilitate metabolism, and hormones - which coordinate a large number of other bodily functions, are just a few examples of proteins and their roles. Proteins consist of chains of amino acids that are folded into three-dimensional shapes. It is the proteins shape which determines the proteins specialized function. Thus, producing images of proteins is essential to understanding their individual purpose. This allows us to learn more about the way that organisms survive.

The biological problem we are trying to solve is determining protein structure by using electron microscopy - the use of a high power microscope which produces magnified images by accelerating electron beams through magnetic lenses onto the specimen.

Electron Microscopy (EM) is a method of resolving detailed images of extremely small structures using a beam of electrons, and specifically for our purposes is used to find the shape and structure of a molecule. All electron microscopes are extremely large and require extensive training to operate properly.

Within the field of EM, there are multiple types of EM. The type that cryo-EM falls into is called Transmission Electron Microscopy (TEM). The principle of TEM is similar to that of a normal light microscope in that electrons are passed through a condenser lens, through the specimen to be examined, through another two sets of lenses called the objective and projection lenses, and onto the fluorescent screen, which gives the image. In simplified terms, cryo-EM takes a purified sample of a specimen in its natural environment and freezes it in a thin film using liquid nitrogen. Then the specimen is put through the electron microscope to develop an image of many individual, identical, discrete proteins in the frozen block. The process that follows this is an analysis of the many identical proteins put in different orientations, coming to a consensus on the shape of the protein.

1.2 Mathematical Model for Electron-Microscopy

1.2.1 Preliminaries for the Model

In this model we define each image corresponding to a given viewing direction as a projection of a formula $\rho : \mathbb{R}^3 \to \mathbb{R}$ and such that $\rho(x, y, z)$ is the electric potential at the point (x, y, z) due to e^-. In this model, a viewing direction, F, is modeled as a

3×3 matrix (or frame), such that;

$$F = \begin{bmatrix} \hat{a} & \hat{b} & \hat{c} \end{bmatrix}$$

where $F \in SO(3) = \{A \in M_{3,3}(\mathbb{R}) | AA^T = Id_3, det(A) = 1\}$. More specifically, the image due to a viewing direction F calculates the density of a pixel along the line to it. This is defined by the function $I_F : \mathbb{R}^2 \to \mathbb{R}$ such that;

$$I_F(x, y) = \int_{-\infty}^{\infty} \rho(ax + by + cz)dz$$

Mathematically, the aim of EM is to find I_F given ρ and F. One method of tackling this problem would be to numerically calculate the integral I_F. Although this seems straightforward, the values of integration of ρ change with every different input, which leads to an unreasonably large $O(n)$. The method in which we use to interpolate aims to eliminate this variability. An alternative method of calculating I_F is to use the Fourier Slice Theorem.

1.2.2 Introduction to Fourier Transforms

Before continuing this method of modeling EM, it is important to understand the concept of Fourier transforms. The study of Fourier transforms developed from the notion of Fourier series. A Fourier series is a method of writing periodic functions as the sum of sines and cosines.

Definition. A function $f(x)$ is *periodic* if there exists an a such that for all x within the domain:

$$f(x + a) = f(x).$$

It follows from the definition above that we can take

$$f(x + an) = f(x)$$

where $n \in \mathbb{Z}$. Now suppose that any periodic function can be written in the form

$$f(x) = \sum_{n=-\infty}^{\infty} f_n e^{\left(\frac{2nx\pi i}{a}\right)}.$$

This can be proved by solving for f_n. Notice that f_n is independent of x. Next we multiply both sides of the equation by the complex conjugate $e^{\frac{-2mx\pi i}{a}}$ and integrate by x from $x = 0$ to $x = a$.

$$\int_0^a e^{\frac{-2mx\pi i}{a}} f(x)dx = \sum_{n=-\infty}^{\infty} f_n \int_0^a e^{\frac{2(n-m)x\pi i}{a}} dx$$

Case 1: $n = m$

$$RHS : \int_0^a e^{\frac{2(m-m)x\pi i}{a}} dx = \int_0^a dx = a$$

Case 2: $n \neq m$

$$\begin{aligned} RHS : \int_0^a e^{\frac{2(n-m)x\pi i}{a}} dx &= \frac{e^{\frac{2(n-m)x\pi i}{a}}}{\frac{2(n-m)\pi i}{a}} \Bigg|_0^a \\ &= \frac{e^{2(n-m)\pi i} - 1}{\frac{2(n-m)\pi i}{a}} \\ &= 0 \end{aligned}$$

Definition. The *Kronecker Delta Function* is

$$\int_0^a e^{\frac{2(n-m)x\pi i}{a}} dx = \delta_{nm} a$$

where δ_{nm} is defined as

$$\delta_{nm} = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

Now we can rearrange the formula on the above page, and substitute the Kronecker Delta Function to get the following

$$\sum_{n=-\infty}^{\infty} f_n \delta_{nm} a = \int_0^a e^{\frac{-2mx\pi i}{a}} f(x) dx$$

$$f_m = \frac{1}{a} \int_0^a e^{\frac{-2mx\pi i}{a}} f(x) dx.$$

Thus, for any periodic function $f(x)$, we can write in the series form:

$$f(x) = \sum_{n=-\infty}^{infty} f_n e^{\frac{2nx\pi i}{a}}.$$

This is known as the Fourier series of $f(x)$. A more general definition is given below with the exponential written in terms of trigonometric functions.

Definition. Let $f(t)$ be a real periodic function such that $f(t)$ is integrable over an interval $[t_0, t_0 + T]$. Then the *Fourier Series (FS)* of $f(t)$ is;

$$f(t) = A_0 + \sum_{k=1}^{\infty} (A_k \cos(kt) + B_k \sin(kt))$$

where;

$$A_0 = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)dt,$$

$$A_k = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\cos(kt)dt, \quad k = 1, 2, \ldots, \infty$$

$$B_k = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\sin(kt)dt, \quad k = 1, 2, \ldots, \infty$$

If the period that the Fourier Series is defined on increases, that is $T \to \infty$, we end up with the Fourier Transform.

Definition. For any $\omega \in \mathbb{R}$, the *Fourier Transform (FT)* of $f(t)$ is defined as;

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)\exp(-2\pi it\omega)dt$$

It is also possible to find $f(t)$ from $f\hat{(}\omega)$ using the Inverse Fourier Transform.

Definition. For any $x \in \mathbb{R}$, the *Inverse Fourier Transform (IFT)* of $f\hat{(}\omega)$ is defined as;

$$f(t) = \int_{-\infty}^{\infty} f\hat{(}\omega)\exp(2\pi it\omega)d\omega$$

By going back to basics, we know that an integral can be thought of as a continuous version of a sum, where we add up infinite amount of infinitely small pieces. In the definition for $f(t)$ we have an oscillatory function $e^{2i\omega t}$ with angular frequency ω which is multiplied by an amplitude $\hat{f}(\omega)$. It follows that the integral of $f(t)$ can be regarded as the 'sum' over all values of ω, of the oscillatory function with angular frequency ω multiplied by an amplitude dependent on ω.

Although the definitions given above are only in one dimension, it is possible to give the Fourier Transform of a point in higher dimensions. This can be done by first supposing that $f(\vec{t})$ is defined on $\mathbb{R}^n$. Now the FT and IFT become:

$$\hat{f}(\vec{\omega}) = \int_{\mathbb{R}^n} f(\vec{t})\exp(-2\pi i\vec{t}.\vec{\omega})d\vec{t} \qquad \text{and}$$

$$f(\vec{t}) = \int_{\mathbb{R}^n} \hat{f}(\vec{\omega})\exp(2\pi i\vec{t}.\vec{\omega})d\vec{\omega}$$

respectively.

Notation: The following notations for the FT of $f(t)$ and the IFT of $\hat{f}(\omega)$ are equivalent:

- $\hat{f}(\omega)$ and $\mathcal{F}\{f(t)\}$
- $f(t)$ and $\mathcal{F}^{-1}\{\hat{f}(\omega)\}$

Definition. Define the *convolution* of two functions $f(t)$ and $g(t)$ as

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(z)g(t-z)dz$$

Note that, by using the substitution $y = t - z$, we have that $(f \star g)(t) = (g \star f)(t)$.

Theorem 2.1. *(The Convolution Theorem)*

$$h(t) = (f \star g)(t) \Leftrightarrow \hat{h}(\omega) = \hat{g}(\omega).\hat{f}(\omega)$$

.

Proof. ($\Rightarrow$)

$$\begin{aligned} & h(t) = (f \star g)(t) \\ \implies \hat{h}(\omega) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(z)g(t-z)dz \right) \exp(-2\pi i\omega.t)dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z)g(t-z) \exp(-2\pi i\omega.t)dtdz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z)g(y) \exp(-2\pi i\omega.(z+y))dydz \\ &= \int_{-\infty}^{\infty} f(z) \exp(-2\pi i\omega.z)dz. \int_{-\infty}^{\infty} g(y) \exp(-2\pi i\omega.y)dy \\ &= \hat{g}(\omega).\hat{f}(\omega) \end{aligned}$$

In the above proof we use the substitution $y = t - z$.

($\Leftarrow$) The proof of the converse uses the uniqueness property that follows from the inverse transform. That is that for continuous functions

$$f(t) = g(t) \iff \hat{f}(\omega) = \hat{g}(\omega).$$

We have seen already that $(f \star g)(t)$ transforms to $\hat{g}(\omega).\hat{f}(\omega)$ and so by using this property, the inverse transform of $\hat{g}(\omega).\hat{f}(\omega)$ is $(f \star g)(t)$.

□

We now aim to show the Fourier Slice Theorem. Suppose that a 2D function f is projected onto a 1D line. We use the notation P_1 for this and denote the one dimensional FT of this projection as $F_1P_1(f)$. Also suppose that the 2D FT of f,

denoted $F_2(f)$, is sliced through its origin (parallel to the projection line). We write this as $S_1F_2(f)$.

Theorem 2.2. *(Fourier Slice Theorem)*

$$F_1P_1(f) = S_1F_2(f)$$

Proof. Take the projection line to be the x-axis. We are allowed to do this since, if the line is shifted in the y direction then the projection does not change and if the line is rotated then the theorem still holds. Given a 2D function $f(x, y)$ then the projection onto the x-axis is

$$P_1(f) = p(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

and the Fourier transform of $f(x, y)$ is

$$\hat{f}(\omega_x, \omega_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\exp(-2\pi i(x\omega_x + y\omega_y))dxdy$$

Now, to take the slice along the same line as projection, we set $\omega_y = 0$ to get

$$\begin{aligned} S_1F_2(f) = \hat{f}(\omega_x, 0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\exp(-2\pi ix\omega_x)dxdy \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x, y)dy\right]\exp(-2\pi i(x\omega_x)dx \\ &= \int_{-\infty}^{\infty} p(x)\exp(-2\pi ix\omega_x)dx \\ &= F_1P_1(f) \end{aligned}$$

□

The theorem can be extended to higher dimensions in the following way:

$$F_mP_m(f) = S_mF_N(f)$$

1.2.3 Method of Calculating I_F

After defining Fourier transforms it is now possible to use their applications to calculate I_F. In particular, for a given viewing direction $F = \begin{bmatrix}\hat{a} & \hat{b} & \hat{c}\end{bmatrix}$, we will apply the Fourier Slice Theorem, giving:

$$\mathcal{F}\{\rho\}\big|_{\text{span}\{\vec{a},\vec{b}\}} = \mathcal{F}\{I_F\}$$

and so, it follows that

$$\hat{\rho}(\vec{\omega})\big|_{\text{span}\{\vec{a},\vec{b}\}} = \hat{\rho}(\omega_x\vec{a} + \omega_y\vec{b}) = \hat{I_F}(\omega_x, \omega_y).$$

Now that we have a form of $\hat{I_F}$, it is possible to take the inverse transform to find I_F. By combining the lines above into one formula, we simply have:

$$I_F = \mathcal{F}^{-1}\{\mathcal{F}\{\rho\}\big|_{\text{span}\{\vec{a},\vec{b}\}}\}.$$

Notation: $\text{span}\{\vec{a},\vec{b}\} = \{\omega_x\vec{a} + \omega_y\vec{b} | \omega_x, \omega_y \in \mathbb{R}\}$

By combining the formula above with the knowledge that ρ is calculable or given, then we can calculate I_F through the following steps:

1. Compute $\hat{\rho}(\vec{\omega})$ using $\rho(\vec{\omega})$.
2. Sample $\hat{\rho}(\vec{\omega})$ along $\text{span}\{\vec{a},\vec{b}\}$
3. Return $\mathcal{F}^{-1}\{\hat{\rho}(\vec{\omega})\big|_{\text{span}\{\vec{a},\vec{b}\}}\}$

Of these three steps, computing the FT of $\rho(\vec{\omega})$ causes the most problems. In order to do this we will first define the Discrete Fourier Transform and discuss it's relation to the corresponding FT.

1.2.4 Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is similar to the FT however, the DFT is not continuous. Instead, the DFT is only equivalent to the FT at N sample points. That is there is only a finite data set to input, much like the real world problem of our EM model.

Definition. Suppose $f : \mathbb{R} \to \mathbb{R}$ and let $y[k] = f(\frac{k}{N})$ where $0 \leq k \leq N$ and N is a fixed number of samples. Then the *Discrete Fourier Transform (DFT)* of y is

$$\hat{y}[\omega] = \sum_{k=0}^{N-1} y[k] \exp\left(-2\pi i.\frac{k\omega}{N}\right).$$

Sometimes this is denoted as $DNF(y, \omega)$.

Definition. Let X and Y be sets. Then a function $f : X \to Y$ is *supported* on $A \subseteq X$ if $f(x) = 0$ for all $x \notin A$.

Although not obvious, it is possible to derive a relationship between the DFT and the FT. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and take $\hat{f}(\omega)$ as the FT of f. Fix a number of samples N and define a periodic summation of $\hat{f}$ as

$$\hat{f}_N(\omega) = \sum_{m=-\infty}^{\infty} f(\omega + mN).$$

But now, since $\hat{f}_N(\omega)$ is periodic, we know that it can be rewritten as a Fourier series, that is:

$$\hat{f}_N(\omega) = \sum_{m=-\infty}^{\infty} f(\omega + mN) = \sum_{k=-\infty}^{\infty} A_k \exp\left(-2\pi i . \frac{k\omega}{N} \right)$$

where

$$\begin{aligned}
A_k &= \frac{1}{N} \int_{-\frac{N}{2}}^{\frac{N}{2}} \hat{f}_n(\omega) \exp\left(2\pi i . \frac{k\omega}{N}\right) d\omega \\
&= \frac{1}{N} \int_{-\frac{N}{2}}^{\frac{N}{2}} \sum_{m=-\infty}^{\infty} \left(f(\omega + mN) \exp\left(2\pi i . \frac{k\omega}{N}\right) \right) d\omega \\
&= \frac{1}{N} \sum_{m=-\infty}^{\infty} \left(\int_{-\frac{N}{2}}^{\frac{N}{2}} f(\omega + mN) \exp\left(2\pi i . \frac{k\omega}{N}\right) d\omega \right) \\
&= \frac{1}{N} \int_{-\infty}^{\infty} f(\omega) \exp\left(2\pi i . \frac{k\omega}{N}\right) d\omega \\
&= \frac{1}{N} f\left(\frac{k}{N}\right)
\end{aligned}$$

Substituting A_k into the summation, we get:

$$\hat{f}_N(\omega) = \sum_{m=-\infty}^{\infty} f(\omega + mN) = \sum_{k=-\infty}^{\infty} \frac{1}{N} f\left(\frac{k}{N}\right) \exp\left(-2\pi i . \frac{k\omega}{N} \right) = \frac{1}{N} DFT(y, \omega)$$

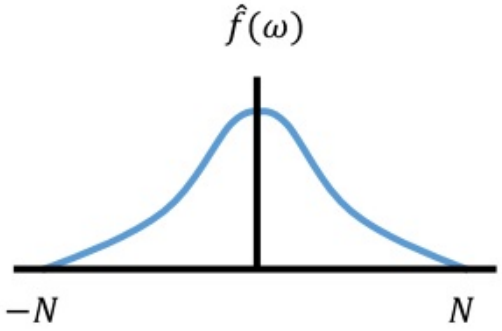
By considering the case where f is supported on $[0, 1)$ we have:

$$\begin{aligned}
\frac{1}{N} DFT(y, \omega) &= \sum_{m=-\infty}^{\infty} f(\omega + mN) \\
&= \hat{f}(\omega) + \sum_{m \in \mathbb{Z}, m \neq 0} f(\omega + mN)
\end{aligned}$$

This is known as the *Poisson Summation Formula* and $\sum_{m \in \mathbb{Z}, m \neq 0} f(\omega + mN)$ is known as the *aliasing error*. We can see from this that if the aliasing error is negligible, then

$DFT(y, \omega) = N\hat{f}(\omega)$. To ensure that this is the case, we notice that the 'smallest' m can be is 1 or -1 and so it is possible to choose a sample size based on the value of $|\omega|$.

Suppose that $\hat{f}(\omega)$ is 'essentially' zero for $|\omega| > Q$, where we define Q as the *reflective band limit*. Now, by referring to the following graph of $\hat{f}(\omega)$, we expect that $\sum_{m\in\mathbb{Z},m\neq 0} f(\omega + mN)$ is negligible if $N > 2Q$. It follows that it is possible to use the DFT as a good approximation for the FT provided a suitable sample size is used.



Now that we have found a calculable sum - the DNF - relating to the FT, we must find an efficient method of computing it. First let's consider the trivial case of direct computation of the DFT, given the formula above. This would mean calculating N^2 terms of sums and so the complexity of computing the DFT in this way would be $O(N^2)$. Thus, while the DFT is applicable to any complex valued series, for large series it can take an inefficient amount of time to compute. Instead, let's consider an algorithm that reduces the complexity of computing the DFT to $O(N \log N)$ - the Fast Fourier Transform.

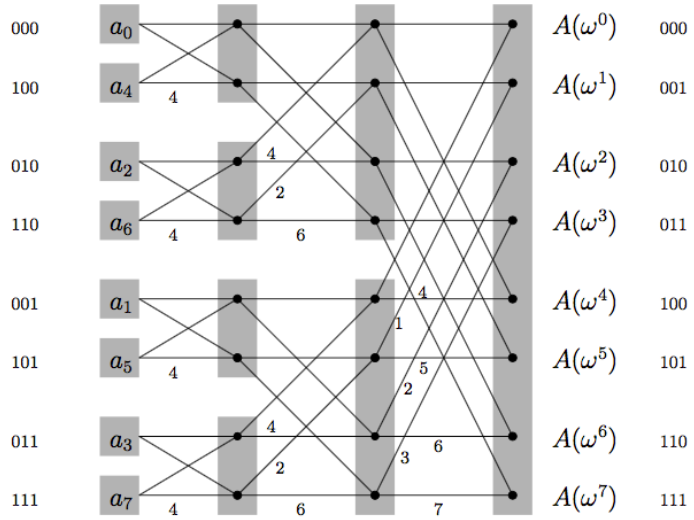
1.2.5 The Fast Fourier Transform

The output of the Fast Fourier Transform (FFT) is the same as the DFT, however, the FFT algorithm removes redundant computations and is therefore a more efficient method of calculating the FT.

Suppose we wanted to multiply two polynomials. Normally, this would involve multiplying coefficients and taking $O(n^2)$ time depending on number of terms. However, we can alternatively represent a polynomial by its value at certain points (because n points uniquely determine a $n-1$ degree polynomial). In that case, it would take only $O(n)$ time to multiply two polynomials, but it would still take longer to find the coefficients and to find values at points by evaluation.

The FFT algorithm notices that we can actually find values at n points in $O(n \log n)$ time if we choose our points wisely, and also interpolate back the coefficients (by

putting the polynomial into a Fourier basis in evaluation, multiplying there, and then converting it back in interpolation).



1.2.6 Back Projection

Back projection is a method of recovering an approximation of the 3-dimensional structure ρ from the 2-dimensional images I_F. Before defining back projection in a mathematically rigorous manner, we give some definitions that are necessary in back projection.

Definition. The *delta function* $\delta(x)$ is 0 for all $x \neq 0$ and satisfies

$$\int_{-\infty}^{\infty} \delta(x) = 1$$

Definition. We define the function *rect* as

$$rect(z^j) = \begin{cases} 1 & \text{if } -D \leq z \leq D, \\ 0 & \text{else} \end{cases}$$

where D is the diameter of ρ.

Definition. If I_F is an image and $F = [\hat{a}\ \hat{b}\ \hat{c}]$, the *back projection* of I_F is a function

$$b(x\hat{a_i} + y\hat{b_i} + z\hat{b_i}) = I_F(x, y) \star (rect(z)\delta(x, y))$$

Taking this definition, we can simplify the equation by changing the indices and setting $\ell_j = \delta(x^j, y^j)rect(z^j)$. This gives us that the back projection of an image I_j is defined as

$$bp_j(x^j, y^j, z^j) = I_j(x^j, y^j) \star \ell_j$$

where the exponents j of x, y, z are indices.

We must describe the back projection equation in relation to ρ in order to be able to utilize it, so we get

$$\begin{aligned} bp(x,y,z) &= \sum_{j=1}^{N} bp_j(x^j, y^j, z^j) \\ &= \sum_{j=1}^{N} \iint I_j(x^j, y^j)\ell(x^j - x'^j, y^j - y'^j, z^j)dx'^j dy'^j \\ &= \sum_{j=1}^{N} \iint \Big[\int \rho(x'^j, y'^j, z'^j)dz'^j \Big] \ell(x^j - x'^j, y^j - y'^j, z^j)dx'^j dy'^j \\ &= \sum_{j=1}^{N} \iiint \rho(x'^j, y'^j, z'^j)\ell(x^j - x'^j, y^j - y'^j, z^j)dx'^j dy'^j dz'^j \end{aligned}$$

With this in mind, we claim that back projecting from $\rho \to b$ is both linear and shift invariant. Thus $\rho = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{bp\}}{\mathcal{F}\{h\}}\right\}$ where $\mathcal{F}\{h\} \neq 0$. Now using the back projection of $\rho = \delta(x, y, z)$ to find h is given by

$$I_j(x^j, y^j) = \int_{-\infty}^{\infty} \rho(x^j, y^j, z^j)dz^j = \delta(x^j, y^j)$$

1.2.7 Application of Back Projection

Given the definition of back projection as $bp_j(x^j, y^j, z^j) = I_j(x^j, y^j) \star \ell_j$, we can take the following steps to use back projection in finding a 3-dimensional model:

1. Compute $\mathcal{F}\{I_j\}$
2. Take $B_i = \mathcal{F}\{I_j\}\ell_i$, where both $\mathcal{F}\{I_j\}$ and ℓ_i are 3-D arrays
3. Rotate B_i to the standard basis according to F_i
4. Return $\mathcal{F}^{-1}\{B_i\}$

We can use this method to aid in producing a 3-dimensional model from a set of 2-dimensional images.

1.3 Statement of the Problem

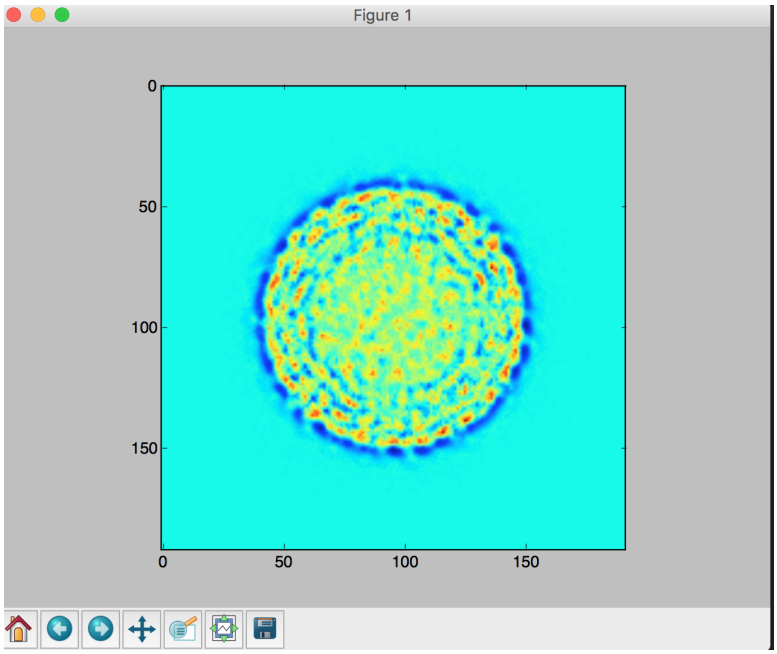
This section reiterates the problem that requires such extensive background mathematical knowledge given in the section above.

1. Part one of our problem was, given a certain sample and a particular orientation, we want to show a simulated electron microscopy image of the sample. In mathematical terms, we compute the Fourier transform of ρ, which is then acted on by the Discrete Fourier transform (Fast Fourier transform), and finally the inverse Fourier transform. We then write a function in Python to produce the final output as the 2-dimensional image.

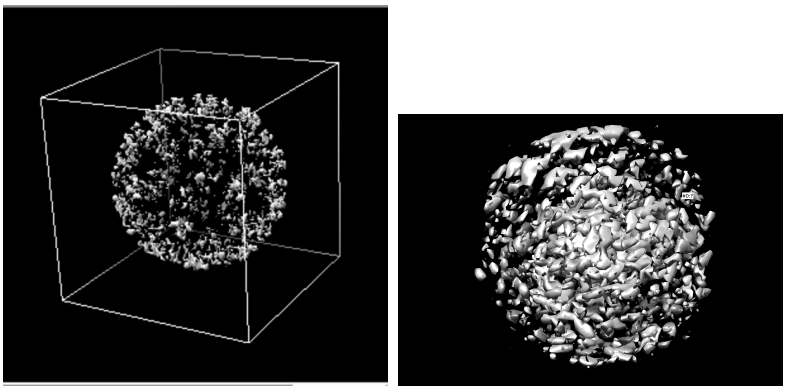
2. The second part of our problem is part one in reverse; we want to produce a 3-dimensional molecule by back projection and a sample of images from different orientations of the molecule. Viewing directions of particular images also have to be approximated in order to produce an accurate 3D model through the different sample slides. The sample of images we take to produce the 3D model are produced from the function written in part 1 of the problem. Part two of our problem is generally the problem that biologists face when conducting cryo-EM imaging.

1.4 Solution to Problem and Results

1. Our first component of attempting to solve the biological problem was to write a function project-fst(mol, R) using the Projection/Fourier Slice Theorem (discussed in Section 1.2.2 as Theorem 2.2) that is able to simulate an EM image of a given molecule from a certain viewing direction that is specified by a rotation matrix, namely R. Below is a simulated image using a viewing direction rotated 45 degrees to the right. For further comments and a look at the source code, please refer to "project-fst.py".

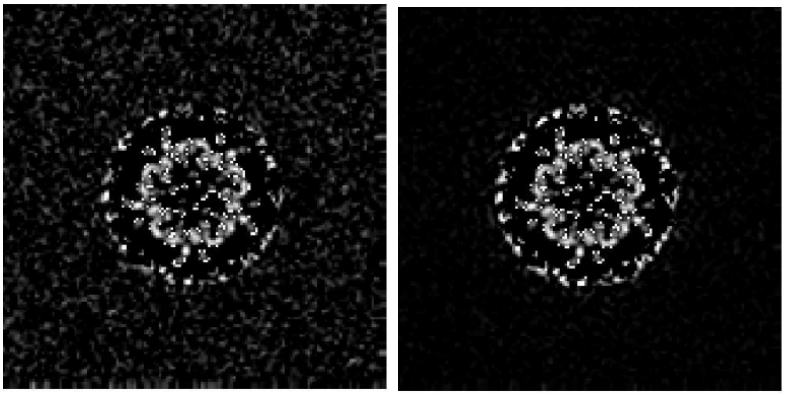


2. Our next component in solving the biological problem was utilize back-projection (discussed in section 1.2.6 and 1.2.7) to implement a function that is able to reconstruct a 3D molecule given a set of images generated using our project-fst function and a set of orientations or viewing directions. After reconstructing the 3D molecule and returning the matrix containing the pertinent data, we were able to write the matrix to an MRC file and view that using a visualization software known as Chimera. Below is an example of the reconstruction of a 3D molecule (zika-153.mrc) that we were able to generate using a set of 10 images. We tested with more (namely 15 and 20 images), however, in terms of runtime, 15 images took about 20 minutes and using 20 images brought the runtime up to 30 minutes. For 10 images, as you can see, we were able to attain pretty reasonable results, keeping aside the loss of resolution. For further comments and a look at the source code, please refer to "reconstruct.py".



d

3. Our final component in our solution to the biological problem was to estimate the viewing directions of certain images of a 3D molecule using common lines and an algorithm known as the angular reconstitution algorithm. We were able to generate a test set of images and reconstruct the 3D molecule. Furthermore, we went in and added various amounts of noise to simulate a more accurate 3D reconstruction. Below are example output with signal to noise ratios of 0.7, 0.8, 0.9, respectively tested on an asymmetric ribosome unit using our back-projection function with known random rotation matrices downloaded from here: http://www.rcsb.org/pdb/explore/explore.do?structureId=5iqr. For further comments and a look at the source code, please refer to "estimate-orientations.py".



2 X-Ray Scattering (Extra Deliverables)

2.1 Introduction to the biological problem of x-ray scattering

We will be solving the same biological problem of determining protein structure, however, now by using X-Ray Diffraction Scattering and Crystallography. A crystal is a repeating unit in three-dimensional space, which can translate over a defined three-dimensional lattice. We will be treating protein as a crystal. First, you start this process by creating a crystal of identical copies of protein that are all oriented the same way - this is called crystallization. Then, x-rays, which are non-visible light waves, scatter off of the crystal and a diffraction pattern can be observed. Next, the diffraction pattern can give us the Fourier Transform of charge density in the crystal via Fourier Optics (Thompson Scattering). The intensity of each spot in the diffraction pattern gives the magnitude of the Fourier Coefficient. However, the phases of the Fourier Coefficient must be found with a different method, this is called the *Phase Problem*.

Another issue with X-Ray Crystallography is that crystallizing proteins is incredibly difficult, or even impossible for many types of proteins - the ribosome took 30 years to crystallize. Thus, Electron Microscopy has the upper hand since EM can determine the 3D structure of proteins without crystallization. EM also doesn't use light rays, rather, EM uses electrons, which react stronger with the charge density in proteins compared to x-rays. However, EM produces unclear, 'noisy' images of protein. Thus, both EM and X-Ray crystallography have their benefits to understanding the structure of proteins.

Understanding the geometry of a crystal is very important to determining its properties, specifically, its symmetry. To begin this discussion, we must consider the mathematical point of view of periodic functions and Fourier Series (shown in section 1.2.2).

2.2 X-Ray Diffraction Scattering and Crystallography

The goal of this model is to be able to create a diffraction pattern, given a known crystal.

2.2.1 Preliminaries

- In a three-dimensional lattice, there are three non-parallel, non-co-planar, *primitive lattice vectors* $\vec{a}, \vec{b}, \vec{c}$.
- The collection of lattice vectors

$$\vec{R}_{uvw} = u\vec{a} + v\vec{b} + w\vec{c}, u, v, w \in \mathbb{Z}$$

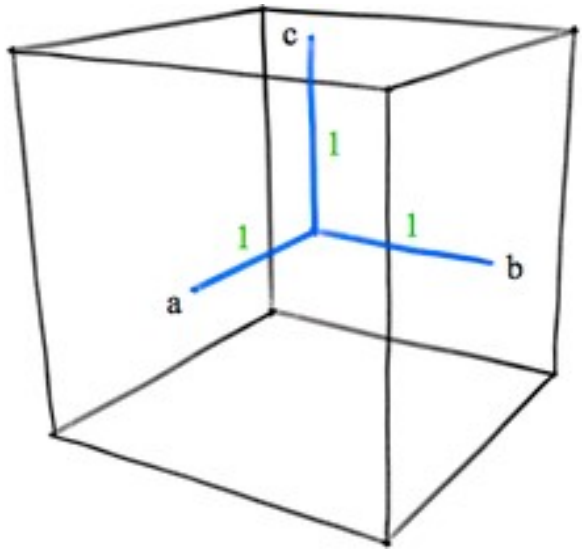
forms a 3D lattice.

- A *position in 3-D space* can be described as a combination of unit vectors in cartesian coordinates:

$$\vec{r} = x\vec{e}_1 + y\vec{e}_2 + z\vec{e}_3$$

or in crystal coordinates:

$$\vec{r} = x\vec{a} + y\vec{b} + z\vec{c}$$



With this nomenclature, we can relate the FS to crystallography. Creating diffraction patterns using X-Ray Scattering and Crystallography utilizes the electron charge density in crystallized objects.

2.2.2 Electron Charge Density

The charge distribution in a crystal is normally periodic in 3D. Since the electron charge density, denoted $\rho(\vec{r})$, is periodic, it satisfies:

$$\rho(\vec{r}) = \rho(\vec{r} + \vec{R}_{uvw})$$

where $\vec{R}_{uvw}$ is a lattice translation of the crystal. Now, using the proof that a periodic function can be written as a FS (section 1.2.2), we can write

$$\rho(\vec{r}) = \sum_G \rho_{\vec{g}} \exp(i\vec{g}.\vec{r})$$

for all $\vec{g} \in G$, where G is some set of vectors. This introduction of G means that we must explore whether $\rho(\vec{r})$ has any constraints on G and so we rearrange to consider:

$$\rho(\vec{r} + \vec{R}_{uvw}) - \rho(\vec{r}) = 0.$$

Now we can write these in their corresponding FS form:

$$\sum_G \rho_{\vec{g}} \exp(i\vec{g}.(\vec{r}+\vec{R}_{uvw}) - \sum_G \rho_{\vec{g}} \exp(i\vec{g}.\vec{r}) = 0$$

$$\Rightarrow \sum_G \rho_{\vec{g}} \exp(i\vec{g}.\vec{r})(\exp(i\vec{g}.\vec{R}_{uvw} - 1) = 0.$$

We find that the plane waves $e^{(i\vec{g}.\vec{r})}$ are linearly independent and so they cannot be written in the form:

$$\sum_{G'} \rho_{\vec{g}'} e(i\vec{g}'.\vec{r})$$

where $G \neq G'$. Thus, we have that

$$e^{(i\vec{g}.\vec{R}_{uvw})} = 1.$$

The constraint further implies that G' must be of the form

$$\vec{g}.\vec{R}_{uvw} = 2k\pi$$

where $k \in \mathbb{Z}$. This set G' is known as the set of *Reciprocal Lattice Vectors.*

2.2.3 Reciprocal Lattice Vectors

Consider a lattice defined by the collection of vectors

$$\vec{R}_{uvw} = u\vec{a} + v\vec{b} + w\vec{c}, \quad u, v, w \in \mathbb{Z}$$

Definition. The *Reciprocal Lattice Generating Vectors* are called $\vec{a}*, \vec{b}*, \vec{c}*$ and satisfy

$$\begin{array}{lll} \vec{a}.\vec{a}* = 2\pi, & \vec{b}.\vec{a}* = 0, & \vec{c}.\vec{a}* = 0, \\ \vec{a}.\vec{b}* = 0, & \vec{b}.\vec{b}* = 2\pi, & \vec{b}.\vec{c}* = 0, \\ \vec{a}.\vec{c}* = 0, & \vec{b}.\vec{c}* = 0, & \vec{c}.\vec{c}* = 2\pi. \end{array}$$

Definition. Define a *reciprocal lattice vector* as

$$\vec{g}_{hkl} = h\vec{a}* + k\vec{b}* + l\vec{c}* \quad h, k, l \in \mathbb{Z}.$$

and the *set of reciprocal lattice vectors* as

$$G' = \left\{\vec{g}_{hkl} : h, k, l \in \mathbb{Z}\right\}.$$

Note now that $\vec{g}_{hkl}.\vec{R}_{uvw} = 2k\pi$, where $k \in \mathbb{Z}$. We can now relate this back to the charge density of a crystal, written as

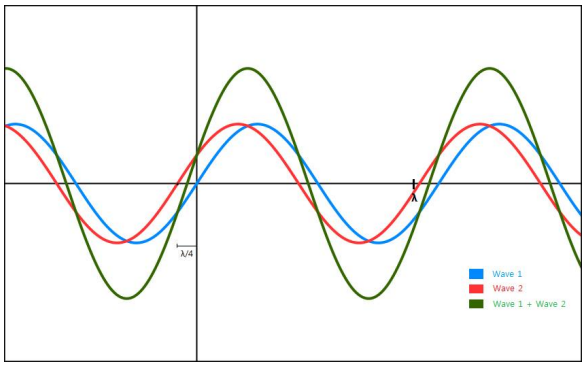
$$\rho(\vec{r}) = \sum_{hkl} \rho_{\vec{g}_{hkl}} e^{(i\vec{g}_{hkl}.\vec{r})}$$

This form will be useful later on in this discussion.

2.2.4 X-Ray Diffraction Scattering

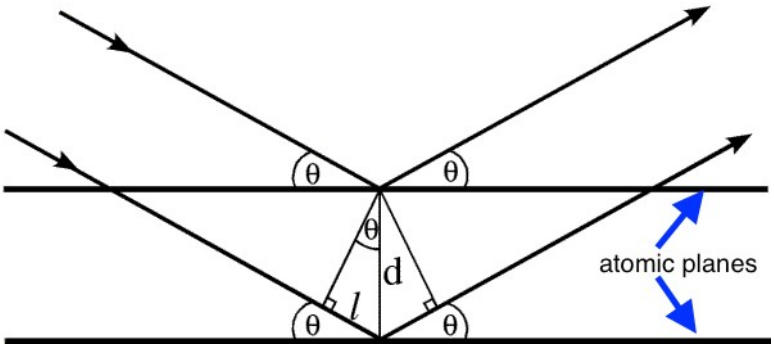
First, let's take two sine waves each with wavelength λ, with starting points apart. Denote the amplitudes of wave 1 and wave 2, respectively, as below:

$$A_1(x) = sin\left(\frac{2\pi\left(x + \frac{\lambda}{4}\right)}{\lambda}\right)$$
$$= sin(\frac{2\pi x}{\lambda} + \frac{\pi}{2})$$
$$A_2(x) = sin(\frac{2\pi x}{\lambda}).$$



NOTES:

- $\frac{\pi}{2}$ in the parentheses is a phase shift.
- The sum of the amplitudes $= A_1(x) + A_2(x)$.
- The phase shift (path length difference) is $\phi = \frac{2\pi}{\lambda}$.
- If $\phi = \lambda$ then the waves add constructively.
- Usually, the magnitude of the *wave vector* is defined as $q = \frac{2\pi}{\lambda}$



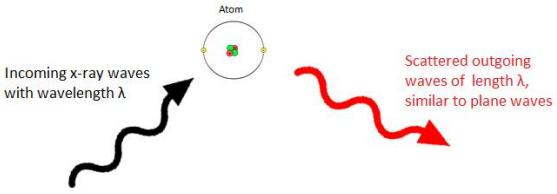
Using the image above, we see that the extra distance that the bottom wave travels is $\Delta = 2d\sin\theta$ and the phase shift by the same lower wave is $\phi = \frac{2\pi}{\lambda}(2d\sin\theta)$. Notice that the waves add constructively if $\phi = 2k\pi$, where $k \in \mathbb{Z}$. Therefore, a diffraction spot is seen when

$$\begin{aligned} 2k\pi &= \frac{2\pi}{\lambda}(2d\sin\theta) \\ \Rightarrow k\lambda &= 2d\sin\theta. \end{aligned}$$

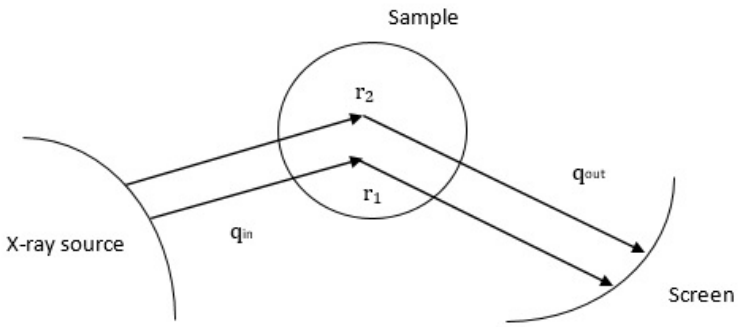
This is known as Bragg's Law. The key point to takeaway from this is that if you know the wavelength, λ, you can measure the distance between lattice planes.

2.2.5 Bragg's Law for Complicated Geometries

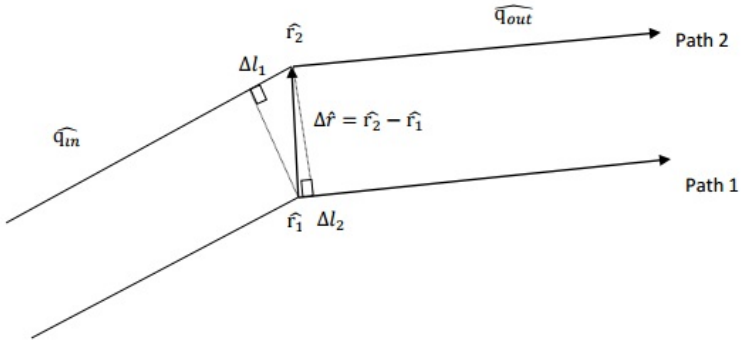
The way x-rays interact with matter is such: x-rays are high frequency electromagnetic waves, which interact with charged particles in an object. When x-ray waves reach an atom, the atoms electron cloud oscillates, which further radiates electromagnetic waves at the frequency of the oscillation frequency.



Consider an experiment where we model x-rays as plane waves, so the amplitude of a wave at a point $\vec{r}$ is written as $Ae^{i\vec{q}.\vec{r}}$, where A is a constant and q is the wave vector.



We must find the path length difference in order to see how the waves reflected by the two different points are different in phase.



Now we can see that the total path length difference is

$$\Delta l = \Delta l_1 - \Delta l_2.$$

From vector calculus, we know that

$$\Delta l_1 = \frac{\vec{q}_{in}.\Delta\vec{r}}{|\vec{q}_{in}|}$$

$$\Delta l_2 = \frac{\vec{q}_{out}.\Delta\vec{r}}{|\vec{q}_{out}|}$$

We can now substitute into the above equation to get

$$\Delta l = (\vec{q}_{in} - \vec{q}_{out}).\frac{\Delta\vec{r}}{|\vec{q}_{in}|}$$

since $|\vec{q}_{in}| = \vec{q}_{out} = \frac{2\pi}{\lambda}$. Now also define $\Delta\vec{q} = (\vec{q}_{out} - \vec{q}_{in})$ to get

$$\Delta l = -\Delta\vec{q}.\frac{(\vec{r}_2 - \vec{r}_1)}{|\vec{q}_{in}|}$$

We now say that the scattered amplitude from the two points is $A_1 + A_2 e^{(-i\Delta\vec{q}(\vec{r}_2 - \vec{r}_1))}$ since the phase shift is $\frac{2\pi}{\lambda}\Delta l$ and $|\vec{q}_{in}| = \frac{2\pi}{\lambda}$. In this case, A_1 is the strength that point 1 scatters x-rays, A_2 is the strength that point 2 scatters x-rays and the amplitude can be written as

$$e^{i\Delta\vec{q}(\vec{r}_1)}\left(A_1 e^{-i\Delta\vec{q}(\vec{r}_1)} + A_2 e^{-i\Delta\vec{q}(\vec{r}_2)}\right).$$

This experiment only takes into account for two points on a solid. To account for the rest of the points, we can use the fact that the signal that is detected is proportional to the complex absolute value of the amplitude squared. That is,

$$Intensity \propto \left| \int_{Solid} A(\vec{r}) e^{-i\Delta\vec{q}\vec{r}} d\vec{r} \right|^2$$

where $A(\vec{r})d\vec{r}$ is the strength that a solid volume $d\vec{r}$ centered at $\vec{r}$ can scatter x-rays.

It is now possible to relate the strength of which x-rays can be scattered to the physical properties of the solid object since electrons scatter electromagnetic radiation. We can assume that $A(\vec{r}) \propto \rho(\vec{r})$, and so

$$Intensity \propto \left| \int_{Solid} \rho(\vec{r}) e^{-i\Delta\vec{q}\vec{r}} d\vec{r} \right|^2$$

Since $\rho(\vec{r})$ is periodic, we can write (as shown above)

$$\rho(\vec{r}) = \sum_{hkl} \rho_{\vec{g}_{hkl}} \exp(i\vec{g}_{hkl}.\vec{r}).$$

Hence, we can substitute into the intensity proportion to see how intensity is written for a crystal, in the following way

$$Intensity(\Delta\vec{q}) = A(\Delta\vec{q}) \propto \left| \sum_{hkl} \rho_{\vec{g}_{hkl}} \int_{Solid} e^{i(\vec{g}_{hkl} - \Delta\vec{q})\vec{r}} d\vec{r} \right|^2$$

Looking only at the integral we find that

$$A(\Delta\vec{q}) = \begin{cases} 0 & \vec{g}_{hkl} \neq \Delta\vec{q} \\ large & \vec{g}_{hkl} = \Delta\vec{q} \end{cases}$$

That is, when $\vec{g}_{hkl} = \Delta\vec{q}$, the integral is 1 and represents the volume of the solid object. The key result here is that we will see bright reflections when $\Delta\vec{q}$ is the reciprocal lattice. This result will help us understand diffraction.

2.2.6 Intensity of a Crystal

$$\begin{aligned} Intensity(\Delta\vec{q}) &\propto \left| \int_{Solid} \rho(\vec{r}) e^{i\vec{q}.\vec{r}} d\vec{r} \right|^2 \\ &= \left| \sum_{uvw} \int_{unitcell} \rho(\vec{r} + \vec{R}_{uvw}) e^{-i\Delta\vec{q}(\vec{r} + \vec{R}_{uvw})} d\vec{r} \right|^2 \\ &= \left| \sum_{uvw} e^{-i\Delta\vec{q}.(\vec{R}_{uvw})} \int_{unitcell} \rho(\vec{r}) e^{-i\Delta\vec{q}.\vec{r}} d\vec{r} \right|^2. \end{aligned}$$

We can approximate $\rho(\vec{r})$ as a sum over atomic charges in the unit cell. That is

$$\rho(\vec{r}) = \sum_{j=1}^{\#ofatoms} \rho_j(\vec{r} - \vec{T_j})$$

where $\vec{T_j} = x_j\vec{a} + y_j\vec{b}$ is a basis vector, which specifies the position of an atom and can be found on the International Tables for Crystallography when given any lattice or definition of a lattice. We now have that

$$Intensity(\Delta\vec{q}) \neq 0$$

when $\vec{g}_{hkl} = \Delta\vec{q}$ and so it follows that

$$Intensity(\vec{g}_{hkl}) \propto \left| \sum_{uvw} e^{-i\vec{g}_{hkl}.(\vec{R}_{uvw})} \int_{unitcell} \rho_j(\vec{r} - \vec{T_j}) e^{-i\vec{g}_{hkl}.\vec{r}} d\vec{r} \right|^2$$

Here it is important to remember that $\vec{g}_{hkl}.\vec{R}_{uvw} = 2k\pi$, where $k \in \mathbb{Z}$. So, from above, it follows that

$$Intensity(\vec{g}_{hkl}) \propto \left| \sum_{uvw} \sum_{j} e^{-i\vec{g}_{hkl}.(\vec{T_j})} \int_{unitcell} \rho_j(\vec{r} - \vec{T_j}) e^{-i\vec{g}_{hkl}.(\vec{r} - \vec{T_j})} d\vec{r} \right|^2$$

We can now extend the the integral to all space, to get the *atomic form factor*, since

$$\int_{unitcell} \rho_j(\vec{r} - \vec{T_j}) e^{-i\vec{g}_{hkl}.(\vec{r} - \vec{T_j})} d\vec{r} \approx \int_{allspace} \rho_j(\vec{r}) e^{-i\vec{g}_{hkl}.(\vec{r})} d\vec{r} = f_j(\vec{g}_{hkl})$$

Hence we can write the intensity as the following

$$\begin{aligned} Intensity(\vec{g}_{hkl}) &\propto \left| \sum_{uvw} \sum_{j} f_j(\vec{g}_{hkl}) e^{-i\vec{g}_{hkl}.(\vec{T_j})} \right|^2 \\ &= N^2 \left| \sum_{j} f_j(\vec{g}_{hkl}) e^{-i\vec{g}_{hkl}.(\vec{T_j})} \right|^2 \end{aligned}$$

From this we have that N is the number of unit cells in the crystal and j is the index over all of the atoms in the unit cell. We also define

$$S(\vec{g}_{hkl}) = \sum_{j} f_j(\vec{g}_{hkl}) e^{-i\vec{g}_{hkl}.(\vec{T_j})}$$

as the *structure factor* of the crystal. Finally we reach the result that

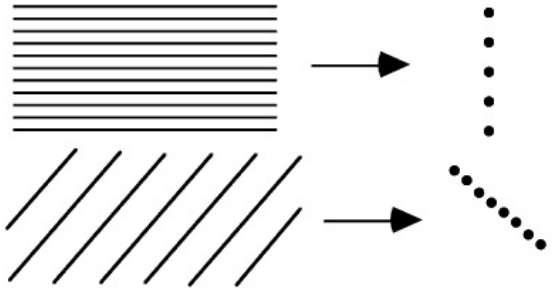
$$Intensity(\vec{g}_{hkl}) \propto N^2 |S(\vec{g}_{hkl})|^2.$$

2.2.7 Use of the Convolution Theorem

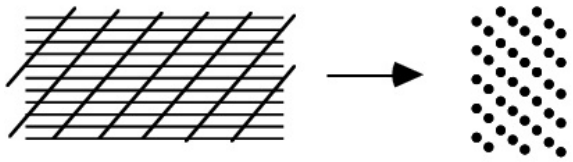
We have seen the convolution theorem for Fourier transforms in section 1.2.2. It turns out that the convolution theorem can be used to describe the reason that diffraction from a lattice in real space of unit cells gives a lattice in reciprocal space.

First, it is important to realize that the Fourier transform of parallel lines is a set of points that are perpendicular to the corresponding lines. The distance between the points and lines is inversely proportional to the space between the parallel lines. This can be shown through the use of Bragg's law in section 2.2.4. and increasing k. This gives us a series of diffraction vectors for the corresponding set of planes used in Bragg's law.

Below, we show that a set of horizontal lines with a small distance between them, produces a vertical row of points that have a large distance between them. Similarly, a set of diagonal lines that are far apart from each other, produces row of points perpendicular to the lines, that are close together.



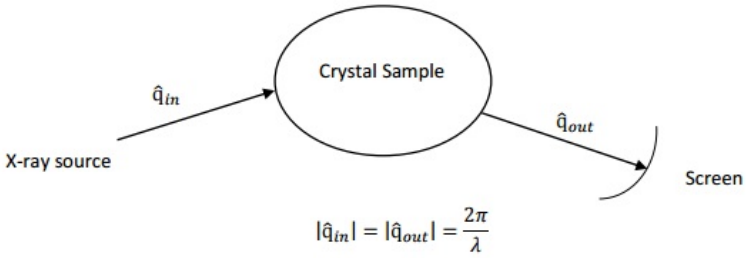
We now consider the multiplication of these sets of lines. This produces a set of points belonging to the intersections of the these lines and so the Fourier transform of this lattice of points is the convolution of the two individual rows of points. This then generates a reciprocal lattice.



This can be extended to three dimensions and so can be related to a crystal that is made up of a series of unit cells. The diffraction pattern of a crystal is therefore the product of the Fourier transform of the 3D lattice and the Fourier transform of a single unit cell. However, we have shown above that the Fourier transform of a lattice in real space is a reciprocal lattice. It follows that the diffraction pattern of a crystal is equivalent to the diffraction pattern of a unit cell at the points of the reciprocal lattice.

2.2.8 Summary

1. The experimental model.



2. We looked at wave scattering and found that

$$\Delta\vec{q} = (\vec{q}_{out} - \vec{q}_{in}) = \vec{g}_{hkl}$$

where $\vec{g}_{hkl}$ is a vector in the collection of the Reciprocal Lattice.

3. We also found that the intensity of the reflection when $\Delta\vec{q} = \vec{g}_{hkl}$, is shown by

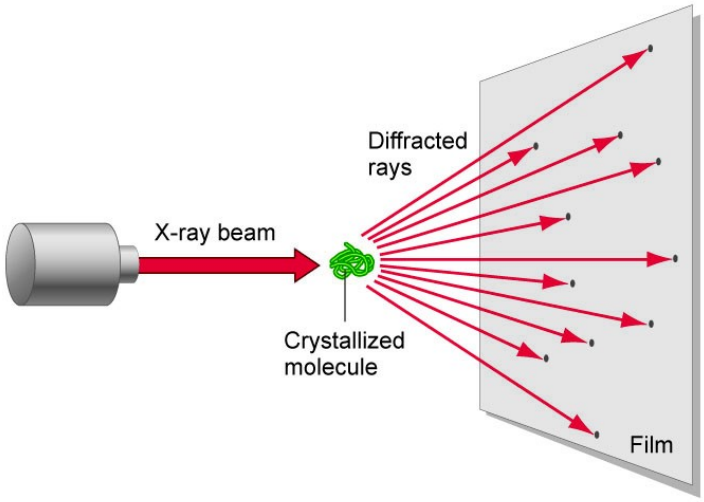
$$Intensity(\vec{g}_{hkl}) \propto N^2|S(\vec{g}_{hkl})|^2$$

where $S(\vec{g}_{hkl}) = \sum_j f_j(\vec{g}_{hkl})e^{-i\vec{g}_{hkl}\cdot(\vec{T}_j)}$

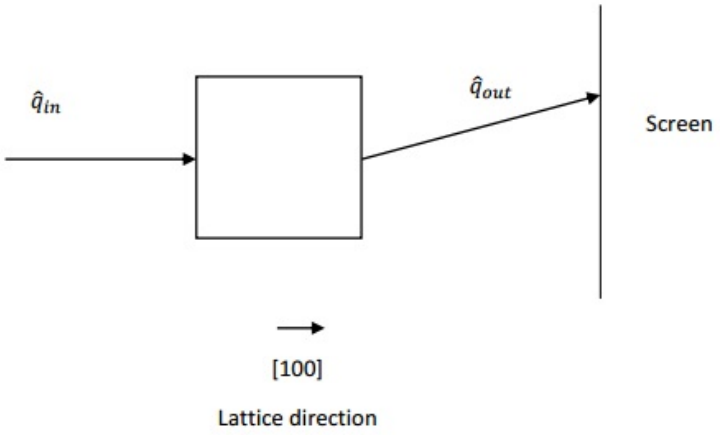
- the sum over the j indexes over all atoms in a unit cell.
- $f_j(\vec{g}_{hkl})$ is the atomic form factor for aton j. These are tabulated and can be found online.
- T_j is the basis vector for atom j. This can be found in the International Tables for Crystallography.

2.3 An Experimental Model

Let's take a look at another experimental x-ray model. An x-ray source is fixed across from a screen detector. Each x-ray beam leaving the crystal corresponds to $\Delta\vec{q} = \vec{g}_{hkl}$. The diffraction pattern shows the symmetry, and therefore geometry, of the crystal.



For example, x-rays go through a cubic p-lattice crystal, which has one atom per cell. (A cubic p-lattice is one of the simplest types of Bravais lattices. A Bravais lattice simply describes a repeating lattice in 3-D space).



Here we have

$$\vec{a}* = \frac{2\pi}{a}\vec{e}_1$$
$$\vec{b}* = \frac{2\pi}{a}\vec{e}_2$$
$$\vec{c}* = \frac{2\pi}{a}\vec{e}_3$$

and so it follows by definition that

$$\vec{g}_{hkl} = h\vec{a}* + k\vec{b}* + l\vec{c}* = \frac{2\pi}{a}(h\vec{e}_1 + k\vec{e}_2 + l\vec{e}_3).$$

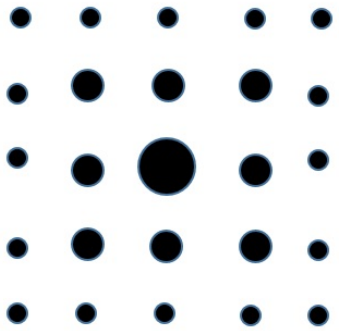
Bright spots appear when $\Delta\vec{q} = (\vec{q}_{out} - \vec{q}_{in}) = \vec{g}_{hkl}$, that is

$$\vec{q}_{out} = \vec{g}_{hkl} + \vec{q}_{in}$$

From the image above we seen that $\vec{q}_{in} = |\vec{g}_{in}|\vec{e}_1$ and so the bright spots appear when

$$\begin{aligned}\vec{q}_{out} &= \big(\frac{2\pi}{a}h + |\vec{q}_{in}|\big)\vec{e}_1 + \frac{2\pi}{a}k\vec{e}_2 + \frac{2\pi}{a}l\vec{e}_3 \\ &= \big(\frac{2\pi}{a}h + \frac{2\pi}{\lambda}\big)\vec{e}_1 + \frac{2\pi}{a}k\vec{e}_2 + \frac{2\pi}{a}l\vec{e}_3\end{aligned}$$

In this case, the diffraction pattern reflects the symmetry of the crystal and the plot looks like the following:

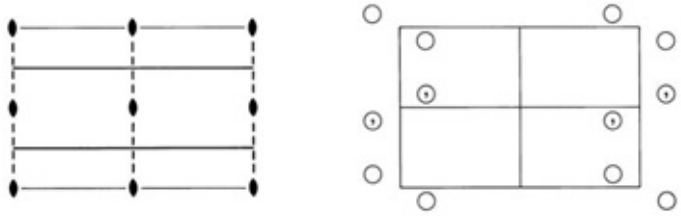


2.3.1 Example 1

Suppose a crystal has a plane group symmetry of type p2mg (from the International Tables for Crystallography). Also suppose that the crystal has Mg atoms at Wyckoff position a and Cl atoms at position d, with $x = 0.1$ and $y = 0.2$. A Wyckoff position is simply a place an atom can be found in a lattice, numerated by letters for each symmetry group. In this case, the Wyckoff letters are a and d. Below is the p2mg plane group on the International Tables for Crystallography.

$p2mg$ $2mm$ Rectangular

No. 7 $p2mg$ Patterson symmetry $p2mm$



Origin at 2 1 g

Asymmetric unit $0 \le x \le \frac{1}{4}$; $0 \le y \le 1$

Symmetry operations

(1) 1 (2) 2 0,0 (3) m $\frac{1}{4},y$ (4) a $x,0$

Generators selected (1); $t(1,0)$; $t(0,1)$; (2); (3)

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates			
4	d	1	(1) x,y	(2) $\bar{x},\bar{y}$	(3) $\bar{x}+\frac{1}{2},y$	(4) $x+\frac{1}{2},\bar{y}$
2	c	. m .	$\frac{1}{4},y$	$\frac{3}{4},\bar{y}$		
2	b	2 . .	$0,\frac{1}{2}$	$\frac{1}{2},\frac{1}{2}$		
2	a	2 . .	0,0	$\frac{1}{2},0$		

Reflection conditions

General:

$h0$: $h = 2n$

Special: as above, plus

no extra conditions

hk: $h = 2n$

hk: $h = 2n$

Maximal non-isomorphic subgroups

I [2] $p\,1\,1\,g$ $(pg, 4)$ 1; 4
[2] $p\,1\,m\,1$ $(pm, 3)$ 1; 3
[2] $p\,2\,1\,1$ $(p2, 2)$ 1; 2

IIa none

IIb [2] $p\,2\,g\,g$ $(\mathbf{b}' = 2\mathbf{b})$ (8)

Maximal isomorphic subgroups of lowest index

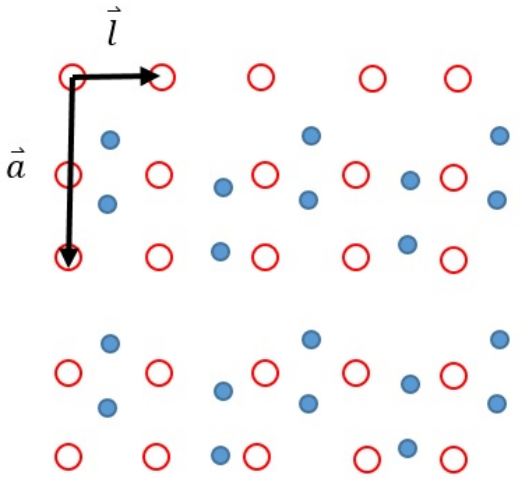
IIc [2] $p\,2\,m\,g$ $(\mathbf{b}' = 2\mathbf{b})$ (7); [3] $p\,2\,m\,g$ $(\mathbf{a}' = 3\mathbf{a})$ (7)

Minimal non-isomorphic supergroups

I none

II [2] $c\,2\,m\,m$ (9); [2] $p\,2\,m\,m$ $(\mathbf{a}' = \frac{1}{2}\mathbf{a})$ (6)

We can use this information to draw the following crystal structure.

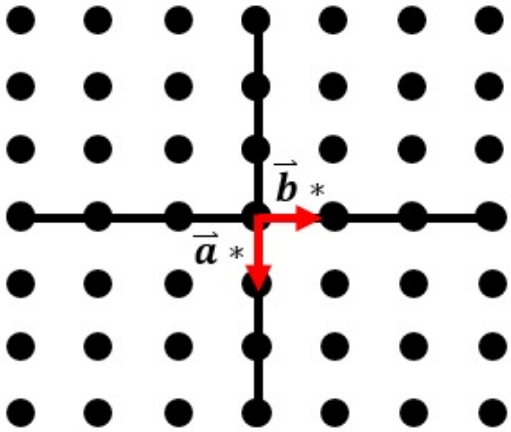


From here, it is possible to draw the reciprocal lattice. In this case, the reciprocal lattice vectors are

$$\vec{a}^* = \frac{-2\pi}{a}\vec{e}_2$$

and

$$\vec{b}^* = \frac{2\pi}{b}\vec{e}_1$$



We now look to compute the structure factor

$$S(\vec{g}_{hk}) = \sum_j (\vec{g}_{hk}) e^{-i\vec{g}_{hk}.\vec{T}_j}$$

In the case of this example, that is

$$\begin{aligned} S(\vec{g}_{hk}) &= f_{Mg}(1 + e^{-i\vec{g}_{hk}\cdot(x\vec{a}+y\vec{b})}) \\ &\quad + f_{Cl}\big(e^{-i\vec{g}_{hk}\cdot(x\vec{a}+y\vec{b})} + e^{i\vec{g}_{hk}\cdot(x\vec{a}+y\vec{b})} + e^{-i\vec{g}_{hk}\cdot((\frac{1}{2}-x)\vec{a}+y\vec{b})} + e^{-i\vec{g}_{hk}\cdot((\frac{1}{2}+x)\vec{a}+y\vec{b})}\big) \\ &= f_{Mg}(1 + e^{-ih\pi}) \\ &\quad + f_{Cl}\big(e^{-2i\pi(hx+ky)} + e^{2i\pi(hx+ky)} + e^{ih\pi}e^{-2i\pi(ky-hx)} + e^{-ih\pi}e^{-2i\pi(hx+ky)}\big) \end{aligned}$$

We now use *Euler's Identity*:

$$e^{i\pi} + 1 = 0$$

and *Euler's Equation*:

$$e^{ix} = \cos(x) + i\sin(x)$$

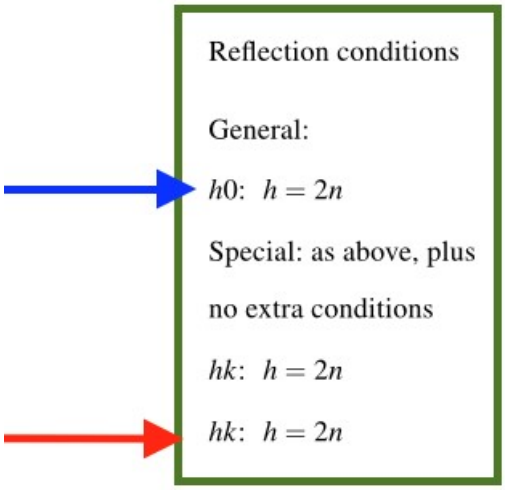
to get

$$S(\vec{g}_{hk}) = f_{Mg}(1 + (-1)^h) + f_{Cl}\big(2\cos(2\pi(hx+ky)) + 2(-1)^h e^{-2iky\pi}\cos(2hx\pi)\big)$$

Notice now that when h is odd we have that

$$S(\vec{g}_{hk}) = f_{Cl}\big(2\cos(2\pi(hx+ky)) + -2e^{-2iky\pi}\cos(2hx\pi)\big)$$

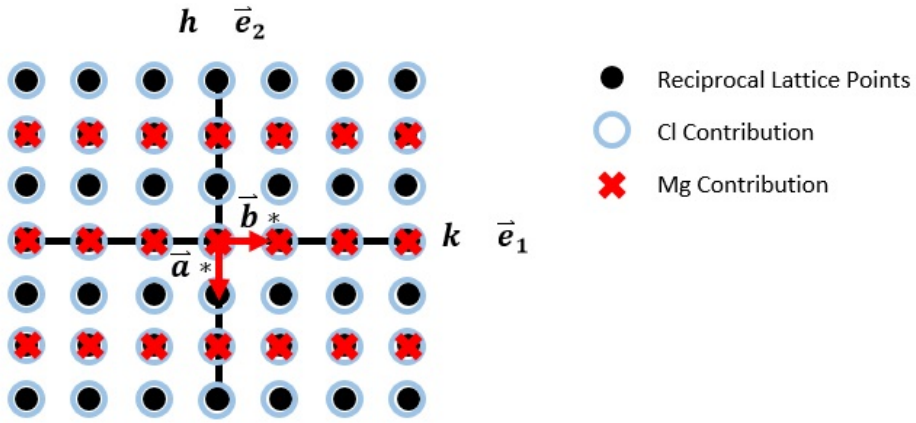
This indicates that there will not be any contribution from Mg atoms at $\vec{g}_{hk}$ and h odd. We can verify this by checking the International Tables of Crystallography for atoms at Wyckoff position a, to the left it reads $hk : h = 2n$ (which means spots only appear for $h = 2n$).



Now consider when $k = 0$:

$$S(\vec{g}_{hk}) = f_{Mg}(1 + (-1)^h) + 2f_{Cl}\cos(2hx\pi)(1 + (-1)^h)$$

This indicates that spots when $k = 0$ will only contribute intensity at $h = 2n$. Which, again, we can check International Tables of Crystallography, it reads $h0 : h = 2n$. Thus the diffraction pattern looks like the following:



2.3.2 Example 2

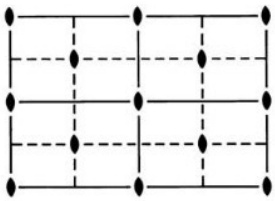
Lets do one last example to tie everything on x-ray Diffraction and Crystallography together. Given:

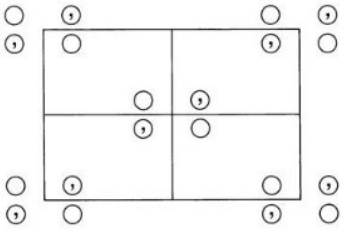
- A crystal with symmetry type c2mm (From the International Tables for Crystallography)
- Primitive lattice vectors are $\vec{a} = 1.1\vec{e}_y$ and $\vec{b} = \sqrt{2}\vec{e}_x$
- There are Ga atoms located at Wyckoff position b
- There are N atoms located at Wyckoff position f with $x = \frac{1}{3}$ and $y = \frac{1}{4}$

International Tables for Crystallography (2006). Vol. A, Plane group 9, p. 100.

$c2mm$ $2mm$ Rectangular

No. 9 $c2mm$ Patterson symmetry $c2mm$





Origin at $2mm$

Asymmetric unit $0 \le x \le \frac{1}{4}$; $0 \le y \le \frac{1}{2}$

Symmetry operations

For $(0,0)+$ set

(1) 1 (2) 2 $0,0$ (3) m $0,y$ (4) m $x,0$

For $(\frac{1}{2},\frac{1}{2})+$ set

(1) $t(\frac{1}{2},\frac{1}{2})$ (2) 2 $\frac{1}{4},\frac{1}{4}$ (3) b $\frac{1}{4},y$ (4) a $x,\frac{1}{4}$

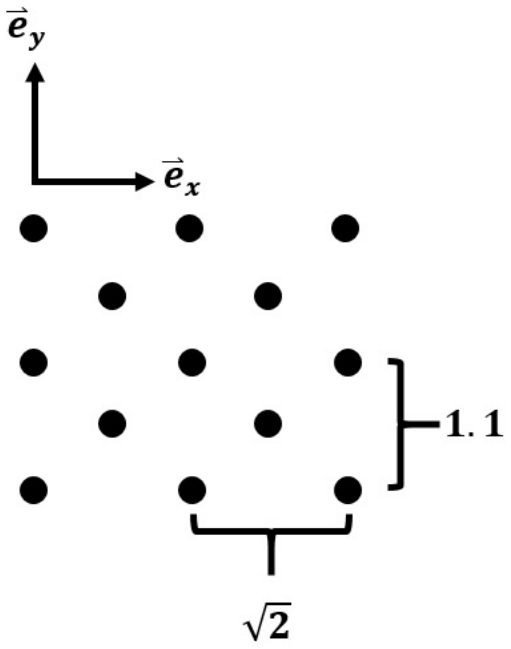
Generators selected (1); $t(1,0)$; $t(0,1)$; $t(\frac{1}{2},\frac{1}{2})$; (2); (3)

Positions

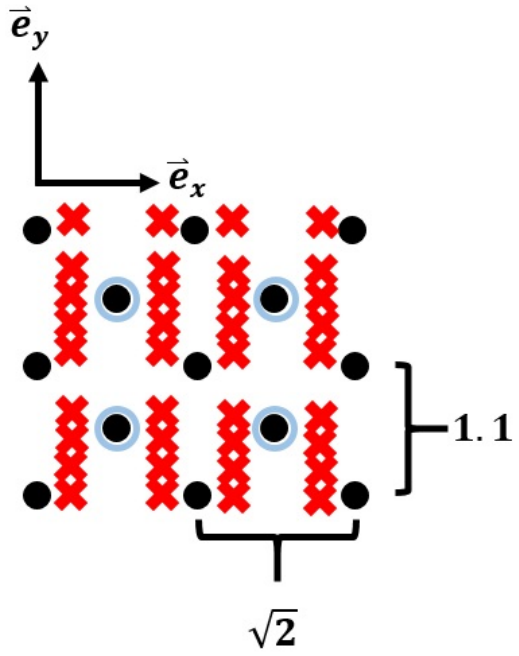
Multiplicity, Wyckoff letter, Site symmetry			Coordinates				Reflection conditions
			$(0,0)+$	$(\frac{1}{2},\frac{1}{2})+$			
8	f	1	(1) x,y	(2) $\bar{x},\bar{y}$	(3) $\bar{x},y$	(4) $x,\bar{y}$	General:
							hk: $h+k=2n$ $h0$: $h=2n$ $0k$: $k=2n$
4	e	$.m.$	$0,y$	$0,\bar{y}$			Special: as above, plus
4	d	$..m$	$x,0$	$\bar{x},0$			no extra conditions
4	c	$2..$	$\frac{1}{4},\frac{1}{4}$	$\frac{3}{4},\frac{1}{4}$			no extra conditions
2	b	$2mm$	$0,\frac{1}{2}$				hk: $h=2n$
2	a	$2mm$	$0,0$				no extra conditions

Solution

- Plot the lattice points using the given the symmetry type, International Tables for Crystallography, and primitive lattice vectors.



- Plot Ga and N atoms given the symmetry type, International Tables for Crystallography the Wyckoff letters (this is our crystal).



- Find the reciprocal lattice vectors using the constraints discussed previously:

– Primitive lattice vectors:

$$\vec{a} = 1.1\vec{e}_y$$
$$\vec{b} = \sqrt{2}\vec{e}_x$$

– Lattice points:

$$\vec{R}_{uv} = u\vec{a} + v\vec{b}$$

– Reciprocal lattice points:

$$\vec{g}_{hk} = h\vec{a}* + k\vec{b}*$$

– Constraints

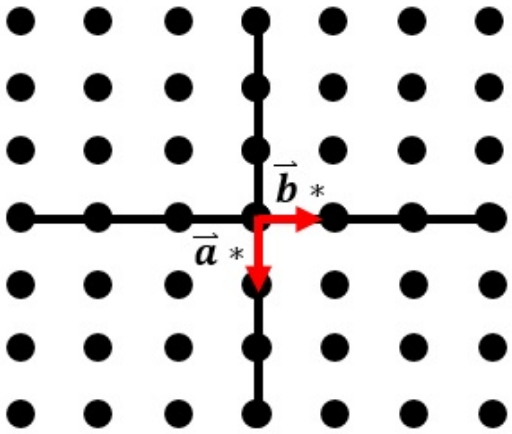
$$\vec{a}.\vec{a}* = 2\pi, \quad \vec{b}.\vec{a}* = 0,$$
$$\vec{a}.\vec{b}* = 0, \quad \vec{b}.\vec{b}* = 2\pi,$$

From this information we get the following:

$$\vec{a}* = \frac{2\pi}{\vec{a}} = \frac{2\pi}{-1.1\vec{e}_x}$$

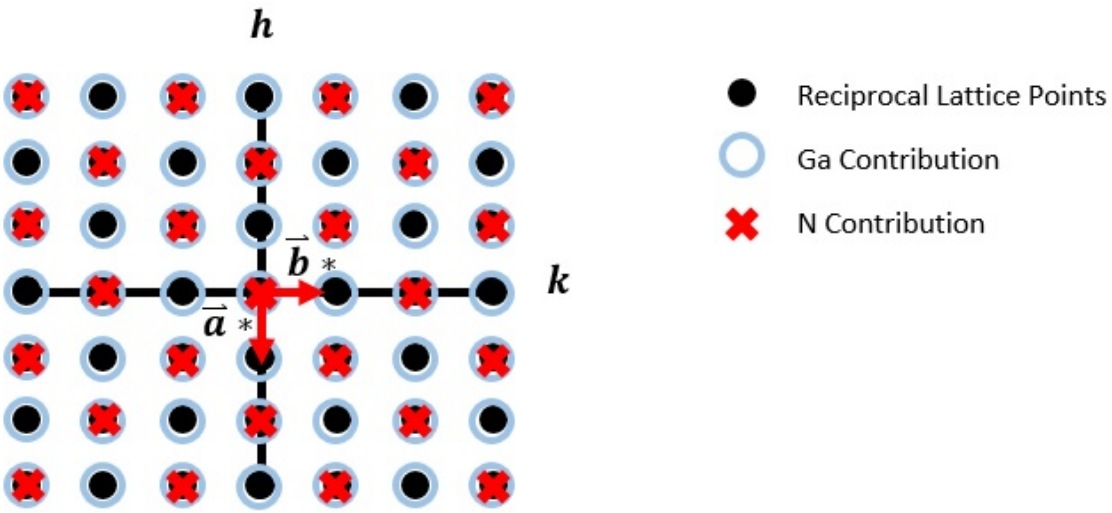
$$\vec{b}* = \frac{2\pi}{\vec{b}} = \frac{2\pi}{\sqrt{2}\vec{e}_x}$$

- Plot the reciprocal lattice points



- Plot each atoms intensity contribution to the diffraction pattern on the reciprocal lattice using the International Tables for Crystallography. (This is the diffraction pattern!)

- $hk : h + k = 2n$ means that for all h and k where $h + k$ is an even number, there will be a bright spot.
- $h0 : h = 2n$ means when $k = 0$, all even h will have a bright spot (h-axis).
- $0k : k = 2n$ means when $h = 0$, all even k will have a bright spot (k-axis).

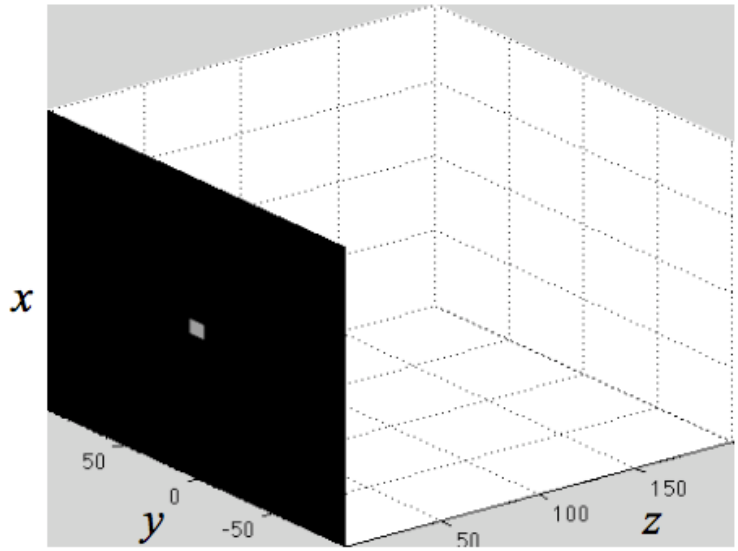


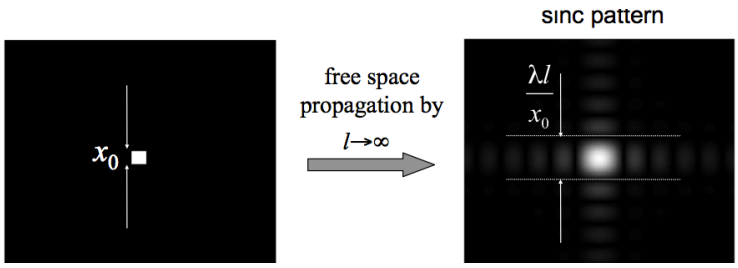
2.4 X-Ray Code

We ran out of time to do the full X-ray diffraction code that we wanted, so we worked on a related problem, Fraunhofer diffraction, or far field diffraction. We have pseudo-code for Fraunhofer.

Consider the classic slit scenario. Fraunhofer diffraction is a model we can use to model the diffraction of waves when the diffraction pattern is viewed at a long distance from the slit.

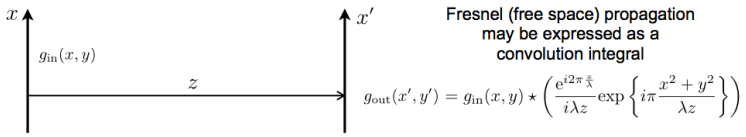
We set it up as follows.





Now, g_{in} is the amplitude of the incoming wave (that is, at the slit), and g_{out} is the amplitude when it hits the far surface. We model g_{in} as a rect function with some width a.

Using the Fresnel model (a model for free space propagation) we see that g_{out} can be expressed as a convolution.



Now from the convolution theorem, we know that whenever we see a convolution we can find a Fourier transform to compute instead for better computational efficiency. If we work through the Fraunhofer model (which includes z), as z gets very large:

$$g_{out}(x', y'; z) \propto \int g_{in}(x, y) \exp(i\pi \frac{(x' - x)^2 + (y' - y)^2}{\lambda z}) dx dy$$

Eventually we come to find out that it's proportional to $G_{in}(u, v)$, that is, the Fourier transform of g_{in} (with some variable changes for $u \equiv \frac{x'}{z\lambda}$, $v \equiv \frac{y'}{z\lambda}$).

So we end up with

$$\begin{aligned} g_{\text{in}}(x, y) &= \text{rect}\left(\frac{x}{x_0}\right)\text{rect}\left(\frac{y}{y_0}\right) \\ G_{\text{in}}(u, v) &= x_0 y_0 \text{sinc}\left(x_0 u\right)\text{sinc}\left(y_0 v\right) \\ g_{\text{out}}(x', y'; z \to \infty) &\propto \text{sinc}\left(\frac{x_0 x'}{\lambda z}\right)\text{sinc}\left(\frac{y_0 y'}{\lambda z}\right). \end{aligned}$$

Note that sinc is the FT of a rect function. At $x = 0$ it is defined to be 1.

$$sinc = \frac{\sin x}{x}$$

Our pseudo-code takes care of 1D: it finds the FT of a rect function, and is able to get g_{out} (also called U sometimes). However, we can't have a continuous rect function, so we create a discrete one and as a result get a discrete sinc to use for our g_{out}.

3 Work Description

Fraser - *Math/Formatting Lead.* Worked on the derivations of the Fourier/Convolution theorem and mathematical model translation. Worked on the statement of the problem as well as typing up some of the extra deliverables in Latex. Also was a part of the pseudocode discussions and proofread/commented code.

CJ - *Math Lead.* Worked on the derivations of the Fourier/Convolution/Fourier Slice theorems and helped type them up in the writeup. Worked on converting many of the MSE extra deliverables from hand drawn pictures to Latex. Worked on the statement of the problem with Fraser and was a part of the pseudocode discussions and proofread/commented code.

Tony - *Presentation/Code Lead.* Worked on the X-ray diffraction pseudocode for extra deliverables as well as was involved in working on the code for project-fst, back-projection, estimate-orientations. Also, helped make the slides for the powerpoint presentation.

Tanmay - *Code Lead.* Primarily worked on the basic deliverables code (project-fst, back-projection, estimate-orientations). Worked on the writeup as well as worked on the X-ray diffraction pseudocode and extra deliverables. Also helped format the writeup and produced the images/results that are presented in the writeup.

Nisha - *MSE/Extra Deliverables Lead.* Was the lead in the extra deliverable portions and was involved in working on the fourier series derivations, images, and examples. Also worked on the introduction to the biological problems and proofread the code.

4 Appendix

4.1 Problem 1

Compute the Fourier series for $f(x) = x^2$, periodic on $[-\pi, \pi]$, and use it to derive the identity

$$\sum_{1}^{\infty} \frac{(-1)^n}{n^2} = -\frac{\pi^2}{12}.$$

Solution.

Notice that, since $f(x) = x^2$ is an even function, we know that the sine coefficients will equal zero and so

$$f(x) = A_0 + \sum_{k=1}^{\infty} (A_k \cos(kx)).$$

Using the definition above we find the following coefficients:

$$A_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} x^2 dx = \frac{\pi^2}{3}$$

$$A_n = \frac{2}{2\pi} \int_{-\pi}^{\pi} x^2 \cos(nx) dx = \frac{4(-1)^n}{n^2}$$

We now substitute these into the FS formula to get the FS for $f(x) = x^2$:

$$f(x) = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos(nx)$$

Now that we have the FS for $f(x)$ we can substitute values of x that are within the domain. We also know that x^2 is continuous on the domain, and that the derivative is also continuous on the domain. Thus, we consider $f(0)$=0, that is:

$$f(0) = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = 0$$

$$\Rightarrow \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = -\frac{\pi^2}{12}.$$

4.2 Problem 2

Suppose f and g are two functions $\mathbb{R} \to \mathbb{R}$ with $\left| \int_{-\infty}^{\infty} f(x)dx \right| < \infty$ and $\left| \int_{-\infty}^{\infty} g(x)dx \right| < \infty$. The convolution of f and g, denoted $f \star g$, is give by

$$(f \star g)(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

If $\mathcal{F}$ is the Fourier Transform of f, prove that

$$\mathcal{F}(f \star g) = \mathcal{F}(f).\mathcal{F}(g),$$

where . denotes pointwise multiplication of the functions. This result is called the convolution theorem.

Solution.

We first notice that we can rewrite the problem as: Show that

$$h(z) = (f \star g)(z) \Leftrightarrow \hat{h}(\omega) = \hat{g}(\omega).\hat{f}(\omega).$$

($\Rightarrow$)

$$\begin{aligned}
& h(z) = (f \star g)(z) \\
\implies \mathcal{F}(f \star g) = \hat{h}(\omega) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x)g(z-x)dx \right) \exp(-2\pi i\omega.z)dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)g(z-x) \exp(-2\pi i\omega.z)dzdx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)g(y) \exp(-2\pi i\omega.(x+y))dydx \\
&= \int_{-\infty}^{\infty} f(x) \exp(-2\pi i\omega.x)dx. \int_{-\infty}^{\infty} g(y) \exp(-2\pi i\omega.y)dy \\
&= \hat{g}(\omega).\hat{f}(\omega) \\
&= \mathcal{F}(f).\mathcal{F}(g)
\end{aligned}$$

In the above proof we use the substitution $y = z - x$.

($\Leftarrow$) The proof of the converse uses the uniqueness property that follows from the inverse transform. That is that for continuous functions

$$f(z) = g(z) \iff \hat{f}(\omega) = \hat{g}(\omega).$$

We have seen already that $(f \star g)(z)$ transforms to $\hat{g}(\omega).\hat{f}(\omega)$ and so by using this property, the inverse transform of $\hat{g}(\omega).\hat{f}(\omega)$ is $(f \star g)(z)$. It follows that

$$h(z) = (f \star g)(z) \Leftrightarrow \hat{h}(\omega) = \hat{g}(\omega).\hat{f}(\omega)$$

and thus:

$$\mathcal{F}(f \star g) = \mathcal{F}(f).\mathcal{F}(g).$$

4.3 Problem 3

The convolution theorem is important in applications because it allows us to compute convolutions by multiplying their Fourier transforms, which is faster than computing the convolution directly. Using big-oh notation, compare the algorithmic complexity of computing $f \star g$ directly with computing $f \star g$ by appealing to the convolution theorem. Don't forget to include the cost of performing the Fourier Transform in your analysis (what happens if you use the naive FT algorithm? what happens if you use the FFT?)

Solution.

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(z)g(t-z)dz$$

Big-O notation means "up to," as opposed to the big-Omega and big-Theta notations which define different bounds. If we say $f(x) = O(n)$ for instance, it roughly means $f(x) \leq kn$ for some constant k.

Convolution is typically defined with the integral definition. We can think of convolution in discrete terms too (with summation). In that case, if we have two signals each with n points, then for each valid t of $(f \star g)(t)$ we are have up to n terms to compute. There are up to n valid t for $(f \star g)(t)$. The runtime as a result is $O(n^2)$.

$$(f \star g)(t) = \sum_{z=-\infty}^{\infty} f(z)g(t-z)$$

For 2D, it becomes $O(n^4)$. There are $O(n^2)$ valid (x, y) pairs, and for each pair there are $O(n^2)$ terms.

$$(f \star g)(x, y) = \sum_{z=-\infty}^{\infty} \sum_{w=-\infty}^{\infty} f(z, w)g(x-z, y-w)$$

The convolution theorem states that

$$h(t) = (f \star g)(t) \Leftrightarrow \hat{h}(\omega) = \hat{g}(\omega).\hat{f}(\omega)$$

If we use the convolution theorem, the calculation of $(f \star g)(t)$ becomes the multiplying of two values of the Fourier transforms of g and f. So we have to calculate the Fourier transforms once (and when we do this we computer each value), and then for each $h(t)$ we just have to fetch the

value and multiply, so we are limited to the speed we can find the Fourier transforms of f and g.

The regular DFT algorithm has a runtime of $O(n^2)$, where n is how many data points in the discrete function. If we have a function f with N data points, then for each ω we have $O(N)$ terms, so we have $O(N^2)$ terms total.

$$\hat{y}[\omega] = \sum_{k=0}^{N-1} y[k] \exp\left(-2\pi i . \frac{k\omega}{N} \right).$$

The FFT is $O(n \log n)$. If we use this, then convolution of two functions with n points becomes $O(n \log n)$ too (by using convolution theorem and replacing naive DFT).

5 Bibliography

1. "Biological Small-angle Scattering." Wikipedia. Wikimedia Foundation, n.d. Web. 1 Dec. 2016.

2. Blanchet, Clement E., and Dmitri I. Svergun. "Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution." European Molecular Biology Laboratory, 3 Dec. 2012. Web. 1 Dec. 2016.

3. "Crystallography & Cryo-electron Microscopy." Methods in Molecular Biophysics. Rutgers, New Jersey. Rutgers University. Web. 1 Dec. 2016.

4. Kelly, Anthony, and G. W. Groves. Crystallography and Crystal Defects. London: Longman, 1970. Print.

5. Nye, J. F. Physical Properties of Crystals: Their Representation by Tensors and Matrices. Oxford: Clarendon, 1985. Print.

6. "Small Angle X-ray Scattering." BIOSAXS. BIOSAXS GmbH, n.d. Web. 1 Dec. 2016.

7. Sundararajan, D. Discrete Fourier Transform: Theory, Algorithms and Applications. N.p.: World Scientific, 2001. Print.

8. Wall, Michael E., Paul D. Adams, James S. Fraser, and Nicholas K. Sauter. "Diffuse X-Ray Scattering to Model Protein Motions." Cell. Structure, 4 Feb. 2014. Web. 1 Dec. 2016.

9. Weeks, S. D., and J. Bouckaert. "Small-angle X-ray Scattering to Obtain Models of Multivalent Lectin-glycan Complexes." Methods in Molecular Biology. U.S. National Library of Medicine, 22 July 2014. Web. 1 Dec. 2016.

10. "X-ray Crystallography." Wikipedia. Wikimedia Foundation, n.d. Web. 1 Dec. 2016.

11. "2.710 Optics Lecture 17 Spring 2009." MIT OCW. Web.

12. Douglass, Kyle M. "Approximating Diffraction Patterns of Rectangular Apertures with the F." Kyle M. Douglass. N.p., 01 Feb. 2016. Web.

13. Dasgupta, Sanjoy, Christos H. Papadimitriou, and Umesh Virkumar. Vazirani. Algorithms. Boston: McGraw-Hill Higher Education, 2008. Print.