

Tutorial 3: Approximate phases and peak search

James Foadi
email j.foadi@bath.ac.uk

June, 2019

This tutorial follows part of the demonstration in reference [crn].

1 Peak search with `local_maxima`

The main goal of structural crystallography is to find the positional coordinates of all atoms in a given structure. The quantity that can be calculated using the experimental data is the electron density. The atomic coordinates can be extracted from the electron density as positions corresponding to the density peaks' maxima. The function in `crone` responsible for extracting such maxima is `local_maxima`. In order to appreciate how this function works, let's load the structure called *cyanate*.

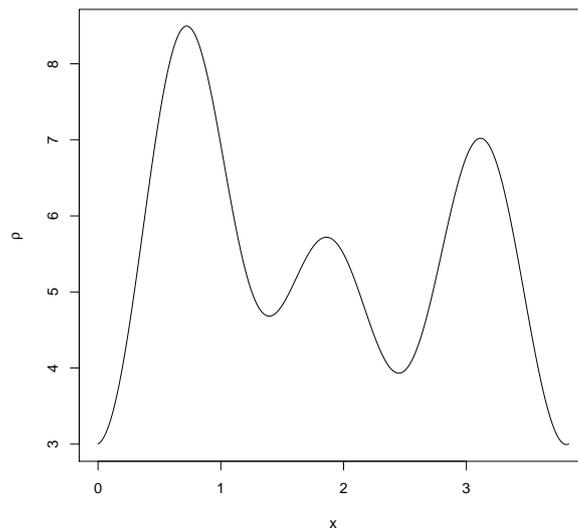
```
library(crone)
sdata <- load_structure("cyanate")
sdata

## $a
## [1] 3.833
##
## $SG
## [1] "P1"
##
## $x0
## [1] 0.707 1.870 3.126
##
## $Z
```

```
## [1] 8 6 7
##
## $B
## [1] 10.000 13.333 11.429
##
## $occ
## [1] 1 1 1
```

From the data in the list `sdata` we can create the exact analytic electron density using function `structure_gauss`.

```
rtmp <- structure_gauss(sdata,N=1000) # Grid with 1000 points
plot(rtmp$x,rtmp$rr,type="l",xlab="x",ylab=expression(rho))
```

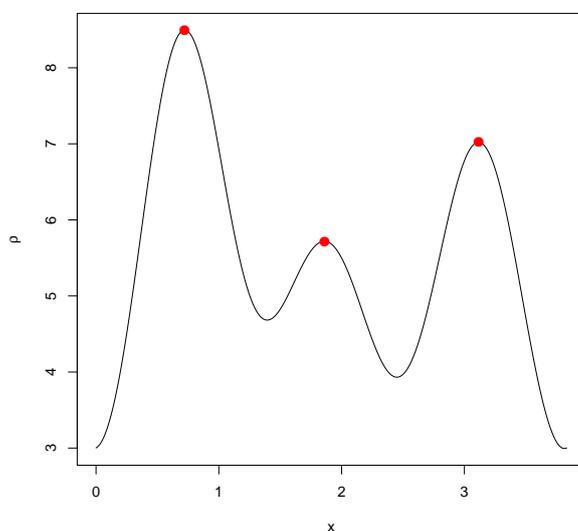


This structure is made of three atoms (three peaks), roughly located between 0 and 1, close to 2 and close to 3. In order to find out the correct location of the peaks, and thus the coordinates of the atoms, we make use of `local_maxima`. The maxima and minima in the electron density are returned as integers corresponding to specific cells of the electron density array.

```
idx <- local_maxima(rtmp$rr)
idx
## [1] 189 486 813
```

Such indices can then be used in conjunction with the electron density to find out which integer corresponds to maxima and, accordingly, display the peaks' location.

```
plot(rtmp$x,rtmp$rr,type="l",xlab="x",ylab=expression(rho))
for (i in 1:length(idx)) {
  points(rtmp$x[idx[i]],rtmp$rr[idx[i]],pch=16,cex=1.5,col=2)
}
```



```
# Peak coordinates found and comparison with published values
for (i in 1:length(idx)) {
  line <- sprintf("%5.3f %5.3f\n",rtmp$x[idx[i]],sdata$x0[i])
  cat(line)
}

## 0.721 0.707
## 1.861 1.870
## 3.116 3.126
```

Peaks coordinates are not exactly equal to the published-structure coordinates because B factors and gaussians' widths slightly shift them. Fixing B to zero and using smaller values of the atomic number (Z) will improve accuracy, as shown in the following snippet where $Z = 1$ and $B = 0$ for all atoms.

```

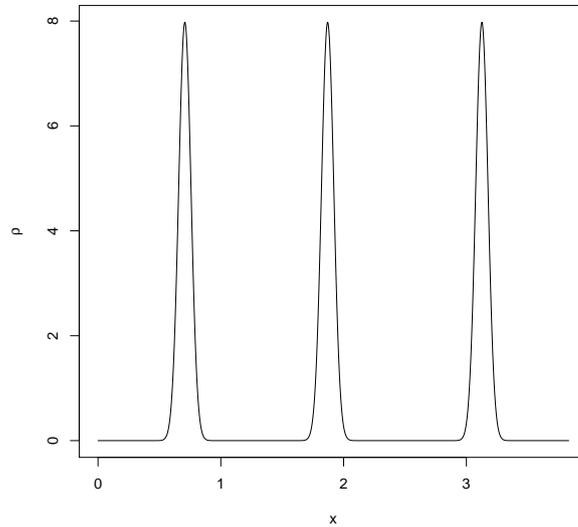
# Change all 3 B factors to 0
sdata$B <- c(0,0,0)

# Change all atoms to hydrogens
sdata$Z <- c(1,1,1)
sdata

## $a
## [1] 3.833
##
## $SG
## [1] "P1"
##
## $x0
## [1] 0.707 1.870 3.126
##
## $Z
## [1] 1 1 1
##
## $B
## [1] 0 0 0
##
## $occ
## [1] 1 1 1

# Electron density
rtmp <- structure_gauss(sdata,N=1000) # Grid with 10000 point
plot(rtmp$x,rtmp$rr,type="l",xlab="x",ylab=expression(rho))

```



```

# Local maxima
idx <- local_maxima(rtmp$rr)
idx

## [1] 185 488 816

for (i in 1:length(idx)) {
  line <- sprintf("%5.3f %5.3f\n",rtmp$x[idx[i]],sdata$x0[i])
  cat(line)
}

## 0.706 0.707
## 1.869 1.870
## 3.127 3.126

```

2 Peak search when the density come from a correct Fourier synthesis

The interest in finding the coordinates of the electron density peaks' coordinates exists because the atoms' positions are obviously unknown. The electron density is in general computed not as an exact analytic expression made out of gaussians, but as a Fourier synthesis in which the Fourier amplitudes are obtained experimentally and the Fourier phases have to be found

with one of the many *phasing methods* existing in structural crystallography. In what follows we will assume that the phases found coincide with the correct phases; in the next section we will explore the effect on the electron density of both errors in the Fourier amplitudes and phases.

To start, let's reload the cyanate data with the unmodified Z and B values. Then let's compute the pristine structure factors for Miller indices $h = 0, 1, \dots, 20$. This will provide us with correct Fourier amplitudes and phases.

```
sdata <- load_structure("cyanate")
hidx <- 0:20
ftmp <- strufac(hidx,sdata)

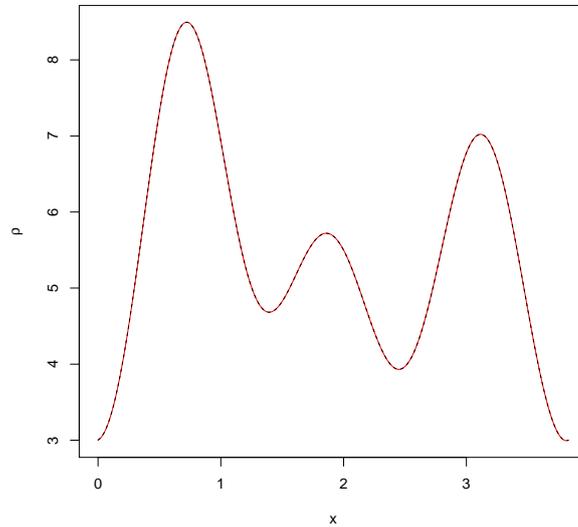
# This should coincide with 8(O) + 6(C) + 7(N) = 21
ftmp$Fmod[1]

## [1] 21

# Exact analytic electron density
rtmp0 <- structure_gauss(sdata,N=1000)

# Electron density as Fourier synthesis
rtmp1 <- fousynth(sdata$a,ftmp$Fmod,ftmp$Fpha,hidx,N=1000)

# The two densities should overlap nicely
plot(rtmp0$x,rtmp0$rr,type="l",col=2,xlab="x",
      ylab=expression(rho))
points(rtmp1$x,rtmp1$rr,type="l",lty=2)
```



We should expect, given the very good overlap of the curves, no much difference in the coordinates of peaks' maxima for both the analytic and Fourier-synthesis densities.

```

idx0 <- local_maxima(rttmp0$rr) # Analytic-density case
idx1 <- local_maxima(rttmp1$rr) # Fourier-synthesis case
for (i in 1:length(idx)) {
  line <- sprintf("%5.3f %5.3f %5.3f\n",
                  rttmp0$x[idx0[i]], rttmp1$x[idx1[i]], sdata$x0[i])
  cat(line)
}

```

```

## 0.721  0.721  0.707
## 1.861  1.859  1.870
## 3.116  3.116  3.126

```

3 Correct phases and biased Fourier amplitudes

The Fourier amplitudes derived from experimental data are in general affected by experimental errors. This is reflected in their values being different

from the corresponding calculated values. As seen in Tutorial 01, experimental (or *observed*) amplitudes can be loaded in memory using the function `load_data`. A comparison between calculated and experimental amplitudes is carried out in the following snippet code.

```
fdata <- load_data(sname="cyanate")

# Names of fdata elements
ntmp <- names(fdata)
for (a in ntmp) {
  ltmp <- sprintf("%s ",a)
  cat(ltmp)
}

## a SG hidx Fobs sigFobs Phicalc

# The experimental amplitudes are available only
# for 10 Miller indices (no h=0 term as it's not
# experimentally available)
hidx <- fdata$hidx
hidx

## [1] 1 2 3 4 5 6 7 8 9 10

# Let's re-compute calculated s.f.
ftmp <- strufac(hidx,sdata)

# The observed amplitudes should be different
# from the correct amplitudes.
# The available experimental precision is to 3 decimals
for (i in 1:length(hidx)) {
  line <- sprintf("%10.3f %10.3f\n",
                  fdata$Fobs[i],ftmp$Fmod[i])
  cat(line)
}

##      1.847      1.237
##      2.645      2.260
##      2.574      2.845
##      0.253      0.176
##      0.062      0.067
```

```
##      0.012      0.008
##     -0.010      0.000
##     -0.007      0.000
##      0.002      0.000
##      0.002      0.000
```

The differences observed above are due to experimental factors affecting the intensities collected. As these, in the specific case, are not real but simulated factors, we should consider them only as an illustration of the fact that observed amplitudes are in general different from calculated amplitudes. The ways in which observed amplitudes are simulated is explained in `crone` documentation. These essentially act on the number of photons scattered (number simulated using the Poisson distribution) and on the average shift of the atoms composing the structure from their equilibrium position as due to lattice imperfections (these add up to the vibrational motion described by the B factors). Some of the observed amplitudes are negative as simulation of the fact that often real 3D small amplitudes are negative as a consequence of the methods of extraction of the intensities from the diffraction images, due to background subtraction.

```
# The phases are the same (as they are calculated)
# The available precision of the phases in the data file
# is to 1 decimal.
for (i in 1:length(hidx)) {
  line <- sprintf("%10.1f  %10.1f\n",
                 fdata$Phicalc[i],ftmp$Fpha[i])
  cat(line)
}

##      80.2      80.2
##     175.4     175.4
##     179.5     179.5
##     -67.4     -67.4
##      -7.3      -7.3
##      14.3      14.3
##     119.3     119.3
##     175.2     175.2
##    -132.0    -132.0
##     -50.0     -50.0
```

Let's compare the electron densities with correct phases using both calculated and observed amplitudes.

```

hidx <- 1:10

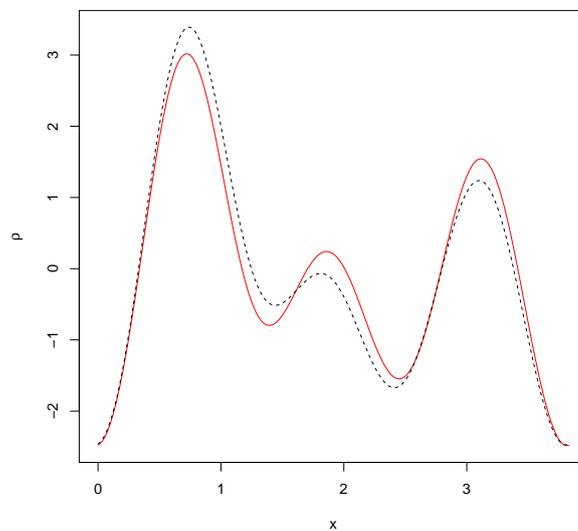
# Correct electron density
rtmp0 <- fousynth(sdata$a,ftmp$Fmod,ftmp$Fpha,hidx,N=1000)

# Biased electron density
rtmp1 <- fousynth(sdata$a,fdata$Fobs,fdata$Phicalc,hidx,N=1000)

# Min and max to include all density in the plot
m <- min(rtmp0$rr,rtmp1$rr)
M <- max(rtmp0$rr,rtmp1$rr)

# Comparison
plot(rtmp0$x,rtmp0$rr,type="l",xlab="x",ylab=expression(rho),
     col=2,ylim=c(m,M))
points(rtmp1$x,rtmp1$rr,type="l",lty=2)

```



Clearly there are differences between the two electron densities. These should be reflected in differences between the coordinates corresponding to the peaks.

```

idx0 <- local_maxima(rtmp0$rr) # Correct amplitudes
idx1 <- local_maxima(rtmp1$rr) # Observed amplitudes

```

```

for (i in 1:length(idx)) {
  line <- sprintf("%5.3f %5.3f %5.3f\n",
                 rtmp0$x[idx0[i]],rtmp1$x[idx1[i]],sdata$x0[i])
  cat(line)
}

## 0.721  0.740  0.707
## 1.859  1.809  1.870
## 3.116  3.101  3.126

```

4 Biased phases and correct Fourier amplitudes

Also phases affect the electron density, like amplitudes (in 3D crystallography they play, in fact, a greater role in determining the electron density). To see this let's try and change the correct and calculated phases with some random errors of approximately 20 degrees and keep the correct amplitudes. We are interested, as before, to explore the variation of the electron density and of the related peaks coordinates.

```

hidx <- 1:10

# Random errors added to phases
set.seed(8761)
pha_new <- ftmp$Fpha + rnorm(length(ftmp$Fpha),mean=30,sd=10)
idx <- which(pha_new < -180)
if (length(idx) > 0) {
  pha_new[idx] <- pha_new[idx]+360
}
idx <- which(pha_new > 180)
if (length(idx) > 0) {
  pha_new[idx] <- pha_new[idx]-360
}

# Correct electron density
rtmp0 <- fousynth(sdata$a,ftmp$Fmod,ftmp$Fpha,hidx,N=1000)

# Biased electron density

```

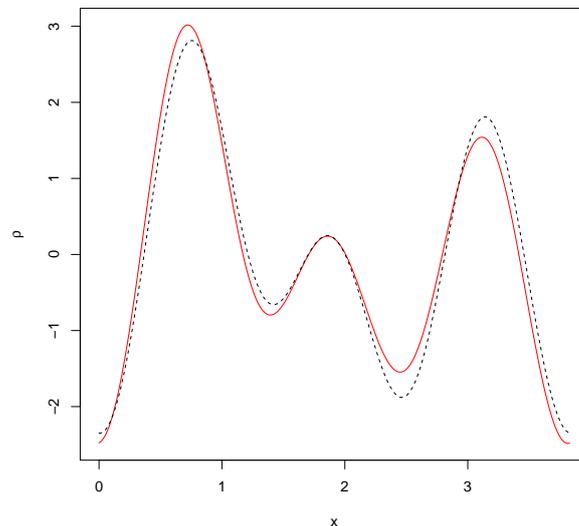
```

rtmp1 <- fousynth(sdata$a,ftmp$Fmod,pha_new,hidx,N=1000)

# Min and max to include all density in the plot
m <- min(rtmp0$rr,rtmp1$rr)
M <- max(rtmp0$rr,rtmp1$rr)

# Comparison
plot(rtmp0$x,rtmp0$rr,type="l",xlab="x",ylab=expression(rho),
      col=2,ylim=c(m,M))
points(rtmp1$x,rtmp1$rr,type="l",lty=2)

```



The biased density seems to include now a "systematic" rather than "random" error (the whole density seems to be slightly shifted towards the right). This type of bias is present in a stronger form in 3D crystallography. Obviously, also the peaks coordinates will be affected by the same systematic bias.

```

idx0 <- local_maxima(rtmp0$rr) # Analytic-density case
idx1 <- local_maxima(rtmp1$rr) # Fourier-synthesis case
for (i in 1:length(idx)) {
  line <- sprintf("%5.3f %5.3f %5.3f\n",
                  rtmp0$x[idx0[i]],rtmp1$x[idx1[i]],sdata$x0[i])
  cat(line)
}

```

```
## 0.721 0.755 0.707
## 1.859 1.859 1.870
## 3.116 3.139 3.126
```

5 Biased phases and biased amplitudes

In real situations both amplitudes and phases are biased; the amplitudes are biased by experimental errors and the phases are biased because they reflect a wrong model. The resulting electron density will contain the effect of both types of bias.

```
# Correct electron density
rtmp0 <- fousynth(sdata$a,ftmp$Fmod,ftmp$Fpha,hidx,N=1000)

# Biased electron density
rtmp1 <- fousynth(sdata$a,fdata$Fobs,pha_new,hidx,N=1000)

# Min and max to include all density in the plot
m <- min(rtmp0$rr,rtmp1$rr)
M <- max(rtmp0$rr,rtmp1$rr)

# Comparison
plot(rtmp0$x,rtmp0$rr,type="l",xlab="x",ylab=expression(rho),
     col=2,ylim=c(m,M))
points(rtmp1$x,rtmp1$rr,type="l",lty=2)
```

