XRD Data File Preparation

Warren A. Carlson

2015-06-17

1   Introduction

Some important aspects of working physics require the cleaning, preprocessing and analysis of experimental data. Analysing data can be significantly more difficult if the data set under study is in a format that is difficult or inconvenient to use or understand. This note describes the cleaning and preprocessing of experimental data taken during an X-ray diffraction experiment in the third year student laboratory in the School of Physics at the University of the Witwatersrand.

The data readout system for the X-ray diffration experiment provides raw data in machine readable, binary format. This data is then repackaged in the human readable .UDF file format. While this format is useful when storing large sets of simple data, it is cumbersome to work with when not using software which accepts this file format. An example of the contents of a .UDF file is shown below.

```
SampleIdent ,HBN2 ,/
Title1 ,b4c ,/
Title2 ,b4c ,/
DiffrType ,PW1710 ,/
DiffrNumber ,1 ,/
Anode ,Cu ,/
LabdaAlpha1 , 1.54056 ,/
LabdaAlpha2 , 1.54439 ,/
RatioAlpha21 , 0.50000 ,/
DivergenceSlit ,Fixed ,1 ,/
ReceivingSlit ,0.2 ,/
MonochromatorUsed ,NO   ,/
GeneratorVoltage ,  40 ,/
TubeCurrent ,  20 ,/
FileDateTime ,22-Feb-2012 10:44 ,/
DataAngleRange ,  20.0100,   99.9900 ,/
ScanStepSize ,   0.020 ,/
ScanType ,CONTINUOUS ,/
ScanStepTime ,    1.00 ,/
RawScan
19,       19,       25,       18,       28,       23,       21,       32
23,       13,       22,       30,       19,       32,       29,       21
30,       25,       18,       28,       25,       30,       25,       25
24,       37,       25,       20,       25,       23,       32,       27
30,       19,       28,       27,       19,       25,       18,       25
19,       20,       36,       25,       32,       14,       21,       30
37,       35,       22,       19,       36,       18,       26,       19
29,       27,       30,       30,       36,       23,       31,       26
19,       31,       31,       27,       26,       31,       23,       32
20,       26,       22,       28,       26,       28,       31,       27
20,       18,       28,       17,       26,       28,       30,       32
27,       20,       31,       29,       29,       38,       22,       23
```

The structure of a .UDF file can be understood as follows. Each .UDF file has two components, a metadata component and a raw data component. The meta data component stores information such as the

apparatus settings, dates, times, and so on. The meta-data components is written to the beginning of the file and is stored as a collection of key-value pairs. These key-value pairs are stored, one pair per line, as a character string and a tuple of one or more comma separated values. Each line with a key-value pair has a / character end of line marker. The end of the meta-data component is marked with the `RawScan` keyword. Following this, raw data measurements are stored as a linear sequence of comma separated integer values with null end of line termination characters. This sequence is stored in rows of eight numbers.

The form of the data in a .UDF file can be understood as follows, the `DataAngleRange` meta-data parameter is stored with two numerical values. These values specify lower and upper bound on the angular range scanned by the apparatus. Together, these values specify the angular parameter domain on which data is measured. This is paired with an increment size denoting the discrete step-size used to traverse this domain. This increment size in stored in the `ScanStepSize` meta-data key-value pair. Each value in raw data component which appears after the `RawScan` keyword corresponds to a unique angular value on the parameter domain. In this way the n-th value in the raw data is associated with the n-th increment on the scan domain.

It is useful to repackage the data stored in a .UDF into a data-file comprising two comma separated columns. These columns specify measuremnt (x, y) value pairs. Once the `DataAngleRange` and `ScanStepSize` values are known, then a sequence of x values may be generated by incrementing the minimum value of the `DataAngleRange` parameter by the value stored in `ScanStepSize` up to the maximum value and storing these values as a column of values in a separate text file. The y vales are obtained by reformatting of the raw data, and applying some simple string replacements to the .UDF to produce a single columns of values (without the null end of line and non-printing characters). The x and y values may be stored as (x, y) values pairs by joining these columns as a comma separated set. This procedure can be performed using a variety of software packages.

Below is the contents of the first few lines of the file `raw-y.dat` containing the raw y values of .UDF file above.

```
19
19
25
18
28
23
```

Given this file `raw-y.dat` the following `octave` may be used to generate a new file named `cleaned-x-y.csv`. *Note:* the values specifying the x-range are read from the `DataAngleRange` and `ScanStepSize` meta-data values and are entered as $x_{min} : x_{step} : x_{max}$.

```
x = 20.01:0.02:99.99;
y = csvread("raw-y.dat");
pairs = [x', y];
csvwrite("cleaned-x-y.csv", pairs);
```

The first few lines of the output file `"cleaned-x-y.csv"` appear below.

```
20.01,19
20.03,19
20.05,25
20.07,18
20.09,28
20.11,23
```

The file `cleaned-x-y.csv` may now be used by a variety of utilities, on a variety of computing systems. *Note:* this precedure is easily replicated in other languages, using other software, and in different environments.