Evaluation of ED-XRF spectra and MA-XRF datasets

Handling the very large spectral datasets resulting from MA-XRF scanning

P. Van Espen\textsuperscript{(1)}
V. Osorio\textsuperscript{(1,2)} & S. Legrand\textsuperscript{(1)}

\textsuperscript{(1)}\textit{AXES, University of Antwerpen}
\textsuperscript{(2)}\textit{Brightspec NV}

MA-XRF scanning in Conservation, Art and Archeology

Tutorial day
ICTP, Sept 24, 2017
MA-XRF has revolutionised the way we handle (store and evaluate) energy-dispersive X-ray spectra

revolution ˌrevəˈlōSH(ə)n
noun
1 a forcible overthrow of a government or social order in favour of a new system.
• (the Revolution) the American Revolution.
• (often the Revolution) (in Marxism) the class struggle that is expected to lead to political change and the triumph of communism
• a dramatic and wide-reaching change in the way something works or is organised or in people's ideas about it: *marketing underwent a revolution*.

*X-ray analysis underwent a revolution*
Dealing with MA-XRF data

Outline

Storage and retrieval of spectra

Some ideas
An implementation

Evaluation of spectra

Some ideas
An implementation
Introduction

What do we have.

What do we want to achieve?

in each pixel
a spectrum
(512×512)
in each channel
an image (1024)

Make informative images based on
the acquired spectrum in each pixel

e.g. element or pigment images

Universiteit Antwerpen
Example  Scanning “large” objects

Panel of the Ghent Altarpiece "The Adoration of the Lamb" painted by the brothers Van Eyck around 1432

Ref: Geert Van der Snickt et. al.

Universiteit Antwerpen
The amount of data is **HUGH**

Number of spectra vs scan size

<table>
<thead>
<tr>
<th>Scan size</th>
<th># spectra</th>
<th>Amount of data*</th>
</tr>
</thead>
<tbody>
<tr>
<td>64×64</td>
<td>4 096</td>
<td>16777216 = 16 Mb</td>
</tr>
<tr>
<td>128×128</td>
<td>16 384</td>
<td>67108864 = 64 Mb</td>
</tr>
<tr>
<td>256×256</td>
<td>65 536</td>
<td>268435456 = 256 Mb</td>
</tr>
<tr>
<td>512×512</td>
<td>262 144</td>
<td>1073741824 = 1 Gb</td>
</tr>
<tr>
<td>1024×1024</td>
<td>1 048 576</td>
<td>4294967296 = 4 Gb</td>
</tr>
<tr>
<td>2048×2048</td>
<td>4 194 304</td>
<td>17179869184 = 16 Gb</td>
</tr>
</tbody>
</table>

* assuming 1024 channels and 4 bytes/ch

storage and retrieval problem!!!
The time required to evaluate is **Enormous**

Time needed to analyse the spectra

<table>
<thead>
<tr>
<th>Scan size</th>
<th># spectra</th>
<th>time*</th>
</tr>
</thead>
<tbody>
<tr>
<td>64×64</td>
<td>4 096</td>
<td>137 min = 2.4 h</td>
</tr>
<tr>
<td>128×128</td>
<td>16 384</td>
<td>546 min = 9.1 h</td>
</tr>
<tr>
<td>256×256</td>
<td>65 536</td>
<td>36.4 h = 1.5 d</td>
</tr>
<tr>
<td>512×512</td>
<td>262 144</td>
<td>145 h = 6 d</td>
</tr>
<tr>
<td>1024×1024</td>
<td>1 048 576</td>
<td>582.5 = 24.3 d</td>
</tr>
<tr>
<td>2048×2048</td>
<td>4 194 304</td>
<td>97.1 d = 3.2 months</td>
</tr>
</tbody>
</table>

*assuming 2 s per spectrum

evaluation problem!!!

Universiteit Antwerpen
We have a space-time problem

Unbinding space from time could solve the biggest conundrums in the Cosmos.

Anil Ananthaswamy

conundrum |ˈkɒnəndrəm|
noun (pl. conundrums)
• a confusing and difficult problem or question: one of the most difficult conundrums for the experts.
• a question asked for amusement, typically one with a pun in its answer; a riddle.

Let’s solve the conundrum!!!
Storage and retrieval

some ideas
Storing the spectra

Historical: each spectrum one file

Serious file system performance degradation if more than 1000 files in one directory!!!
Better: one file per line scanned

~1000 spectra per file

Retrieval of arbitrary spectrum (x,y) for inspection not trivial
Content of each channel stored in 4 bytes (values 0 — 4 294 967 295 counts)

One spectrum = 2048 channel = 8192 bytes = 65536 bits

168×8 + 1880×4 = 1344 + 7520 = 8864 bits needed

factor: 65536 / 8864 = 7.4
Lossless spectrum compression

Original: constant length $2048 \times 4$ bytes

Variable length compression:
use $\frac{1}{2}$ byte when possible, 1 byte if needed or $1\frac{1}{2}$, or 2…

<table>
<thead>
<tr>
<th>data size</th>
<th>values</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 nibble</td>
<td>1 — 15</td>
</tr>
<tr>
<td>2 nibbles</td>
<td>16 — 255</td>
</tr>
<tr>
<td>3 nibbles</td>
<td></td>
</tr>
<tr>
<td>4 nibbles</td>
<td>1 — 65231</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

data size values

nibble $= \frac{1}{2}$ byte $= 4$ bits
Storage of compressed spectra in one file

Original data (ascii, ESRF-EDF) → Compressed Spectroscopic Image Data (*.CSID)

special format for efficient storage and retrieval makes **interactive** analysis possible
Directory of 37,511,813,448 bytes (37 Gb) to a file of 759,505,916 bytes (759 Mb) compression of a factor 47.5
Storage and retrieval

implementation
Example: painting on oak panel by Pieter Eyskens

202 × 201 pixels
40602 spectra
Storage and retrieval as implemented in bAxilConvertSID
Evaluation

some ideas
Historical

Evaluation with user interaction of individual spectra
e.g. using computer programs like

- Axil, QXAS
- WinAxil, WinQXAS
- PyMCA
- bAxil
- ...

Universiteit Antwerpen
Purpose relating spectral data to physical properties

e.g. net peak area ⇔ amounts of the element

Different terms and techniques

Spectrum integration
Fitting
deconvolution
...

Linear least squares
Non-linear (iterative) least squares
Alternating least squares
They all have something in common: use of a “profile”

Consider simple peak integration (region of interest, ROI)

![Graph showing peak integration with integration limits](image)

\[ A_{Fe Kα} = \sum_{i=i_{beg}}^{i=i_{end}} y_i \]

integration limits are based on assumed profile
Characteristics of the profile

Shape
Gaussian

Parameters
position \((i_p)\)
width \((s_{ch})\)
height (area, \(A\))

mathematical

in channels
\[
y(i) = \frac{A}{s_{ch}\sqrt{2\pi}} \exp \left[-\frac{(i - i_p)^2}{2s_{ch}^2}\right]
\]

\(s_{ch}\) width (sigma) of peak in channels
\[
s_{ch} = \frac{\text{FWHM}}{2\sqrt{2\ln 2}} = \frac{\text{FWHM}}{2.35}
\]

in keV
\[
y(i) = \frac{A \times \text{Gain}}{s_E\sqrt{2\pi}} \exp \left[-\frac{(E_i - E_p)^2}{2s_E^2}\right]
\]

\(s_E\) width of peak in keV
\(E_i\) energy of channel \(i\), \(E_i = \text{zero} + \text{gain} \times i\)

Universiteit Antwerpen
Profile of an X-ray line (Zn Kα)  
Profile of an element (Pb)

different spectrum evaluation methods = different applications of the profiles
1 “Traditional” non-linear least squares

Iterative process
Adjusting the profiles

<table>
<thead>
<tr>
<th>n</th>
<th>$\chi^2$</th>
<th>Zero eV</th>
<th>$A_{Fe}$</th>
<th>$A_{Zn}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.6</td>
<td>-100.00</td>
<td>469</td>
<td>1602</td>
</tr>
<tr>
<td>n</td>
<td>$\chi^2$</td>
<td>Zero ev</td>
<td>$A_{Fe}$</td>
<td>$A_{Zn}$</td>
</tr>
<tr>
<td>----</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>1</td>
<td>44.6</td>
<td>-100.00</td>
<td>469</td>
<td>1602</td>
</tr>
<tr>
<td>2</td>
<td>44.5</td>
<td>-0.98</td>
<td>1224</td>
<td>4495</td>
</tr>
</tbody>
</table>

---

Universiteit Antwerpen
<table>
<thead>
<tr>
<th>n</th>
<th>$\chi^2$</th>
<th>Zero ev</th>
<th>$A_{Fe}$</th>
<th>$A_{Zn}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.6</td>
<td>-100.00</td>
<td>469</td>
<td>1602</td>
</tr>
<tr>
<td>2</td>
<td>44.5</td>
<td>-0.98</td>
<td>1224</td>
<td>4495</td>
</tr>
<tr>
<td>3</td>
<td>22.6</td>
<td>-62.62</td>
<td>1500</td>
<td>5607</td>
</tr>
</tbody>
</table>

data fit (profile)
<table>
<thead>
<tr>
<th>n</th>
<th>$\chi^2$</th>
<th>Zero ev</th>
<th>$A_{Fe}$</th>
<th>$A_{Zn}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.6</td>
<td>-100.00</td>
<td>469</td>
<td>1602</td>
</tr>
<tr>
<td>2</td>
<td>44.5</td>
<td>-0.98</td>
<td>1224</td>
<td>4495</td>
</tr>
<tr>
<td>3</td>
<td>22.6</td>
<td>-62.62</td>
<td>1500</td>
<td>5607</td>
</tr>
<tr>
<td>4</td>
<td>18.9</td>
<td>71.03</td>
<td>2680</td>
<td>10922</td>
</tr>
<tr>
<td>n</td>
<td>$\chi^2$</td>
<td>Zero</td>
<td>$A_{\text{Fe}}$</td>
<td>$A_{\text{Zn}}$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>44.6</td>
<td>-100.00</td>
<td>469</td>
<td>1602</td>
</tr>
<tr>
<td>2</td>
<td>44.5</td>
<td>-0.98</td>
<td>1224</td>
<td>4495</td>
</tr>
<tr>
<td>3</td>
<td>22.6</td>
<td>-62.62</td>
<td>1500</td>
<td>5607</td>
</tr>
<tr>
<td>4</td>
<td>18.9</td>
<td>71.03</td>
<td>2680</td>
<td>10922</td>
</tr>
<tr>
<td>5</td>
<td>4.97</td>
<td>-3.64</td>
<td>2883</td>
<td>12005</td>
</tr>
<tr>
<td>6</td>
<td>1.22</td>
<td>3.81</td>
<td>2867</td>
<td>12013</td>
</tr>
<tr>
<td>7</td>
<td>0.814</td>
<td>1.28</td>
<td>2917</td>
<td>12119</td>
</tr>
<tr>
<td>8</td>
<td>0.822</td>
<td>1.26</td>
<td>2917</td>
<td>12129</td>
</tr>
<tr>
<td>9</td>
<td>0.822</td>
<td>1.21</td>
<td>2917</td>
<td>12129</td>
</tr>
</tbody>
</table>

Data fit (profile)
Non-linear Least Squares

- Takes profile changes into account (energy and resolution calibration)
- Account for the continuum (background) present
- Allows to model other artefacts (sum peaks)

**BUT**

- Is inherently slow
  mathematical complexity and iterative
  (Marquardt - Leverberg algorithm)
2. Linear Least Squares

position and width of peaks know estimate the area

\[ y(i) = A_{Fe} p_{Fe}(i) + A_{Ni} p_{Ni}(i) \]

\[ = a_{Fe} \]

\[ + a_{Ni} \]

is like “fitting a straight line”
“Simple” matrix algebra

\[ y(i) = A_1 P_1(i) + A_2 P_2(i) + \ldots + A_n P_m(i) \]

\[ y_{n \times 1} = P_{n \times m} A_{m \times 1} \]

General linear model

Least squares solution

Finding the \( A_j \)’s

\[ A = (P^T P)^{-1} \times P^T y \]

Direct solution using matrix multiplication

- Fast

BUT

- Systematic errors if energy or resolution calibration changes
- Cannot take continuum (background) into account
- Cannot deal with non-linear effects (absorption, sum peaks)
## Summary

<table>
<thead>
<tr>
<th>Level of knowledge about the profile</th>
<th>Need to estimate</th>
<th>Method</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>I know: shape</td>
<td>Area, position and width</td>
<td>Non-linear least squares</td>
<td>Slow</td>
</tr>
<tr>
<td>I know: shape, position and width</td>
<td>Area</td>
<td>Linear least squares</td>
<td>Vulnerable to systematic errors</td>
</tr>
<tr>
<td>“I know nothing”</td>
<td>Shape and area</td>
<td>Multivariate statistics</td>
<td>Results need interpretation</td>
</tr>
</tbody>
</table>

position and width = energy and resolution calibration
Combining speed with accuracy

Hybrid least squares fitting HLS

At start

Use non-linear least squares to determine the profiles \( p_j(i) \)
(is energy and resolution calibration)

Calculate the matrix

\[ M = \left( P^T P \right)^{-1} P^T \]

For each spectrum \( y \)

Remove the non-additive continuum (stripping)
Calculate the elemental contributions \( A_j \)

\[ A = M y \]

Repeat non-linear fitting if necessary

 ONE matrix multiplication per spectrum!!!
Evaluation
implementation
Evaluation using bAxil
Conclusions:

- Reduced the data size on disk by a factor of ~ 10
- Reduced the spectrum evaluation time by a factor of ~ 1000
- While maintaining the interactive capabilities of spectrum evaluation
- While obtaining results very comparable with non-linear least squares fitting

A useful method for the processing of large MA-XRF datasets
Thanks to BRIGHTSPEC

http://www.brightspec.be/

Special thanks for your attention