Model independent pre-processing of X-ray powder diffraction profiles

M. Ladisa *

*Istituto di Cristallografia (IC-CNR), Via Amendola 122/O, 70126 Bari, Italy

A. Lamura

*Istituto Applicazioni Calcolo (IAC-CNR), Sezione di Bari, Via Amendola 122/D, 70126 Bari, Italy

T. Laudadio

*Katholieke Universiteit Leuven, Department of Electrical Engineering, Division ESAT-SCD (SISTA), Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium

G. Nico

*Istituto Applicazioni Calcolo (IAC-CNR), Sezione di Bari, Via Amendola 122/D, 70126 Bari, Italy

Abstract

Precise knowledge of X-ray diffraction profile shape is crucial in the investigation of the properties of matter in crystals powder. Line-broadening analysis is a pre-processing step in most of the full powder pattern fitting softwares. Final result of line-broadening analysis strongly depends on preliminary three steps: Noise filtering, removal of background signal and peak fitting. In this work a new model independent procedure for two of the aforementioned steps (background suppression and peak fitting) is presented. The former is dealt with by using morphological mathematics, while the latter relies on the Hankel Lanczos Singular Value Decomposition technique. Real X-ray powder diffraction (XRPD) intensity profiles of Ceria samples are used to test the performance of the proposed procedure. Results show the robustness of this approach and its capability of efficiently improving the disentangling of instrumental broadening. These features make the proposed approach an interesting and user-friendly tool for the pre-processing of XRPD data.

Key words: Hankel Lanczos Singular Value Decomposition (HLSVD), Morphological filtering, X-ray powder diffraction
1 Introduction

X-ray powder diffraction (XRPD) technique is nowadays a well known tool to study crystalline properties, which provide important information for applications in fields such as nanotechnology [1,2]. All the applications benefit from a reliable pre-processing aimed at enhancing the quality of XRPD data. The pre-processing procedure consists of four steps: Denoising, background suppression, peak fitting and signal deblurring, also known as line-broadening (see Fig. 1). Many techniques can be applied to remove noise from XRPD data (see ref. [3] and references therein). The step of background suppression is needed to emphasize the peak features of the sample. Traditional techniques are: Young’s polynomials, Chebyshev approximation and linear interpolation [4]. Peak fitting is a very challenging step to extract information about the properties of polycrystalline powder [5]. The final stage in XRPD data pre-processing is the signal deblurring aiming at disentangling the instrumental line-broadening out of data. On this topic we spot a recent review in ref. [6]. In this paper we propose a new approach for two of the above pre-processing steps of XRPD data. The signal background is removed by means of a procedure based on morphological mathematics. The peak shape profile fitting is carried out by using a subspace-based parameter estimation method called Hankel Lanczos Singular Value Decomposition (HLSVD) technique [7]. The main advantage of this approach is twofold: While both background suppression and peak shape profile fitting are model independent, the signal deblurring procedure, employed thereafter, crucially benefits from the model independence of the previous two steps. Indeed, our signal deblurring steps over the major drawback of most of the XRPD analysis tools available in the literature: The peak overlapping problem, namely the difficulty in singling the peak out of the full XRPD profile. This problem is mainly due to the model-induced bias in background/peak profile reconstruction. As to the denoising we use a wavelet based filter, a popular tool available in several software packages. Real XRPD intensity profiles of Ceria samples [6] are used to test the performances. Four different raw datasets were used in pairs [8]. For each pair, one dataset was collected on the annealed Ceria specimen (representing the instrumental broadening) and the other was collected on the broadened sample. The selected pairs are those measured at the University of Birmingham (a high resolution X-ray laboratory) and at the National Synchrotron Light Source (NSLS X3B1).

* Corresponding Author: Massimo Ladisa (m.ladisa@ic.cnr.it)

Email addresses: m.ladisa@ic.cnr.it (M. Ladisa), a.lamura@ba.iac.cnr.it (A. Lamura), Laudadio@esat.kuleuven.ac.be (T. Laudadio), g.nico@ba.iac.cnr.it (G. Nico).

URLs: www.ic.cnr.it (M. Ladisa), www.iac.cnr.it (A. Lamura), www.kuleuven.ac.be (T. Laudadio), www.iac.cnr.it (G. Nico).
The structure of the paper is as follows. Section 2 is devoted to the description of the proposed method and of the results. Conclusions are drawn in Section 3.

2 The method

Two different raw datasets were downloaded for the Ceria sample: The instrumental standard representing the instrumental broadening and the Ceria XRPD pattern of the broadened sample.

Our strategy in disentangling the profile broadening out of the experimental sample relies on a three step procedure which is sketched in the sequel.

2.1 Denoising

The noise was removed by applying wavelet transforms to the full XRPD spectrum and subtracted prior to the background suppression.

Among the many applications of wavelets, signal denoising has been deeply investigated and the wavelet filter can be considered as the state of art on this subject. The discrete wavelet transform (DWT) is a linear operator which modifies the data vector in a similar way as the discrete Fourier transform (DFT). In both cases, the transform, given by a $N \times N$ matrix acting on the input $N$-vector data, is invertible [9]. The matrix entries are combinations of basis functions (the familiar sines and cosines in the case of DFT). Tables 1-2 summarize the main properties of some wavelet bases. An interesting property of wavelet basis is the localization in both space and frequency domain. This means that they have a finite support or a decay in both domains. Regularity is another important property of wavelet basis. Regularity $r$ means that the $r^{th}$ derivative exists almost everywhere (see Table 1). In the case of Daubechies wavelet, regularity depends on the order $N$ (see Table 2). A wavelet is defined by particular set of numbers $\{c_k\}_{k=0, \ldots , 2N-1}$, called wavelet filter coefficients. They are determined by imposing the constraints of $N$ vanishing moments, $\sum_{k=0}^{2N-1} (-1)^k k^m c_k = 0$ ($m = 0, \ldots , 2N - 1$), see Table 1, and orthogonality $\sum_{k=0}^{2N-1} c_k c_{k + 2m} = 2 \delta_{0,m}$. The coefficients $c_k$ characterize a low-pass filter while the coefficients $b_k = (-1)^k c_{2N-1-k}$ results in a high-pass filter. Coefficients $c_k$ give the entries of a $N \times N$ matrix, which iteratively applies to the $N$-data vector thus resulting in the $N$-vector of detail coefficients. The whole procedure described above is the wavelet transform (see [9].
for further details). Generally speaking, the denoising procedure involves three steps. The basic version of the procedure is the following:

- Calculate the wavelet transform of XRPD profile and sort the components of the output vector by increasing frequency. This shall result in $N$-vector containing the XRPD profile average coefficient and a set of detail coefficients.
- Noise thresholding, calculated on the highest frequency detail coefficient of the wavelet spectrum.
- Signal reconstruction by using the average coefficient and thresholded detail coefficients.

In this paper we choose the Daubechies wavelet basis with $N = 2$. We address the reader to Daubechies [10] for further details.

### 2.2 Background suppression

The background was determined by means of morphological transforms for the full XRPD spectrum and subtracted prior to the signal deblurring.

The morphological mathematics is based on the language of set theory. Considered a discrete binary image $I \in \mathbb{R}^2$ and a structuring element $S \in \mathbb{R}^2$, the four basic morphological mathematical operations on $I$ by $S$ are:

\[
\begin{align*}
\text{Dilation} & : & I \oplus S &= \bigcup_{s \in S} I_s \\
\text{Erosion} & : & I \ominus S &= \bigcap_{s \in S} I_{-s} \\
\text{Opening} & : & I \circ S &= (I \ominus S) \oplus S \\
\text{Closing} & : & I \bullet S &= (I \oplus S) \ominus S
\end{align*}
\]

(1)

where $I_s$ denotes the translation of $I$ by $s$, namely $I_s = \{x + s | x \in I\}$. The value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbours, whose number and location is given by the structuring element. Generally, dilation expands image objects, whereas erosion shrinks them. In practice, dilation and erosion are employed in pairs. Opening is the erosion of an image followed by the dilation of the eroded image, and closing is the dilation of an image followed by the erosion of the dilated image. Opening eliminates sharp peaks smaller then the structuring element while the closing fills in the small holes and gaps. The binary morphological operations of dilation, erosion, opening and closing can be extended to grey-scale images. Let $E_I$ and $E_S$ be the domains of the gray-scale image $I$ and the grey-scale structuring element $S$ respectively. The
grey-scale dilation and erosion can be computed by

\[
\text{Dilation: } (\mathcal{I} \oplus S)(x, y) = \max \{ \mathcal{I}(x - m, y - n) + S(m, n) \}
\]

\[
\text{Erosion: } (\mathcal{I} \ominus S)(x, y) = \min \{ \mathcal{I}(x - m, y - n) - S(m, n) \}
\]

(2)

where \((x - m, y - n) \in \mathcal{E}_I\) and \((m, n) \in \mathcal{E}_S\). For such images, the minimum and maximum values are computed within neighbourhood represented by the structuring element (see [11] for details).

In our background suppression procedure the XRPD pattern is reshaped and padded into a 2-D image. A disk with a radius of three pixels is used as structuring element both for erosion and for dilation. As to the erosion (dilation), pixels beyond the image border are assigned the maximum (minimum) value afforded by the data type. The morphological opening removes small objects from the image while preserving the shape and size of larger objects in the image. The overall result is a peak smearing effect while the background intensity remains unaltered. Restoring the original 1-D pattern provides the XRPD spectrum background. Figure 2 summarizes the whole procedure. We compared our findings to the traditional interpolation method and we found a satisfactory agreement (the percentage difference between the background computed by traditional techniques and our finding is below 3 %). Up to our knowledge, this technique has never been applied to XRPD spectrum background suppression and it provides a reliable and user independent estimate of it.

2.3 Peak fitting

The main problem in analysing an XRPD spectrum is the peak search, since the exact \(\vartheta\) position is crucial in the extraction of the relevant microstructural information. Had the peak well defined, its shape would be straightforwardly achieved (for instance by a high resolution interpolation/fit by means of a model). Unfortunately the data resolution is rarely high enough to reach the goal of a well profiled peak. In that respect several methods have been devised so far to accomplish the peak fitting by means of gaussian, lorentzian, voigt, pseudo-voigt, Pearson VII and other models [5]. The main drawback of the aforementioned methods cited above is the dependence of results on the model used in the fit procedure itself and the poor description of the real peak profile shape (for instance asymmetry). Here we use HLSVD method to the purpose [7]. The main advantage of this method is the flexibility since the number of parameters is not fixed and it can be chosen to achieve a more satisfactory agreement between the model and the real peak profile shape. The HLSVD method works as follows. Let us model the XRPD intensity samples \(I_n\) col-
lected at angles $\vartheta_n$, $n = 0, \ldots, N - 1$ as the sum of $K$ exponentially damped complex sinusoids

$$I_n \simeq \sum_{k=1}^{K} a_k \exp(-d_k \vartheta_n) \cos[2\pi f_k \vartheta_n + \varphi_k], \quad (3)$$

$a_k$ is the amplitude, $\varphi_k$ the phase, $d_k$ the damping factor and $f_k$ the frequency of the $k$th sinusoid, $k = 1, \ldots, K$, with $K$ the number of damped sinusoids. The $N$ data points defined in (3) are arranged into a Hankel matrix $H \overset{def}{=} H_{m+l}^{m} = I_{m+l}$, $m = 0, \ldots, M - 1$, $l = 0, \ldots, L - 1$, with $L + M = N + 1$ ($M \simeq L \simeq N/2$). The SVD of the Hankel matrix is computed as $H_{L \times M} = U_{L \times L} \Sigma_{L \times M} V_{M \times M}^H$, where $\Sigma = \text{diag} \{ \lambda_1, \lambda_2, \ldots, \lambda_r \}$, $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r$, $r = \min(L, M)$, $U$ and $V$ are orthogonal matrices and the superscript $H$ denotes the Hermitian conjugate. The Lanczos bidiagonalization algorithm with partial reorthogonalization is used to compute SVD. This algorithm, based on FFT, computes the two matrix-vector products which are performed at each step of the Lanczos procedure in $O((L + M) \log_2(L + M))$ rather than in $O(LM)$. In order to obtain the “signal” subspace, the matrix $H$ is truncated to a matrix $H_K$ of rank $K$, $H_K = U_K \Sigma_K V_K^H$, where $U_K$, $V_K$, and $\Sigma_K$ are defined by taking the first $K$ columns of $U$ and $V$, and the $K \times K$ upper-left matrix of $\Sigma$, respectively. As subsequent step, the least-squares solution of the following over-determined set of equations is computed $V_K^{(\text{top})} E^H \simeq V_K^{(\text{bottom})}$, where $V_K^{(\text{bottom})}$ and $V_K^{(\text{top})}$ are derived from $V_K$ by deleting its first and last row, respectively. The $K$ eigenvalues $\hat{\beta}_k$ of matrix $E$ are used to estimate the frequencies $\hat{f}_k$ and damping factors $\hat{d}_k$ of the model damped sinusoids from the relationship

$$\hat{\beta}_k = \exp \left[ (-\hat{d}_k + i2\pi \hat{f}_k) \Delta \vartheta \right], \quad (4)$$

with $k = 1, \ldots, K$. Values so obtained are inserted into the model equation (3) which yields the set of equations

$$I_n \simeq \sum_{k=1}^{K} a_k \exp(-\hat{d}_k \vartheta_n) \cos[2\pi \hat{f}_k \vartheta_n + \hat{\varphi}_k], \quad (5)$$

with $n = 0, \ldots, N - 1$. The least-squares solution of (5) provides the amplitude $\hat{a}_k$ and phase $\hat{\varphi}_k$ estimates of the model sinusoids which are used in the next step.
2.4 Deblurring

The XRPD pattern has to be corrected for the instrumental broadening. Several methods have been devised so far to deal with this problem. Among them we quote the Stokes method [12] and the Bayesian approach [13,14]. The main drawbacks of these methods stem from the difficulty in evaluating the background level, mainly due to peak overlapping.

The XRPD pattern plugged in the deblurring algorithm is noise-background free since it has been already pre-processed by the wavelets filter and the morphological operator.

The technique proposed in this paper is a modified version of the one presented in ref. [6]. A blurred or degraded XRPD pattern can be approximately described by a Volterra equation \( g = H \otimes f + n \), where \( g \) is the blurred XRPD pattern and \( H \) is the distortion operator due to several causes, namely the point spread function (PSF), and \( n \) is an additive noise, introduced in the XRPD acquisition, that corrupts the signal. Strictly speaking in an XRPD experimental setup we deal with a poissonian noise which is a multiplicative noise. However the Poisson distribution function resembles the Gauss one provided a sufficiently large statistics in photons counting.

As to the deblurring procedure we implement the damped Lucy–Richardson algorithm. This function performs multiple iterations, using optimization techniques and Poisson statistics. In our approach the PSF is the raw dataset downloaded for the Ceria sample - the instrumental standard - resembling the instrument profile [6]. The algorithm maximizes the likelihood that the resulting image, when convolved with the PSF, is an instance of the blurred image, assuming Poisson noise statistics. This function can be effective when the PSF is known but the knowledge about the additive noise in the image is poor. The Lucy–Richardson algorithm introduces several adaptations to the original maximum likelihood algorithm that addresses complex image restoration tasks. By using these adaptations, the effect of noise amplification on image restoration can be reduced, nonuniform image quality can be accounted for (e.g., bad pixels, flat-field variation) and the restored image resolution can be improved by subsampling.

Due to the denoising/background suppression, the original Volterra equation is readily simplified: \( g = H \otimes f \), where \( g, H \) and \( f \) have been already defined. Was the inverse \( H^{-1} \) explicitly known, we would solve the former equation at a glance. Unfortunately this is not the case and the solution has to be approximated as follows. \( g, H \) and \( f \) are positive and the PSF cannot change
the $g$ norm, i.e. $\|g\| = \|f\|$. Thus:

$$\sum_i f_i = \sum_i g_i \iff \sum_i f_i = \sum_i g_i \frac{(H \otimes f)_i}{(H \otimes f)_i} = \sum_i f_i (\hat{g} \otimes H)^i,$$

where $\hat{g}_k = \frac{g_k}{(H \otimes f)_k}$; the solution can be found by an iterative procedure:

$$f^{n+1} = f^n \frac{g \otimes H}{H \otimes f^n},$$

where the initial guess XRPD spectrum is uniform. As already stressed in ref. [6], the main drawbacks in applying such an algorithm to the single peak deconvolution are the noise amplification and the peak fitting bias. Noise amplification is dramatically reduced by both the denoising procedure and the small (some five) number of iterations used in the algorithm. Moreover the damp in the algorithm specifies the threshold level for the deviation of the resulting image from the original image, below which damping occurs. For pixels that deviate in the vicinity of their original values, iterations are suppressed.

As to the peak fitting bias, unlike the Balzar approach, our procedure uses the instrumental standard pattern (with no overlapping) as the PSF to deconvolve the full XRPD pattern and then we extract the deconvoluted/deblurred XRPD pattern in the same range of the PSF used for the deconvolution itself. The rationale of this choice relies on the fact that while the PSF peaks have no overlap, this is not the case for the broadened sample peaks and, thus, the peak ranges can be defined starting on the annealed sample rather than the broadened one. Moreover the discrete Fourier transform (DFT), used by the deblurring functions, assumes that the frequency pattern of an image is periodic. This assumption creates a high-frequency drop-off at the edges of an overlapping peaks cluster [15]. This high-frequency drop-off can create an effect called boundary related ringing in deblurred images, that is a systematic error affecting any further investigation on the physical meaning of the deconvolved spectrum. To reduce ringing, our full pattern deconvolution, as described above, resembles an edgetaper function. It removes the high-frequency drop-off at the edge of an image by blurring the entire image and then replacing the center pixels of the blurred image with the original one. In this way, the edges of the image taper off to a lower frequency.

3 Conclusions

In this paper we presented a new approach for background removal and peak fitting of XRPD profiles. Such operations are crucial in the line-broadening analysis of X-ray diffraction profiles, an important pre-processing step in the
investigation of the crystal powder samples by means of XRPD data. Background suppression relies on the use of morphological mathematics while peak fitting is carried out by means of HLSVD technique. In order to enhance the signal-to-noise ratio of XRPD profiles a wavelet based filter is preliminarily applied to XRPD data. The output of the proposed procedure is supplied to a damped Lucy–Richardson algorithm for deblurring. The main advantage of this approach is twofold: Background suppression and peak shape profile fitting are model independent. Real XRPD intensity profiles of Ceria samples are used to test performances. Results show that the background estimate is in agreement with that computed by traditional interpolation methods with a percentage difference below 3%. Further, the output XRPD profile has narrower peaks, located at the same position and with the same shape as the original one. It is worth noting that once the deblurred, noise-background free XRPD spectrum is convoluted back to the PSF and added to the removed noise and background signals, the resulting XRPD pattern resembles the original one.
References

[1] Cervellino, A., Giannini, C., Guagliardi, A. & Ladisa, M. (2005), Nanoparticle size distribution estimation by full-pattern diffraction analysis, submitted to Physical Review B; arXiv:cond-mat/0502583.

[2] Cervellino, A., Giannini, C., Guagliardi, A. & Ladisa, M. (2005), Disentangling instrumental broadening, submitted to Journal of Applied Crystallography; arXiv:cond-mat/0502582.

[3] Ladisa, M., Lamura, A., Laudadio, T., and Nico, G. (2005), Application of the HLSVD technique to the filtering of X-ray diffraction data, submitted to Signal Processing; arXiv:math.SP/0502539.

[4] Jenkis, R., and Snyder, R.L. (1996), Introduction to X-ray Powder Diffractometry, New York: Wiley.

[5] Estevez–Rams, E., Penton, A., Martinez–Garcia, J & Fuess, H. (2005), The use of analytical peak profile functions to fit diffraction data of planar faulted layer crystals, Cryst. Res. Technol., 40, 1/2, 166-176.

[6] Balzar, D., Audebrand, N., Daymond, M.R., Fitch, A., Hewat, A., Langford, J.I., Le Bail, A., Louer, D., Masson, O., McCowan, C.N., Popa, N.C., Stephens, P.W. & Toby, B.H. (2004), Journal of Applied Crystallography, 37, 911-924.

[7] Laudadio, T., Mastronardi, N., Vanhamme, L., Van Hecke, P. and Van Huffel, S. (2002), Improved Lanczos Algorithms for Blackbox MRS Data Quantitation, Journal of Magnetic Resonance, 157, 292-297.

[8] Data are available at the URLs: www.du.edu/%7ebalzar/s-srr.htm, www.ccp14.ac.uk, www.boulder.nist.gov/div853/balzar.

[9] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992), Numerical recipes in C, Cambridge University Press.

[10] Daubechies, I. (1992), Ten lectures on wavelets, SIAM.

[11] Serra, J. (1994) Morphological filtering: an overview, Signal Processing, 38, 3-11.

[12] Stokes, A.R. (1948), Proc. Phys. Soc. (London), 61, 382-391.

[13] Lucy, L.B. (1972) An iterative technique for the rectification of observed distributions, The Astronomical Journal, 8, 243-246.

[14] Richardson, W.H. (1972), Journal of Optical Society of America A, 62, 55-59.

[15] Biemond, J., Lagendijk, R.L., Mersereau, R.M. (1990), Iterative methods for image deblurring, Proceedings of the IEEE, 78, 5, 856-883.
Table 1  
Properties of some wavelet bases.

| wavelet          | t-localization | f-localization | # zero moments | r  |
|------------------|----------------|----------------|----------------|----|
| Haar             | [0,1]          | 1/f            | 1              | 0  |
| Sinc             | 1/t            | [0,1]          | ∞              | ∞  |
| Daubechies (N)   | [0,2N-1]       | 1/f            | N              | α(N) |

Table 2  
Regularity $r = \alpha(N)$ of a Daubechies wavelet basis of order $N$.

| N    | 2   | 3   | 4   | 5   | 6   | N   |
|------|-----|-----|-----|-----|-----|-----|
| α(N) | 0.500 | 0.915 | 1.275 | 1.596 | 1.888 | 0.2075 N |
Fig. 1. Block diagram of the pre-processing procedure. The proposed approach refers to the background suppression and peak fitting steps (shadowed boxes).
Fig. 2. Background evaluation procedure. Left top: original XRPD pattern. Right top: XRPD pattern after two-dimensional reshaping. Right bottom: two-dimensional reshaped XRPD pattern after morphological operations. Left bottom: background after one-dimensional reshaping.
Fig. 3. Plots. Left top: original XRPD pattern. Right top: XRPD pattern after denoising and background suppression. Left bottom: noise-background free XRPD pattern after deblurring. Right bottom: residue between the final XRPD pattern re-convoluted to the PSF together with noise and background and the original XRPD pattern.