

Research Article

Application of the HLSVD Technique to the Filtering of X-Ray Diffraction Data

M. Ladisa,¹ A. Lamura,² T. Laudadio,³ and G. Nico²

¹Istituto di Cristallografia (IC), Consiglio Nazionale delle Ricerche (CNR), Via Amendola 122/O, 70126 Bari, Italy

²Istituto Applicazioni del Calcolo Mauro Picone (IAC), Consiglio Nazionale delle Ricerche (CNR), Via Amendola 122/D, 70126 Bari, Italy

³SISTA, SCD Division, Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium

Received 6 February 2006; Revised 21 December 2006; Accepted 2 February 2007

Recommended by Jacques G. Verly

A filter based on the Hankel-Lanczos singular value decomposition (HLSVD) technique is presented and applied for the first time to X-ray diffraction (XRD) data. Synthetic and real powder XRD intensity profiles of nanocrystals are used to study the filter performances with different noise levels. Results show the robustness of the HLSVD filter and its capability to extract easily and efficiently the useful crystallographic information. These characteristics make the filter an interesting and user-friendly tool for processing of XRD data.

Copyright © 2007 M. Ladisa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

In many applications of X-ray diffraction (XRD) techniques to the study of crystal properties, a key step in the data processing chain is an effective and adaptive noise filtering [1–4]. A correct noise removal can facilitate the separation of the useful crystallographic information from the background signal, and the estimation of crystal structure and domain size. Important issues of XRD data filtering are performances in noise suppression, capability to preserve the peak position, computational cost, and finally, the possibility of being used as a blackbox tool. Different digital filters have been applied to XRD data, in spatial and frequency domains. Simple procedures are based on polynomial filtering (and fitting) in the spatial domain [1]. A standard practice when working in frequency domain is to use Fourier smoothing. It consists in removing the high-frequency components of the spectrum [5]. Since the truncation of high-frequency components can be problematic in the case of high-level noise, a different approach based on the Wiener-Fourier (WF) filter has been proposed to clean XRD data [6]. A different approach, which makes use of the singular value decomposition (SVD), has been successfully applied to time-resolved XRD data to reduce noise level [3, 4].

In this work, we describe an application of the Hankel-Lanczos singular value decomposition (HLSVD) algorithm

to filter XRD intensity data. The proposed filter is based on a subspace-based parameter estimation method, called Hankel singular value decomposition (HSVD) [7], which is currently applied to nuclear magnetic resonance spectroscopy data for solvent suppression [8]. The HSVD method computes a “signal” subspace and a “noise” subspace by means of the SVD of the Hankel matrix H , whose entries are the noisy signal data points. Its computationally most intensive part consists of the computation of the SVD of the matrix H . Recently, several improved versions of the algorithm have been developed in order to reduce the needed computational time [8]. In this paper, we choose to apply the HSVD method based on the Lanczos algorithm with partial reorthogonalization (HLSVD-PRO), which is proved to be the most accurate and efficient version available in the literature. A comparison in terms of numerical reliability and computational efficiency of HSVD with its Lanczos-based variants can be found in [8].

A criterion is presented to facilitate the separation of noise from the useful crystallographic signal. It is completely user-independent since it is based on a numerical method. It will be described in more detail in Section 4. It enables the design of a blackbox filter to be used in the processing of XRD data. Here, the filter is applied to nanocrystalline XRD data. Nanocrystals are characterized by chemical and physical properties different from those of the bulk [9].

At a scale of a few nanometers, metals can crystallize in a structure different from that of bulk. Nowadays, different branches of science and engineering are benefiting from the properties of nanocrystalline materials [10]. In particular, recent XRD experiments have shown that intensities, measured as a function of the scattering angle, could be useful to extract structural and domain size information about nanocrystalline materials. Experimental XRD data were collected on the XRD beam line at the Brazilian Synchrotron Light Facility (LNLS-Campinas, Brazil) using 8.040 keV photons at room temperature (for further details, see [11]). In this case, samples with different diameters of powder of gold nanoparticles underwent diffraction measurements in the X-ray domain. The diffracted intensity was recorded varying the diffraction angle, namely, the angle between the incident beam and the scattered one. Synthetic XRD datasets are generated by computing the X-ray scattered intensity from nanocrystalline samples of different sizes and properties by using an analytic expression (see (6)). Synthetic datasets are processed and filter performance is studied when considering different levels of noise. Numerical tests on real XRD data of Au nanocrystalline samples of different sizes and properties show the robustness of the proposed filter and its capability to extract easily and efficiently the useful crystallographic information. These characteristics make this filter an interesting and user-friendly tool for the interactive processing of XRD data.

The paper has the following structure. Section 2 is devoted to the theoretical aspects of the proposed approach. The dataset used to study the filter properties is described in Section 3. Numerical results are reported in Section 4. Finally, some conclusions are drawn in Section 5.

2. THE SUBSPACE-BASED PARAMETER ESTIMATION METHOD HSVD

Let us denote with I_n the samples of the diffracted intensity signal collected at angles ϑ_n , $n = 0, \dots, N - 1$. They are modelled as the sum of K exponentially damped complex sinusoids

$$I_n = I_n^0 + e_n = \sum_{k=1}^K a_k \exp(i\varphi_k) \exp[(-d_k + i2\pi f_k)\vartheta_n] + e_n, \quad (1)$$

where I_n and I_n^0 , respectively, represent the measured and modelled intensities at the n th scattering angle $\vartheta_n = n\Delta\vartheta + \vartheta_0$, with $\Delta\vartheta$ the sampling angular interval and ϑ_0 the initial scattering angular position, a_k is the amplitude, φ_k the phase, d_k the damping factor, and f_k the frequency of the k th sinusoid, $k = 1, \dots, K$, with K the number of damped sinusoids, and e_n is complex white noise with a circular Gaussian distribution. It is worth noting that the value of K increases or decreases by 2 in order to guarantee that the modelled intensity is real. This constraint is enforced in the filtering process. The algorithm is described in detail in Appendix A.

It allows to estimate the parameters \hat{d}_k and \hat{f}_k appearing in (1). These are inserted into the model (1), which yields the

set of equations

$$I_n \simeq \sum_{k=1}^K a_k \exp(i\varphi_k) \exp[(-\hat{d}_k + i2\pi\hat{f}_k)\vartheta_n] + e_n, \quad (2)$$

with $n = 0, 1, \dots, N - 1$. The least-squares solution $\hat{c}_k = \hat{a}_k \exp(i\hat{\varphi}_k)$ of (2) provides the amplitude \hat{a}_k and phase $\hat{\varphi}_k$ estimates of the model sinusoids. The computationally most intensive part of this method is the computation of the SVD of the Hankel matrix H . Various algorithms are available for computing the SVD of a matrix. The most reliable algorithm for dense matrices is due to Golub and Reinsch [12] and is available in LAPACK [13]. The Golub-Reinsch method computes the full SVD in a reliable way and takes approximately $2LM^2 + 4M^3$ complex multiplications for an $L \times M$ matrix. However, when only the computation of a few largest singular values and corresponding singular vectors is needed, the method is computationally too expensive. More suitable algorithms exist which are based on the Lanczos method. The proposed approach relies on the latter.

A key step in the filtering procedure is the selection of the number K of damped sinusoids characterizing the model function of the HLSVD-PRO filter. Here, a possible approach is presented, which is based on the following frequency selection criterion: the singular values λ_k , $k = 1, \dots, r$, are plotted versus the corresponding frequencies f_k of the sinusoids in (1). This choice facilitates a direct comparison of the results of the proposed filter with those obtained by other filters based on a frequency approach. It was observed that, generally, crystallographic XRD intensity signals show a clear transition from a low-frequency region, characterized by high singular values λ_k , to a high-frequency region with small singular values whose variability is below 10% with respect to the asymptotic value λ_r , namely $(\lambda_K - \lambda_r)/\lambda_K < 0.1$. We exploit this feature in order to automatically find a threshold. The index K of the frequency f_K corresponding to the transition provides the number of damped sinusoids to be used in the HLSVD-PRO filter. The filter performance was evaluated using the measure

$$\mathcal{E} = \frac{\|I_{\text{exp}} - I_{\text{th}}\|}{\|I_{\text{fil}} - I_{\text{th}}\|}, \quad (3)$$

where $I_{\text{exp/th/fil}}$ are the experimental/theoretical/filtered intensities, respectively.

3. DATASET

The HLSVD-PRO filter was applied to synthetic as well as real XRD data. In this section, the generation of XRD intensity profiles and the experimental setup for the acquisition of real data are described. Both synthetic and real XRD data refer to Au nanocrystalline samples. Nanocrystals are made of clusters of three different structure types: cuboctahedral \mathcal{C} , icosahedral \mathcal{I} , and decahedral \mathcal{D} . For each fixed structure type \mathcal{X} ($\mathcal{X} = \mathcal{C}, \mathcal{I}, \mathcal{D}$), the size n of clusters follows a log-normal distribution

$$f_{\mathcal{X}}(n) = \frac{\exp(-s_{\mathcal{X}}/2)}{\sqrt{2\pi\xi_{\mathcal{X}}s_{\mathcal{X}}}} \exp\left[-\frac{(\log n - \log \xi_{\mathcal{X}})^2}{2s_{\mathcal{X}}^2}\right], \quad (4)$$

with mode $\xi_{\mathcal{X}}$ and logarithmic width $s_{\mathcal{X}}$. Structural distances for the different structure types \mathcal{X} are generally studied independently of the actual nanomaterial. The nearest distance between atoms in the crystals is chosen as a reference length and is arbitrarily set to $1/\sqrt{2}$, a constant in various structures \mathcal{X} and for all sizes n of the clusters. Actual distances in nanoclusters are then recovered by applying a correction factor $a_{\mathcal{X}}(n)$ for strain, supposed to be uniform and isotropic. A convenient description of the strain factor as a function of the structure type and cluster size is

$$a_{\mathcal{X}}(n) = \Omega_{\mathcal{X}} + (\Xi_{\mathcal{X}} - \Omega_{\mathcal{X}}) \times \frac{\pi + 2 \operatorname{atan}((n_{\mathcal{X}}^0 - n)/w_{\mathcal{X}})}{\pi + 2 \operatorname{atan}((n_{\mathcal{X}}^0 - 1)/w_{\mathcal{X}})}, \quad (5)$$

given in terms of the four parameters $[n_{\mathcal{X}}^0, \Omega_{\mathcal{X}}, \Xi_{\mathcal{X}}, w_{\mathcal{X}}]$. Intensities scattered by nanoclusters with size n and structure type \mathcal{X} are computed by using the diffractive model based on the Debye function method [14, 15]:

$$I_{\mathcal{X},n}(q) = A \left\{ N_{\mathcal{X}}(n) + \sum_{i \neq j}^{N_{\mathcal{X}}(n)} \frac{\sin[2\pi q u_{i,j}^{\mathcal{X},n} a_{\mathcal{X}}(n)]}{2\pi q u_{i,j}^{\mathcal{X},n} a_{\mathcal{X}}(n)} \right\}, \quad (6)$$

where I_0 is the incident X-ray intensity, $T(q')$ is the Debye-Waller factor, $f(q)$ is the atomic form factor, $A = I_0 [T(q') f(q)]^2$, $q = 2a_{\text{f.c.c.}} \sin \vartheta / \lambda$ and $q' = q/a_{\text{f.c.c.}}$ are, respectively, the dimensionless and the usual scattering vector lengths with $a_{\text{f.c.c.}}$ being the face-centered cubic (f.c.c.) bulk lattice constant; $N_{\mathcal{X}}(n)$ is the number of atoms in the cluster, $u_{i,j}^{\mathcal{X},n}$ the distance between the i th and j th atom, $a_{\mathcal{X}}(n)$ the strain factor. The total scattered intensity is computed as

$$I(q) = \sum_{\mathcal{X}} x_{\mathcal{X}} \sum_{n=1}^{S_{\mathcal{X}}} f_{\mathcal{X}}(n) I_{\mathcal{X},n}(q), \quad (7)$$

where $S_{\mathcal{X}}$ denotes the size of the largest cluster of type \mathcal{X} , $x_{\mathcal{X}}$ is the number fraction of each structure type ($\sum_{\mathcal{X}} x_{\mathcal{X}} = 1$), and $f_{\mathcal{X}}(n)$ is the value of log-normal size distribution (4). It is worth noting that both intensities in (6) and (7) are actually functions of the scattering angle ϑ being $q = 2a_{\text{f.c.c.}} \lambda^{-1} \sin \vartheta$. Experimental XRD intensity profiles are collected by counting, at each scattering angle ϑ_n , the number of scattered photons giving the diffracted intensity signal I_n . For such events, data are affected by Poisson noise. Since the number of photons scattered at each angle ϑ_n is large, the Poisson-distributed noise can be approximated by a Gaussian-distributed noise as required in Section 2 [16].

Noisy synthetic XRD intensity profiles were built by generating Poisson-distributed random profiles with intensity I (7) taken as the mean value of the Poisson process. As a measure of the noise level, the noise-to-signal ratio (NSR) was defined as follows:

$$\text{NSR} = \frac{\|\sqrt{\mathcal{P}(F \times I)}\|}{\|\mathcal{P}(F \times I)\|}, \quad (8)$$

where $\mathcal{P}(I)$ denotes a Poisson process with mean value I . Different NSR values were obtained by scaling the scattered intensity (7) by a factor F . Figure 1 displays XRD intensity

TABLE 1: Values of parameters used in (6) to compute synthetic XRD intensity profiles. The wavelength and the f.c.c. bulk lattice constant were set to $\lambda = 0.15418$ nm and $a_{\text{f.c.c.}} = 0.40786$ nm, respectively.

Parameter	$\mathcal{X} = \mathcal{C}$	$\mathcal{X} = \mathcal{I}$	$\mathcal{X} = \mathcal{D}$
$\xi_{\mathcal{X}}$	5.0	5.0	5.0
$s_{\mathcal{X}}$	0.3	0.3	0.3
$n_{\mathcal{X}}^0$	4.0	4.0	6.0
$\Omega_{\mathcal{X}}$	1.0	1.0	1.0
$\Xi_{\mathcal{X}}$	1.0	1.0	1.0
$w_{\mathcal{X}}$	0.5	0.5	0.5

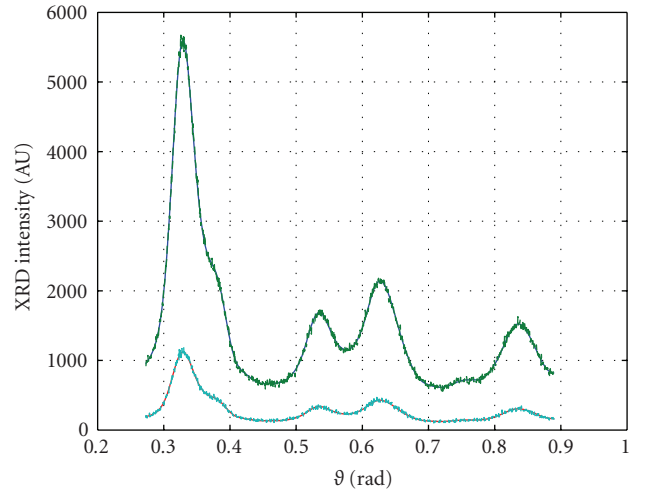


FIGURE 1: Synthetic XRD intensity profiles as a function of the scattering angle. From the upper to the lower profile, the NSR increases from 2% to 5% (see Figure 2 and text for details).

profiles with increasing NSRs. They were obtained by setting $\lambda = 0.15418$ nm and $a_{\text{f.c.c.}} = 0.40786$ nm in (6). The set of parameters used to compute the synthetic profiles are summarized in Table 1. Figure 2 shows the NSR of the synthetic profiles as a function of the scaling factor F ranging from 0 to 2. This range contains the NSR values usually measured in experimental profiles.

We also considered real data in order to validate our method. Three different samples were prepared with resultant mean diameters of 2.0, 3.2, and 4.1 nm, respectively (as measured by transmission electron microscope). The size distributions were approximately characterized by the same full width at half-maximum (≈ 1 nm) for all three systems.

4. NUMERICAL RESULTS

Noisy synthetic XRD patterns were generated corresponding to nanocrystalline samples of increasing size from 2 to 4 nm, and Poisson-distributed noise with increasing NSR from 2% to 10%. The HLSVD-PRO filter was then applied to the noisy synthetic XRD signals in order to study their

TABLE 2: Measure \mathcal{E} (see (3)) of the filter performance as a function of the order K of the filter. The synthetic XRD intensity data refer to different sample sizes and NSRs. The best performance corresponds to the order K reported in the middle row of each NSR value.

		2 nm	3 nm	4 nm
NSR = 10%	$K - 2$	1.86 ± 0.16	1.25 ± 0.09	1.67 ± 0.10
	$K = 9$	2.49 ± 0.16	2.34 ± 0.20	1.89 ± 0.19
	$K + 2$	2.43 ± 0.42	2.28 ± 0.21	1.73 ± 0.18
NSR = 5%	$K - 2$	2.17 ± 0.18	1.81 ± 0.16	1.52 ± 0.11
	$K = 11$	2.34 ± 0.18	1.87 ± 0.16	1.56 ± 0.12
	$K + 2$	2.22 ± 0.28	1.87 ± 0.16	1.48 ± 0.09
NSR = 2%	$K - 2$	1.80 ± 0.21	1.37 ± 0.32	1.13 ± 0.14
	$K = 15$	1.89 ± 0.28	1.54 ± 0.39	1.25 ± 0.06
	$K + 2$	1.86 ± 0.18	1.46 ± 0.38	1.18 ± 0.09

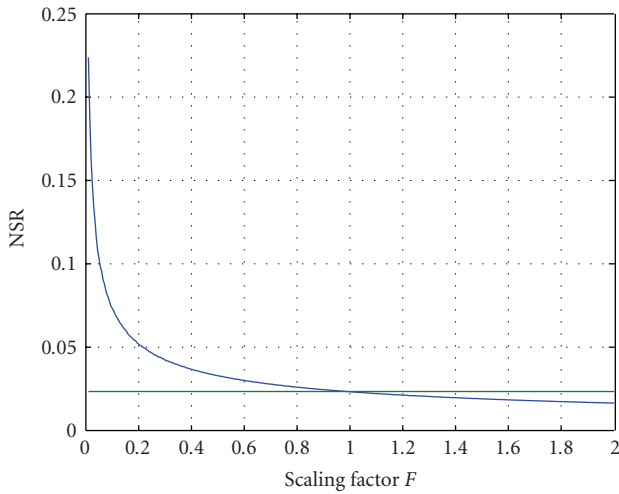


FIGURE 2: NSR as a function of the factor F , see text for details. The horizontal line separates the NSR values above and below $F = 1$.

properties under controlled conditions. Figure 3 displays an example of application of the HLSVD-PRO filter. A noisy synthetic XRD intensity profile is shown at the top of the figure. It corresponds to X-ray scattering from an Au sample having a 3 nm size with a Poisson-like noise with NSR = 10%. The filtered signal shown in the middle of the figure was obtained by setting $K = 9$ in the HLSVD-PRO filter. In the following (see Table 2 for results), we discuss in more detail the performance of the method when the values $K = 7$ and $K = 11$ are used. This value was estimated according to the criterion illustrated in Section 2. The values of λ_k were plotted by first sorting frequencies f_k in ascendant order. Specifically, a transition from high to small λ_k was observed at frequency $f_K = 35 \text{ rad}^{-1}$, which represents the K th frequency in the set of the sorted frequencies starting from the smallest one. For a comparison, the discrete Fourier transform (DFT) of the noisy synthetic XRD signal is reported at the bottom of Figure 4. Again, a phenomenon of transition from high to small singular values occurs in the same region of the spectrum, as observed at the top of the figure. However, the tran-

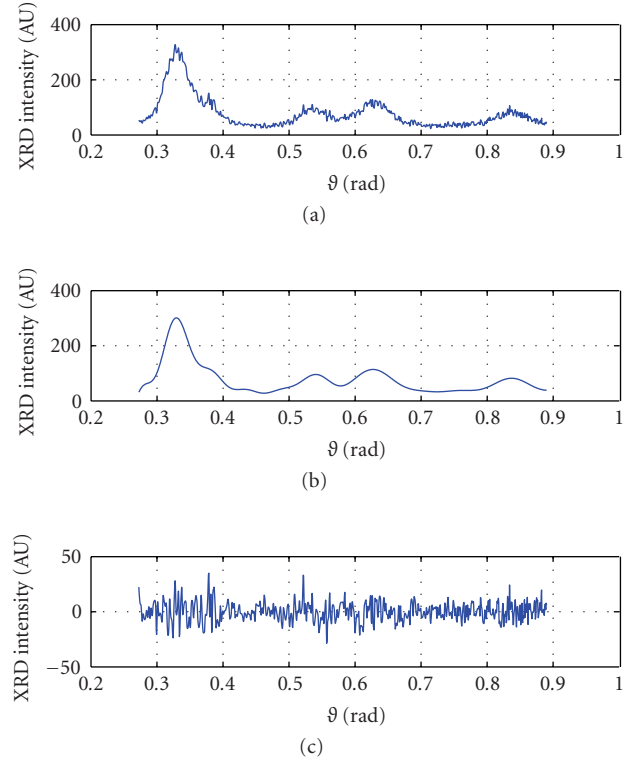


FIGURE 3: Three nm Au synthetic sample: (a) noisy (NSR = 10%) synthetic XRD intensity profile as a function of the scattering angle ϑ ; (b) filtered XRD intensity profile; (c) difference between measured and filtered profiles.

sition frequency is much more difficult to localize than in the HLSVD-PRO filter case. The same behavior is observed when using the WF filter. This makes troublesome the application of DFT and WF filters to clean noisy XRD data. It is worth noting that this difference between the HLSVD-PRO and Fourier-frequency-based filters is relevant when the filter is intended to be used during interactive XRD data analysis. In this case, the successful application of an easy-to-use blackbox filter becomes crucial.

Coming back to Figure 3, the difference between the values of noisy and filtered profiles is shown at the bottom. To quantify the performance of the filter, the filtered signal was compared with the noiseless synthetic XRD signal (see Figure 5). For the sake of completeness, we also report in Figure 5 the residue between the noiseless and the filtered signals. This can be done only with synthetic signals as experimental XRD data without noise are not available. To give a statistical significance to these measures a Monte Carlo experiment was carried out. More precisely, the HLSVD-PRO was applied to 1000 noisy synthetic profiles generated by considering samples of the same size undergoing different NSRs. For each filtered profile, the filter performance measure \mathcal{E} , defined in (3), was estimated by calculating the mean value and the standard deviation. For each sample size and NSR, the mean and standard deviation are obtained using 1000 synthetic XRD intensity profiles with different

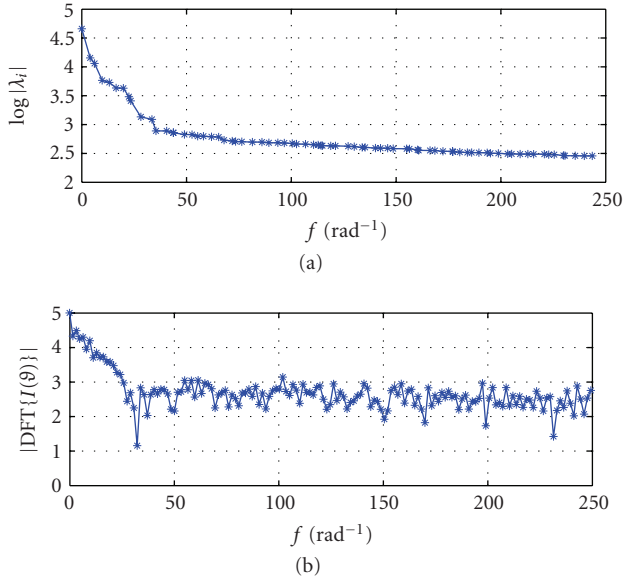


FIGURE 4: Three nm Au synthetic sample with NSR = 10%: (a) amplitude of eigenvalues λ_k versus frequency f_k , $k = 1, \dots, r$; (b) portion of the DFT amplitude spectrum of the noisy synthetic XRD intensity profile. Both plots refer to the XRD intensity profile shown at the top of Figure 3.

noise realizations having the same NSR. The sensitivity to the number K of sinusoids of the HLSVD-PRO filter was also studied. This number was slightly varied around the optimal K value selected by using the threshold criterion. The performance results were compared in order to validate the choice of the optimal K value. In particular, K was increased and decreased by 2, as discussed in Section 1. The results of such a comparison are summarized in Table 2 and they show that the proposed threshold criterion provides the value of K corresponding to the best performance of the HLSVD-PRO filter.

The filter was also applied to real XRD intensity profiles of Au samples of sizes 2, 3.2, and 4.1 nm. Figure 6 shows at top the profile of a 3.2 nm Au sample with NSR = 2.3%. The latter is computed as $\|\sigma\|/\|I\|$, where σ and I are vectors with the measured error and the intensity values, respectively. Since in the case of XRD signals, the noise follows the Poisson distribution, σ is given by \sqrt{I} . The result obtained by HLSVD-PRO is displayed in the middle of the figure. At bottom, the plot of singular values is depicted versus the frequency. Components with a frequency higher than $f_K = 34 \text{ rad}^{-1}$, due to noise, were removed. Denoising a real XRD profile of 500 intensity data samples, as typical ones used in the present study, requires about 11 seconds, using Matlab 7 on a machine with an Intel Xeon 2.80 GHz processor and a 512 KB cache size.

Finally, as a matter of comparison, we applied two well-known parametric algorithms that are commonly used for spectral analysis: MUSIC and ESPRIT [17]. Such methods are generally expected to be more effective spectral tools compared to DFT since they rely on the use of a model func-

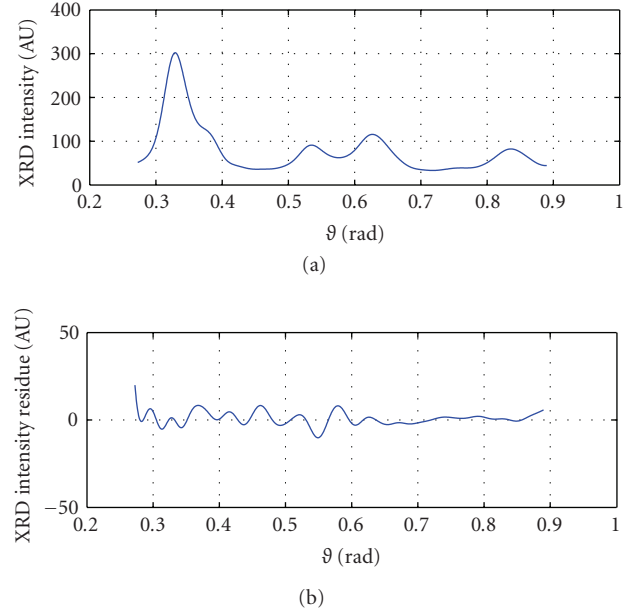


FIGURE 5: Three nm Au synthetic sample with NSR = 10%: (a) noiseless synthetic XRD intensity profile as a function of the scattering angle θ ; (b) difference between noiseless and filtered (see middle plot of Figure 3) profiles.

tion. However, in the present case where the signal is better modelled with damped sinusoids, the aforementioned methods are not able to correctly filter the signal. This limitation comes from the use of prescribed model functions that do not account for damping. Extensive simulation studies by using synthetic as well as real data show that MUSIC and ESPRIT fail. For instance, for the real XRD intensity reported in Figure 6, we computed the residue-to-signal ratio (RSR). We obtain the following results: RSR = 54% (ESPRIT), 51% (MUSIC), 2% (HLSVD).

5. CONCLUSIONS

A filter based on the HLSVD-PRO method has been presented. It has been applied to filter XRD patterns of nanocluster powders. The filter performance has been studied on synthetic and real XRD patterns with different NSRs. Results show that the proposed filter is robust and computationally efficient. A further advantage is its user-friendliness that makes it a useful blackbox tool for the processing of XRD data.

APPENDICES

HSVD is a subspace-based parameter estimation method in which the noisy signal is arranged in a Hankel matrix H . Its SVD allows to compute a “signal” subspace and a “noise” subspace. In fact, if H is constructed from a noiseless signal, the data matrix H has exactly rank equal to K , the number of exponentials that models the underlying signal. Due to the presence of the noise, H becomes a full-rank matrix.

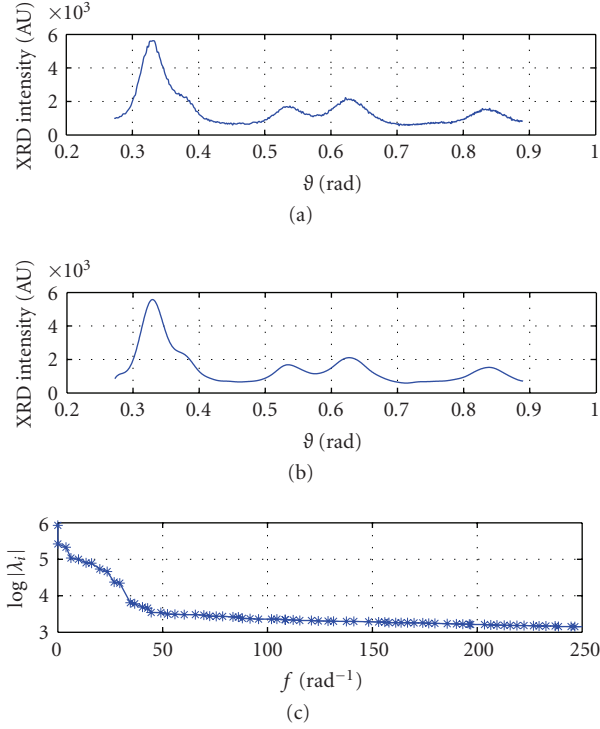


FIGURE 6: Au real sample of 3.2 nm: (a) noisy (NSR = 2.3%) XRD intensity profile as a function of the scattering angle ϑ ; (b) filtered XRD intensity profile; (c) amplitudes of eigenvalues λ_k versus frequency f_k , $k = 1, \dots, q$.

However, as long as the SNR of the signal is not too low, one can still define the “numerical” rank being approximately equal to K . Then, the “signal” subspace is found by truncating the SVD of the matrix H to rank K .

In the following subsections, the method will be derived in the context of linear algebra.

A. HSVD: THE ALGORITHM

The N data points defined in (1) are arranged into a Hankel matrix H of dimensions $L \times M$, with $L + M = N + 1$ and $L \simeq N/2$,

$$H_{L \times M} = \begin{bmatrix} I_0 & I_1 & \cdots & \cdots & I_{M-1} \\ I_1 & I_2 & \cdots & \cdots & I_M \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{L-1} & I_{L-2} & \cdots & \cdots & I_{N-1} \end{bmatrix}_{L \times M}. \quad (\text{A.1})$$

The SVD of the Hankel matrix is computed as

$$H_{L \times M} = U_{L \times L} \Sigma_{L \times M} V_{M \times M}^H, \quad (\text{A.2})$$

where $\Sigma = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_r\}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$, $r = \min(L, M)$, U and V are orthogonal matrices, and the superscript H denotes the Hermitian conjugate. The SVD is computed by using the Lanczos bidiagonalization algorithm with partial reorthogonalization [18]. This algorithm computes two matrix-vector products at each step. Exploiting the

structure of the matrix (A.1) by using the FFT, the latter computation requires $O((L+M) \log_2(L+M))$ rather than $O(LM)$.

In order to obtain the “signal” subspace, the matrix H is truncated to a matrix H_K of rank K ,

$$H_K = U_K \Sigma_K V_K^H, \quad (\text{A.3})$$

where U_K , V_K , and Σ_K are defined by taking the first K columns of U and V , and the $K \times K$ upper-left matrix of Σ , respectively. The way of choosing K is described at the beginning of Section 4. As a subsequent step, the least-squares solution E of the following overdetermined set of equations is computed as

$$U_K^{(\text{top})} \simeq U_K^{(\text{bottom})} E, \quad (\text{A.4})$$

where $U_K^{(\text{bottom})}$ and $U_K^{(\text{top})}$ are derived from U_K by deleting its last and first rows, respectively. Equation (A.4) follows from the shift-invariance property holding for the Vandermonde decomposition of the Hankel matrix H [7]. The K eigenvalues \hat{z}_k of the matrix E are used to estimate the frequencies \hat{f}_k and the damping factors \hat{d}_k of the model damped sinusoids from the relationship

$$\hat{z}_k = \exp [(-\hat{d}_k + i2\pi \hat{f}_k) \Delta \vartheta], \quad (\text{A.5})$$

as

$$\begin{aligned} \hat{d}_k &= -\frac{\Re[\log(\hat{z}_k)]}{\Delta \vartheta}, \\ \hat{f}_k &= \frac{\Im[\log(\hat{z}_k)]}{(2\pi \Delta \vartheta)}, \end{aligned} \quad (\text{A.6})$$

with $k = 1, \dots, K$.

B. HSVD: NOISELESS DATA

Arrange the N noiseless data points I_n^0 defined in (1) in a Hankel matrix H of dimensions $L \times M$, with L and M greater than K and $N = L + M - 1$,

$$H = \begin{bmatrix} I_0^0 & I_1^0 & \cdots & I_{M-1}^0 \\ I_1^0 & I_2^0 & \cdots & I_M^0 \\ \vdots & \vdots & \vdots & \vdots \\ I_{L-1}^0 & I_{L-2}^0 & \cdots & I_{N-1}^0 \end{bmatrix}. \quad (\text{B.1})$$

The model of (1) can be rewritten in terms of complex amplitudes c_k and signal poles z_k as follows:

$$I_n^0 = \sum_{k=1}^K c_k z_k^n, \quad n = 0, \dots, N-1, \quad (\text{B.2})$$

where $c_k = a_k \exp(i\varphi_k) \exp(-\hat{d}_k + i2\pi \hat{f}_k) \vartheta_0$ and $z_k = \exp(-\hat{d}_k + i2\pi \hat{f}_k) \Delta \vartheta$. Using this model function, the Hankel

matrix H can be factorized as follows:

$$H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_K \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{L-1} & z_2^{L-1} & \cdots & z_K^{L-1} \end{bmatrix} \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_K \end{bmatrix} \\ \times \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_K \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{M-1} & z_2^{M-1} & \cdots & z_K^{M-1} \end{bmatrix}^T = SCT^T. \quad (\text{B.3})$$

This factorization is called Vandermonde decomposition and from it the signal parameters can immediately be derived. A well-known algorithm to directly compute the Vandermonde decomposition is available in the literature and is called Prony's method [19–22]. Here, a more reliable approach, based on an indirect computation of the parameters, is adopted. This approach is described below. From (B.3), it can be easily proved that the matrix S satisfies the so-called shift-invariance property, that is,

$$S^1 = S_1 Z, \quad (\text{B.4})$$

where S^1 and S_1 are derived from S by deleting its first and last rows, respectively, and Z is a $K \times K$ complex diagonal matrix with entries equal to the K signal poles z_k , $k = 1, \dots, K$. The rank of the matrix H is equal to K , and thus, its SVD has the following form:

$$H = U\Sigma V^H = \begin{bmatrix} U_K & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_K & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_K & V_2 \end{bmatrix}^H = U_K \Sigma_K V_K^H, \quad (\text{B.5})$$

where $U_K \in \mathbb{C}^{L \times K}$, $U_2 \in \mathbb{C}^{L \times (L-K)}$, $\Sigma_K \in \mathbb{C}^{K \times K}$, $V_K \in \mathbb{C}^{M \times K}$, $V_2 \in \mathbb{C}^{M \times (M-K)}$. From the comparison of (B.3) and (B.5), it follows that S and U_K span the same column space, and hence are equal up to a multiplication by a nonsingular matrix $Q \in \mathbb{C}^{K \times K}$, that is,

$$U_K = SQ. \quad (\text{B.6})$$

Using (B.6), the shift-invariance property of (B.4) becomes

$$U_K^1 = U_{K1} Q^{-1} Z Q. \quad (\text{B.7})$$

The matrix $Q^{-1} Z Q$ can be determined as the least-squares solution of (B.7). Several reliable and efficient algorithms are available in the literature and they exploit well-known algebraic tools such as the QR decomposition, the SVD decomposition, and so forth. The reader is referred to [12, 23], where an exhaustive overview on the computation of the least-squares solution of a system of equations is provided. Since the eigenvalues of $Q^{-1} Z Q$ and Z are equal, the signal poles are easily derived as

$$\{z_k\}_{k=1}^K = \text{eig}(Q^{-1} Z Q) = \text{eig}(Z), \quad (\text{B.8})$$

where the function $\text{eig}(\cdot)$ determines the eigenvalues of the matrix between brackets.

From the signal poles, frequency and damping factors are estimated. By filling in these estimates into the model function (B.2), a new system of equations is obtained with unknowns equal to the complex variables c_k . Its solution provides estimates for the amplitudes and the phases.

C. HSVD: NOISY DATA

When noise affects the data, as in real MRS signals, relation (B.5) no longer holds. Although no exact solution of the shift-invariance property exists, if the noise is small compared to the signal, H can be approximated by the truncated SVD, that is,

$$H = U\Sigma V^H \approx U_K \Sigma_K V_K^H = H_K, \quad (\text{C.1})$$

where U_K and V_K are the first K columns of U and V , respectively, and Σ_K is the $K \times K$ upper-left submatrix of Σ .

The matrix H_K has rank K but its Hankel structure has been destroyed by the truncation of the SVD. Therefore, there exists no exact solution of the system in (C.1). However, estimates of the signal poles can still be obtained by solving the aforementioned system in an LS sense and the signal parameters can be derived from such estimates as in the noiseless case. Further details about the derivation of HSVD can be found in [7, 24].

ACKNOWLEDGMENTS

The authors thank A. Cervellino, C. Giannini, and A. Guagliardi for kindly providing us with experimental XRD data.

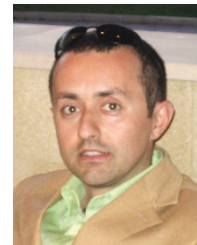
REFERENCES

- [1] B. Mierzwa and J. Pielaszek, "Smoothing of low-intensity noisy X-ray diffraction data by Fourier filtering: application to supported metal catalyst studies," *Journal of Applied Crystallography*, vol. 30, no. 5, pp. 544–546, 1997.
- [2] A. Hieke and H.-D. Dörfner, "Methodical developments for X-ray diffraction measurements and data analysis on lyotropic liquid crystals applied to K-soap/glycerol systems," *Colloid and Polymer Science*, vol. 277, no. 8, pp. 762–776, 1999.
- [3] M. Schmidt, S. Rajagopal, Z. Ren, and K. Moffat, "Application of singular value decomposition to the analysis of time-resolved macromolecular X-ray data," *Biophysical Journal*, vol. 84, no. 3, pp. 2112–2129, 2003.
- [4] S. Rajagopal, M. Schmidt, S. Anderson, H. Ihee, and K. Moffat, "Analysis of experimental time-resolved crystallographic data by singular value decomposition," *Acta Crystallographica Section D*, vol. 60, no. 5, pp. 860–871, 2004.
- [5] E. E. Aubanel and K. B. Oldham, "Fourier smoothing without the fast Fourier transform," *Byte*, vol. 10, no. 2, pp. 207–222, 1985.
- [6] C. Wooff, "Smoothing of data by least squares fitting," *Computer Physics Communications*, vol. 42, no. 2, pp. 249–251, 1986.
- [7] H. Barkhuijsen, R. de Beer, and D. van Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," *Journal of Magnetic Resonance*, vol. 73, no. 3, pp. 553–557, 1987.

- [8] T. Laudadio, N. Mastronardi, L. Vanhamme, P. van Hecke, and S. van Huffel, "Improved Lanczos algorithms for black-box MRS data quantitation," *Journal of Magnetic Resonance*, vol. 157, no. 2, pp. 292–297, 2002.
- [9] D. J. Wales, "Structure, dynamics, and thermodynamics of clusters: tales from topographic potential surfaces," *Science*, vol. 271, no. 5251, pp. 925–929, 1996.
- [10] R. W. Siegel, E. Hu, D. M. Cox, et al., "Nanostructure Science and Technology. A Worldwide Study," The Interagency Working Group on NanoScience, Engineering and Technology, <http://www.wtec.org/loyola/nano/>.
- [11] D. Zanchet, M. B. D. Hall, and D. Ugarte, "Structure population in thioi-passivated gold nanoparticles," *Journal of Physical Chemistry B*, vol. 104, no. 47, pp. 11013–11018, 2000.
- [12] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, no. 5, pp. 403–420, 1970.
- [13] E. Anderson, Z. Bai, C. Bischof, et al., *LAPACK Users' Guide*, SIAM, Philadelphia, Pa, USA, 1995.
- [14] R. A. Young, *The Rietvel Method*, Oxford University Press, New York, NY, USA, 1993.
- [15] A. Cervellino, C. Giannini, and A. Guagliardi, "Determination of nanoparticle structure type, size and strain distribution from X-ray data for monatomic f.c.c.-derived non-crystallographic nanoclusters," *Journal of Applied Crystallography*, vol. 36, no. 5, pp. 1148–1158, 2003.
- [16] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, University Scientific Books, Sausalito, Calif, USA, 1997.
- [17] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.
- [18] H. D. Simon, "The Lanczos algorithm with partial reorthogonalization," *Mathematics of Computation*, vol. 42, no. 165, pp. 115–142, 1984.
- [19] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
- [20] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.
- [21] B. J. C. Baxter and A. Iserles, "On approximation by exponentials," *Annals of Numerical Mathematics*, vol. 4, pp. 39–54, 1997, The heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T. J. Rivlin, hskip 1em plus 0.5em minus 0.4em.
- [22] G. Beylkin and L. Monzon, "On approximation of functions by exponential sums," *Applied and Computational Harmonic Analysis*, vol. 19, no. 1, pp. 17–48, 2005.
- [23] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, Pa, USA, 1996.
- [24] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *Journal of the Optical Society of America*, vol. 73, no. 12, pp. 1799–1811, 1983.

M. Ladisa received the Laurea and Ph.D. degrees in physics from the University of Bari, Bari, Italy, in 1997 and 2001, respectively. He is currently a Researcher with the Istituto di Cristallografia (IC), National Research Council (CNR), Bari, Italy.

A. Lamura received the Laurea and Ph.D. degrees in physics from the University of Bari, Bari, Italy, in 1994 and 2000, respectively. He is currently a Researcher with the Istituto per le Applicazioni del Calcolo (IAC), National Research Council (CNR), Bari, Italy.



T. Laudadio received the Laurea degree in mathematics from the University of Bari, Bari, Italy, in 1992, and the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in 2005. She is currently a Research Fellow with the Istituto di Studi sui Sistemi Intelligenti per l'Automazione (ISSIA), National Research Council (CNR), Bari, Italy.



G. Nico received the Laurea and Ph.D. degrees in physics from the University of Bari, Bari, Italy, in 1993 and 1999, respectively. He is currently a Researcher with the Istituto per le Applicazioni del Calcolo (IAC), National Research Council (CNR), Bari, Italy.