INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS D: APPLIED PHYSICS

J. Phys. D: Appl. Phys. **34** (2001) 209–217 www.iop.org/Journals/jd PII: S0022-3727(01)18079-4

Extraction of the fine structure from x-ray absorption spectra

K V Klementev

Moscow State Engineering Physics Institute, 115409 Kashirskoe sh. 31, Moscow, Russia

Received 17 October 2000

Abstract

We propose two methods for obtaining the atomic-like background for the x-ray absorption fine structure (XAFS): the methods of smoothing spline and of Bayesian smoothing. Both are capable of using the prior information, calculated or experimental, about the background. The XAFS signals obtained by these techniques are shown to be significantly corrected in comparison with standard methods. The method of Bayesian smoothing is the only method that gives the errors of approximation of the atomic-like background by an artificial smooth function. These errors are shown to be the main source of the uncertainty of the XAFS function.

1. Introduction

The x-ray absorption fine structure (XAFS), χ, is determined in E- or k-space ($k = \sqrt{2m_e(E - E_0)}/\hbar$—$E_0$ being the energy of the corresponding absorption edge) as [1]:

$$\chi = (\mu - \mu_0)/\mu_0 \tag{1}$$

where $\mu = \mu_{\mathrm{exp}} - \mu_{\mathrm{b}}$ is the measured absorption minus pre-edge background (the latter is usually approximated by a Victoreen law), and μ_0 is the 'atomic' absorption due to electrons of considered atomic level (post-edge background). Since the electronic state of an embedded atom is, in general, different from its state in gaseous phase, μ_0 is not the same as for an isolated atom and cannot be found experimentally. Therefore, a demand arises for an artificial construction of μ_0. This is the most difficult procedure in the extraction of XAFS from the measured absorption because one cannot definitely distinguish the environmental-born part of absorption from the atomic-like one. Most methods for determination of the post-edge background are based on the assumption of its smoothness, and the only criterion for its validity is the absence of low-frequency structure in χ. From the critical review of existing post-edge background methods that we make in this paper, it follows that their main disadvantages are: (i) instability and ambiguity near the absorption edge; and (ii) inability to take into account independent information about μ_0, calculated or measured. In principle, modern methods for theoretical XAFS calculations (e.g. FEFF8 [2]) are able to calculate μ_0. However, these calculations may be not accurate at high energies and therefore cannot be used for direct background removal.

In this paper, we present methods for extracting the post-edge background from experimental data with the use of calculated μ_0. Besides the significant influence on the XAFS signal, especially first coordination sphere, the use of the prior information raises the accuracy of the sought μ_0 near the absorption edge.

Estimation of the uncertainties of μ_0 construction has always been an unresolvable problem for any method. The problem is much deeper than the simple determination of the numerical errors of μ_0 drawing; the problem is to find out how much the artificially constructed μ_0 differs or can differ from the real atomic-like absorption. The standard practice in XAFS analysis is to ignore these uncertainties. However, it is well known that all sources of errors affect the accuracy of the final conclusions for the structural information. When this information happens to conflict with the data of other experimental techniques, often vague 'systematic errors' are mentioned. While many sources of systematic errors are well understood—(i) acquisition-related ones are removed by proper sample preparation or can be calculated and (ii) imperfection of *ab initio* references can be accounted for in the fitting procedure [3]—the errors of approximation of μ_0 have not been discussed yet and not even defined.

We believe that the errors of μ_0, always neglected because undefined, contribute essentially to systematic errors. In this paper, we also show how this contribution can be determined. For this purpose, we propose the method of Bayesian smoothing that gives the errors within which lie all functions of specified smoothness and, among them, the artificial Bayesian μ_0 and the real atomic-like absorptance.

2. Review of post-edge background removal methods

A rich variety of computer programs for XAFS spectra processing are collected on the web site of the International

XAFS Society [4]. The vast majority of these use, as an approximation for the atomic-like absorption, a smoothing spline or more general piecewise-polynomial representation. For example, by the method of [5], the construction of μ_0 includes several preliminary stages: μ_0 is approximated by a low-degree polynomial; through obtained $\chi(k)k^w$, an additional $\mu_0^{\rm I}$ is drawn again as a low-degree polynomial and subtracted; a smoothing spline then approximates one more additional $\mu_0^{\rm II}$. The sum of all μ_0s gives the total atomic-like background. From our experience, we know that the spline function is very unstable with respect to the small variations of the $k_{\rm min}$, the left end of the XAFS range, and near $k_{\rm min}$ the spline can bend rapidly. Although the preliminary stages with several auxiliary parameters (degrees of polynomials and k^w-weighting) can give a corrected χ, they cannot make the spline be stable with respect to $k_{\rm min}$.

In [6], an iterative approach to 'atomic background' removal was developed. First, splines or low-order polynomials are used to obtain a rough estimate of the background; this alone is enough to have a reliable χ at $k > 5$–6 Å^{-1}. Over that range, the theoretical $\chi_{\rm th}$ is fitted to the χ obtained. The resulting fit parameters are then used to generate $\chi_{\rm th}(k)$ that extends down to low k. This function is transformed back into E-space and μ_0 is obtained as $\mu_0 = \mu/(\chi_{\rm th}+1)$ that need be a little smoothed or fitted by an additional spline. Since the logic of reasoning was inversed: not 'to find μ_0 for χ', but 'to find χ for μ_0', the method is suited for the quest of peculiarities on the μ_0 curve, not for structural XAFS researches. In addition, the accuracy of the model parameters appears to be unknown in principle: all that is not described by the model is included into μ_0.

For the determination of the background absorption μ_0, Boland *et al* [7] considered the damping of the XAFS amplitude resulting from the measurements with low resolution (with a large slit width). The superimposition of two spectra measured with different energy resolution gives the intersection points, the part of which belong to μ_0. Then, through the obtained nodal points, a spline is drawn. As Boland *et al* [7] noted, the measurements of the spectra with worsened resolution are not necessary: the spectra could be damped by the convolution with a 'rocking curve', approximated by a Gaussian function. Of course, the method works correctly only with a small variation of the Gaussian curve width since for the large width not only the XAFS amplitude is damped but the very edge is washed out. Because of this, only the extended part of a spectrum could be reliably determined.

The damping of the XAFS amplitude can be due to other reasons. For instance, as was pointed out in [7], the nodal points may be obtained from the variable-temperature study. This idea was realized in [8] and relies on the assumption that, while the XAFS is temperature dependent, the atomic-like background is not. But, for all that, it is important that the phase difference between XAFS of different temperatures was negligible, which is true only for low wavenumbers. Furthermore, the method is suitable only for some particular cases (to say nothing of the need for a measured temperature series of spectra). In [8], it was demonstrated for the x-ray absorption data for the L_3 edge of solid Pb. In those spectra, the first crossing of μ and μ_0 occurs already at ~15 eV above the edge. In our sample spectra, considered below, the first crossing occurs only at ~30 eV, which allows one to find at most 2–3 points, the first of these being situated at $k \gtrsim 2.5$ Å^{-1}.

Newville *et al* [9] suggested the subtraction of a spline that best eliminates the nonstructural, low-r, portion of $\chi(r)$, the Fourier transform of $\chi(k)k^w$. The spline is drawn through knots, equally spaced in k-space. The ordinates of these are varied to minimize $\chi(r)$ or $|\chi(r) - \chi_{\rm st}(r)|$ in the chosen low-r region $0 \leqslant r \leqslant r_0$, where $\chi_{\rm st}$ is a 'standard' χ, calculated or experimental. In [9], it was asserted that one need know the 'standard' $\chi_{\rm st}$ merely approximately since it is used only to get an estimate of the leakage from the first shell to the region minimized. Having omitted the question of the accuracy of found knots, Newville *et al* [9] made a fine comparison between several theoretical models for χ calculations. However, it is not evident that the knot positions are uniquely determined by this procedure. Moreover, when the number of knots is close to $N_{\rm max} = 2r_0\Delta k/\pi$, where Δk is the k range of useful data, correlations between knot positions become very high. We have used this method many times and have found that the pair-correlation coefficients between first knots can be as high as 0.997, so the variation of one of the knots can be compensated by the movement of another one. Fortunately, this ambiguity is not pronounced in r-space, and $\chi(r)$ can be regarded as reliable.

One more approach to the problem of μ_0 determination was reported in [10]. It is based on the simple identity that relates the Fourier transform (FT) of some function with the FT of its nth derivative:

$$\mathrm{FT}[f^{(n)}(k)] = (2\mathrm{i}r)^n \mathrm{FT}[f(k)] \qquad (2)$$

where the conjugate variables are k and $2r$. Assuming that for higher derivatives $\mu_0^{(n)} \ll \mu^{(n)}$, one obtains $\mathrm{FT}[\mu^{(n)}] \approx (2\mathrm{i}r)^n \mathrm{FT}[\chi\mu_0(k)]$; see figure 1 for our sample spectrum described below. The only way to obtain χ from this formula is to neglect the k-dependence of μ_0 and normalize χ to the absorptance step at the edge. In [10], the low-r part (which in our example is $0 \leqslant r \lesssim 1.1$) was cut off, and then the back FT was done. As a result, one obtains the fine structure and, having subtracted it from μ, the atomic-like background, on which some peculiarities due to multi-electron excitations can be distinguished. Like the method of [6], this method is suited for the quest of peculiarities on the μ_0 curve, not for structural XAFS researches, because of the evident distortion of the first peak on the FT by the contribution from the atomic-like background. To illustrate this assertion, in figure 1 we show the FT of the second derivative of the $\mu_0(k)$ that was found by the present method. As seen, this contribution is not as small in the vicinity of the first structural peak at $1 \lesssim r \lesssim 2$ Å.

If the electronic states of absorbing atoms in gaseous phase and in the compound of interest may be considered as equivalent, μ_0 can be set equal to the measured absorption in gas, as was done in [11] for solid, liquid and gaseous Kr. Some differences in energy positions and relative weights of double-electron excitation channels were taken into account by a model using simple empirical functions which were transferred then to the spectra of liquid and solid Kr.

3. Smoothing spline method

Owing to fast algorithm and easy program realization, the approximation of μ_0 by the smoothing spline has become

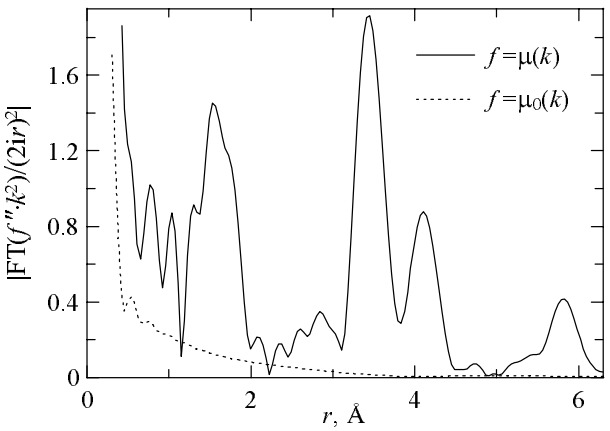

Figure 1. On the method of [10]. The solid curve is nearly proportional to the $|\mathrm{FT}[\chi(k)\cdot k^2]|$. The dotted curve is the contribution from the atomic-like absorption.

widespread. Let $N+1$ experimental values of μ_i be defined on the mesh E_i. The smoothing spline μ_0 minimizes the functional

$$J(\mu_0,\mu)=\int_{E_{\min}}^{E_{\max}}\left[\mu_0''(E)\right]^2\,\mathrm{d}E+\frac{1}{\alpha}\sum_{i=0}^{N}\left[\mu_{0i}-\mu_i\right]^2. \quad (3)$$

The smoothing parameter (or regularizer) α is the measure of the compromise between the smoothness of μ_0 and its deviation from μ. At $\alpha=0$, the smoothing spline exactly coincides with μ, at $\alpha\to\infty$ it degenerates to $\mu_0=$ const. The optimal regularizer should lead to μ_0 containing only low-frequency oscillations and, hence, to χ containing only structural oscillations. We shall consider below the formulation of a criterion for optimal α.

First, we address another problem of how to take into account the prior information about μ_0. Let μ_0 be approximately known in advance as μ_0^*, being measured in gaseous phase or calculated. In general, μ_0^* cannot be directly subtracted from μ. Measured spectra for different samples have different rates of decrease of μ, and the superimposition of one spectrum on another should be done after relevant shifting and/or scaling. If the condensed sample has holes, which is frequently the case, such a simple correction fails. On the other hand, the calculated μ_0 (we used the FEFF8.10 code) may give errors. Full multiple scattering results are not accurate at high energies; the multiple scattering expansion sometimes does not converge well in the XANES energy region.

Nevertheless, the prior information can be used. For that we should slightly modify the smoothing spline method. Now we will tend the second derivative of the sought $\mu_0(E)$ not to zero (at the specified deviation of μ_0 from μ) but to the second derivative of $\mu_0^*(E)$. The sought $\mu_0(E)$ now minimizes the functional

$$J^*(\mu_0,\mu)=\int_{E_{\min}}^{E_{\max}}\left[\mu_0''(E)-\mu_0^{*\prime\prime}(E)\right]^2\,\mathrm{d}E+\frac{1}{\alpha}\sum_{i=0}^{N+1}\left[\mu_{0i}-\mu_i\right]^2. \quad (4)$$

As seen, in fact there is no need to know μ_0^* itself, its second derivative is sufficient. The explicit presence of μ_0^* in the following formulae should be taken as a consequence of the technical trick applied; at first, μ_0^* is subtracted from the data, then it is added to the found spline.

Introducing $\tilde{\mu}_{0i}=\mu_{0i}-\mu_{0i}^*$, one obtains:

$$J^*(\mu_0,\mu)=\int_{E_{\min}}^{E_{\max}}\left[\tilde{\mu}_0''(E)\right]^2\,\mathrm{d}E+\frac{1}{\alpha}\sum_{i=0}^{N+1}\left[\tilde{\mu}_{0i}-(\mu_i-\mu_{0i}^*)\right]^2$$
$$=J\left(\tilde{\mu}_0,\mu_i-\mu_{0i}^*\right). \quad (5)$$

Thus, the problem is reduced to the preceding one in which, instead of initial data μ, the difference $\mu-\mu_0^*$ appears. Then, μ_0 is found from the smooth $\tilde{\mu}_0$ as $\mu_0=\tilde{\mu}_0+\mu_0^*$. Figure 2 shows the example of the atomic-like absorption approximation by the smoothing spline with and without the use of prior function. The latter was calculated by the FEFF8.10 program using self-consistent potential calculations with the ground-state exchange correlation potential, full multiple scattering approach and a 57-atom cluster of simple cubic perovskite structure with the lattice parameter 4.275 Å. We used the fine radial grid for calculations (rgrid = 0.01), otherwise at high energies the background curves upward. The calculated background was then multiplied by a constant factor to match the step height. Our calculated μ_0 has less of a slope at high energies than the measured absorption (see the dotted curve in figure 2(*a*)). Additional features due to possible multi-electron excitation channels should be added to the $\mu_0^*(E)$, if known. However, we do not intend to discuss here the validity of the prior function, we just demonstrate how the prior knowledge can be used for XAFS post-edge background removal. As seen, the use of $\mu_0^*(E)$ has led to considerable correction of the first shell signal.

Incidentally, we have solved another problem. It is well known that a spline is unstable with respect to the small variations of input parameters: number of nodes, nodal values of the processed function and limits of the integral. In our case, the spline is most sensitive to $E_{\min}$ due to the fast growth of μ in the edge. It turns out that the use of prior function has reduced the influence of the left hanging end: now it is no longer hanging but constrained by the prior function. Thus, the preliminary stages for smoothing spline, mentioned in the previous section, have become redundant.

Let us now define the criterion for the determination of the smoothing parameter. An attempt to solve the problem was made in [12], where the requirement $H_{\mathrm{R}}-H_{\mathrm{N}}\geqslant 0.05H_{\mathrm{M}}$ was proposed; H_{R} is the average value of the weighted Fourier transform magnitude between 0–0.25 Å, H_{M} is the maximum value of the transform magnitude between 1–5 Å and H_{N} is the average value of the transform magnitude between 9–10 Å attributed to the noise. Obviously, this criterion cannot pretend to the generality since it depends on the weighting (see [12], k^3) and the relative contribution of noise and the first coordination shell into spectra.

Here, we propose another approach to the problem, also based on the consideration of the heights of the FT peaks as functions of the regularizer α (see figure 3). On increasing α from zero, μ_0 starts to deviate from the experimental absorption μ, and $|\chi(r)|$ grows and then saturates, the peaks at larger r being saturated earlier. Clearly, α should be determined by the first peak height since it is the last to saturate. Let us define the start of saturation on the minimum of second derivative of the squared first peak with respect to $\ln\alpha$, the

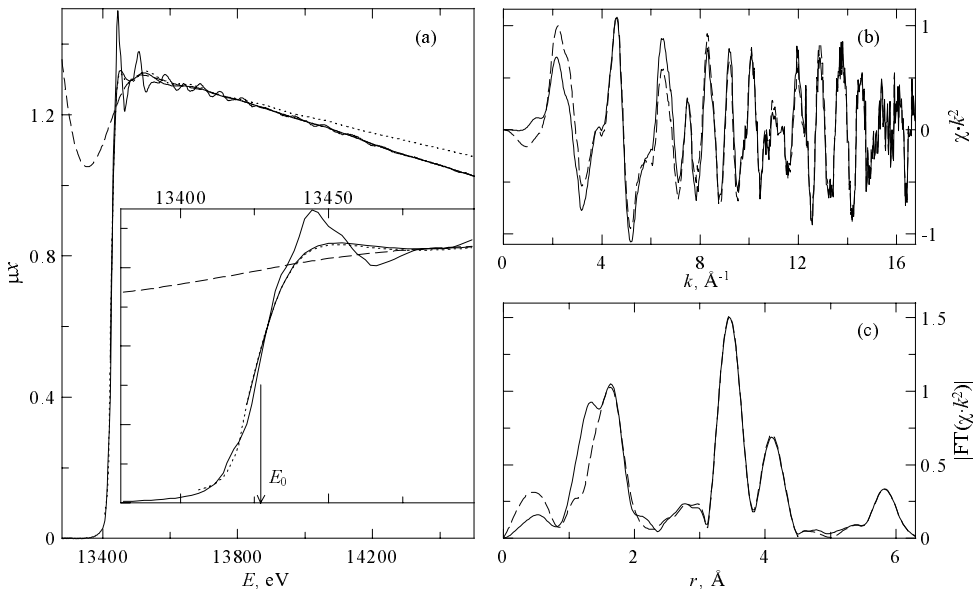

Figure 2. Extraction of XAFS from the measured absorption using the smoothing spline: (*a*) absorption coefficient μ, prior function μ_0^* calculated by FEFF8 (dotted curve) and various post-edge backgrounds; (*b*) XAFS functions obtained with these backgrounds; and (*c*) their module of FT. $\mu_0(E)$, $\chi(k)\cdot k^2$ and $\chi(r)$ obtained (——) with use of the prior function, and (- - - -) without prior function. The regularizer α is the same for both cases. For examples, we used the spectrum at the Bi L_3 absorption edge for $Ba_{0.6}K_{0.4}BiO_3$ at 50 K recorded in transmission mode at D-21 line of DCI (LURE). Energy step ∼1 eV, counting time 1 s. The double-crystal Si (311) monochromator was detuned to minimize harmonic contamination.

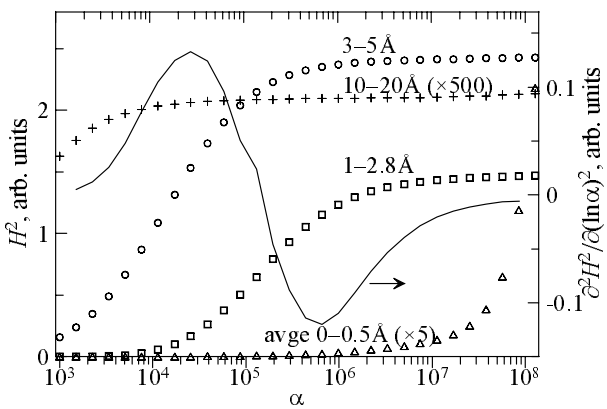

Figure 3. The $|\chi(r)|$ peak heights squared, H^2, maximal in the indicated areas and average over the range $0 < r < 0.5$ Å, as functions of α. The right axis relates to the second derivative of the square of first peak height with respect to $\ln\alpha$.

value of α in the minimum being declared to be optimal. It is seen that the increase of α from the optimal leads to unwanted rapid growth of $|\chi(r)|$ at low r. In the example in figure 2, the regularizer is optimized following our new criterion.

4. Bayesian smooth curve

The method of Bayesian smoothing is ideologically similar to the smoothing spline method, but has one global advantage: within no other existing approach can one define and determine the errors of the μ_0 construction. Since the Bayesian smoothing method finds the posterior distributions for all μ_{0i}, one can find not only average values but also standard deviations and any desirable moments. In addition, within the framework of the method it is possible (before the post-edge background construction) to deconvolve μ with the monochromator resolution curve. The weakness of the method is its low speed. On a modern PC, it takes a few minutes for the curve drawn through $N \sim 500$ points to be smoothed.

Detailed formalism of the Bayesian smoothing is given in the appendix. To determine $\bar{\mu}_{0i}$ and $\delta\mu_{0i}$ one should find eigenvalues and eigenvectors of a special five-diagonal square $N \times N$ matrix. In figure 4, the Bayesian smoothing was done on the mesh of $N = 536$ experimental points above the absorption edge, with and without the prior function described in the preceding section. In addition, we demanded from the Bayesian curve in figure 4 to pass through a point nearest at left to E_0 (for this, the five-diagonal matrix should be changed, see the appendix); this requirement does not practically affect μ_0, but affects $\delta\mu_{0i}$ in the low-k region, as we shall see below. The values $\bar{\mu}_{0i}$ and $\delta^2(\mu_{0i})$ were found by formulae (A.15) and (A.16). Since the smoothed values $\bar{\mu}_{0i}$ do not lie within the limits $\pm\delta\mu_{0i}$ from μ_i, we did not look for the most probable smoothness (see the appendix). Instead, we considered the regularizer to be known and equal to the optimal one found for the smoothing spline method. As seen, we have obtained completely the same XAFS function as that given by the previous method. So, the only thing that warrants such a slow method is that it gives the errors of the μ_0 construction.

5. The errors of χ extraction

To derive the expression for the error of χ extraction, one should note that because initially the pre-edge background was subtracted from $\mu_{\rm exp}$, the experimentally obtained μ_0 is dependent on $\mu_{\rm b}$ in such a way that the sum $\mu_0 + \mu_{\rm b}$ is always, independently of $\mu_{\rm b}$, a middle line of $\mu_{\rm exp}$. Therefore, the

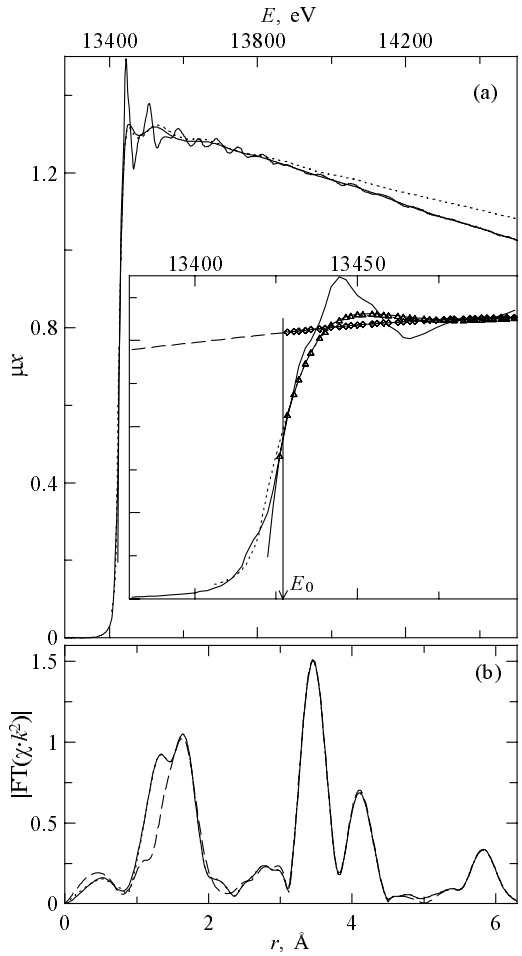

Figure 4. Extraction of XAFS from the measured absorption using Bayesian smoothing. Prior function $\mu_0^*(E)$ for the atomic-like absorption is represented by dots. $\mu_0(E)$, $\chi(k)\cdot k^2$ and $\chi(r)$ obtained (——) with the use of the prior function and (- - - -) without prior function. The dotted curve in (*b*), hardly seen in the low-r region only, is obtained without additional requirement for μ_0 to pass through a point immediately before E_0. The regularizer α is the same for all cases and equals the optimal one found for the smoothing spline.

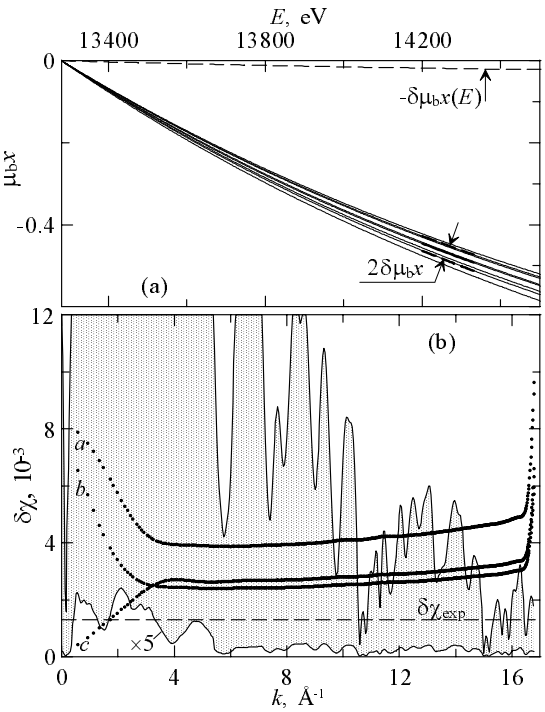

Figure 5. (*a*) Different experimental pre-edge backgrounds obtained for eight different scans, used to determine the contribution of $\delta\chi_b$. (*b*) The errors of χ extraction. The solid curve with grey shading represents the envelope of unweighted $\chi(k)$. The solid curve below is $\delta\chi_b$ magnified by a factor of five. The dashed line represents the noise estimate $\delta\chi_{exp}$ from FT. The dotted curves represent $\delta\chi_0$ obtained by the method of Bayesian smoothing (curve a) without and (curves b and c) with the prior information specifying the second derivative of μ_0. In addition, for curve c, we have used the information that μ_0 passes through a point immediately before E_0.

error of this sum is the error $\delta\mu_0$ of the post-edge background construction. Thus, from $\chi = [\mu_{exp} - (\mu_0 + \mu_b)]/[(\mu_0 + \mu_b) - \mu_b)]$ follows that $\delta\chi$ is the root-mean-square function of $\delta\chi_b = \chi\cdot\delta\mu_b/\mu_0$, $\delta\chi_{exp} = \delta\mu_{exp}/\mu_0$, $\delta\chi_0 = \delta\mu_0/\mu_0$ and $\chi\cdot\delta\mu_0/\mu_0$. We neglect the latter contribution in comparison with the third one. Let us consider the others.

(i) μ_b is obtained via extrapolation of the spectrum pre-edge part by a chosen law. There, a small random variation of μ_{exp} leads to a considerable dispersion in the extended region. An example of this random dispersion is shown in figure 5(*a*) for eight different scans taken at various temperatures; $\delta\mu_b$ was calculated as the standard deviation from the average background. Instead of χ in the formula $\delta\chi_b = \chi\cdot\delta\mu_b/\mu_0$, we used the envelope of χ, drawn in figure 5(*b*) with grey shading. Even at high energies, the ratio μ_b/μ_0 equals a few per cent, so $\delta\chi_b$ is the smallest contribution.

(ii) To determine that the experimental noise is a straightforward task for r-space, where EXAFS signals at high r clearly have noise character. Assuming noise in k-space and r-space to be constant, via Parseval's identity one obtains [13]:

$$(\delta\chi_{exp})^2 = n_r^2\frac{\pi}{dk}\frac{2w+1}{k_{max}^{2w+1} - k_{min}^{2w+1}} \qquad (6)$$

where as n_r the mean value of $|\chi(r)|$ serves, usually over the range $15 < r < 25$ Å. As seen from the formula, $\delta\chi_{exp}$ depends on dk, the size of the evenly-spaced k-grid. Although we have already used the Fourier transform, the question of the choice of dk was not yet raised. The algorithm of fast FT needs the transformed function to be set on a uniform grid. Having chosen a small dk, we artificially obtain a large number of 'experimental' values. Naturally, this trick would not give more information than we have from experiment and a small dk must give a large $\delta\chi_{exp}$. Conversely, we can obtain a small $\delta\chi_{exp}$ at the expense of loss of experimental points at large dk. In our example, the choice of $dk = 0.03$ Å^{-1} was based on the equality of numbers of experimental points and the nodes of the grid. The noise calculated via equation (6) is shown in figure 5(*b*) by dashes.

(iii) One may argue that the errors of μ_0 are not as important since they are of low frequency and presumably are outside (in r-space) the region of XAFS analysis. This may be true only if one performs a fitting of the $\chi(r)$ or

filtered $\chi(k)$. But many researchers try to fit the entire $\chi(k)$ function, together with possibly undetermined slow varying background. This is the first reason why we have to know the errors of μ_0. The second, and the main, reason is that the μ_0 drawn is not the atomic background itself but an artificial representation of it. Therefore, even if our μ_0 is of zero frequency, considerable spectral range (especially the signal from the first shell) may be distorted by this representation. Speaking of the errors of μ_0 construction, we mean the deviation of our artificial μ_0 from the real background. The real background must be *nearly* a middle line of the oscillating part of μ with the second derivative which is either *nearly* small or *nearly* close to the second derivative of the prior function. The Bayesian analysis gives errors $\delta\mu_{0i}$ within which lie *all* such functions, and, among them, the real post-edge background. Thus, the errors $\delta\mu_{0i}$ are very important because they account for possible deviation from the true atomic background.

Having determined $\bar{\mu}_{0i}$ and $\delta\mu_{0i}$, we express the errors $\delta\chi_0$ as $\delta\mu_{0i}/\bar{\mu}_{0i}$. For our example spectrum they are shown in figure 5(*b*) by dots. As seen, the introduction of the prior information has significantly diminished $\delta\chi_0$. This is quite natural, since any decrease of our ignorance about μ_0 should narrow the posterior distribution of μ_{0i} for all i. Of course, this concerns the experimental information as well: $\delta\chi_0$ are the smaller (among equal-length spectra) the more measured points the spectrum has.

Near the hanging ends of the spectrum, the errors of μ_0 construction are, as expected, significantly larger than in the inner area. The ends should be deleted after $\chi(k)$ extraction or treated in a different way; we can take into account other information that the atomic-like absorption must coincide with the total absorption (minus pre-edge background) at energies $E < E_0$. As seen (curve c in figure 5(*b*)), in this case the left spectrum end is saved for further analysis.

Now we can see that the main contribution to the error of χ extraction comes from the errors of the post-edge background construction. It is quite reasonable to demand that the errors of $\delta\chi$ were less than the XAFS signal (the envelope of χ). For the Bayesian curve a, this range is $0 \leqslant k \lesssim 14$ Å^{-1}. For the Bayesian curves b and c, this range is wider, $0 \leqslant k \lesssim 16$ Å^{-1}.

For correct subsequent fitting of the XAFS signal, one should determine, ideally, all sources of errors. However, practically all programs of XAFS spectra processing [4] use the noise, determined from FT, as uncertainties ε_i of $\chi(k)$ determination in the definition of χ^2 statistics:

$$\chi^2 = \frac{N_{\max}}{M}\sum_{i=1}^{M}\frac{\left[(\chi_{\exp})_i - (\chi_{\mathrm{mod}})_i\right]^2}{\varepsilon_i^2}. \qquad (7)$$

It would be more correct to consider as ε_i the larger function from the above listed contributions to $\delta\chi$. In our example, this is $\delta\chi_0$. Recently, Krappe and Rossner [3] have come yet further. Into ε_i they include the noise, the error from the truncation in the EXAFS formula summation, and the errors derived from the theoretically calculated values. Among the latter errors, the errors due to some uncertainties in scattering amplitudes and phases can be easily defined because the amplitudes and phases are explicitly contained in the EXAFS formula. The backgrounds μ_b and μ_0 are also calculated in [3], but since they contribute to χ in a complicated way, their errors, as before, are considered as vague 'systematic errors' and modelled as convenient. In [3], they were modelled by an unjustified rectangular function and placed in a low-k region.

The values of ε_i are of certain importance because it is well known that the understated ε_i give too optimistic uncertainties of structural parameters. More realistic uncertainties are given by more realistic errors of EXAFS function extraction.

6. Conclusion

In this paper, we have considered one of the most important stages of XAFS function extraction from the measured absorption, construction of the atomic-like absorption μ_0.

For the widespread method of approximation of μ_0 by a smoothing spline we have proposed a way to take into account the prior information about μ_0. More specifically, one can use the energy dependence of the second derivative of μ_0, calculated or measured in gaseous phase. The usage of the prior information not only significantly corrects the signal from first coordination sphere but also raises the stability of the spline with respect to the variation of $k_{\min}$, the left limit of XAFS region. In addition, we have proposed a new criterion for determination of the smoothing regularization parameter.

A new method for approximation of μ_0 is proposed, the method of Bayesian smoothing which gives almost the same result as the smoothing spline method but is much slower. However, the Bayesian smoothing is the only method that gives the errors of approximation of the atomic-like background by an artificial smooth function. As we have shown, these errors are greater than the experimental noise determined from Fourier analysis, and the use of these errors instead of the noise in the definition of χ^2 statistics leads to more realistic uncertainties of structural parameters inferred in fitting procedures. In addition, these errors impose more severe restrictions on the EXAFS spectrum length than the noise.

Proposed in the paper, methods for background removal and other stages of XAFS analysis are realized in the freeware program VIPER [14].

Appendix. Bayesian smoothing and deconvolution

Here, we consider the general linear problem of data smoothing with the use of statistical methods (for an introduction, see the review by Turchin *et al* [15]). Let data $\boldsymbol{d}$ be defined on the mesh $x_1, \ldots, x_N$ and consist of the true values $\boldsymbol{t}$ and the additive noise $\boldsymbol{n}$:

$$d_i = t_i + n_i \qquad i = 1, \ldots, N. \qquad (A.1)$$

The problem of smoothing is to find the best estimates for $\boldsymbol{t}$. For an arbitrary node j, we find the probability density function for t_j given the data $\boldsymbol{d}$:

$$P(t_j|\boldsymbol{d}) = \int P(\boldsymbol{t}|\boldsymbol{d})\mathrm{d}\boldsymbol{t}_{\neq j} \qquad (A.2)$$

where $P(\boldsymbol{t}|\boldsymbol{d})$ is the joint probability density function for all values $\boldsymbol{t}$, and the integration is done over all $t_{i\neq j}$. According

to Bayes theorem,

$$P(\boldsymbol{t}|\boldsymbol{d}) = \frac{P(\boldsymbol{d}|\boldsymbol{t})P(\boldsymbol{t})}{P(\boldsymbol{d})} \tag{A.3}$$

$P(\boldsymbol{t})$ being the joint prior probability for all t_i, and $P(\boldsymbol{d})$ is a normalization constant. Assuming that the values n_i are independent in different nodes and normally distributed with zero expected values, the probability $P(\boldsymbol{d}|\boldsymbol{t})$, the so-called likelihood function, is given by

$$P(\boldsymbol{d}|\boldsymbol{t},\sigma) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(d_k - t_k)^2\right) \tag{A.4}$$

where the standard deviation of the noise, σ, appears as a known value. Later, we apply the rules of probability theory to remove σ from the problem.

Now we define prior probability $P(\boldsymbol{t})$. Let the function $t(x)$ be known in advance to be smooth enough. To specify this information, we introduce the norm of the second derivative and indicate its expected approximate value:

$$\Omega(t(x)) = \int \left(\frac{\mathrm{d}^2 t}{\mathrm{d}x^2}\right)^2 \mathrm{d}x \approx \omega \tag{A.5}$$

or, denoting $\Delta_i = x_{i+1} - x_i, i = 1, \ldots, N-1$ and representing the second derivative in the finite-difference form:

$$\begin{aligned}\Omega(t(x)) &\equiv \Omega(\boldsymbol{t}) \\ &= \sum_{i=2}^{N-1} \Delta_i^{-1}\left[t_{i-1}\Delta_{i-1}^{-1} - t_i\left(\Delta_{i-1}^{-1} + \Delta_i^{-1}\right) + t_{i+1}\Delta_i^{-1}\right]^2 \\ &\equiv \sum_{k,l=1}^{N} \Omega_{kl} t_k t_l.\end{aligned} \tag{A.6}$$

Ω_{kl} is a five-diagonal symmetric matrix with the following non-zero elements:

$$\begin{aligned}\Omega_{11} &= \Delta_1^{-2}\Delta_2^{-1} \\ \Omega_{22} &= \Delta_2^{-1}\left(\Delta_1^{-1} + \Delta_2^{-1}\right)^2 + \Delta_2^{-2}\Delta_3^{-1} \\ \Omega_{12} &= -\left(\Delta_1\Delta_2\right)^{-1}\left(\Delta_1^{-1} + \Delta_2^{-1}\right) \\ \Omega_{ii} &= \Delta_i^{-1}\left(\Delta_i^{-1} + \Delta_{i-1}^{-1}\right)^2 + \Delta_i^{-2}\Delta_{i+1}^{-1} + \Delta_{i-1}^{-3} \\ \Omega_{i-1,i} &= -\Delta_{i-1}^{-2}\left(\Delta_{i-1}^{-1} + \Delta_{i-2}^{-1}\right) - \left(\Delta_{i-1}\Delta_i\right)^{-1}\left(\Delta_{i-1}^{-1} + \Delta_i^{-1}\right) \\ \Omega_{i-2,i} &= \Delta_{i-2}^{-1}\Delta_{i-1}^{-2} \\ \Omega_{NN} &= \Delta_{N-1}^{-3} \\ \Omega_{N-1,N-1} &= \Delta_{N-1}^{-1}\left(\Delta_{N-1}^{-1} + \Delta_{N-2}^{-1}\right)^2 + \Delta_{N-2}^{-3} \\ \Omega_{N-1,N} &= -\Delta_{N-1}^{-2}\left(\Delta_{N-1}^{-1} + \Delta_{N-2}^{-1}\right).\end{aligned} \tag{A.7}$$

In order to introduce the minimum information in addition to that contained in (A.6), from all normalized to unity functions $P(\boldsymbol{t})$ which satisfy the condition (A.6) we choose a single one that contains minimum information about $\boldsymbol{t}$, i.e. minimizes the functional

$$\begin{aligned}I[P(\boldsymbol{t})] = &\int P(\boldsymbol{t}) \ln P(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} + \beta\left[1 - \int P(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}\right] \\ &+\gamma\left[\omega - \int \Omega(\boldsymbol{t})\,P(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}\right]\end{aligned} \tag{A.8}$$

where β and γ are the Lagrange multipliers. Minimizing $I[P(\boldsymbol{t})]$, one obtains the equation set

$$\begin{aligned}\ln P(\boldsymbol{t}) + 1 - \beta - \gamma\Omega(\boldsymbol{t}) &= 0 \\ \int P(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} &= 1 \\ \int \Omega(\boldsymbol{t})\,P(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} &= \omega\end{aligned} \tag{A.9}$$

which has a solution:

$$P(\boldsymbol{t}) = \left(\lambda_1 \cdots \lambda_N\right)^{-1/2}\left(\frac{2\pi\sigma^2}{\alpha}\right)^{-N/2} \exp\left(-\frac{\alpha}{2\sigma^2}\Omega(\boldsymbol{t})\right) \tag{A.10}$$

where $\alpha/2\sigma^2 = \gamma = N/2\omega$, and $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of the matrix Ω_{kl}. The regularizer α will be used to control the smoothness of $\boldsymbol{t}$. The prior distribution obtained is a 'soft' one, i.e. does not demand from the solution to have a strictly prescribed form.

Thus, we have for the posterior probability density function:

$$\begin{aligned}P(t_j|\boldsymbol{d},\sigma,\alpha) &\propto \int \mathrm{d}\boldsymbol{t}_{\neq j}\,\sigma^{-2N}\alpha^{N/2} \\ &\times \exp\left(-\frac{1}{2\sigma^2}\left[\boldsymbol{d}^2 - 2\sum_{k=1}^{N} d_k t_k + \sum_{k,l=1}^{N} g_{kl} t_k t_l\right]\right)\end{aligned} \tag{A.11}$$

where

$$g_{kl} = \alpha\Omega_{kl} + \delta_{kl} \qquad \boldsymbol{d}^2 = \sum_{k=1}^{N} d_k^2. \tag{A.12}$$

In most real problems, σ and α are not known. To eliminate σ is quite a straightforward problem:

$$P(t_j|\boldsymbol{d},\alpha) = \int \mathrm{d}\sigma\,P(t_j,\sigma|\boldsymbol{d},\alpha) = \int \mathrm{d}\sigma\,P(\sigma)P(t_j|\boldsymbol{d},\sigma,\alpha) \tag{A.13}$$

one needs only to know a prior probability $P(\sigma)$. Having no specific information about σ, a Jeffreys prior $P(\sigma) = 1/\sigma$ is assigned [16]. Then

$$\begin{aligned}\bar{t}_j = \int t_j P(t_j|\boldsymbol{d},\alpha)\mathrm{d}t_j &\propto \int \mathrm{d}\boldsymbol{t}\,\mathrm{d}\sigma\,\sigma^{-2N-1}t_j \\ &\times \exp\left(-\frac{1}{2\sigma^2}\left[\boldsymbol{d}^2 - 2\sum_{k=1}^{N} d_k t_k + \sum_{k,l=1}^{N} g_{kl} t_k t_l\right]\right).\end{aligned} \tag{A.14}$$

Performing the diagonalization, one obtains:

$$\bar{t}_j = \sum_{i=1}^{N} \frac{h_i e_{ij}}{\sqrt{\lambda_i'}} \tag{A.15}$$

and for the variance $\delta^2(t_j)$:

$$\delta^2(t_j) = \frac{\left[\boldsymbol{d}^2 - \boldsymbol{h}^2\right]}{N-2}\sum_{i=1}^{N}\frac{e_{ij}^2}{\lambda_i'} \tag{A.16}$$

where

$$\boldsymbol{h}^2 = \sum_{i=1}^{N} h_i^2 \qquad h_i = \frac{1}{\sqrt{\lambda_i'}}\sum_{k=1}^{N} d_k e_{ik} \tag{A.17}$$

and λ_i' and $\boldsymbol{e}_i$ are the eigenvalues and eigenvectors, respectively, of matrix g.

Thus, we have lost unknown σ and found the expressions for mean values t_j and their dispersions at known regularizer α. In modern Bayesian methods, α is itself determined by Bayesian arguments that maximize the posterior probability of α given the data:

$$P(\alpha|\boldsymbol{d}) = \int \mathrm{d}\boldsymbol{t}\,\mathrm{d}\sigma\, P(\alpha,\sigma,\boldsymbol{t}|\boldsymbol{d}) = \int \mathrm{d}\boldsymbol{t}\,\mathrm{d}\sigma\, P(\alpha,\sigma)\,P(\boldsymbol{t}|\alpha,\sigma,\boldsymbol{d}). \tag{A.18}$$

Assuming that α and σ are independent and using Bayes theorem (A.3), one obtains:

$$P(\alpha|\boldsymbol{d}) \propto \int \mathrm{d}\boldsymbol{t}\,\mathrm{d}\sigma\, P(\alpha)P(\sigma)P(\boldsymbol{t}|\alpha,\sigma)P(\boldsymbol{d}|\boldsymbol{t},\alpha,\sigma). \tag{A.19}$$

Substituting (A.10) for the prior probability $P(\boldsymbol{t}|\alpha,\sigma)$, (A.4) for the likelihood, and a Jeffreys prior $P(\sigma) = 1/\sigma$ and $P(\alpha) = 1/\alpha$, one obtains the posterior distribution for the regularizer α:

$$P(\alpha|\boldsymbol{d}) \propto \left(\lambda_1' \cdots \lambda_N'\right)^{-1/2} \alpha^{N/2-1} \left[\boldsymbol{d}^2 - \boldsymbol{h}^2\right]^{-N/2}. \tag{A.20}$$

Having found the maximum of the posterior probability (A.20), at this α one has the sought $\boldsymbol{t}$ with the most probable smoothness. However it is necessary to point out that this procedure narrows the applicability of the Bayesian smoothing down to the class of tasks where the smoothed values lie mostly within the limits $\pm\sigma$ from the most probable. In practice, there are other possible tasks where the condition (A.1) is treated more widely and the smoothed values exceed the bounds of noise.

Addenda to the Bayesian smoothing.

(i) Let the curvature of the function $t(x)$ be approximately known in advance. To specify this information, introduce the norm of the difference between $\mathrm{d}^2t/\mathrm{d}x^2$ and approximately known second derivative $\mathrm{d}^2f/\mathrm{d}x^2$:

$$\Omega(t(x)) = \int \left(\frac{\mathrm{d}^2t}{\mathrm{d}x^2} - \frac{\mathrm{d}^2f}{\mathrm{d}x^2}\right)^2 \mathrm{d}x \approx \omega. \tag{A.21}$$

Notice that there is no need to know $f(x)$ itself, its second derivative is sufficient. The explicit presence of $f(x)$ in the following formulae should be taken as a consequence of the technical trick applied: at first $f(x)$ is subtracted from the data, then it is added to the found solution.

Everywhere in formulae (A.6)–(A.20) we make the substitutions:

$$\tilde{t}_i = t_i - f_i \qquad \tilde{d}_i = d_i - f_i. \tag{A.22}$$

Performing the described above procedure for smoothing, one finds $\tilde{t}_i$, from which by inverse transformation the sought vector is given by $\boldsymbol{t} = \tilde{\boldsymbol{t}} + \boldsymbol{f}$.

(ii) In some tasks, the value on the starting (zero) node is known without measurement. This sort of prior information represents a 'hard' one, i.e. it restricts the class of possible solutions. In the given case, the solution must pass through the known zero node. The quadratic form $\Omega(\boldsymbol{t})$ (or $\Omega(\tilde{\boldsymbol{t}})$ in the case of approximately known second derivative) in the expression for the prior probability has now changed:

$$\Omega(\boldsymbol{t}) = \sum_{i=1}^{N-1} \left[t_{i-1}\Delta_{i-1}^{-1} - t_i\left(\Delta_{i-1}^{-1} + \Delta_i^{-1}\right) + t_{i+1}\Delta_i^{-1}\right]^2$$
$$\equiv \sum_{k,l=1}^{N} \Omega_{kl}t_kt_l + \Omega_{00}t_0^2 + 2\Omega_{01}t_0t_1 + 2\Omega_{02}t_0t_2. \tag{A.23}$$

The first matrix elements of Ω_{kl} now are:

$$\begin{aligned}
\Omega_{00} &= \Delta_0^{-2}\Delta_1^{-1} \\
\Omega_{01} &= -\left(\Delta_0\Delta_1\right)^{-1}\left(\Delta_0^{-1} + \Delta_1^{-1}\right) \\
\Omega_{02} &= \Delta_0^{-1}\Delta_1^{-2} \\
\Omega_{11} &= \Delta_1^{-1}\left(\Delta_0^{-1} + \Delta_1^{-1}\right)^2 + \Delta_1^{-1}\Delta_2^{-2} \\
\Omega_{12} &= -\Delta_1^{-2}\left(\Delta_1^{-1} + \Delta_0^{-1}\right) - \left(\Delta_1\Delta_2\right)^{-1}\left(\Delta_1^{-1} + \Delta_2^{-1}\right) \\
\Omega_{22} &= \Delta_2^{-1}\left(\Delta_2^{-1} + \Delta_1^{-1}\right)^2 + \Delta_2^{-2}\Delta_3^{-1} + \Delta_1^{-3}.
\end{aligned} \tag{A.24}$$

If $t_0 = 0$ (or $\tilde{t}_0 = 0$ if $\Omega(\tilde{\boldsymbol{t}})$ is used), no further changes to the formulae of smoothing are needed. At $t_0 \neq 0$, the changes are evident; instead of the scalar product $\boldsymbol{dt}$ in (A.11) will be $(\boldsymbol{d}-\hat{\boldsymbol{d}})\boldsymbol{t}$, where $\hat{d}_1 = \alpha t_0\Omega_{01}$, $\hat{d}_2 = \alpha t_0\Omega_{02}$, all remaining $\hat{d}_i = 0$; to the $\boldsymbol{d}^2$ the term $\alpha t_0^2\Omega_{00}$ will be added.

(iii) Making small changes in the above considered problem of smoothing allows one to solve the problem of deconvolution. If the experimental value d_j on some node j is determined not only by t_j but also by the values of some neighbouring nodes, then instead of (A.1) we have:

$$d_i = \sum_{j=1}^{N} r_{ji}t_j + n_i \qquad i = 1,\ldots,N \tag{A.25}$$

where r_{ji} is the grid representation of the impulse response function $\left(\sum_{j=1}^{N} r_{ji} = 1\right)$. Instead of expression (A.4), for the likelihood we now have:

$$P(\boldsymbol{d}|\boldsymbol{t},\sigma) = (2\pi\sigma^2)^{-N/2} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{N}\left[d_k - \sum_{i=1}^{N} r_{ik}t_i\right]^2\right) \tag{A.26}$$

and instead of (A.11), the posterior probability for t_j is now expressed as:

$$P(t_j|\boldsymbol{d},\sigma,\alpha) \propto \int \mathrm{d}\boldsymbol{t}_{\neq j}\,\sigma^{-2N}\alpha^{N/2} \times \exp\left(-\frac{1}{2\sigma^2}\left[\boldsymbol{d}^2 - 2\sum_{k=1}^{N} D_kt_k + \sum_{k,l=1}^{N} G_{kl}t_kt_l\right]\right) \tag{A.27}$$

where

$$G_{kl} = \alpha\Omega_{kl} + \sum_{i=1}^{N} r_{ki}r_{li} \qquad D_k = \sum_{i=1}^{N} r_{ki}d_i \tag{A.28}$$

and matrix G, in general, has more than five diagonals. Further steps for finding $\boldsymbol{t}$ are analogous to those described above.

Acknowledgments

The author is indebted to A P Menushenkov for his support, and to A V Kuznetsov and A A Ivanov for valuable comments. The work is supported in part by RFBR grant No. 99-02-17343.

References

[1] Lytle F W, Sayers D E and Stern E A 1975 *Phys. Rev.* B **11** 4825–35
[2] Ankudinov A L, Ravel B, Rehr J J and Conradson S D 1998 *Phys. Rev.* B **58** 7565–76
[3] Krappe H J and Rossner H H 2000 *Phys. Rev.* B **61** 6596–610
[4] Catalogue of XAFS Analysis Programs, http://ixs.csrri.iit.edu/catalog/XAFS_Programs
[5] Kuzmin A 1995 *Physica* B **208–209** 175–6
[6] Bridges F, Booth C H and Li G G 1995 *Physica* B **208–209** 121–4
[7] Boland J J, Halaka F G and Baldeschwieler J D 1983 *Phys. Rev.* B **28** 2921–6
[8] Stern E A, Līviņš P and Zhang Z 1991 *Phys. Rev.* B **43** 8850–60
[9] Newville M, Līviņš P, Yacoby Y, Rehr J J and Stern E A 1993 *Phys. Rev.* B **47** 14 126–31
[10] Hu T D, Xie Y N, Jin Y La and Liu T 1997 *J. Phys.: Condens. Matter* **9** 5507–15
[11] Di Cicco A, Filipponi A, Itié J P and Polian A 1996 *Phys. Rev.* B **54** 9086–98
[12] Cook J W Jr and Sayers D E 1981 *J. Appl. Phys.* **52** 5024–31
[13] Newville M, Boyanov B I and Sayers D E 1999 *J. Synchrotron Radiat.* **6** 264–5
[14] Klementev K V 2000 *VIPER for Windows (Visual Processing in EXAFS Researches)* freeware, http://www.crosswinds.net/∼klmn/viper.html
[15] Turchin V F, Kozlov V P and Malkevich M S 1971 *Sov. Phys. Usp.* **13** 681–840
[16] Jeffreys H 1939 *Theory of Probability* (London: Oxford University Press) later editions in 1948, 1961 and 1983