Error analysis of XAFS measurements

H. J. Krappe and H. H. Rossner
Hahn-Meitner-Institut Berlin, Glienicker Straße 100, D-14109 Berlin, Germany
(Received 9 July 1999)

We propose a method to analyze x-ray-absorption fine-structure data that avoids an arbitrary restriction of the size of the model-parameter space. It starts with an *a priori* guess of the model parameters which is introduced into the fitting procedure by Bayesian arguments. Two different descriptions are discussed to determine the relative impact of the *a priori* and experimental information on the fit. The resulting algorithms are tested by application to three simulated experiments at the Ta L_3-edge and to Cu *K*-edge data.

I. INTRODUCTION

Traditionally measured extended or near-edge x-ray-absorption fine-structure (EXAFS or NEXAFS) raw data are subject to various manipulations like Fourier filtering and background subtraction, often on the basis of the individual experience of the experimentalist. This makes it difficult to follow the error propagation from the measured input data to the output of the data analysis in a systematic, quantitatively reliable way. Furthermore, the effects of systematic errors resulting from shortcomings of the many-body solid-state theory used in EXAFS are difficult to quantify. We feel, therefore, that a more systematic investigation of the reliability of EXAFS data analyses would be worthwhile. This requires reliable estimates of both the errors of the XAFS measurements as well as those connected with the theoretical model on which the data analysis is based. Although both types of errors are not always known with the desirable reliability, it may be useful to show what could be gained by a more dependable error estimate of the input data than available at present, in order to motivate a careful analysis of the various sources of systematic errors in future work on XAFS data.

In the following we shall try to identify and quantify the various sources of statistical and systematic errors entering the data processing. Using well-defined statistical procedures the effect of these input errors on the resulting fit is determined. As in most high-precision data analyses, Bayesian arguments will be used. Special effort is made not to compromise the genuine information content in the data by *a priori* assumptions. The algorithm is tested by application to four sets of data: the copper *K*-edge data of Newville[1] and three sets of computer-generated tantalum L_3-edge data.

II. FORMULATION OF THE PROBLEM

The following discussion will be based on the standard formula for x-ray absorption on a polycrystalline or amorphous sample[2,3]

$$\frac{\mu(k)-\mu_{\rm back}(k)-\mu_0(k)}{\mu_0(k)} = \chi(k) = \frac{S_0^2}{k}\sum_i N_i \frac{|f_i(k,R_i)|}{R_i^2} e^{-2k^2\sigma_i^2-2R_i/\lambda(k)} \times \sin[2kR_i+\phi_i(k)] \quad (1)$$

with the wave number

$$k=\sqrt{\frac{2m}{\hbar^2}(h\nu-E_0)},$$

where E_0 refers to the Fermi energy reckoned from the muffin-tin potential zero. In Eq. (1) $\mu(k)$ is the measured absorption signal. The background absorption $\mu_{\rm back}(k)$ is assumed to contain all known systematic disturbing contributions to the absorption, like tails of nearby *K* and *L* edges of other atoms, signals from photons scattered in the monochromator or the target, and other errors to be determined in principle by appropriate reference measurements, and in practice often by the Victoreen procedure.[4] Remaining errors in the difference $\mu(k)-\mu_{\rm back}(k)$ are therefore supposed to have numerous, statistically independent errors of comparable size. By virtue of the central limit theorem they are lumped together in a Gaussian distribution with width $\Delta\mu(k)$. The absorption coefficient $\mu_0(k)$ of the absorbing atom, with no other atoms present in the lattice, is an artificial quantity to be obtained from theory. The same holds for the effective curved-wave backscattering amplitude $f_i(k,\mathbf{x})$ and the net phase shift $\phi_i(k,\mathbf{x})$, connected with the *i*th shell of neighboring atoms. The *N*-component vector $\mathbf{x}$ stands for the half-path lengths R_i and coordination numbers N_i, to be determined by the fit. Also the mean free path $\lambda(k)$ and the ''many-body'' amplitude reduction-factor S_0^2 are in principle obtained from an approximate treatment of the many-electron system surrounding the absorbing atom. Finally, the temperature-dependent and angular-averaged variances σ_i^2 of the atomic positions in the *i*th shell with respect to the absorbing atom can be calculated from the optical branch of the phonon spectrum or, if the latter is not available, approximately from the spring constants between neighboring pairs of atoms. It should be noted that all cumulants of the peaks in the pair-distribution function above the second one are neglected in Eq. (1).

The quantities $\mu_{\rm back}$, μ_0, $|f_i(R_i)|$, ϕ_i, λ, S_0^2, σ_i^2, and E_0 are represented by the *M* component vector $\mathbf{y}(k,\mathbf{x})$. Because of the practical difficulty to account exhaustively for all contributions to $\mu_{\rm back}$ and the limitations in the theory of the electronic many-body system in a solid, all components of $\mathbf{y}$ are subject to systematic errors. In order to achieve a quantitative error analysis, it is obviously necessary to quan-

tify one's confidence in evaluating these quantities. For this purpose we shall introduce probability distributions $P_j(\mathbf{y})dy_j$ to find the value y_j of the jth component of $\mathbf{y}$ to lie between y_j and y_j+dy_j when the other components have values $y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_M$. We shall not discuss here how these—in general non-normal—distributions can be obtained (for instance, by performing calculations in more reliable, but more expensive approximation schemes). Instead, the functions P_j will be treated as an input of Gaussian shape. It is often difficult enough to obtain an estimate for a width Δy_j in addition to a mean value $y_j^{(0)}$. Given these two numbers, a Gaussian is the most probable distribution function from the point of view of maximizing the information-theoretical entropy. For the sake of simplicity we assume that the errors for different components of $\mathbf{y}$ are statistically independent. Therefore,

$$P(\mathbf{y})d^M\mathbf{y}\approx\prod_{j=1}^{M} e^{-1/2(y_j-y_j^{(0)})^2/\Delta y_j^2}d^M\mathbf{y}. \tag{2}$$

In the following it is convenient to rewrite Eq. (1) as

$$\mu(k)=g(k;\mathbf{x},\mathbf{y}(k,\mathbf{x})),$$

the ''model'' being defined by the function g. The observable $\mu(k)$ is assumed to be measured at the L values $k_1,\ldots,k_L$ of its argument. We order the terms in the sum of Eq. (1) according to decreasing average amplitude and truncate the sum after the Ith term. Calling the truncated sum $\boldsymbol{\chi}_I(k)$, the model function

$$g(k)=\mu_{\rm back}(k)+\mu_0(k)+\mu_0(k)\boldsymbol{\chi}_I(k),$$

therefore, contains a truncation error. It is useful to consider the differences $X_i(k_l)=\boldsymbol{\chi}_{I+i}(k_l)-\boldsymbol{\chi}_I(k_l)$ for $i=1,\ldots,I_{\max}$. If $I_{\max}$ is chosen sufficiently large to accumulate reasonable statistics, the numbers $X_i(k_l)$ at each fixed k_l turn out to have roughly a Gaussian distribution. We may then define average values

$$\overline{X(k_l)}=\frac{1}{I_{\max}}\sum_{i=1}^{I_{\max}}X_i(k_l)$$

and a correlation matrix

$$\Sigma_{ll'}=\overline{X(k_l)X(k_{l'})}-\overline{X(k_l)}\cdot\overline{X(k_{l'})}.$$

We found for all examples investigated so far that

$$\Sigma_{ll}>\overline{X(k_l)}^2,\quad l=1,\ldots,L.$$

We, therefore, model the truncation error, i.e., the distribution of the stochastic vector $\mathbf{g}'$ with components

$$g'(k_l)=g(k_l)+\mu_0X(k_l),$$

by the distribution

$$P_{\rm model}(\mathbf{g}')\approx e^{-(1/2)\chi^2_{\rm model}(\mathbf{g},\mathbf{g}')};$$

here, in matrix notation,

$$\chi^2_{\rm model}=(\mathbf{g}'-\mathbf{g})^TB'(\mathbf{g}'-\mathbf{g})$$

in terms of a symmetric matrix B' whose inverse is given by

$$(B'^{-1})_{ll'}=\mu_0(k_l)\mu_0(k_{l'})\Sigma_{ll'}.$$

The total uncertainty due to systematic errors is then

$$P_{\rm syst}(\mathbf{g}',\mathbf{y})=P_{\rm model}(\mathbf{g}')\prod_{j=1}^{M}P_j(\mathbf{y}).$$

It may be desirable to determine some of the components of $\mathbf{y}$ by fitting the data, rather than relying on model calculations. In this case they have to be added to the vector $\mathbf{x}$, thus increasing the dimension of the space of model parameters. In general it cannot be expected that a given set of measured data allows the determination of vectors $\mathbf{x}$ with too many components. However, it is not easy to know in advance what the maximal admissible number of components of $\mathbf{x}$ is that can reliably be determined by a given set of data. In fact, due to cross correlations between the errors in the components of $\mathbf{x}$, one expects that the data determine the fit only in some subspace of the whole parameter space, the complementary subspace being ''drowned'' in the combined noise of the data and the systematic error of $\mathbf{y}$. One, therefore, needs a general algorithm to determine these subspaces.

III. EXTENDED χ^2 FIT

It is inconceivable that a precision measurement is done on a sample whose structure is totally unknown. In a typical case some estimate of the components of $\mathbf{x}$ is known *a priori* although, perhaps, with large uncertainty.[5] Calling this estimate $\mathbf{x}^{(0)}$, the aim of the fit procedure is to find the probability distribution for the relative differences $(x_n-x_n^{(0)})/\Delta x_n$, $n=1,\ldots,N$, on the basis of the information gained by the measurement. The order of magnitude of the range over which one expects the x_n to differ from $x_n^{(0)}$ is called Δx_n. To simplify the notation the symbol $\mathbf{x}$ shall in the following be used to represent these relative differences. Just as one knows *a priori* an estimate $\mathbf{x}^{(0)}$, one may model one's uncertainty of this guess by a Gaussian probability distribution

$$P_{\rm prior}(\mathbf{x})=N_{\rm prior}e^{-(1/2)\chi^2_{\rm prior}(\mathbf{x})} \tag{3}$$

with the normalization factor $N_{\rm prior}$ and the quadratic form

$$\chi^2_{\rm prior}(\mathbf{x})=\sum_{n,n'=1}^{N}A_{nn'}x_nx_{n'}. \tag{4}$$

This ansatz can also be obtained from the requirement that $P_{\rm prior}$ should maximize the information-theoretical entropy $\int P_{\rm prior}(\mathbf{x})\ln P_{\rm prior}(\mathbf{x})dm(\mathbf{x})$ under the constraint that

$$\langle\mathbf{x}\rangle_{\rm prior}=0\quad\text{and}\quad\langle x_nx_{n'}\rangle_{\rm prior}=A^{-1}_{nn'},$$

and that the probability measure in $\mathbf{x}$ space is given by

$$dm(\mathbf{x})=\prod_{n=1}^{N}dx_n.$$

Note that by itself the maximum-entropy principle neither yields $\mathbf{x}^{(0)}$, nor the metrical tensor $A_{nn'}$ or the metric $m(\mathbf{x})$. Conditions for a reasonable choice of the quadratic form $A_{nn'}$ shall be discussed in the next section.

There are often good reasons to modify the ansatz Eq. (3). There exist, e.g., lower and upper limits for the distances R_i beyond which no result of a fit would be acceptable, no matter how well it fits the data. Therefore, the *a priori* probability Eq. (3) may be substituted by the cutoff Gaussian

$$P_{\text{prior}}=N'_{\text{prior}}\prod_{n=1}^{N}\Theta(x_n-x_n^l)\Theta(x_n^u-x_n)e^{-(1/2)\chi^2_{\text{prior}}}, \quad (5)$$

where Θ is the step function, N'_{prior} is a new normalization factor and

$$x_n^l\leqslant x_n\leqslant x_n^u, \quad n=1,\ldots,N,$$

defines the permissible range of model parameters.

After the measurement, experimental values $\mu_l=\mu(k_l)$, $l=1,\ldots,L$ are known, with normal-distributed errors $\Delta\mu'_l$ at wave vectors k_l, determined with errors Δk_l, which we assume to be also normal distributed and statistically independent. Then the conditional probability to find $\boldsymbol{\mu}$ once $\mathbf{g}'$ is given is

$$P(\boldsymbol{\mu}|\boldsymbol{g}')=\prod_{l=1}^{L}(2\pi\Delta\mu_l^2)^{-1/2}e^{-(1/2)\chi^2_{\text{exp}}}$$

with

$$\chi^2_{\text{exp}}=\sum_{l=1}^{L}\left[\frac{\mu_l-g'(k_l)}{\Delta\mu_l}\right]^2$$

and

$$\Delta\mu_l^2=\Delta\mu_l'^2+\left(\frac{dg'}{dk}\bigg|_{k=k_l}\Delta k_l\right)^2. \quad (6)$$

To account also for the systematic errors, this function must be weighted with P_{syst}

$$P_{\text{cond}}(\boldsymbol{\mu}|\boldsymbol{x})=\int P_{\text{syst}}(\mathbf{g}',\mathbf{y};\mathbf{g}(\mathbf{x},\mathbf{y}))P(\boldsymbol{\mu}|\boldsymbol{g}')d^M\mathbf{y}d^L\mathbf{g}'. \quad (7)$$

To perform the integrations in Eq. (7) it is convenient to introduce the rescaled vectors $\boldsymbol{\mu}$, $\mathbf{g}$, and $\mathbf{g}'$ with components $\mu_l/\Delta\mu_l$, $g(k_l;\mathbf{x},\mathbf{y}(k_l,\mathbf{x}))/\Delta\mu_l$, and $g'(k_l)/\Delta\mu_l$, respectively, and the matrix B^{-1} with components

$$(B^{-1})_{ll'}=\frac{(B'^{-1})_{ll'}}{\Delta\mu(k_l)\Delta\mu(k_{l'})}=\frac{\mu_0(k_l)}{\Delta\mu(k_l)}\frac{\mu_0(k_{l'})}{\Delta\mu(k_{l'})}\Sigma_{ll'}.$$

In terms of these rescaled quantities we have

$$\chi^2_{\text{exp}}=(\boldsymbol{\mu}-\mathbf{g}')^T(\boldsymbol{\mu}-\mathbf{g}')$$

and

$$\chi^2_{\text{model}}+\chi^2_{\text{exp}}=\mathbf{g}'^T(1+B)\mathbf{g}'-2(\boldsymbol{\mu}^T+\mathbf{g}^TB)\mathbf{g}'+\mathbf{g}^TB\mathbf{g}+\boldsymbol{\mu}^T\boldsymbol{\mu}.$$

Using the formula[6]

$$\int e^{-(1/2)\mathbf{x}^TA\mathbf{x}+\mathbf{b}^T\mathbf{x}}d^L\mathbf{x}=\left[\frac{(2\pi)^L}{\det(A)}\right]^{1/2}e^{(1/2)\mathbf{b}^TA^{-1}\mathbf{b}}, \quad (8)$$

which is valid for any symmetric $L\times L$ matrix A possessing an inverse A^{-1}, and L-dimensional vectors $\mathbf{x}$ and $\mathbf{b}$, the $\mathbf{g}'$ integration in Eq. (7) yields

$$P_{\text{cond}}(\boldsymbol{\mu}|\boldsymbol{x})=\int e^{-(1/2)\chi^2_{\text{inter}}-1/2(\mathbf{y}-\mathbf{y}^{(0)})^TD(\mathbf{y}-\mathbf{y}^{(0)})}d^M\mathbf{y} \quad (9)$$

in terms of the diagonal matrix D with components $D_{jj'}=\Delta y_j^{-2}\delta_{jj'}$ $(j=1,\ldots,M)$ and

$$\boldsymbol{\chi}^2_{\text{inter}}=-(\boldsymbol{\mu}^T+\mathbf{g}^TB)(1+B)^{-1}(\boldsymbol{\mu}+B\mathbf{g})+\mathbf{g}^TB\mathbf{g}+\boldsymbol{\mu}^T\boldsymbol{\mu}. \quad (10)$$

Here we have assumed Gaussian distributions for the P_j and disregarded the cutoff factors in Eq. (5). Keeping only terms up to linear order in B^{-1} gives

$$\boldsymbol{\chi}^2_{\text{inter}}=\mathbf{g}^T(1-B^{-1})\mathbf{g}-2\boldsymbol{\mu}^T(1-B^{-1})\mathbf{g}+\boldsymbol{\mu}^T(1-B^{-1})\boldsymbol{\mu}.$$

According to Bayes' theorem the *a posteriori* probability to find $\mathbf{x}$ is given by[5,7]

$$P_{\text{post}}=\frac{P_{\text{prior}}(\mathbf{x})P_{\text{cond}}(\boldsymbol{\mu}|\mathbf{x})}{\int P_{\text{prior}}(\mathbf{x})P_{\text{cond}}(\boldsymbol{\mu}|\mathbf{x})d^N\mathbf{x}}. \quad (11)$$

This theorem expresses the modification of the *a priori* expectation of the model parameters $\mathbf{x}$ as a result of the experiment.

Traditionally one first determines μ_{back} by fitting the preedge spectrum to a Victoreen function, then one determines μ_0 by passing a smoothing spline through the data above the edge, finally one fixes E_0, thus constructing the EXAFS signal $\boldsymbol{\chi}(k)$ which is then Fourier transformed to obtain the R_j and N_j from peak positions and zero moments of the peaks, respectively. In contrast, we have proposed above to calculate the various fit parameters in Eq. (1) in *one* step. This is done with the purpose of keeping control of the correlations between their errors.

Our approach is basically a k-space fit. It may be useful to compare it with R-space fits which are frequently used in various variants to analyze XAFS data. If $g(k)$ is measured at $L=2L_k+1$ equidistant values k_l in the interval $k_{\text{min}}\leqslant k_l\leqslant k_{\text{max}}$ one may represent $k\boldsymbol{\chi}(k)$ in this interval by the Fourier series

$$(k'+k_0)\boldsymbol{\chi}(k'+k_0)=\sum_{\nu=-\infty}^{\infty}c_\nu e^{i\nu\delta rk'}$$

with $c_\nu=c^*_{-\nu}$, the midpoint $k_0=\frac{1}{2}(k_{\text{max}}+k_{\text{min}})$ and $k=k'+k_0$, where $\delta r=2\pi/(k_{\text{max}}-k_{\text{min}})$ is the optical resolution. The expansion coefficients c_ν are given by

$$c_\nu=\frac{\delta r}{2\pi}\int_{-\pi/\delta r}^{\pi/\delta r}(k'+k_0)\boldsymbol{\chi}(k'+k_0)e^{-i\nu\delta rk'}dk'$$

$$\approx\frac{1}{2L_k+1}\sum_{l=-L_k}^{L_k}k_l\boldsymbol{\chi}(k_l)e^{-i\nu\delta r(k_l-k_0)}, \quad \nu\geqslant 0. \quad (12)$$

The c_ν are a discrete representation of the EXAFS signal in r space, $F[k\boldsymbol{\chi}(k);r]$, with $r=\nu\delta r$, $\nu\geqslant 0$.

Rewriting Eq. (1) as

$$k\chi(k)=\sum_j A_j \sin(2kR_j+\phi_j)$$

and neglecting the k dependence of the amplitudes A_j and phases ϕ_j, insertion into Eq. (12) yields

$$c_\nu=\sum_j A_j e^{i(2k_0R_j+\phi_j-\pi/2)}\frac{\sin[(2R_j-\nu\delta r)\pi/\delta r]}{2R_j-\nu\delta r}.$$

The last factor on the right-hand side (rhs) of this equation is strongly peaked at $\nu=2R_j/\delta r$ if $R_j(k_{\max}-k_{\min})\gg 1$. Therefore, $|c_\nu|^2$ peaks strongly at these ν values, which allows one to obtain the R_j (and A_j^2) from the c_ν. But since the phases and amplitudes are actually k dependent, and $\mu_0(k)$ may have Fourier components in the EXAFS regime, such an analysis can only yield a first estimate of the R_j. A set of empirical rules[8] was introduced to correct the number, position, and width of the peaks of the Fourier-transformed EXAFS signal for these k dependences. A systematic analysis of error propagation under these rules would require still more empirical rules. One may therefore wonder whether the Fourier technique is really the adequate tool for a high-precision analysis of EXAFS data. Furthermore, one is often interested only in a couple of numbers R_j and integer coordination numbers N_j, rather than in the whole continuous function of $rF[k\chi(k);r]$. To extract this information out of the signal in k space more specific ways than a (truncated) Fourier transform appear to be preferable.

IV. LINEAR MODELING

The simplest case arises when g is a linear function of $\mathbf{x}$, i.e., when the initial guess $\mathbf{x}^{(0)}$ is sufficiently close to the final solution to justify the linear expansion

$$g(k;\mathbf{x},\mathbf{y})=g(k;\mathbf{0},\mathbf{y}(k,0))+\sum_{n=1}^{N} g_n(k,\mathbf{y})x_n, \qquad (13)$$

with

$$g_n(k;\mathbf{y})=\left.\frac{\partial g}{\partial x_n}\right|_{\mathbf{x}=0}+\sum_{j=1}^{M}\left.\frac{\partial g}{\partial y_j}\right|_{\mathbf{y}=\mathbf{y}(k,0);\ \mathbf{x}=0}\left.\frac{\partial y_j(k,\mathbf{x})}{\partial x_n}\right|_{\mathbf{x}=0}.$$

In this approximation $\chi^2_{\exp}$ becomes a quadratic form of $\mathbf{x}$.

To simplify the situation further we shall assume that our ability to determine the components of $\mathbf{y}$ from many-body theory is sufficiently advanced that the systematic errors Δy_j are small and allow an expansion of $g(k;\mathbf{x},\mathbf{y})$ to linear order in $\mathbf{y}-\mathbf{y}^{(0)}$,

$$g(k;\mathbf{x},\mathbf{y})=g(k;\mathbf{x},\mathbf{y}^{(0)})+\sum_{j=1}^{M}\left.\frac{\partial g}{\partial y_j}\right|_{\mathbf{y}=\mathbf{y}^{(0)}}(y_j-y_j^{(0)}). \qquad (14)$$

In terms of the rectangular matrices

$$G_{ln}=g_n(k_l)/\Delta\mu_l,\quad n=1,\ldots,N,\quad l=1,\ldots,L,$$

$$T_{lj}=\frac{1}{\Delta\mu_l}\left.\frac{\partial g(k_l)}{\partial y_j}\right|_{\mathbf{y}=\mathbf{y}^{(0)},\mathbf{x}=0},\quad j=1,\ldots,M,\quad l=1,\ldots,L,$$

and the vector $\mathbf{g_0}=\mathbf{g}(\mathbf{0},\mathbf{y}^{(0)})$ we have, again, in matrix notation

$$\mathbf{g}(\mathbf{x},\mathbf{y})=\mathbf{g_0}+G\mathbf{x}+T(\mathbf{y}-\mathbf{y}^{(0)}).$$

This expansion is substituted into Eq. (10) and the $\mathbf{y}$ integration in Eq. (9) is performed using again Eq. (8). The conditional probability is then Gaussian in $\mathbf{x}$, $P_{\rm cond}\propto\exp(-1/2\chi^2_{\rm cond})$ with

$$\chi^2_{\rm cond}=\mathbf{x}^TQ\mathbf{x}-2\mathbf{b}^T\mathbf{x}+(\boldsymbol{\mu}-\mathbf{g}_0)^TC(\boldsymbol{\mu}-\mathbf{g}_0)$$

in terms of the information matrix Q, the vector $\mathbf{b}^T$, and the $L\times L$ matrix C,

$$Q=G^TCG, \qquad (15a)$$

$$\mathbf{b}^T=(\boldsymbol{\mu}-\mathbf{g}_0)^TCG, \qquad (15b)$$

$$C=\{1-(1+B^{-1})^{-1}T[D+T^T(1+B^{-1})^{-1}T]^{-1}T^T\}\times(1+B^{-1})^{-1}. \qquad (15c)$$

If the systematic errors described by the matrices B^{-1} and D^{-1} are small compared to the experimental errors $\Delta\mu_l$, the brackets in Eq. (15c) may be expanded in powers of these matrices. To linear order in B^{-1} and D^{-1} one obtains $C=1-B^{-1}-TD^{-1}T^T$ or to the same order

$$C=(1+B^{-1}+TD^{-1}T^T)^{-1}.$$

If, in addition, one neglects the off-diagonal matrix elements of $C_{ll'}$, one arrives at the rather simple expression

$$\chi^2_{\rm cond}=\sum_{l=1}^{L}\left[\frac{\mu_l-g(k_l;\mathbf{x},\mathbf{y}^{(0)})}{\Delta\mu_l^{\rm eff}}\right]^2 \qquad (16)$$

with

$$(\Delta\mu_l^{\rm eff})^2=(\Delta\mu_l)^2+\sum_{j=1}^{M}\left[\frac{\partial g(k_l;\mathbf{0},\mathbf{y}^{(0)})}{\partial y_j}\Delta y_j\right]^2+\mu_0(k_l)^2\Sigma_{ll},$$

which generalizes the usual expression for $\chi^2_{\exp}$ to the case of a model with intrinsic uncertainties.

The *a posteriori* distribution Eq. (11) becomes $P_{\rm post}\propto\exp(-\frac{1}{2}\chi^2_{\rm post})$ with

$$\chi^2_{\rm post}=\chi^2_{\rm prior}+\chi^2_{\rm cond}=\mathbf{x}^T(A+Q)\mathbf{x}-2\mathbf{b}^T\mathbf{x}+(\boldsymbol{\mu}-\mathbf{g}_0)^TC\times(\boldsymbol{\mu}-\mathbf{g}_0). \qquad (17)$$

The *a posteriori* expectation value of the model parameters,

$$\overline{\mathbf{x}}:=\langle\mathbf{x}\rangle_{\rm post}=\int\mathbf{x}P_{\rm post}(\mathbf{x})d^N\mathbf{x},$$

follows from the normal equations

$$\frac{\partial\chi^2_{\rm post}}{\partial x_n}=0,\quad n=1,\ldots,N. \qquad (18)$$

More explicitly, one obtains the set of linear equations

$$\sum_{n'=1}^{N}(Q_{nn'}+A_{nn'})\overline{x_{n'}}=b_n, \qquad (19)$$

and the *a posteriori* variance matrix is given by

$$\langle (x_n-\overline{x_n})(x_{n'}-\overline{x_{n'}})\rangle_{\text{post}}=(Q_{nn'}+A_{nn'})^{-1}. \tag{20}$$

The interpretation of Eq. (19) is particularly transparent when $A_{nn'}$ commutes with $Q_{nn'}$, which, in particular, is the case when $A_{nn'}$ is proportional to the unit matrix. This is in fact the choice most frequently made. Denoting the eigenvalues of Q and A by q_n and a_n, respectively, and the components of $\mathbf{x}$ and $\mathbf{b}$ in the eigenrepresentation of Q by $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{b}}$, respectively, the linear system of equations (19) decouples in this representation,

$$(q_n+a_n)\tilde{x}_n=\tilde{b}_n, \quad n=1,\ldots,N. \tag{21}$$

The eigenvalues q_n may be ordered in decreasing size, $|q_n|\geqslant|q_{n+1}|$. The condition number of the matrix Q, q_1/q_N, increases with the size N of the model-parameter space and reaches values between 10^3 and 10^{15} in the examples to be discussed later. Since the model parameters $\tilde{x}_n$ are finite, the $\tilde{b}_n$ also have to decrease with increasing n, roughly in parallel with q_n. However, the $\tilde{b}_n$ are subject to ''noise,'' because they depend through $\mathbf{b}$ on the observables μ_l. Since there is no reason why the noise level should also decrease, the solution of Eq. (21) becomes meaningless for sufficiently large n unless the eigenvalues a_n prevent the sum q_n+a_n from becoming too small. The matrix A therefore has to be bounded from below to serve this purpose. The precise value of the bound shall be determined below. At this point we only want to point out that the *a priori* probability is seen to introduce the matrix A into Eq. (21), which formally regularizes the ill-posed inversion problem $\mathbf{Q}\overline{\mathbf{x}}=\mathbf{b}$ in the sense of Tikhonov.[9]

Those directions in model-parameter space for which $q_n \geqslant a_n$ are defined to span the subspace $\mathcal{R}$. Its orthogonal complement is called $\mathcal{P}$. So the data dominate the fit in $\mathcal{R}$, and the *a priori* assumptions are dominant in $\mathcal{P}$. The simplest case arises when the eigenvalues of Q decrease in size by an order of magnitude going from one eigenvalue to the next in the relevant range of eigenvalues defining the boundaries between $\mathcal{R}$ and $\mathcal{P}$ spaces,

$$q_n \gg a_n \quad \text{in } \mathcal{R}, \tag{22a}$$

$$q_n \ll a_n \quad \text{in } \mathcal{P}. \tag{22b}$$

This is frequently observed in fit problems involving model spaces of dimension larger than ten.[10] Then Eqs. (22) can be satisfied by the choice $A_{nn'}=\alpha\delta_{nn'}$. The overall weight factor α controls the impact of the *a priori* information on the fit compared to the experimental information.

Two methods will be discussed in the following to determine α. Both were proposed by Turchin[6,11] with the intention not to compromise the information in the data by too strong *a priori* conditions. In the first method the parameter α is determined in the fitting procedure by imposing Turchin's condition[11]

$$\langle \chi^2_{\text{cond}}\rangle_{\text{post}}=L. \tag{23}$$

This equation is to be added to the linear equations (19) to determine the N model parameters x_n and α simultaneously. Note that this system of equations is nonlinear since Eq. (23) is nonlinear. It is to be solved by an appropriate iteration scheme.

The meaning of Eq. (23) becomes more transparent by noticing that from

$$\langle \chi^2_{\text{cond}}(\mathbf{x})\rangle_{\text{post}}-\chi^2_{\text{cond}}(\overline{\mathbf{x}})=\sum_{n,n'} Q_{n,n'}(\langle x_n x_{n'}\rangle-\overline{x_n}\,\overline{x_{n'}}),$$

one obtains with Eqs. (20) and (23)

$$\chi^2_{\text{cond}}(\overline{\mathbf{x}})=L-\text{tr}\frac{Q}{Q+A}. \tag{24}$$

If $[Q,A]=0$, the rhs of Eq. (24) can be written in the form

$$L_{\text{eff}}=L-\sum_{n=1}^{N}\frac{q_n}{q_n+\alpha}.$$

Instead of Eq. (23) the more convenient expression Eq. (24) will be used together with Eq. (19) to determine the solution $\mathbf{x}^*=\overline{\mathbf{x}}$ and $\alpha^*=\alpha$. In $\mathcal{R}$ space the terms in the sum are approximately equal to unity, and they are approximately zero in $\mathcal{P}$. The trace therefore measures roughly the dimension of $\mathcal{R}$. The expression on the rhs of Eq. (24) thus extends the concept of the number of degrees of freedom $L-N$ in well-posed inversion problems to the present case of an ill-posed problem. We, therefore, call this expression L_{eff}, the effective number of degrees of freedom.

There is an alternative interpretation of Turchin's condition which is also valid when Q and A do not commute. All points $\mathbf{x}$ in model-parameter space for which $\chi^2_{\text{cond}}(\mathbf{x})\leqslant L_{\text{eff}}$ form an ellipsoid. All of them are compatible with the data within one standard deviation and must therefore be considered as indistinguishable on the basis of the measurement. If the initial guess $\mathbf{x}^{(0)}$ does not happen to lie inside this ellipsoid, it is easy to show that Eqs. (19) and (23) determine that point $\mathbf{x}^*$ on the surface of the ellipsoid which is closest to the origin, i.e., to the *a priori* assumption, distances being measured with the metric $A_{nn'}$.[12,13] In fact, minimization of the distance $\Sigma_{nn'}A_{nn'}x_n x_{n'}$ with the constraint Eq. (24) may be written as

$$\partial_{x_n}[\chi^2_{\text{prior}}(\mathbf{x})+\Lambda\chi^2_{\text{cond}}(\mathbf{x})]=0,$$

where Λ is a Lagrange multiplier. In view of Eqs. (17), (18), and (19), $\mathbf{x}=\overline{\mathbf{x}}$ is seen to solve this equation with $\Lambda=1$, and Eq. (24) is satisfied for $\overline{\mathbf{x}}=\mathbf{x}^*$. The determination of α by Turchin's condition therefore ensures that the information contained in the data is not distorted by *a priori* assumptions since $\mathbf{x}^*$ still belongs to the one-standard-deviation ellipsoid. But of all the points of the ellipsoid the one closest to the *a priori* guess is selected.

Our procedure is closely related to a treatment of the χ^2 fit by the singular-value-decomposition method with a substitution of the small singular values of the matrix Q by infinitely large ones.[14] This amounts to an exact projection of the model space onto the space $\mathcal{R}$. What might appear there as a mathematical trick is seen in our stochastic interpretation of the regularization to lead to a straightforward deter-

mination of the cutoff parameter α^*, independent of the numerical accuracy of the inversion algorithm used in the calculation.

When the sequence of eigenvalues of Q does not decrease as fast as assumed above, it is not desirable to define A essentially as a projector onto $\mathcal{R}$ space. In fact, the boundary between $\mathcal{R}$ and $\mathcal{P}$ spaces is then somewhat blurred and one needs a smoother onset of the regularizing effect of A. Instead of formal recipes to construct A it is advisable in this case to incorporate as much intuitive understanding of the physical system under consideration as possible in constructing the *a priori* error correlation-matrix $A_{nn'}$ in model parameter space. One should then normalize A by extracting the largest eigenvalue α as a factor. Writing

$$A = \alpha A', \tag{25}$$

the largest eigenvalue of A' is defined to be unity. Once one has assumed some reasonable form for A' the overall weight factor α can again be determined from Eq. (24). This prevents the information in the experimental data to be seriously distorted. Only the details of the smooth transition from $\mathcal{R}$ to $\mathcal{P}$ spaces depend on the choice of A'.

It is important to stress that the *a posteriori* probability does not imply a statement about the closeness of the solution $\mathbf{x}^*$ to the true solution. Instead it is the probability that the *a priori* model parameters $\mathbf{x}^{(0)}$ have to be modified as a consequence of the information gained by the experimental data. If, for example, $\mathbf{x}^{(0)}$ would be so close to the true solution that $\chi^2_{\mathrm{cond}}(0) = L$, α^* could be chosen infinitely large and Eqs. (19) and (24) could be satisfied with $\mathbf{x}^* = 0$, i.e., without changing the *a priori* model parameters at all. In this case the variance matrix $(Q+\alpha^*)^{-1}$ becomes infinitely small. This does not mean that the solution $\mathbf{x}^* = 0$ is particularly accurate, but that there is no need to modify any component of the *a priori* vector $\mathbf{x}^{(0)}$ as a consequence of the information gained by the experimental data.

It is useful to compare this concept of an *a posteriori* error matrix with the error matrix obtained in traditional χ^2 fits. In such a fit one decides *a priori* which of the components of $\mathbf{x}$ one considers to be determined completely by the data, e.g., besides S_0^2 and E_0 the first three half-path lengths R_1, R_2, and R_3 and the corresponding Debye-Waller parameters σ_1^2, σ_2^2, and σ_3^2, resulting in the eight-dimensional space $\mathcal{R}$ spanned by these parameters. The rest of the model parameters are treated as undetermined by the data. In our terminology this amounts to a choice of the regularization matrix $A_{nn'}$ as a projector onto the space $\mathcal{R}$ of the well-determined model parameters. In general, the information matrix Q does not commute with such a projector. The projection of the *a posteriori* error matrix $(Q+A)^{-1}$ onto the space $\mathcal{R}$ would then be the error matrix in the traditional χ^2 fit. Uncertainties of the model, which are included in our definition of Q, are not considered in the standard fit.

The regularization parameter α^* which satisfies Eq. (23) represents the strongest regularization still compatible with the data. As we shall see, there are situations in which one would like to decrease the effect of the *a priori* choice of the model parameters $\mathbf{x}^{(0)}$ on the fit. For such situations Turchin suggested[6] to use instead of α^* the maximum of the following function of α:

$$P(\alpha|\mathbf{g}) = \mathrm{const}\cdot \alpha^{N/2} e^{1/2 \sum_{nn'} b_n (Q+\alpha I)^{-1}_{nn'} b_{n'}} [\det(Q+\alpha I)]^{-1/2}. \tag{26}$$

It represents an estimate of the most probable value of the regularization parameter α inferred from the data in a Bayesian sense. The zero of the logarithmic derivative of Eq. (26) with respect to α is given by

$$\partial_\alpha \ln P = \frac{N}{2\alpha} - \frac{1}{2}|\mathbf{x}(\alpha)|^2 - \frac{1}{2}\mathrm{tr}(Q+\alpha I)^{-1} = 0. \tag{27}$$

We shall call the solution of this equation α'^*. The meaning of the function Eq. (26) implies that $\alpha'^* \leqslant \alpha^*$.

V. NONLINEAR INVERSION PROBLEMS

If the *a posteriori* errors in $\mathbf{x}$ and $\mathbf{y}$ are not so large that they lead out of the range of validity of the linear expansions (13) and (14), and if the higher cumulants in the distributions P_{prior} and P_j are negligible in the vicinity of the solution $\mathbf{x}^*$, the initial guess $\mathbf{x}^{(0)}$ may still be outside the range of validity of Eq. (13). In this case the guess $\mathbf{x}^{(0)}$ has first to be improved before the formalism of the last section can be applied. The most convenient, but not necessarily convergent, method for this purpose is Newton's algorithm. Since the normal equations in each iteration step are likely to be ill-conditioned, a regularization of the matrix inversion at each step is necessary. Again the matrix $A = \alpha A'$ can be used for this purpose with a parameter α typically several orders of magnitude larger than the optimal α satisfying Turchin's condition Eq. (24).[12] In the resulting iteration scheme the solution vector $\mathbf{x}^{(\nu+1)}$ is obtained from the preceding solution $\mathbf{x}^{(\nu)}$ by solving the regularized linear equations

$$\sum_{n'=1}^{N} [Q^{(\nu)}_{nn'}(x^{(\nu+1)} - x^{(\nu)})_{n'} + A^{(\nu)}_{nn'} x^{(\nu+1)}_{n'}] = b^{(\nu)}_n, \quad \mathbf{x}^{(0)} = 0. \tag{28}$$

The matrix $Q^{(\nu)}$ and the vector $\mathbf{b}^{(\nu)}$ are given by Eqs. (15), in terms of the vector $\mathbf{g}^{(\nu)}$ with components

$$g^{(\nu)}_l = \frac{g[k_l; \mathbf{x}^{(\nu)}, \mathbf{y}(k_l, \mathbf{x}^{(\nu)})]}{\Delta\mu_l}$$

instead of $\mathbf{g}_0$ and the rectangular matrices

$$T^{(\nu)}_{l,j} = \frac{1}{\Delta\mu_l} \frac{\partial g(k_l)}{\partial y_j}\bigg|_{\mathbf{y}=\mathbf{y}(k_l,\mathbf{x}^{(\nu)}),\mathbf{x}=\mathbf{x}^{(\nu)}}$$

and

$$G^{(\nu)}_{l,n} = \left[\frac{\partial g(k_l)}{\partial x_n}\bigg|_{\mathbf{x}=\mathbf{x}^{(\nu)}} + \sum_{j=1}^{M} \frac{\partial g(k_l)}{\partial y_j}\bigg|_{\mathbf{y}=\mathbf{y}(k_l,\mathbf{x}^{(\nu)});\mathbf{x}=\mathbf{x}^{(\nu)}} \cdot \frac{\partial y_j(k_l,\mathbf{x})}{\partial x_n}\bigg|_{\mathbf{x}=\mathbf{x}^{(\nu)}} \right] \Big/ \Delta\mu_l .$$

This scheme is just the Levenberg-Marquardt algorithm.[15] If necessary, its convergence properties may be improved by the under-relaxation technique.

A strategy is required to decrease the regularization parameter α during the iteration. We shall use below the following scheme: Starting with a value of the order of magnitude of the largest eigenvalue of the matrix Q, α is decreased in each step of the iteration by the same factor. At the end of the Levenberg-Marquardt regime the condition number of the regularized matrix $Q+\alpha A'$ shall still be well above the rounding limit of the computer.

The iteration scheme (28) can be rather cumbersome numerically when the matrix dimensions N or L are large. Three simplifications of the scheme have, therefore, been tested in several examples and have been found to lead to acceptable results: (i) Except for the last step in the iteration scheme, where Eq. (24) or (27) are to be satisfied, the rather complicated form of $\chi^2_{\rm cond}$ in terms of the matrix C, Eq. (15c), has been simplified to the expression (16); (ii) the partial derivatives $\partial y_j(k_l,\mathbf{x})/\partial x_n|_{\mathbf{x}=\mathbf{x}^\nu}$ have been substituted by the derivatives taken at $\mathbf{x}=\mathbf{x}^{(0)}=0$, and the derivatives with respect to the σ^2 components of $\mathbf{x}$ are taken to be zero; (iii) the truncation error $\Sigma_{ll'}$ has been calculated only once at the start of the iteration at $\mathbf{x}=0$, and the differences X_i have been defined with respect to a change of 0.1% in the curved-wave amplitudes.

The *a posteriori* errors depend on the combined statistical and systematic errors and in addition on the *a priori* probability. This is unavoidable for an ill-posed inversion problem. But nevertheless it is an undesirable feature of the *a posteriori* distribution that different researchers may have different opinions about $P_{\rm prior}$ and, therefore, obtain different results from fitting the same data. Although the options in the choice of A are reduced by determining α from Eq. (23), the choice of $\mathbf{x}^{(0)}$ still contains an element of considerable arbitrariness. However, as we have seen, in $\mathcal{R}$ space it is essentially the data which determine the result of the fit. For a function $F(\mathbf{P}_{\mathcal{R}}(\mathbf{x}))$ depending only on the projection $\mathbf{P}_{\mathcal{R}}$ of the model space on the subspace $\mathcal{R}$, errors can be defined which depend only on the experimental and systematic errors. One, therefore, expects

$$\overline{F}=\lim_{\alpha\to 0}\langle F\rangle_{\rm post}$$

and

$$\Delta F=\Big(\lim_{\alpha\to 0}\langle (F-\overline{F})^2\rangle_{\rm post}\Big)^{1/2}$$

to exist, which then are the *a priori*-independent average value and error of F, respectively.

If, however, the *a posteriori* errors are so large that the fluctuations around the solution $\mathbf{x}^*$ lead out of the range of the linear expansion Eq. (13), or if a possible skewness or kurtosis of the distributions P_j plays a role, analytical methods are rather useless and the *a posteriori* distribution has to be constructed by Monte Carlo simulations using Eqs. (2), (4) or (5), (11), and (25). It is not clear how to generalize Eq. (23) to obtain the optimal α in this case. However, it is clear that a condition like Eq. (23) or Eq. (27) is needed to determine the relative weight of the *a priori* information compared to the experimental information entering the fitting procedure.

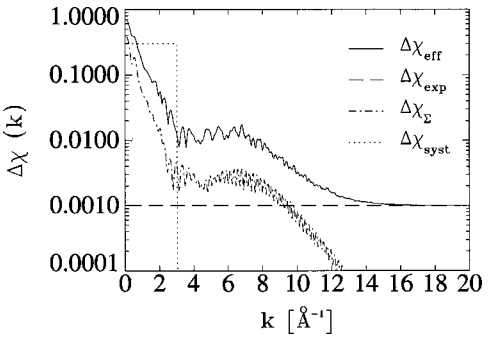

FIG. 1. The total error $\Delta\boldsymbol{\chi}_{\rm eff}$ (solid line) as function of k, the experimental error $\Delta\boldsymbol{\chi}_{\rm exp}$ (dashed line), the truncation error $\Delta\boldsymbol{\chi}_{\Sigma}=\sqrt{\Sigma_{ll}}$ (dash-dotted line), and the systematic error $\Delta\boldsymbol{\chi}_{\rm syst}$ (dotted line) for the Cu data.

As an example for the algorithm described in this section we analyze the K-edge copper data from Ref. 1 obtained at 10 K for k values between 0.05 and 24.95 Å^{-1} in steps of 0.05 Å^{-1}. We use these $\boldsymbol{\chi}_{\rm exp}(k)$ data in the k range from 0.1 to 19.9 Å^{-1} and assign an absolute statistical error of $\Delta\boldsymbol{\chi}_{\rm exp}=0.001$ to these data, which is consistent with the rms value of $F[\boldsymbol{\chi}_{\rm exp}(k);r]$ between 20 and 25 Å. In this test calculation we do not analyze the procedures used in Ref. 1 to obtain $\boldsymbol{\chi}(k)$ from the observable $\mu(k)$. As the *a priori* model-parameter set we intentionally choose incorrect lattice parameters: R_i values of an ideal fcc lattice with lattice constant 3.614 96 Å, valid at 291 K,[16] and Debye-Waller parameters from the correlated Debye model at $\theta=291$ K with $\theta_{\rm Debye}=315$ K.[17] As initial values for E_0 and S_0^2 we took $E_0=8979$ eV,[18] $S_0^2=0.9$. The coordination numbers N_i are not included in the fit, however the half-path lengths R_i are varied independently and not just via the lattice constant. To calculate the XAFS signal from the model-parameter set, the FEFF7 code[17] is used with 78 paths included in the sum of Eq. (1), ordered according to increasing path length and accounting for multiple scattering paths. They correspond to a filter of the curved wave-amplitude ratio in FEFF7 of 4% within a prefixed cluster radius of 8 Å. The largest half-path length in this set is $R_{\rm max}=7.7$ Å.

The truncation error $\Sigma_{ll'}$ is determined as described in Sec. II with $I=78$, $I_{\rm max}=1000$. The square roots of its diagonal matrix elements $\Delta\boldsymbol{\chi}_{\Sigma}=\sqrt{\Sigma_{ll}}$ are plotted in Fig. 1 together with the experimental errors $\Delta\boldsymbol{\chi}_{\rm exp}$. The vector $\mathbf{y}$ has the 157 components $\lambda,|f_1(R_1)|,\phi_1,\ldots|f_{78}(R_{78})|,\phi_{78}$. The

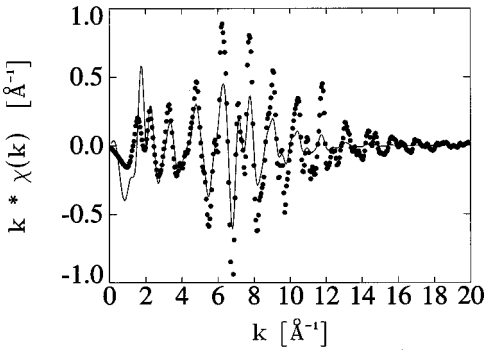

FIG. 2. EXAFS oscillations of the experimental $\boldsymbol{\chi}_{\rm exp}(k_l)$ (dots) and the *a priori* $\boldsymbol{\chi}_{\rm prior}(k_l)$ (line) for the 10 K Cu data.

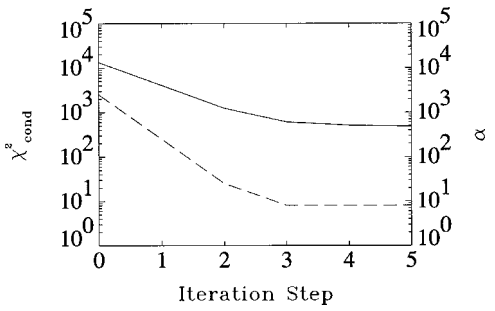

FIG. 3. Regularization parameter α (dashed line) and $\chi^2_{\rm cond}$ (solid line) of the first five iteration steps for the Cu data.

errors associated with them are assumed to be $\Delta\lambda/\lambda = 10\%$, $\Delta f_i/f_i = 7\%$, and $\Delta\phi_i = 0.07$ rad. The resulting effective error

$$\Delta\chi^{\rm eff}(k_l) = \sqrt{\Delta\chi_{\rm exp}(k_l)^2 + \sum_{j=1}^{157}\left(\frac{\partial\chi(k_l;\mathbf{y})}{\partial y_j}\Delta y_j\right)^2 + \Sigma_{ll}} \tag{29}$$

is represented by a solid curve in Fig. 1. Superimposed on the curve's general falloff with increasing k are rapid oscillations. They are mainly due to the derivatives $(\partial\chi/\partial y_j)^2$. All of them contain terms of the form $\Sigma_i[a_i\sin(2kR_i) + b_i\cos(2kR_i)]$. The square of this expression gives rise to all sum and difference frequencies contained in the EXAFS signal Eq. (1). Another systematic error (short dashed line) is added to this quantity to account for deficiencies of the theory for small k, in particular for the truncation in r space at 8 Å, and for uncertainties in $\mu_{\rm back}$ at small k values. We found that without this extra error the fitting procedure yields a considerable number of half-path lengths R_i which are shifted towards larger values than the *a priori* distances. In Fig. 2 the quantity $k\chi(k)$, calculated from the *a priori* parameter set is shown as full line and compared with the experimental data.

In the Levenberg-Marquardt regime of the iteration scheme the regularization parameter α is decreased as shown in Fig. 3. The corresponding decrease of the function $\chi^2_{\rm cond}(\mathbf{x}^\nu)$ is displayed as well. The eigenvalues of the 158 dimensional information matrix Q are plotted in Fig. 4. Also indicated are the values α^* and α'^* of the regularization parameter which satisfy conditions (24) and (27), respectively. The corresponding $\mathcal{R}$ spaces have dimensions of 21 and 40. In the remaining subspaces $\mathcal{P}$ of the model-

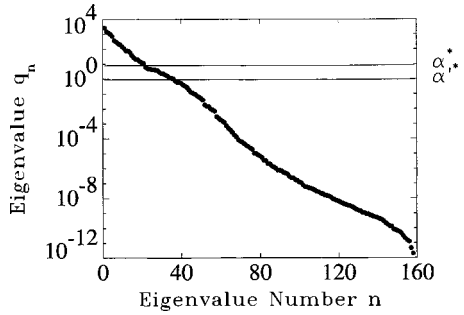

FIG. 4. Eigenvalues q_n of the information matrix $Q(\mathbf{x}^*)$ (dots) for the Cu data plotted with the cutoff lines $\alpha=\alpha^*$ and $\alpha=\alpha'^*$.

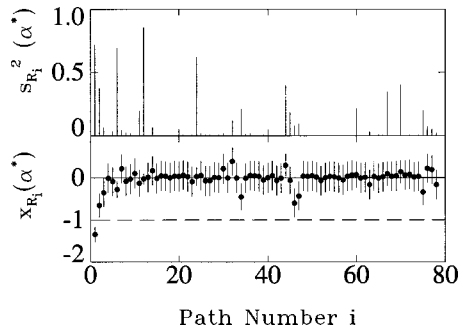

FIG. 5. Components of the solution vector $\mathbf{x}(\alpha^*)$ for the half-path lengths R_i of the Cu data together with the one-standard-deviation *a posteriori* error-band (lower part). The corresponding lengths s_i of the projection of the components of $\mathbf{x}^*$ into the space $\mathcal{R}$ are shown in the upper part.

parameter space the data do not influence the fit. These numbers are smaller than the ''number of independent data points'' traditionally used in Fourier analysis $N_d \approx (2/\pi)\Delta k\Delta R = 55$, with $\Delta k = k_{\rm max} - k_{\rm min} = 16.8$ Å^{-1} and $\Delta R = R_{\rm max} - R_1 = 5.1$ Å, where the effective narrowing of the bandwidth Δk due to the large extra *a priori* error for small k values has been taken into account. The larger size of N_d is not surprising since according to information theory N_d represents the number of items that can be encoded on the bandwidth Δk if the information is ideally packed. In general, however, it is not ideally packed.

One may expect to obtain a qualitative estimate for the degree to which the nth component of the solution vector $\mathbf{x}^*$ is modified by the data by taking a unit vector in the direction n and calculating the length s_n of its projection into the space $\mathcal{R}$. In terms of the components u_{ni} of the ith eigenvector of the matrix $Q_{nn'} + \alpha^*\delta_{nn'}$ one finds

$$s_n^2 = \sum_{i=1}^{N} u_{ni}^2\frac{q_i}{q_i+\alpha^*} = \sum_{n'=1}^{N} Q_{nn'}(Q+\alpha^* I)^{-1}_{n'n} \approx \sum_{i=1}^{i_{\mathcal{R}}} u_{ni}^2,$$

where $i_{\mathcal{R}}$ is the number of eigenvalues q_i for which $q_i > \alpha^*$. Note that the determination of α^*, α'^*, $\mathbf{x}^*$, and $\mathbf{s}$ requires only the inversion of the regularized matrix $Q+A$, and not its diagonalization. The latter was introduced here only for illustrative purposes. Values of s_n^2 close to unity indicate parameters which are in this sense well determined

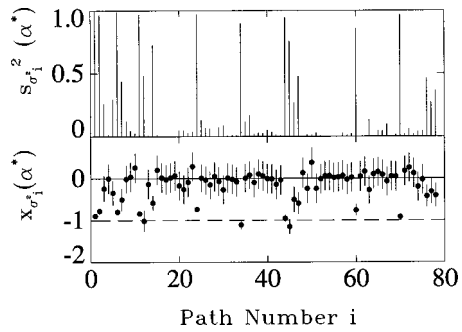

FIG. 6. Components of the solution vector $\mathbf{x}(\alpha^*)$ for the Debye-Waller parameters σ_i^2 of the Cu data together with the one-standard-deviation *a posteriori* error-band (lower part). The corresponding lengths s_i of the projection of the components of $\mathbf{x}^*$ into the space $\mathcal{R}$ are shown in the upper part.

by the data. The quantities s_n^2 are plotted in Figs. 5 and 6. One sees that large values of s_n^2 for certain R_i go together with large values of s_n^2 for the corresponding Debye-Waller parameter σ_i^2. Somewhat surprisingly, the σ_i^2 seem to be better determined than the corresponding R_i.

The components of the solution vector **x*** are shown in the lower frame of Figs. 5 and 6 together with their *a posteriori* errors. The scaling factors Δx_n of the components of **x** are defined using the estimates $\Delta S_0^2 = 0.1$ and $\Delta E_0 = 2$ eV. The values of ΔR_i were set equal to the difference of the half-path lengths corresponding to the lattice constants for a temperature of 291 K ($a = 3.61496$ Å) and of 10 K ($a = 3.6032$ Å).[17] Similarly the $\Delta\sigma_i^2$ values were taken as differences of the results obtained from the correlated Debye model at 291 and 10 K. Using these scaling factors the x_n^* values for $n > 2$ should be -1 within the *a posteriori* error margin, if the data would actually determine all R_i and σ_i^2, and if their values would be those of the ideal fcc lattice with lattice constant $a = 3.6032$ Å and of the correlated Debye model at 10 K. Actually only the fitted values of R_1 and R_2, and of σ_i^2 with $i = 1, 2, 6, 11, 12, 14, 24, 34, 44, 45, 47, 60, 70$, differ by more than two standard deviations from the *a priori* assumptions. Note that the numbers $i = 1, 2, 6, 9, 24, 44, 63$ label single scattering paths, $i = 11$ marks a focusing double scattering path, $i = 12$ a path with four atoms in a row. Surprisingly, the σ_i^2 seem to be somewhat more sensitive to the experimental (rather than the *a priori*) input of the fit than the R_i. It is remarkable that the fit yields $R_1 = 2.5450 \pm 0.0016$ Å for the first half-path length, which is by two standard deviations smaller than the nearest-neighbor distance of 2.5479 Å in the fcc lattice with $a = 3.6032$ Å. Since the *a priori* assumption was even larger than this distance, the result of the fit is certainly not due to the *a priori* input. It is, however, subject to the appropriateness of the FEFF7 code and our assumptions on the size of its systematic errors.

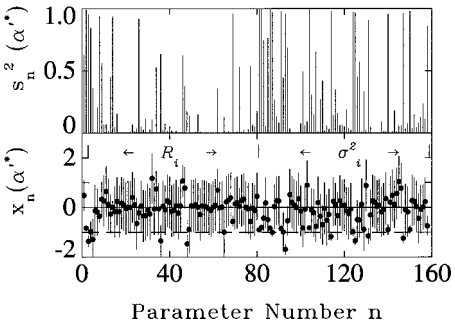

FIG. 7. Components of the solution vector $\mathbf{x}(\alpha'^*)$ for the Cu data together with the one-standard-deviation *a posteriori* error-band (lower part). The lengths s_n of the projection of the components of $\mathbf{x}'^*$ into the space $\mathcal{R}$ are shown in the upper part. The parameter sequence starting with $n = 1$ is $S_0^2,\ E_0,\ R_1, \ldots R_{78},\ \sigma_1^2, \ldots \sigma_{78}^2$.

The regularization parameter α^* was determined from the condition that the solution should be as close to the *a priori* parameter set as compatible with the data. A smaller weight is given to the *a priori* input when the ''most probable'' regularization parameter α'^* is used. The components of the solution vector $\mathbf{x}'^*$ corresponding to $\alpha = \alpha'^*$ are displayed in Fig. 7. Since the dimension of the $\mathcal{R}$ space is larger now, the s_n are generally larger than in Figs. 5 and 6. Since the *a*

TABLE I. Scattering-path features of the *a priori* Ta crystal strucure. For each scattering path i the component numbers n_R and n_{σ^2} are given in columns 2 and 3, where the corresponding R_i and σ_i^2 are located, respectively, in the vector **x**. The fourth column shows the values of R_i (in Å) for the ideal lattice and the fifth column gives the vertex angles for each scattering path i.

i	n_R	n_{σ^2}	R_i	Angles	i	n_R	n_{σ^2}	R_i	Angles
1	3	44	2.858	180, 180	22	24	65	5.815	151, 55, 155
2	4	45	3.300	180, 180	23	25	66	6.158	180, 55, 180, 55
3	5	46	4.508	125, 125, 109	24	26	67	6.158	125, 55, 125, 55
4	6	47	4.508	125, 109, 125	25	27	68	6.158	55, 180, 55, 180
5	7	48	4.667	180, 180	26	28	69	6.158	55, 180, 55, 180
6	8	49	5.191	145, 145, 71	27	29	70	6.158	180, 125, 180, 125
7	9	50	5.191	145, 71, 145	28	30	71	6.158	55, 125, 55, 125
8	10	51	5.472	180, 180	29	31	72	6.158	90, 125, 71, 125
9	11	52	5.633	135, 90, 135	30	32	73	6.499	149, 121, 90
10	12	53	5.716	180, 180	31	33	74	6.499	149, 90, 121
11	13	54	5.716	180, 180, 0	32	34	75	6.499	121, 90, 149
12	14	55	5.716	180, 0, 180	33	35	76	6.600	180, 180
13	15	56	5.716	180, 0, 180, 0	34	36	77	6.600	180, 180, 0
14	16	57	5.716	180, 180, 180, 180	35	37	78	6.600	180, 0, 180
15	17	58	5.716	0, 180, 0, 180	36	38	79	6.600	180, 0, 180, 0
16	18	59	5.716	109, 180, 109, 180	37	39	80	6.600	0, 180, 0, 180
17	19	60	5.716	71, 180, 71, 180	38	40	81	6.841	90, 125, 0, 145
18	20	61	5.716	109, 71, 109, 71	39	41	82	6.841	145, 0, 125, 90
19	21	62	5.716	71, 109, 71, 109	40	42	83	6.841	90, 145, 0, 125
20	22	63	5.815	155, 151, 55	41	43	84	6.841	135, 55, 109, 90
21	23	64	5.815	155, 55, 151					

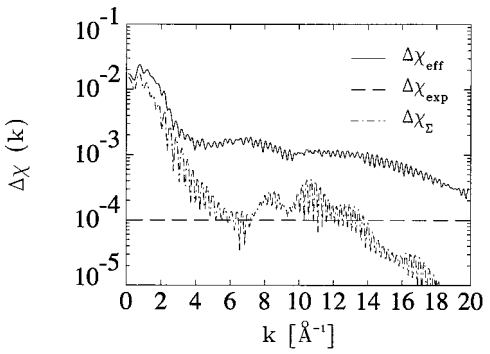

FIG. 8. The total error $\Delta\boldsymbol{\chi}_{\text{eff}}$ (solid line) as function of k, the experimental error $\Delta\boldsymbol{\chi}_{\text{exp}}$ (dashed line), and the truncation error $\Delta\boldsymbol{\chi}_{\Sigma}=\sqrt{\Sigma_{ll}}$ (dash-dotted line) for the first computer experiment.

posteriori errors are also larger in this case, many more of the R_i and σ_i^2 are compatible with the expectation for the fcc lattice at 10 K. However, R_1 is still smaller by more than one standard deviation.

VI. EXPERIMENT DESIGN

Before an experiment starts, a number of decisions have to be made. There is the trivial question to optimize $\Delta\mu_l$ since according to Eq. (6) choosing smaller values Δk, i.e., a narrower slit setting, means reduced statistics under the constraint of constant beam time for each l value, thus increasing $\Delta\mu_l'$. The optimal choice depends on $\partial g(k;\mathbf{x},\mathbf{y})/\partial k$. Then there is the less trivial question how to distribute the k_l optimally and to find the optimal L, again under the constraint of constant total beam time. Further one needs to know to which extent it is desirable to reduce the $\Delta\mu_l$ in view of the assumed systematic errors, which limit the value of the measurement anyhow. Finally, one may want to achieve a certain minimal accuracy for at least some model parameters and therefore the question arises whether this accuracy can be achieved at all in view of the systematic errors and, if it can be achieved, what the minimal beam time is, assuming L and $\Delta\mu_l$ to be already optimized.

Since most of these optimization problems are nonlinear, Monte Carlo simulations, preceding the experiment, are most appropriate. They all follow the same strategy: Using a supposedly true $\mathbf{x}^{(\text{true})}$ for the model parameters and the model g, values $\mu(k_l)$ are calculated at prescribed values k_l. Based on the simplified version for χ^2_{cond} given by Eq. (16) Gaussian-

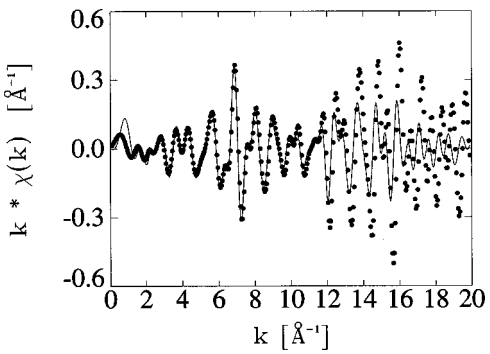

FIG. 9. EXAFS oscillations of the ''experimental'' $\boldsymbol{\chi}_{\text{exp}}(k_l)$ (dots) and the *a priori* $\boldsymbol{\chi}_{\text{prior}}(k_l)$ (line).

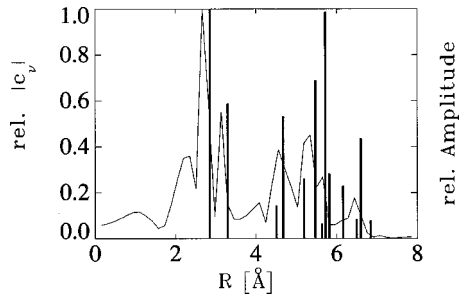

FIG. 10. Radial distribution of the absolute value of the Fourier coefficient c_ν of $k\boldsymbol{\chi}_{\text{exp}}(k_l)$ normalized to its maximum value (thin line) compared to the relative amplitudes of the next-neighbor shells (thick bars).

distributed random numbers with width $\Delta\mu_l^{\text{eff}}$ are added to the μ_l. The resulting ''pseudoexperimental'' μ_l are used as input for the fit procedure and it is checked how accurately the original model parameters $\mathbf{x}^{(\text{true})}$ are reproduced. Also the *a posteriori* distribution of the fitted model parameters is obtained, which allows one to decide whether the required accuracy can be achieved with the assumed experimental setting and with the estimated systematic errors.

A. Computer experiment 1

As an example for the procedure proposed in this paper we simulate an experiment on tantalum with a bcc lattice. We assume that the lattice is actually strongly distorted, but that the *a priori* guess corresponds to the distances R_i and coordination numbers N_i of the ideal lattice. In order to show the robustness of the iteration procedure we shall assume that also the first two components of the vector $\mathbf{x}$, E_0, and S_0^2 and the Debye-Waller parameters σ_i^2 are actually, in our computer example, very far away from their standard values. Specifically, the *a priori* vector of model parameters $\mathbf{x}^{(0)}$ has components: $E_0=9881$ eV, the L_3 edge of tantalum, $S_0^2=1$, all half-path lengths R_i are computed for the ideal bcc lattice with lattice constant 3.3 Å and all σ_i^2 are assumed to be 0.003 Å^2. For the actual, distorted system we assume $E_0=9879$ eV, $S_0^2=0.8$, the lattice constant is reduced by 1% compared to the ideal lattice values, and all σ_i^2 are assumed to be 0.001 Å^2. We shall refer to this set of model parameters in the following as the ''true'' parameter set $\mathbf{x}^{\text{true}}$. It describes a strongly compressed system, rarely met in practice. But here we want to find out whether the procedure finds an acceptable solution even in an extreme situation, where the *a priori* guess $\mathbf{x}^{(0)}$ is far from the true solution. The coordination numbers N_i shall be always those of the ideal lattice and are therefore not included in the fit and in the parameter set $\mathbf{x}$.

To calculate scattering amplitudes and phases we use the FEFF7 code[17] for a Ta cluster with radius 7 Å, and select the 41 paths of single and multiple scattering with an average contribution of more than 4% to the sum in Eq. (1). The vector $\mathbf{x}$ then consists of the $N=84$ quantities $S_0^2, E_0, R_1, \ldots R_{41}, \sigma_1^2, \ldots \sigma_{41}^2$, where the half-path lengths R_i are assumed to be arranged in increasing order. The largest R_i is 6.84 Å. Table I shows for these 41 paths the component numbers n of the vector $\mathbf{x}$ for the corresponding R_i

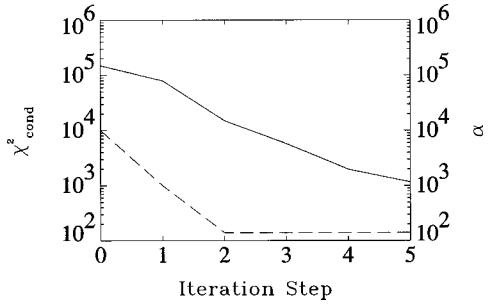

FIG. 11. Regularization parameter α (dashed line) and $\chi^2_{\rm cond}$ (solid line) of the first five iteration steps for the first computer experiment.

and σ_i^2 together with the half-path lengths $R_i^{(0)}$ and the vertex angles of the scattering paths: a single scattering path is a single line traversed back and forth and has the two vertex angles 180, 180; a stretched double scattering path has angles 180, 0, 180, and so on. The vector $\mathbf{y}$ has the $M=83$ components $\lambda, |f_1(R_1)|, \phi_1, \ldots |f_{41}(R_{41})|, \phi_{41}$. A first computer experiment is defined by the k window $k_{\rm min}=0.1$ Å^{-1}, $k_{\rm max}=19.9$ Å^{-1}, and $L=397$, i.e., measurements are assumed to be done at k values in steps of 0.05 Å^{-1}. We do not consider the reduction of the observable $\mu(k)$ to $\chi(k)$ in this example.

Using $\mathbf{x}^{\rm true}$ as input for the FEFF7 code, we calculate ''true'' values $\boldsymbol{\chi}_{\rm true}(k_l)$, $l=1, \ldots 397$ to which Gaussian-distributed errors with width $\Delta\boldsymbol{\chi}_l^{\rm eff}$ are added at random resulting in ''experimental'' $\boldsymbol{\chi}_{\rm exp}(k_l)$. To obtain the width $\Delta\boldsymbol{\chi}_l^{\rm eff}$ of Eq. (29) we assume in this example ''experimental'' errors $\Delta\boldsymbol{\chi}_l$ of 10^{-4} at each k_l and model uncertainties $\Delta y_j/y_j$ of 0.3% for the mean free path λ, for all amplitudes $|f(R_i)|$, and for all the phases ϕ_i. The truncation error Σ_{ll} was determined from comparing FEFF7 results received from a curved wave amplitude filter of 4% with a sequence of results between the path number $I=41$ and the maximum path number $I_{\rm max}=1000$. These errors are plotted in Fig. 8. For $k<2$ Å^{-1} the error $\Delta\boldsymbol{\chi}^{\rm eff}$ is seen to be dominated by the truncation error. In the FEFF7 code the k dependence of S_0 for small k is neglected. The corresponding uncertainty is

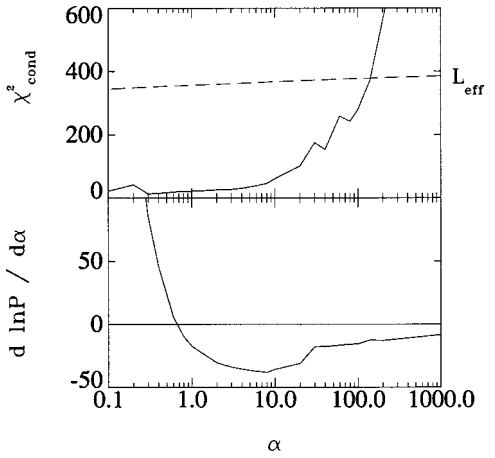

FIG. 12. Upper part: $\chi^2_{\rm cond}(\overline{\mathbf{x}})$ (solid line) and the effective number of degrees of freedom $L_{\rm eff}$ (dashed line) as function of the regularization parameter α. Lower part: Logarithmic derivative of Eq. (26) with respect to α.

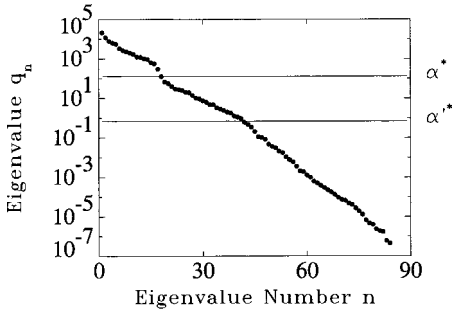

FIG. 13. Eigenvalues q_n of the information matrix $Q(\mathbf{x}^*)$ (dots) for the first computer experiment plotted with the cutoff lines $\alpha=\alpha^*$ and $\alpha=\alpha'^*$.

assumed to be absorbed in Σ_{ll}. Our choice for the small size of the errors $\Delta\boldsymbol{\chi}_l$ and $\Delta\boldsymbol{\chi}_l^{\rm eff}$ was motivated by an attempt to be equally far away from the two (trivial) limiting cases where either the $\mathcal{P}$ or the $\mathcal{R}$ space exhausts the whole model parameter space. The scaling vector $\Delta\mathbf{x}$ is chosen in such a way that all components of the scaled vector $\mathbf{x}$ have the same size: $x_i=-1$ for $1\leqslant i\leqslant 84$. With $\mathbf{x}^{(0)}$ as input FEFF7 yields the EXAFS signal for the undisturbed lattice $\boldsymbol{\chi}_{\rm prior}(k_l)$. The functions $\boldsymbol{\chi}_{\rm exp}(k_l)$ and $\boldsymbol{\chi}_{\rm prior}(k_l)$ are shown in Fig. 9. It should be pointed out that for our example the experimental data are well described by the *a priori* parameters within the k range 3 Å$^{-1}\leqslant k\leqslant 11$ Å^{-1}. An EXAFS analysis based on this limited data range would yield false structure informations.

From Eq. (12) the Fourier transform c_ν of $k_l\boldsymbol{\chi}_{\rm exp}(k_l)$ is obtained and its absolute value is plotted vs r in Fig. 10, in which also the input values R_i are indicated by the thick bars. Their length is proportional to N_i. The highest peak at $r=2.67$ Å is identified with the first shell radius R_1. It is shifted with respect to the true value due to the k dependence of the phase ϕ_1. Note that the peak of the Fourier transform at 2.3 Å is not connected with any of the half-path lengths R_i, but is known to result from the k dependence of the amplitude A_1. We do not try to identify any of the other R_i with the peaks in Fig. 10 beyond $r=2.67$ Å. Also no attempt was made here to ''improve'' the Fourier transform by using as input some k^n-weighted $\boldsymbol{\chi}$ with $n>1$, instead of $k\boldsymbol{\chi}$.

With our initial values $\mathbf{x}^{(0)}$ we start an iteration as described in Sec. V. The regularization parameter α and $\chi'^2_{\rm cond}$ are shown in Fig. 11 along the sequence of the first five

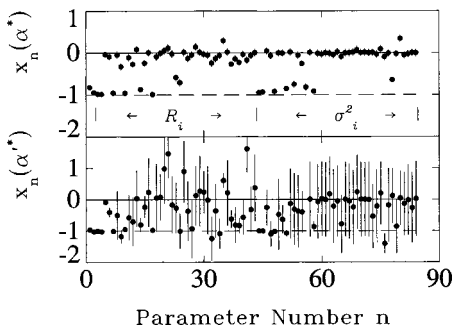

FIG. 14. Components of the solution vectors $\mathbf{x}(\alpha^*)$ (upper part) and $\mathbf{x}(\alpha'^*)$ (lower part) together with the one-standard-deviation *a posteriori* error-band. The parameter sequence starting with $n=1$ is $S_0^2,\ E_0,\ R_1, \ldots R_{41},\ \sigma_1^2, \ldots \sigma_{41}^2$.

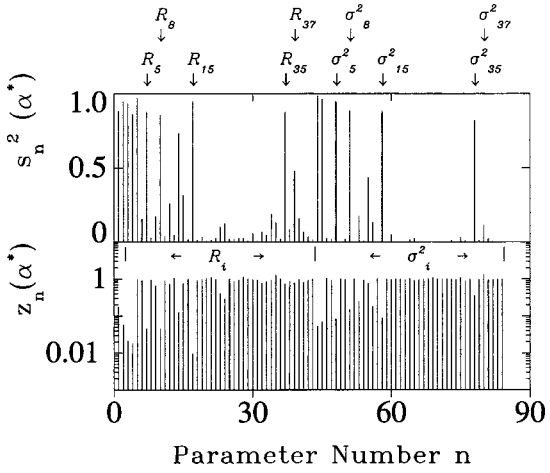

FIG. 15. Relative deviation z_n between the components of the true model-parameter vector $\mathbf{x}^{\text{true}}$ and the solution vector $\mathbf{x}^*$ (lower part). The lengths s_n of the projection of the components of $\mathbf{x}^*$ into the space $\mathcal{R}$ are shown in the upper part. The parameter sequence starting with $n=1$ is $S_0^2,\ E_0,\ R_1, \ldots R_{41},\ \sigma_1^2, \ldots \sigma_{41}^2$.

iterations. In Fig. 12 we plot $\chi'^2_{\text{cond}}(\overline{\mathbf{x}})$ as function of the regularization parameter α. It shows the result of iterating with respect to $\mathbf{x}$ for fixed α. The iteration is stopped when the average value of $|(x^{(\nu+1)}-x^{\nu})_i|$ is less than 10^{-6} or when a local minimum is reached. Also plotted as function of α is the rhs of Eq. (24). The point of intersection yields $\alpha^*=130$. For $\alpha<100$ convergence of the iteration deteriorates and requires increasingly larger under-relaxation of the step size. In the lower frame of Fig. 12 the lhs of Eq. (27) is shown as function of α. The zero point of this function yields $\alpha'^*=0.7$. As expected $\alpha^*>\alpha'^*$.

In Fig. 13 the eigenvalues q_n of the information matrix $Q(\overline{\mathbf{x}})$ are plotted together with the cutoff lines $\alpha=\alpha^*$ and $\alpha=\alpha'^*$. The eigenspaces belonging to the eigenvalues above these lines span the spaces $\mathcal{R}$ in which the ''data'' determine the fit in these two cases. Their dimensions $i_{\mathcal{R}}$ are seen to be 17 and 42. It may be compared with the ''number of independent data points'' $N_d \approx (2/\pi)\Delta k \Delta R = 50$. As could be expected, even with the very small experimental and systematic errors used, the $\mathcal{R}$ space consists of only a

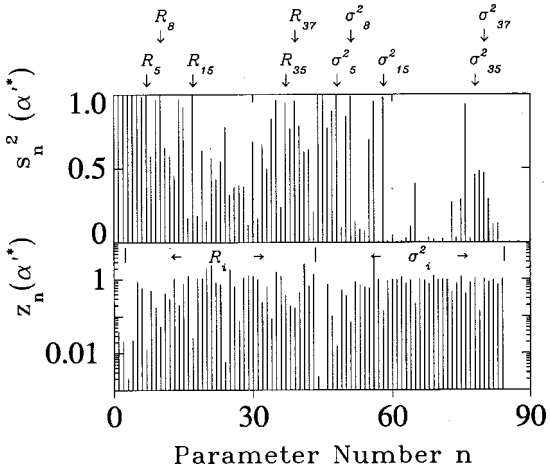

FIG. 16. Same as Fig. 15 for the solution vector $\mathbf{x}'^*$ which corresponds to the regularization parameter $\alpha=\alpha'^*$.

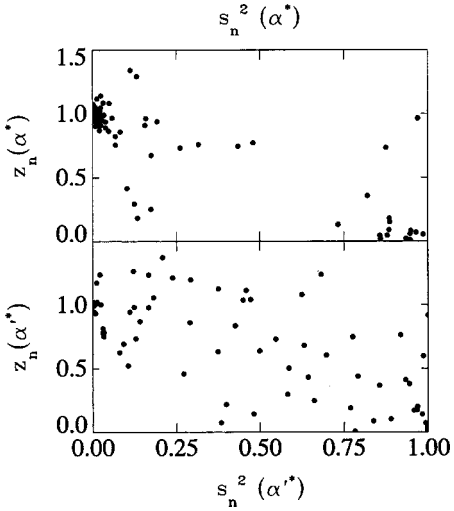

FIG. 17. Correlation between s_n and z_n for the solution vectors $\mathbf{x}(\alpha^*)$ (upper part) and $\mathbf{x}(\alpha'^*)$ (lower part).

fraction of the 84-dimensional model space with which we started. We chose its dimension so large on purpose in order to demonstrate the ability of the method to always identify the relevant parameter subspace $\mathcal{R}$.

In Fig. 14 we show the components of the solution vectors $\mathbf{x}^*$ corresponding to α^* and to α'^*, together with the one-standard-deviation *a posteriori* error band which, at $\alpha=\alpha^*$ is smaller than the dot size. In Figs. 15 and 16 we present for α^* and α'^* the ratios $z_n=|(x_n^*-x_n^{\text{true}})/x_n^{\text{true}}|$, which measure the degree to which the data allowed one to correct the ''error'' in the *a priori* assumption. It is seen that the first two R_i are rather well determined by the data and so are E_0, σ_1^2, and σ_2^2. Note that R_3 corresponds to the first double-scattering path as can be seen from Table I. The path lengths R_5 and R_8 belong to single scattering paths. They are fairly well determined together with σ_5^2 and σ_8^2. In comparison with neighboring paths, R_{12}, and R_{35} are fairly reproduced as well. They correspond to multiple scattering paths with three Ta atoms in a line. The same holds for R_{15} and R_{37} representing multiple scattering with four Ta atoms in a line. Compared with E_0 the many-body factor S_0^2 is obviously somewhat more difficult to reproduce.

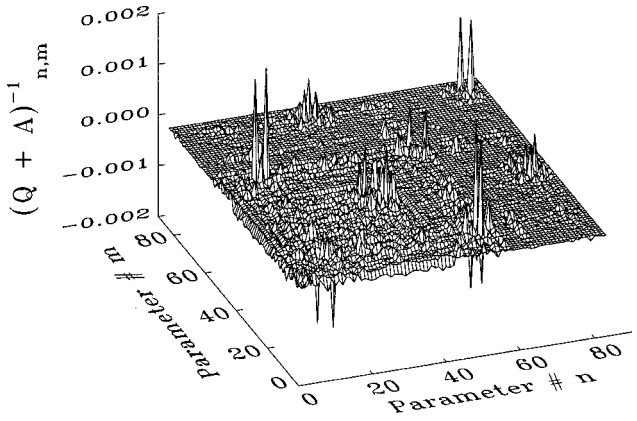

FIG. 18. Matrix elements of the regularized variance matrix $(Q+\alpha^* I)^{-1}_{nm}$; diagonal elements are suppressed. Parameter sequence starting from $n=0$: $S_0^2,\ E_0,\ R_1, \ldots R_{41},\ \sigma_1^2, \ldots \sigma_{41}^2$.

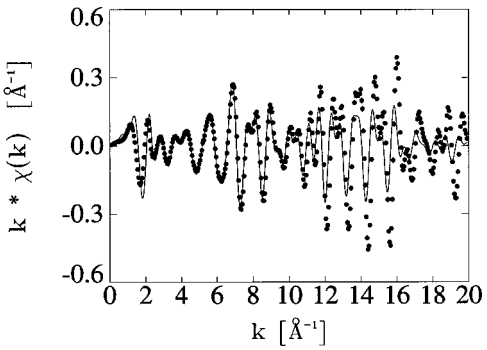

FIG. 19. EXAFS oscillations of the ''experimental'' $\boldsymbol{\chi}_{\mathrm{exp}}(k_l)$ (thick line) and the *a priori* $\boldsymbol{\chi}_{\mathrm{prior}}(k_l)$ (thin line) of the simulated experiment 2a.

In an actual experiment the quantities z_n are of course not available since $\mathbf{x}^{\mathrm{true}}$ is not known in that case. Always available, however, are the s_n, which are shown in Figs. 15 and 16. Note that the sequence of model parameters still is $E_0,\ S_0^2,\ R_1,\ldots,R_{41},\ \sigma_1^2,\ldots,\sigma_{41}^2$. Although $1-s_n^2$ is not exactly equal to z_n, these quantities, nevertheless, have in general a rather similar behavior when the regularization parameter α^* is used. In the upper part of Fig. 17 we show the correlation between s_n and z_n. This is less obvious for the regularization parameter α'^*, as shown in the lower part of Fig. 17.

Since $i_{\mathcal{R}}$ is smaller for α^* than for α'^*, more of the s_n^2 come close to one when the regularization parameter α'^* is used as seen by comparing Fig. 16 with Fig. 15. In particular the first four single scattering path parameters $R_1,\ \sigma_1^2,\ R_2,\ \sigma_2^2,\ R_5,\ \sigma_5^2,\ R_8,\ \sigma_8^2$ are fairly well determined by the fit with α'^*. Although the z_n are in general smaller in Fig. 16 than in Fig. 15, there are a few exceptions, e.g. the focusing scattering path parameters $R_{12},\ \sigma_{12}^2,\ R_{15},\ \sigma_{15}^2$, even though the corresponding s_n^2 are still close to unity. One should, however, be aware that the *a posteriori* errors, plotted together with the fitted model-parameter values x_n in Fig. 14, are much larger for the smaller regularization parameter, and so are the error correlations. The fitted model parameters are therefore sometimes more accurately determined with α'^*, but with smaller reliability.

The matrix elements of the regularized variance matrix $(Q+A)^{-1}_{nn'}$ are plotted in Fig. 18 for $\alpha=\alpha^*$, where the diagonal matrix elements have been suppressed. A strong correlation between neighboring R_i and neighboring σ_i^2 leads to the rugged mountain region in the diagonal direction. In addition, correlations between R_i and σ_j^2 with $i\approx j$ give rise to the ''side bands'' in this figure.

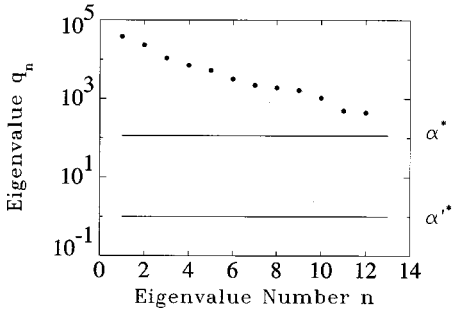

FIG. 20. Eigenvalues q_n of the information matrix $Q(\mathbf{x}'^*)$ (dots) plotted with the cutoff lines $\alpha=\alpha^*$ and $\alpha=\alpha'^*$ of the simulated experiment 2a.

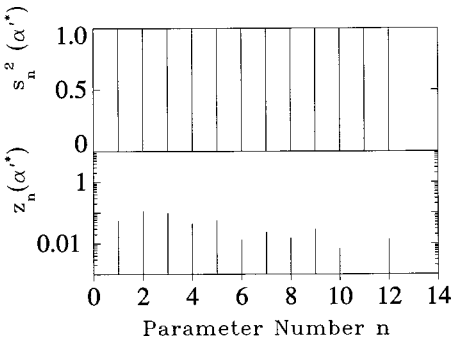

FIG. 21. Same as Fig. 16 for the simulated experiment 2a. Parameter sequence starting from $n=1$: $S_0^2,\ E_0,\ R_1,\ldots R_5,\ \sigma_1^2,\ldots\sigma_5^2$.

B. Computer experiment 2

Naturally one would like to see whether in the limit of vanishing errors $\Delta\boldsymbol{\chi}_l$ the true model-parameter set could always be recovered, irrespective of the choice of the *a priori* $\mathbf{x}^{(0)}$. In the framework of the first computer experiment this is unfortunately not possible because finite grid-size limitations in the FEFF7 code do not allow one to go with Δy_j to zero. Furthermore, the minimal step size in the k space of 0.05 Å^{-1} turns out to be not sufficient to fully resolve the structure of $\boldsymbol{\chi}(k)$ in this computer experiment. We, therefore, define a second computer experiment in which only the single scattering paths $R_1,\ldots,R_5$ of the first five shells are taken into account. The model-parameter vector $\mathbf{x}$ now has only 12 components, corresponding to the quantities $S_0^2,\ E_0,\ R_1,\ldots,R_5,\ \sigma_1^2,\ldots,\sigma_5^2$. The grid in the k space and the errors are the same as before: $\Delta\boldsymbol{\chi}_l=10^{-4}$ for $l=1,\ldots 397$. The vector $\mathbf{y}$ consists of the 11 components $\lambda,\ t_1,\ \phi_1,\ldots,t_5,\ \phi_5$ and the errors $\Delta y_j/y_j$ are again 0.3% in each component. Since we consider just the first five single scattering paths, we have $\Sigma_{ll'}=0$. This would, of course, not be possible when actually measured data were to be analyzed. The choice of the *a priori* vector $\mathbf{x}^{(0)}$ is the

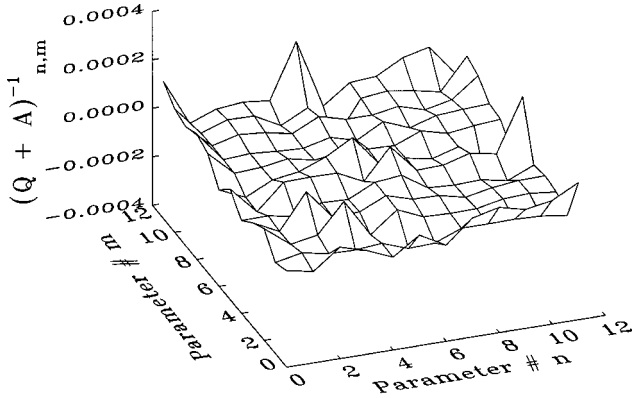

FIG. 22. Same as Fig. 18 for the simulated experiment 2a. Parameter sequence starting from $n=0$: $S_0^2,\ E_0,\ R_1,\ldots R_5,\ \sigma_1^2,\ldots\sigma_5^2$.

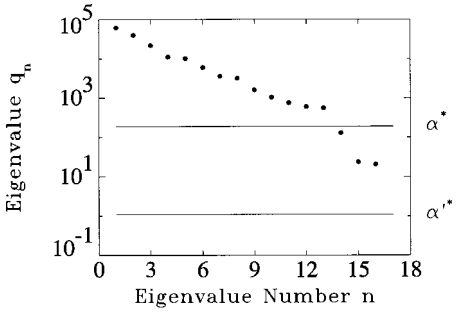

FIG. 23. Eigenvalues q_n of the information matrix $Q(\mathbf{x}'^*)$ (dots) plotted with the cutoff lines $\alpha=\alpha^*$ and $\alpha=\alpha'^*$ of the simulated experiment 2b.

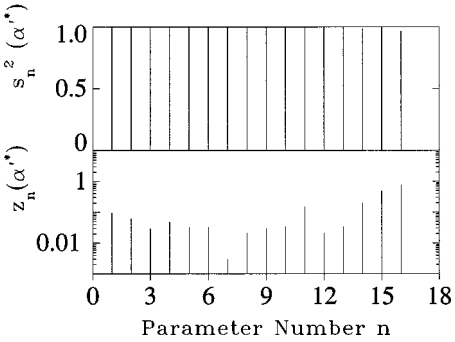

FIG. 24. Same as Fig. 21 for the simulated experiment 2b. Parameter sequence starting from $n=1$: $E_0,\ R_1,\ldots R_5,\ \sigma_1^2,\ldots\sigma_5^2,\ N_1,\ldots N_5$.

same as before, i.e., it corresponds to the ideal lattice, and $\mathbf{x}^{\text{true}}$ describes again the compressed system.

Figure 19 shows χ_{exp}, which is essentially χ_{true} in this case, and $\chi(k)$ calculated with the *a priori* model-parameter set. Comparing with Fig. 9 it is clear that now the highest frequency oscillations are missing. Figure 20 indicates that α^* and α'^* are smaller than all 12 eigenvalues q_ν of Q so that the data determine the fit in the whole model-parameter space. This finding is confirmed by the behavior of the quantities s_n^2 shown in the upper part of Fig. 21. That the true solution was actually found can be seen from the behavior of the z_n displayed in the lower part of Fig. 21. The largest deviations of about 10% occur for x_{E_0} and x_{R_1}, corresponding to deviations of $\Delta E_0=0.22$ eV and $\Delta R_1=0.0025$ Å. Due to the small amount of well-determined model parameters taken in this example the error correlation matrix shown in Fig. 22 (diagonal matrix elements again suppressed) has very small matrix elements compared to the first computer experiment shown in Fig. 18.

It is instructive to consider a variant of this second computer experiment where also the five coordination numbers N_i are included in the model-parameter vector $\mathbf{x}$, which then has 17 components. For $\mathbf{x}^{\text{true}}$ we choose again the compressed system, where, in addition, we assume that one atom is missing in each shell. In this example one eigenvalue of the information matrix Q becomes zero, which reflects the fact that S_0^2 and N_i are not linear independent variables in the model equation (1). We, therefore, take S_0 equal to the ''true'' value of 0.8 and drop it from the list of model parameters. There are then 16 remaining model parameters. All 16 eigenvalues of the matrix Q are again larger than α'^* (see Fig. 23), i.e., the problem is well posed. Therefore, all s_n shown in the upper part of Fig. 24, are close to unity. Nevertheless, all of the N_i are not very well determined as is seen from the rather large values of z_{14} to z_{16} in the lower part of Fig. 24. As shown in Fig. 25 there are strong error correlations between these components of the *a posteriori* error-correlation matrix. In Table II, we display the numerical values of the 16 model parameters in their initial units for the *a priori*, the *a posteriori*, and the ''true'' set.

VII. SUMMARY

We have presented a method to analyze EXAFS data in terms of model parameters in which *a priori* knowledge about the model is introduced through Bayes' theorem. Turchin's conditions were used to ensure that the information contained in the data is not distorted by *a priori* input into the fitting procedure. Besides the statistical errors in the data, a reliability estimate of the model was introduced. Together with the *a priori* assumptions and the errors of the data this estimate determines the *a posteriori* error of the deduced best model parameters. The *a posteriori* error is a measure of the probability with which the data require a modification of the *a priori* parameter set. The method does not require an *a priori* restriction of the dimension of the model-parameter space to the size one expects to be determined by the data. Instead, the fitting procedure itself allows one to calculate the subspace of the model-parameter space in which the data affect the outcome of the fit. It can, in particular, simulate a model-independent data analysis to any desired degree of accuracy by choosing the model space sufficiently large. One also obtains for each model parameter the degree to which the data, rather than *a priori* assumptions, contribute to its fitted value.

The method has been applied to Cu data measured at a temperature of 10 K. To demonstrate the robustness of the analysis we used 158 fit parameters starting from the *a priori* data set appropriate for a temperature of 291 K. The dimensions of the model-parameter subspace in which the data

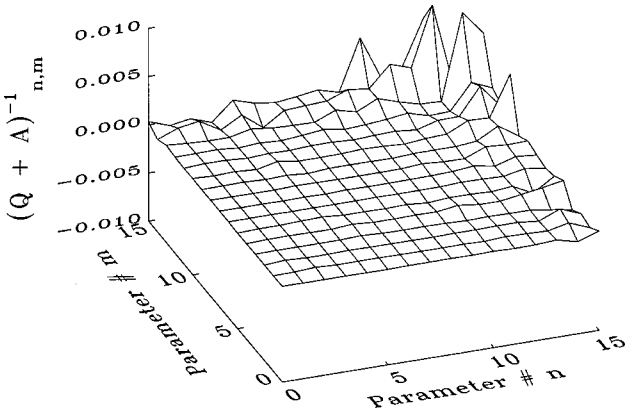

FIG. 25. Same as Fig. 22 for the simulated experiment 2b. Parameter sequence starting from $n=0$: $E_0,\ R_1,\ldots R_5,\ \sigma_1^2,\ldots\sigma_5^2,\ N_1,\ldots N_5$.

TABLE II. *A priori*, *a posteriori*, and ''true'' parameter values of the simulated Ta experiment 2a. The component numbers *n* are given in column 1, the corresponding parameter names, *a priori* parameters, fitted parameter values including their *a posteriori* errors, and the ''true'' parameter values are shown in columns 2, 3, 4, and 5, respectively.

n	Name	*a priori*	*a posteriori*	''True''
1	E_0	9881 eV	9878.81±0.01 eV	9879 eV
2	R_1	2.8579 Å	2.8275±0.0005 Å	2.8293 Å
3	R_2	3.3000 Å	3.2650±0.0006 Å	3.2670 Å
4	R_3	4.6669 Å	4.6180±0.0009 Å	4.6202 Å
5	R_4	5.4724 Å	5.4159±0.0009 Å	5.4177 Å
6	R_5	5.7158 Å	5.6567±0.0022 Å	5.6586 Å
7	σ_1	0.003 Å^2	0.001 01±0.000 01 Å^2	0.001 Å^2
8	σ_2	0.003 Å^2	0.000 96±0.000 03 Å^2	0.001 Å^2
9	σ_3	0.003 Å^2	0.001 06±0.000 03 Å^2	0.001 Å^2
10	σ_4	0.003 Å^2	0.001 07±0.000 02 Å^2	0.001 Å^2
11	σ_5	0.003 Å^2	0.001 30±0.000 08 Å^2	0.001 Å^2
12	N_1	8	7.02±0.04	7
13	N_2	6	4.97±0.04	5
14	N_3	12	11.20±0.10	11
15	N_4	24	23.50±0.20	23
16	N_5	8	7.78±0.18	7

dominate the fit were $i_{\mathcal{R}}(\alpha^*)=21$ and $i_{\mathcal{R}}(\alpha'^*)=40$, depending on the assumed weight of the *a priori* assumptions. These numbers are smaller than the ''number of independent data points'' $N_d \approx (2/\pi)\Delta k \Delta R = 55$, traditionally used in Fourier analysis. The example further showed that only a few half-path lengths and Debye-Waller parameters were determined by the data indicating the locality of the EXAFS method.

We have also shown how the computer simulation of a planned experiment can be used to find out whether a given size and accuracy of the expected data allow the determination of any model parameter within a given error margin. We gave a few examples for estimates of the uncertainties connected with the FEFF code. However, we are aware that much more needs to be done in this direction. The same holds for an estimate of the errors in μ_0 and in μ_{back}. More extensive applications of the fitting procedure to real data and additional efforts to estimate the errors of the model parameters are needed to improve the present situation.

ACKNOWLEDGMENTS

We thank R. Lipperheide and R. Silver for valuable discussions, and J. Rehr for his continuing encouragement and support of this investigation.

[1] M. Newville, Ph.D. thesis, University of Washington, 1994.
[2] J. Mustre de Leon, J. J. Rehr, S. I. Zabinsky, and R. C. Albers, Phys. Rev. B **44**, 4146 (1991).
[3] E. A. Stern, in *X-Ray Absorption*, edited by D. C. Koningsberger and R. Prins (Wiley, New York, 1988), p. 3.
[4] J. A. Victoreen, J. Appl. Phys. **19**, 855 (1948); *International Tables for X-Ray Cristallography III*, edited by K. Lonsdale *et al.* (Kynoch, Birmingham, England, 1962), Sec. 3.2.
[5] J. Skilling, in *Maximum Entropy in Action*, edited by B. Buck and V. A. Macanlay (Oxford University, Oxford, 1991), Chap. 2.
[6] V. F. Turchin and V. Z. Nozik, Izv. Atmospheric and Oceanic Physics **5**, 29 (1969).
[7] H. Richter, *Wahrscheinlichkeitstheorie* (Springer, Berlin, 1966), Chap. 3.
[8] D. E. Sayers and B. A. Bunker, in *X-Ray Absorption*, edited by D. C. Koningsberger and R. Prins (Wiley, New York, 1988), p. 211.
[9] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Winston, Washington D.C., 1977).
[10] H. J. Krappe and H. H. Rossner, Z. Phys. A **314**, 149 (1983).
[11] V. F. Turchin, V. P. Kozlov, and M. S. Malkevich, Usp. Fiz. Nauk **102**, 345 (1970) [Sov. Phys. Usp. **13**, 681 (1971)].
[12] H. J.Krappe, H. Massmann, and H. H. Rossner, in *Inverse Problems: An Interdisciplinary Study*, edited by P. C. Sabatier (Academic, London, 1987), p. 129.
[13] H. J. Krappe and H. H. Rossner, in *Advanced Methods in the Evaluation of Nuclear Scattering Data*, edited by R. Lipperheide, Lecture Notes in Physics Vol. 236 (Springer, Berlin, 1985), p. 242.
[14] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, NJ, 1977), Chap. 9.
[15] D. W. Marquardt, J. Soc. Ind. Appl. Math. **11**, 431 (1963).
[16] R. W. G. Wyckoff, *Crystal Structures* (Wiley, New York, 1963), Vol. 1.
[17] S. I. Zabinsky, J. J. Rehr, A. Ankudinov, R. Albers, and M. J. Eller, Phys. Rev. B **52**, 2995 (1995); A. L. Ankoudinov, Ph.D. thesis, University of Washington, 1996.
[18] Code ATOMS version 2.46 written by Bruce Ravel. This code is part of the UWXAFS analysis programs.