



Chapter 5.13

Received 26 October 2016
Accepted 22 April 2020**Keywords:** EXAFS equation; nonlinear
least-squares fitting.

Nonlinear least-squares fitting

Matthew Newville*

Center for Advanced Radiation Sources, The University of Chicago, Chicago, IL 60637, USA. *Correspondence e-mail:
newville@cars.uchicago.edu

Solutions to the EXAFS equation allow the determination of the structural parameters R , N and σ^2 from EXAFS data, provided that the scattering phase shifts, amplitudes and mean-free path terms are known. Even in the simplest form, the oscillatory nature of the EXAFS equation makes it intrinsically nonlinear in the structural parameters, so that nonlinear least-squares methods are needed to find the most likely set of values for the structural parameters to best fit the EXAFS data. Such analysis methods have a substantial literature, are used in a broad range of scientific fields and have been demonstrated to lead to robust results for EXAFS analysis problems. In addition to finding the best-fit values, the methods generally give reliable estimates for the uncertainty in each of the variable parameters, taking into account the correlations between parameters. Still, the use of these methods requires some care, including legitimate concerns over uniqueness and the possibility of finding ‘false minima’: solutions which match the data well but with parameter values that are not the correct solution. In this chapter, an overview of the practical application of these methods to EXAFS will be provided.

1. Nonlinearity of the EXAFS equation

Extended X-ray absorption fine structure (EXAFS) is an oscillatory function of photoelectron wavenumber k and interatomic distance R , with complex scattering factors that also depend on both k and R , as can be seen in a simple form of the EXAFS equation,

$$\chi(k) = \sum_i \frac{N_i f_i(k) \exp[-2R_i/\lambda(k)] \exp(-2k^2)\sigma_i^2}{kR_i^2} \times \sin[2kR_i + \delta_i(k)], \quad (1)$$

where the summation is over shells of atomic neighbours or, more generally, *scattering paths* of the emitted photoelectron. Even when considering only a single scattering path or nearest neighbour shell, the EXAFS function is obviously not linear in k or any simple power of k . In addition, each of $f(k)$, $\delta(k)$ and $\lambda(k)$ has a strong and nonlinear dependence on k . Having multiple terms in the summation will give overlapping contributions for all but the simplest of spectra, making any attempt to linearize the EXAFS equation impossible. Thus, in order to extract the structural parameters R , N and σ^2 from EXAFS data in general, nonlinear analysis methods are required. The inherent interdependence and correlation of the structural parameters N , R and σ^2 does complicate the analysis and can potentially lead to ambiguities when trying to find optimal parameter values from an EXAFS spectrum. Fortunately, a good deal of mathematical work has been performed to help find the best solutions to nonlinear fitting problems. In addition, knowledge about the nature of the spectra and our understanding of the structures of real solids and molecules can help avoid some of the pitfalls associated with data fitting.

Related chapters

Volume I: 2.1, 2.13, 4.6,
4.7, 5.7

That is, rather than being afraid of this confusion, ‘no matter how thick’ (Dylan, 1989), we can rely on nonlinear data-analysis methods that are well understood from a statistical framework, with robust and efficient algorithms that have been developed over the past several decades.

2. Data fitting for EXAFS

General data-fitting analysis starts with a parameterized model for a measured signal and seeks to find optimal values for these parameters. The most common and statistically justified way to express ‘optimal’ is to minimize the sum of squares of the difference between a parameterized model $y(\mathbf{x})$ and measured data z with N different samples,

$$\chi^2 = \sum_{j=1}^N \left[\frac{y_j(\mathbf{x}) - z_j}{\varepsilon_j} \right]^2, \quad (2)$$

where \mathbf{x} is a particular set of varying parameters and ε is the estimated uncertainty in the data, all of which have N observations.¹

Equation (2) describes data modelling in general, and there are several choices that must be made for each term in equation (2) for the particular case of analyzing EXAFS data. Briefly, we need to decide how to represent the data and model [such as deciding whether to fit $\chi(k)$, the Fourier-transformed $\chi(R)$ or something else], how many data points there are in that data form and what the uncertainty in that form of the data is.

While the model for the data will derive from the EXAFS equation, we have a choice of how to represent the data. As discussed elsewhere in this volume, EXAFS data are routinely transformed from X-ray energy E to wavenumber k , Fourier-transformed to R -space, and sometimes Fourier-filtered back from R -space to k -space using a limited R range. In principle, EXAFS data could be modelled in any of these spaces. In practice, EXAFS is limited in its useful R range, and the number of scattering paths grows exponentially with R (Zabinsky *et al.*, 1995). Therefore, when building a model using the EXAFS equation, we generally know how far in R that model extends. Using unfiltered k -space or E -space data gives no guarantee that the data are restricted to the same extent as the model. Fitting in R -space or filtered k -space is usually preferred as it allows us to select both the k and R range for the data being modelled and to filter out the parts of the signal that we do not model.

Because the Fourier transform is complex, the data in R -space and filtered k -space can be separated into pairs of real and imaginary components or magnitude and phase components. Some EXAFS analysis procedures (such as the log-ratio method) exploit this to separately analyze the magnitude and phase components of the data, recognizing that the structural parameters in equation (1) mainly affect either the magnitude (N , σ^2) or the phase (R). This separation is not complete (notably, the magnitude depends on R) and cannot disen-

tangle contributions from multiple paths or shells, but does mean that the correlation between R and N will be much smaller than the correlation between σ^2 and N .

Modern analysis are based on fitting EXAFS to models using first-principles calculations of EXAFS from computer codes such *FEFF* (Zabinsky *et al.*, 1995; Kas *et al.*, 2024), *GNXAS* (Filipponi *et al.*, 1995; Filipponi, Di Cicco *et al.*, 2024; Filipponi, Natoli *et al.*, 2024) or *EXCURVE* (Binsted & Hasnain, 1996; Feiters *et al.*, 2024). With these, there is no need to separately analyze the magnitude and phase terms or to distinguish single- and multiple-scattering paths. These tools calculate $\chi(k)$, including the R dependence of $f(k)$ and $\delta(k)$, and can include more complex disorder terms that mix amplitude and phase terms. By calculating $\chi(k)$ from a model structure or a set of scattering paths, one then can transform the model and data $\chi(k)$ in the same way either to R -space or filtered k -space for evaluation and fitting. When using such calculations for data modelling, there is little distinction between using filtered k -space and R -space, as both explicitly select a finite k and R range of the data and model to fit. Indeed, one can also use wavelet transforms for EXAFS analysis (Muñoz *et al.*, 2003; Funke *et al.*, 2005; Timoshenko & Kuzmin, 2009), which mix k -space and R -space. Of course, whether the data is represented in k -space, R -space, back-transformed k -space or ‘wavelet-transform space’, the uncertainty ε in these data for equation (2) needs to be in the same space as the data, which we will return to below.

The choice of different data spaces to use in the analysis requires us to carefully consider what the value of N should be in equation (2). While many discussions of data-analysis methods implicitly assume that all measurements are independent of one another, we should not necessarily make that assumption here. For one thing, the $\chi(k)$ data to be analyzed with the EXAFS equation have already been extracted from the raw signals measuring X-ray absorption, so some interpolation or resampling of the raw data will already have happened. Considering that the energy resolution of EXAFS measurements are typically in the 1 eV range, it is apparent that sampling $\mu(E)$ in 1 meV steps may allow a precise energy calibration but will not give 1000 times more independent measurements of near-neighbour distance compared with data sampled in 1 eV steps. In addition, while the raw signals for $\mu(E)$ may independently measure the X-ray absorption by the sample, they do not necessarily independently measure the local structure: $\mu(E)$ below the absorption edge is not sensitive to near-neighbour distance. The number of points to be used in equation (2) should account for the actual number of independent points in the EXAFS spectrum to be modelled.

The Fourier transforms used for EXAFS (see Bunker, 2024) typically pad the $\chi(k)$ data with zeros well past 50 \AA^{-1} or even 100 \AA^{-1} , so that the $\chi(R)$ from the Fourier transform is very finely spaced in R (say, at 0.03 \AA) with many more data points than would be present in the Fourier transform of the unpadding data. This padding smooths the $\chi(R)$, but it also greatly oversamples the measured data. Explicitly stating the k and R ranges for the actual data being modelled clarifies the information content of the data and the number of parameters

¹ Note that the χ^2 statistic is *not* the EXAFS χ ! (Bevington & Robinson, 2002).

that can be independently extracted from it. The degree of freedom or the number of independent points is given by (Shannon, 1949; Stern, 1993)

$$N_{\text{indep}} \simeq \frac{2\Delta k \Delta R}{\pi} + 1. \quad (3)$$

That is, the finite k and R range of the data each limit how many parameters can be reliably determined from an EXAFS spectrum. It is important to recognize that this limitation is not because the data have been Fourier-transformed, but rather is intrinsic to the nature of any oscillatory signal with limited extent. Because the measured EXAFS signal rarely persists above typical noise levels past 20 \AA^{-1} or past 10 \AA , the number of available parameters in an EXAFS spectrum is relatively small. It is also important to recall that the Fourier transform gives a complex quantity and both real and imaginary (or magnitude and phase) must be considered, even though many EXAFS works only present the magnitude for $\chi(R)$.

This discussion may suggest that zero-padding in the Fourier transform should be avoided or limited. Although it does oversample the EXAFS signal, zero-padding of the Fourier transform conveniently smooths the resulting $\chi(R)$, which can avoid problems of numerical stability during the fit. Instead, we can rescale the sum in equation (2) to be

$$\chi^2 = \frac{N_{\text{indep}}}{N} \sum_{j=1}^N \left[\frac{y_j(\mathbf{x}) - z_j}{\varepsilon_j} \right]^2, \quad (4)$$

which allows the use of finely spaced $\chi(R)$ for stability but does not overcount the number of independent samples.

The parameters \mathbf{x} used in the EXAFS model are traditionally the near-neighbour distance R , the coordination number N , the mean-square displacement σ^2 and the energy origin E_0 that defines $k = 0$ (see Newville, 2024), with the scattering amplitude $f(k)$, phase shift $\delta(k)$ and mean-free path $\lambda(k)$ assumed to be known for each path. For analyses with many scattering paths (either shells of different atom types or distances or multiple-scattering paths), the number of parameters that might be determined could easily grow to exceed N_{indep} . For highly disordered systems using σ^2 to represent disorder may not be sufficient, and other terms such as higher-order cumulants (Bunker, 1983) may be needed to parameterize the contribution from even a single path.

With even a modest number of shells or scattering paths, the limit of N_{indep} variables becomes critically important in EXAFS analysis. To handle this limit of extractable information, it is important to be able to reduce the number of completely independent variables. For example, one may simply assert that some values are known. It is often (but not always, especially when using multiple theoretical calculations) the case that the E_0 parameters for multiple paths can be set to the same value. In order to limit the number of independently varying values of R and N , one can often use knowledge or assumptions about the structure or chemical environment, such as a known crystal structure for the system, or the valence state determined by XANES and bond valence

theory (Brown & Altermatt, 1985). More generally, one may write the parameters in the EXAFS equation as mathematical expressions of a set of generalized variables (Newville, 2001), and therefore place complex constraints between the physical parameters, and limit the number of independent variables used in the fit. In a practical analysis, it can also be helpful to place upper and lower bounds on each of the variables (James & Roos, 1975; Newville, 2013) so as to prevent nonphysical or absurd values, such as requiring that all values for σ^2 be positive.

3. Nonlinear data-fitting methods

There are several iterative algorithms that find parameter values by minimizing χ^2 for nonlinear problems. For data-fitting problems it is helpful to select a method that considers not only the value of χ^2 itself, but also the full residual array, the scaled difference between the model and the data. Many such methods can also use the derivatives of this residual with respect to the variables in the fit, which can lead to more efficient and stable solutions.

One of the most commonly used algorithms for general-purpose curve-fitting, which is heavily used in EXAFS analysis, is the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963), which dynamically adjusts the importance of the Jacobian matrix (the first partial derivative of the residual with respect to the variables \mathbf{x} , which gives a ‘steepest-descent’ approach to a solution) and the Hessian matrix (the second partial derivative of the residual with respect to \mathbf{x} , which gives the curvature of the parameter space near the solution). This combination gives an efficient and robust method and has a few convenient features notable for EXAFS analysis. Firstly, analytic derivatives of the residual array with respect to the variables are not needed, and the relative scaling of the variables and the derivatives with respect to them are dynamically adjusted. Secondly, the evaluation of these partial derivatives near the best-fit solution can be easily converted to the covariance matrix, which gives both the uncertainties in the variables and the correlations between pairs of variables (Bevington & Robinson, 2002). One popular implementation of the Levenberg–Marquardt method is given in the *MINPACK* code (Moré *et al.*, 1980), which is used in many popular programming languages and environments.

Like most other optimization methods, the Levenberg–Marquardt algorithm finds a local solution for parameter values given a particular set of starting values, and cannot guarantee that there is not a better solution with very different parameter values. Finding the global minimum is, of course, a much harder task. For parameter spaces that are unusually complex or high-dimensional (that is, with a large number of variables), global optimization methods exist but are generally extremely slow compared with algorithms that find a local solution.

Fortunately for EXAFS analysis, the oscillatory nature of the signal and the inherent understanding that interatomic distances tend to be between 1 and 5 Å (coincidentally, the

range to which EXAFS is most sensitive) tends to mean that becoming stuck in false minima is not as common as one might fear. Indeed, for many EXAFS curve-fitting problems one has a good idea of the plausible values and ranges for parameter values and can place constraints and bounds on them. In addition, when using first-principles calculations in the analysis, one starts with a model structure that is expected to be close to the unknown structure, so that the nonlinear fit is a refinement rather than a global search for the best values. Still, problems can occur, especially for cases of multiple neighbouring atoms at similar distances, and it can be important to compare the results from many initial but reasonable starting values and to check that fits converge to consistent results that are sensible with as many tools as are available.

4. Fit statistics and parameter uncertainties

For any analysis of scientific data it is necessary to provide reliable statistics about the quality of the fit and to provide estimates for the uncertainties and interdependence of the variable parameters. The χ^2 statistic from equation (2) is clearly an important measure of the quality of an individual fit and is widely used in the statistical treatment of many kinds of data.

The definition of χ^2 scales with the uncertainty in the data, ε , which has not yet been discussed. There are many possible strategies for estimating uncertainties in EXAFS data, usually based on the statistical treatment of many repeated measurements. These can be very useful, but may not be readily applied to every EXAFS spectrum. A simple approach that can be applied to each spectrum relies on the observation that the EXAFS signal decays rapidly with R . Thus, the portion of the data at very high R (say above 15 Å) will likely be dominated by noise. This assumes that the noise is independent of R (white noise), which is not easy to verify for a single spectrum. Tests and experience have shown that it gives a reasonable estimate for data of low-to-normal quality, but generally underestimates the noise level for very good data. Fourier analysis and Parseval's theorem can be used to give a simple relationship between ε_R , the noise estimate in $\chi(R)$, and ε_k , the noise in $\chi(k)$ (Newville *et al.*, 1999). If properly scaled, a good fit should have a value for χ^2 of around $N_{\text{indep}} - N_{\text{varys}}$, where N_{varys} is the number of variables in the fit. In practice this is rarely the case, which we will return to below.

One should expect that adding more variable parameters should improve the fit, so it is useful to have statistics that reflect both the fit quality and the number of free parameters. The simplest of these is the reduced χ^2 , which simply scales χ^2 by the number of free parameters,

$$\chi_v^2 = \chi^2 / (N_{\text{indep}} - N_{\text{varys}}), \quad (5)$$

where N_{varys} is the number of variables in the fit. In principle, a good fit with properly estimated uncertainties should give $\chi_v^2 \simeq 1$. While the χ^2 statistic gives a useful measure of the absolute fit quality, χ_v^2 is more readily used to compare fits with different numbers of variables, thereby helping to decide

whether adding a particular variable to a fit is justified or not. Both χ^2 and χ_v^2 are recommended for all reported EXAFS analyses (Lytle *et al.*, 1989). Other statistics such as the Akaike information criterion (Akaike, 1974), which can be related to χ^2 by

$$\text{aic} = N \ln(\chi^2) + 2N_{\text{varys}}, \quad (6)$$

are very useful for comparing fits with different numbers of variables. In addition, following the common practice for X-ray diffraction data, it is common to report an \mathcal{R} factor that scales the misfit to the magnitude of the data itself instead of the estimated uncertainty in the data,

$$\mathcal{R} = \sum_j [y_j(\mathbf{x}) - z_j]^2 / \sum_j z_j^2, \quad (7)$$

where, as in equation (2), $y(\mathbf{x})$ is the parametrized model and z are the measured data. This statistic has the advantage of a simple interpretation as a fractional misfit, independent of the uncertainty in the data itself, and is typically found to be below 0.05 or better for good fits.

A few methods can be used to estimate uncertainties in the parameter values around the best-fit values and to determine the interdependence or correlation of parameters. Generally, the uncertainty for a parameter is assigned as the change from its optimal value for which χ^2 is increased by 1 from its best value. This gives so-called 1σ uncertainties, which give a confidence level of 0.683, and are most commonly used in presenting results from EXAFS analysis. For higher confidence, 3σ uncertainties (with a confidence level of 0.997) can be reported. To estimate the uncertainty, it is important to allow all other variables to be re-optimized in response to the changing value. The correlations between variables measures how much and in what direction the other variables respond to changing the value of the parameter in question.

As mentioned above, a fortunate result of the Levenberg–Marquardt algorithm is that the partial derivatives used to find the solution for the variables can be used to build the covariance matrix, from which the correlations between pairs of variables and the uncertainties in the variables that take those correlations into account can readily be extracted. The diagonal elements of the covariance matrix give the uncertainties, while the off-diagonal elements give the correlations between variables. This approach does assume that the fit becomes worse symmetrically and quadratically as each parameter moves away from its optimal value and that the correlations between variables are also symmetric. To test these assumptions, a more detailed exploration of the parameter values near the best solution can be made either by brute force trying multiple values or by using Monte Carlo methods. For most EXAFS problems, it is generally found that the assumptions of simple and symmetric deviations from the optimal solution are pretty good (Curis & Bénazeth, 2000) and that the estimates from the Levenberg–Marquardt method are of the correct scale. Still, it must be acknowledged that this may not always be the case. As with the best-fit values themselves, the uncertainties from any single fit should be tested by running multiple fits with small variations in starting values, data-

processing parameters, data ranges and Fourier transform parameters. This practice can help to identify unreliable results, ‘false minima’, and give more confidence in the values for uncertainties.

Another important caveat is that for most EXAFS analyses using theoretical standards and reasonably high-quality data, χ^2 is much larger than $N_{\text{indep}} - N_{\text{varys}}$. This is due in part to systematic errors that are not accounted for in ε as estimated from white noise and in part to due to incomplete treatment in the calculations of the scattering factors. As mentioned above, this means even the best fits to real EXAFS data often exceed $\chi_v^2 \simeq 1$ by an order of magnitude or more. With χ^2 much larger than $N_{\text{indep}} - N_{\text{varys}}$, increasing this value by 1 will give very small estimates of the parameter uncertainties. It is common in the EXAFS literature and recommended to report parameter values that increase χ^2 by χ_v^2 as the 1σ uncertainties (Lytle *et al.*, 1989). This effectively asserts that the fit is ‘good’ and that ε should be adjusted so that χ_v^2 is 1. While this clearly leaves open questions about why the uncertainty in the data is underestimated, it does produce more realistic values for the parameter uncertainties.

EXAFS analysis has benefited greatly from a variety of well proven general statistical treatments and data-analysis methods. The description here attempts to capture the most commonly used and best practices. There is no doubt further room for improvement and the application of new and more refined approaches to data analysis in the years ahead.

References

- Akaike, H. (1974). *IEEE Trans. Autom. Contr.* **19**, 716–723.
- Bevington, P. R. & Robinson, D. K. (2002). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw–Hill.
- Binsted, N. & Hasnain, S. S. (1996). *J. Synchrotron Rad.* **3**, 185–96.
- Brown, I. D. & Altermatt, D. (1985). *Acta Cryst.* **B41**, 244–247.
- Bunker, G. (1983). *Nucl. Instrum. Methods Phys. Res.* **207**, 437–444.
- Bunker, G. (2024). *Int. Tables Crystallogr. I*, ch. 5.3, 639–644.
- Curis, E. & Bénazeth, S. (2000). *J. Synchrotron Rad.* **7**, 262–266.
- Dylan, B. (1989). *Most of the Time*. Special Rider Music.
- Feiters, M. C., Strange, R. W. & Binsted, N. (2024). *Int. Tables Crystallogr. I*, ch. 6.5, 744–751.
- Filipponi, A., Di Cicco, A. & Natoli, C. R. (1995). *Phys. Rev. B*, **52**, 15122–15134.
- Filipponi, A., Di Cicco, A. & Natoli, C. R. (2024). *Int. Tables Crystallogr. I*, ch. 6.12, 787–790.
- Filipponi, A., Natoli, C. R. & Di Cicco, A. (2024). *Int. Tables Crystallogr. I*, ch. 6.11, 782–786.
- Funke, H., Scheinost, A. & Chukalina, M. (2005). *Phys. Rev. B*, **71**, 094110.
- James, F. & Roos, M. (1975). *Comput. Phys. Commun.* **10**, 343–367.
- Kas, J. J., Vila, F. D. & Rehr, J. J. (2024). *Int. Tables Crystallogr. I*, ch. 6.8, 764–769.
- Levenberg, K. (1944). *Q. Appl. Math.* **2**, 164–168.
- Lytle, F. W., Sayers, D. E. & Stern, E. A. (1989). *Physica B*, **158**, 701–722.
- Marquardt, D. (1963). *J. Soc. Ind. Appl. Math.* **11**, 431–441.
- Moré, J. J., Garbow, B. S. & Hillstom, K. E. (1980). *User Guide for MINPACK-1*. Argonne National Laboratory Technical Report ANL-80-74.
- Muñoz, M., Argoul, P. & Farges, F. (2003). *Am. Mineral.* **88**, 694–700.
- Newville, M. (2001). *J. Synchrotron Rad.* **8**, 96–100.
- Newville, M. (2013). *J. Phys. Conf. Ser.* **430**, 012007.
- Newville, M. (2024). *Int. Tables Crystallogr. I*, ch. 5.1, 631–635.
- Newville, M., Boyanov, B. I. & Sayers, D. E. (1999). *J. Synchrotron Rad.* **6**, 264–265.
- Shannon, C. E. (1949). *Proc. Inst. Radio Eng.* **37**, 10–21.
- Stern, E. A. (1993). *Phys. Rev. B*, **48**, 9825–9827.
- Timoshenko, J. & Kuzmin, A. (2009). *Comput. Phys. Commun.* **180**, 920–925.
- Zabinsky, S. I., Rehr, J. J., Ankudinov, A., Albers, R. C. & Eller, M. J. (1995). *Phys. Rev. B*, **52**, 2995–3009.