Journal of
Applied
Crystallography
ISSN 0021-8898

Received 1 February 2001
Accepted 9 March 2001

Background estimation using a robust Bayesian analysis

W. I. F. David* and D. S. Sivia

ISIS Facility, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England. Correspondence e-mail: bill.david@rl.ac.uk

A novel method for the estimation of the background in a powder diffraction pattern has been developed using a robust Bayesian analysis. In formulating a probabilistic approach to background fitting, the diffraction peaks are considered to be nuisance data that must be taken into account. The underlying probability theory is discussed in terms of going beyond the Gaussian approximation normally associated with counting statistics and least-squares analysis. Various examples are presented that illustrate the general applicability of this approach.

1. Introduction

The successful removal of a background signal from a diffraction pattern without the availability of a suitable calibration measurement relies heavily on the ability to distinguish between broad and sharp features. While a procedure for the simultaneous estimation of the slowly varying contribution and deconvolution of the sharp structure has been put forward (Sivia, 1990), most of the work in this area has focused solely on the extraction of the background signal. One of the early proposals within crystallography was proffered by Steenstrup (1981), who developed an iterative procedure for fitting a low-order polynomial to the data using a least-squares analysis. The success of this method hinged on the successive removal of more and more data that were deemed to include a significant contribution from a Bragg peak. Other approaches have been suggested and a recent ingenious algorithm developed by Brückner (2000) that is based upon an adaptive low-pass filter is worthy of mention.

In this paper, a more formal probabilistic approach to the problem is developed, which correctly takes into account the error-bar on each data point while allowing for an unknown contribution from a Bragg peak that may or may not be present at that point in the diffraction pattern. This approach leads to a functional minimization that deviates from least-squares analysis in an asymmetrical manner. It is not iterative in the Steenstrup (1981) sense because no data are removed from the analysis. Indeed, the principal appeal of this algorithm is that all the measurements are treated on an equal footing. The approach developed in this paper bears a close similarity to the one described by Fischer *et al.* (2000) and, though not as rigorous in the full optimization of the ideal background parameterization, it is, we believe, more intuitively accessible.

2. Dealing with outlier data: beyond least squares

In a powder diffraction experiment, the uncertainty associated with a single observation in a diffraction pattern is generally considered to result solely from Poisson counting statistics. If more than around 20 counts have been measured at a particular point, then the Gaussian approximation to counting statistics is quite sufficient and least-squares analysis is the appropriate minimization procedure for fitting the diffraction pattern. [The issue of dealing with low count rates (<20 counts per point) has been treated elsewhere (Antoniadis *et al.*, 1990) and will not be discussed in this paper.] Least-squares minimization is often taken to be a basic tenet of data analysis. The almost universal use of least-squares analysis is because the underlying statistical assumption is a very general one and has its origins in the assumption that the uncertainties associated with an observation follow a Gaussian probability distribution. In other words, the observation of a data point, D, with an error bar, σ, can be stated mathematically in terms of a likelihood that follows a Gaussian probability distribution function:

$$p(D|\mu,\sigma) = [1/\sigma(2\pi)^{1/2}]\exp[-(D-\mu)^2/2\sigma^2]. \qquad (1)$$

The quoted error bar is usually based upon ideal conditions and in the case of a powder diffraction pattern is generally associated with counting statistics alone. Occasionally, however, rogue data points resulting from, for example, detector problems can occur. Based upon a Gaussian probability distribution derived solely from counting statistics, these points must be extremely unlikely. However, their very appearance testifies to the fact that conditions other than just counting statistics are causing uncertainties in the data value. When this occurs, an alternative approach would be to state that the error bar, σ, is a lower bound, $\sigma_{\min}$, on the uncertainty