STRUCTURAL BIOLOGY

ISSN 2059-7983

Evaluating crystallographic likelihood functions using numerical quadratures

Petrus H. Zwart[a,b]* and Elliott D. Perryman[a,b,c]

[a]Center for Advanced Mathematics in Energy Research Applications, Computational Research Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA, [b]Molecular Biophysics and Integrative Bioimaging Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA, and [c]The University of Tennessee at Knoxville, Knoxville, TN 37916, USA. *Correspondence e-mail: phzwart@lbl.gov

Received 13 January 2020
Accepted 23 June 2020

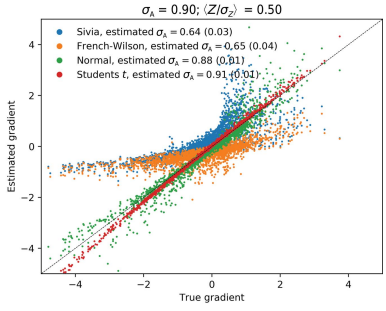
Edited by R. J. Read, University of Cambridge, United Kingdom

Keywords: maximum likelihood; refinement; numerical integration.

Intensity-based likelihood functions in crystallographic applications have the potential to enhance the quality of structures derived from marginal diffraction data. Their usage, however, is complicated by the ability to efficiently compute these target functions. Here, a numerical quadrature is developed that allows the rapid evaluation of intensity-based likelihood functions in crystallographic applications. By using a sequence of change-of-variable transformations, including a nonlinear domain-compression operation, an accurate, robust and efficient quadrature is constructed. The approach is flexible and can incorporate different noise models with relative ease.

1. Introduction

The estimation of model parameters from experimental observations plays a central role in the natural sciences, and the use of likelihood-based methods has been shown to yield robust estimates of 'best guess' values and their associated confidence intervals (Rossi, 2018). Maximum-likelihood estimation goes back to sporadic use in the 1800s by Gauss (Gauss, 1809, 1816, 1823) and Hagen (1867), and was further developed by Fisher (1915), Wilks (1938), and Neyman and Pearson (Neyman & Scott, 1948; Pearson, 1970). In the crystallographic community, Beu *et al.* (1962) were the first to explicitly use maximum-likelihood estimation, applying it to lattice-parameter refinement in powder diffraction. In a late reaction to this work, Wilson (1980) states that 'the use of maximum likelihood is unnecessary, and open to some objection', and subsequently recasts the work of Beu *et al.* (1962) into a more familiar least-squares framework. It is important to note that least-squares estimation methods are equivalent to a likelihood formalism under the assumption of normality of the random variables. The use of maximum-likelihood-based methods using non-normal distributions in structural sciences took off after making significant impacts in the analysis of macromolecules. For these types of samples, structure solution and refinement problems were often problematic owing to very incomplete or low-quality starting models, making standard least-squares techniques underperform. In the 1980s and 1990s, likelihood-based methods became mainstream, culminating in the ability to routinely determine and refine structures that were previously thought to be problematic (Lunin & Urzhumtsev, 1984; Read, 1986; Bricogne & Gilmore, 1990; de La Fortelle & Bricogne, 1997;



Pannu & Read, 1996; Murshudov *et al.*, 1997). A key ingredient to this success was the development of cross-validation techniques to reduce bias in the estimation of hyper-parameters that govern the behavior of the likelihood functions (Lunin & Skovoroda, 1995; Pannu & Read, 1996). At the beginning of the 21st century, Read and coworkers extended the likelihood formalism to molecular-replacement settings as well, resulting in a significant improvement in the ability to solve structures from marginal starting models (McCoy *et al.*, 2005; Storoni *et al.*, 2004; Read, 2001). The first use of approximate likelihood methods for the detection of heavy atoms from anomalous or derivative data originates from Terwilliger & Eisenberg (1983), who used an origin-removed Patterson correlation function for substructure solution. This approach was shown by Bricogne (1997) to be equivalent to a second-order approximation of a Rice-based likelihood function. A more recent development is the use of a more elaborate likelihood formalism in the location of substructures (Bunkóczi *et al.*, 2015), showing a dramatic improvement in the ability to locate heavy atoms. In density modification, the use of the likelihood formalism has significantly increased its radius of convergence (Terwilliger, 2000; Cowtan, 2000; Skubák *et al.*, 2010).

As the above examples illustrate, impressive progress has been made by the application of likelihood-based methods to a wide variety of crystallographic problems. In all of the described scenarios, key advances were made by deriving problem-specific likelihood functions and applying them to challenging structure-determination problems. In the majority of these cases, a thorough treatment of experimental errors has only a secondary role, resulting in approximations that work well in medium- or low-noise settings. The principal challenge in the handling of random noise in crystallographic likelihood functions is how to efficiently convolve Rice-like distribution functions modeling the distribution of a structure factor from an incomplete model with errors with the appropriate distribution that models the experimental noise. In this manuscript, we develop quadrature approaches to overcome these difficulties. We accomplish this by using a sequence of changes of variables that are amenable to straightforward numerical integration using standard methods. The approach derived has direct applications in model refinement and molecular replacement, while the general methodology can also be extended to other crystallographic scenarios. In the remainder of this paper, we will provide a general introduction to likelihood-based methods, provide a relevant background into numerical integration techniques, develop an adaptive quadrature approach, apply it to Rice-type likelihood functions and validate its results.

1.1. Maximum-likelihood formalism

The estimation of model parameters $\boldsymbol{\theta}$ given some data set $\mathcal{X} = \{x_1, \ldots, x_j, \ldots, x_N\}$ via the likelihood formalism is performed in the following manner. Given the probability density function (PDF) $f(x_j|\boldsymbol{\theta})$ of a single observation x_j given a model parameter $\boldsymbol{\theta}$, the joint probability of the entire data set is, under the assumption of independence of the observations, equal to the product of the individual PDFs,

$$f(\mathcal{X}|\boldsymbol{\theta}) = \prod_{j=1}^{N} f(x_j|\boldsymbol{\theta}). \tag{1}$$

The probability of the data $\mathcal{X}$ given the model parameters $\boldsymbol{\theta}$ is known as the likelihood of the model parameters given the data:

$$L(\boldsymbol{\theta}|\mathcal{X}) = f(\mathcal{X}|\boldsymbol{\theta}). \tag{2}$$

A natural choice for the *best estimate* of the model parameters is obtained by finding the $\boldsymbol{\theta}$ that maximizes the likelihood function. This choice is called the maximum-likelihood estimate (MLE). The likelihood function itself $L(\boldsymbol{\theta}|\mathcal{X})$ can be seen as a probability distribution, allowing one to obtain confidence-limit estimates on the MLE (Rossi, 2018). The determination of the MLE is typically performed by optimizing the log-likelihood:

$$\ln L(\boldsymbol{\theta}|\mathcal{X}) = \sum_{j=1}^{N} \ln f(x_j|\boldsymbol{\theta}). \tag{3}$$

Often, the distribution needed for the likelihood function has to be obtained via a process known as marginalization. During this integration, a so-called nuisance parameter is integrated out,

$$f(x|\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f(x, y|\boldsymbol{\theta})\,\mathrm{d}y, \tag{4}$$

where, under the assumption of conditional independence,

$$f(x, y|\boldsymbol{\theta}) = f(x|\boldsymbol{\theta})f(y|x, \boldsymbol{\theta}). \tag{5}$$

Depending on the mathematical form of the distributions involved, this marginalization can range from a trivial analytical exercise to a numerically challenging problem. In likelihood functions in a crystallographic setting, this marginalization is required to take into account the effects of experimental noise, and its efficient calculation is the focus of this communication.

1.2. Motivation

The most common likelihood function used in crystallographic applications specifies the probability of the *true* structure-factor amplitude given the value of a calculated structure factor originating from a model with errors (Sim, 1959; Srinivasan & Parthasarathy, 1976; Luzzati, 1952; Woolfson, 1956; Lunin & Urzhumtsev, 1984):

$$f_{\mathrm{a}}(F|F_{\mathrm{C}}, \alpha, \beta) = \frac{2F}{\varepsilon\beta}\exp\left(-\frac{F^2 + \alpha^2 F_{\mathrm{C}}^2}{\varepsilon\beta}\right) I_0\left(\frac{2\alpha F F_{\mathrm{C}}}{\varepsilon\beta}\right), \tag{6}$$

$$f_{\mathrm{c}}(F|F_{\mathrm{C}}, \alpha, \beta) = \left(\frac{2}{\varepsilon\pi\beta}\right)^{1/2}\exp\left(-\frac{F^2 + \alpha^2 F_{\mathrm{C}}^2}{2\varepsilon\beta}\right)\cosh\left(\frac{2\alpha F F_{\mathrm{C}}}{2\varepsilon\beta}\right). \tag{7}$$

f_{a} and f_{c} are the distributions for acentric and centric reflections (the so-called Rice distribution), ε is a symmetry-enhancement factor, F is the true structure-factor amplitude

and F_C is the current model structure-factor amplitude, while α and β are likelihood distribution parameters (Lunin & Urzhumtsev, 1984). For the refinement of atomic models given experimental data, the likelihood of the model-based structure-factor amplitudes given the experimental data is needed and can be obtained from a marginalization over the unknown, error-free structure-factor amplitude. Following Pannu & Read (1996) and assuming conditional independence between the distributions of the experimental intensity I_o and amplitude F, we obtain

$$L(F_C|I_o) = f(I_o|F_C, \alpha, \beta, \sigma_I^2) = \int_0^\infty f(I_o|\sigma_I^2, F) f(F|F_C, \alpha, \beta)\, dF, \quad (8)$$

where $f(F|F_C, \alpha, \beta)$ is given by expressions (6) or (7) and $f(I_o|\sigma_I^2, F)$ is equal to a normal distribution with mean F^2 and variance σ_I^2. This integral is equivalent to the MLI target function derived by Pannu & Read (1996). Because there is no fast-converging series approximation or simple closed-form analytical expression for this integral, various approaches have been developed, as excellently summarized by Read & McCoy (2016), including a method-of-moments-type approach to find reasonable analytical approximations to the intensity-based likelihood function.

In this work, we investigate the use of numerical integration methods to obtain high-quality approximations of integral (8) while also taking into account uncertainties in the estimated standard deviation. The approach outlined above, in which a Rice function is convoluted with a Gaussian distribution, essentially assumes that the standard deviation of the mean is known exactly. Given that both the standard deviation and the mean are derived from the same experimental data, this assumption is clearly suboptimal, especially when the redundancy of the data is low. In order to take into account possible errors in the observed standard deviation, we will use a *t*-distribution instead of a normal distribution, which arises as the distribution choice when the true standard deviation is approximated by an estimate from experimental data (Student, 1908). The aim of this work is to derive an efficient means of obtaining target functions that can provide an additional performance boost when working with very marginal data, such as those obtained from time-resolved serial crystallography or free-electron laser data, in which the experimental errors are typically larger than those obtained using standard rotation-based methods or have nonstandard error models (Brewster *et al.*, 2019). Furthermore, high-quality data sets are rarely resolution-limited by the diffraction geometry alone, indicating that many more marginal data are readily available that can potentially increase the quality of the final models if appropriate target functions are used. In the remainder of this manuscript, we develop and compare a number of numerical integration schemes aimed at rapidly evaluating intensity-based likelihood functions and their derivatives that take into account the presence of experimental errors, both in the mean intensity and in its associated standard deviation.

2. Methods

In order to evaluate a variety of numerical integration schemes and approximation methods, the equations are first recast into a normalized structure-factor amplitudes E and normalized intensities Z_o framework, with the use of the σ_A formulation of the distributions involved, assuming a $P1$ space group, such that $\varepsilon = 1$ (Read, 1997). The joint probability distribution of the error-free structure-factor amplitude E and the experimental intensity Z_o, given the calculated normalized structure factor E_C, the model-quality parameter σ_A, the estimated standard deviation of the observation σ_Z and the effective degrees of freedom ν, reads

$$f_a(E, Z_o|E_C, \sigma_A, \sigma_Z^2, \nu) = \frac{2E}{1-\sigma_A^2} \exp\left(-\frac{E^2+\sigma_A^2 E_C^2}{1-\sigma_A^2}\right) \times I_0\left(\frac{2\sigma_A E E_C}{1-\sigma_A^2}\right) f(Z_o|E, \sigma_Z^2, \nu) \quad (9)$$

for acentric reflections and

$$f_c(E, Z_o|E_C, \sigma_A, \sigma_Z^2, \nu) = \left[\frac{2}{\pi(1-\sigma_A^2)}\right]^{1/2} \exp\left(-\frac{E^2+\sigma_A^2 E_C^2}{2-2\sigma_A^2}\right) \times \cosh\left(\frac{\alpha E E_C}{1-\sigma_A^2}\right) f(Z_o|E, \sigma_Z^2, \nu) \quad (10)$$

for centric reflections. When the distribution of the observed mean intensity Z_o is modeled by a *t*-distribution (Student, 1908) with a location parameter equal to E^2, we have

$$f(Z_o|E, \sigma_Z^2, \nu) = \Gamma\left(\frac{\nu+1}{2}\right)\left(\frac{1}{\nu\pi\sigma_Z^2}\right)^{1/2}\left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-1} \times \left[1+\frac{(Z_o-E^2)^2}{\nu\sigma_Z^2}\right]^{-(\nu+1)/2}, \quad (11)$$

where ν is the effective degrees of freedom of the observation, which is related to the effective sample size $N_{\rm eff}$,

$$\nu = N_{\rm eff} - 1. \quad (12)$$

The effective sample size can be taken as the redundancy of an observed intensity, or can be estimated during data processing using the Welch–Satterthwaite equation (Welch, 1947) to take into account the weighting protocols implemented in data processing (Brewster *et al.*, 2019). The *t*-distribution arises as the distribution of choice given a sample mean and sample variance from a set of observations (Student, 1908). The use of a normal distribution essentially assumes no uncertainty in the variance σ_Z^2, but only in the observed mean Z_o. The *t*-distribution is similar to a normal distribution, but has heavier tails and therefore will be expected to result in likelihood functions that are less punitive to larger deviations between observed and model intensities. When ν tends to infinity, the above distribution converges to a normal distribution,

$$f(Z_o|E^2, \sigma_o^2) = \frac{1}{\left(2\pi\sigma_Z^2\right)^{1/2}} \exp\left[-\frac{(Z_o-E^2)^2}{2\sigma_Z^2}\right]. \quad (13)$$

The above joint probability distributions need to be marginalized over E in $\mathbb{R}^+$ to obtain the distribution of interest:

$$f(Z_o|E_C, \sigma_A, \sigma_Z^2, \nu) = \int_0^\infty f(E, Z_o|E_C, \sigma_A, \sigma_Z^2, \nu)\,\mathrm{d}E. \quad (14)$$

2.1. Variance inflation

A common approach to avoid performing the integration specified above is to inflate the variance of the Rice function ($1 - \sigma_A^2$) by the variance of the 'observed structure-factor amplitude', yielding ($1 - \sigma_A^2 + \sigma_E^2$) (Green, 1979). This approach circumvents the need to perform an integration, but is suboptimal in a number of different ways. Because we do not observe amplitudes, we are required to estimate the amplitude and its variance from observed intensity data. A common way to perform the intensity-to-amplitude conversion is via a Bayesian estimate (French & Wilson, 1978) under the assumption of a uniform distribution of atoms throughout the unit cell. Although this so-called Wilson prior is used in most cases, a slightly different result can be obtained when using a constant, improper prior on the possible values of the structure-factor amplitudes on the positive half-line (Sivia & David, 1994). This results in an intensity-to-amplitude conversion that does not rely on the accurate estimation of the mean intensity, possibly complicated by the effects of pseudo-symmetry, diffraction anisotropy or twinning:

$$E_o = \left\{\frac{1}{2}\left[Z_o + (Z_o^2 + 2\sigma_Z^2)^{1/2}\right]\right\}^{1/2}, \quad (15)$$

$$\sigma_E^2 = \frac{\sigma_Z^2}{4\left(Z_o^2 + 2\sigma_Z^2\right)^{1/2}}. \quad (16)$$

Further details are given in Appendix *E*. While this procedure allows a straightforward intensity-to-amplitude conversion, even when intensities are negative, and can subsequently be used to inflate the variance of the Rice function, it is no substitute for the full integration. Given the simplicity of the variance-inflation approach and its wide usage in a number of crystallographic applications, we will use this approach as a benchmark, using conversion schemes based both on the Wilson prior (denoted French–Wilson) and on the outlined uniform, non-informative prior (denoted Sivia).

2.2. Approaches to numerical integration

Several conventional numerical integration approximations exist for improper integrals such as expression (8). Standard methods include trapezoidal-based methods with a truncated integration range, the use of Laplace's method, Monte Carlo-based methods or approaches based on orthogonal polynomials (Davis & Rabinowitz, 1984). Whereas a straightforward use of a trapezoidal integration scheme is tempting, the shape of the integrand for certain combinations of distribution parameters will result in a fair chance of missing the bulk of the mass of the function unless a very fine sampling grid is used. When using the Laplace approximation, in which the integrand is approximated by an appropriately scaled and translated Gaussian function, the integrand can deviate significantly from a Gaussian, also resulting in a poor performance. These challenges are summarized in Fig. 1, where a number of typical integrand shapes are visualized for different parameter choices. A number of numerical integration and approximation methods will be outlined below, including a discussion of how *ground truth* is established as a basis for comparison. Here, we will limit ourself to the Laplace approximation owing to its simplicity and the trapezoidal rules because of their excellent convergence properties when applied to analytic functions on the real line and their close relation to classical Gauss quadratures (Trefethen & Weideman, 2014). The use of (quasi) Monte Carlo schemes will not be considered, since these methods are typically used as a 'method of last resort' for very high dimensional integrals (Cools, 2002).

2.3. Change of variables and the Laplace approximation

Analytical and numerical integration is often greatly simplified by a change of variables of the integrand (Davis & Rabinowitz, 1984). The change-of-variable theorem relates the integral of some function $h(u)$ under a change of variables $u = \psi(x)$,

$$\int_a^b h(u)\,\mathrm{d}u = \int_{\psi^{-1}(a)}^{\psi^{-1}(b)} h[\psi(x)]\frac{\mathrm{d}\psi(x)}{\mathrm{d}x}\,\mathrm{d}x. \quad (17)$$

The modified shape of the integrand by a change of variables makes the use of the so-called Laplace approximation appealing. In a Laplace approximation, the integrand is approximated by a scaled squared exponential function with a suitably chosen mean and length scale (Peng, 2018). The Laplace approximation can be derived from truncated Taylor expansion of the logarithm of the integrand:

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &= \int_a^b \exp[g(x)]\,\mathrm{d}x \\ &= \int_a^b \exp\left[\sum_{n=0}^{\infty}\frac{g^{(n)}(x_0)}{n!}(x - x_0)^n\right]\mathrm{d}x \\ &\simeq \int_{-\infty}^{\infty} \exp[g(x_0)]\exp\left[\tfrac{1}{2}g''(x_0)(x - x_0)^2\right]\mathrm{d}x, \end{aligned} \quad (18)$$

where $g(x) = \ln[f(x)]$ and x_0 is the location of the maximum of $f(x)$, implying that $g''(x_0) = 0$. Note that in the last step in equation (18) the assumption is made that $f(x)$ goes to 0 when not near x_0 quickly enough that integrating over $[a, b]$ yields the same results as integrating over $\mathbb{R}$. Although this approximation does not work for all possible choices of $g(x)$, it has proven to be a successful tool in marginalizing distributions in Bayesian analysis (Kass & Steffey, 1989) and crystallographic applications (Murshudov *et al.*, 2011).

The above expression thus yields

$$\int_a^b f(x)\,\mathrm{d}x \simeq f(x_0)\left[\frac{2\pi}{-g''(x_0)}\right]^{1/2}. \quad (19)$$

The effectiveness of this approximation hinges on the location of x_0 (it should be contained within the original integration domain), the magnitude of $g''(x_0)$ and how rapidly higher-order derivatives of $g(x)$ vanish around x_0. The change-of-

variable strategy outlined above can aid in increasing the performance of approximation to expression (8).

2.4. Quadrature methods

Even though the change-of-variables approach combined with the Laplace approximation has the potential to yield accurate integral approximations, obtaining reasonable estimates of the derivative of the log-likelihood, as needed for difference maps or for first or higher-order optimization methods, seems less straightforward using the Laplace approach. The difficulty arises from the need to obtain the derivative of the location of the maximum of integrand, as this value is a function of the variables for which derivatives are computed. In addition, the introduction of *t*-based noise models introduces heavy tails in the distribution for which Gaussian approximations can have a poor performance. For this reason, the use of a quadrature approach is of interest. In a numerical integration with a quadrature, the integral of interest is approximated by a weighted sum of selected function values. The Laplace approximation outlined above can thus be seen as a one-point quadrature, where the location of the function value is located at the maximum of the integrand, and the associated weight is derived from a local Gaussian fit to the integrand. An expanded quadrature approach provides an easy way to increase the precision of the integral by increasing the number of sampling points, but also circumvents issues with computing derivatives of the location of the maximum of the integrand that are encountered when using the Laplace approximation. Quadrature approaches have, however, been assumed to need a large number of terms to obtain sufficient precision (Read & McCoy, 2016), possibly making them an unattractive target for practical crystallographic applications. In order to circumvent or at least ameliorate these issues, we design quadratures that combine the benefits of a Laplace approximation and basic numerical quadratures (Appendix *A*).

A high-level overview of our integration approach is depicted in Fig. 2. By combining a power transform followed by a hyperbolic transform of the integrand, we transform the integration domain from [0, ∞] onto [0, 1]. While the first power transform (Appendix *C*) allows the integrand to have more Gaussian-like character, the second change-of-variable operation nonlinearly compresses low-mass regions onto relatively small line segments, while approximately linearly transforming high-mass areas of the

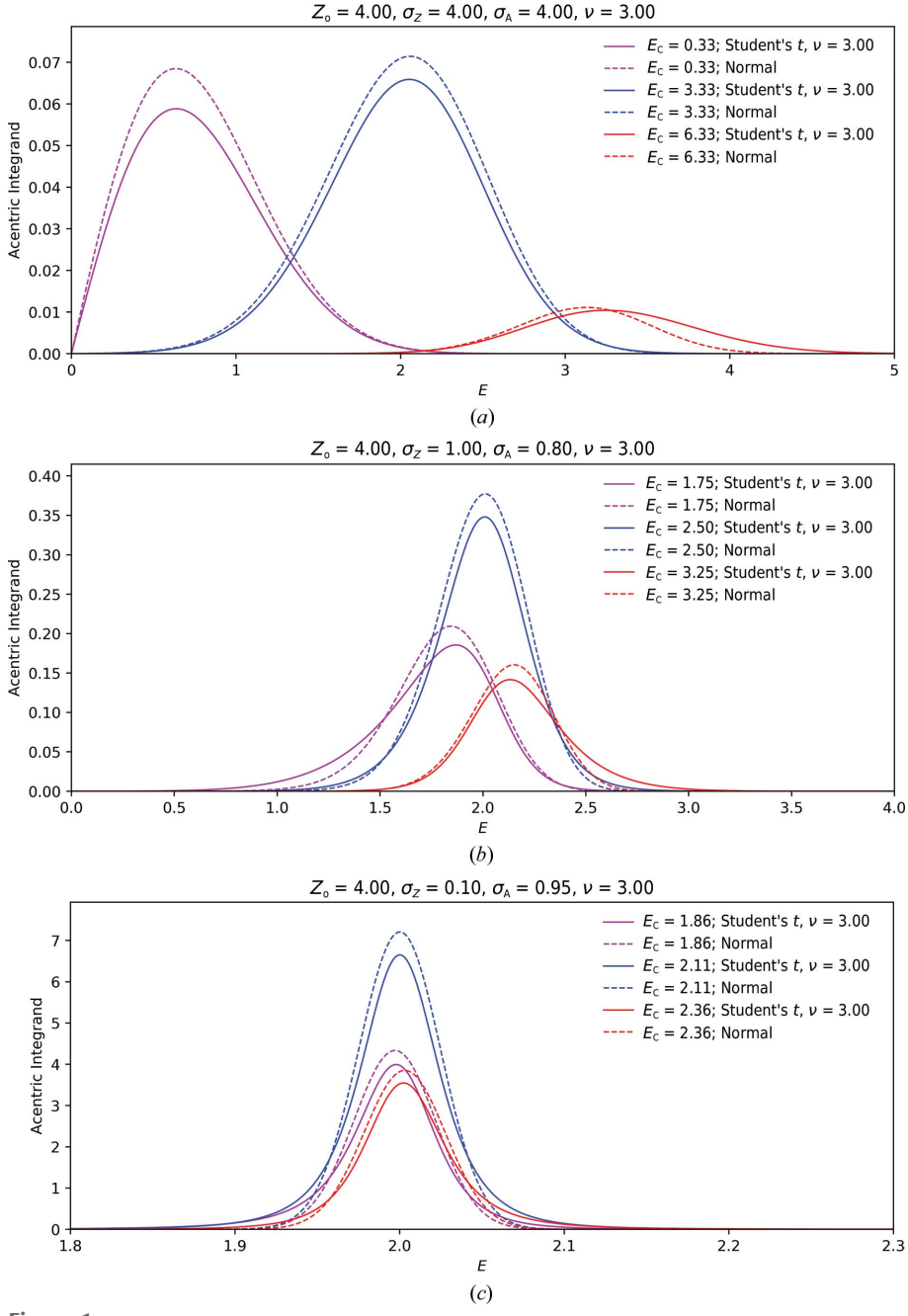


Figure 1
Integrand shapes for the acentric and centric distribution for different parameter settings show the variety of function shapes that occur when computing the marginal likelihood. When the experimental error is relatively large with respect to the intensity, high-mass areas of the function span a decent portion of the integration domain for $E \leq 6$ (*a*). When the error on the experimental data is relatively small, the bulk of the integrand mass is concentrated in smaller areas (*b*, *c*). In the case of a *t*-distribution-based noise model, the tails of the distribution are lifted compared with the normal noise model. The variety of these shapes makes the uniform application of a standard quadrature or Laplace approximation inefficient and suboptimal.

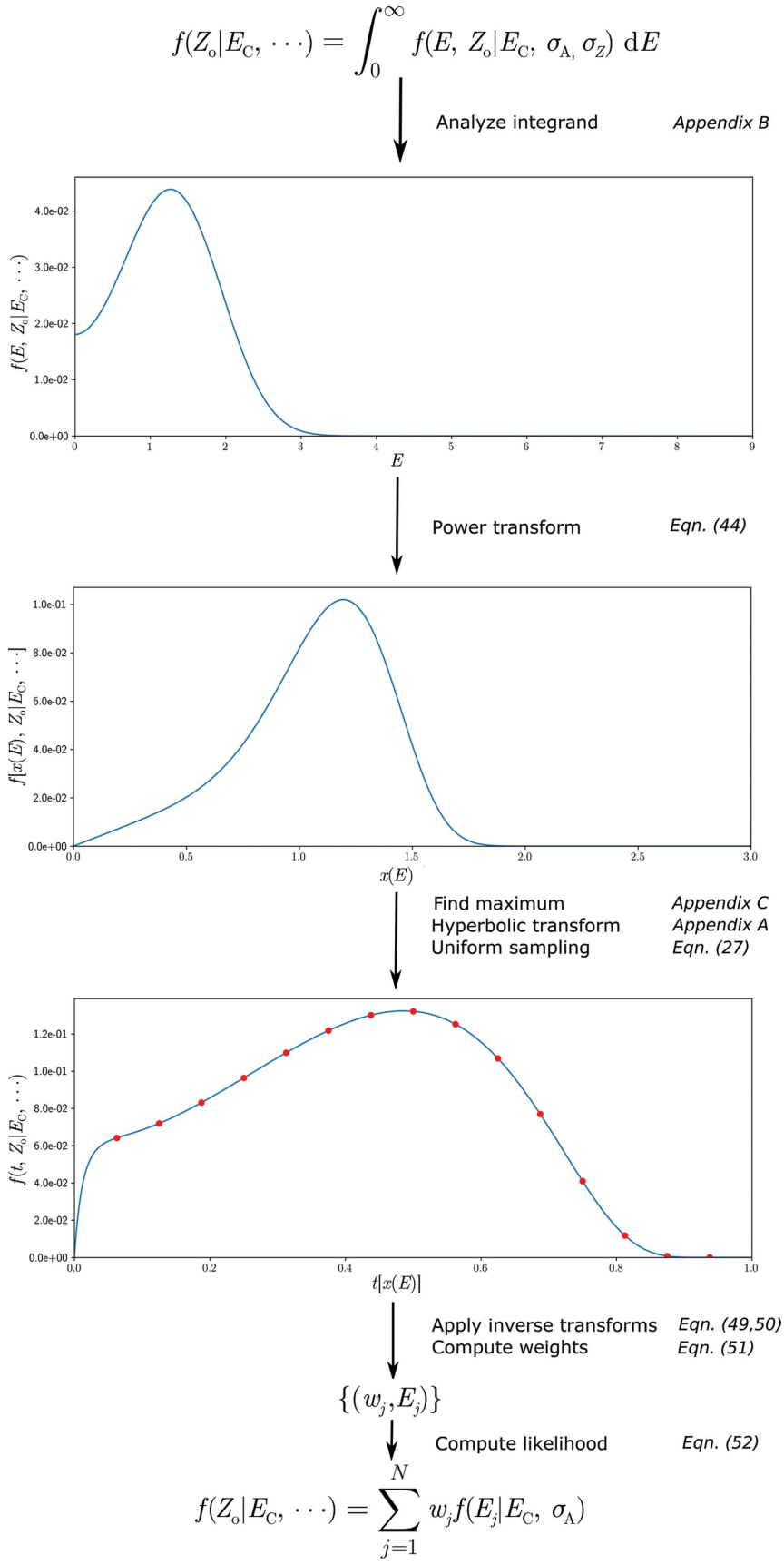


Figure 2
The numerical integration procedure developed is depicted as a sequence of steps. The general idea is to use a sequence of variable transformations that result in a smooth function on [0, 1] which can be easily integrated via a trapezoidal integration scheme. Once quadrature points have been established, the integration can be written as a sum of weighted Rice functions. See the main text for details.

Table 1
Parameter bounds for comparing integration methods.

Parameter	Start	End	Sampling points
E_C	0.1	6.0	20
σ_A	0.0	0.95	10
Z_o	−5.0	50.0	20
Z/σ_Z	0.5	10.0	20

Table 2
Integration results: acentric distribution.

The mean error and standard deviation of the relative log-likelihood over the full parameter range are reported as percentages.

Method	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
Laplace approximation	−0.142/0.874	0.294/0.971	0.485/1.135
Quadrature ($N = 3$)	0.191/0.778	0.152/0.831	0.281/1.058
Quadrature ($N = 5$)	0.130/0.377	0.126/0.481	0.196/0.627
Quadrature ($N = 7$)	0.085/0.218	0.074/0.309	0.116/0.428

Table 3
Integration results: centric distribution.

The mean error and standard deviation of the relative log-likelihood over the full parameter range are reported as percentages. Quadrature results for $\gamma = 1$ are absent because the function is not guaranteed to be zero at the origin as required by the hyperbolic quadrature scheme.

Method	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
Laplace approximation	−1.766/8.170	0.357/1.729	0.738/1.841
Quadrature ($N = 3$)	—	0.300/1.617	0.725/1.850
Quadrature ($N = 5$)	—	0.391/0.990	0.438/1.183
Quadrature ($N = 7$)	—	0.269/0.750	0.311/0.943

integrand to the middle of the new integration domain (Appendix *A*). This double-transformed function can subsequently be integrated using an equidistant trapezoidal integration scheme. The second change-of-variable operation requires, just like the Laplace approximation, knowledge of the maximum of the power-transformed integrand, which can be obtained using standard optimization methods (Appendices *B* and *C*). In a final step, the resulting quadrature expressed on the domain of the doubly transformed variable can be rewritten in the original variables by applying inverse transforms. A subsequent further simplification allows us to recast the whole integration as a sum of weighted Rice functions, where the effects of noisy observations and other errors are *hidden* in the sampling of E on $\mathbb{R}^+$ and the associated weights (Appendix *D*),

$$\begin{aligned} Q(E_C|Z_o, \sigma_A, \sigma_Z^2, \nu) &= \ln L_\cdot(E_C|Z_o, \sigma_A, \sigma_Z^2, \nu) \\ &= \ln\left[\sum_{j=1}^{N} w_j f_\cdot(E_j|E_C, \sigma_A)\right], \end{aligned} \tag{20}$$

where E_j are the quadrature sampling points and w_j are the associated weights. The sampling points and weights are dependent on E_C, Z_o, σ_A, σ_Z and ν. The quadrature sampling used can either be an N-point power-transformed hyperbolic quadrature or a single-point quadrature on the basis of a (power-transformed) Laplace approximation. Further details are given in Appendices *A*–*D*.

2.5. Derivatives

The practical use of a likelihood-based target function requires the calculation of its derivatives so that it can be used in gradient-based optimization methods. From expression (20), derivatives with respect to $Y \in \{E_C, \sigma_A, \nu\}$ can be obtained as follows:

$$\begin{aligned} Q'_Y(E_C|Z_o, \sigma_A, \sigma_Z^2, \nu) &= \frac{\mathrm{d}}{\mathrm{d}Y} Q(E_C|Z_o, \sigma_A, \sigma_Z^2, \nu) \\ &= \exp\left[-Q(E_C|Z_o, \sigma_A, \sigma_Z^2, \nu)\right] \\ &\quad \times \sum_{j=1}^{N} \frac{\mathrm{d}w_j f_\cdot(E_j|E_C, \sigma_A)}{\mathrm{d}Y}. \end{aligned} \tag{21}$$

The derivatives of the Rice components $f_\cdot(E_j|E_C, \sigma_A)$ with respect to E_C are listed in Appendix *B*.

3. Results and discussion

The first step in evaluating the proposed integration methods is to establish the ground truth of the integral that we wish to approximate. To this end, an equispaced, non-power-transformed trapezoidal quadrature was constructed integrating the function from $E = 0$ to $E = 6$ using 50 000 sampling points using all combinations of distribution parameters, as listed in Table 1, under the assumption of Gaussian errors on the intensities. Comparing the results of this integration with those obtained using a hyperbolic quadrature with 1500 points indicates that these two integration methods give similar results. We therefore take the ground truth as the results obtained with a hyperbolic quadrature using 1500 or more sample points. For both the acentric and the centric distributions, setting the power-transform variable γ to 2 provides good results, as shown in Tables 2 and 3, where the mean and standard deviation of the relative error in the log integrand are reported (as percentages). A number of different approximation schemes were tested, comparing results using the mean relative error in the log integrand. Because the variance-inflation approximation does not actually perform a marginalization, but performs a more *ad hoc* correction to incorporate low-fidelity measurements, its relative error against the log-likelihood is not a fair measure of its performance, nor does it provide insights into its strengths and drawbacks. Instead, we will compare the gradients of the log-likelihood target function with respect to E_C for all approximations, as this measure is independent of the different normalizations that arise when computing the full integral as compared with the variance-inflation approaches. Furthermore, given that the gradients of the log-likelihood function form are the Fourier coefficients of the 3D difference or gradient maps used to complete or rebuild structures, comparing the gradients of various approximations with those obtained from the full likelihood function can provide valuable insights into the strengths of different approximations. The use of gradients is of course predicated on being able to

estimate the value of σ_A, which in this case can be performed using a simple line search in fixed resolution shells. Details of these tests and their results can be found below.

3.1. Comparing integration methods

A comparison of the integration results using a number of different approximations are visualized in Fig. 3 for data sets

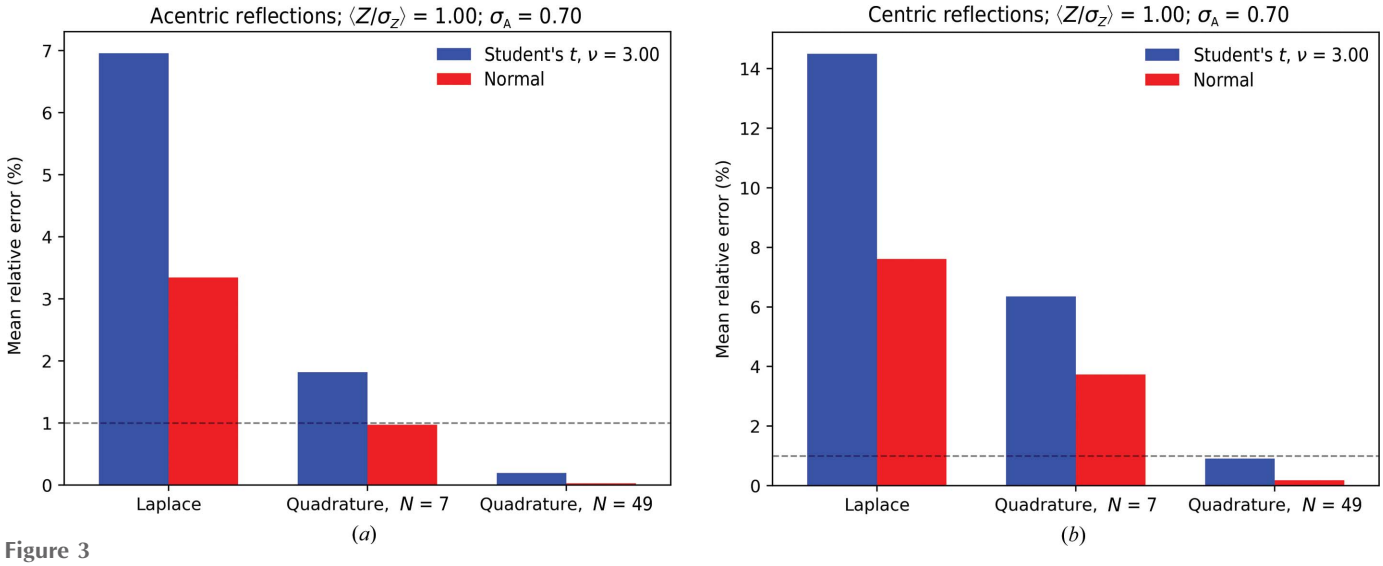


Figure 3
The relative mean error of the likelihood functions using a Laplace approximation and quadrature-based methods for normal and *t*-based noise models, for acentric (*a*) and centric reflections (*b*). The dotted horizontal line is set at 1% as a visual reference.

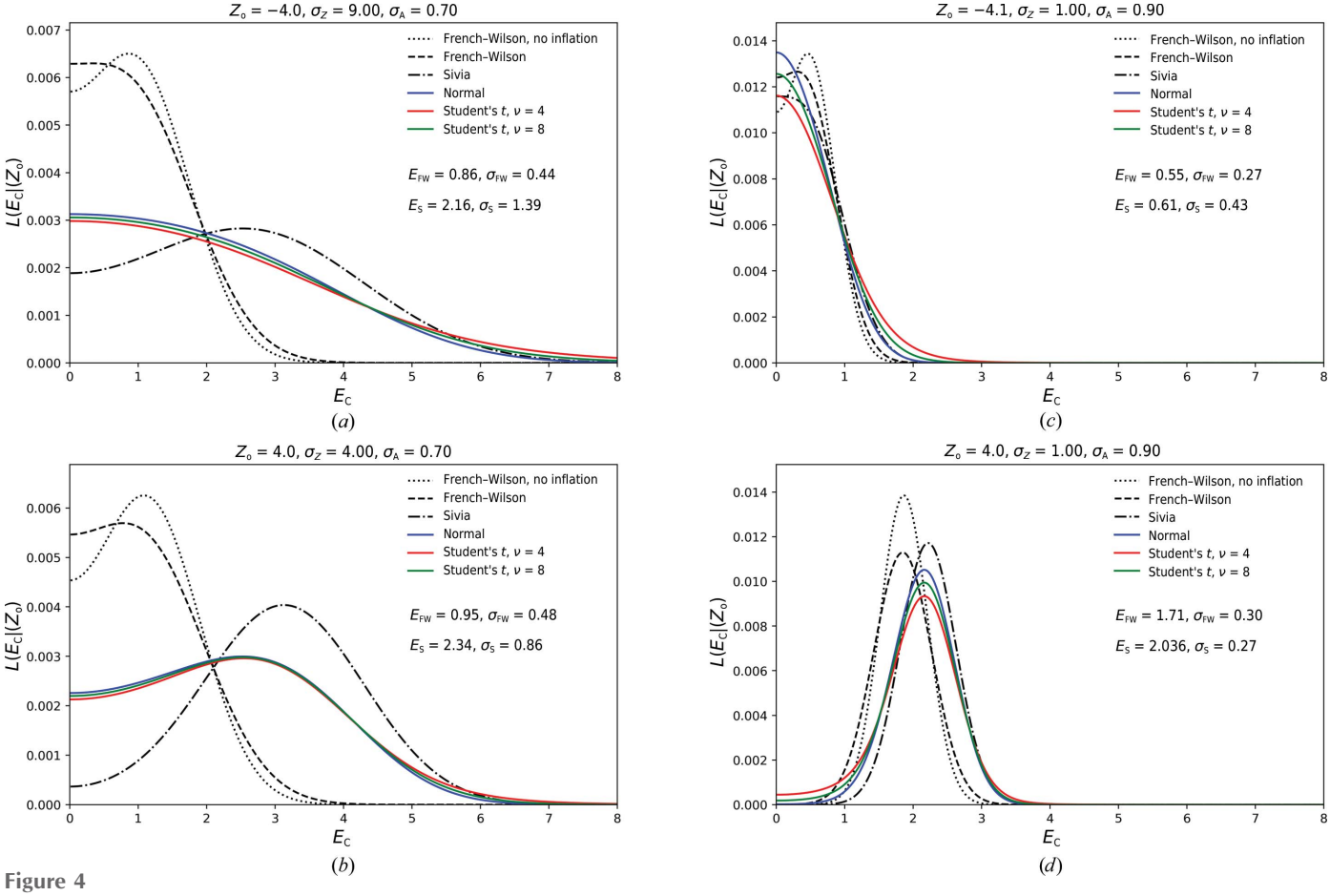


Figure 4
The shape of normalized likelihood functions under a number of different approximations for different input parameters indicates that the use of a point estimate for negative intensities or those with high noise values results in significant deviations from the ideal likelihood function. The difference between a *t*-based noise model and a normal noise model is small, but significantly affects the tail behavior of the likelihood function. Amplitude and standard deviation estimates for both the French–Wilson and Sivia approaches are given in the figure.

generated according to the procedure outlined in Appendix *F*. For the results shown the value of σ_A was set to 0.70, and a fixed error ratio was chosen such that $\langle Z/\sigma_Z \rangle = 1.0$. The redundancy was set to 4, resulting in $\nu = 3$. For the Z_o, E_C pairs, a likelihood function computed using a 1500-point hyperbolic quadrature was treated as the ground truth both for a *t*-distribution and an error model assuming a normal distribution. These values were compared with the Laplace approximation (a one-point quadrature) and seven-point and 49-point quadratures for both error models. While the Laplace approximation behaves relative well for the normal error model, it fails to deal properly with the elevated tails of the *t*-distribution, and better results are obtained using a quadrature. Satisfactory results are obtained using quadratures composed of seven or more sampling points. General heuristics can in principle be developed to tailor the specific accuracy of the quadrature on the basis of the hyperparameter of the error model. As expected, *t*-distributions with low ν values require a larger quadrature to get to a comparable error compared with those originating from a normal distribution owing to the presence of heavier tails.

3.2. Comparing likelihood functions

In order to obtain a better intuition of the behaviors of the target functions, we directly plot them for a few input parameters. Fig. 4 depicts the likelihood function $L(E_C|Z_o)$ for acentric reflections using just the French–Wilson protocol to estimate the amplitude while not inflating the variance and using the variance-inflation method with both the French–Wilson and the Sivia approaches, as well as the full likelihood functions using both a Gaussian error model and a *t*-distribution variant. All functions shown have been numerically normalized over $0 \leq E_C \leq 12$. When comparing the curves for weak and negative intensities, there is a remarkably large difference between techniques that use an estimate of E_o on the basis of a non-informative prior (French–Wilson & Sivia) versus those derived by the full integration (Figs. 4*a*, 4*b* and 4*c*). In the case of an observation with lower associated standard deviation, the differences between the approximations are smaller. The differences between a normal error model and a *t*-distribution manifest themselves in the tail behavior of the likelihood-function approximations, while the locations of the maxima seem relatively unchanged (Fig. 4*d*). The practical effects of the mismatch between an assumed normal error model and the *t*-type error models become apparent in the estimation of σ_A on the basis of the corresponding likelihood approximation. A synthetic data set with errors was constructed using the protocol outlined in Appendix *F*. The errors were chosen using a fixed error level such that the expected $\langle Z_o/\sigma_Z \rangle$ was 0.5 when $\nu \rightarrow \infty$ (see Appendix *F*). The resulting Z_o, σ_Z and E_C values were used to determine σ_A via a golden section-driven likelihood-maximization procedure (Kiefer, 1953). The resulting estimates of σ_A and their associated estimated standard deviation for different redundancy values ($\nu + 1$) are shown in Fig. 5. While for large values of ν the estimated values of σ_A are equivalent for both error models, at lower redundancy values the normal error model systematically underestimates σ_A. When the French–Wilson protocol is used, the resulting σ_A estimates are underestimated even more (Fig. 6).

3.3. Comparing log-likelihood gradients

Additional insights into the behavior of the likelihood-function approximation can be obtained by directly comparing its gradients for a selected set of parameter combinations. Numerical tests indicate that gradients computed using a 1500-point hyperbolic quadrature of the power-transformed function (with γ set to 2 for both the acentric and centric distribution) are indistinguishable from finite-difference gradients computed with a 50 000-point trapezoidal approach. In order to investigate the quality of the various approximations under common refinement scenarios, we construct a synthetic data set using random sampling methods as outlined in Appendix *F*. A redundancy of 4 ($\nu = 3$) was used in these tests. Gradients were computed using a 49-point quadrature, using a value of σ_A estimated from the corresponding approximation to the likelihood function. The results of these

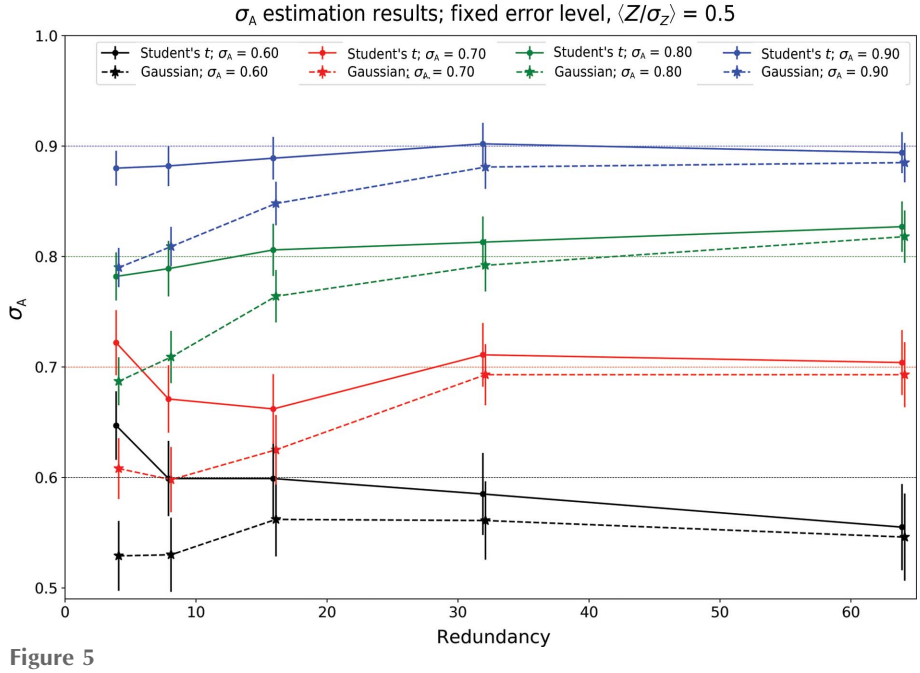


Figure 5
The behavior of a likelihood-based σ_A-estimation procedure when data with a *t*-based noise model are treated with a likelihood-based approach using normal noise: a negative bias is introduced in the estimate of σ_A at low redundancies.

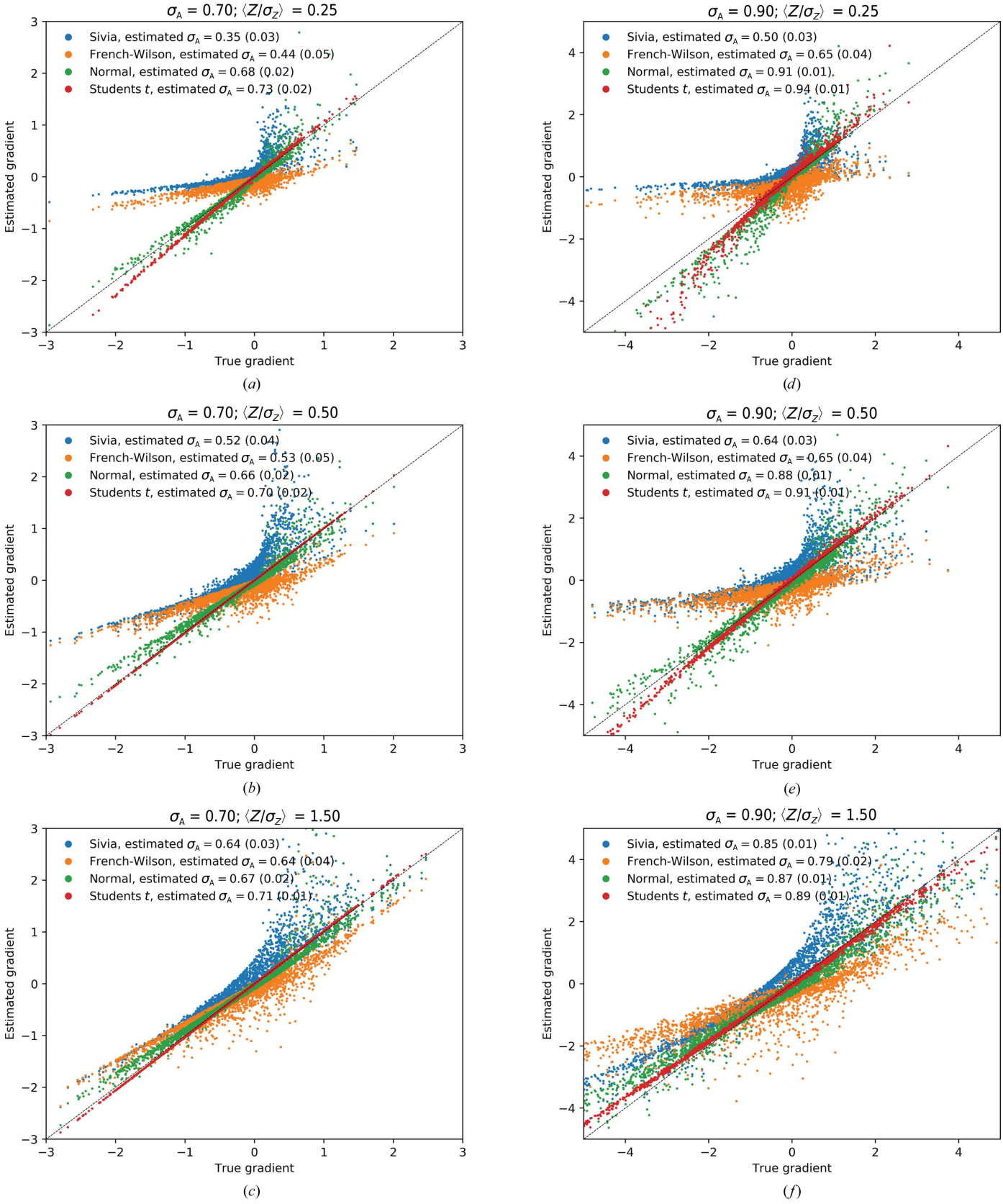


Figure 6
A comparison of gradients computed using different approximation schemes for *t*-based noise with $\nu = 3$. (*a*)–(*c*) and (*d*)–(*f*) depict the behavior of the gradient approximations with a decreasing noise level. Gradients were computed using a maximum-likelihood estimate of σ_A using their corresponding approximations. Both the normal model and the *t*-based model clearly outperform the French–Wilson and the Sivia approaches, while a marginal improvement over the normal noise model is obseved when the *t*-based model is used.

Table 4
Comparing likelihood gradients for simulated data by computing correlations of gradients computed using a 1500-point quadrature with the correct σ_A value (0.70) and those obtained using four different approximation methods, as outlined in the main text, and the maximum-likelihood estimate of σ_A given the approximation of the likelihood function.

The reported entries are estimated values of σ_A and the gradient correlation.

Method	$\langle Z/\sigma_Z\rangle = 0.25$	$\langle Z/\sigma_Z\rangle = 0.5$	$\langle Z/\sigma_Z\rangle = 1.5$
Sivia	0.35/68.8%	0.52/82.1%	0.64/93.5%
French–Wilson	0.44/80.5%	0.53/87.7%	0.64/94.3%
Normal	0.68/96.9%	0.66/97.4%	0.67/97.9%
Student's *t*	0.73/100%	0.70/100%	0.71/100%

Table 5
Comparing likelihood gradients for simulated data by computing correlations of gradients computed using a 1500-point quadrature with the correct σ_A value (0.90) and those obtained using four different approximation methods, as outlined in the main text, and the maximum-likelihood estimate of σ_A given the approximation of the likelihood function.

The reported entries are estimated values of σ_A and the gradient correlation.

Method	$\langle Z/\sigma_Z\rangle = 0.25$	$\langle Z/\sigma_Z\rangle = 0.5$	$\langle Z/\sigma_Z\rangle = 1.5$
Sivia	0.50/63.1%	0.64/73.6%	0.85/93.2%
French–Wilson	0.64/60.3%	0.65/74.9%	0.79/88.4%
Normal	0.91/97.0%	0.88/96.9%	0.87/97.6%
Student's *t*	0.94/98.7%	0.91/99.9%	0.89/99.9%

comparisons are shown in Fig. 6 and summarized in Tables 4 and 5. The quality of the gradients is gauged by a correlation coefficient to the true value. The results indicate that for data for which $\langle Z_0/\sigma_Z\rangle$ is large, all gradient-calculation methods converge to those obtained using the full intensity-based likelihood function with experimental errors and a Student's *t* noise model, but that for high and intermediate noise levels the variance-inflation method significantly underperforms. While differences between normal and *t*-style noise models seem small on the basis of the correlation coefficients, significant deviations are seen in individual reflections under high-noise and low-redundancy settings. These aberrant gradients can potentially negatively influence the quality of gradient maps for structure completion.

4. Conclusions

Numerical procedures for the efficient determination of intensity-based likelihood functions and their gradients are developed and compared. Whereas the Laplace approximation behaves reasonably well for the estimation of the likelihood function itself under a normal noise model, our results show that the both the likelihood and its associated gradients can be significantly improved upon by using a numerical quadrature. Given that the derivative of the log-likelihood function is the key ingredient in gradient-based refinement methods and is used to compute difference maps for structure completion, the proposed approach could improve the convergence of existing refinement and model-building methods. Although it is unclear what the optimal quadrature order or noise model should be in a practical case, our results suggest that it is likely to be below 15 sampling points for normal noise and below 49 for *t*-type errors. Algorithmically, the most costly operation is the iterative procedure for finding the maximum of the integrand. The proposed Newton-based method typically converges well within 50 function evaluations, even in the absence of a predetermined good starting point for the line search. The construction of the hyperbolic quadrature does not require any iterative optimization, nor does the subsequent calculation of the associated gradient and function values. Given the large additional overhead in refinement or other maximum-likelihood applications in crystallography, the use of the presented methodology to compute target functions is likely to have only have a minimal impact on the total run time of the workflow, while providing a rapidly converging approximation to a full intensity-based likelihood that takes experimental errors in both the estimate of the mean intensity and its variance into account. Although only a full integration into a crystallographic software package can determine the situations under which a practical benefit can be obtained from using the outlined approach, the tests here indicate that significant improvements are possible. Furthermore, the ease with which the proposed quadrature method can be adapted to a different of choice of error model is a large benefit over existing approximation methods, making it for instance possible to use experiment-specific noise models in refinement and phasing targets (Sharma *et al.*, 2017).

APPENDIX *A*
A hyperbolic quadrature

Given a function $g(x)$, with $x \geq 0$ and $g(x) \geq 0$, we seek to compute its integral over the positive half-line:

$$G = \int_0^\infty g(x)\,\mathrm{d}x. \tag{22}$$

Set

$$h(x) = \ln g(x). \tag{23}$$

Define the supremum of $g(x)$ by x_0 such that $h'(x_0) = 0$. For the class of functions we are interested in, $g(0)$ is equal to 0, for instance owing to the power transform outlined in the main text, and $\lim_{x\to\infty} g(x)$ is 0 as well. Define the following change of variables on the basis of a shifted and rescaled logistic function,

$$t = \frac{\exp(kx) - 1}{\exp(kx) + \exp(kx_0)}, \tag{24}$$

where k is a positive contant. Note that $t(x = 0) = 0$ and $\lim_{x\to\infty} t(x) = 1$. The inverse function is

$$x(t) = x_0 - \frac{1}{k}\ln\left[\frac{\exp(x_0k)(1-t)}{1 + t\exp(kx_0)}\right] \tag{25}$$

and has a derivative with respect to t equal to

$$x'(t) = \frac{\exp(-kt)[\exp(kx_0) + \exp(kt)]^2}{k[\exp(kx_0) + 1]}. \tag{26}$$

The value x_0 determines the approximate 'inflection' point of the hyperbolic compression scheme and the constant k controls the slope around the midpoint. An N-point quadrature can now be constructed by uniformly sampling t between 0 and 1,

$$t_j = \frac{j}{N+1}, \tag{27}$$

for $1 \leq j \leq N$. Given that both $g(0)$ and $\lim_{x\to\infty} g(x)$ are zero, the integral G can now be computed via a trapezoidal integration rule,

$$G = \frac{1}{N+1}\sum_{j=1}^{N} g[x(t_j)]x'(t_j). \tag{28}$$

If k is chosen to be

$$k = \left[\frac{-2h''(x_0)}{\pi}\right]^{1/2} \tag{29}$$

then the above quadrature for $N = 1$ yields the Laplace approximation when $x_0|h''(x_0)|^{1/2}$ is large, as $|x_0 - x(1/2)|$ goes to zero. If a hyperbolic quadrature is constructed on a distribution of power-transformed variables, these derived weights can be multiplied by the Jacobian of that transformation, so that the final numerical evaluation can be carried out in the original set of variables.

APPENDIX *B*
Distributions and derivatives

B1. Rice functions, acentrics

The logarithms of the acentric Rice distribution and its derivatives with respect to E and E_C are given below.

$$\begin{aligned} h_{a,\mathrm{Rice}}(E, E_C, \sigma_A) &= \ln f_a(E|E_C, \sigma_A) \\ &= \ln 2 + \ln E - \ln\left(1-\sigma_A^2\right) - \frac{(E-\sigma_A E_C)^2}{1-\sigma_A^2} \\ &\quad + \ln e I_0\left(\frac{2\sigma_A E E_C}{1-\sigma_A^2}\right), \end{aligned} \tag{30}$$

$$\begin{aligned} h'^{E}_{a,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}}{\mathrm{d}E}\ln f_a(E|E_C, \sigma_A) \\ &= \frac{1}{E} - \frac{2(E-\sigma_A E_C)}{1-\sigma_A^2} \\ &\quad + \frac{2E_C\sigma_A}{1-\sigma_A^2}\left[\frac{I_1\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}{I_0\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)} - 1\right], \end{aligned} \tag{31}$$

$$\begin{aligned} h''^{E}_{a,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}^2}{\mathrm{d}E^2}\ln f_a(E|E_C, \sigma_A) \\ &= -\frac{1}{E^2} - \frac{2}{1-\sigma_A^2} + \left(\frac{2E_C\sigma_A}{1-\sigma_A^2}\right)^2 \\ &\quad - \frac{2E_C\sigma_A}{E(1-\sigma_A^2)}\left[\frac{I_1\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}{I_0\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}\right] \\ &\quad - \left(\frac{2E_C\sigma_A}{1-\sigma_A^2}\right)^2\left[\frac{I_1\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}{I_0\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}\right]^2, \end{aligned} \tag{32}$$

$$\begin{aligned} h'^{E_C}_{a,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}}{\mathrm{d}E_C}\ln f_a(E|E_C, \sigma_A) \\ &= \frac{2\sigma_A(E-\sigma_A E_C)}{1-\sigma_A^2} \\ &\quad + \frac{2E\sigma_A}{1-\sigma_A^2}\left[\frac{I_1\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)}{I_0\left(\frac{2EE_C\sigma_A}{1-\sigma_A^2}\right)} - 1\right]. \end{aligned} \tag{33}$$

B2. Rice functions, centrics

The logarithms of the centric Rice distribution and its derivatives with respect to E and E_C are given below.

$$\begin{aligned} h_{c,\mathrm{Rice}}(E, E_C, \sigma_A) &= \ln f_c(E|E_C, \sigma_A) \\ &= \frac{1}{2}[\ln 2 - \ln\pi - \ln(1-\sigma_A^2)] - \frac{E^2+\sigma_A^2E_C^2}{2(1-\sigma_A^2)} \\ &\quad + \ln\cosh\left(\frac{\sigma_A E E_C}{1-\sigma_A^2}\right), \end{aligned} \tag{34}$$

$$\begin{aligned} h'^{E}_{c,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}}{\mathrm{d}E}\ln f_c(E|E_C, \sigma_A) \\ &= \frac{E}{\sigma_A^2-1} + \frac{\sigma_A E_C}{1-\sigma_A^2}\tanh\left(\frac{\sigma_A E E_C}{1-\sigma_A^2}\right), \end{aligned} \tag{35}$$

$$\begin{aligned} h''^{E}_{c,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}^2}{\mathrm{d}E^2}\ln f_c(E|E_C, \sigma_A) \\ &= \frac{1}{\sigma_A^2-1} + \left(\frac{\sigma_A E_C}{1-\sigma_A^2}\right)^2 \\ &\quad - \left(\frac{\sigma_A E_C}{1-\sigma_A^2}\right)^2\left[\tanh\left(\frac{\sigma_A E E_C}{1-\sigma_A^2}\right)\right]^2, \end{aligned} \tag{36}$$

$$\begin{aligned} h'^{E_C}_{c,\mathrm{Rice}}(E, E_C, \sigma_A) &= \frac{\mathrm{d}}{\mathrm{d}E_C}\ln f_c(E|E_C, \sigma_A) \\ &= \frac{\sigma_A^2 E_C}{\sigma_A^2-1} + \frac{\sigma_A E}{1-\sigma_A^2}\tanh\left(\frac{\sigma_A E E_C}{1-\sigma_A^2}\right). \end{aligned} \tag{37}$$

B3. Student's *t*-distribution

The logarithm of the *t*-distribution as specified in the main text and its derivatives with respect to E are given below.

$$\begin{aligned} h_{\rm o}(Z_{\rm o}, E, \sigma_Z^2, \nu) &= \ln f(Z_{\rm o}|E, \sigma_{\rm o}^2, \nu) \\ &= \ln\Gamma\left(\frac{\nu+1}{2}\right) - \frac{1}{2}\ln\left(\nu\pi\sigma_{\rm o}^2\right) - \ln\Gamma\left(\frac{\nu}{2}\right) \\ &\quad - \frac{\nu+1}{2}\ln\left[1 + \frac{(Z_{\rm o}-E^2)^2}{\nu\sigma_Z^2}\right], \end{aligned} \tag{38}$$

$$\begin{aligned} h_{\rm o}^{\prime E}(Z_{\rm o}, E, \sigma_Z^2, \nu) &= \frac{\rm d}{{\rm d}E}\ln f(Z_{\rm o}|E, \sigma_{\rm o}^2, \nu) \\ &= \frac{2(\nu+1)E(Z_{\rm o}-E^2)}{\sigma_Z^2\nu + (Z_{\rm o}-E^2)^2}, \end{aligned} \tag{39}$$

$$\begin{aligned} h_{\rm o}^{\prime\prime E}(Z_{\rm o}, E, \sigma_Z^2, \nu) &= \frac{{\rm d}^2}{{\rm d}E^2}\ln f(Z_{\rm o}|E, \sigma_{\rm o}^2, \nu) \\ &= 2(\nu^2+1) \\ &\quad \times \frac{[\sigma_Z^2\nu(Z_{\rm o}-3E^2) + (Z_{\rm o}-E^2)^2(E^2+Z_{\rm o})]}{[\sigma_Z^2\nu + (Z_{\rm o}-E^2)^2]^2}. \end{aligned} \tag{40}$$

When ν is large, the *t*-distribution can be approximated with a normal distribution:

$$h_{\rm o}(Z_{\rm o}, E, \sigma_Z^2) = -\frac{1}{2}\ln 2\pi + \ln\sigma_Z - \frac{(Z_{\rm o}-E^2)^2}{2\sigma_Z^2}, \tag{41}$$

$$h_{\rm o}^{\prime E}(Z_{\rm o}, E, \sigma_Z^2) = \frac{2E(Z_{\rm o}-E^2)}{\sigma_Z^2}, \tag{42}$$

$$h_{\rm o}^{\prime\prime E}(Z_{\rm o}, E, \sigma_Z^2) = \frac{6E^2}{\sigma_Z^2}. \tag{43}$$

APPENDIX C
Finding x_0

As outlined in the main text, numerical integration via the hyperbolic quadrature is greatly assisted by the change of variables

$$E = x^\gamma \tag{44}$$

with Jacobian (see expression 17)

$$\frac{{\rm d}E(x)}{{\rm d}x} = \gamma x^{\gamma-1}. \tag{45}$$

The location of the maximum value of the integrand, x_0, is found using a straightforward application of the Newton root-finding algorithm using the first and second derivatives that are outlined below. Set

$$h_{t,{\rm a}}(E) = h_{\rm a}(E, \cdots) + h_{\rm o}(E, \cdots) \tag{46}$$

and

$$h_{t,{\rm c}}(E) = h_{\rm c}(E, \cdots) + h_{\rm o}(E, \cdots). \tag{47}$$

Using the shorthand $h_t(E, \cdots)$ to indicate either of these functions, we first apply a power transform (expression 44), as outlined above. Because we need to locate the maximum of this function, we need the derivatives with respect to x as well. The resulting function and its first and second derivatives with respect to x after the change-of-variable operation are given by

$$\begin{aligned} \hat{h}_t(x, \gamma) &= \ln\gamma + (\gamma-1)\ln x + h_t(E = x^\gamma, \cdots), \\ \hat{h}_t^{\prime}(x, \gamma) &= \frac{(\gamma-1)}{x} + \gamma x^{\gamma-1} h_t^{\prime}(E = x^\gamma, \cdots), \\ \hat{h}_t^{\prime\prime}(x, \gamma) &= \frac{(1-\gamma)}{x^2} + \gamma^2 x^{2\gamma-2} h_t^{\prime\prime}(E = x^\gamma, \cdots) \\ &\quad + \gamma(\gamma-1)x^{\gamma-2} h_t^{\prime}(E = x^\gamma, \cdots). \end{aligned} \tag{48}$$

The analytic forms for $h_t(\cdots)$, $h_t^{\prime}(\cdots)$ and $h_t^{\prime\prime}(\cdots)$ are given in Appendix *B*. Decent starting values for the Newton search can be found by performing a single Newton-based update on a set of (say) 15 equispaced values of x sampled between 0 and $x = 6^{1/\gamma}$. The integrand-weighted mean of the resulting updated sampling points typically refines within ten iterations to the supremum.

APPENDIX D
Likelihood synthesis

Using the above approaches, the full likelihood function can be expressed as a sum of weighted Rice functions (expressions 30 and 34), where E is sampled on the basis of a quadrature derived from a power-transformed variable $E = x^\gamma$ using the hyperbolic sampling scheme outlined above. Taking into account the combination of the power transform and the hyperbolic quadrature, the sampling nodes of the quadrature are equal to

$$E_j = x_j^\gamma, \tag{49}$$

$$x_j = x_0 - \frac{1}{k}\ln\left[\frac{\exp(x_0 k)(1-t_j)}{1 + t_j\exp(kx_0)}\right], \tag{50}$$

where t_j, x_0 and k are defined and computed as outlined in Appendices *A* and *C* and $1 \le j \le N$. The quadrature weights can now be set to absorb the hyperbolic sampling, the power transform and the error model acting on the observed intensity and its associated standard deviation:

$$\begin{aligned} w_j &= \gamma x_j^{\gamma-1} \times \frac{\exp(-kt_j)[\exp(kx_0) + \exp(kt_j)]^2}{k[\exp(kx_0)+1]} \\ &\quad \times f(Z_o|E_j, \sigma_Z^2, \nu) \times \frac{1}{N+1}. \end{aligned} \tag{51}$$

This thus yields a sum of weighted Rice functions that approximates the full likelihood function,

$$L_{\cdot}(E_{\rm C}|Z_{\rm o}, \sigma_{\rm A}, \sigma_Z^2) = \sum_{j=1}^{N} w_j f_{\cdot}(E_j|E_{\rm C}, \sigma_{\rm A}), \tag{52}$$

where $f_{\cdot}(\cdots)$ is the acentric or centric Rice function.

When the likelihood function is approximated using the power-transformed Laplace approximation instead of using the quadrature approach, we obtain a weighted Rice function

$$L_{\cdot}(E_{\rm C}|Z_{\rm o},\sigma_{\rm A},\sigma_Z^2) = w_0 f_{\cdot}(E_j|E_{\rm C},\sigma_{\rm A}), \tag{53}$$

with the weight given by

$$w_0 = \gamma x_0^{\gamma-1} \times f(Z_{\rm o}|E_j,\sigma_Z^2,\nu) \times \left[\frac{2\pi}{\hat{\boldsymbol{h}}_{\cdot}''(x_0)}\right]^{1/2}, \tag{54}$$

where $E_0 = x_0^{\gamma}$ and $\hat{\boldsymbol{h}}_t''(x_0)$ is defined in expression (48).

APPENDIX *E*
Structure-factor amplitude estimation

In order to use an inflated variance modification as an approximation to the full numerical integration, we need to be able to estimate reflection amplitudes and their standard deviations from observed intensities and their standard deviations. While this process is normally performed using a standard French–Wilson estimation procedure, another route can be adopted following an approach developed by Sivia & David (1994). Assume a uniform, improper prior on E, such that

$$f(E) = \begin{cases} 1 & E \geq 0 \\ 0 & E < 0 \end{cases}, \tag{55}$$

resulting in a conditional distribution

$$f(E|Z_{\rm o},\sigma_Z^2) \propto E\exp\left[-\frac{(E^2-Z_{\rm o})^2}{2\sigma_Z^2}\right]. \tag{56}$$

A normal approximation to this distribution can be obtained by the method of moments or, as performed here, by a maximum *a posteriori* approximation with a mean equal to the mode of the above distribution and a standard deviation estimated on the basis of the second derivative of the log-likelihood at the location of the mode:

$$E_{\rm o} = \sup_E \ln\left[f(E|Z_{\rm o},\sigma_Z^2)\right], \tag{57}$$

$$\sigma_E^2 = -\left\{\frac{{\rm d}^2}{{\rm d}E^2}\ln\left[f(E|Z_{\rm o},\sigma_Z^2)\right]\right\}^{-1}_{E=E_{\rm o}}. \tag{58}$$

An analytic expression is readily obtained, resulting in

$$E_{\rm o} = \left\{\frac{1}{2}[Z_{\rm o} + (Z_{\rm o}^2 + 2\sigma_Z^2)^{1/2}]\right\}^{1/2}, \tag{59}$$

$$\sigma_E^2 = \frac{\sigma_Z^2}{4\left(Z_{\rm o}^2 + 2\sigma_Z^2\right)^{1/2}}. \tag{60}$$

The quality of this approximation will critically depend on the values of $Z_{\rm o}$ and σ_Z^2. Note that standard error propagation on $E_{\rm o} = Z_{\rm o}^{1/2}$ yields

$$\sigma_E^2 = \frac{\sigma_Z^2}{4Z_{\rm o}}, \tag{61}$$

which can be seen to be converge to expression (59) for the case where $Z_{\rm o}$ is significantly larger then σ_Z.

APPENDIX *F*
Simulating synthetic data

Data for the numerical tests and benchmarks were obtained by sampling from the underlying distribution using the following procedure. For acentric reflections, 'true' complex structure factors and 'perturbed' complex structure factors are generated by successive draws from normal distributions with zero mean and specified variance:

$$\begin{aligned} \mathbf{E}_{\rm True} &= (X_{\Re}, X_{\Im}); f(X_{\cdot}) = N(0, 1/2), \\ \boldsymbol{\Delta}_{\rm C} &= (X_{\Re}, X_{\Im}); f(X_{\cdot}) = N[0, (1-\sigma_{\rm A}^2)/2], \\ \mathbf{E}_{\rm C} &= \sigma_{\rm A}\mathbf{E}_{\rm True} + \boldsymbol{\Delta}_{\rm C}, \end{aligned} \tag{62}$$

where $N(\mu, \sigma^2)$ denotes the normal distribution. For centric reflections, the following procedure is used:

$$\begin{aligned} \mathbf{E}_{\rm True} &= (X_{\Re}, 0); f(X_{\Re}) = N(0, 1), \\ \boldsymbol{\Delta}_{\rm C} &= (X_{\Re}, 0); f(X_{\Re}) = N[0, (1-\sigma_{\rm A}^2)], \\ \mathbf{E}_{\rm C} &= \sigma_{\rm A}\mathbf{E}_{\rm True} + \boldsymbol{\Delta}_{\rm C}. \end{aligned} \tag{63}$$

Noise is added in the following fashion. Given a target variance $\sigma_{\rm target}^2$, we generate $\nu + 1$ normal random variates to compute the sample mean and sample variance:

$$\begin{aligned} Z_{\rm True} &= |\mathbf{E}_{\rm True}|^2, \\ Z_j &= X; f(X) = N[Z_{\rm True}, (\nu+1)\sigma_{\rm target}^2], \\ Z_{\rm o} &= [1/(\nu+1)]\textstyle\sum Z_j, \\ \sigma_Z^2 &= \{1/[\nu(\nu+1)]^2\}\textstyle\sum(Z_j - Z_{\rm o})^2. \end{aligned} \tag{64}$$

Two error models are adopted in the described tests, namely a *fixed error ratio* or a *fixed error level*. In the fixed error ratio method, $\sigma_{\rm target}$ is different for every simulated intensity, and is chosen to be $Z_{\rm True}/\tau$, where τ is equal to the desired $\mathbb{E}(Z_{\rm True}/\sigma_{\rm target})$ level. For the fixed error level method, $\sigma_{\rm target}$ is fixed at $1/\tau$ for all intensities. A complete data set is simulated assuming a 9:1 ratio of acentric to centric reflections.

Acknowledgements

The above algorithms are implemented in a set of Python3 routines and are available upon request. Some parts of this work were prepared in partial fulfillment of the requirements of the Berkeley Laboratory Undergraduate Research (BLUR) Program, managed by Workforce Development and Education at Berkeley Laboratory. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Funding information

This research was supported in part by the Advanced Scientific Computing Research and the Basic Energy Sciences programs, which are supported by the Office of Science of the US Department of Energy (DOE) under Contract DE-AC02-05CH11231. Further support originates from the National

Institute of General Medical Sciences of the National Institutes of Health (NIH) under Award 5R21GM129649-02.

References

Beu, K. E., Musil, F. J. & Whitney, D. R. (1962). *Acta Cryst.* **15**, 1292–1301.

Brewster, A. S., Bhowmick, A., Bolotovsky, R., Mendez, D., Zwart, P. H. & Sauter, N. K. (2019). *Acta Cryst.* D**75**, 959–968.

Bricogne, G. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 159–178. Warrington: Daresbury Laboratory.

Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

Bunkóczi, G., McCoy, A. J., Echols, N., Grosse-Kunstleve, R. W., Adams, P. D., Holton, J. M., Read, R. J. & Terwilliger, T. C. (2015). *Nat. Methods*, **12**, 127–130.

Cools, R. (2002). *J. Comput. Appl. Math.* **149**, 1–12.

Cowtan, K. (2000). *Acta Cryst.* D**56**, 1612–1621.

Davis, P. J. & Rabinowitz, P. (1984). *Methods of Numerical Integration*, 2nd ed. New York: Academic Press.

Fisher, R. A. (1915). *Biometrika*, **10**, 507–521.

French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Hamburg: Perthes & Besser.

Gauss, C. F. (1816). *Z. Astronom. Verwandte Wiss.* **1**, 1816.

Gauss, C. F. (1823). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*. Göttingen: Henricus Dieterich.

Green, E. A. (1979). *Acta Cryst.* A**35**, 351–359.

Hagen, G. (1867). *Grundzüge der Wahrscheinlichkeits-Rechnung.* Berlin: Ernst & Korn.

Kass, R. E. & Steffey, D. (1989). *J. Am. Stat. Assoc.* **84**, 717–726.

Kiefer, J. (1953). *Proc. Am. Math. Soc.* **4**, 502–506.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Lunin, V. Y. & Skovoroda, T. P. (1995). *Acta Cryst.* A**51**, 880–887.

Lunin, V. Y. & Urzhumtsev, A. G. (1984). *Acta Cryst.* A**40**, 269–277.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Neyman, J. & Scott, E. L. (1948). *Econometrica*, **16**, 1–32.

Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.

Pearson, E. S. (1970). *Studies in the History of Statistics and Probability*, edited by E. S. Pearson & M. G. Kendall, pp. 411–413. London: Charles Griffin.

Peng, R. D. (2018). *Advanced Statistical Computing*. https://leanpub.com/advstatcomp.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Read, R. J. (1997). *Methods Enzymol.* **277**, 110–128.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Read, R. J. & McCoy, A. J. (2016). *Acta Cryst.* D**72**, 375–387.

Rossi, R. J. (2018). *Mathematical Statistics: An Introduction to Likelihood Based Inference.* Hoboken: John Wiley & Sons.

Sharma, A., Johansson, L., Dunevall, E., Wahlgren, W. Y., Neutze, R. & Katona, G. (2017). *Acta Cryst.* A**73**, 93–101.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Sivia, D. S. & David, W. I. F. (1994). *Acta Cryst.* A**50**, 703–714.

Skubák, P., Waterreus, W.-J. & Pannu, N. S. (2010). *Acta Cryst.* D**66**, 783–788.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography*, 1st ed. Oxford: Pergamon Press.

Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.

Student (1908). *Biometrika*, **6**, 1–25.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Cryst.* A**39**, 813–817.

Trefethen, L. N. & Weideman, J. A. C. (2014). *SIAM Rev.* **56**, 385–458.

Welch, B. L. (1947). *Biometrika*, **34**, 28–35.

Wilks, S. S. (1938). *Ann. Math. Stat.* **9**, 60–62.

Wilson, A. J. C. (1980). *Acta Cryst.* A**36**, 937–944.

Woolfson, M. M. (1956). *Acta Cryst.* **9**, 804–810.