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. E-mail: PHZwart@lbl.gov*

Maximum Likelihood, Refinement, Numerical Integration

Abstract

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 targets functions. Here a numerical quadrature is developed that allows for the rapid evaluation of intensity-based likelihood functions in crystallographic applications. By using a sequence of change of variable transformations, including a non-linear 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 have 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 early 1800s by Gauss (Gauss, 1809; Gauss, 1816; Gauss, 1823) and Hagen (1867), and was further developed by Fisher (1915), Wilks (1938) and Neyman and Pearson (Neyman *et al.*, 1948; Pearson, 1970). In the crystallographic community, Beu *et al.* (1962) were the first to explicitly use maximum likelihood estimation, applying it on 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 really took off after making significant impacts in the analysis of macromolecules. For these type of samples, structure solution and refinement problems were often problematic due to very incomplete or low quality starting models, making standard least squares techniques under-perform. In the 1980's and 1990's, likelihood based methods became mainstream, culminating in the ability to routinely determine and refine structures that were previously thought of as 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 the success was the development of cross-validation techniques to reduce bias in the estimation of hyper parameters that govern behavior of the likelihood functions (Lunin & Skovoroda, 1995; Pannu & Read, 1996). In the beginning of the 21st century,

Read and coworkers extended the likelihood formalism to the 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, an approach that is equivalent to a second-order approximation of a Rice-based likelihood function (Bricogne, 1997). 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 scenarios described, key advances were made by deriving problem-specific likelihood functions and applying them on challenging structure determination problems. In the majority of those 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 convolute 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 said difficulties. The approach derived has direct applications in structure refinement and molecular replacement, while the general methodology can be extended to other crystallographic scenarios as well. In the remainder of this paper we will provide a general introduction into likelihood based

methods, list relevant background into numerical integration techniques, develop a adaptive quadrature approach, apply it to a Rice-type likelihood functions, and validate its results.

1.1. Maximum Likelihood Formalism

The estimation of model parameters $\boldsymbol{\theta}$ given some data set $\mathcal{X} = \{x_1, \cdots, x_j, \cdots, x_N\}$ via the likelihood formalism is done 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 done by finding that $\boldsymbol{\theta}$ 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, in which a so-called nuisance parameter is integrated out:

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

where, under 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 anywhere between 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, 1976; Luzzati, 1952; Woolfson, 1956; Lunin & Urzhumtsev, 1984):

$$f_a(F|F_C, \alpha, \beta) = \frac{2F}{\epsilon\beta} \exp\left[-\frac{F^2 + \alpha^2 F_C^2}{\epsilon\beta}\right] I_0\left(\frac{2\alpha F F_C}{\epsilon\beta}\right) \tag{6}$$

$$f_c(F|F_C, \alpha, \beta) = \left(\frac{2}{\epsilon\pi\beta}\right)^{1/2} \exp\left[-\frac{F^2 + \alpha^2 F_C^2}{2\epsilon\beta}\right] \times \cosh\left(\frac{2\alpha F F_C}{2\epsilon\beta}\right) \tag{7}$$

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 the true structure factor amplitude, F_C is the current model structure factor amplitude, while α and β are likelihood distribution parameters (Lunin & Urzhumtsev, 1984). For the refinement of structures 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

$$\begin{aligned} L(F_C|I_o) &= f(I_o|F_C, \alpha, \beta, \sigma_I^2) \\ &= \int_0^\infty f(I_o|\sigma_I^2, F) f(F|F_C, \alpha, \beta) dF \end{aligned} \tag{8}$$

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, 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 taking into account uncertainties of the estimated standard deviation as well. 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 normal distribution, which arises as the distribution choice when the true standard deviation is approximated by an estimate from experimental data. The aim of this work is to derive efficient means to obtain target functions that can provide an additional performance boost when working with very marginal data, such as obtained from time-resolved serial crystallography or Free Electron Laser data in which experimental errors are typically worse than obtained using standard rotation-based methods or have non-standard error models (Brewster *et al.*, 2019). Furthermore, high-quality datasets are rarely resolution limited by the diffraction geometry alone, indicating

that much more marginal data is 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 an intensity based likelihood functions and its derivatives that take into account the presence of experimental errors, both in the mean intensity and its associated standard deviation.

2. Methods

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

$$\begin{aligned} f_a(E, Z_o|E_C, \sigma_A, \sigma_Z^2) &= \frac{2E}{1-\sigma_A^2} \exp\left[-\frac{E^2+\sigma_A^2 E_C^2}{1-\sigma_A^2}\right] \\ &\times I_0\left(\frac{2\sigma_A E E_C}{1-\sigma_A^2}\right) \\ &\times f(Z_o|E, \sigma_Z^2, \nu) \end{aligned} \tag{9}$$

for acentric reflections, and

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

for centrics. When the distribution of the observed mean intensity Z_0 is modeled by a t-distribution, we have

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

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

$$\nu = N_{\text{eff}} - 1 \tag{12}$$

The effective sample size can be taken as the redundancy of an observed intensity, or be estimated using the Welch-Satterthwaite equation (Welch, 1947) to take into account 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 observations (**?**). The use of normal distribution on the other hand 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:

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

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

$$f_{\cdot}(Z_o|E_C,\sigma_A,\sigma_Z^2) = \int_0^\infty f_{\cdot}(E,Z_o|E_C,\sigma_A,\sigma_Z^2)dE \tag{14}$$

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" (Green, 1979) $(1-\sigma_A^2+\sigma_E^2)$. This approach circumvents the need to perform an integration, but is suboptimal in a number of different ways. Because we don't 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 doesn't rely on the accurate estimation of the mean intensity, possibly complicated by effects of pseudo-symmetry, diffraction anisotropy or twinning:

$$E_{\mathrm{o}} = \left(\frac{1}{2}\left(Z_{\mathrm{o}} + \sqrt{Z_{\mathrm{o}}^2 + 2\sigma_Z^2}\right)\right)^{1/2} \tag{15}$$

$$\sigma_E^2 = \frac{\sigma_Z^2}{4\sqrt{Z_{\mathrm{o}}^2 + 2\sigma_Z^2}} \tag{16}$$

Further details are listed in Appendix E. While this procedure allows for a straightforward intensity to amplitude conversion, even when intensities are negative, and can be subsequently used to inflate the variance of the Rice function, it is no substitute for the full integration. Given the simplicity of variance inflation approach and its wide usage in a number of crystallographic applications, we will use this approach as a benchmark, using both the conversion schemes based on the Wilson prior (denoted by French-Wilson) and the outlined uniform prior (denoted by 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 or Monte Carlo based methods or approaches based on orthogonal polynomials (Davies & 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, where 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 on how *ground truth* is established as a basis for comparison. Here we will limit ourself to the Laplace approximation due to its simplicity and the trapezoidal rules because excellent convergence properties when applied to analytic functions on the real line and its close relation to classic 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 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 (Davies & Rabinowitz, 1984). The change of variable theorem relates the integral of some function $g(u)$ under a change of variables $u = \psi(x)$:

$$\int_a^b g(u)du = \int_{\psi^{-1}(a)}^{\psi^{-1}(b)} g(\psi(x))\frac{d\psi(x)}{dx}dx \tag{17}$$

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 suitably chosen mean and length scale (Peng, 2018). The Laplace approximation can be derived from truncated Taylor expansion of the logarithm of the integrand:

$$\begin{aligned}\int_a^b f(x)dx &= \int_a^b e^{g(x)}dx \\ &= \int_a^b \exp\left[\sum_{n=0}^{\infty}\frac{g^{(n)}(x_0)}{n!}(x-x_0)^n\right]dx \\ &\approx \int_{-\infty}^{\infty} e^{g(x_0)}e^{-\frac{M}{2}|g''(x_0)|(x-x_0)^2}dx \end{aligned} \tag{18}$$

where $g(x) = \ln(f(x))$, and x_0 is the location of the maximum of $f(x)$ such that $g'(x_0) = 0$. Note that in the last step in equation (19), the assumption is made that $g(x)$ goes rapidly to $-\infty$, which allows the approximation that integrating over $[a, b]$ yields the same results as integrating over $\mathbb{R}$. Although this approximation doesn't work for all possible choices of $g(x)$, it has proven to be a successful tool in marginalising distributions in Bayesian analysis (Kass & Steffey, 1989) and crystallographic applications (**?**) .

This yields

$$\int_a^b f(x)dx \approx f(x_0)\sqrt{\frac{2\pi}{-g''(x_0)}} \tag{19}$$

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 first or higher-order optimization methods seems less straightforward using the Laplace approach. The difficulty arises in 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, which provides not only an easy way to increase the precision of the integral by increasing the number sampling points but also circumvents issues with computing derivatives encountered when using the Laplace approximation. Quadrature approaches have been assumed to need a large number of terms to get sufficient precision (Read & McCoy, 2016), possibly making them an unattractive target for practical crystallographic applications. In order to circumvent these issues, we combine the benefits from a Laplace approximation and a numerical quadrature and overcome drawbacks associated with both approaches. Here we construct a quadrature on the basis of a power transform followed by a logistic transform of the integrand, that maps the domain from $[0, \infty)$ onto $[0, 1]$, non-linearly compressing low-mass integrand regions on relatively small line segments, while approximately linearly transforming high-mass areas of the integrand to the middle of the new integration domain. Once a quadrature has been established, the logarithm of the integral can be recast as the log of the sum of weighted Rice

functions as outlined in Appendix D:

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

where E_j are the quadrature sampling points and w_j the associated weights and 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 Appendix 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 (21), derivatives with respect to $Y \in \{E_C, \sigma_A, \nu\}$ can be obtained as follows:

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

The derivatives of the Rice components $f.(E_j|E_C,\sigma_A)$ with respect to E_C are listed in Appendix C.

3. Results and Discussion

The first step to evaluate the proposed integration methods is to establish ground truth of the integral we wish to approximate. To this extend, an equispaced, non power-transformed trapezoidal quadrature was constructed integrating the function from $E = 0$ to $E = 6$ using fifty-thousand sampling points using all combination of distribution parameters as listed in Table 1 for the distributions under the assumption of

Gaussian intensity errors. Comparing the results of this integration to those obtained using a hyperbolic quadrature with fifteen-hundred points indicates that these two integration methods give virtually the same results. We therefore take ground truth as the results obtained with a hyperbolic quadrature using fifteen-hundred or more sample points. For both the acentric and the centric distributions, setting the power transform variable γ to 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 (in percent). A number of different approximation schemes were tested, comparing results using the mean relative error in the log-integrand. Because the variance inflation approximation doesn't 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 in its strengths and drawbacks. Instead, we will compare 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 to the variance inflation approaches. Furthermore, given that the gradients of the log-likelihood function form the basis of 3D difference or gradient maps that are used to complete or rebuild structures, comparing various approximations to the full likelihood function can provide insights when certain approximations fail. 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. Details of these tests and their results are found below.

3.1. Comparing integration methods

A comparison of the integration results using a number of different approximations visualized in Figure 2 for data sets generated according to a 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_0, E_C pairs, a likelihood computed using a 1500 point hyperbolic quadrature was treated as ground truth, both for a t-distribution and an error model assuming a normal distribution. These values were compared to the Laplace approximation and 7- 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 7 or more sampling points. General heuristics can in principle be developed to tailor the specific accuracy of the quadrtaure 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 as compared to those originating from a normal distribiution due to the presence of heavier tails.

3.2. Comparing likelihood functions

In order to get a better intuition of the behaviors of the target functions, we directly plot them for a few input parameters. Figure 3 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, the variance inflation method using both the French-Wilson and the Sivia approach, 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_0 on the basis of a non-informative prior (French-Wilson & Sivia) versus those derived by the full integration, Figure 3 a–c. In the case of an

observation with lower associated standard deviation, differences between 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 location of the maxima seem relatively unchanged, Figure 3d. 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 a 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 $\nu \to \infty$, 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 Figure 4. 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, Figure 5.

3.3. Comparing log-likelihood gradients

Additional insights in the behavior of the likelihood function approximations can be obtained by directly comparing its gradients for a selected set of parameter combinations. Numerical tests indicate that gradients computed using an fifteen-hundred-point hyperbolic quadrature of the power-transformed function (with γ set to 2 for both the acentric and centric distributions respectively) are indistinguishable from finite difference gradients computed with an fifty-thousand-point trapezoidal approach. In order to investigate the quality of the various approximations under common refinement scenarios, we construct a synthetic datasets using random sampling methods as out-

lined in Appendix F. A redundancy of 4 (ν=3) was used in these test. Gradients were computed using an 49 point quadrature, using a value of σ_A estimated from the corresponding approximation to the likelihood function. The results of these comparisons is shown in figure 5 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 gradients 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 under performs. 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 its 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 are 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 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 absence of predetermined good starting point of the line search. The construction of the hyperbolic quadrature doesn't require any iterative optimization, nor does the subsequent calculation of 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 will likely only have a minimal impact on the total run-time of the workflow, while providing a fast 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 under what situations 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 adopted to a different of choice of error model is a big 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).

The above algorithms are implemented in a set of python3 routines, and are available upon request. Some parts of this work was prepared in partial fulfillment of the requirements of the Berkeley Lab Undergraduate Research (BLUR) Program, managed by Workforce Development & Education at Berkeley Lab. 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. The content of this article is solely

the responsibility of the authors and does not necessarily represent the official views of the NIH.

Appendix A
A hyperbolic quadrature

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

$$G = \int_0^\infty g(x)dx \tag{22}$$

Set

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

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 due 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:

$$t = \frac{\exp[kx] - 1}{\exp[kx] + \exp[kx_0]} \tag{24}$$

Note that $t(x = 0) = 0$, and $\lim_{x\to\infty} t(x) = 1$. The inverse function is

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

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

$$k = \sqrt{\frac{-2h''(x_0)}{\pi}} \tag{29}$$

the above quadrature for $N = 1$ yields the Laplace approximation for 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 variable, 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

B.1. Rice functions, acentrics

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

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

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

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

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

B.2. Rice functions, centrics

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

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

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

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

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

B.3. Student's t-distribution

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

$$\begin{aligned} h_o(Z_o, E, \sigma_z^2, \nu) &= \ln f(Z_o|E, \sigma_o^2, \nu) \\ &= \ln \Gamma\left[\frac{\nu+1}{2}\right] \\ &- \frac{1}{2}\ln\left(\nu\pi\sigma_o^2\right) \\ &- \ln\Gamma\left[\frac{\nu}{2}\right] \\ &- \frac{\nu+1}{2}\ln\left(1+\frac{\left(Z_o-E^2\right)^2}{\nu\sigma_Z^2}\right) \end{aligned} \tag{38}$$

$$\begin{aligned} h_o'^E(Z_o, E, \sigma_z^2, \nu) &= \frac{d}{dE}\ln f(Z_o|E, \sigma_o^2, \nu) \\ &= \frac{2(\nu+1)E(Z_o-E^2)}{\sigma_Z^2\nu+(Z_o-E^2)^2} \end{aligned} \tag{39}$$

$$\begin{aligned} h_o''^E(Z_o, E, \sigma_z^2, \nu) &= \frac{d^2}{dE^2}\ln f(Z_o|E, \sigma_o^2, \nu) \\ &= 2(\nu^2+1) \\ &\times \frac{\left(\sigma_Z^2\nu(Z_o-3E^2)+(Z_o-E^2)^2(E^2+Z_o)\right)}{\left(\sigma_Z^2\nu+(Z_o-E^2)^2\right)^2} \end{aligned} \tag{40}$$

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

$$h_o(Z_o, E, \sigma_z^2) = -\frac{1}{2}\ln 2\pi + \ln\sigma_Z - \frac{\left(Z_o-E^2\right)^2}{2\sigma_Z^2} \tag{41}$$

$$h_o'^E(Z_o, E, \sigma_z^2) = \frac{2E\left(Z_o-E^2\right)}{\sigma_Z^2} \tag{42}$$

$$h_o''^E(Z_o, E, \sigma_z^2) = \frac{6E^2}{\sigma_Z^2} \tag{43}$$

Appendix C
Finding x_0

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

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

with Jacobian

$$\frac{dE(x)}{dx} = \gamma x^{\gamma-1} \tag{45}$$

The location of the maximum value of the integrand, x_0, is found using a straightforward application of Newton root finding algorithm. Set

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

and

$$h_{t,c}(E) = h_c(E, \cdots) + h_o(E, \cdots) \tag{47}$$

using the shorthand $h_t(E, \cdots)$ to indicate either of these function, we obtain the following functions and its derivatives under the power transform change of variables:

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

where the derivatives on the RHS are taken with respect with E as listed 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 is typically refines within 10 iterations to the supremums.

Appendix D
Likelihood synthesis

Using the above approaches, the full likelihood function can be expressed as a sum of weighted Rice functions (31,35), 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 $1 \leq j \leq N$. The quadrature weights can be now set to absorb the the hyperbolic sampling, the power transform and the observed intensity and its associated standard deviation

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

which yields a sum of weighted Rice functions that approximates the full likelihood function:

$$L.(E_C|Z_o,\sigma_A,\sigma_Z^2) = \sum_{j=1}^{N} w_j f.(E_j|E_C,\sigma_A) \tag{52}$$

where $f.(\cdots)$ is the appropriate Rice function.

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

$$L.(E_C|Z_o, \sigma_A, \sigma_Z^2) = w_0 f.(E_j|E_C, \sigma_A) \tag{53}$$

with the weight given by

$$\begin{aligned} w_0 &= \gamma x_0^{\gamma-1} \\ &\times f(Z_o|E_j, \sigma_Z^2, \nu) \\ &\times \sqrt{\frac{2\pi}{\hat{h}_{\cdot}''(x_0)}} \end{aligned} \tag{54}$$

where $E_0 = x_0^{\gamma}$, and $\hat{h}_t''(x_0)$ is defined in expression (49).

Appendix E
Structure factor amplitudes 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 deviation. While this process is normally performed using a standard French-Wilson estimation procedure, an other 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

$$f(E|Z_o, \sigma_Z^2) \propto E \exp\left(-\frac{(E^2 - Z_o)^2}{2\sigma_Z^2}\right) \tag{56}$$

A normal approximation to this distribution can be obtained by the method of moments or, as done 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:

$$E_o = \sup_E \ln\left[f(E|Z_o,\sigma_Z^2)\right] \tag{57}$$

$$\sigma_E^2 = -\left[\frac{d^2}{dE^2}\ln\left[f(E|Z_o,\sigma_Z^2)\right]\right]^{-1}_{E=E_o} \tag{58}$$

An analytic expression is readily obtained, resulting in

$$E_o = \left(\frac{1}{2}\left(Z_o+\sqrt{Z_o^2+2\sigma_Z^2}\right)\right)^{1/2} \tag{59}$$

$$\sigma_E^2 = \frac{\sigma_Z^2}{4\sqrt{Z_\mathrm{o}^2+2\sigma_Z^2}} \tag{60}$$

The quality of this approximation will critically depend on the value of Z_o and σ_Z^2. Note that standard error propagation on $E_o=\sqrt{Z_o}$ yields

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

which can be seen to be converge to expression (60) for the case that Z_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:

- $\mathbf{E}_\mathrm{True} = (X_\Re, X_\Im)\,; f(X.) = N(0,1/2)$
- $\mathbf{\Delta}_C = (X_\Re, X_\Im)\,; f(X.) = N(0,(1-\sigma_A^2)/2)$
- $\mathbf{E}_C = \sigma_A\mathbf{E}_\mathrm{True} + \mathbf{\Delta}_C$

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

- $\mathbf{E}_{\text{True}} = (X_{\Re}, 0)\,; f(X_{\Re}) = N(0, 1)$
- $\mathbf{\Delta}_C = (X_{\Re}, 0)\,; f(X_{\Re}) = N(0, (1 - \sigma_A^2))$
- $\mathbf{E}_C = \sigma_A \mathbf{E}_{\text{True}} + \mathbf{\Delta}_C$

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

- $Z_{\text{True}} = |\mathbf{E}_{\text{True}}|^2$
- $Z_j = X; f(X) = N(Z_{\text{True}}, (\nu + 1)\sigma^2_{\text{target}})$
- $Z_o = (1/(\nu + 1)) \sum Z_j$
- $\sigma_Z^2 = (1/(\nu(\nu + 1))^2) \sum (Z_j - Z_o)^2$

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, σ_{target} is different for every simulated intensity, and chosen to be Z_{True}/τ, where τ is equal to the desired $\mathbb{E}\left[Z_{\text{True}}/\sigma_{\text{target}}\right]$ level. For the fixed error level method, σ_{target} is fixed at $1/\tau$ for all intensities. A complete dataset is simulated assuming 9 to 1 ratio of acentric to centric reflections.

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. Reported are the mean error and standard deviation of the relative log-likelihood over the full parameter range (in percent).*

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. Reported are the mean error and standard deviation of the relative log-likelihood over the full parameter range (in percent). 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 ($N = 3$)	–	0.30 / 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

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%

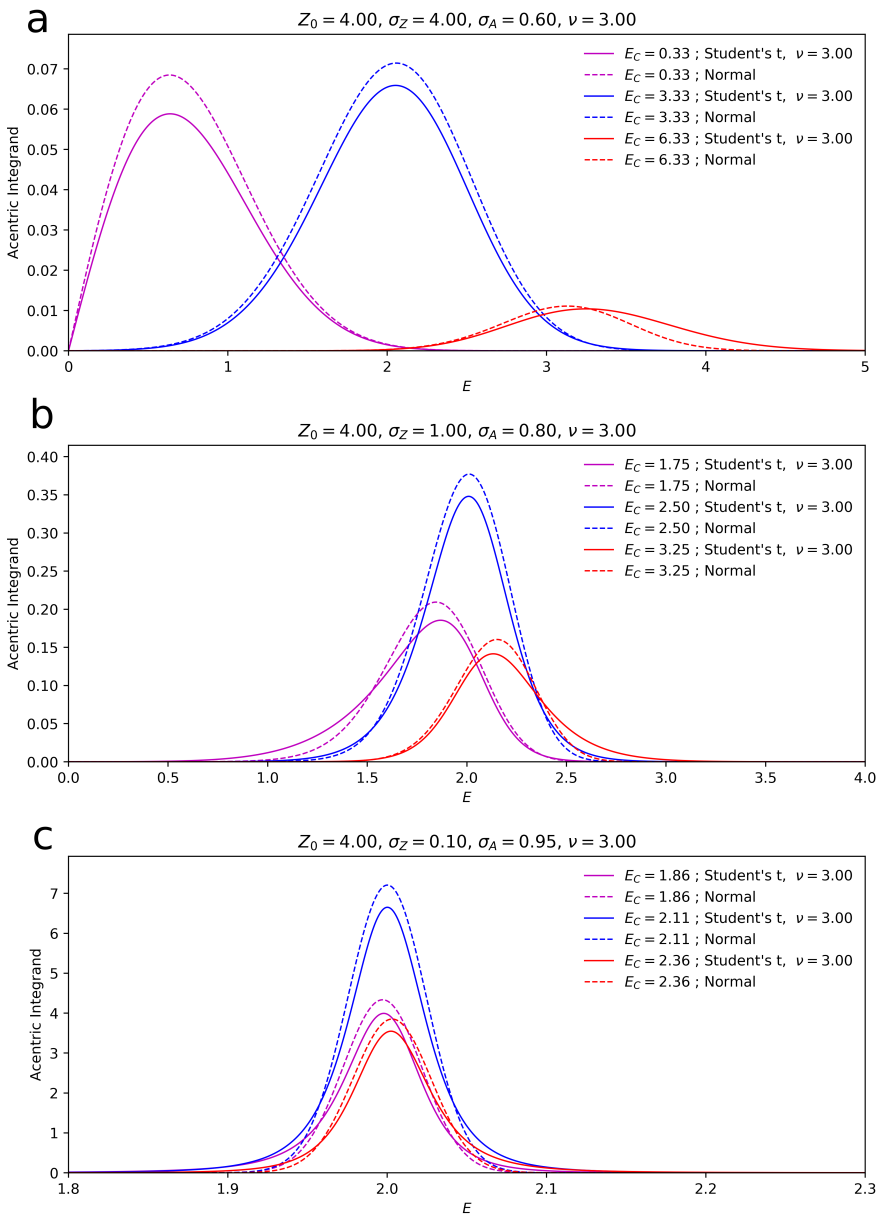

Fig. 1. Integrand shapes for the acentric and centric distribution for different parameter settings show the variety of function shapes that occur when computing the marginal likelihood. When the experimental error is relative large with respect to the intensity, high-mass areas of the function span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

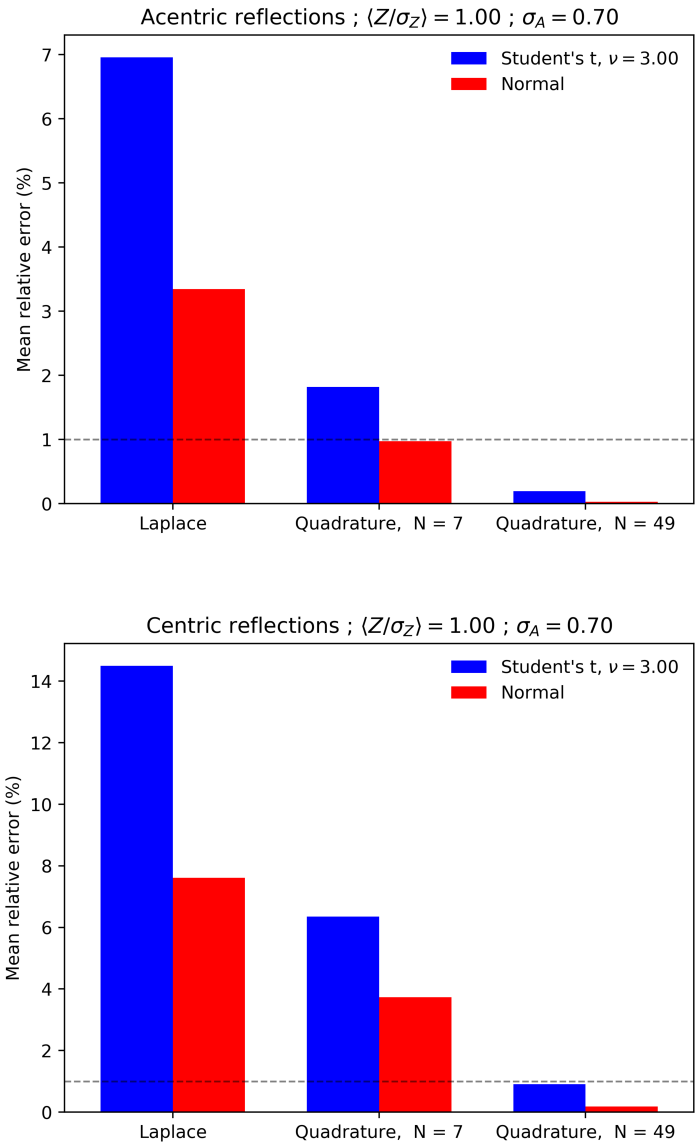

Fig. 2. The relative mean error of the likelihood functions using a Laplace approximation and quadrature based methods for normal and t-based noise models, for acentric (a) and centric reflections (b). The dotted horizontal like is set at 1% as a visual reference.

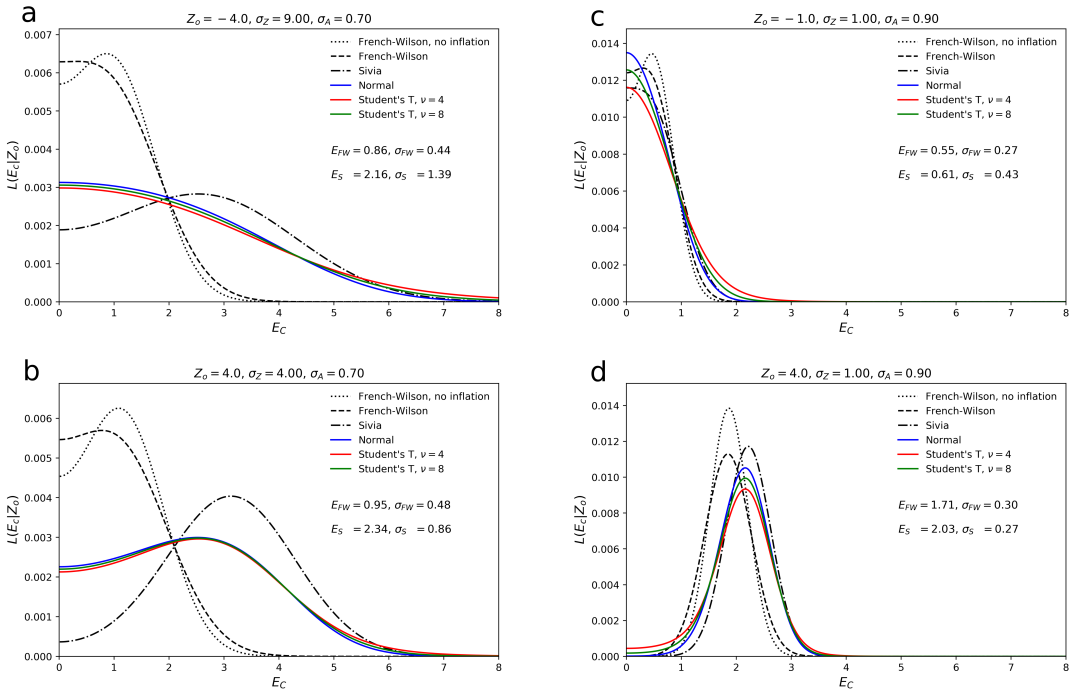

Fig. 3. The shape of normalized likelihood functions under a number of different approximations for different input parameters indicates that the use of a point estimate for negative intensities or those with high noise values results in significant deviations from the ideal likelihood function. The difference between a t-based noise model and a normal noise model is small, but significantly effects 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.

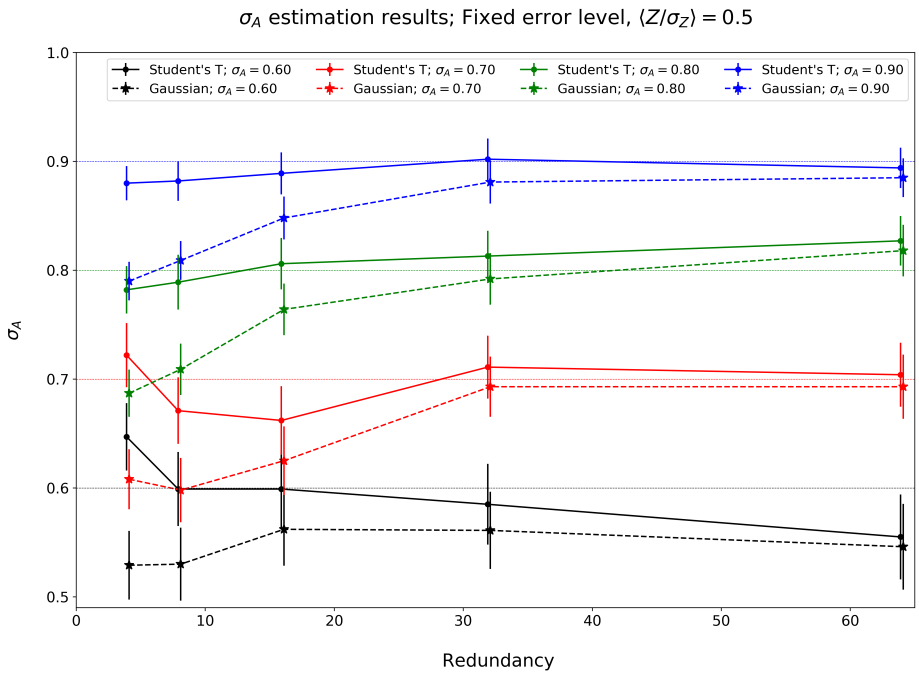

Fig. 4. The behavior of a likelihood-based σ_A estimation procedure when data with a t-based noise model is treated with a likelihood based approach using normal noise introduces a negative bias in the estimate of σ_A at low redundancies.

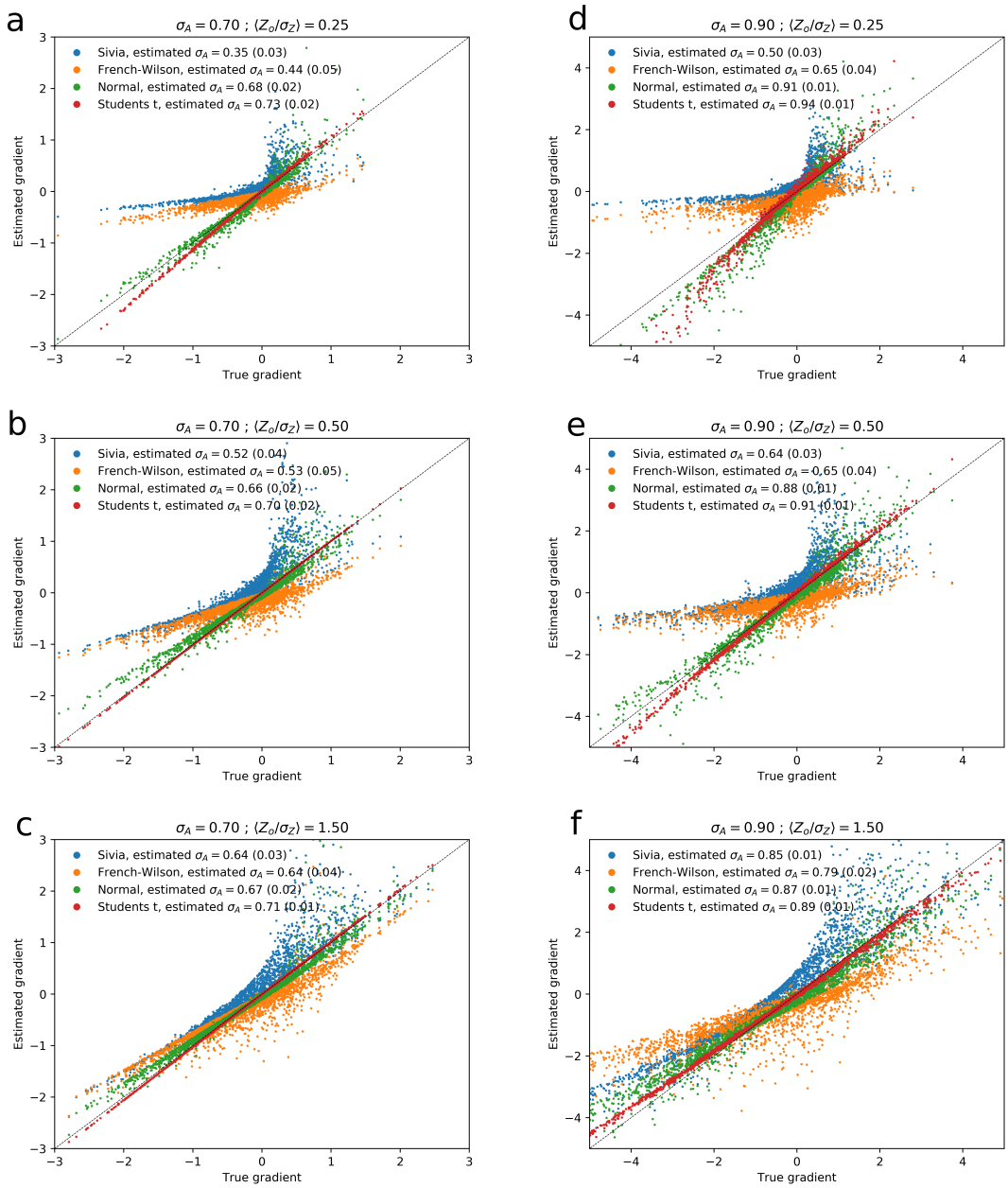

Fig. 5. A comparison of gradients computed using different approximation schemes for t-based noise with $\nu = 3$. Plots a–c and d–f depict the behavior of the gradient approximations with a decreasing noise level. Gradients were computed using an maximum likelihood estimate of σ_A using their corresponding approximations. Both the normal model and the t-based model clearly outperform the French-Wilson and the Sivia approach, while a marginal improvement over the normal noise model is obseved when the t-based model is used.

References

Beu, K. E., Musil, F. J. & Whitney, D. R. (1962). *Acta Crystallographica*, **15**(12), 1292–1301.

Brewster, A. S., Bhowmick, A., Bolotovsky, R., Mendez, D., Zwart, P. H. & Sauter, N. K. (2019). *Acta Crystallogr D Struct Biol*, **75**(Pt 11), 959–968.

Bricogne, G. (1997). *Proceedings of the CCP4 Study Weekend*.

Bricogne, G. & Gilmore, C. J. (1990). *Acta Crystallographica Section A*, **46**(4), 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**(2), 127–30.

Cools, R. (2002). *Journal of Computational and Applied Mathematics*, **149**(1), 1 – 12.

Cowtan, K. (2000). *Acta Crystallographica Section D*, **56**(12), 1612–1621.

Davies, P. J. & Rabinowitz, P. (1984). *Methods of Numerical Integration*. Elsevier.

Fisher, R. A. (1915). *Biometrika*, **10**(4), 507–521.

French, S. & Wilson, K. (1978). *Acta Crystallographica Section A*, **34**(4), 517–525.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, vol. 7. Perthes et Besser.

Gauss, C.-F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*, vol. 1. Henricus Dieterich.

Gauss, K. F. (1816). *Zeitschrift für Astronomie und verwandte Wissenschaften, herausgegeben von B. von Lindenau und JG F. Bohnenberger*, **1**(158), 1816.

Green, E. A. (1979). *Acta Crystallographica Section A*, **35**(3), 351–359.

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

Kass, R. E. & Steffey, D. (1989). *Journal of the American Statistical Association*, **84**(407), 717–726.

Kiefer, J. (1953). *Proceedings of the American Mathematical Society*, **4**(3), 502–502.

de La Fortelle, E. & Bricogne, G. (1997). In *Macromolecular Crystallography Part A*, vol. 276 of *Methods in Enzymology*, pp. 472 – 494. Academic Press.

Lunin, V. Y. & Skovoroda, T. P. (1995). *Acta Crystallographica Section A Foundations of Crystallography*, **51**(6), 880–887.

Lunin, V. Y. & Urzhumtsev, A. G. (1984). *Acta Crystallographica Section A*, **40**(3), 269–277.

Luzzati, V. (1952). *Acta Crystallographica*, **5**(6), 802–810.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Crystallographica Section D*, **61**(4), 458–464.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Crystallogr D Biol Crystallogr*, **53**(Pt 3), 240–55.

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

Pannu, N. S. & Read, R. J. (1996). *Acta Crystallographica Section A Foundations of Crystallography*, **52**(5), 659–668.

Pearson, E. (1970). *Studies in the History of Statistics and Probability (London, 1970)*, pp. 411–13.

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

Read, R. J. (1986). *Acta Crystallographica Section A*, **42**(3), 140–149.

Read, R. J. (1997). In *Macromolecular Crystallography Part B*, vol. 277 of *Methods in Enzymology*, pp. 110 – 128. Academic Press.

Read, R. J. (2001). *Acta Crystallographica Section D*, **57**(10), 1373–1382.

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

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

Sharma, A., Johansson, L., Dunevall, E., Wahlgren, W. Y., Neutze, R. & Katona, G. (2017). *Acta Crystallographica Section A*, **73**(2), 93–101.

Sim, G. A. (1959). *Acta Crystallographica*, **12**(10), 813–815.

Sivia, D. S. & David, W. I. F. (1994). *Acta Crystallographica Section A*, **50**(6), 703–714.

Skubák, P., Waterreus, W.-J. & Pannu, N. S. (2010). *Acta Crystallographica Section D*, **66**(7), 783–788.

Srinivasan, R. Parthasarathy, S. (1976). *Some statistical applications in X-ray crystallography.* New York: Pergamon Press, 1st ed.

Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Crystallographica Section D*, **60**(3), 432–438.

Terwilliger, T. C. (2000). *Acta Crystallographica Section D*, **56**(8), 965–972.

Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Crystallographica Section A*, **39**(5), 813–817.

Trefethen, L. N. & Weideman, J. A. C. (2014). *SIAM Review*, **56**(3), 385–458.

Welch, B. L. (1947). *Biometrika*, **34**(1/2), 28–35.

Wilks, S. S. (1938). *Ann. Math. Statist.* **9**(1), 60–62.

Wilson, A. J. C. (1980). *Acta Crystallographica Section A*, **36**(6), 937–944.

Woolfson, M. M. (1956). *Acta Crystallographica*, **9**(10), 804–810.

Synopsis

A quadrature is developed that allows for the efficient evaluation of an intensity-based likelihood target function that includes experimental errors.