Least-squares and maximum-likelihood in Computed Tomography

Murdock G. Grewar[a], Glenn R. Myers[a,b], and Andrew M. Kingston[a,b]

[a]Dept. of Applied Mathematics, RSPhys, The Australian National University, Canberra, ACT 2601, Australia

[b]CTLab: National Laboratory for Micro Computed-Tomography, Advanced Imaging Precinct, The Australian National University, Canberra, ACT 2601, Australia.

ABSTRACT

Statistical reconstruction methods in X-ray Computed Tomography (XCT) are well-regarded for their ability to produce more accurate and artefact-free reconstructed volumes, in the presence of measurement noise. Maximum-likelihood methods are particularly salient and have been shown to result in superior reconstruction quality, compared with methods that minimise the ℓ^2 residual between measured and projected line attenuations. Least-squares more generally may refer to the minimisation of quadratic forms of the projected attenuation residuals. Early maximum-likelihood methods showed promising reconstruction capabilities but were not practical to implement due to very slow convergence, especially compared with least-squares methods. More recently, least-squares methods have been adapted to minimise quadratic approximations to (negative) log-likelihood, thereby attaining the speed of least-squares minimisation in service of likelihood maximisation for superior reconstruction fidelity.

Quadratic approximation to the log-likelihood under Poisson measurement statistics has been demonstrated several times in the literature. In this publication we describe an approach to quadratically expanding log-likelihood under an arbitrary noise model, and demonstrate via simulation that this can be implemented practically to maximise likelihood under mixed Poisson-Gaussian models that describe a broad range of transmission XCT imaging systems.

Keywords: computed tomography, Maximum Likelihood, Quadratic Form, Least Squares, Generalised Least Squares

1. INTRODUCTION

The highly-penetrating nature of X-rays enables X-ray Computed Tomography (XCT) scans to perform projection measurements on a physical sample. The (transformed) measurement vector, $\mathbf{m}$, is comprised of the integrated X-ray attenuation coefficients of the object along lines from the X-ray source, through the object, and to the X-ray detector. This transformation from a volume $\mathbf{x}$ to measurements $\mathbf{m}$ is modelled by the projection operator $\mathcal{A}$, i.e. $\mathbf{m} = \mathcal{A}(\mathbf{x})$. $\mathcal{A}$ is a mathematical encoding of the physical computation which takes place during X-ray scanning. Real measurements are always subject to errors and noise, at the very least due to the quantum nature of light which induces Poisson statistics in the measured X-ray flux ('shot noise'). Under the simplest mathematical model of attenuation-based monochromatic X-ray scanning, $\mathcal{A}$ becomes a noiseless, linear operator, $\mathbf{A}$, and the XCT problem is reduced to the linear equation $\mathbf{m} = \mathbf{A}\mathbf{x}$, in which we desire the solution for the volume $\mathbf{x}$, thereby characterising the physical sample. Inescapable noise in the measurement vector $\mathbf{m}$ always deprives us of an exact solution for $\mathbf{x}$, and therefore necessitates a regularisation scheme, i.e. in practice we solve for $\mathbf{m} \approx \mathbf{A}\mathbf{x}$. For our purposes, a 'regularisation scheme' is the method and conditions under which this inexact equation is considered 'solved'. Among the simplest and most computationally amenable regularisation schemes are those which minimise a quadratic form of the projection residual $(\mathbf{A}\mathbf{x} - \mathbf{m})$. Quadratic forms are valued for their global convexity and derivative-linearity, which allow them to be rapidly minimised. They can represent

Further author information: (Send correspondence to A. Kingston)
A. Kingston e-mail: andrew.kingston@anu.edu.au

Developments in X-Ray Tomography XIII, edited by Bert Müller
Ge Wang, Proc. of SPIE Vol. 11840, 1184014 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2595559

log-likelihood under multivariate Gaussian models of the measured line attenuations, in which case minimising the quadratic form maximises the likelihood of the reconstructed volume under the corresponding Gaussian model. Maximum-likelihood methods are generally regarded to provide better reconstruction quality, both in practice and from a theoretical perspective. Historical maximum-likelihood methods were based on the famous 'expectation-maximisation' algorithm, and seem to be of primarily theoretical interest due to their slothful convergence. More recent maximum-likelihood methods for Poisson noise have used approximate quadratic models in order to leverage the desirable properties of quadratic forms to produce statistically-optimal reconstructions.

Our goal in this publication is to illuminate the general method for producing maximum-likelihood reconstructions via quadratic approximations to log-likelihood. This procedure is not wholly original, but neither do we think that this excellent technique been discussed with the attention and generality that it warrants. We demonstrate its efficacy with two novel applications in monochromatic transmission tomography, producing algorithms to rapidly maximise likelihood under (1) a convolved Poisson-Gaussian distribution on individual measurements, and (2) a Poisson model which incorporates intensity blur on the X-ray detector.

Our discussion begins in §2 where we describe the theoretical origins of the linear transmission model $\mathbf{m} = \mathbf{A}\mathbf{x}$. We also give a brief review of some existing methods in the XCT literature minimising quadratic forms of the the projection residual $(\mathbf{A}\mathbf{x} - \mathbf{m})$, some of which date to 1972 among the earliest published algorithms in XCT. Following in §3, we enunciate the disruption caused to the linear model in the presence of measurement noise. Most early XCT reconstruction methods had essentially undefined behaviour when the linear model could not be satisfied, leading to favouritism among algorithms meant for the exact reconstruction problem. We also discuss the advent of maximum-likelihood methods as well a prescient publication for the minimisation of diagonal quadratic forms. Subsequently in §4 we describe explicitly what is meant by a quadratic approximation to log-likelihood, and how quadratic minimisation may be used to maximise likelihood under essentially arbitrary measurement noise models. In §5 we present an explicit quadratic expansion under a general convolved Poisson-Gaussian noise model in measurements. We also provide suggestions on how best to compute the expansion numerically. A reconstruction is illustrated using this method on simulated data (with the commensurate noise model) and compared with the outcomes from Poisson likelihood maximisation, Gaussian likelihood maximisation, and ℓ^2 minimisation. Next in §6 we present the explicit expansion in the case of Poisson noise, where the detector X-ray flux is blurred prior to the application of Poisson statistics. We illustrate and compare reconstruction results on a simulation, similar to the Poisson-Gaussian example. In §7 we describe the simulation method, depict additional information regarding the reconstructions, and discuss the results of the simulations. Finally in §8 we conclude the paper and suggest future research opportunities.

2. BACKGROUND

Germinating in medical practice,[1] X-ray computed tomography (XCT) leverages the penetrative power of high-energy photons to non-destructively image samples *in vivo*, in 3 dimensions. Since this initial development, XCT has been applied more generally to materials study where it serves as an indispensable tool for volumetric non-destructive evaluation. XCT was ideated at the latest by the 1960s[2,3] and exploded in popularity by the early 1970s after the first tomogram of a human brain was produced by *Hounsfield et. al.*[1] In order to institute theory into practice, there are two necessary ingredients: (1) a mathematical-physical model which relates the properties of a prospective volume to the expected radiographs (measured X-ray projection signal) of said volume, and (2) a computational algorithm which will transform the recorded radiographs into a 3-dimensional (3D) data structure representing volume information. The algorithm (2) serves as a numerical computational method to invert the mathematical model (1). A perennial trade-off lies in increasing the complexity of the physical model, which improves quantitative accuracy at the cost of increasing the complexity of the inversion problem, often to the extent that it becomes intractable. Examples of this complexity are difficulty in accounting for polychromatic X-ray attenuation (i.e. 'beam hardening')[4–6] and multiple Compton scatter in metallic samples.[7]

The most basic physical model, which is still largely in use, is the empirically-established *Beer-Lambert* model. This model captures quantitatively only the most essential qualitative feature of X-rays: that they are *locally* and *linearly* attenuated as they pass through a volume. The resulting prediction is that X-rays should be attenuated exponentially with distance through the sample. We assume that X-rays pass through the object without deflection, i.e. along linear paths. Consider the intensity I (i.e. photon flux) of X-rays along one such

linear path $\mathbf{p}(s)$ through the volume, parametrised by distance s. According to the assumptions of linearity and locality, the intensity must obey the differential equation

$$\frac{\mathrm{d}I}{\mathrm{d}s} = -I \cdot \boldsymbol{\mu}(\mathbf{p}(s)) \tag{1}$$

where $\boldsymbol{\mu}$ is a scalar field representing some physical property of the volume. This property is empirically measured, and is termed the *attenuation coefficient* for the specific X-ray energy in question. Under the *Beer-Lambert* model, a linear equation is recovered relating the attenuation coefficient $\boldsymbol{\mu}$ of the volume to a transformation of the measured intensity X-ray intensity I^m compared with its expected value I^c if no attenuating sample were present (clearfield intensity):

$$-\ln \frac{I^m}{I^c} = \int_s \boldsymbol{\mu}(\mathbf{p}(s))\,\mathrm{d}s\,. \tag{2}$$

The goal of X-ray transmission tomography is to reassemble the 3D quantity $\boldsymbol{\mu}$ from our knowledge of the measurements I^m and expected clearfield intensities I^c. In order to make this problem numerically approachable, the measurements and volume are discretised, so that the equation may be rewritten in discretised, vectorised form as

$$\mathbf{m} = \mathbf{A}\mathbf{x} \tag{3}$$

where $\mathbf{m}$ is a vector of transformed measurements as per the left-hand-side (LHS) of (2), $\mathbf{x}$ is a discrete vectorised form of the volume attenuation coefficient $\boldsymbol{\mu}(\mathbf{p})$, and each row of the linear operator $\mathbf{A}$ is a discretised form of the corresponding integral in (2).

We described earlier that the XCT problem has two ingredients: (1) a mathematical-physical model, and (2) an inversion algorithm for that model. We have just described the model that was used in the first XCT reconstructions. Since it admits a linear equation, it is amenable to inversion via numerous common methods.[8,9] Nonetheless, bespoke algorithms emerged in the early 1970s which would invert this model and so produce an estimate of the attenuation volume $\mathbf{x}$ ($\mathbf{x}$ represents the same physical information as the 3D quantity $\boldsymbol{\mu}$, but in vector form). Among the most seminal are *Gordon et. al*'s *Algebraic Reconstruction Technique* (ART) (1970),[10] *Peter Gilbert*'s *Simultaneous Iterative Reconstruction Technique* (SIRT) (1972),[11] later *Andersen & Kak*'s *Simultaneous Algebraic Reconstruction Technique* (SART) (1984),[12] and filtered backprojection (FBP) methods.[13] These algorithms work well when the acquired X-ray data is consistent with the linear model. In particular, SART and FBP arrive very quickly at a solution. Nonetheless, the SIRT method has remained popular since it is robust to imaging noise with early termination of iterations being a form of regularisation.[14,15] In other terms, SIRT produces better solutions for $\mathbf{x}$ when the data $\mathbf{m}$ is subject to *measurement noise*.

3. REGULARISED INVERSION AND MAXIMUM LIKELIHOOD

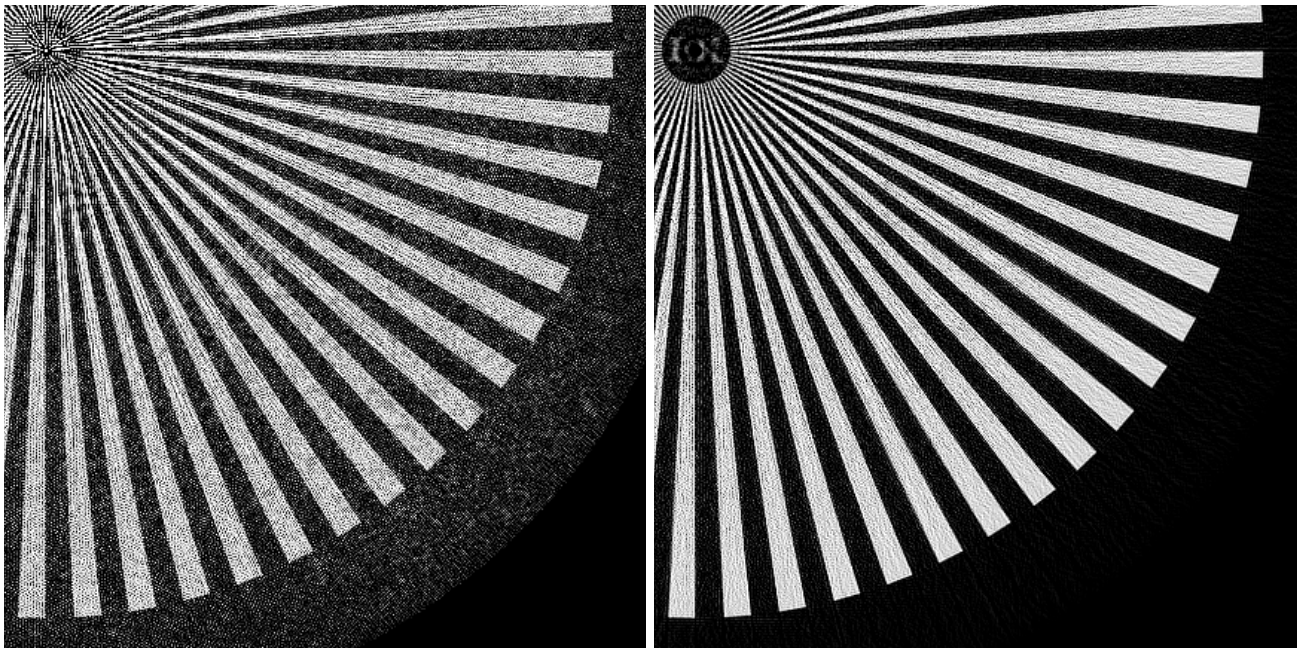
Measurement noise may be modelled as a random component added to the transformed measurements $\mathbf{m}$. Due to the geometry of imaging, the projection operator $\mathbf{A}$ has some linearly dependent rows in all physical experiments, so that the equation $\mathbf{m} = \mathbf{A}\mathbf{x}$ is partially overdetermined for a general measurement vector $\mathbf{m}$. When measurements are noise-free, then $\mathbf{m} \in \mathrm{range}(\mathbf{A})$ and this does not pose an issue for the inversion problem– as long as sufficient data is captured, then a left-pseudoinverse[16] $\mathbf{A}^L$ can be applied to measurements to recover the volume:

$$\mathbf{x} = \mathbf{A}^L\mathbf{m}\,. \tag{4}$$

Effectively, the aforementioned reconstruction algorithms are algorithmic implementations of $\mathbf{A}^L$. They recover an estimate of the volume attenuation $\mathbf{x}$ from the measurements $\mathbf{m}$, by applying a sequence of operations which can invert *in silico* the projection operator after it has been applied to the volume via a physical calculation (i.e. experiment). When noise is added to $\mathbf{m}$ however, then $\mathbf{m} \notin \mathrm{range}(\mathbf{A})$ with probability 1, and there exists no volume solution $\mathbf{x}$ which can explain the measurements per $\mathbf{m} = \mathbf{A}\mathbf{x}$. At face value, the behaviour of a left-pseudoinverse $\mathbf{A}^L$ is undefined for elements which are not in the range of $\mathbf{A}$– the outcome of applying a reconstruction algorithm $\mathbf{A}^L$ is dependent on the specific implementation, i.e. the specific algorithm used. This explains why some algorithms (e.g. FBP, SART) are found to be less favourable than others (e.g. SIRT) though they were designed to solve the same problem. The SART algorithm does not converge in the presence of

noise, while SIRT and FBP methods act as Moore-Penrose left-pseudoinverses[16,17] which minimise the ℓ^2 norm $||\mathbf{m} - \mathbf{A}\mathbf{x}||^2$ (discounting the numerical instability of FBP).[17,18] The way in which an inversion algorithm deals with input which has no solution (such as measurements $\mathbf{m}$ with noise included) comprises its *regularisation*, at least for our purposes.

Since every experiment has measurement noise, and the presence of noise is an ever-pressing concern when attempting to minimise patient dose or scanning time, the question arises: how best do we regularise a reconstruction algorithm to be robust to noise? This is an important question which, if answered appropriately, can have a dramatic impact on the quality of reconstructions– an example is depicted in Fig. 1. A common approach to dealing with measurement noise is to model the statistical distribution of $\mathbf{m}$ explicitly as a probability density function (PDF) of the attenuating volume $\mathbf{x}$ and produce a *maximum-likelihood* solution for the volume. The first seminal papers on this technique, as applied to XCT, appear to date back to the early 1980s[19–21] and utilise the famous expectation-maximisation (EM)[22] algorithm published in 1977. However, a prescient paper by *Michael Goitein* (1972)[23] also maximised likelihood. This method involved minimising a weighted least-squares (WLS) expression $||\mathbf{D}(\mathbf{m} - \mathbf{A}\mathbf{x})||^2 = ||\mathbf{m} - \mathbf{A}\mathbf{x}||^2_{\mathbf{D}^2}$ where $\mathbf{D}$ is a diagonal matrix. The components of $\mathbf{D}$ represent the relative uncertainty in each measurement component in $\mathbf{m}$. This is a maximum-likelihood method when the noise model considers each measurement m_k as a Gaussian random variable with mean $(\mathbf{A}\mathbf{x})_k$ and standard deviation d_k (the kth component of the diagonal of $\mathbf{D}$). The standard ℓ^2 metric minimised by FBP and SIRT produces a maximum-likelihood solution precisely when these Gaussian distributions have uniform variance ($d_i = d_j$). Unfortunately, a uniform Gaussian distribution on the measured line attenuations $\mathbf{m}$ is not a real occurrence. This conclusion can be motivated as follows: the measured intensities $-\ln I^m/I^c$ are *noiselessly transformed as the logarithm* of measured X-ray intensity measurements I^m. These measurements are subject to Poisson statistics in all cases due to the quantum nature of light, and in most instances the other sources of noise may be reduced to an additive Gaussian component.[24] To a reasonable approximation, then, an all-encompassing noise archetype for most XCT imaging systems is the *convolved Poisson-Gaussian* noise model on individual intensity measurements, the logarithm of which is certainly not Gaussian. This model still neglects effects such as penumbral blurring from a finitely-sized X-ray source, and other effects that mix measurements.



(a) ℓ^2 with 0-photon measurements removed (b) Quadratic approx. to Maximum Likelihood

Figure 1: Simulated example of two converged reconstructions on identical data. Each solves $\mathbf{m} = \mathbf{A}\mathbf{x}$, but with a different regularisation scheme. Each minimises a quadratic form indicated in the corresponding caption. Data was subject to Poisson statistics in measured photon counts.

The question may arise in the reader: if the uniform Gaussian model on line attenuations is so physically implausible, why then does ℓ^2 minimisation via SIRT often work so well for combating noise? The likely answer is that SIRT is rarely iterated enough times to reach convergence– in fact there is a misconception that SIRT *diverges* due to the noise introduced in the limit of many iterations (e.g. see Fig. 1). However this noise is bounded and does not appear to increase further once the algorithm has converged upon a minimum value for the ℓ^2 norm. The 'salt-and-pepper' noise is a bona fide part of the minimum ℓ^2 solution and *not a numerical artefact*. Recall that both FBP and SIRT minimise the ℓ^2 norm. When numerical instabilities are resolved and the output of both algorithms are compared, they produce almost identical tomograms. Assuming that projection data is complete (i.e. $\mathbf{A}$ is injective) then SIRT and iterated FBP:

$$\Delta\mathbf{x} = \text{diag}(\mathbf{A}^{\mathrm{T}}\mathbf{A}\underline{\mathbf{1}}_v)^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{m}-\mathbf{A}\mathbf{x}) \tag{5}$$

$$\Delta\mathbf{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{m}-\mathbf{A}\mathbf{x}) \tag{6}$$

are both gradient descent algorithms, with the latter more strongly conditioned so that it reaches a solution (ideally) in a single iteration. Iterated FBP methods apply the Moore-Penrose left-inverse of $\mathbf{A}$, $\mathbf{A}^{\mathrm{MPL}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$, more-or-less in its exact form, directly to the projected attenuation residual $(\mathbf{m}-\mathbf{A}\mathbf{x})$ and so minimise ℓ^2 in a single iteration, if we set aside issues of numerical instability. SIRT instead applies an (poor) approximation $\mathbf{A}^{\mathrm{MPL}} \approx \text{diag}(\mathbf{A}^{\mathrm{T}}\mathbf{A}\underline{\mathbf{1}}_v)^{-1}\mathbf{A}^{\mathrm{T}}$, reducing the convergence speed from a single iteration to hundreds, thousands or more. Both methods minimise the ℓ^2 metric via gradient descent ($\nabla_{\mathbf{x}}[||\mathbf{m}-\mathbf{A}\mathbf{x}||^2_{\ell^2}] = -2\mathbf{A}^{\mathrm{T}}(\mathbf{m}-\mathbf{A}\mathbf{x})$) but with different 'preconditioners' applied. This modifies the path and rate of convergence, but not the ultimate outcome, assuming that the projection data is complete.[17] It is ironic that the reputation SIRT has gained for its stability is actually a side effect of its poor convergence behaviour, and the tendency for users to regularise the algorithm by cutting it short before it 'diverges' by resolving fine details. However we see clearly in Fig. 1 (and later figures) that with the appropriate quadratic form, there is no need to regularise by early stopping, as alternative algorithms can achieve superior reconstructions when given the time to converge.

4. QUADRATIC APPROXIMATION

A particularly interesting class of noise models are those that predict a distribution on measurements $\mathbf{m}$ which depends only on the projected attenuations $\mathbf{A}\mathbf{x}$:

$$\ln P(\mathbf{m}|\mathbf{A}\mathbf{x}) = \ln \mathcal{L}(\mathbf{A}\mathbf{x}|\mathbf{m}) = \dots . \tag{7}$$

Every noise model can be expressed this way provided the scanning geometry contains complete data (as $\mathbf{A}$ is injective). Practically speaking however, regardless of whether data is complete, this is probably the most natural way to express any reasonable noise model, as the expected measurement intensities are typically considered a function of the line attenuations of the sample. For example, the log-likelihood of a candidate tomogram $\mathbf{x}$ under the ubiquitous Poisson noise model is given by

$$\ln P(\mathbf{I}^m|\mathbf{A}\mathbf{x}) = \ln \mathcal{L}(\mathbf{A}\mathbf{x}|\mathbf{I}^m) = \sum_i (I^c)_i e^{-m_i}\left(\ln(I^c)_i - (\mathbf{A}\mathbf{x})_i\right) - (I^c)_i e^{-(\mathbf{A}\mathbf{x})_i} - \ln\Gamma\left((I^c)_i e^{-m_i}+1\right) . \tag{8}$$

The summands (indexed by i) give the log-likelihood associated with each individual measurement. Γ is the gamma function, i.e. $\Gamma(k+1) = k!$ where $k \in \mathbb{N}$, although the term involving Γ may be ignored in the interest of maximising likelihood since it is independent of the choice of $\mathbf{x}$. Equation (8) encapsulates the statistics of X-ray intensity measurement under the simplest sensible noise model– one which accounts for the random arrival of discrete photons. This noise model is not just important in low-flux experiments, but in any case where the dynamic range in X-ray intensity is large across measurements. Practically speaking, any sample which is significantly attenuating to the X-rays (e.g. metals, but perhaps not biological tissue) will suffice for Poisson noise to be a worthy consideration. In Fig. 1 we depict one example of the difference in the reconstructed tomogram when the ℓ^2 norm is minimised cf. maximising likelihood under the Poisson noise model.

Noise models written in the form (7) may be expanded quadratically in the components of $\mathbf{A}\mathbf{x}$, so that the *log-likelihood residual* $\ln\mathcal{L} - \ln\mathcal{L}_{\max}$ is approximated by a quadratic form of the projected attenuations $\mathbf{A}\mathbf{x}$:

$$\ln \mathcal{L}(\mathbf{A}\mathbf{x}|\mathbf{m}) - \ln \mathcal{L}(\mathbf{M}|\mathbf{m}) \approx -\tfrac{1}{2}(\mathbf{A}\mathbf{x}-\mathbf{M})^{\mathrm{T}}\mathbf{Q}(\mathbf{A}\mathbf{x}-\mathbf{M}) \tag{9}$$

where $\mathbf{M}$ is a maximum-likelihood value* for $\mathbf{Ax}$, $^{\mathrm{T}}$ denotes a transpose, and $\mathbf{Q}$ is the negative Hessian matrix (symmetric positive semi-definite). Once a quadratic expansion is made, the problem reduces to maximising likelihood under a multivariate Gaussian model, for which the theory of generalised least-squares (GLS)[25–27] is applicable. There is a further reduction in complexity if each component of $\mathbf{m}$ is distributed independently as a function of the respective component of $\mathbf{Ax}$– in that case $\mathbf{Q}$ is diagonal and maximising likelihood amounts to a weighted least-squares (WLS) problem which is a subset of GLS. In either case, likelihood may be maximised under the quadratic expansion by solving the least-squares problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} -\tfrac{1}{2}||\mathbf{M} - \mathbf{Ax}||^2_{\mathbf{Q}} \tag{10}$$

where $||\mathbf{y}||_{\mathbf{Q}}$ indicates the '$\mathbf{Q}$'-norm of $\mathbf{y}$, equal to $\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{Q}\mathbf{y}} = \sqrt{\langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle}$. Quadratic expansion translates the likelihood-maximisation problem to a GLS problem by approximating the log-likelihood function as a concave paraboloid in $(\mathbf{Ax})-$space (right-hand-side (RHS) of (9) and of (10)). In WLS problems, the principal axes of that paraboloid are aligned with the coordinate axes of $(\mathbf{Ax})-$space. The reconstruction problem seeks to maximise the paraboloid subject to the constraint that $\mathbf{Ax}$ is the projection $\mathbf{A}$ of some candidate volume $\mathbf{x}$. In essence, the problem is quadratic optimisation constrained to the affine subspace range($\mathbf{A}$) of $(\mathbf{Ax})-$space. With the ready availability of advanced square-minimisation techniques, converting the maximum-likelihood problem to a constrained quadratic optimisation problem (10) makes reconstruction significantly more expedient than EM-based maximum likelihood (ML) algorithms.[19–21] Several more recent ML reconstruction methods are based upon quadratic approximation[28–32] as described here, and such methods are still an active area of research. All of these cited methods maximise likelihood under the Poisson model (8). The resulting objective function is an intensity-weighted sum-of-squares,

$$||\mathbf{m} - \mathbf{Ax}||^2_{\mathrm{diag}(\mathbf{I}^m)} \tag{11}$$

where $\mathrm{diag}(\mathbf{I}^m)$ is a diagonal matrix of measured photon counts. Any method minimising this objective function will achieve approximate maximum likelihood under the Poisson noise model. For example, the SIRT (5) algorithm can be adapted[17] to maximise likelihood like so:

$$\Delta\mathbf{x} = \mathrm{diag}(\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}\underline{\mathbf{1}}_v)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Q}(\mathbf{m} - \mathbf{Ax}) \tag{12}$$

where $\mathbf{Q}$ is substituted for $\mathrm{diag}(\mathbf{I}^m)$ for the Poisson noise model. The volume vector $\underline{\mathbf{1}}_v$ is filled with 1s in the domain of reconstruction, and 0 elsewhere. A vector $\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}\underline{\mathbf{1}}_v$ is formed by projecting that volume vector, applying the quadratic form $\mathbf{Q}$ as a linear operator, then backprojecting. The resulting vector acts as a multiplicative mask which divides the backprojected weighted projection residual $\mathbf{A}^{\mathrm{T}}\mathbf{Q}(\mathbf{m} - \mathbf{Ax})$. When $\mathbf{Q}$ is the identity operator, this reduces to the SIRT algorithm and solves for exact maximum likelihood under a noise model with uniform additive Gaussian noise in *transformed* intensity measurements (i.e. line attenuations).

The quadratic expansion method produces a sum-of-squares objective function (RHS of (10)) which is the *best possible quadratic approximation to log-likelihood*. (The objective function is the tangible quantity that the algorithm seeks to minimise/maximise.) This ensures that close to the expansion point $\mathbf{M}$, the likelihood of a prospective solution $\mathbf{Ax}$ is represented as well as possible. The direct correspondence between the objective function and likelihood is appealing since the former can be monitored during the course of iterative square-minimisation algorithms, providing a direct (if approximate) measure of log-likelihood during the course of reconstruction. It also means that the objective function is aligned with the goal of likelihood maximisation. Approximating log-likelihood quadratically ensures global concavity and a predictable gradient, which makes inversion theoretically much faster and more stable. In fact, the solution can in principle be computed directly in the style of FBP, though this remains an open problem and will likely always be subject to serious numerical stability issues.[17] This approach may be extended to maximise/minimise any chosen objective function of $\mathbf{Ax}$ (in light of measurements $\mathbf{m}$). There is no need to restrict ourselves to likelihood maximisation, although it is certainly an appealing application.

*Note that $\mathbf{Ax}$ is restricted to a subspace whereas $\mathbf{M}$ is freely chosen as a vector which maximises likelihood, even if there does not exist $\mathbf{x}$ such that $\mathbf{M} = \mathbf{Ax}$.

5. MAXIMUM LIKELIHOOD UNDER CONVOLVED POISSON-GAUSSIAN NOISE

Measurement noise in transmission X-ray imaging is generally well-modelled by a convolution of Poisson noise with Gaussian noise on each measurement.[24] The Poisson noise component is inherent in the quantum nature of photon detection, while the Gaussian component serves as a 'catch-all' for various other sources of noise, including noise introduced by instrumentation. The approach to quadratically expanding this highly general convolved Poisson-Gaussian model in measured intensities $\mathbf{I}^m$ is procedurally not very different to the Poisson case, and a full derivation (including practical numerical methods) is given in the first author's (MG) MPhil dissertation.[17] The model can be extended further by adding a shift-variant convolution kernel to the means of the Poisson distributions; this can represent phenomena such as adjacent pixel interference, 'cross-talk' in the X-ray detector, and potentially even refraction fringes in a propagation based phase-contrast regime.[33] We will outlay the results of this expansion in this section.

The convolved Poisson-Gaussian model considers the intensity measurements on a sensor to be subject to the usual Poisson noise on the number of photons, plus an independent additive Gaussian component which may have a separate– but known– standard deviation σ_k and mean μ_k for each measurement. We deal here with the raw (or *untransformed*) measured photon counts I_k^m and *clearfield* photon counts I_k^c expected in the absence of an attenuating sample. The vectors of all such measurements are denoted $\mathbf{I}^m$ and $\mathbf{I}^c$ respectively. We consider the measured value I_k^m to be the *realised sum* of the Poisson and Gaussian random variables for that measurement, $P_k + G_k$. It is convenient to pretend that the Poisson distribution is continuous rather than discrete, though in doing so we must introduce a normalisation factor $N(\lambda)^{-1}$. In addition, we stipulate that a model is known for the how the line attenuations $\mathbf{Ax}$ of an object determine the means of P_k. With this, we have enough information to construct the *a posteriori* probability density function of the Poisson variable. In the 'plain' convolved Poisson and Gaussian (P+G) model we say $\lambda_k = I_k = I_k^c e^{-(\mathbf{Ax})_k}$. The RHS defines the symbol I_k, while λ_k refers to the mean of the Poisson distribution for measurement k, i.e. the mean of P_k where $I_k^m = P_k + G_k$. Generally, the relationship between $\mathbf{I}$ and $\boldsymbol{\lambda}$ (vector of Poisson means) may be more complicated (e.g. if sensor cross-talk or source blurring is incorporated), so we express $\lambda_k(\mathbf{I})$ as a function of $\mathbf{I}$. The *a posteriori* probability density for the Poisson variable $p = P_k$ given measurement information, knowledge of the Gaussian means and standard deviations, and an understanding of how the Poisson mean λ_k is determined from $\mathbf{Ax}$, is given by

$$P(p|\lambda_k, \mu_k, \sigma_k) \propto f_k(p; \lambda_k) = \frac{1}{\sigma_k\sqrt{2\pi}} \frac{1}{N(\lambda_k(\mathbf{I}))} \exp\left\{ p \ln \lambda_k(\mathbf{I}) - \lambda_k(\mathbf{I}) - \ln\Gamma(p+1) - \frac{1}{2}\left(\frac{I_k^m - p - \mu_k}{\sigma_k}\right)^2 \right\} \tag{13}$$

where the normalisation factor $N(\lambda_k)^{-1}$ is defined by the integrated unnormalised probability density of the continuous Poisson distribution:

$$N(\lambda) = \int_0^\infty \exp\left\{p \ln \lambda - \lambda - \ln\Gamma(p+1)\right\} \, \mathrm{d}p \,. \tag{14}$$

We have denoted the RHS of (13) by the function $f(p; \lambda_k)$ because it, while requiring normalisation to be a probability density over p, represents in unnormalised form an important quantity which we repeatedly refer to. The likelihood of $\mathbf{Ax}$ associated with a single measurement I_k^m, and the log-likelihood in light of the combined measurements, are given respectively by

$$\mathcal{L}(\mathbf{Ax}|I_k^m) = \int_0^\infty f(p; \lambda_k) \, \mathrm{d}p \,, \qquad \ln \mathcal{L}(\mathbf{Ax}|\mathbf{I}^m) = \sum_k \ln\left[\int_0^\infty f(p; \lambda_k) \, \mathrm{d}p\right] . \tag{15}$$

It is assumed at all times that the Gaussian means σ_k and standard deviations μ_k are known for every measurement. If they are not known, then they might be measured or estimated from the statistics of the clearfield or darkfield (X-ray source turned off) radiographs.

In order to produce a quadratic expansion to log-likelihood in (15), we need to compute $\mathbf{M}$ and $\mathbf{Q}$ introduced in (9). $\mathbf{M}$ is defined as the value of $\mathbf{Ax}$ for which the derivatives of log-likelihood are zero. $\mathbf{Q}$ is the negative

Hessian matrix (matrix of negative second partial derivatives). We omit the lengthy derivation of the derivatives[17] and present them here without showing our work:

$$\frac{\partial \ln \mathcal{L}}{\partial (\mathbf{Ax})_i} = -I_i \sum_k \frac{\partial \lambda_k}{\partial I_i} \left(\frac{1}{\lambda_k} \mathrm{E}_k\left[p\right] - 1 - \frac{N'(\lambda_k)}{N(\lambda_k)} \right) \tag{16a}$$

$$\begin{aligned} \frac{\partial^2 \ln \mathcal{L}}{\partial (\mathbf{Ax})_j \partial (\mathbf{Ax})_i} &= I_i I_j \sum_k \frac{\partial^2 \lambda_k}{\partial I_j \partial I_i} \left(\frac{1}{\lambda_k} \mathrm{E}_k\left[p\right] - 1 - \frac{N'(\lambda_k)}{N(\lambda_k)} \right) \\ &+ \frac{\partial \lambda_k}{\partial I_i} \frac{\partial \lambda_k}{\partial I_j} \left(\frac{1}{\lambda_k^2} \Big(\mathrm{Var}_k\left[p\right] - \mathrm{E}_k\left[p\right] \Big) - \frac{N''(\lambda_k)}{N(\lambda_k)} + \left(\frac{N'(\lambda_k)}{N(\lambda_k)} \right)^2 \right) \\ &+ \frac{\delta_{ij}}{I_i} \frac{\partial \lambda_k}{\partial I_i} \left(\frac{1}{\lambda_k} \mathrm{E}_k\left[p\right] - 1 - \frac{N'(\lambda_k)}{N(\lambda_k)} \right) \end{aligned} \tag{16b}$$

where δ_{ij} is the Kronecker delta function on i and j. The expectation values $\mathrm{E}_k\left[\circ\right]$ and variances $\mathrm{Var}_k\left[\circ\right]$ are taken with respect to the normalised probability distributions in (13). Under the simplifying assumption that $\boldsymbol{\lambda}(\mathbf{I}) = \mathbf{I}$, we have

$$\frac{\partial \lambda_k}{\partial I_i} = \delta_{ik}, \qquad \frac{\partial^2 \lambda_k}{\partial I_j \partial I_i} = 0. \tag{17}$$

and the derivatives of $\ln \mathcal{L}$ reduce to

$$\frac{\partial \ln \mathcal{L}}{\partial (\mathbf{Ax})_i} = I_i \left(1 + \frac{N'(I_i)}{N(I_i)} \right) - \mathrm{E}_i\left[p\right] \tag{18a}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial (\mathbf{Ax})_j \partial (\mathbf{Ax})_i} = \delta_{ij} \left(\mathrm{Var}_i\left[p\right] - I_i \left(1 + \frac{N'(I_i)}{N(I_i)} \right) - I_i^2 \left(\frac{N''(I_i)}{N(I_i)} - \left(\frac{N'(I_i)}{N(I_i)} \right)^2 \right) \right) . \tag{18b}$$

In the event where the primary signal $\mathbf{I}$ is expected to be nowhere small (e.g. greater than 10 photons will suffice), then $N'(I_i), N''(I_i) \approx 0$ and these reduce further to

$$\frac{\partial \ln \mathcal{L}}{\partial (\mathbf{Ax})_i} = I_i - \mathrm{E}_i\left[p\right] \tag{19a}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial (\mathbf{Ax})_j \partial (\mathbf{Ax})_i} = \delta_{ij} \left(\mathrm{Var}_i\left[p\right] - I_i \right) . \tag{19b}$$

These three expressions for the derivatives of log-likelihood have been listed in order of decreasing complexity. The first (16) is capable of capturing phenomena such as interference between adjacent pixels (e.g. point spread associated with the detector), and if the X-rays are at least partially coherent, also some refraction effects resulting from X-ray propagation in the near Fresnel region after exiting the sample. The second (18) cannot model interference between pixels, but is otherwise fully general. It can even model negative intensity measurements which come about due to the Gaussian noise. The third (19) is the same as the second, but with an approximation applied in the case that transmitted intensities $\mathbf{I}$ are all expected to be > 10 photons.

The most difficult part of computing these derivatives is the evaluation of $\mathrm{E}_k\left[\circ\right]$ and $\mathrm{Var}_k\left[\circ\right]$. We discuss a method to do this in the next subsection. Whichever model is used ((16), (18) or (19)), the expansion point $\mathbf{M}$ must be solved as that value of $\mathbf{Ax}$ that nullifies the first derivatives of log-likelihood. This value for $\mathbf{Ax}$ should then by substituted into the expressions for the second derivatives to compute $\mathbf{Q}$ explicitly as the negative Hessian. A quadratic approximation to likelihood is then given by (9) and any square-minimisation technique may be applied to solve for an approximate ML solution $\mathbf{x}$ per (10). Various methods are available, including modified SIRT (12), modified ordered-subsets SIRT (12),[34] the Conjugate Gradient Method (CGM),[35] and undoubtedly others with which the authors are unfamiliar.[8,36]

5.1 Numerical procedure & simulation results

As the Poisson-Gaussian model poses several numerical sensitivities in practice, we suggest a practical approach which we have used with success in simulations of 2-dimensional slices of 3D volumes.

- Use a cubic spline of interpolated values for $N(\lambda)$ in the region $\lambda \in [10^{-5}, 25]$. For $\lambda \in (25, 30]$ use an interpolation between that cubic spline and the expression $\mathrm{erf}(\sqrt{\lambda})$. For $\lambda > 30$, use $\mathrm{erf}(\sqrt{\lambda})$. Ensure that the interpolation is twice-differentiable.

- Define finite bounds on the domain of $f(p; \lambda_k)$ so that numerical quadrature can be used to compute expectation values and variances as required in (16), (18), or (19). We suggest the following bounds,[17] sufficient when $\sigma_k \gtrsim \pi/10$:

$$p \in \begin{cases} p_k^{MAP} \geq 0: & \left[\max\{0, p_k^{MAP} - 4s_k\}, p_k^{MAP} + 8s_k\right] \text{ with } s_k = \left(\frac{1}{\sigma_k^2} + \psi_1(p_k^{MAP} + 1)\right)^{-1/2} \\ p_k^{MAP} < 0: & \left[0, \frac{2\alpha}{\beta}\left(1 - \sqrt{1 + 20\frac{\beta}{\alpha^2}}\right)\right] \text{ with } \alpha = \gamma + \frac{\lambda_k - \mu_k}{\sigma_k^2} + I_k^m \text{ and } \beta = \frac{\pi^2}{6} + \frac{1}{\sigma_k^2} \end{cases} \tag{20}$$

$$\text{where } p_k^{MAP} = \sigma_k^2 \,\mathrm{M}\left(\frac{1}{\sigma_k^2}\left(I_k^m - \mu_k + \frac{1}{2}\right) + \ln\frac{\lambda_k}{\sigma_k^2}\right) - \frac{1}{2} \tag{21}$$

$$\text{where } \mathrm{M}(x) = \begin{cases} 0 < x < 7e: & x - \ln(x - \ln(x - \ln(x - \ln(x)))) \\ 7e \leq x: & \ln(e^x + 1)\frac{1 + (123/40)\exp(x) + (21/10)\exp(2x)}{1 + (143/40)\exp(x) + (713/240)\exp(2x)} \end{cases}. \tag{22}$$

  where ψ_1 is the trigamma function (second derivative of $\ln\Gamma$) and γ is the Euler-Mascheroni constant.

- Use a generic root-finding or scalar minimisation algorithm (e.g. methods in *SciPy*[37]) to solve for the root $\mathbf{M} = \mathbf{Ax}$ of (16a)/(18a)/(19a). This is substantially easier for the simple P+G models (18) and (19) where each equation, indexed by k, is separable from the others and may be solved independently with a single-parameter scan of $(\mathbf{Ax})_k$. We have not attempted the more general P+G model (16) due to the difficulty in simultaneously solving the inseparable implicit equations.

- Substitute $\mathbf{Ax}$ in (16b)/(18b)/(19b) for the solved root $\mathbf{M}$ and compute the negative Hessian $\mathbf{Q}$.

This procedure has been applied to (18) to produce the reconstructions depicted in Fig. 2. The Conjugate Gradient Method[8,35] was used to minimise the quadratic forms, with a SIRT-like (12) preconditioner. The clearfield intensities $\mathbf{I}^c$ varied from 0 to 5,625,000 photons. 2000 equispaced parallel projection angles (0 to π) were taken for reconstruction on a 890×890 grid. The Gaussian means and standard deviations were uniformly equal to $\mu_k = 300$ and $\sigma_k = 90$ photons respectively. The modal non-zero attenuation through the object was ≈ 7.5 for a modal transmittance of $e^{-7.5} \approx 0.00055$. Projection data was downsampled from a simulation of a 5000×5000 sample. Unabridged information on the simulation and reconstructions may be found in MG's MPhil dissertation.[17]

6. MAXIMUM LIKELIHOOD UNDER LINEARLY TRANSFORMED POISSON NOISE

The P+G model for photon measurements, discussed in the previous section, is quite general and will approximately model many if not most transmission XCT experiments. However, we were unable to demonstrate an implemented example for the unsimplified equations (16) due to the simultaneous implicit equations which need to be solved in finding a root $\mathbf{I}$ to (16a). This means we restricted our discussion to the case where $\boldsymbol{\lambda}(\mathbf{I}) = \mathbf{I}$– i.e. where X-ray flux on the detector is not in any way blurred or mixed across measurement sites. In this section we demonstrate a model which does incorporate the mixing of measurement intensities across the sensor. We discard the Gaussian component and consider only Poisson statistics.

The probability density function of a measurement I^m given expected intensities $\mathbf{I}$ is modelled per

$$\ln P(\mathbf{I}^m|\mathbf{I}) = \ln \mathcal{L}(\mathbf{I}|\mathbf{I}^m) = \sum_k I_k^m \ln(\mathbf{CI})_k - (\mathbf{CI})_k - \ln\Gamma(I_k^m + 1) \tag{23}$$

which is identical to the Poisson model, with mean intensities $\mathbf{I}$ substituted with $\mathbf{CI}$ where $\mathbf{C}$ is an *invertible, linear operator*. Technically this is a subset of the Poisson-Gaussian model with $\boldsymbol{\lambda}(\mathbf{I}) = \mathbf{CI}$ and $\sigma_k = \mu_k = 0$.

The expansion point $\mathbf{M}$ and negative Hessian $\mathbf{Q}$ (evaluated at that expansion point) are given by

$$\mathbf{M} = -\ln \frac{\mathbf{C}^{-1}\mathbf{I}^m}{\mathbf{I}^c} \tag{24a}$$

$$\mathbf{Q} = \operatorname{diag}(\mathbf{C}^{-1}\mathbf{I}^m)\mathbf{C}^{\mathrm{T}}\operatorname{diag}\left(\frac{1}{\mathbf{I}^m}\right)\mathbf{C}\operatorname{diag}(\mathbf{C}^{-1}\mathbf{I}^m) \tag{24b}$$

where $1/\mathbf{I}^m$ is the component-wise multiplicative inverse of $\mathbf{I}^m$. Obviously this is not well defined when $I^m_k = 0$, however the limiting solution can be well-defined. For example, when $\mathbf{C}$ is a diagonal operator $\mathbf{C} = \operatorname{diag}(\mathbf{c})$ (which may model a quantum efficiency that varies across detector pixels, i.e. c_k is the quantum efficiency of pixel k), the equations reduce to

$$\mathbf{M} = -\ln \frac{\mathbf{I}^m/\mathbf{c}}{\mathbf{I}^c} \tag{25a}$$

$$\mathbf{Q} = \operatorname{diag}(\mathbf{I}^m) \tag{25b}$$

where $\mathbf{I}^m/\mathbf{c}$ is a component-wise division of vectors.

6.1 Simulation results

We have applied (24) to produce a maximum-likelihood solution on simulated data subject to a noise model where the incident X-ray flux on the detector is blurred before the Poisson statistics are realised. Also of interest is the case where the quantum efficiency of each detector pixel varies (25). We have opted to demonstrate the former example because the resulting quadratic form $\mathbf{Q}$ is non-diagonal. The reconstructions are depicted in Fig. 3. The Conjugate Gradient Method[8,35] was used to minimise the quadratic forms, with a SIRT-like (12) preconditioner.

The clearfield intensities $\mathbf{I}^c$ were uniformly equal to $6,250,000$ photons. 2000 equispaced parallel projection angles (0 to π) were taken for reconstruction on a 890×890 grid. The linear operator $\mathbf{C}$ was a linear combination of 20% the identity operation $\mathbf{Id}$ and 80% a Gaussian blur with a standard deviation of 3.2 pixels. The modal non-zero attenuation through the object was ≈ 7.75 for a modal transmittance of $e^{-7.75} \approx 0.00043$. This projection data was downsampled from a simulation of a 5000×5000 sample. Unabridged information on the simulation and reconstructions may be found in MG's MPhil dissertation.[17]

7. RESULTS & DISCUSSION

We have illustrated maximum-likelihood reconstructed tomograms based upon quadratic approximation in Fig. 1, Fig. 2 and Fig. 3. The input data to many of the reconstruction algorithms had to be normalised in order to allow quadratic-form expansions to be used; we elaborate presently. The ℓ^2 metric is not defined when there are measurement components equal to zero, $I^m_k = 0$, because this corresponds to an infinite attenuation. In order to minimise the ℓ^2 metric in these reconstructions, we either removed measurements $I^m_k = 0$ or rounded them to $I^m_k = 1$ before passing the data to the ℓ^2-minimising algorithm. We labelled those modified ℓ^2 norms as "ℓ^2 (prune)" and "ℓ^2 (round)" respectively. The reconstruction under a Gaussian model on measured intensities (Fig. 2b) does not have a real-valued quadratic-form expansion when there are measurements less than the corresponding mean, i.e. $I^m_k < \mu_k$. For this reason, we rounded such measurements up to the corresponding mean. The quadratic expansion under the Poisson model is not real-valued or well-defined when a measurement $I^m_k < 0$. For this reason, we rounded measurements I^m_k up to 0 where appropriate. The affected reconstructions are Fig. 2c, Fig. 3c and Fig. 3d. In the case of the latter two, it was not the direct measurements I^m_k that were rounded up to 0, but the components of the deblurred intensity, $(\mathbf{C}^{-1}\mathbf{I}^m)_k$.

In all tomograms, the quadratic form used as an objective function was able to converge. We simulated 2000 equispaced parallel projections with angle 0 to π, and the reconstruction domain was an inscribed circle within a 890×890 square grid. The large number of equispaced angles (compared with the size of the reconstruction domain) ensures that the projection operator $\mathbf{A}$ is injective, which in turn ensures that the quadratic-form

minimising solutions are unique. Generally, the minimising solutions to a quadratic form constitute an affine subspace,

$$X = \min_{\mathbf{x}'} \left\{ ||\mathbf{A}\mathbf{x}' - \mathbf{m}||^2_{\mathbf{Q}} \right\} = \{\mathbf{x}^* + \mathbf{x}' : \mathbf{x}' \in \ker(\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A})\}, \tag{26}$$

where $\ker(\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A})$ is the kernel of $\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}$, and $\mathbf{x}^* \in X$, e.g. $\mathbf{x}^* = \min_{\mathbf{x}' \in X} ||\mathbf{x}'||^2$. It is understood that $\min_{\mathbf{x}'}\{\dots\}$ in the equation above yields a *set* of minimising choices for $\mathbf{x}'$. When sufficient scanning angles are taken, $\mathbf{A}$ is injective and so too is $\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}$ when $\mathbf{Q}$ is positive-definite. When $\mathbf{Q}$ is positive semi-definite, (as in the Poisson noise case with measurements of 0 photons) then more scanning data may be necessary before $\mathbf{A}\mathbf{Q}\mathbf{A}$ is injective (i.e. has trivial kernel, resulting in a unique solution for $\mathbf{x}$). We simulated ample projection angles to ensure that the kernel is trivial, and for this reason we may say that *each reconstructed tomogram converged upon the unique solution minimising its quadratic form.*

Quadratic forms were minimised via the Conjugate Gradient Method (CGM) as described by *Yousef Saad*.[8] The preconditioning matrix $\mathbf{M}^{-1}$ described in that citation needs to be chosen as an approximate inverse of $\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}$. More exact choices will converge in fewer iterations (the exact choice in a single iteration) but will suffer from greater numerical instability. An open problem is the efficient construction of preconditioners which achieve a practical balance between exactness and numerical stability. We used the SIRT-like approximate inverse from (12) as our choice of preconditioner, i.e. $\mathbf{M}^{-1} := \mathrm{diag}(\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A}\underline{\mathbf{1}}_v)^{-1} \approx (\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A})^{-1}$. Each reconstruction was performed with 5000 iterations of the CGM, except that the cache of conjugate vectors was cleared every 100 iterations, effectively resetting the CGM that often. More rapid reconstruction (fewer iterations) could be achieved if a stable approximation to $(\mathbf{A}^{\mathrm{T}}\mathbf{Q}\mathbf{A})^{-1}$ is found that is more accurate than the SIRT-like choice.

Each simulated dataset was subject to a known noise model. When the correct quadratic expansion was applied to approximate the log-likelihood of $\mathbf{A}\mathbf{x}$ under the known noise model, we observed tomograms with improved feature resolvability, and reduced high-frequency noise. In each case we see that the ℓ^2-minimising solution is plagued with high-frequency 'salt-and-pepper' noise. This overwhelms the high-frequency detail in the tomogram and effectively limits spatial resolution. We found that a Gaussian blur with $\sigma = 0.8$ was just about sufficient to clear up the the high-frequency noise in Fig. 2a, but the spokes toward the centre still cannot be resolved. This is depicted in Fig. 4a. For a fair comparison, we reproduce the quadratic P+G reconstruction in Fig. 2d at the same larger size in Fig. 4b. It is evident that the quadratic P+G reconstruction was able to resolve finer details with significantly less reconstruction artefacts. There is also an appearance of 'brushed metal' in the ℓ^2-minimising solution which is absent in the P+G ML solution.

In Fig. 3 the simulated dataset was subject to a blurring operation $\mathbf{C}$ previously described. Two reconstructions deblurred the measurement data by applying $\mathbf{C}^{-1}$, and one of them maximised the (non-diagonal) quadratic expansion (24) to log-likelihood. In both cases, it was necessary to compute $\mathbf{C}^{-1}\mathbf{I}^m$. The inversion of the blurring operation $\mathbf{C}$ is a highly unstable problem, and a number of the components of the resulting vector $\mathbf{C}^{-1}\mathbf{I}^m$ were negative. This is possible even if the inverse is stable, due to the Poisson noise in measurements. As described earlier, the components of $(\mathbf{C}^{-1}\mathbf{I}^m)_k$ less than 0 were rounded up to 0, or to 1 for the ℓ^2 (round) method in Fig. 3b. Despite the instability of $\mathbf{C}^{-1}$, we observed superior results with the full quadratic form, compared with maximising Poisson likelihood on the deblurred data. The empty regions of the tomogram have less noise, and the spokes are more clearly distinguished. Inaccuracies of attenuation within the spokes, such as the long black 'gouges', were also somewhat reduced from the deblurred Poisson model. We have depicted enlarged versions of Fig. 3c and Fig. 3d in Fig. 5.

We also have a record of the measurement-averaged negative log-likelihood residuals (MANLLRs) at each iteration of each reconstruction. The MANLLR is equal to $(\ln \mathcal{L}_{\max} - \ln \mathcal{L})$ (negative LHS of (9)) divided by the number of measurements. It is computed according to the likelihood model with which the measurement data is generated. Let us spend a moment clarifying the significance of this metric. We expect MANLLRs to be strictly positive, with lower values indicating more likely (i.e. better) tomograms. The MANLLR cannot equal zero, because this would require $\mathcal{L} = \mathcal{L}_{\max}$, and $\mathcal{L}_{\max}$ represents the maximum likelihood achievable with arbitrary adjustments to $\mathbf{A}\mathbf{x}$. Tomograms are volumes $\mathbf{x}$ and the projections of those volumes $\mathbf{A}\mathbf{x}$ are constrained to range($\mathbf{A}$). Since the expansion point $\mathbf{M}$ is chosen as a global maximum of likelihood in $\mathbf{A}\mathbf{x}$-space without this constraint, it is generally not possible to choose a volume $\mathbf{x}$ achieving that maximum likelihood, therefore MANLLR > 0. There exists some ML solution(s), and so it follows that MANLLR should converge

upon an *unknown positive real value* if a reconstruction is achieving ML. (Actually, something can be said about the minimum MANLLR if the noise model is known; see appendix). We average the negative log-likelihood residual by dividing it by the number measurements (i.e. by the number of components of $\mathbf{Ax}$) because this makes the metric a more sensible choice for comparing reconstructions which may contain differing numbers of measurement data. MANLLRs were computed using exact expressions for likelihood; we did not invoke quadratic approximations when computing this metric. All MANLLRs were computed according to the known noise model with which the relevant measurement data was generated.

The MANLLR converged upon a minimum value for all those reconstructions where the quadratic expansion technique was applied as appropriate to the noise model (e.g. Fig. 1b, Fig. 2d and Fig. 3d). Most other reconstructions either did not have a converging MANLLR (i.e. oscillating or increasing), or reached some minimum value before converging upon a larger one. An exception is the Gaussian-noise-based reconstruction on P+G data (Fig. 2b) which appeared to be converging to a minimum MANLLR, though not as small a value as the P+G reconstruction (Fig. 2d). We have tabulated data regarding the MANLLRs in Fig. 6. The tabulated data shows clearly that the reconstructions based upon quadratic approximations to log-likelihood are, as expected, more effective at maximising likelihood than their counterparts whose objective functions are not modelled on the correct likelihood function. Less predictably, the tomograms with least MANLRR also appear to have the best visual quality by a significant margin.

8. CONCLUSION

Methods of tomographic reconstruction in XCT that maximise likelihood are regarded to produce higher quality tomograms than those that minimise the ℓ^2 norm of the residual projected attenuation vector. Earlier methods of likelihood maximisation were based upon the expectation-maximisation algorithm and converged too slowly for practical use. More recently, quadratic expansion of log-likelihood has been applied to allow square minimisation techniques to achieve approximate maximum-likelihood solutions. We are aware of several examples in the literature of quadratic expansion applied to Poisson noise. Here we have extended that idea to arbitrary noise models, with simulated examples for the Poisson-Gaussian and blurred Poison cases.

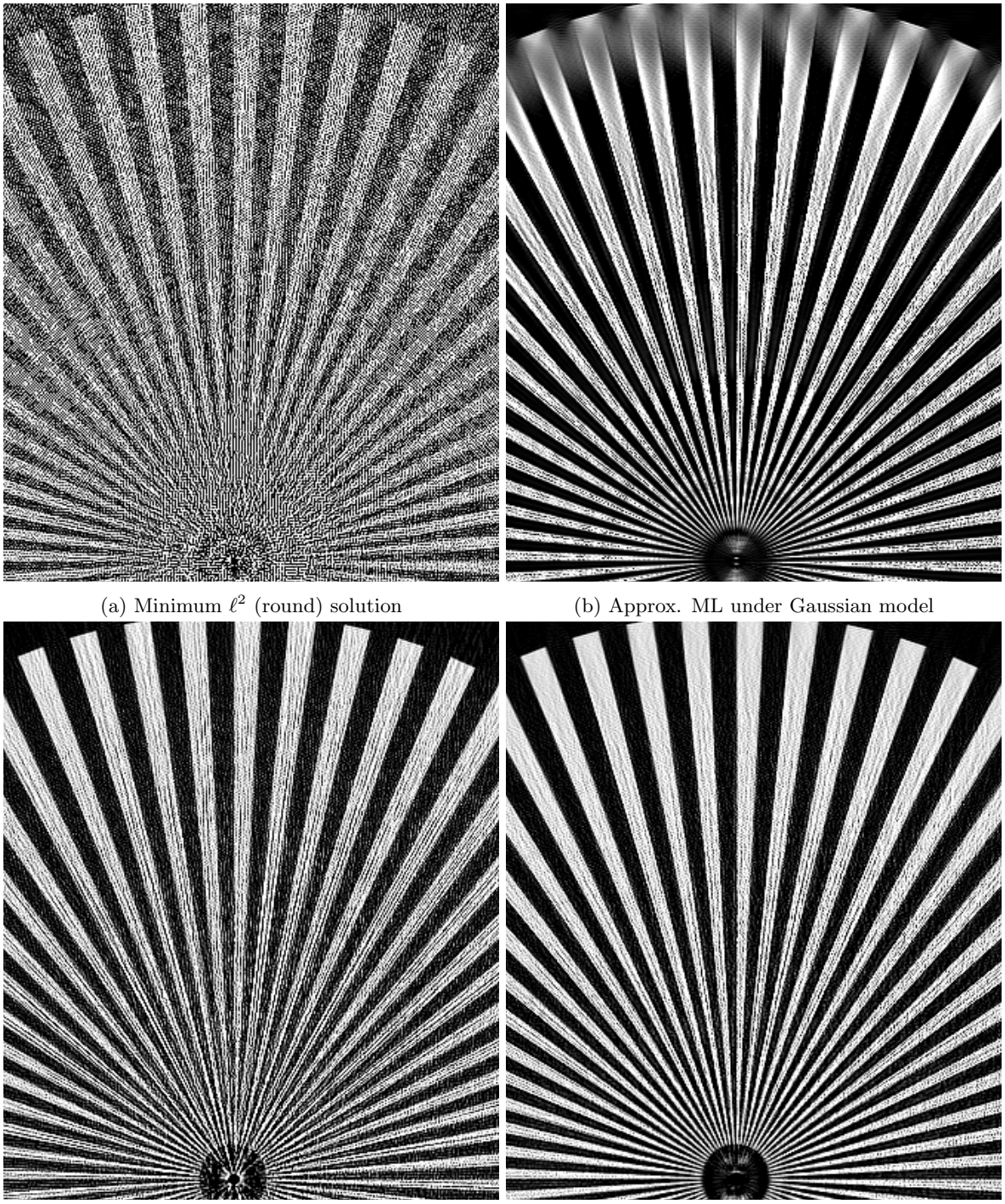
The positive results observed by others in the context of Poisson noise promise to extend to more complex noise models which swathe a great range of realistic physical experiment scenarios. Generally a choice of 'preconditioner' will be key to timely convergence, and we have described and demonstrated one plausible choice, albeit far from optimal. This preconditioner choice is a generalisation of the preconditioner implicit in the SIRT algorithm, once it has been extended to arbitrary (nonnegative) quadratic forms, i.e. arbitrary semi-inner products aside from ℓ^2.

Consequently, higher quality tomograms maximising likelihood may be produced in a practical manner for arbitrary noise models, utilising any of the full gamut of linearly-constrained quadratic optimisation algorithms in the literature. Based on our preliminary simulation results, we expect these methods to substantially improve reconstruction quality over ℓ^2-minimisation methods when measurement noise is substantial, allowing accurate reconstruction under more 'hostile' scanning conditions.

Future work involves applying this method to experimental data for an imaging system whose noise model is characterised and known to fit the blurred Poisson-Gaussian model, or another model altogether. Although very promising, it remains to be observed in practice whether these methods offer compelling tomogram quality in the presence of complex and otherwise oppressive measurement noise. It also remains a question: how can we construct better stable preconditioners under arbitrary noise models and scanning geometries?

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council (ARC) and Thermo Fisher Scientific through the linkage project LP150101040 as well as by the ARC industrial transformation training centre IC18010008 with financial support from the Defence Science Technology Group (DSTG) project number MyIP 9227. MGG received a Research Training Program Scholarship. We also acknowledge support from ANU and Australia's National Computational Infrastructure (NCI) by providing access to supercomputer resources through merit allocation schemes.



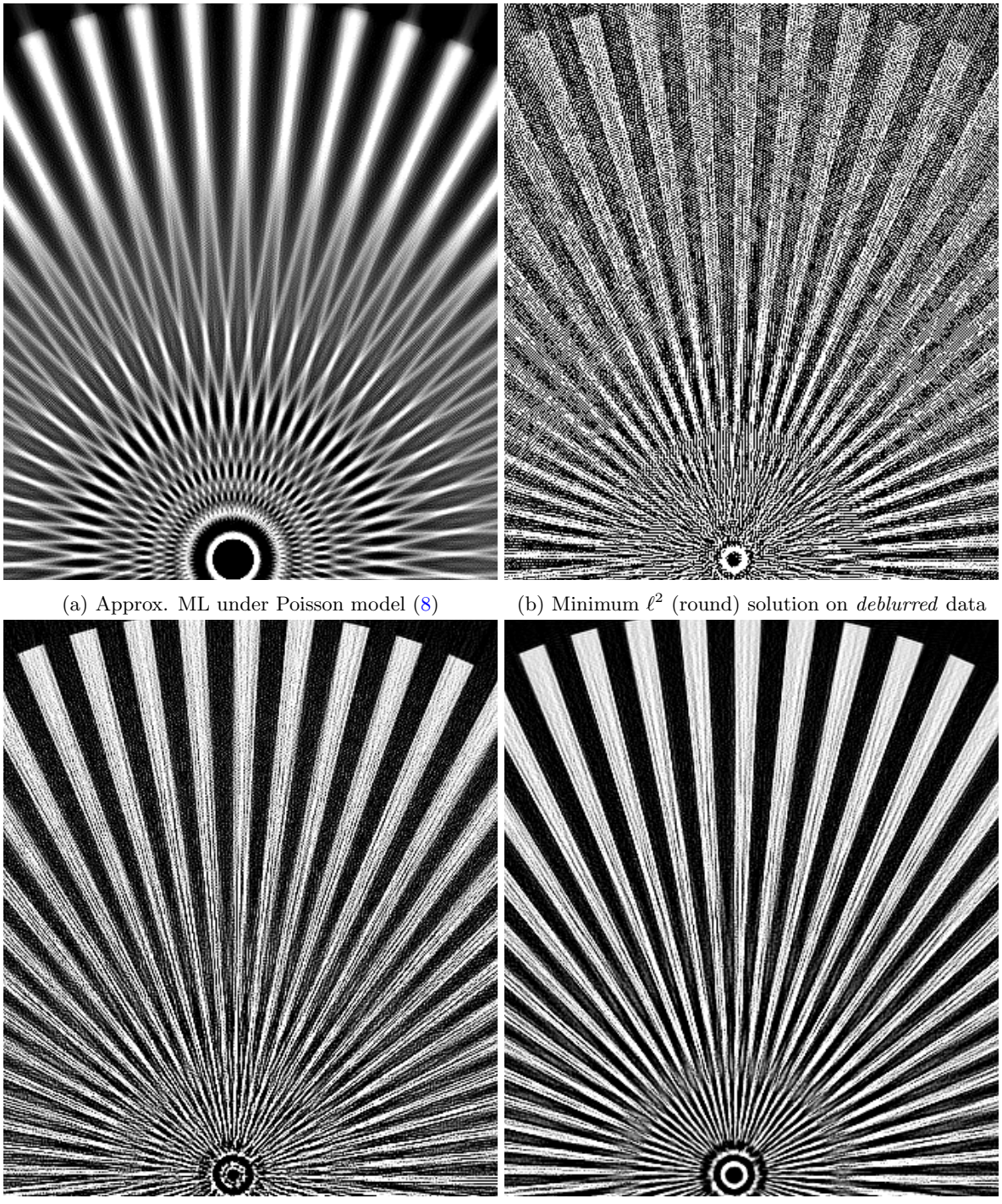
(a) Minimum ℓ^2 (round) solution

(b) Approx. ML under Gaussian model

(c) Approx. ML under Poisson model (8)

(d) Approx. ML in Poisson-Gaussian model (15), (18)

Figure 2: Example of four reconstructions performed on identical simulated data. The data was generated with convolved Poisson+Gaussian noise. The Gaussian noise was uniform, with $\sigma_k = 90$ and $\mu_k = 300$ (measured in photons for each detection site). Each reconstruction solves $\mathbf{m} = \mathbf{Ax}$, but with a different regularisation scheme. Each minimises a quadratic form indicated in the caption. **(b):** Models the Gaussian component of noise but ignores the Poisson component. **(c):** Models the Poisson component of noise but ignores the Gaussian component.



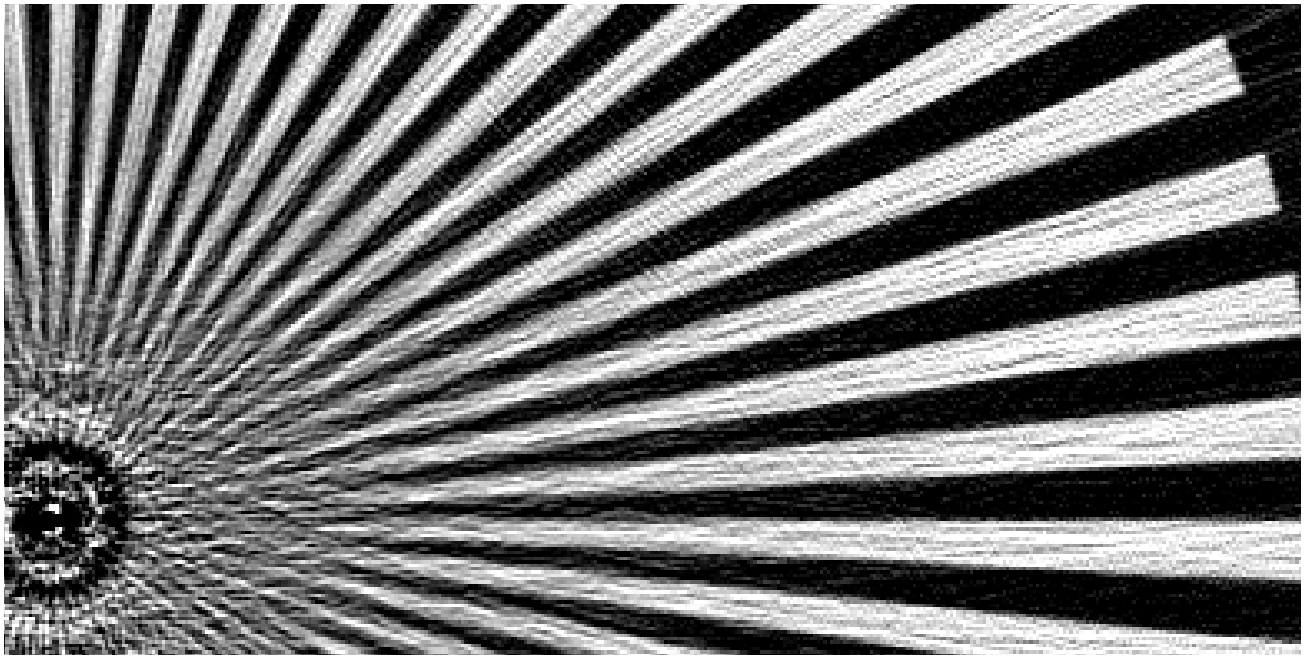
(a) Approx. ML under Poisson model (8)

(b) Minimum ℓ^2 (round) solution on *deblurred* data

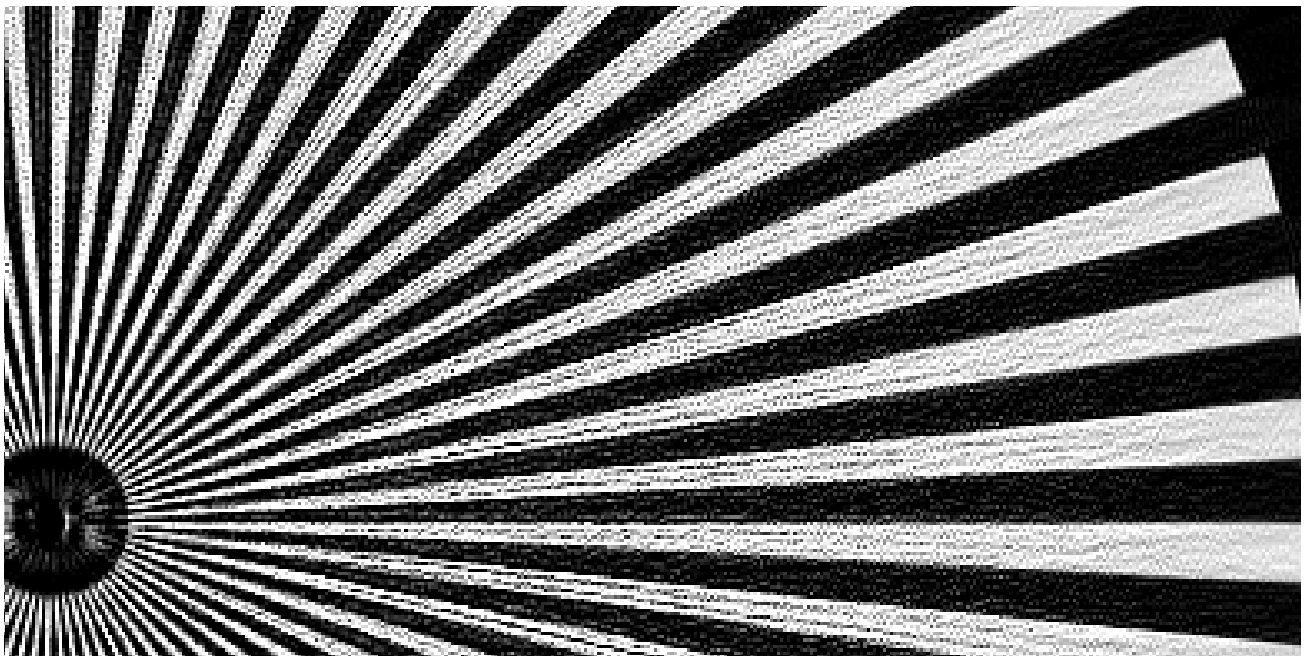
(c) Approx. ML in Poisson model (8) on *deblurred* data

(d) Approx. ML under blurred-Poisson model (24)

Figure 3: Example of four quadratic-form-optimising reconstructions performed on identical simulated data. The data was generated with blurred Poisson statistics per (23). '*Deblurred* data' refers to intensity measurements which have been transformed with a best-attempt inverse, $\mathbf{I}^m \mapsto \mathbf{C}^{-1}\mathbf{I}^m$. **(a):** Ignorant of blurring. Demonstrates the severity of the blurring.

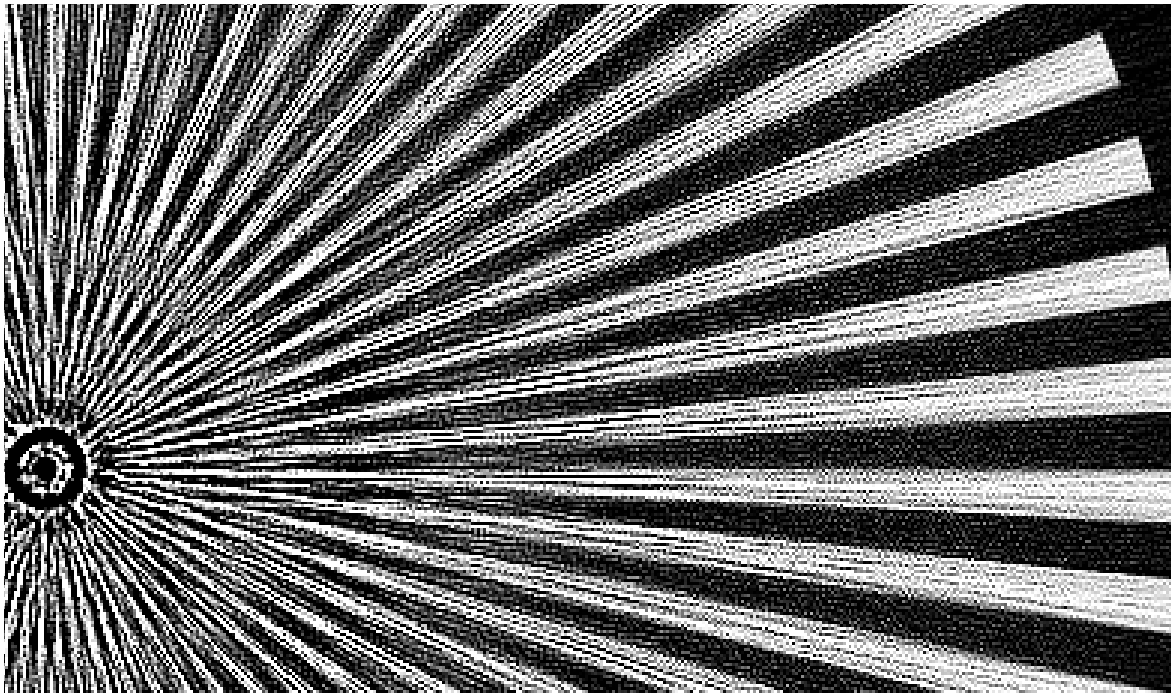


(a) Depiction of the tomogram from Fig. 2a with a Gaussian blur applied with $\sigma = 0.8$. This is roughly the minimum blur required to eliminate the high frequency noise in both the non-attenuating regions and in the attenuating spokes. This image gives an idea of the information which can be recovered from that tomogram, if the higher spatial frequencies are filtered out. We see that toward the centre of the tomogram, the spokes are indistinguishable– this happens once they are ≈ 5 pixels wide and the same distance apart.

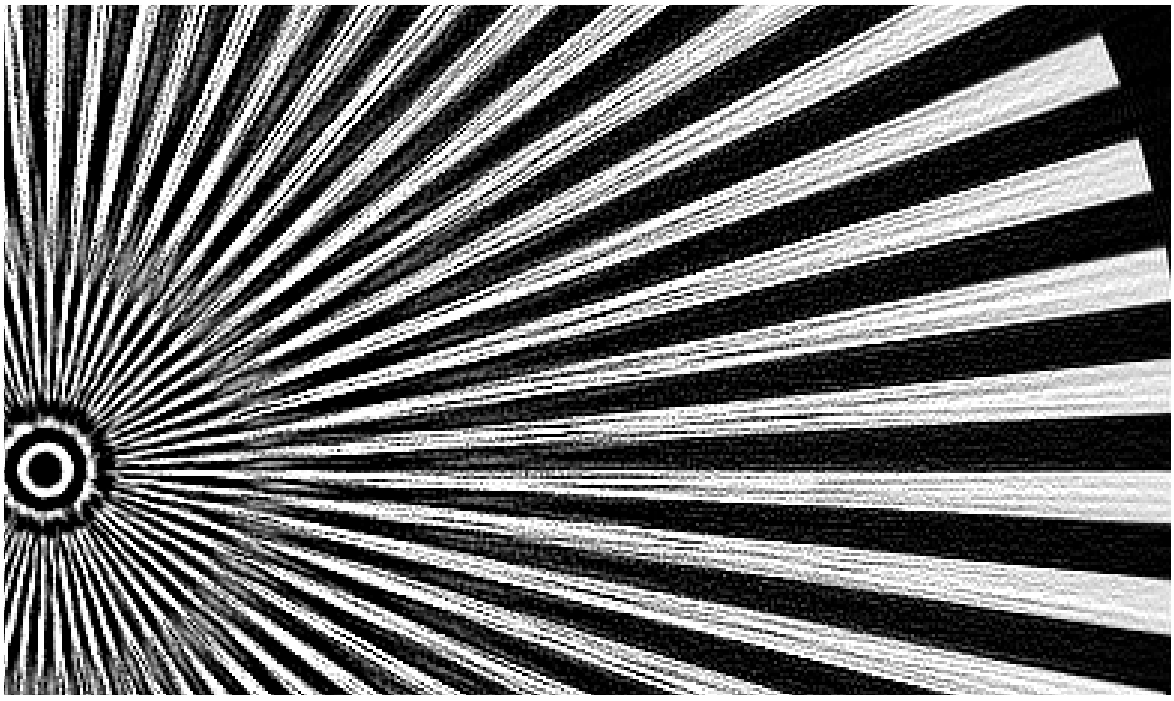


(b) A reproduction of Fig. 2d at a larger size for comparison with the filtered (i.e. blurred) ℓ^2 (round) reconstruction in Fig. 4a. This figure has not been filtered.

Figure 4: Comparison of two tomograms reconstructed from simulated data subject to convolved Poisson-Gaussian noise (15). One is a filtered ℓ^2 (round) minimising tomogram (a) and the other is an approximate Poisson+Gaussian Maximum Likelihood (P+G ML) tomogram (b).



(a) An enlarged reproduction of Fig. 3c for comparison with Fig. 5b.



(b) An enlarged reproduction of Fig. 3d for comparison with Fig. 5a.

Figure 5: Comparison of two tomograms reconstructed from simulated data subject to blurred-Poisson noise (23). One is an approximate Poisson ML tomogram (a) acting on deblurred data, and the other is a true approximate ML solution to the blurred-Poisson model (b).

Iterations to achieve likelihood residual			
2×MANLLR	Pois.	ℓ^2 (round)	ℓ^2 (prune)
50	11	27	25
40	13	31	27
30	16	40	31
25	19	62	34
20	22	—	39
15	27	—	47
10	36	—	59
8.0	41	—	77
3.0	75	—	—
2.0	101	—	—
1.05	781	—	—
1.01	2177	—	—
1.00008	4614	—	—

(a) Iteration counts by MANLLR thresholds for the Poisson noise model. Final tomograms after 5000 iterations depicted in Fig. 1.

Iterations to achieve likelihood residual				
2×MANLLR	P.G.	Pois.	Gaus.	ℓ^2 (round)
20	72	48	524	—
10	118	125	1762	—
8.5	132	951	2293	—
5.5	172	—	4792	—
2.5	239	—	—	—
2.0	289	—	—	—

(b) Iteration counts by MANLLR thresholds for the Poisson-Gaussian noise model. Final tomograms after 5000 iterations depicted in Fig. 2.

Iterations to achieve likelihood residual				
2×MANLLR	B. Pois.	Pois.	Debl. Pois.	Debl. ℓ^2 (round)
100	36	—	32	—
50	50	—	43	—
20	78	—	64	—
10	129	—	97	—
4.7	466	—	804	—
4.0	983	—	—	—
3.5	2403	—	—	—
3.4	3184	—	—	—
3.3	4609	—	—	—

(c) Iteration counts by MANLLR thresholds for the blurred Poisson noise model. Final tomogarms after 5000 iterations depicted in Fig. 3.

Figure 6: Tabulation of measurement-averaged negative log-likelihood residuals (MANLLRs; lower is better; LLR is LHS of (9)) of the quadratic reconstructions performed on data subject to Poisson noise (a), Poisson-Gaussian noise (b) and blurred Poisson noise (c). The integer numbers in the table are the number of iterations of the respective algorithm before that value for the MANLLR was subceeded. A dash indicates that the reconstruction did not achieve the listed MANLLR. More data available in MG's MPhil dissertation.[17] **(a):** Notice that the MANLLR is converging to 1/2 for the Poisson ML reconstruction. This should be the case for any ML solution under Poisson noise (except in the most extremely photon-starved cases) and can be used as a check that a tomogram is approaching maximum likelihood when the data is subject to perfect Poisson noise; see the appendix.

APPENDIX

We noted in the caption of Fig. 6 that the reconstruction algorithm which maximised likelihood under the Poisson noise model was converging to a MANLLR value of nearly exactly 1/2. The same will always be said of a Poisson ML reconstruction as long as (1) the data is truly subject to pure Poisson noise, and (2) the Poisson ML solution predicts that only a negligible fraction of measurement sites will receive a mean photon count λ less than ≈ 10. If a Poisson ML algorithm does not converge to a MANLRR value of 1/2 then we can surmise that the measurement data is not perfectly Poisson distributed, or that the algorithm is faulty. We justify this below.

Suppose we have a ground truth for the volume $\mathbf{x}$. We then have a collection of mean photon-counts on the measurements sites as $\mathbf{I} = \mathbf{I}^c \exp\{-\mathbf{A}\mathbf{x}\}$. Further let us assume that the measured photon counts $\mathbf{I}^m$ are indeed distributed according to the discrete Poisson distribution (8). The NLLR of $\mathbf{x}$ associated with a single measurement I_k^m is a randomly distributed variable, dependent on the realised photon count I_k^m.

$$\begin{aligned} \mathrm{NLLR}_k &= \ln \mathcal{L}_{\max} - \ln \mathcal{L}(\mathbf{I}|I_k^m) && (27)\\ &= \ln \mathcal{L}(\mathbf{I}^m|I_k^m) - \ln \mathcal{L}(\mathbf{I}|I_k^m) && (28)\\ &= (I_k - I_k^m) + I_k^m \ln(I_k^m/I_k)\,, && (29) \end{aligned}$$

When $I_k^m = 0$, we take the limit $I_k^m \to 0^+$, and the result is $\mathrm{NLLR}_k = I_k$. As a random variable, the mean and variance of NLLR_k depend on I_k and both have been calculated numerically and plotted in Fig. 7. When $I_k \gtrsim 10$, $\mathrm{NLLR}_k \approx 1/2$. The variance converges similarly to 1/2.

Let us suppose that the Poisson ML tomogram predicts only a small proportion of measurement sites to have a mean photon count of less than 10 (say $< 1\%$). In many experiments this will be a reasonable assumption. Further suppose that we have a moderate number of measurements ($N \gtrsim 1000$). The central limit theorem determines that, if the measurement data is Poisson-distributed, the mean NLLR across a sampling of measurements, (i.e. the MANLLR) will be randomly distributed with a mean of 1/2 and a variance of $1/(2N)$.

We have shown that a Poisson ML tomogram reconstructed from Poisson-distributed data will have a MANLLR of nearly exactly 1/2, assuming that only a small fraction of measurement sites expect less than 10 photons.

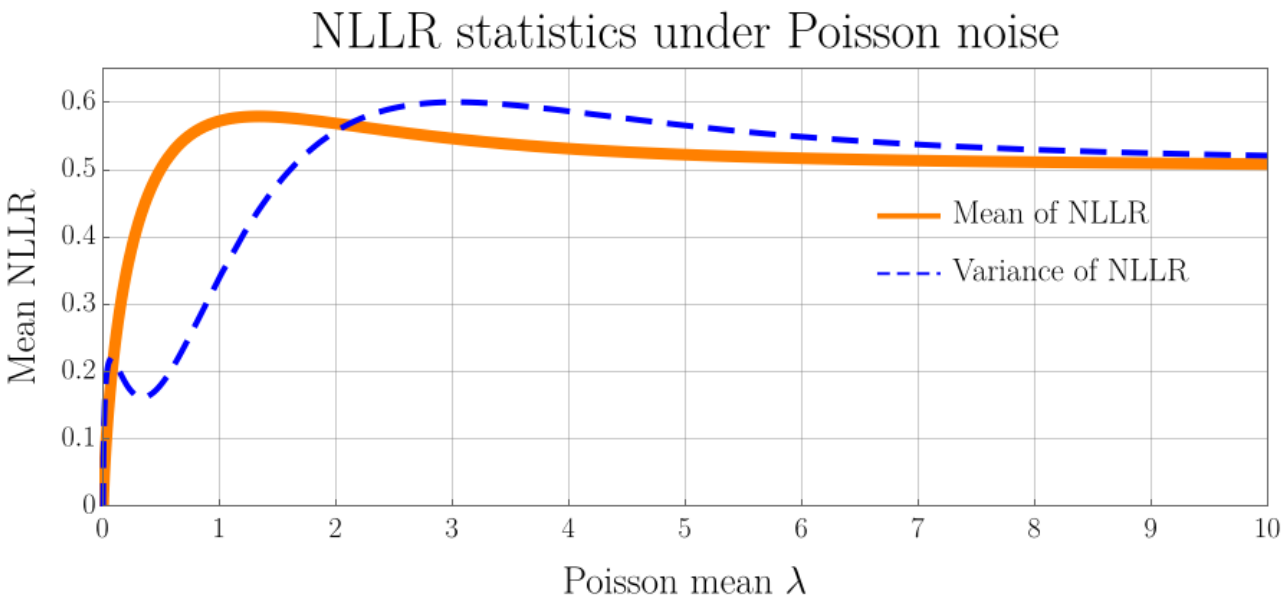


Figure 7: Plot of the mean and variance of the negative log-likelihood residual (NLLR) of a single measurement under the Poisson noise model. We generate measurements I^m from a Poisson mean λ, and compute the likelihood of that same Poisson mean under the generated measurements, i.e. $\mathrm{NLLR} = \mathcal{L}(\lambda|I^m)$. This is a randomly distributed variable whose mean and variance we have plotted.

REFERENCES

[1] Hounsfield, G. N., "Computed medical imaging," *Medical Physics* **7**, 283–290 (July 1980).

[2] Cormack, A. M., "Representation of a function by its line integrals, with some radiological applications," *Journal of Applied Physics* **34**, 2722–2727 (September 1963).

[3] Cormack, A. M., "Representation of a function by its line integrals, with some radiological applications. ii," *Journal of Applied Physics* **35**, 2908–2913 (October 1964).

[4] De Man, B., Nuyts, J., Dupont, P., Marchal, G., and Suetens, P., "An iterative maximum-likelihood polychromatic algorithm for ct," *IEEE transactions on medical imaging* **20**(10), 999–1008 (2001).

[5] Van Gompel, G., Van Slambrouck, K., Defrise, M., Batenburg, K. J., de Mey, J., Sijbers, J., and Nuyts, J., "Iterative correction of beam hardening artifacts in CT: Iterative correction of beam hardening artifacts in CT," *Medical Physics* **38**, S36–S49 (July 2011).

[6] Yang, Q., Fullagar, W., Myers, G., Latham, S., Varslot, T., Sheppard, A., and Kingston, A., "X-ray attenuation models to account for beam hardening in computed tomography," *Applied Optics* **59**(29), 9126–9136 (2020).

[7] Rührnschopf, E.-P. and Klingenbeck, K., "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. par 1: Scatter compensation approaches," *Medical Physics* **38**(7), 4296–4311 (2011).

[8] Saad, Y., [*Iterative methods for sparse linear systems*], Society for Industrial and Applied Mathematics, 2 ed. (2003).

[9] Golub, H. G. and Van Loan, C. F., [*Matrix Computations*], Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins Universtiy Press, 4 ed. (February 2004/2013).

[10] Gordon, R., Bender, R., and Herman, G. T., "Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography," *Journal of Theoretical Biology* **29**, 471–481 (December 1970).

[11] Gilbert, P., "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology* **36**, 105–117 (July 1972).

[12] Andersen, A. H. and Kak, A. C., "Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm," *Ultrason Imaging* **6**, 81–94 (January 1984).

[13] Kak, A. C., Slaney, M., and Wang, G., [*Principles of computerized tomographic imaging*], Wiley Online Library (2002).

[14] Norton, S. J., "Iterative reconstruction algorithms: convergence as a function of spatial frequency," *JOSA A* **2**(1), 6–13 (1985).

[15] Boschi, L. and Dziewonski, A. M., "High-and low-resolution images of the earth's mantle: Implications of different approaches to tomographic modeling," *Journal of Geophysical Research: Solid Earth* **104**(B11), 25567–25594 (1999).

[16] Ben-Israel, A. and Greville, T. N., [*Generalized inverses: theory and applications*], vol. 15, Springer Science & Business Media (2003).

[17] Grewar, M. G., *Quadratic Form Minimisation in X-ray Computed Tomography*, mphil thesis, Australian National University, Research School of Physics, Department of Applied Mathematics (2021). Pending publication as an ANU open access thesis.

[18] Lakshminarayanan, A. V. and Lent, A., "The simultaneous iterative reconstruction technique as a least-squares method," *Application of Optical Instrumentation in Medicine V* **96**, 108–116 (December 1976).

[19] Shepp, L. A. and Vardi, Y., "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging* **1**, 113–122 (Oct 1982).

[20] Lange, K., Carson, R., et al., "Em reconstruction algorithms for emission and transmission tomography," *J Comput Assist Tomogr* **8**(2), 306–16 (1984).

[21] Vardi, Y., Shepp, L. A., and Kaufman, L., "A statistical model for positron emission tomography," *Journal of the American Statistical Association* **80**(389), 8–20 (1985).

[22] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22 (1977).

[23] Goitein, M., "Three-dimensional density reconstruction from a series of two-dimensional projections," *Nuclear Instruments and Methods* **101**(3), 509 – 518 (1972).

[24] Nuyts, J., Man, B. D., Fessler, J. A., Zbijewski, W., and Beekman, F. J., "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Physics in Medicine and Biology* **58**(12), R63 (2013).

[25] Kariya, T. and Kurata, H., [*Generalized least squares*], John Wiley & Sons (2004).

[26] Yalavarthy, P. K., Pogue, B. W., Dehghani, H., and Paulsen, K. D., "Weight-matrix structured regularization provides optimal generalized least-squares estimate in diffuse optical tomography," *Med. Phys.* **34**, 2085–2098 (May 2007).

[27] Dillon, O., "Probabilistic approximations of matrix decompositions for inverse problems," (2019).

[28] Sauer, K. and Bouman, C., "A local update strategy for iterative reconstruction from projections," *IEEE Transactions on Signal Processing* **41**, 534–548 (February 1993).

[29] Fessler, J. A., "Penalized weighted least-squares image reconstruction," *IEEE Transactions on Medical Imaging* **13**, 290–300 (June 1994).

[30] Fessler, J. A., "Hybrid poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Transactions on Image Processing* **4**, 1439–1450 (October 1995).

[31] Bouman, C. A. and Sauer, K., "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing* **5**, 480–492 (March 1996).

[32] Thibault, J.-B., Sauer, K. D., Bouman, C. A., and Hsieh, J., "A three-dimensional statistical approach to improved image quality for multslice helical ct," *Medical Physics* **34**(11), 4526–4544 (2007).

[33] Myers, G. R., Mayo, S. C., Gureyev, T. E., Paganin, D. M., and Wilkins, S. W., "Polychromatic cone-beam phase-contrast tomography," *Phys. Rev. A* **76**, 045804 (Oct 2007).

[34] Hudson, H. M. and Larkin, R. S., "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging* **13**, 601–609 (Dec 1994).

[35] Scales, J. A., "Tomographic inversion via the conjugate gradient method," *GEOPHYSICS* **52**, 179–185 (February 1987).

[36] Necoara, I., Nesterov, Y., and Glineur, F., "A random coordinate descent method on large optimization problems with linear constraints," *Technical Report, University Politehnica Bucharest* (2011).

[37] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods* **17**, 261–272 (2020).