Maximum-likelihood refinement for coherent diffractive imaging

P Thibault$^{1,3}$ and M Guizar-Sicairos$^{2}$

$^1$ Physics Department, Technische Universität München, 85748 Garching, Germany
$^2$ Paul Scherrer Institut, 5232 Villigen PSI, Switzerland
E-mail: pierre.thibault@tum.de

New Journal of Physics 14 (2012) 063004 (20pp)
Received 27 March 2012
Published 7 June 2012
Online at http://www.njp.org/
doi:10.1088/1367-2630/14/6/063004

Abstract. We introduce the application of maximum-likelihood (ML) principles to the image reconstruction problem in coherent diffractive imaging. We describe an implementation of the optimization procedure for ptychography, using conjugate gradients and including preconditioning strategies, regularization and typical modifications of the statistical noise model. The optimization principle is compared to a difference map reconstruction algorithm. With simulated data important improvements are observed, as measured by a strong increase in the signal-to-noise ratio. Significant gains in resolution and sensitivity are also demonstrated in the ML refinement of a reconstruction from experimental x-ray data. The immediate consequence of our results is the possible reduction of exposure, or dose, by up to an order of magnitude for a reconstruction quality similar to iterative algorithms currently in use.

$^3$ Author to whom any correspondence should be addressed.
1. Introduction

The last decade has seen a dramatic expansion of synchrotron-based x-ray microscopy techniques. Diffraction microscopy, holography and ptychography, to name a few, all have in common their exploitation of the transverse coherence properties of third-generation synchrotron x-ray beams to encode image information and overcome the technical limitations caused by imperfect x-ray optics (for reviews, see, e.g., [1, 2]). The wide variety of coherent diffractive imaging (CDI) techniques bears witness to both the vitality and the infancy of the field.

All CDI techniques ultimately rely on the collection of intensity distributions after a macroscopic propagation distance. In the best-case scenario, the measured intensities are given by photon counts, and can be interpreted as the squared amplitude, plus noise, of the wavefield that reaches the detector after interacting with the sample upstream. Reconstructing the image or density of the sample is then equivalent to recovering the phases that are lost in the measurement. Whereas holography [3–6] encodes the phase in the measurement, other CDI techniques face the more difficult ‘phase problem’. By far the most successful reconstruction strategy so far is to use iterative algorithms, such as Fienup’s hybrid input–output algorithm [7] or more recent alternatives such as Elser’s difference map [8] and Luke’s relaxed averaged alternating reflections technique [9]. These reconstruction algorithms can be expressed in a general formalism based on projections onto constraint sets. While convenient and powerful, the notion of set membership implies an either/or constraint that can be difficult to adapt to experimental knowledge that is probabilistic in nature because of noise and other measurement uncertainties. Typical iterative reconstruction algorithms never truly converge in the presence of noise, thus making ambiguous the definition of a unique solution. One can use corrected estimators [10] or relax the constraints [11, 12] to improve convergence properties but in the presence of noise the possibility for non-unique solutions remains. A common strategy to increase confidence in
a reconstruction is to average multiple solution attempts \[11, 13\]. Unless the formulation of the algorithm is strongly biased, averaging sufficiently many reconstructions ensures that random fluctuations are averaged out, with the additional consequence of a resolution loss.

Viewed from a broader perspective, reconstruction from noisy data is a problem met in many corners of experimental sciences, and multiple standard tools are available to solve it. Likelihood maximization may be one of the most general and powerful principles available to tackle such problems. The absence of maximum-likelihood (ML) concepts in most of the recent CDI works stems from the intrinsic difficulty of the reconstruction process. In general, cost-function optimization formulations of the phase problem lead to strong non-convexity and ensuing multiple local minima. In essence, CDI is plagued by the same difficulties as structure determination from crystallographic data. Yet, even though ML typically provides little help for \textit{ab initio} phasing from diffraction data, it is nowadays widely used by the crystallographic community for the refinement of the atomic parameters once a model has been built.

In this paper, we propose the use of ML principles as a refinement step for CDI after a reconstruction that is close to a solution that has been obtained with projection-based techniques. While the general concepts are meant to apply to CDI at large, we concentrate most of the developments and demonstrations to ptychography. The reasons for this choice are twofold. Firstly, the approach is known to give good results for ptychography. It was shown in \[14\] that ptychography can be stated as a phase retrieval problem with transverse translation diversity and that gradient-based minimization approaches are surprisingly well behaved for this case, even for essentially one-dimensional (1D) problems \[15, 16\]. This behaviour was also observed in other reconstruction problems where sufficient diversity of measurement is attainable, for instance when repeated measurements can be taken at different known configurations of an optical setup \[17–19\]. Secondly, the theoretical developments for ptychography are more complex than most other techniques, making a detailed discussion of these points more relevant. In section \ref{sec:simplifications}, we discuss the simplifications needed on the presented theory in order for this refinement step to be applied to CDI with a support constraint, either for the case of planar or curved wavefront illumination \[20, 21\]. The reconstruction procedure developed for ptychography also encompasses other imaging schemes with little modification, such as techniques based on phase modulating optics \[22, 23\].

In the following, we present statistical models appropriate for most experimental settings and provide relevant implementation details, including preconditioning strategies, regularization and examples where the model is either further restricted or relaxed in accordance with experimental conditions. The theoretical developments are then demonstrated and compared with difference map reconstructions for both simulated and experimental diffraction data. Our findings indicate that the best strategy for ptychographic reconstructions is to obtain a first result estimate using the difference map, thereby utilizing its ability to search efficiently the solution space, and follow it by an ML refinement to account for the statistical nature of the data on the final solution.

\section*{2. Fundamentals}

Image reconstruction through ML estimation necessitates a model of the propagation of x-rays and their interaction with the sample, and a probabilistic description of the data acquisition. The ptychography model has been described in previous publications, but is restated here for completeness. The illumination function \( P_r \) (the ‘probe’) is a 2D complex-valued quantity that
represents the amplitude and phase of the illumination at the plane of the sample. The interaction of the sample with the incident wavefield is modelled by a multiplicative 2D complex-valued transmission function $O_r$ (the ‘object’). At the exit of the sample, the wavefield is therefore

$$\psi_{jr} = P_{r-r_j} O_r,$$

where $r = (x, y)$ is a coordinate in the plane perpendicular to the propagation direction and $j$ is the label of the probe position relative to the object. In the absence of noise and other experimental errors, the intensity at the position $q$ in the far-field plane is

$$I_{jq} = |\tilde{\psi}_{jq}|^2 = |\mathcal{F}[P_{r-r_j} O_r]|^2,$$

where $\tilde{\psi}_q = \mathcal{F}[\psi_q]$ represents a Fourier transformation operation. For all practical applications, $r$ and $q$ are discrete coordinates on a finite rectangular grid, and the Fourier transform is replaced with a discrete Fourier transform. The reconstruction problem is to find the pair $(P_r, O_r)$ that satisfies the system of equation (2). In the best scenario, the only source of noise is the counting statistics, giving a Poisson distribution. The probability of measuring $n_{jq}$ photons given $P_r$ and $O_r$ is

$$p(n_{jq}|P_r, O_r) = \frac{(I_{jq})^{n_{jq}}}{n_{jq}!} e^{-I_{jq}},$$

where $I_{jq}$ is a function of $(P_r, O_r)$ as defined above. The negative log-likelihood function associated with this probability distribution is a function of the parameters, given the observed data $n_{jq}$:

$$\mathcal{L} = -\log \prod_j \prod_q p(n_{jq}|P_r, O_r)$$

$$= -\sum_j \sum_q w_{jq} \left[ n_{jq} \log(I_{jq}) - I_{jq} - \log(n_{jq}!) \right].$$

The sum over $q$ must include only valid measurements. In practice, it is convenient to let the sum cover the complete plane and embed instead the domain of integration as a mask $w_{jq}$ that is equal to 1 for valid pixels and to 0 for unmeasured regions, as well as bad or dead pixels. The last term containing the factorial is a constant offset that can be either ignored or computed prior to the reconstruction. The maximization of the likelihood, equivalent to the minimization of this function, requires the calculation of its gradient for any $(P_r, O_r)$. Because the intensity is a function of the product (1), the derivatives\(^4\) can be written as

$$g_{O_r} = \frac{\partial \mathcal{L}}{\partial O_r} = \sum_j P_{r-r_j} x_{jr}^*,$$

$$g_{P_r} = \frac{\partial \mathcal{L}}{\partial P_r} = \sum_j O_{r+r_j} x_{jr}^*,$$

\(^4\) We use Wirtinger derivatives\(^24\) instead of computing the gradient with respect to real and imaginary parts of the complex-valued quantities $P_r$ and $O_r$.\n
where $\ast$ denotes the complex conjugation and the auxiliary function $\chi_{jr}$ is most simply expressed in Fourier space:

$$
\tilde{\chi}_{jq} = \frac{\partial L}{\partial I_{jq}} \tilde{\psi}_{jq} = w_{jq} \left( 1 - \frac{n_{jq}}{I_{jq}} \right) \tilde{\psi}_{jq}.
$$

(7)

In many situations, additional sources of error are present, and the appropriate likelihood function is instead a Gaussian distribution, where the spatially dependent measurement uncertainties are $\sigma_{jq}$. The corresponding negative log-likelihood takes the familiar form of a weighted sum of squares:

$$
L = \sum_j \sum_q \frac{w_{jq}}{2\sigma_{jq}^2} (I_{jq} - n_{jq})^2.
$$

(8)

Computing the gradient of $L$ also leads to terms in the form (5) and (6), this time with

$$
\tilde{\chi}_{jq} = w_{jq} \frac{I_{jq} - n_{jq}}{\sigma_{jq}^2} \tilde{\psi}_{jq}.
$$

(9)

In the particular case where only the Poisson variance contributes to the measurement uncertainty, one may substitute $\sigma_{jq}^2 = n_{jq}$ to obtain a form similar to (7):

$$
\tilde{\chi}_{jq} = w_{jq} \left( \frac{I_{jq}}{n_{jq}} - 1 \right) \tilde{\psi}_{jq}.
$$

(10)

This approximation is, however, valid only for moderately large photon counts, where the Poisson distribution approaches a Gaussian. To use the Gaussian model even for low counts (say, 5 and below), one may resort to a better estimation of $\sigma_{jq}^2$ using a simple prior model (see appendix A).

Yet another possible path to approximating the Poisson likelihood function is to substitute $F_{jq} = \sqrt{I_{jq}}$ into (4) and expanding around $F_{jq} = \sqrt{n_{jq}}$, leading to, up to an additive constant,

$$
L \approx \sum_j \sum_q 2w_{jq} (F_{jq} - \sqrt{n_{jq}})^2.
$$

(11)

This likelihood function is seen to correspond to a Gaussian distribution of $F_{jq} = \sqrt{I_{jq}}$ centred on $\sqrt{n_{jq}}$ and with a variance of $1/4$, independent of $n_{jq}$. It fortunately has the property of being identical to the cost function found throughout the phase retrieval literature, where it is formulated as a Euclidean metric in the space spanned by the wavefields $\psi_{jr}$. Reconstruction techniques based on such Euclidean metrics are therefore especially well suited to measurements where Poisson statistics are dominant, and it is plausible that the noise robustness of these techniques can be partly explained by this connection. This metric has already been used and found to behave very well for ptychographic reconstruction in [14]. The gradient of (11) can again be written in the form (5) and (6) with

$$
\tilde{\chi}_{jq} = 2w_{jq} \left( 1 - \frac{\sqrt{n_{jq}}}{|\psi_{jq}|} \right) \tilde{\psi}_{jq}.
$$

(12)

The case when a single diffraction pattern appears in (11) corresponds to conventional diffractive imaging. Since the log-likelihood coincides with the metric of the search space, the
gradient has a clear geometric interpretation: its negative points in the direction of the projection of $\psi_{jr}$ onto the set of elements satisfying $|\psi_{jq}| = F_{jq}$.

Ptychographic reconstruction—or refinement—including measurement statistics amounts to minimizing equation (4), (8) or (11). In the following section, we give a description of the important implementation details of the algorithms designed for this task.

3. Implementation details

The optimization problem described above can be tackled with a variety of numerical methods. Typical reconstruction problems lead to the search for a solution in high-dimensional spaces that does not lend itself easily to second-order techniques involving the computation of a Hessian matrix. Here we describe briefly the implementation of a nonlinear conjugate gradient (CG) minimization of the negative log-likelihood functions described above.

Conjugate gradient minimization is realized iteratively. At the $n$th iteration step, a new gradient is computed according to equations (5) and (6),

$$g^{(n)} = \left( g_{P}^{(n)} , g_{O}^{(n)} \right) = \left( \sum_{j} O_{r_{r}r}^{(n)} \chi_{jr}^{(n)u} , \sum_{j} P_{r_{r}r}^{(n)} \chi_{jr}^{(n)u} \right). \quad (13)$$

The essential property of CG methods is that the new search direction $\Delta^{(n)} = (\Delta P_{r}^{(n)}, \Delta O_{r}^{(n)})$ is in general not parallel to $-g^{(n)}$, but rather in a conjugate direction induced by the local curvature of the function to optimize. One popular way to update $\Delta^{(n)}$ is

$$\Delta^{(n)} = -g^{(n)} + \beta^{(n)} \Delta^{(n-1)}, \quad (14)$$

with $\beta^{(n)}$ computed here using the Polak–Ribière formula (see, e.g., [25]):

$$\beta^{(n)} = \frac{\langle g^{(n)} , g^{(n)} \rangle - \langle g^{(n)} , g^{(n-1)} \rangle}{\langle g^{(n-1)} , g^{(n-1)} \rangle}. \quad (15)$$

The form $\langle , \rangle$ is the scalar product in the space spanned by $(P_{r}, O_{r})$. Its natural definition—we will consider other possibilities below—can be written as the dot product between column vectors $g_{P}, g_{O}$ and their Hermitian transpose $g_{P}^{\dagger}, g_{O}^{\dagger}$, with indices running over all positions $r$:

$$\langle g, g \rangle = g_{P}^{\dagger} g_{P} + g_{O}^{\dagger} g_{O}. \quad (16)$$

To complete the iteration step, a new point is found by minimizing $\mathcal{L}$ in the given search direction. How this line search is accomplished depends mostly on computing time considerations. For the Gaussian model, the 1D function to minimize takes the form of an order eight-polynomial,

$$\mathcal{L}(P_{r} + \gamma \Delta P_{r}^{(n)}, O_{r} + \gamma \Delta O_{r}^{(n)}) = \sum_{i=0}^{8} c_{i} \gamma^{i}, \quad (17)$$

where $\gamma$ is a real-valued step parameter. The coefficients $c_{i}$ can be readily computed (see appendix B). Here, line minimization simply amounts to selecting the relevant root computed from this polynomial. A possible strategy for reducing computing time is to neglect the coefficients above the second order and compute the minimum uniquely from the remaining quadratic form. This is equivalent to performing a single step in a Newton–Raphson line
minimization. For the Poisson and Euclidean metric cases, $\mathcal{L}$ does not reduce to a finite-order polynomial, but a few Newton–Raphson steps can be used instead.

The convergence speed for one iteration depends on the method used for line minimization and on the time required to compute the gradient. Also critical is the detailed formulation of the update rule for the determination of the consecutive search directions. The efficiency of the successive line minimizations in reaching convergence is ultimately determined by the metric of the embedding space or, in other words, the definition of the scalar product $\langle \cdot, \cdot \rangle$. Departing from the ‘native’ metric definition is a common procedure in optimization called preconditioning.

3.1. Preconditioning

An in-depth convergence study of ML-based Ptychographic reconstruction remains to be conducted. For the present work, we have developed two different \textit{ad hoc} forms of preconditioners. The first one comes from the mere realization that $O_r$ and $P_r$ are different physical quantities and that their relative multiplicative scaling is a free parameter of the system. A second preconditioner originates from the observation that low spatial frequencies of the object tend to evolve slowly despite their overall large power. Both preconditioners are special cases arising from a change of variable of the form

$$
\hat{L}(\hat{P}_r, \hat{O}_r) = L(U P_r, V O_r),
$$

where $U$ and $V$ are invertible linear operators acting on $P_r$ and $O_r$, respectively. The effect of a minimization with respect to $(\hat{P}_r, \hat{O}_r)$ can be absorbed in the conjugate direction calculation through the transformation of the scalar product of the gradient $\hat{g} = \nabla \hat{L}$

$$
\langle \hat{g}, \hat{g} \rangle = g_P^\dagger U g_P + g_O^\dagger V g_O.
$$

The first case described above corresponds to $U = s, V = 1$, where $s$ is a positive scalar. The update rule thus becomes

$$
\hat{\beta}^{(n)} = \frac{g_O^{(n)} g_P^{(n)} g_O^{(n)} g_P^{(n)} - s^2 g_P^{(n)} g_O^{(n)} g_O^{(n)} g_P^{(n)} - s^2 g_P^{(n-1)} g_O^{(n-1)} g_O^{(n)} g_P^{(n-1)}}{s^2 g_P^{(n)} g_O^{(n)} + g_O^{(n)} g_O^{(n)} g_P^{(n)} + g_O^{(n)} g_O^{(n)} g_P^{(n)}},
$$

with

$$
\Delta P_r^{(n)} = -s^2 g_P^{(n)} \hat{\beta}^{(n)} \Delta P_r^{(n-1)},
$$

$$
\Delta O_r^{(n)} = -g_O^{(n)} \hat{\beta}^{(n)} \Delta O_r^{(n-1)}.
$$

The scale $s$ is in principle a fixed parameter of the reconstruction. It is, however, possible to define \textit{ad hoc} rules that redefine $s$ as a function of the current estimates $O_r^{(n)}$ and $P_r^{(n)}$. We have observed that convergence could be substantially improved for a weakly scattering object—i.e., such that $O_r$ deviates only weakly from unity—by enforcing the near equality of the probe and object gradients,

$$
|\hat{g}_P|^2 = |\hat{g}_O|^2,
$$

leading to the formula

$$
s^2 = \frac{|g_O|^2}{|g_P|^2}.
$$

New Journal of Physics 14 (2012) 063004 (http://www.njp.org/)
In the second case, \( V \) is a low-pass filter operating on \( O_r \), with a preconditioned gradient best expressed in Fourier space:

\[
\hat{G}(\mathbf{q}) \hat{g}_{\mathbf{q}} = G(\mathbf{q}) g_{\mathbf{q}},
\]

This formulation, which can be understood as an application of Grenander’s method of sieves [26], is designed to favour minimization of the low spatial frequencies at first, but because the gradient spatial spectrum is not forced to be zero at high frequencies, these are also eventually refined. A simple choice for \( G \) is the 2D Hann window,

\[
G(\mathbf{q}) = \cos^2 \left( \frac{q_x \Delta x}{2} \right) \cos^2 \left( \frac{q_y \Delta y}{2} \right),
\]

where \( \Delta x \) and \( \Delta y \) are the pixel dimensions of the object array \( O_r \). Convolution with a Hann window is especially efficient as a real-space convolution since its kernel has only nine non-null entries. Other filters are also possible, and one can use for instance a Gaussian filter,

\[
G(\mathbf{q}) = \exp\left( -D |\mathbf{q}|^2 / 2 \right),
\]

where \( D \) is an adjustable threshold length scale between low and high spatial frequencies.

With \( U = 1 \), the update rule for the search direction becomes

\[
\Delta \beta^{(n)} = \frac{g_P^{(n)^\dagger}}{g_P^{(n-1)^\dagger}} + \frac{\hat{g}_{\mathbf{q}}^{(n)^\dagger}}{g_O^{(n-1)^\dagger}} \frac{g_{\mathbf{q}}^{(n)}}{g_O^{(n-1)^\dagger}} - \frac{g_P^{(n-1)^\dagger}}{g_P^{(n-1)^\dagger}} \frac{\hat{g}_{\mathbf{q}}^{(n-1)^\dagger}}{g_O^{(n-1)^\dagger}} \frac{\hat{g}_{\mathbf{q}}^{(n)}}{g_O^{(n-1)^\dagger}}
\]

and

\[
\Delta O_{\mathbf{q}}^{(n)} = -G(\mathbf{q}) \hat{g}_{\mathbf{q}}^{(n)} + \hat{\beta}^{(n)} \Delta O_{\mathbf{q}}^{(n-1)}.
\]

### 3.2. Regularization

Regularization is a set of techniques used to reduce difficulties arising from ill-posed or ill-conditioned problems. Like most physical inverse problems, our ML formulation is probably formally ill-posed. In addition, despite the numerical robustness of the method, we have observed a tendency to amplify high-spatial-frequency noise, a probable indication of a mild ill-conditioning. In the present context, regularization takes the form of a term added to the negative log-likelihood function:

\[
\mathcal{L}_R(O_r, P_r) = \mathcal{L}(O_r, P_r) + \frac{\mu}{K} R(O_r, P_r).
\]

Here \( \mu \) is an adjustable parameter and \( K \) is a renormalization constant, both described below. Such a formulation can be motivated by the addition to the original model of a priori information on the distribution of final solutions, an approach termed interchangeably ‘penalized likelihood’ or ‘maximum a posteriori’ in the literature. It can be seen that the regularizer \( R \) is analogous to an energy term that competes with the log-likelihood term. Here this term is used to impose a form of smoothness in the final reconstruction, in a way completely analogous to statistical reconstruction methods in computed tomography [27].

We use the simplest and most common regularizer on the object only, which would take the form \( R(O_r) \sim |\nabla O_r|^2 \) in the continuous case. This formulation penalizes any fluctuations in an image. Its discrete form is

\[
R(O_r) = \sum_r |O_{r+\Delta x} - O_r|^2 + |O_{r+\Delta y} - O_r|^2,
\]
where \( \Delta x \) and \( \Delta y \) are the grid spacing in \( x \) and \( y \), respectively. The derivative of this sum of squared discrete differences, to be added to \( g_O \) in the reconstruction algorithm, is

\[
\frac{\partial R}{\partial O_r} = 4O^*_r - O^*_{r+\Delta x} - O^*_{r-\Delta x} - O^*_{r+\Delta y} - O^*_{r-\Delta y},
\]

which is readily identified as a common discretized version of the Laplace operator.

A critical detail is the definition of a reasonable rule to set the amplitude \( \mu/K \) of the regularization term. The approach taken here is most directly obtained by considering the Gaussian model \((8)\). The expectation value of \( L \), evaluated at the solution and for an ensemble of possible experimental realizations, is seen to be

\[
\langle L \rangle = \sum_j \sum_q \frac{w_{jq}}{2\sigma_{jq}^2} \left( (I_{jq} - n_{jq})^2 \right)
\]

\[
= \frac{1}{2} N_m,
\]

where \( N_m \) is the total number of independent measurements (the number of valid detector pixels times the number of diffraction patterns). The expectation value of \( R \) depends on the actual object \( O_r \) but is bounded below by the variance \( \Delta O^2 \) originating from noise propagation:

\[
\langle R \rangle = 4 \sum_r \langle |O_r - \langle O_r \rangle|^2 \rangle
\]

\[
\leq 4 N_{\text{pix}} \Delta O^2,
\]

where \( N_{\text{pix}} \) is the total number of grid points (pixels) on which \( O_r \) is evaluated. The reconstruction error may be difficult to evaluate, but its lower bound is dictated by the finite number of photons measured in the complete experiment (i.e. summed over all diffraction patterns), \( N_{\text{phot}} \). Assuming a uniform fluence on the reconstructed area, the number of photons scattered within a single reconstruction pixel is \( N_{\text{phot}}/N_{\text{pix}} \) and thus

\[
\Delta O^2 \approx \frac{N_{\text{pix}}}{N_{\text{phot}}},
\]

These order of magnitude calculations lead us to define the renormalization constant \( K \) as

\[
K = 8 \frac{N_{\text{pix}}}{N_m N_{\text{phot}}},
\]

which gives the same order of magnitude and the same scaling, to \( \langle L \rangle \) and \( \langle R \rangle/K \). The amplitude \( \mu \) can therefore be interpreted as the relative power given to the regularization term, independent of experimental conditions. A value of \( \mu = 1\% \) was used throughout the reconstructions shown below. Much smaller values, for our choice of demonstrations, create ill-conditioned behaviour that manifests itself as strong high-spatial-frequency noise.

### 3.3. Alternative models

The Poisson, Gaussian and Euclidean models presented in section 2, although surprisingly robust in many situations, may be too simple for more complex experimental reality. Within the ML framework developed in this paper, the proper way to include additional constraints, as well as lack of knowledge, necessarily calls for a reformulation of the physical and data acquisition model. A proper model definition provides the essential basis for ulterior interpretation of the
reconstructed images. We show here two simple examples of such a model reformulation. The first one restricts the spatial extent of the probe, while the second relaxes the model and accommodates intensity fluctuations in the incoming beam.

The physical extent of the probe is in practice limited by far-field sampling considerations, and is most of the time naturally enforced by the finite number of pixels over which diffraction patterns are measured. If the scanned area is large, i.e., if the extent of the array representation of the object is much larger than the probe array, no support constraint on the illumination is generally needed. In the cases where the scanned area is just slightly larger than the probe extent however, spurious under-constrained amplitudes may start developing at the corners of the probe array. Avoiding such behaviour is in practice easily accomplished by imposing a loose support on the probe array. This operation is of course readily incorporated into iterative approaches. In the present optimization context, a finite probe extent is also simply enforced by setting to zero the values outside the support and removing altogether the degrees of freedom corresponding to the probe amplitudes lying outside the support during the optimization. In practice, the latter can be accomplished by setting to 0 all elements outside the support $S$ in the probe gradient:

$$g_{Pr}^{\text{support}} = \begin{cases} g_{Pr} & \text{if } r \in S, \\ 0 & \text{otherwise}. \end{cases}$$ (37)

Unlike for other CDI techniques, for ptychography the support does not need to be tight, and an extent of twice the size of the illumination or more is generally sufficient to avoid the appearance of the above-mentioned spurious amplitudes at the edges of the window, while allowing the probe to develop natural side lobes.

A similar constraint can be applied to the object array for areas that are known to be smooth and flat, such as the air surrounding a sample. Imposing this ‘flatness’ constraint also amounts to maintaining $g_o$ to zero over a given region, and is found to be especially useful in reducing reconstruction artefacts arising from raster scan patterns [28].

A model can be relaxed as much as it can be constrained. A valuable example of relaxation is the case of unknown shot-to-shot intensity fluctuations in the incoming beam. Adapting the likelihood model for this situation implies the introduction of new degrees of freedom quantifying the renormalization of the incoming intensity. For the Poisson model, this can be written as

$$L = -\sum_j \sum_q w_{jq} \left[ n_{jq} \log(\alpha_j I_{jq}) - \alpha_j I_{jq} - \log(n_{jq})! \right],$$ (38)

where $\alpha_j$ is a scaling parameter, to be optimized along with $P_r$ and $O_r$. Instead of adding $\alpha_j$ to the CG routine, the optimal value of $\alpha_j$ can be computed at every step. Setting $\partial L/\partial \alpha_j = 0$ in equation (38) leads to

$$\alpha_j = \frac{\sum_q w_{jq} n_{jq}}{\sum_q w_{jq} I_{jq}}.$$ (39)

Using this value, equation (38) becomes a likelihood metric that is independent of fluctuations in overall intensity [29]. Note that we achieve this while avoiding multiplicative scaling of the data $n_{jq}$, which would negatively affect our modelling of Poisson statistics. Similarly, by using this scaling the error metric contribution for a diffraction pattern is proportional to its overall incident intensity, which in the case of strong intensity fluctuations naturally de-emphasizes data that had a smaller incident intensity. This quantity can be computed from the current estimates
\((P_r^{(n)}, O_r^{(n)})\) for each computation of the likelihood metric. Using the chain rule it can be shown that the gradient may still be expressed as (5) and (6) with

\[
\tilde{\chi}_{j,q} = w_{j,q} \left( \alpha_j - \frac{n_{j,q}}{I_{j,q}} \right) \tilde{\psi}_{j,q}
\]

\[
= w_{j,q} \left( \frac{\sum_q n_{j,q}}{\sum_q I_{j,q}} - \frac{n_{j,q}}{I_{j,q}} \right) \tilde{\psi}_{j,q}.
\]  \(\text{(40)}\)

For the Gaussian model, the rescaling takes the form

\[
\alpha_j = \frac{\sum_q w_{j,q} n_{j,q} I_{j,q} / \sigma_{j,q}^2}{\sum_q w_{j,q} I_{j,q}^2 / \sigma_{j,q}^2}
\]  \(\text{(41)}\)

and the gradient is computed with

\[
\tilde{\chi}_{j,q} = \alpha_j w_{j,q} \frac{\left( \alpha_j I_{j,q} - n_{j,q} \right)}{\sigma_{j,q}^2} \tilde{\psi}_{j,q}.
\]  \(\text{(42)}\)

4. Results

The image reconstruction improvement by using ML optimization has been quantified on realistic simulations and verified with experimental data. Not all of the strategies presented above are demonstrated here. A critical ingredient, especially for the experimental data, is regularization, which is used for all reconstructions below with a parameter \(\mu = 1\%\). The preconditioners described above were not used for the reconstructions presented here. All difference map reconstructions were performed according to the previously published procedure [28], i.e. in the case of \(\beta = 1\) and using a relaxed Fourier constraint as in [12].

For the simulations, the test sample is the image of a histology slice from a rat embryo (http://www.brainmaps.org/), shown in figure 1(a). Organic structures have a characteristic distribution of length scales that provide relevant simulation conditions. The object transmission function is computed by assigning a phase shift and absorption value corresponding to the interpretation of the original image as a thickness map. Here we assumed that the sample is made of carbon and reaches a maximum thickness of 1 \(\mu m\). As is often the case with real samples, the real and imaginary parts of the resulting transmission function are strongly correlated; yet this fact is not enforced in the reconstructions. The illumination function is the free-space propagation over 0.7 mm of a 1.5 \(\mu m\) diameter pinhole whose edges have been slightly tapered, in accordance with typical experimental conditions. An x-ray energy of 6.2 keV is used throughout the simulations. In the simulations, a ptychography scan is created from probe positions lying on concentric circles with radii increasing in steps of 0.7 \(\mu m\), until a field of view of 12 \(\mu m \times 12 \mu m\) is covered. The resulting 232 diffraction patterns are sampled on a 256 \(\times\) 256 array. To keep conditions similar to the experimental data used below, we have set the size of the square detector pixels to 172 \(\mu m\) and the propagation distance to 7.05 m, resulting in a direct-space pixel size of 32 nm. The only source of noise is Poisson counting statistics, simulated using a pseudo-random number generator on the computed intensities rescaled to be consistent with a predefined photon flux.

Reconstructions shown here were performed in Python, taking advantage of the message passing interface (MPI) on a 2.27 GHz eight-core computer, using less than 1 GByte of RAM. Typical reconstruction times were a few minutes.
Figure 1. Simulation results. (a) The original image (histology section of a rat embryo (http://www.brainmaps.org/)), scaled to produce a phase shift of up to about $\pi/4$. (b) Difference map reconstruction. (c) ML reconstruction (Gaussian model). (d) ML reconstruction (Poisson model). All three images show the phase part of the reconstructions coming from the same dataset, in which Poisson noise was simulated assuming $10^6$ incident photons per diffraction pattern.

Three reconstruction techniques—the difference map, Gaussian approximation to Poisson and Poisson ML—are compared in figure 1 for a simulated dataset with $10^6$ incident photons per diffraction pattern. The qualitative superiority of ML refinement is clearly observed. While all three approaches seem to provide low- and mid-resolution details with similar accuracy, recovery of high spatial frequencies is significantly better for the ML cases.

The deviation of the reconstruction $O_r$ from the original image $O_{\text{model},r}$ (the ‘ground truth’) provides a quantitative evaluation of the reconstruction noise. The noise power spectrum can be computed from the absolute value squared of this deviation,

$$N(q) = |\tilde{O}_q - \tilde{O}_{\text{model},q}|^2.$$  

(43)

In practice, to avoid accounting for errors introduced by inherent ambiguities in the reconstruction, special care has to be taken to eliminate incompatible linear and constant phase terms and to register the images with sub-pixel precision. The Fourier-space signal-to-noise ratio (SNR) is the ratio of the original image power spectrum to the noise power spectrum:

$$\text{SNR}(q) = \frac{|\tilde{O}_{\text{model},q}|^2}{N(q)}.$$  

(44)
Figure 2. (a) Computed SNR of the reconstructions as a function of spatial frequency for the three reconstructions shown in figure 1. The SNR is seen to be one to two orders of magnitudes higher for the ML reconstructions than for the difference map one. (b) Noise power spectrum of the Gaussian ML reconstructions (thin lines) compared with the image power spectrum (thick line), for a range of incident photon counts per diffraction pattern. The dashed lines show the theoretical noise power levels for each incident photon count, computed assuming white Poisson noise spread uniformly over the imaged region.

An azimuthal average of this quantity is shown in figure 2(a). It is seen that, in terms of SNR, the reconstructions using ML (Poisson and Gaussian models) are in this case essentially equivalent and are able to recover high-frequency details that are not observed in the difference map reconstruction. The gain in signal to noise is close to two orders of magnitude, and the resolution for this ideal simulation is seen to increase roughly by a factor of 3, from about 300 nm to about 100 nm.\(^5\) Figure 2(b) shows a direct comparison of \(|\hat{O}_{\text{model},q}|^2\) (thick line) with \(N(q)\) from the

\(^5\) We define resolution as the smallest full period observable in the image, so the best achievable resolution is twice the pixel size.
Figure 3. Reconstructions of a hydroxyapatite sphere sample from x-ray data using the difference map (left column) and ML (right column). (a, b) The phase part of the object transmission function. The difference between the two images is not discernible with a colour scale covering the full range of more than $2\pi$. (c, d) Amplitude part of the transmission function. Improvements in the ML reconstruction are clearly visible. (e, f) Probe, reconstructed along with the transmission function. Higher spatial frequencies are retrieved more efficiently with ML. Inset: the reconstructed probe after numerical propagation back in the pinhole plane.

Gaussian ML reconstructions for a range of incident photon counts per diffraction pattern. It is seen that the ML reconstruction reproduces correctly the expected linear scaling of the noise power with the number of incident photons.

The performance of the Gaussian ML optimization on experimental data is illustrated in figure 3. The diffraction data for this reconstruction were acquired at the cSAXS beamline at the
Swiss Light Source at an x-ray energy of 6.2 keV. The sample, a porous hydroxyapatite sphere, was scanned with a $\sim 2 \mu m$ coherent illumination over a $30 \mu m \times 36 \mu m$ area. A total of 387 diffraction patterns were measured, each cropped to a $128 \times 128$ array, giving a reconstruction pixel size of 65.3 nm. The total number of detected photons varied from $7 \times 10^5$ in the most absorbing parts to $1.45 \times 10^6$ in the empty region surrounding the sample. With an average fluence of about 1000 photons per reconstruction pixel, the experimental conditions are similar to the simulation shown in figure 1.

A first reconstruction was obtained with the difference map algorithm. A second image was produced with the Gaussian model ML as a refinement of the difference map solution. Variances for this model were computed using the least informative prior (see appendix A), and the model was relaxed according to equation (38). The results of the two reconstruction steps are shown in figure 3. Figures 3(a) and (b) display the phase part of the reconstruction after linear phase removal and phase unwrapping, with no discernible improvement. The reconstructed specimen absorption, as depicted in figures 3(c) and (d), has a much lower quality than the phase part. While fluctuations in the beam intensities sometimes cause stronger degradation of the absorption part compared to the phase part, in the current situation both signals have almost identical noise amplitudes and the quality difference is mostly attributable to the large difference in the sample’s natural dynamic range. Figures 3(e) and (f) show the probe reconstructed simultaneously with the object, with a clear improvement in the higher spatial frequencies of the ML reconstruction. Numerical propagation back to the pinhole plane confirms that the additional details in figure 3(f) are consistent with the experimental conditions.

Figures 4(a)–(b) show once more the phase part of the object, this time with a greyscale reduced to emphasize the small fluctuations in the air surrounding the sample to help distinguish the quality improvement following an ML refinement. The improved image quality is clearly seen in the enlargement in figures 4(c)–(d), which shows a faint filament of the glue used to fix the specimen. Without a ‘ground truth’ as in the simulations above, a full noise analysis is not possible on the experimental data. However, assuming that the noise is isotropic throughout the image, an estimate of the noise power spectrum can be computed using a flat region of the reconstruction. Such estimates, shown in figures 4(e)–(f), indicate that the noise in the DM reconstruction is mostly located in the mid-spatial frequencies, while the ML reconstruction has a lower, flatter power spectrum except for a few low-spatial-frequency peaks that originate from the regular scanning pattern in the region used for this analysis. The variance within the white dashed region in figures 4(a)–(b), or equivalently the integrated power from figures 4(e)–(f), shows nearly an order of magnitude reduction in noise by ML optimization of the solution from the DM, giving estimates of the phase uncertainty (mean standard deviation of the noise in the phase) from $\Delta \phi_{DM} = 0.11$ to $\Delta \phi_{ML} = 0.036$. Assuming that the object scattering power follows a $q^{-4}$ power law, the increase by one order of magnitude should lead to a nearly twofold increase in resolution.

It is not possible to guarantee that a similar or higher gain in SNR can be expected with other ptychography datasets. Most certainly, experimental measurements can contain other sources of data degradation which are unaccounted for in the simple physical model described above. In the future, a more careful modelization of the x-ray–matter interaction and of further degradation of data by instrumental imperfections—such as sample vibrations and the detector point-spread function—could help reduce further noise and artefacts in reconstructions.
Figure 4. Phase part of the hydroxyapatite sphere reconstruction depicted in a narrow greyscale to show the reduction in noise and artefact in the ML reconstruction (right column) compared to DM (left column). (a, b) Full reconstruction, with the same field of view as in figure 3. (c, d) Enlargement of the region highlighted by a black frame in (a) and (b), showing the strong reduction of noise in the flat air surrounding the sample and improved resolution of a weakly scattering feature. (e, f) Squared amplitude of the Fourier transform of the flat region highlighted by a dashed white frame in (a) and (b), shown on a logarithmic greyscale. These two images give a good estimate of the noise power spectrum in the reconstruction.

5. Discussion

5.1. Convergence and computation time

The above results were concerned mostly with the properties of the final images, but little was said about the convergence properties of ML reconstructions. While progress has been made in recent years [30], convergence is not yet guaranteed and there is no proof of the uniqueness
of the minimum for noisy data. Fortunately, empirical evidence with ptychographic datasets indicates for now that ML refinement rarely leads to stagnation or to false minima. However, the convergence time of ML methods depends strongly on the initial conditions and additional care has to be taken to produce a reasonable initial estimate. In particular, we have observed in a few instances that the reconstruction progress was greatly slowed down when phase vortices, not part of the solution, were present in the probe or the object. Such vortices may be introduced unintentionally in the initial iterates if both the object and the probe differ significantly from the truth, as is often the case in practical scenarios. The difference map, designed to be robust and avoid stagnation, typically converges within 10–100 iterations when provided with a much rougher initial estimate—a Gaussian probe and a flat object typically suffice—and is in fact much more efficient in removing the spurious vortices from the reconstruction when present. Consequently, we found that the most efficient strategy for ptychographic reconstruction is to use both reconstruction procedures successively. The total computation time can be optimized by choosing the number of initial difference map iterations prior to the ML procedure. If the number of DM iterations is too short, the initial conditions provided to the ML algorithm are not close enough to the solution, and convergence happens late. On the other hand, a very large number of DM iterations becomes a waste of computation time since the difference map stops progressing towards the ML optimum. As a consequence, there exists an optimal number of prior difference map iterations that will lead to the quickest convergence. While this number depends on many parameters, we have observed that typical reconstruction scenarios require between 50 and 100 difference map iterations prior to ML refinement.

5.2. Application to other coherent diffractive imaging techniques

The formalism presented in this paper can be straightforwardly applied to diffraction microscopy by restricting to a single term the sum over \( j \) in equations (4), (8) or (11), with \( O_r = 1 \) and consequently \( P_r = \psi_r \). Obviously, the system is then strongly underconstrained, and without further model restriction, optimization of any of the likelihood functions is done in a single step, where the Fourier amplitudes of the current estimate of the probe are replaced with \( \sqrt{n_q} \). The support constraint as implemented in equation (37), typically not necessary for ptychography as mentioned above, becomes essentially the only physical input in the likelihood formulation of diffraction microscopy. Then ML refinement of a solution from a projection-based algorithm, and using a tight support constraint, may be a feasible route for the application of the concepts presented here to CDI. The Euclidean version of the negative log-likelihood, equation (11), was studied by Fienup [7] many decades ago, who found that its minimization is equivalent to the error-reduction (ER) procedure. How ML refinement can help in the quality of the reconstruction remains to be investigated and will depend strongly on the amount of information fed into the physical model. Likelihood maximization is most useful when a well-constrained physical model is to be adjusted to noisy data. It is expected to be effective for image reconstruction problems where many measurements are combined in a model, but may provide a smaller gain in simpler cases where the model permits less information mixing. In diffraction microscopy, the support imposes localized correlations in diffraction patterns in the form of speckles. Therefore, ML refinement should prove especially useful with highly oversampled diffraction patterns, where multiple data pixels contribute to a single speckle.
6. Conclusion

ML estimation is used for a wide range of statistical models. Provided that the physical and noise models are appropriate, the optimization principle described in this work could be most efficient in extracting from diffraction data all the information available. The resulting reconstructions have a higher SNR than those obtained from ‘statistics-unaware’ methods such as the difference map and the extended ptychographic iterative engine (ePIE). The observed decrease in reconstruction noise implies that future experiments could reduce by a factor up to ten the exposure times while reaching the same quality level as iterative algorithms. Shorter acquisition times increase the throughput of CDI experiments and reduce the impact of slow drifts in the experimental setup. Our results also indicate that ML principles will be instrumental in pushing the resolution of CDI to its theoretical limit, especially for dose-limited imaging applications.

From a methodological point of view, optimization approaches have the benefit of separating unambiguously the model formulation from the implementation, thereby lifting in part uncertainties regarding the ambiguities, validity and interpretation of a reconstruction. Of course, these advantages only have value as long as one is confident about the uniqueness of the solution. While no proof of this uniqueness exists yet for ptychography, mounting empirical evidence from experimental data provides some confidence that ML estimation can be safely applied in many cases.

The requirement for a precise a priori model formulation also suggests that much more complicated systems could be treated with the same strategy. Obvious novel possibilities include the formulation of a more complete scattering model that incorporates incoherent effects such as Compton scattering, fluorescence and electron emission. Since ptychography is a scanning technique, such complementary imaging modality could in principle be combined to form heterogeneous datasets bound together by an appropriate model.

Among other research avenues, it seems probable that faster and more efficient implementations could be obtained, thanks to the large body of literature on optimization techniques.

Acknowledgments

Special thanks are due to A Menzel for his help in improving the manuscript. We are also grateful to M Dierolf, J R Fienup and O Bunk for many fruitful discussions. We thank M Kerschnitzki for providing the hydroxyapatite sample and A Diaz for support with the cSAXS beamline. M Dierolf, B Enders and A Fehringer helped in the acquisition of the x-ray data. This work was supported in part by the ERC Starting Grant, under the project OptimaX (number 279753), and by the DFG Cluster of Excellence MAP (Munich-Center for Advanced Photonics).

Appendix A. Estimating Gaussian parameters from Poisson measurements

Applying the Gaussian model equation (8) to datasets obtained from photon-counting detector brings the question of how to define the variances $\sigma_{ij}^2$ from the measured discrete photon counts. A proper estimate can be computed from a prior probability distribution for $I_{ij}$. In the cases where the envelope of the diffracted intensities is mostly radially symmetric, intensities in a
given resolution ring are expected to follow the Wilson statistics [31]:

\[ p(I) = \exp(-I_{jq}/\langle I_j \rangle_{|q|}), \quad (A.1) \]

where \( \langle I_j \rangle_{|q|} \) is the average of the intensity within a ring of radius \(|q|\) in the diffraction pattern. Using Bayes rules, the resulting estimate for the variance is

\[ \hat{\sigma}^2_{jq} = \frac{n_{jq} + 1}{(1 + \langle I_j \rangle_{|q|})^2}. \quad (A.2) \]

Computing \( \langle I_j \rangle_{|q|} \) may be difficult however. Not only does one have to choose the appropriate width of the integration ring, but even the centre of a continuous diffraction pattern is not necessarily well defined. Fortunately, it is sufficient in most cases to simply use a less informative prior, given by the limit \( \langle I \rangle \rightarrow \infty \), which corresponds to a uniform distribution for all positive values of \( I \). In this limit, one finds that \( \sigma^2_{jq} = n_{jq} + 1 \), thus alleviating the risk of divisions by zero in equation (9).

**Appendix B. Polynomial expansion of the Gaussian model**

The polynomial coefficients in (17) for the line minimization can be computed in a straightforward manner. It suffices to note that a displacement in the direction \((\Delta P_r, \Delta O_r)\) in (8) gives

\[ \mathcal{L}(P_r + \gamma \Delta P_r, O_r + \gamma \Delta O_r) = \sum_{jq} w_{jq} \left( \frac{4}{2\sigma^2_{jq}} \right) \left( \sum_{n=0}^{4} A_{n,jq} \gamma^n \right)^2. \quad (B.1) \]

The coefficients are obtained by summing over \( j \) and \( q \) the nine terms resulting from the expansion of the square within the sum. The arrays \( A_{n,jq} \) can be written as

\[
\begin{align*}
A_{0,jq} &= (|\tilde{\psi}_{jq}|^2 - I_{jq}), \\
A_{1,jq} &= 2\Re\left\{\tilde{\psi}_{jq} a_{jq}^*\right\}, \\
A_{2,jq} &= 2\Re\left\{\tilde{\psi}_{jq} b_{jq}^*\right\} + |a_{jq}|^2, \\
A_{3,jq} &= 2\Re\left\{a_{jq} b_{jq}^*\right\}, \\
A_{4,jq} &= |b_{jq}|^2,
\end{align*}
\]

with \( \tilde{\psi}_{jq}, a_{jq} \) and \( b_{jq} \) given by

\[
\begin{align*}
\tilde{\psi}_{jq} &= \mathcal{F}[P_{r-r}, O_r], \\
a_{jq} &= \mathcal{F}[P_{r-r}, \Delta O_r + O_r \Delta P_{r-r}], \\
b_{jq} &= \mathcal{F}[\Delta P_{r-r}, \Delta O_r]. \quad (B.3)
\end{align*}
\]

Assuming that the number of fast Fourier transform calls per iteration is a good estimate for the scaling of the total computation time, computing the polynomial coefficients is found to be six times slower than computing the log-likelihood and three times slower than computing the gradient. This rough scaling can be used to choose the best line minimization strategy.
References

[1] Nugent K A 2010 Adv. Phys. 59 1–99
[2] Thibault P and Elser V 2010 Annu. Rev. Condens. Matter. Phys. 1 237–55
[3] McNulty I, Kirz J, Jacobsen C, Anderson E H H, Howells M R R and Kern D P P 1992 Science 256 1009–12
[4] Eisebitt S, Luning J, Schlotter W F, Lorgen M, Hellwig O, Eberhardt W and Stühr J 2004 Nature 432 885–8
[5] Marchesini S et al 2008 Nature Photonics 2 560–3
[6] Zhu D et al 2010 Phys. Rev. Lett. 105 43901
[7] Fienup J R 1982 Appl. Opt. 21 2758–69
[8] Elser V 2003 J. Opt. Soc. Am. A 20 40–55
[9] Russel Luke D 2005 Inverse Probl. 21 37–50
[10] Dilanian R A, Williams G J, Whitehead L W, Vine D J, Pelee A G, Balaur E, McNulty I, Quiney H M and Nugent K A 2010 New J. Phys. 12 093042
[11] Chapman H N et al 2006 J. Opt. Soc. Am. A 23 1179–200
[12] Giewekemeyer K, Thibault P, Kalbfleisch S, Beerlink A, Kewish C M, Dierolf M, Pfeiffer F and Salditt T 2010 Proc. Natl Acad. Sci. USA 107 529
[13] Thibault P, Elser V, Jacobsen C, Sayre D and Shapiro D A 2006 Acta Crystallogr. A 62 248–61
[14] Guizar-Sicairos M and Fienup J R 2009 Opt. Express 17 2670–85
[15] Kewish C M et al 2010 Opt. Express 18 23420–7
[16] Guizar-Sicairos M, Narayanan S, Stein A, Metzler M, Sandy A R, Fienup J R and Evans-Lutterodt K 2011 Appl. Phys. Lett. 98 111108
[17] Fienup J R 1993 Appl. Opt. 32 1737–46
[18] Brady G R and Fienup J R 2006 Opt. Express 14 474–86
[19] Brady G R, Guizar-Sicairos M and Fienup J R 2009 Opt. Express 17 624–39
[20] Miao J, Charalambous P, Kirz J and Sayre D 1999 Nature 400 342–4
[21] Abbey B, Nugent K A, Williams G J, Clark J N, Pelee A G, Pfeifer M A, de Jonge M and McNulty I 2008 Nature Phys. 4 394–8
[22] Zhang F, Pedrini G and Osten W 2007 Phys. Rev. A 75 43805
[23] Johnson I, Jefimovs K, Bunk O, David C, Dierolf M, Gray J, Renker D and Pfeiffer F 2008 Phys. Rev. Lett. 100 155503
[24] Kaup L., Kaup B and Barthel G 1983 Holomorphic Functions of Several Variables: An Introduction to the Fundamental Theory (New York: Walter de Gruyter)
[25] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 Numerical Recipes in C: The Art of Scientific Computing 2nd edn (Cambridge: Cambridge University Press)
[26] Grenander U 1981 Abstract Inference (New York: Wiley)
[27] Geman S and McClure D E 1985 Bayesian image analysis: an application to single photon emission tomography Proc. Statistical Computing Section (American Statistical Association) pp 12–8
[28] Thibault P, Dierolf M, Menzel A, Bunk O and Pfeiffer F 2009 Ultramicroscopy 109 338–43
[29] Thurman S T and Fienup J R 2009 J. Opt. Soc. Am. A 26 1008–14
[30] Yang C, Qian J, Schirotzek A, Maia F and Marchesini S 2011 arXiv:1105.5628v1 32
[31] Wilson A J C 1949 Acta Crystallogr. 2 318–21