Noise models for low counting rate coherent diffraction imaging

Pierre Godard,[1,2,*] **Marc Allain,**[1] **Virginie Chamard,**[1] **and John Rodenburg**[2]

[1]*Institut Fresnel - Université Aix Marseille, CNRS, Faculté de St Jérôme, 13397 Marseille Cedex 20, France*

[2]*Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, UK*

[*]*pierre.godard@fresnel.fr*

Abstract: Coherent diffraction imaging (CDI) is a lens-less microscopy method that extracts the complex-valued exit field from intensity measurements alone. It is of particular importance for microscopy imaging with diffraction set-ups where high quality lenses are not available. The inversion scheme allowing the phase retrieval is based on the use of an iterative algorithm. In this work, we address the question of the choice of the iterative process in the case of data corrupted by photon or electron shot noise. Several noise models are presented and further used within two inversion strategies, the ordered subset and the scaled gradient. Based on analytical and numerical analysis together with Monte-Carlo studies, we show that any physical interpretations drawn from a CDI iterative technique require a detailed understanding of the relationship between the noise model and the used inversion method. We observe that iterative algorithms often assume implicitly a noise model. For low counting rates, each noise model behaves differently. Moreover, the used optimization strategy introduces its own artefacts. Based on this analysis, we develop a hybrid strategy which works efficiently in the absence of an informed initial guess. Our work emphasises issues which should be considered carefully when inverting experimental data.

© 2012 Optical Society of America

OCIS codes: (030.4280) Noise in imaging systems; (100.5070) Phase retrieval; (110.0180) Microscopy.

References and links

1. K. A. Nugent, "Coherent methods in the X-ray sciences," Adv. Phys. **59**, 1–100 (2010).
2. J. M. Rodenburg, "Ptychography and related diffracted imaging methods," in "Advances in Imaging and Electron Physics," **150**, P. W. Hawkes, ed. (Elsevier, 2008), 87–184.
3. J. M. Rodenburg, A. C. Hurst, and A. G. Cullis, "Transmission microscopy without lenses for objects of unlimited size," Ultramicroscopy **107**, 227–231 (2007).
4. D. Claus, A. M. Maiden, F. Zhang, F. G. R. Sweeney, M. Humphry, H. Schluesener, and J. M. Rodenburg, "Quantitative phase contrast optimised cancerous cell differentiation via ptychography," Opt. Express **20**, 9911–9918 (2012).
5. M. Beckers, T. Senkbeil, T. Gorniak, M. Reese, K. Giewekemeyer, S. C. Gleber, T. Salditt, and A. Rosenhahn, "Chemical constrasts in soft X-ray ptychography," Phys. Rev. Lett. **107**, 208101 (2011).
6. J. M. Rodenburg, A. C. Hurst, A. G. Cullis, B. R. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson, "Hard-X-ray lensless imaging of extended objects," Phys. Rev. Lett. **98**, 34801 (2007).

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

7. P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, "High-resolution scanning X-ray diffraction microscopy," Science **321**, 379–382 (2008).
8. M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer, "Ptychographic X-ray computed tomography at the nanoscale," Nature **467**, 436–439 (2010).
9. P. Godard, G. Carbone, M. Allain, F. Mastropietro, G. Chen, L. Capello, A. Diaz, T. H. Metzger, J. Stangl, and V. Chamard, "Three-dimensional high-resolution quantitative microscopy of extended crystals," Nat. Commun. **2**, 1569 (2011).
10. F. Hue, J. M. Rodenburg, A. M. Maiden, F. Sweeney, and P. A. Midgley, "Wave-front phase retrieval in transmission electron microscopy via ptychography," Phys. Rev. B **82**, 121415 (2010).
11. M. J. Humphry, B. Kraus, A. C. Hurst, A. M. Maiden, and J. M. Rodenburg, "Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging," Nat. Commun. **3**, 730 (2012).
12. C. T. Putkunz, A. J. D'Alfonso, A. J. Morgan, M. Weyland, C. Dwyer, L. Bourgeois, J. Etheridge, A. Roberts, R. E. Scholten, K. A. Nugent, and L. J. Allen, "Atom-scale ptychographic electron diffractive imaging of boron nitride cones," Phys. Rev. Lett. **108**, 73901 (2012).
13. S. Eisebitt, J. Lüning, W. F. Schlotter, M. Lörgen, O. Hellwig, W. Eberhardt, and J. Stöhr, "Lensless imaging of magnetic nanostructures by X-ray spectro-holography," Nature **432**, 885–888 (2004).
14. V. Chamard, J. Stangl, D. Carbone, A. Diaz, G. Chen, C. Alfonso, C. Mocuta, G. Bauer, and T. H. Metzger, "Three-dimensional x-ray Fourier transform holography: the Bragg case," Phys. Rev. Lett. **104**, 165501 (2010).
15. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," Ultramicroscopy **109**, 1256–1262 (2009).
16. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264–7276 (2008).
17. A. M. Maiden, M. J. Humphry, M. C. Sarahan, B. Kraus, and J. M. Rodenburg, "An annealing algorithm to correct positioning errors in ptychography," Ultramicroscopy **120**, 64–72 (2012).
18. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," New J. Phys. **14**, 063004 (2012).
19. R. A. Fisher, *Statistical Methods and Scientific Inference* (Oliver & Boyd, 1956).
20. M. G. Kendall and A. Stuart, *The advanced theory of statistics* **2a** (Griffin, 1963).
21. F. Livet, F. Bley, J. Mainville, R. Caudron, S. G. J. Mochrie, E. Geissler, G. Dolino, D. Abernathy, G. Grübel, and M. Sutton, "Using direct illumination CCDs as high resolution area detector for X-ray scattering," Nucl. Instr. Meth. A **451**, 596–609 (2000).
22. C. Ponchut, J. Clément, J.-M. Rigal, E. Papillon, J. Vallerga, D. LaMarra, and B. Mikulec, "Photon-counting X-ray imaging at kilohertz frame rates," Nucl. Instrum. Meth. A **576**, 109–112 (2007).
23. C. Broennimann, E. F. Eikenberry, B. Henrich, R. Horisberger, G. Huelsen, E. Pohl, B. Schmitt, C. Schulze-Briese, M. Suzuki, T. Tomizaki, H. Toyokawa, and A. Wagner, "The Pilatus 1M detector," J. Synchrotron Rad. **13**, 120–130 (2006).
24. K. Lange and R. Carson, "EM reconstruction algorithm for emission and transmission tomography," IEEE Trans. Med. Imag. **8**, 306–316 (1984).
25. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imag. **1**, 113–122 (1982).
26. L. B. Lucy, "An iterative technique for the rectification of observed distribution," New Astron. Rev. **79**, 745–754 (1974).
27. G. Williams, M. Pfeifer, I. Vartanyants and I. Robinson, "Effectiveness of iterative algorithms in recovering phase in the presence of noise," Acta Cryst. **A63**, 36–42 (2007).
28. P. Godard, M. Allain, and V. Chamard, "Imaging of highly inhomogeneous strain field in nanocrystals using x-ray Bragg ptychography: A numerical study," Phys. Rev. B **84**, 144109 (2011).
29. J. Vila-Comamala, A. Diaz, M. Guizar-Sicairos, A. Mantion, C. M. Kewish, A. Menzel, O. Bunk, and C. David, "Characterization of high-resolution diffractive X-ray optics by ptychographic coherent diffractive imaging," Opt. Express **19**, 21333–21344 (2011).
30. Provided that the fluctuations in one measurement are accurately described by a Poisson PDF, then this PDF is defined by a single (positive) parameter that is the mean and the variance.
31. J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phys. Lett. **85**, 4795–4797 (2004).
32. M. F. Freeman and J. W. Tuckey, "Transformations related to the angular and the square root," Ann. Math. Statist. **21**, 607–611 (1950).
33. C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," IEEE Trans. Image Process. **5**, 480–492 (1996).
34. L. Bouchet, "A comparative-study of deconvolution methods for gamma-ray spectra," Astron. Astrophys. **113**, 167–183 (1995).
35. M. Allain and J.-P. Roques, "High resolution techniques for gamma-ray diffuse emission: application to INTEGRAL/SPI," Astron. Astrophys. **43**, 1175–1187 (2006).
36. By definition, the likelihood is the PDF of the noise model seen as a function of the unknown parameters ρ. In

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

practice, the opposite of the logarithm of the likelihood is rather considered. However, the logarithm function being a monotonic increasing function, the minimiser of the neg-loglikelihood is also the maximiser of the likelihood, *i.e.*, the ML estimator.

37. Following [33], it is shown that a second order Taylor expansion around $h_{m,j} = y_{m,j}$ of the Poissonian fitting function $\mathscr{L}_{\mathscr{P}}$ leads to (10e).
38. M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging* (Institute of Physics Publishing, 1998).
39. Since the condition $h_{m,j} > 0$ is enforced if $b_{m,j} > 0$ [cf. Eq. (1)], an arbitrary small background component can be introduced, hence allowing all the fitting functions and gradients to be well defined.
40. J. R. Fienup and C. C. Wackerman, "Phase-retrieval stagnation problems and solutions," J. Opt. Soc. Am. A **3**, 1897–1907 (1986).
41. In the optimization literature, OS algorithms are also known as *incremental gradient methods* or *block iterative methods*, see for instance [42, Sec. 1.5.2] or [43] for details.
42. D. P. Bertsekas, *Nonlinear programming*, 2nd ed. (Athena Scientific, 1999).
43. Y. Censor, D. Gordon, and R. Gordon, "BICAV: a block-iterative parallel algorithm for sparse systems with pixel-related weighting," IEEE Trans. Med. Imag. **20**, 1050–1060 (2001).
44. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered-subset of projection data," IEEE Trans. Med. Imag. **13**, 601–609 (1994).
45. S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subset algorithms," IEEE Trans. Med. Imag. **22**, 613–626 (2003).
46. The original version of the PIE introduced by Rodenburg and Faulkner in [31] considers another definition for $\boldsymbol{D}_j$.
47. A. M. Maiden, Humphry, and J. M. Rodenburg, "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," J. Opt. Soc. Am. **29**, 1606–1614 (2012).
48. C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini, "Iterative algorithms for ptychographic phase retrieval," arXiv:optics (2011).
49. Since the three fitting functions $\mathscr{L}_{\mathscr{P}}$, $\mathscr{L}_{\mathscr{G}}$ and $\mathscr{L}_{\mathscr{R}}$ are equivalent w.r.t. a nil data, only the data such that $y_{m,j} \neq 0$ should be considered in order to discriminate the noise-models.
50. This non-monotonic behaviour of the relative error is standard when inverse problems (*e.g.*, image restoration or tomographic reconstruction) are solved with gradient optimization technics, see for instance [38, Chap. 6].
51. Ph. Réfrégier, *Noise Theory and Application to Physics: From Fluctuation to Information* (Springer, 2004).
52. V. Elser, "Phase retrieval by iterated projections," J. Opt. Soc. Am. A **20**, 40–55 (2003).
53. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," Ultramicroscopy **109**, 338–343 (2009).
54. From [52, p. 339], one notes that the constraint defined by the data set in the DM strategy takes the form of Eq. (12b), suggesting that the data fluctuations are described by the Gaussian model defined in Sec. 2.2.
55. J. F. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," Biometrika **35**, 246–254 (1948).

1. Introduction

Coherent diffraction imaging (CDI) is a class of microscopy method that circumvents the need of high quality optics. It is based on the calculation of a numerical lens to retrieve the quantitative sample image from coherently diffracted intensity measurements. The information obtained contains both the amplitude and phase distributions of the exit-wave field. This quantity can be related to various structural parameters such as absorption, dispersion, magnetization state, crystalline structure, etc. [1]. Among the proposed CDI approaches, ptychography is particularly attractive since it allows the reconstruction of non-isolated objects, without *a priori* restrictions on the field of view [2] and without requiring any specific sample preparation. The ptychographic approach consists in scanning a sample across a finite-support beam and recording a diffraction intensity pattern for each probe position; assuming that the scan step is small enough, each point of the sample is encoded several times and in a different way. This redundancy ensures that the phase retrieval of the complex-valued diffracted field can be achieved. It is usually performed by iterative algorithms that combine the intensity patterns.

Successful results have been obtained with visible light [3, 4], with soft [5] and hard [6–9] x-rays, and in electron microscopy [10–12]. Major advantages of the ptychography approach are linked to the absence of serious physical aberrations: the method is lens-less, does not require any reference beam or sample [13, 14], and is robust to inaccurately known parameters that

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

can be retrieved simultaneously with the object image. Examples of this last issue include the illumination function [7, 15, 16], the probe positions [16, 17] and intensity fluctuations in the incoming beam [18].

However, as the approach is based on an iterative algorithm, it can face problems with convergence, uniqueness of the solution, *etc*. The successive iterations lead to a solution which is reached when the constraints resulting from the overlapping condition and the intensity measurements are satisfied simultaneously. In the presence of shot noise, such a solution does not exist as the different intensity patterns are not anymore consistent one with another. Low counting statistics are of key importance in the study for instance of radiation-sensitive objects (especially biological structures), or when the object scatters weakly, or when one attempts to obtain very high-resolution images although only few photons are scattered at the needed high angles.

In this work, we address precisely the question of the degradation of the solution that is obtained in a phase retrieval approach in presence of photon noise. While we study specifically the ptychographic scheme as an example, our methods and conclusions can be extended to the other iterative algorithm based lens-less imaging microscopies. We believe that the interested reader will find herein the material needed to adapt our approach to the case of support-based phase retrieval algorithm.

We begin by defining some common noise models in order to derive a fitting function by the mean of the *maximum likelihood* principle [19, 20]. A noise-model dependent reconstruction is thereby obtained by the minimization of the corresponding fitting functions. For this purpose, two different optimization strategies are examined, namely the ordered subset (OS) and the scaled gradient (SG). The former strategy is equivalent to the well known ptychographical iterative engine (PIE) when the additional assumption of a Gaussian noise model is considered. It has the advantage of fast convergence in the early iterations, but its final convergence is precluded by the inconsistencies in the different diffraction patterns. In contrast, the latter is slower in the early iterations, but its asymptotic convergence remains in presence of noise. For the different inversion schemes, a Monte-Carlo analysis is conducted for different noise levels, allowing a direct comparison of the solutions. The quantitative evaluation of each pair "noise model/optimization strategy" is done through quality indicators like the bias and standard deviation. Our results demonstrate the large variety of trade-offs resulting directly from the use of inversion schemes and from the implicit physical models. These are discussed in detail. The conclusions we reach have important implications for experimental applications of diffractive imaging.

The next section of this article presents the noise models that are considered for a CDI experiment. Section 3 gives the fitting functions that are derived from the maximum likelihood principle. Then, two iterative strategies that can be used for retrieving the object from the chosen fitting function are described. Section 4 presents the main results of this study: first, the definition of some performance indicators together with the description of the numerical sample are given; second, the convergence properties of the iterative strategies are briefly discussed; finally, Monte-Carlo analysis of the reconstruction algorithms are considered with regard to the selected noise models.

2. Noise models for ptychographic data sets

The ptychography approach requires the description of the exit field as a function of the probe $p(\boldsymbol{r})$ and the sample scattering function $\rho(\boldsymbol{r})$, named the object in the following. In the multi-

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

plicative approximation, the j-th exit-field $\psi_j(\boldsymbol{r})$ is given by

$$\psi_j(\boldsymbol{r}) \;:=\; p_j(\boldsymbol{r}) \times \rho(\boldsymbol{r})$$

where ρ is unknown and $p_j(\boldsymbol{r}) := p(\boldsymbol{r}-\boldsymbol{r}_j)$ is the probe function shifted in $\boldsymbol{r}_j$. From a practical viewpoint, the reconstruction from ptychographic data requires the object and the probe to be discretised. In what follows, we denote by

$$\rho = \{\rho_n\}_{n=1}^{N} \qquad \text{with} \qquad \rho_n := \rho(\boldsymbol{r}_n),$$

the object to be retrieved; N is thus the number of pixels in the object plane. This object is illuminated by a support-limited probe $\boldsymbol{p}_j := \{p_j(\boldsymbol{r}_n)\}_{n=1}^{M}$, where M is the number of pixels in the camera. This vector is converted into an $M \times N$ matrix $\boldsymbol{P}_j$ so that the exit field is expressed in vector form by

$$\psi_j := \boldsymbol{P}_j \rho$$

where the index j refers to the position of the probe. The corresponding far-field $\Psi_j := \{\Psi_{m,j}\}_{m=1}^{M}$ is computed from the exit-field by

$$\Psi_j := \boldsymbol{W} \psi_j$$

where $\boldsymbol{W}$ is the discrete Fourier transform operator. Provided that the size of the camera pixel or detector is small enough, the *expected* number of photons in the m-th detector reads

$$h_{m,j} := \mathscr{A} \left|\Psi_{m,j}\right|^2 + b_{m,j} \qquad \forall (m,j) \tag{1}$$

where $b_{m,j}$ is the expected number of *background events* and $\mathscr{A}$ is the area of the detector. Since $\mathscr{A}$ can be incorporated into the probe, one can set $\mathscr{A} = 1$ without loss of generality, so that $h_{m,j} = \left|\Psi_{m,j}\right|^2 + b_{m,j}$ is the *expected* number of events for the m-th measurement in the j-th illumination.

The above relations give a deterministic relationship between the object ρ and the expected (noise-free) data set $\{h_{m,j}\}$ that is at the basis of any reconstruction numerical scheme. However, when realistic data are considered, the presence of photon noise results in a substantial degradation of the measured data set $\boldsymbol{y} := \{y_{m,j}\}$ relative to $\{h_{m,j}\}$. In order to take into account the noise issue in a ptychographic experiment, three distinct noise models are introduced in the following sections. Each of them leads to a specific criterion that links the unknown object to the measured data. We will show that these criteria are fitting functions that provide an estimate of the object further obtained *via* a minimization algorithm.

2.1. Noise Model $\mathscr{P}$: The standard photon counting model

The far-field intensity is a quantity with nonnegative real number values; however, a detector collects a finite number of photons: this number takes integer values that can be considered as a random variable. The standard probability distribution function (PDF) considered for particle counting is the Poisson probability law. Assuming independent measurements $y_{m,j}$, the probability that the entire data set $\boldsymbol{y}$ is collected reads

$$f_{\mathscr{P}}(\boldsymbol{y};\rho) \;=\; \prod_j \prod_m e^{-h_{m,j}} \times \frac{[h_{m,j}]^{y_{m,j}}}{y_{m,j}\,!}\,. \tag{2}$$

For experiments performed with a single photon counting detector, like a cooled charge coupled device camera [21] or a pixel camera (*e.g.* the Maxipix [22] or the Pilatus [23]), the main noise encountered during the measurement is indeed the Poisson shot noise.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

The PDF given in Eq. (2) is standard in many applications dealing with low counting rates: for instance in transmission or emission tomography [24, 25] or in astronomy [26]. Although Poisson shot noise is sometimes used in the CDI community for testing algorithms [15, 27, 28], the noise model given in Eq. (2) has only recently been introduced in a phase retrieval algorithm [18, 29].

2.2. *Noise Model $\mathscr{G}$: stabilizing the variance of the counting process*

Even if one deals with counting statistics, it is often convenient to consider that the data are corrupted by an additive Gaussian (thermal) noise. Such a (standard) noise model is built with the following observation equation

$$y_{m,j}^{1/2} \;=\; h_{m,j}^{1/2} + \varepsilon_{m,j} \qquad \forall (m,j) \tag{3}$$

with $\varepsilon_{m,j}$ an independent centered fluctuation drawn from Gaussian random vector with constant variance σ^2, $\forall (m,j)$. Under these hypothesis, it is deduced that the PDF of the *transformed* data set $\left\{\sqrt{y_{m,j}}\right\}$ is also Gaussian and reads

$$f_{\mathscr{G}}(\boldsymbol{y};\rho) \;=\; \prod_j \prod_m (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2} \left[\frac{y_{m,j}^{1/2} - h_{m,j}^{1/2}}{\sigma} \right]^2 \right). \tag{4}$$

With this model, the transformed measurement $y_{m,j}^{1/2}$ has a standard deviation σ independent from its expected value $h_{m,j}^{1/2}$, while these two quantities should be linked for a photon counting process [30]. Therefore, it is clear that a model mismatch exists in the noise model $f_{\mathscr{G}}$. In practice, however, this Gaussian approximation works well. The proof is given by the presence of several ptychographic reconstruction algorithms in the literature [16, 31], which are related to this simple noise model, as shown in the section 3.2. This results from the fact that the square-root transformation applied to the photon noise is known as a "variance stabilization" transform that allows, in a first order approximation, the variance and the expected value of the transformed data to be independent parameters [32]. A proof of the variance stabilization of the photon noise by the square-root transform is provided in appendix A.

2.3. *Noise Model $\mathscr{Q}$: An approximation of the counting model*

Finally, the following observation equation is considered

$$y_{m,j} \;=\; h_{m,j} + \varepsilon_{m,j} \qquad \forall (m,j) \tag{5}$$

where $\varepsilon_{m,j}$ is an independent centred fluctuation drawn from Gaussian random vector with variance $\sigma_{m,j}^2$. As we are considering photon counting, the fluctuation variance $\sigma_{m,j}^2$ should be set to the unknown expected-value $h_{m,j}$. It leads to the following PDF for the data set $\boldsymbol{y}$

$$f(\boldsymbol{y};\rho) \;=\; \prod_{\substack{(m,j) \\ h_{m,j}\neq 0}} (2\pi h_{m,j})^{-1/2} \exp\left(-\frac{1}{2} \left[\frac{y_{m,j} - h_{m,j}}{h_{m,j}^{1/2}} \right]^2 \right). \tag{6}$$

Provided that the number of expected counts $\{h_{m,j}\}$ is "large enough", the *central limit theorem* (Ref. [20], Sec. 8.47) ensures that the Gaussian PDF in Eq. (6) is a good approximation of its Poissonian counterpart given in Eq. (2). Hence, from the ptychographic image reconstruction viewpoint, Eq. (6) is a fair noise model that could be used for the design of a reconstruction

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

algorithm. However, simpler iterative algorithms are readily obtained if the additional approximation $h_{j,m} \approx y_{m,j}$ is used for the variance so that the resulting noise model finally reads

$$f_{\mathscr{Q}}(\boldsymbol{y};\rho) \;=\; \prod_{\substack{(m,j)\\ y_{m,j}\neq 0}} (2\pi y_{m,j})^{-1/2} \exp\left(-\frac{1}{2}\left[\frac{y_{m,j}-h_{m,j}}{y_{m,j}^{1/2}} \right]^2 \right). \tag{7}$$

Note that data with no detected photon have been suppressed to avoid division by zero. Since the standard deviation depends on the data, this last noise model is no longer Gaussian. It is used for imaging reconstruction issues with photon noise in *e.g.*, Ref. [33–35].

3. Ptychographic image reconstruction by the maximum likelihood principle

The estimation of the unknown object ρ from a noisy data set is now introduced. Following the standard statistical inference literature, the so-called maximum likelihood (ML) principle can be used to estimate the object. It derives directly from the noise model. In the case of the ptychographical reconstruction problem, the ML estimator for ρ is the quantity that maximizes (with respect to ρ) the PDF of the chosen noise model. In more formal terms, this ML estimate reads

$$\rho_{\bullet} = \underset{\rho\in\mathbb{C}^N}{\arg\min}\, \mathscr{L}_{\bullet}(\rho) \tag{8}$$

where "$\bullet$" stands for $\mathscr{P}$, $\mathscr{G}$ or $\mathscr{Q}$ (*i.e.*, the noise model under consideration), and with

$$\mathscr{L}_{\bullet} := -\log f_{\bullet} \tag{9}$$

the neg-loglikelihood [36], which is a fitting function adapted to the noise model $f_{\mathscr{P}}$, $f_{\mathscr{G}}$ or $f_{\mathscr{Q}}$; for more details concerning the ML principle the reader is referred to *e.g.*, Ref. [20] (Chap. 18). For the noise models considered in this article, these fitting functions split as a sum over all the probe positions:

$$\mathscr{L}_{\bullet} = \sum_j \mathscr{L}_{\bullet;j} \tag{10a}$$

where $\mathscr{L}_{\bullet;j}$ is given by (up to irrelevant constant terms)

$$\mathscr{L}_{\mathscr{P};j}(\rho) := \sum_m h_{m,j}(\rho) - y_{m,j}\log\left[h_{m,j}(\rho)\right] \tag{10b}$$

$$\mathscr{L}_{\mathscr{G};j}(\rho) := \sum_m \left[y_{m,j}^{1/2} - h_{m,j}^{1/2}(\rho) \right]^2 \tag{10c}$$

$$\mathscr{L}_{\mathscr{Q};j}(\rho) := \frac{1}{2} \sum_{\substack{m\\ y_{m,j}\neq 0}} \left[\frac{y_{m,j}-h_{m,j}(\rho)}{y_{m,j}^{1/2}} \right]^2 \tag{10d}$$

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

where the dependencies with respect to (w.r.t.) the unknown object ρ are made explicit. From these expressions, one notes that the value $y_{m,j} = 0$ leads to a contribution $h_{m,j}(\rho)$ in the summands of both Eq. (10b) and Eq. (10c). As a result, the fitting functions $\mathscr{L}_{\mathscr{P}}$ and $\mathscr{L}_{\mathscr{G}}$ are equivalent w.r.t. the camera pixels that do not detect any photon. On the contrary, zero intensity camera pixels are discarded from $\mathscr{L}_{\mathscr{Q}}$ (Eq. (10d)) which is expected consequently to lead to very noisy solutions since these pixels are legitimate constraints for the final solution (see Sec. 4.5 for an example). This problem is clearly circumvented if Eq. (10d) is modified so that the empty pixels are accounted for, *i.e.*,

$$\mathscr{L}_{\mathscr{R};j}(\rho) := \mathscr{L}_{\mathscr{Q};j}(\rho) + \sum_{\substack{m \\ y_{m,j}=0}} h_{m,j}(\rho). \tag{10e}$$

The accuracy of this approximation [37] w.r.t. the Poissonian fitting function $\mathscr{L}_{\mathscr{P}}$ is studied in [33]. When the counting process is Poissonian, $\mathscr{L}_{\mathscr{P}}$ is expected to be the "best" fitting function since it is perfectly adapted to the data fluctuations. With photon noise, the ML estimator drawn from $\mathscr{L}_{\mathscr{P}}$ is attractive because it benefits from good asymptotic properties: for high counting rates, the ML estimator is free of systematic errors and presents the best variance estimation (Ref. [20], p. 56). For limited counting rates, however, the situation can be different and another fitting function may be more appropriate. We also stress that (by definition) the ML does not account for any additional *a priori* constraints concerning the electronic density to be retrieved (*e.g.*, support constraint, positivity). If the oversampling is too low and/or the number of diffraction patterns is limited (possibly equal to one), the ML may perform poorly and such additional constraints may be desirable (or even mandatory). This situation appears in support-based phase retrieval problems. However, since the present study aims at evaluating noise models for diffraction-pattern information only, the addition of object constraints has to be avoided because it may most probably blur the analysis. Hence, the ML is the appropriate tool to be considered. For sake of completeness, we also note that one can resort to the maximum *a posteriori* principle [38, p. 183] to introduce additional constraints within a statistical framework.

Finally, we note that the assumption $h_{m,j} > 0$, $\forall(m,j)$, is mandatory in order to ensure that $\mathscr{L}_{\mathscr{P}}$ given in Eq. (10b) is always defined. Indeed, the same condition results in the existence of the $\mathscr{L}_{\mathscr{P}}$ and $\mathscr{L}_{\mathscr{G}}$ gradients, allowing the iterative algorithms introduced in the next section to be defined. For the sake of simplicity, we assume in the following that the assumption $h_{m,j} > 0$ holds [39].

3.1. Computing the ML estimate

From a practical viewpoint, computing a solution defined by Eq. (8) requires an iterative algorithm in order to minimize one fitting function among the ones given by Eq. (10). This computation reduces to an unconstrained optimization problem, the aim being to find a solution that makes the gradient of $\mathscr{L}_{\bullet}$ vanish. As a result, gradient-based algorithms are natural candidates for the optimization of the chosen likelihood. The gradient of the likelihoods given in Eq. (10) reads

$$\partial_{\bullet} := \sum_{j} \partial_{\bullet;j} \tag{11a}$$

where the gradient for the j-th probe position $\partial_{\bullet;j}$ is given by

$$\partial_{\bullet;j}(\rho) = \boldsymbol{P}_j^{\dagger} \boldsymbol{W}^{\dagger} [\Psi_j(\rho) - \Psi_{\bullet;j}(\rho)] \tag{11b}$$

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

with "†" the conjugate-transpose operator, and where $\Psi_{\bullet;j} := \{\Psi_{\bullet;m,j}\}_{m=1}^{M}$ is the *corrected* far-field that depends on the chosen fitting function

$$\Psi_{\mathscr{P};m,j} := \Psi_{m,j} \times \frac{y_{m,j}}{h_{m,j}} \tag{12a}$$

$$\Psi_{\mathscr{G};m,j} := \Psi_{m,j} \times \frac{y_{m,j}^{1/2}}{h_{m,j}^{1/2}} \tag{12b}$$

$$\Psi_{\mathscr{Q};m,j} := \begin{cases} \Psi_{m,j} \times \left(2 - \dfrac{h_{m,j}}{y_{m,j}}\right) & \text{if } y_{m,j} \neq 0, \\ \Psi_{m,j} & \text{otherwise,} \end{cases} \tag{12c}$$

$$\Psi_{\mathscr{R};m,j} := \begin{cases} \Psi_{m,j} \times \left(2 - \dfrac{h_{m,j}}{y_{m,j}}\right) & \text{if } y_{m,j} \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{12d}$$

The functions given by Eq. (10) being *not* strictly convex, local minima may exist and can trap gradient algorithms. Moreover, it is well known that ambiguous solutions exist so that a unique ML cannot be defined for the ptychographical problem [16, 40]. From a computational viewpoint, the gradients given in Eq. (11) are the basic ingredients in the design of iterative reconstruction algorithms dedicated to the noise models. Two different classes of iterative algorithms are considered in the next subsections. In section 4.5 we also consider a hybrid algorithm that uses the best properties of both strategies.

3.2. Ordered-subset optimization strategies

Within the ptychographic experiments, the successive acquisition of intensity patterns for different but overlapping illumination positions on the sample naturally defines a partitioning in the data set. *Ordered-subset* (OS) algorithms [41] [44, 45] rest upon such a partitioning in order to update the object in a two nested loop process. Whereas the inner loop runs over the probe position $j = 1 \cdots J$, updating consecutively the illuminated portion of the object, one full iteration $k \to k+1$ occurs once the J probes are considered. Thus, for a given initial guess $\rho^{(0)}$, the algorithm is defined by the following updates for $k = 0, 1, \cdots$

$$\begin{aligned} & & \rho^{(k,0)} &:= \rho^{(k)} \\ j = 1, \cdots, J & & \rho^{(k,j)} &:= \rho^{(k,j-1)} - \beta\, \boldsymbol{D}_j \times \partial_{\bullet;j}\big(\rho^{(k;j-1)}\big) \\ & & \rho^{(k+1)} &:= \rho^{(k,J)} \end{aligned} \tag{13}$$

with $\boldsymbol{D}_j$ a diagonal scaling matrix and where $\beta > 0$ is the step-length. One may note that the classical *ptychographical iterative engine* (PIE) is a special case of this generic OS strategy. For instance, the choice

$$\partial_{\bullet;j} \equiv \partial_{\mathscr{G};j} \tag{14a}$$

$$\boldsymbol{D}_j = \big(1/\max_m |p_{m,j}|^2\big) \times \boldsymbol{I} \tag{14b}$$

(where $\boldsymbol{I}$ is the identity matrix) is precisely [46] the version of the PIE introduced in Ref. [15] for the object update, stressing that the PIE is a reconstruction algorithm relying (implicitly) on the noise model given in Sec. 2.2. Obviously, the choices $\partial_{\bullet;j} \equiv \partial_{\mathscr{P};j}$, $\partial_{\bullet;j} \equiv \partial_{\mathscr{Q};j}$ or $\partial_{\bullet;j} \equiv \partial_{\mathscr{R};j}$ provide natural extensions of the PIE algorithm to the noise models $f_{\mathscr{P}}$, $f_{\mathscr{Q}}$ and $f_{\mathscr{R}}$, respectively; such extensions are considered in Sec. 4.5.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

In practice, OS strategies (like the PIE) have appealing properties for several reasons. Firstly, image reconstructions from large data sets are made computationally more compact because each object update (in the nested loop) uses a subset of the data set; this means also that the field of view can be changed dynamically in the case of a real-time reconstruction. Secondly, the unique 3D propagation and evolution of the probe as it passes through a thick object at each position can be modelled and inverted easily [47]. Finally, these algorithms usually benefit from a fast convergence in the early iterations, hence providing an efficient means for the object estimate to "get into the right ball park" (see for instance Ref. [45]). However, some convergence issues exist for these algorithms. For instance, following [45], Godard *et al.* [28] show that the iteration (13) is not convergent to a local minimum of the fitting function $\mathscr{L}_\bullet$ (Eq. (10)) if the scaling matrix $\boldsymbol{D}_j$ depends on the probe position. Moreover, even when $\boldsymbol{D}_j$ is constant over the probe position (*i.e.*, if $\boldsymbol{D}_j \equiv \boldsymbol{D}, \forall j$), the authors find that the convergence toward a minimum of $\mathscr{L}_\bullet$ occurs *only* with noise-free data set. For noisy data, OS algorithms quickly find a relatively correct solution, but start to loop around after some iterations because the set of diffraction patterns is inconsistent, as a consequence of the presence of noise (see Sec. 4.4). Hence, at a given probe position, the algorithm "undoes" what it just did at the preceding probe position, reintroducing fully the noise within the associated diffraction pattern. In the next subsection, an iterative strategy that solves this convergence problem is considered.

3.3. Scaled-gradient optimization strategies

Given an initial guess $\rho^{(0)}$, the following *scaled-gradient* (SG) strategy is defined for $k = 0, 1, \cdots$

$$\rho^{(k+1)} := \rho^{(k)} - \beta \times \Lambda^{-1} \partial_\bullet\left(\rho^{(k)}\right) \qquad \beta > 0 \tag{15}$$

where the gradient $\partial_\bullet$ given by Eq. (11a) accounts for all the probes, and $\Lambda \in \mathbb{R}^{N\times N}$ is a diagonal scaling matrix chosen as

$$\Lambda := \textstyle\sum_j \boldsymbol{P}_j^\dagger \boldsymbol{P}_j + \alpha \qquad \alpha \geq 0. \tag{16}$$

As underlined in Ref. [48], the iteration (15) is a natural extension of the Error Reduction algorithm to the ptychographical approach. Since β and Λ are not dependent on the iteration number, the condition $\rho^{(k)} \to \rho^\infty$ implies $||\partial_\bullet\left(\rho^{(k)}\right)|| \to 0$: the convergence toward a limit point implies that this point is a local optimum of $\mathscr{L}_\bullet$. In practice, the step-length $\beta > 0$ is adjusted in order to generate a sequence converging toward a global or (at least) a local minimum of the fitting function. To our best knowledge, no result exists that gives admissible values for β ensuring the (local) convergence of Eq. (15). However, the tuning $\beta \approx 1$ was found to ensure convergence in most cases investigated in the present study. Indeed, provided that β is properly tuned, the SG iteration was always found to be a convergent algorithm.

For OS algorithms, the order in which the subsets are treated is often critical. This is clearly not the case with the SG strategy since the update (15) requires the full set of probe positions. The SG strategy is usually slower than the PIE in the early iterations because the latter performs J updates when the former performs only one. However, the SG strategy converges to a (local or global) minimum of $\mathscr{L}_\bullet$, even with a noisy data set. An illustration of these distinct convergence behaviors is given in Sec. 4.4.

4. Data inversion: a resolution *vs.* robustness trade-off

Clearly, the ML solution (Eq. (8)) is subject to fluctuations induced by the random nature of the measurements $\boldsymbol{y}$. Because four distinct fitting functions are discussed here, it is appropriate to search for the "best" model for the reconstruction purpose. This task requires to first define how the estimators will be compared.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

4.1. *Some quality indicators*

In the statistical literature, the accuracy of estimation is evaluated *via* two standard indicators, the *estimation bias* and the *estimation standard deviation*. Let $\langle\cdot\rangle$ be the expectation (*i.e.*, average over several realizations of the noise) operator, and $\rho_{\bullet;n}$ and $\rho_{\star;n}$ denote the n-th component of the ML solution $\rho_\bullet$ and the true object $\rho_\star$, respectively. Then, the estimation bias reads

$$\forall n, \qquad \mathrm{BIAS}_{\bullet;n} := \overline{\rho}_{\bullet;n} - \rho_{\star;n} \tag{17}$$

with $\overline{\rho}_{\bullet;n}$ the n-th component of the averaged solution

$$\forall n, \qquad \overline{\rho}_{\bullet;n} := \langle \rho_{\bullet;n} e^{\imath\phi_\bullet} \rangle \tag{18}$$

where

$$\phi_\bullet = \arg\left(\rho_\bullet^\dagger \rho_\star\right) \tag{19}$$

aims at compensating a global phase ambiguity. For complex random variables, the standard deviation of the estimation is defined by

$$\forall n, \qquad \mathrm{STD}_{\bullet;n} := \left\langle \left| \rho_{\bullet;n} e^{\imath\phi_\bullet} - \overline{\rho}_{\bullet;n} \right|^2 \right\rangle^{1/2} \quad . \tag{20}$$

Note that Eq. (17) and Eq. (20) are intuitive quality indicators for the (noise model dependent) estimator $\rho_\bullet$: whereas the bias (17) gives the systematic error, the variance (20) tells if the estimator is robust w.r.t. the noise. A third indicator is interesting to introduce: the *mean square error* (MSE)

$$\forall n, \qquad \mathrm{MSE}_{\bullet;n} := \left\langle \left| \rho_{\bullet;n} e^{\imath\phi_\bullet} - \rho_{\star;n} \right|^2 \right\rangle \tag{21}$$

$$= \mathrm{STD}^2_{\bullet;n} + |\mathrm{BIAS}_{\bullet;n}|^2 \tag{22}$$

that combines conveniently the preceding indicators. While a general closed-form expression for the bias and the standard deviation is not available, the computation of the averaged quantities given by Eq. (17) and Eq. (20) can however be achieved *via* Monte-Carlo simulations.

4.2. *Some implicit effects induced by the noise models*

This section aims at deriving typical features contained in the calculated solutions resulting from the choice of the noise model itself.

The *asymptotic* case of an arbitrary large signal to noise ratio (SNR) is first investigated: since we are dealing with photon noise, this results in $y_{m,j} \to h_{m,j}(\rho_\star)$. Consequently from Eq. (11), the gradient evaluated in $\rho_\star$ vanishes whatever the noise model is. In this context, the true object $\rho_\star$ minimizes the four fitting functions and the bias vanishes, *i.e.* the four noise models are equivalent. Therefore, consideration of the four noise models is only relevant at low SNR. In particular, as Eq. (12) gives the following relation

$$\Psi_{\mathscr{P};j} = \mathrm{Diag}\left(\frac{y_{m,j}^{1/2}}{h_{m,j}^{1/2}}\right) \times \Psi_{\mathscr{G};j} \tag{23}$$

between the corrected exit-field drawn from the models $\mathscr{P}$ and $\mathscr{G}$, it shows that the contribution in the final solution of a low SNR measurement [49] $y_{m,j} \sim 1$ is *enhanced* with the noise model $\mathscr{P}$ because its typical expected value is then $h_{m,j} < 1 \leq y_{m,j}$. Such measurements being spread over the borders of the intensity pattern, one expects that the noise model $\mathscr{P}$ enhances

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

the spatial resolution (*i.e.* reduces the bias) w.r.t. the noise model $\mathscr{G}$. However, this gain has necessarily a cost: because these low SNR measurements are plagued by large fluctuations, the model $\mathscr{P}$ should also lead to larger estimation variance. The *opposite* arguments holds for the noise model $\mathscr{R}$ since the condition $h_{m,j} \ll 1$ leads to

$$\Psi_{\mathscr{R};j,m} \simeq 2 \left(\frac{h_{m,j}^{1/2}}{y_{m,j}^{1/2}} \right) \times \Psi_{\mathscr{G};j,m} \tag{24}$$

i.e. the model $\mathscr{R}$ should lead to higher biases and to lower variances as compared to the noise model $\mathscr{G}$. In summary, we can see that the specific behavior of each model is dominated by the set of pixels that collects the lowest number of photons.

Finally, we also note that a photon noise, in the low SNR regime, produces very sparse intensity patterns. As these empty pixels are usually at the very edge, high-frequency components are missing in each intensity pattern and one can assume that the retrieved object is, more or less, a *low-pass filtered version* of the original object (with a loss in resolution being driven by the SNR). This result should hold whatever is the considered noise model.

4.3. A test-chart that highlights the predicted effects

To highlight these specific behaviors, a numerical test is now presented, which involves the evaluations of the estimation bias and standard deviation from Monte-Carlo simulations. The choice of the object is primarily motivated by its ability to illustrate the predicted "cut-off frequency" effect of each noise model explained above. For instance, a suitable object is a part of a Fresnel Zone Plate (see Fig. 1). The transmission coefficient of the object is set to one, while

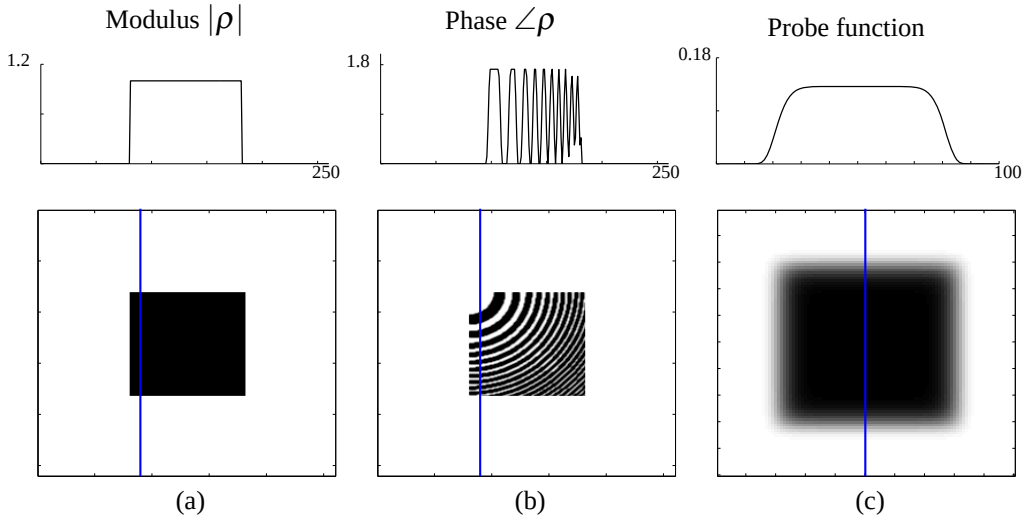


Fig. 1. The test object is a (support-limited) quadrant of a Fresnel zone-plate extending over 100×100 pixels within a 260×260 pixel image. The modulus (a) is 1 within the support of the object and the phase (b) ranges from 0 to 1.72 rad. The corresponding cross-sections are plotted along the 86-th column of the image. A real probe function (c) is chosen so that it extends over 58×58 pixels (full-width at half maximum) within a 100×100 pixel image, corresponding in an oversampling ratio of 1.7; the corresponding cross-section is plotted along the 50-th column of the image.

it vanishes outside the object window. The object extends over 100×100 pixels in a numerical

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

window of 260×260 pixels. The phase shift encountered by the beam is 1.72 radians and the radial frequency varies from 0.07 to 0.3 pixel^{-1}. The ptychographical data-set is composed of a total of 81 diffraction patterns, each one of size 100×100 pixels. The choice of a step size of about 20 pixels in both directions leads to an overlap ratio of 65%. In addition, two SNRs are considered in these simulations: the highest SNR provides a maximum of 10^6 expected counts over the 2D camera; the lowest provides 10^3 expected photons over the camera.

The Monte-Carlo analysis presented below is based on a set of 100 noisy (photon noise) ptychographic data sets.

4.4. Some issues concerning the iterative strategy

It is clear from Sec. 3 that distinct iterative strategies can be derived for the minimization of the *same* fitting function. It is therefore appropriate to investigate the impact of the iterative strategy on the retrieved object. For that purpose, the inversion of a single ptychographic data set by either the OS or the SG strategy is now considered. For sake of simplicity, the fitting function $\mathscr{L}_{\mathscr{G}}$ is considered, but similar results are obtained with the other fitting functions.

When a noise-free data set is considered, Fig. 2(b) shows that the OS strategy converges toward a minimum of the fitting function since the gradient norm decreases toward zero. With noisy data (*i.e.*, corrupted by photon noise) however, the gradient norm starts to decrease before it reaches a stagnation, such that the convergence does not occur. Furthermore, Fig. 2(b, c) shows that the OS strategy should be stopped early in the iteration process [50] in order to pick the best solution w.r.t. the *relative error* (in the object plane) defined by

$$\mathrm{Err}(\rho_k) \;:=\; \frac{||\rho_\star - \rho_k e^{\imath\phi_k}||}{||\rho_\star||} \qquad (\text{with } \imath = \sqrt{-1}) \tag{25}$$

where $\rho_\star$ is the true object and where

$$\phi_k = \arg \rho_k^\dagger \rho_\star. \tag{26}$$

When the SG strategy is considered, Fig. 2(a, b) shows that the iterations converge toward a (local or global) minimum of $\mathscr{L}_\bullet$, even with a noisy data set. This minimum defines an estimate which is a *global* trade-off over the set of inconsistent diffraction patterns, leading to a lower relative error than the best relative error reached by the OS strategy (see Fig. 2(c) and the reconstructions shown in Fig. 2(d, e, f)).

4.5. Monte-Carlo analysis

Investigation of the figure of merit for each noise model In an attempt to define the *intrinsic* merit of each noise-model, the impact of the minimization method has to be as low as possible. In other words, the minima of $\mathscr{L}_\bullet$ can be only compared w.r.t. the noise models if one ensures that the reconstruction quality is not affected by the way the data are handled along the iterative process. Therefore, the use of the (convergent) SG strategy is mandatory, with the additional condition of an initialization as close as possible to the minima. For that purpose, the true solution is chosen as initial guess, *i.e.*, $\rho^{(0)} = \rho_\star$. The algorithms are stopped when the norm of the gradient reaches a conveniently small value.

For each fitting function, and for the lowest SNR, the numerical evaluation of the averaged solution $\overline{\rho}_\bullet$ (both modulus and phase) given in Eq. (18) and the standard deviation given in Eq. (20) are provided in Fig. 3 and Fig. 4, respectively. The predicted cut-off frequency effect is clearly visible in Fig. 3: for $\mathscr{L}_{\mathscr{R}}$, the edges of the modulus are smoothed and the phase is damped whereas they remain much more resolved (undamped) for $\mathscr{L}_{\mathscr{P}}$. For this low SNR, the

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

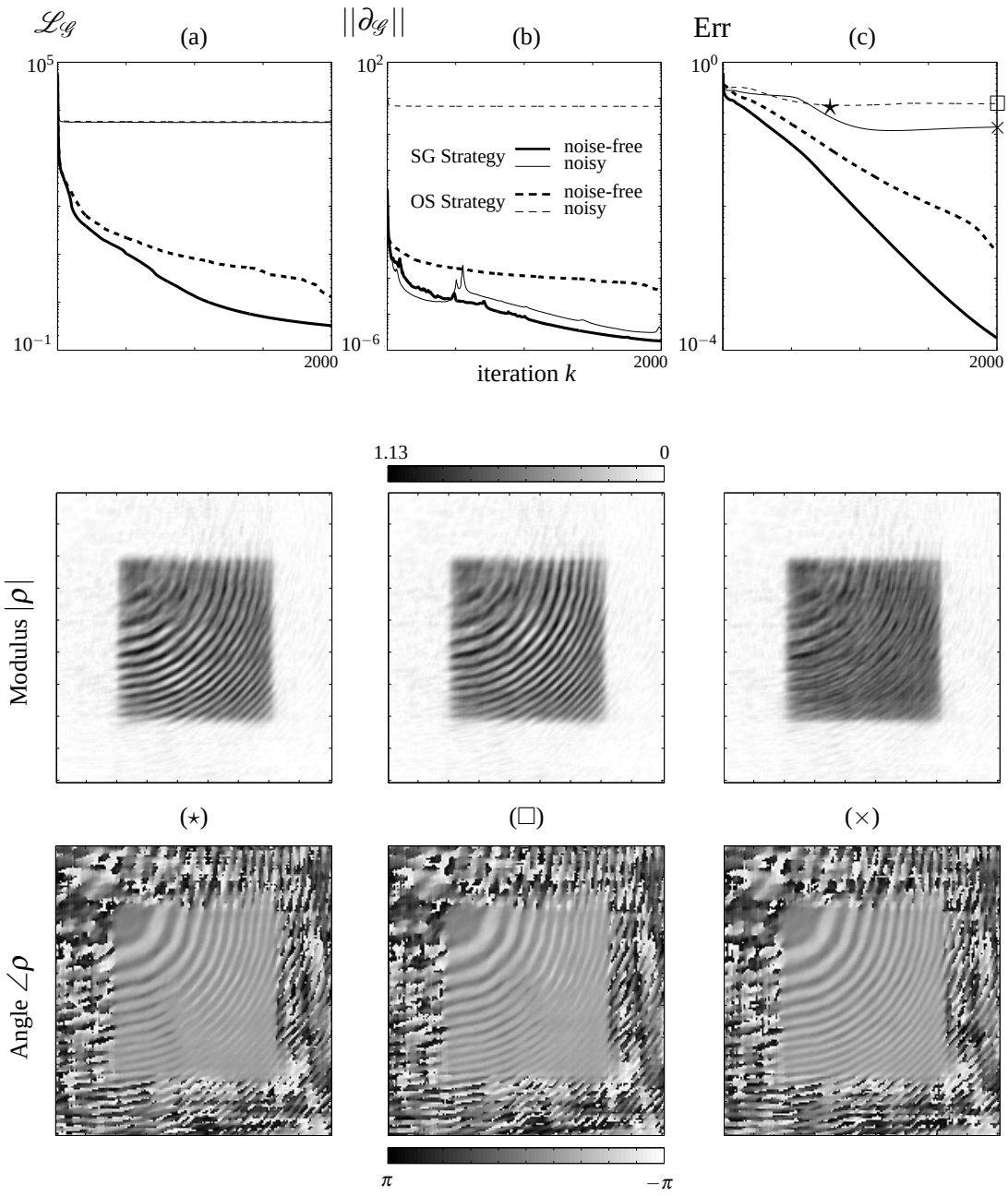


Fig. 2. A ptychographical reconstruction illustrates the convergence behavior of the OS and the SG strategies. In these examples, the fitting function is $\mathscr{L}_{\mathscr{G}}$ given in (10c), such that the OS strategy corresponds to the standard PIE. Top line: for the OS (dashed line) and the SG (solid line), (a) evolution of the fitting function $\mathscr{L}_{\mathscr{G}}$ w.r.t. the iteration k for a noise-free data set (thick line) and an example of a noisy data set (thin line) with a maximum of 10^3 photons on the camera; (b) *idem* for the gradient norm $||\partial_{\mathscr{G}}(\rho^{(k)})||$; (c) *idem* for the error $\mathrm{Err}(\rho^{(k)})$ defined by (25). Second and third lines: reconstruction from a noisy data set; with the OS strategy, ($\star$) is the estimate that minimizes the error depicted in (c) and ($\square$) is the estimate obtained after $k=2000$; with the SG strategy, ($\times$) is the estimate after $k=2000$. The shown results correspond to a 180×180 window centered around the object central pixel. The respective color scales are indicated on the figure.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

modulus is contaminated by fluctuations that come from the object phase function. The relative amplitude of these fluctuations is 8, 10, 17 and 10 % for $\mathscr{L}_{\mathscr{P}}$, $\mathscr{L}_{\mathscr{G}}$, $\mathscr{L}_{\mathscr{Q}}$ and $\mathscr{L}_{\mathscr{R}}$, respectively. They reduce when the SNR increases and they become negligible (around 1%) for 10^6 photons. As explained in Sec. 4.2, such artifacts appear because the retrieved object is essentially a low-pass filtered version of the original object (see Fig. 5). In the case of the present object, it results from a mixing between the real and imaginary parts of the object, leading to the observation of a phase-like structure in the modulus component.

Finally, the standard deviation depicted in Fig. 4 confirms that $\mathscr{L}_{\mathscr{R}}$ has the highest robustness w.r.t. the photon noise whereas $\mathscr{L}_{\mathscr{P}}$ has the lowest. For all the fitting functions, the standard deviation grows with the collected number of photons, which is a standard result when one deals with photon noise (Ref. [51], p. 181). As expected, the fitting function $\mathscr{L}_{\mathscr{G}}$ reaches a trade-off w.r.t. these two fitting functions, as expected from Sec. 4.2. Quantitatively, the numerical evaluation of the quality indicators defined in Sec. 4.1 are reported in Table 1. For both SNRs, although the variance is lower with $\mathscr{L}_{\mathscr{G}}$ or $\mathscr{L}_{\mathscr{R}}$, one notes that $\mathscr{L}_{\mathscr{P}}$ gives however the best results w.r.t. the MSE and the error in the object plane. The fitting function $\mathscr{L}_{\mathscr{Q}}$ being much less robust to the noise than the other three fitting functions (see Sec. 3), it is not considered as a valuable alternative for CDI in the low counting rate regime.

Table 1. Figure of merit of each noise model. The l^2-norms of the bias, standard deviation (STD) and mean-square error (MSE) as well as the error (Err) in the object plane for the fitting functions defined in section 2 are given. The SG strategy is used with the true object as initial guess.

SG optimization					
SNR	$\mathscr{L}_\bullet$	$\|\|\mathbf{BIAS}_\bullet\|\|$	$\|\|\mathbf{STD}_\bullet\|\|$	$\|\|\mathbf{MSE}_\bullet\|\|$	$\mathrm{Err}_\bullet$
10^6 photons	$\mathscr{P}$	1.5	3.3	0.1	2.2
	$\mathscr{G}$	3.2	2.4	0.3	10
	$\mathscr{R}$	4.1	2.4	0.4	17
	$\mathscr{Q}$	3.3	18	2.0	11
10^3 photons	$\mathscr{P}$	21	31	12	430
	$\mathscr{G}$	31	16	14	940
	$\mathscr{R}$	37	14	18	1 400
	$\mathscr{Q}$	50	290	490	2 500

Starting from a coarse initial guess In practice, the chosen initial guess is often a rough estimate and the iterative strategy adds its own bias and variance. For this reason, it is appropriate to investigate how the reconstruction quality deteriorates when one uses a coarse initialization with either the OS or the SG strategy. We further assume that no *a priori* object information can be used, resulting in the choice of a free-space estimate for the initial guess. The quality indicators achieved with 10^3 photons are reported in the Table 2; as the OS strategy does not lead to converging iterations (see Sec. 3.2), the iteration that gives the best (*i.e.*, the smallest) error in the object plane is selected for each data set.

To summarize, the behaviors exhibited in the preceding section are still valid here: the fitting function $\mathscr{L}_{\mathscr{P}}$ offers the lowest bias but the worst variance while the fitting function $\mathscr{L}_{\mathscr{R}}$ has the opposite characteristics. Moreover, every criteria is improved when using the SG strategy, the gain being more clearly evidenced with the noise model $\mathscr{P}$. For the sake of completeness, the difference-map (DM) iteration [52] for ptychographic image reconstruction, as described in

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

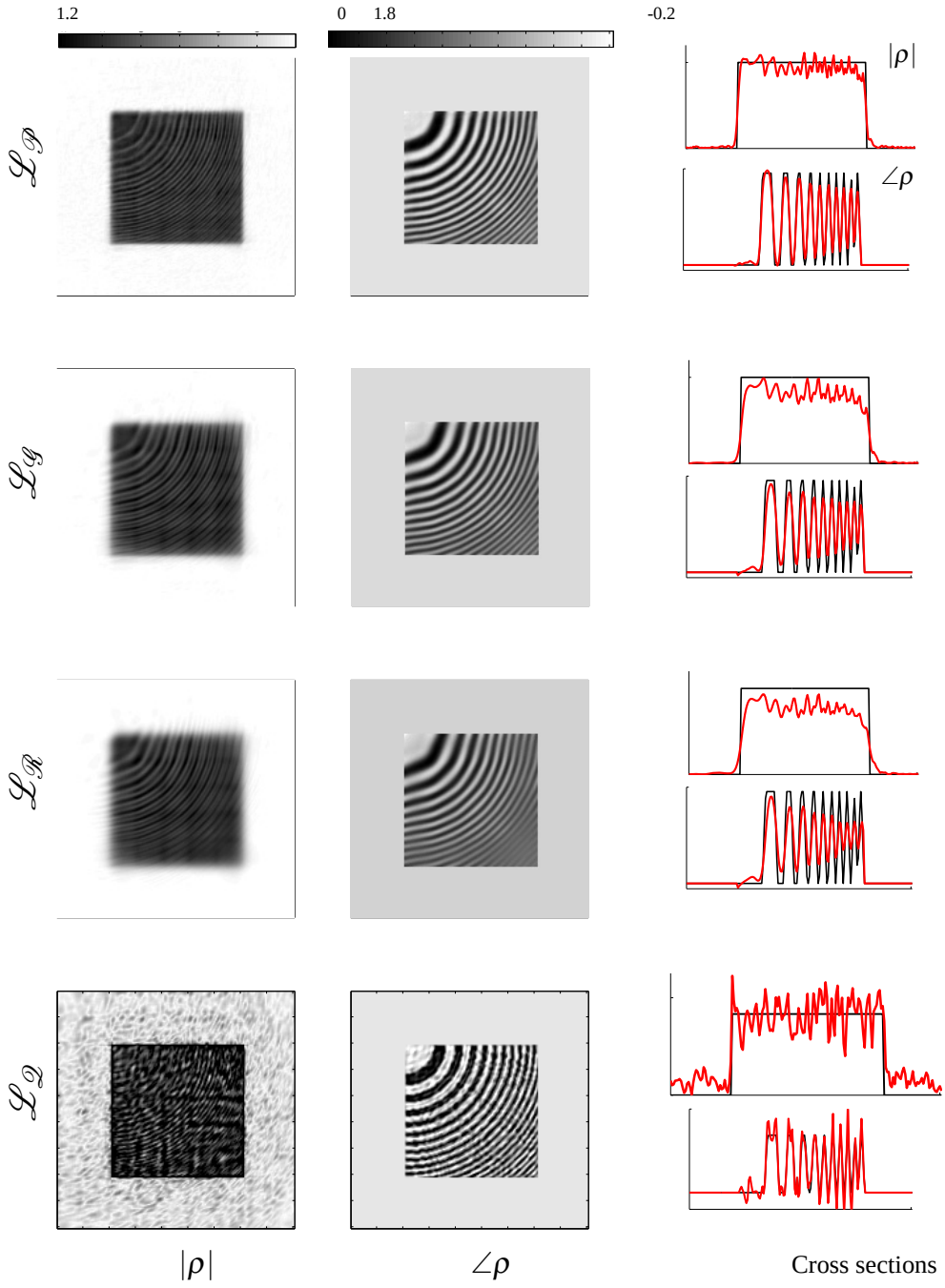


Fig. 3. The average solution for each noise model evaluated over a series of 100 noisy data sets. The initial guess is the true object and the SG strategy is used for the optimization of the fitting function. For each noisy data set, no more than 10^3 photons impinge on the detector. The grey level scaling in each column shares the same linear scale. The shown results correspond to a 180×180 window centered around the object central pixel.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

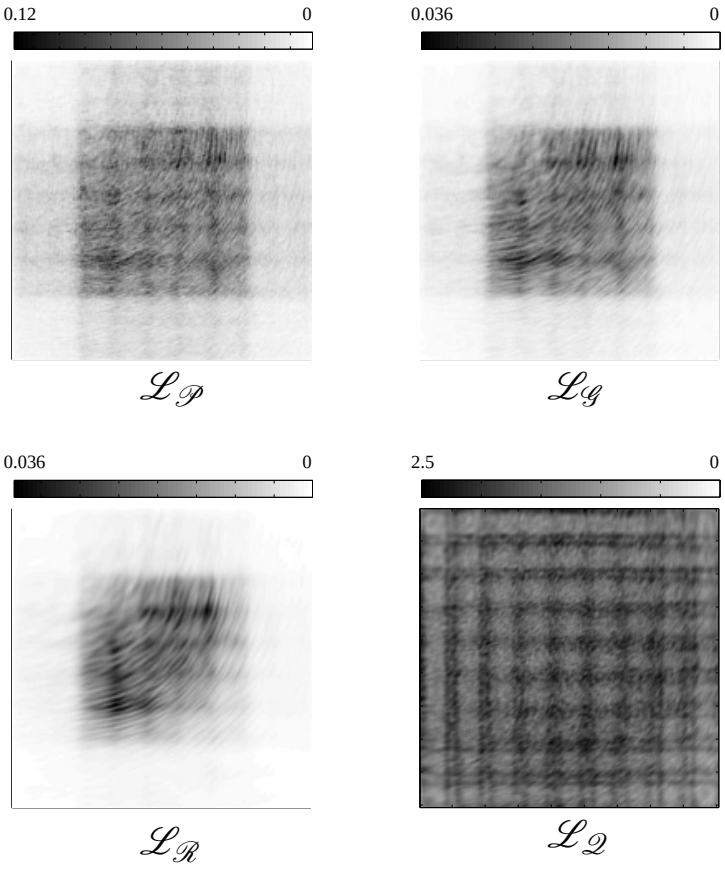


Fig. 4. Evaluation of the standard deviation for each noise model computed over a series of 100 noisy data sets. The initial guess is the true object and the SG strategy is used for the optimization of the fitting function. For each noisy data set, no more than 10^3 photons impinge on the camera area.

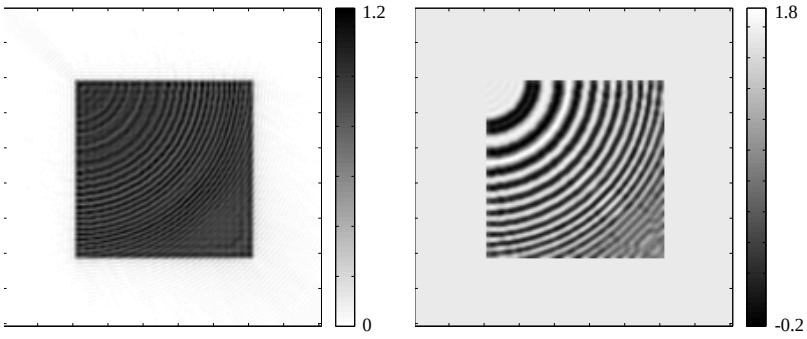


Fig. 5. The modulus (left) and the phase (right) that are obtained when high-frequency components are filtered out of the original object shown in Fig. 1.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

Ref. [53] is also implemented and compared. In all the tests performed, the DM iteration and the OS strategy with $\mathscr{L}_{\mathscr{G}}$ perform equivalently [54]. In Fig. 6, the results of the SG strategy obtained from a single noisy data set is shown for the various fitting functions; these illustrate how a typical reconstruction looks like for each noise model when the SG strategy is used. Finally, the algorithms presented in this work have been tested on several object classes: phase

Table 2. The l^2-norms of the bias, STD and MSE as well as Err in the object plane achieved by the fitting functions $\mathscr{L}_{\mathscr{P}}$, $\mathscr{L}_{\mathscr{G}}$ and $\mathscr{L}_{\mathscr{R}}$ when either the SG or the OS strategy is used with a free-space as initial guess. The results achieved by the DM and the hybrid method are also presented.

SNR = 10^3 photons					
Iterative strategy	$\mathscr{L}_\bullet$	$\|\|\mathbf{BIAS}_\bullet\|\|$	$\|\|\mathbf{STD}_\bullet\|\|$	$\|\|\mathbf{MSE}_\bullet\|\|$	$\mathrm{Err}_\bullet$
SG strategy	$\mathscr{P}$	26	41	21	680
	$\mathscr{G}$	32	20	15	1000
	$\mathscr{R}$	37	14	18	1400
OS strategy	$\mathscr{P}$	29	48	28	860
	$\mathscr{G}$	32	22	17	1100
	$\mathscr{R}$	37	14	18	1400
Difference Map	—	32	21	16	960
Hybrid	$\mathscr{P}$	22	32	13	470

objects, absorption objects, objects with low or high contrasts, *etc*. It is always the case that the Poissonian noise model $\mathscr{P}$ presents the least systematic errors, whereas the noise model $\mathscr{R}$ is the most robust, the Gaussian model $\mathscr{G}$ reaching a trade-off between the two others. It is also observed that the differences between all these algorithms tend to vanish when the SNR increases.

The minimization of $\mathscr{L}_{\mathscr{P}}$: a hybrid optimization strategy. It is clear from Tables 1 and 2 that $\mathscr{L}_{\mathscr{P}}$ is the fitting function that undergoes the strongest degradation if the initialization is far from the final solution. On the contrary, the minimization of $\mathscr{L}_{\mathscr{R}}$ is very robust w.r.t. the starting point. Moreover, the OS and the SG strategies are mostly equivalent for that fitting function. Hence, it is appropriate to search for a hybrid strategy that profits from both fitting functions. Therefore, we propose to use the OS strategy starting with the fitting function $\mathscr{L}_{\mathscr{R}}$ in order to get quickly to a first estimate which is subsequently introduced as an initial guess for the further minimization of the fitting function $\mathscr{L}_{\mathscr{P}}$. As an example, one can perform 1000 OS iterations with $\mathscr{L}_{\mathscr{R}}$ followed by 1000 SG iterations with the fitting function $\mathscr{L}_{\mathscr{P}}$. The quality indicators obtained with this strategy are shown in Table 2. One notes that these indicators are improved: they reach values similar to the ones obtained with the true object (see Table 1). Figure 7 also shows the reconstructed phase obtained by either the "hybrid" strategy or by the SG strategy with the true object as an initial guess. These phases are similar, showing that the hybrid strategy is a valuable technique for the optimization of the Poissonian fitting function.

5. Conclusion

In summary, we have addressed the question of the choice of the iterative process for coherent diffraction imaging in the case of data corrupted by noise. Several noise models compatible with photon (or electron) shot noise have been presented and further used within two inversion

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

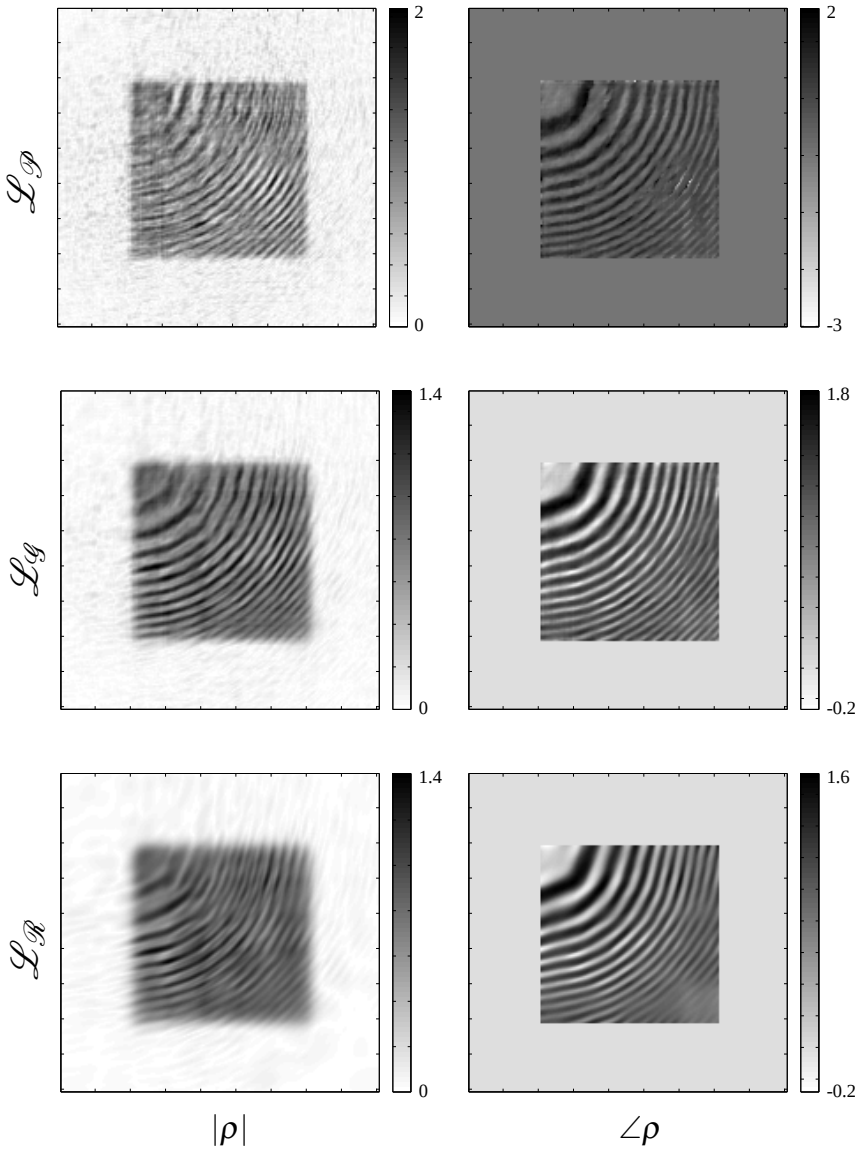


Fig. 6. Object reconstruction by means of the minimization of the fitting functions $\mathscr{L}_{\mathscr{P}}$, $\mathscr{L}_{\mathscr{G}}$ and $\mathscr{L}_{\mathscr{R}}$ given in Eq. (10). The optimization is performed with the SG strategy presented in Sec. 3.3 and an initial guess defined as an uniform object. The phase is set to zero outside of the object support for visualization purpose.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

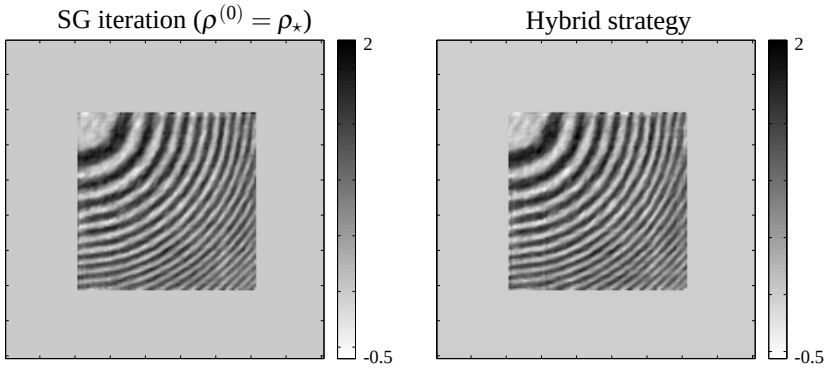


Fig. 7. The retrieved phase obtained by the minimization of $\mathscr{L}_{\mathscr{P}}$ with, either the hybrid strategy (right), or the SG strategy with the true object as initial guess (left).

strategies, the OS and the SG. We have shown that any physical interpretation drawn from a CDI iterative technique requires a detailed understanding of this iterative technique. Our analysis emphasizes that iterative reconstruction algorithms for CDI often assume implicitly a noise model that may be more or less a coarse approximation of the data fluctuations. While standard asymptotic results for photon noise foresee that high SNR measurements should be handled in the same way by any model, each model has the ability to enhance or inhibit the weight of low SNR measurements in the final reconstruction. From this viewpoint, the noise models presented in this paper reach their own resolution *vs.* robustness trade-off. The merit of each noise model may be user and/or object dependent and, from an experimental perspective, the impact of the intensity fluctuations w.r.t. the noise model has to be tested on numerical samples prior to the inversion. An efficient strategy to circumvent the problem in the case of experimental intensity analysis consists in building a set of data for a model sample, designed as close as possible to the available experimental data set (Fourier space resolution, number of probe positions, SNR, *etc*.). This numerical data set can then be used to test the different noise model approaches and emphasizes the possible reconstruction artifacts.

Whereas it is not a surprise that in presence of shot noise the initial object guess has a strong impact on the final solution obtained with CDI, the employed optimization strategy (OS or SG) generates its own artifacts. Clearly, algorithms that reach the minimum of the fitting function defined by the noise model should be used. On the contrary, if non-converging algorithms are employed, some additional reconstruction degradation is expected. Finally, based on this detailed study, a hybrid strategy has been presented that improves the convergence towards the minimum of the Poissonian fitting function when a good initial guess is missing.

The ML principle adopted in this work does not rely on any prior model related to the unknown object. However, when such information is available, such models provide additional constraints that may enhance the resolution and the robustness of the reconstructions. In this context, the maximum *a posteriori* would be the natural extension of the ML principle when prior models are accessible. It would lead to a *penalized* fitting function as discussed elsewhere in *e.g.,* Ref. [9, 18]. Projective algorithms (like ER or HIO) are also natural means to handle additional constraints concerning the unknown object. Another interesting perspective consists in the adaptation of these standard algorithms in order to cope with the various noise models presented in this article.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA

A. The variance stabilization transform

Let y be a random variable with mean $\langle y\rangle = \mu$ and variance $\mathrm{VAR}(y) = \sigma^2$, and suppose that σ and μ are related by $\sigma = f(\mu)$ for some function f. A variance-stabilization transform aims at constructing a function h such that the random variable $h(y)$ has an almost constant variance, without loosing any information (*i.e.*, h has to be injective in the range of y).

The Taylor expansion of h around μ in the first order is

$$h(y) - h(\mu) = (y-\mu)h'(\mu) + R \tag{27}$$

where R stands for higher order terms. One then has

$$\begin{aligned}\mathrm{VAR}\big(h(y)\big) &= \mathrm{VAR}\big(h(y) - h(\mu)\big)\\ &= \mathrm{VAR}\big((y-\mu)\,h'(\mu) + R\big)\\ &= \mathrm{VAR}\big((y-\mu)\,h'(\mu)\big) + \mathrm{VAR}(R) + 2\langle\big((y-\mu)\,h'(\mu)\,R\big)\rangle\end{aligned} \tag{28}$$

where $\langle\big((y-\mu)\,h'(\mu)\big)\rangle = 0$ is used in the last line. Neglecting all the contributions from the terms higher than the first order gives:

$$\begin{aligned}\mathrm{VAR}\big(h(y)\big) &\sim \big(h'(\mu)\big)^2\,\mathrm{VAR}(y-\mu)\\ &= \big(h'(\mu)\big)^2\,\mathrm{VAR}(y)\\ &= \big(h'(\mu)\,f(\mu)\big)^2.\end{aligned} \tag{29}$$

Thus, within the first order approximation, the variance of $h(y)$ is independent of μ if a function h is exhibited such that $h'(x)\big|_{x=\mu} = b/f(\mu)$ for a constant b in $\mathbb{R}$. The obvious candidate $h(x) = bx/f(\mu)$ is of no interest, being a linear function. A suitable choice for the Poissonian case in which $f(x) = \sqrt{x}$ is the function $h(x) = \sqrt{x}$; we then find $b = 1/4$. This is the variance stabilization used in the Section 2.2. Anscombe, in [55], showed that the function $h(x) = (x + 3/8)^{1/2}$ has a better variance-stabilization capability than the square-root transform.

Acknowledgments

We are grateful to Ph. Réfrégier for fruitful discussions. This work was partially funded by the French ANR under project number ANR-11-BS10-0005. The postdoctoral fellowship of P. Godard in Sheffield was funded by the EPSRC Basic Technology Grant No. EP/E034055/1; 'Ultimate Microscopy: Wavelength-Limited Resolution Without High Quality Lenses'.

#173585 - $15.00 USD Received 31 Jul 2012; revised 11 Sep 2012; accepted 11 Sep 2012; published 1 Nov 2012
(C) 2012 OSA