Compressed sensing for phase retrieval

Marcus C. Newton*

London Centre for Nanotechnology, University College London, United Kingdom and
Present address, Laboratory of Coherent X-Ray Optics, Research Institute for Electronic Science, Hokkaido University, Sapporo, Japan

(Rceived 20 February 2012; published 22 May 2012)

To date there are several iterative techniques that enjoy moderate success when reconstructing phase information, where only intensity measurements are made. There remains, however, a number of cases in which conventional approaches are unsuccessful. In the last decade, the theory of compressed sensing has emerged and provides a route to solving convex optimisation problems exactly via $\ell_1$-norm minimization. Here the application of compressed sensing to phase retrieval in a nonconvex setting is reported. An algorithm is presented that applies reweighted $\ell_1$-norm minimization to yield accurate reconstruction where conventional methods fail.

DOI: 10.1103/PhysRevE.85.056706 PACS number(s): 02.70.—c, 87.59.—e, 61.43.Bn, 71.15.Dx

I. INTRODUCTION

The problem of recovering a sparse signal by $\ell_1$-norm minimization has received significant attention in recent years as applications continue to expand [1,2]. This renewed interest in the $\ell_1$ norm can be attributed to the emergent field of compressive sensing (CS). The theory of compressed sensing dictates that an appropriately subsampled (or compressed) signal can be recovered exactly through various routes to minimizing the $\ell_1$ norm of that signal. This method is so named compressed sensing as the signal is compressed in a recoverable manner on acquisition. Practical solutions can be obtained by using the techniques of linear programming. This approach has proven effective in solving a broad range of convex optimisation problems and is robust against measurement noise [3,4].

When performing spectral acquisition measurements, phase information is in general lost as only intensity measurements can be made. This is, for example, a routine occurrence in x-ray microscopy [5]. It therefore becomes necessary to reconstruct the phase information subject to some a priori information. In general phase retrieval methods operate by the application of iterative projective algorithms that cause the objective function to converge toward the optimal solution by the repeated application of constraints at each iteration [6,7].

One important application of phase retrieval methods is to obtain real-space images of nanoscale structures by inverting coherent x-ray diffraction (CXD) measurement data [5]. In CXD imaging, phase information is obtained by alternating between real and reciprocal space while applying constraints to the objective function at each turn. In real space, a support or solvent constraint is applied that defines a subset of the function space for which the object should exist. The reciprocal space constraint ensures that the modulus of the objective function equals that of the measured diffraction amplitude, while the phase remains unchanged. A solution is found when both conditions are reasonably well satisfied. While this approach works well for the majority of cases, some limitations persist. Namely, that the support constraint can place a restriction on the ability to find an iterative path to the true global solution. Also, to date, $\ell_1$-norm minimization methods have limited application as they are generally based on linear programming and are effective for convex optimization problems but are not readily applied to the phase retrieval problem. This is largely due to the Fourier space projection onto the modulus set being nonconvex.

In this study, we report on an iterative algorithm based on compressed sensing that is able to reconstruct phase information in a robust manner where other current methods fail. Moreover, we show that our approach achieves reconstruction that closely resembles the global minimum of the initial object. We also demonstrate application to the reconstruction of data from highly strained crystals [8]. Our results are significant as they demonstrate the first direct application of compressed sensing to the phase retrieval problem.

II. MAIN ALGORITHM

A. Iterative phase retrieval

When intensity measurements are performed phase information is lost and in general we obtain the reciprocal space intensity distribution $I(k) = |\hat{\rho}(k)|^2 = |\hat{\rho}(k)e^{i\phi(k)}|^2$, where $\hat{\rho}$ is the complex density. Iterative reconstruction algorithms then proceed by traversing back and forth between direct and Fourier space while applying a constraint at each turn. A successful approach is to define a subset or support region $S$ in direct space for which the amplitude of the object density is unrestricted. Outside this region the amplitude of the density $\rho(r)$ is minimized in some way. The Fourier space constraint is described by the operator $F|\rangle$ and requires the object amplitude to be proportional in some way to the original measurement on some set $M$ such that $P|\rangle|\hat{\rho}(k)e^{i\phi(k)}\rangle = |\hat{\rho}(k)e^{i\phi(k)}\rangle$, $\forall k \in M$. A well-known example is the hybrid input-output (HIO) algorithm which can be written as

$$\rho^{(n+1)}(r) = \begin{cases} P_M \rho^{(n)}(r), & \forall r \in S, \\ (I - \beta P_M) \circ \rho^{(n)}(r), & \forall r \notin S, \end{cases}$$

where $P_M = F^{-1}P|\rangle F$ is the modulus projection operator, $\beta \in [0,1]$ is a relaxation parameter, and $I$ is the identity. The direct and inverse Fourier transform are represented by $F$ and $F^{-1}$, respectively.

Fienup’s HIO algorithm and its generalizations have enjoyed significant application in numerous fields including...
protein crystallography, diffraction microscopy, sparse-aperture imagery, and x-ray astronomy [7,9–13].

B. Compressed sensing phase retrieval

The main results of the compressed sensing theory dictate that it is possible to reconstruct a signal exactly from fewer samples than Shannon’s sampling theory permits by exploiting the signal’s inherent sparsity. Robust strategies for achieving this are based on convex linear optimization that minimize the $\ell_1$ norm to promote sparsity. For the case of phase retrieval such direct techniques do not translate well as the modulus projection operator $P_M$ is nonconvex. An alternative approach, developed here, employs an iterative reweighting (IR) scheme to promote sparsity in the reconstruction. IR algorithms can be classed as a subset of majorization-minimization (MM) algorithms [2,14]. Candes et al. have recently shown that reweighted $\ell_1$-norm minimization can in many cases outperform its unweighted counterpart [15]. Early IR schemes have focused on least-squares reweighting to solve underdetermined systems. A well-known example is that of Lawson’s algorithm (and its various extensions) for uniform approximation via $\ell_p$, $p \in [2,\infty]$ Tchebyscheff approximations [16]. Here we apply IR $\ell_p$, $p \in [0,1]$ to the nonconvex problem. In doing so, we define the following error metric:

$$
\epsilon_w = \|\Gamma\|_p, \quad \Gamma = W|\rho^{(n)}(r)|^2 - |\rho^{(n)}(r)|, \\
W = [(|\rho^{(n-1)}(r)|^2 + \epsilon)^{-\frac{1}{2}},
$$

where $\epsilon$ is a positive relaxation parameter and $W$ are real positive weights. The rationale here is to minimize the $p$-norm sparsity measure between the current and previous iterations. This was found to promote sparsity in the amplitude distribution of the direct space density $|\rho(r)|$. We also note that a narrow or sparse density amplitude distribution is commensurate with Bragg diffraction from a compact periodic crystalline material. Values of $W$ that deviate significantly from the inverse norm $[(|\rho^{(n)}(r)|^{-1})]$ will cause those elements to have a lesser influence in the current reconstruction step. Constructing weights in this way at each iteration allows for successive improvement in agreement with the current and previous iterate as large changes in $|\rho(r)|$ are penalized. As a result, earlier iterations may lead to inaccurate signal estimates while at later iterations there is improved sensitivity to determining smaller nonzero coefficients of the signal.

Following Fienup [6,17], we define also the modulus $(\epsilon_m)$ and support $(\epsilon_s)$ error metrics

$$
\epsilon_m[\rho(r)] = \|P_M\rho(r) - \rho(r)\|_2, \\
\epsilon_s[\rho(r)] = \|P_S\rho(r) - \rho(r)\|_2.
$$

In the above, $P_S$ is the support projection operator which equals the identity $I$ for all $r \in S$ and is zero otherwise. We then seek to minimize the total error by constructing the following Lagrangian:

$$
\mathcal{L}[\rho(r)] = \epsilon_m^2[\rho(r)] + \epsilon_s^2[\rho(r)] - \epsilon_p[\rho(r)].
$$

We are then able to derive the step length $\Delta \rho^{(n)}(r)$ to the next iterate $\rho^{(n+1)}(r)$ as

$$
\Delta \rho^{(n)}(r) = \frac{1}{2} \nabla_\rho \mathcal{L}[\rho(r)] = (P_S - P_M + P_w) \circ \rho(r),
$$

where we have taken the functional derivative of Eq. (2), yielding

$$
\nabla_\rho \epsilon_m^2[\rho(r)] = \frac{\Gamma}{|\Gamma|^2 - p} \nabla_\rho \epsilon_m \equiv 2P_w \rho(r)
$$

and we have utilized the following results [6,17]:

$$
\nabla_\rho \epsilon_s^2[\rho(r)] = -2(P_M - I) \circ \rho(r), \\
\nabla_\rho \epsilon_p[\rho(r)] = -2(P_S - I) \circ \rho(r).
$$

Introducing the relaxation term [18] $[\beta(I - P_S) - P_S]$ with parameter $\beta \in [0,1]$ into Eq. (6) we arrive at the following for fixed point evaluation:

$$
\rho^{(n+1)}(r) = \rho^{(n)}(r) + \Delta \rho^{(n)}(r) = (P_S - P_M - P_w) \circ \rho^{(n)}(r),
$$

which can be written as

$$
\rho^{(n+1)}(r) = \begin{cases} 
(P_M - P_w) \circ \rho^{(n)}(r), & \forall r \in S, \\
[I - \beta(P_M - P_S)] \circ \rho^{(n)}(r), & \forall r \notin S.
\end{cases}
$$

In the absence of the operator $P_w$, Eq. (11) reduces to the HIO algorithm of Eq. (1). We therefore denote this as the compressed sensing HIO (CSHIO) algorithm. One immediate and distinctive feature of this algorithm is that the current ($n$) and previous ($n - 1$) iterate are required to calculate the next ($n + 1$) iterate. This is a direct consequence of the reweighting scheme as $P_w$ is a function of both $\rho^{(n)}(r)$ and $\rho^{(n-1)}(r)$. Using pseudocode, Fig. 1 illustrates how one might implement the CSHIO algorithm. We now show how this algorithm is effective in reconstructing phase information and able to overcome limitations of other methods.

III. CSHIO PHASE RECONSTRUCTION

To demonstrate the effectiveness of the CSHIO algorithm, numerical experiments were performed where simulated data from coherent x-ray diffraction (CXD) measurements on highly strained material was reconstructed using the CSHIO algorithm. In CXD measurements on crystalline materials, atoms displaced from their equilibrium positions give rise to strained material [5,7,19,20]. This is manifested as phase information in the reconstructed image. As the phase information is lost during data acquisition, this phase must be reconstructed. Current methods for doing so are generally able to reconstruct phase information with a high degree of accuracy, although the true global minimum cannot be determined with absolute certainty.

Recent studies have shown that a situation may arise when phase information fails to reconstruct. This might be due to atoms displaced toward half an atomic spacing due to grain
boundaries and dislocations [21,22]. In such cases, the phase is outside the range of \( \pm \pi/2 \) and ambiguities arise as to the appropriate displacement vector [8,20]. Such materials exhibit a strong phase structure and are labeled as highly strained. Other examples where phase information cannot be recovered include inappropriately sampled data or partial coherence effects [23].

To simulate highly strained material, we began by defining a three-dimensional pyramidal-shaped object on a space of 128\(^3\) units with base side lengths of 44 units and top side lengths of 32 units. The object amplitude was then convolved with a normal Gaussian function (\( \sigma_g = 0.25 \)) and given phase information defined as:

\[
\phi(r_i) = \phi_0 \left( 1 - \prod_{i=1}^{3} \cos(2\pi \frac{r_i - L_i}{2L_i}) \right),
\]

where \( \phi_0 = 5.0 \), \( r_i \in [0, L_i] \) is the index along axis \( i \) and \( L_i \) is the axis length which is 128 units (Fig. 2). To simulate diffraction data, the object was then inverse Fourier transformed and Gaussian noise with amplitude \( \alpha \sigma \) added to the amplitude, where \( \sigma \) is the standard deviation of \( |\hat{\rho}(k)| \) and \( \alpha \) is a constant. Data below a value of 30.0 was then removed along with all phase information. Figures 2(a) and 2(f) show the amplitude and phase, respectively, of the original object in the absence of noise. The object amplitude with increasing levels of Gaussian noise are shown in Figs. 2(b) through 2(e), for which \( \alpha = 0.1, 0.2, 0.4, \) and 0.8, respectively, at an isosurface of 50. Here the original phase information is used to Fourier transform the object back to real-space. The corresponding Fourier-space amplitudes of the object are shown in Figs. 2(g) through 2(j).

With increasing \( \alpha \), an increase in the object’s amplitude range is seen in Fig. 2, as might be expected.

Phase reconstruction was then performed on the simulated diffraction amplitude using both the HIO and CSHIO algorithms. For each case, reconstruction commenced initially with random phase and with a real-space amplitude of 200. A total of 4000 iterations were used, with a rectangular support \( S \) of 62 \( \times \) 62 \( \times \) 64 units and a relaxation parameter of \( \beta = 0.9 \). For the CSHIO algorithm, the initial weights and \( p \) were set to unity. There was also some freedom in selecting the parameter \( \epsilon \). We have found that starting the reconstruction with a large value for \( \epsilon \) and gradually reducing that value can aid in preventing stagnation and local minimum “traps.” For this reason we started with \( \epsilon = 4000.0 \) and reduced the value of \( \epsilon \) by \( d = 2.0 \) at the start of each iteration when the following condition was met:

\[
\| \rho^{(n)}(r) \|_2 - \| \rho^{(n-1)}(r) \|_2 \| < \frac{\sqrt{\epsilon}}{\eta},
\]

where the value of \( \eta = 1000.0 \) was chosen in a heuristic manner. We found that early into the reconstruction process, the CSHIO algorithm would occasionally stagnate as a result with a normal Gaussian function.
FIG. 3. (Color online) Reconstructed object. Reconstructed amplitude of the object with the (a) HIO and (b) CSHIO algorithms, respectively. Cross sections of the reconstructed phase information for the (c, e) HIO and (d, f) CSHIO algorithms, respectively (x-y and x-z planes). The HIO algorithm fails to reconstruct the amplitude and phase of the object. The CSHIO algorithm adequately reconstructs both the amplitude and phase of the object. Extraneous phase variations that appear in the HIO phase reconstruction are also suppressed.

of being caught in a local minimum. To avoid this and improve the repeatability of the reconstruction process, the Fourier space constraint $P_{\parallel}$ was relaxed and therefore only applied when the intensity difference $|\hat{\rho}(k)|^2 - I(k)$ fell below the value of $\sqrt{I(k)} \exp(\epsilon(n-1))$. This was at the expense of increasing the number of iterations required for convergence. Here, $\epsilon(n-1)$ is the error metric at the previous iteration $(n-1)$, initialized to unity. The error metric is defined in the usual way as $\epsilon = \sum_k |\hat{\rho}(k)|^2 - |\hat{\rho}_0(k)|^2|/\sum_k |\hat{\rho}_0(k)|^2$. Relaxation of the Fourier space modulus constraint was not applied for the HIO algorithm as no appreciable improvement was observed.

Figure 3 shows the reconstructed amplitude and phase for both the HIO and CSHIO ($\alpha = 0.4$) algorithms. The HIO algorithm clearly fails to correctly reconstruct the object’s amplitude [Fig. 3(a)] and phase [Figs. 3(c) and 3(e)]. Regions of reduced amplitude where the phase is changing most quickly are visible. These gaps where the density is reduced are a common occurrence in the reconstruction of diffraction data from highly strained crystals. More severely, artifacts of this variation can be seen within the reconstructed phase.

The CSHIO algorithm, however, was able to adequately reconstruct both the amplitude [Fig. 3(b)] and phase [Figs. 3(d) and 3(f)] of the object to a high level of accuracy. This accurate reconstruction also persisted for all levels of noise used here. Moreover, extraneous phase variations that appear in the HIO phase reconstruction are completely suppressed. The phase, however, does appear with a constant arbitrary offset which we understand to be possible as an absolute phase is not preserved [24]. The CSHIO algorithm appears therefore to closely approach the global solution.

In addition the twin stagnation problem, where the object and its inverted conjugate become superimposed, was overcome in our results [18]. It was found that, although object twinning might arise during reconstruction, after the reconstruction error had converged this was always suppressed by the CSHIO algorithm with exclusively either the object or its conjugate reconstructed.
Figure 4 shows the reconstruction error residual $\varepsilon$ (solid blue line) and relaxation parameter $\epsilon$ (dashed green line) for each noise level. Each error residual in figure Figs. 4(a,i) through 4(d,i) is shown with the reconstructed object, in Figs. 4(a,ii) through 4(d,ii), with the corresponding phase information mapped onto the surface. A quarter section has been removed and the object shown in two halves to aid in visualizing the phase information. In the following we consider a reconstruction with error $\varepsilon < 5.0 \times 10^{-4}$ as converged. In all simulations performed, it was found that the HIO algorithm did not converge below, but rather oscillated around, $\epsilon \simeq 0.17$ (not shown in Fig. 4). For the CSHIO algorithm and for noise levels of $\alpha = 0.1$ to 0.4 [Figs. 4(a) through 4(c)] the error $\varepsilon$ is reduced to below $5.0 \times 10^{-4}$ within the first 2000 iterations, with the least-noisy dataset converging the earliest. For the $\alpha = 0.8$ noise level, reconstruction had not completely converged after 4000 iterations. This is also evident in the reconstruction, shown in Figs. 4(a,ii) to 4(d,ii). In all cases, the relaxation parameter $\epsilon$ rapidly falls as the error $\varepsilon$ reduces.

Table 1 shows the rate of successful reconstruction for both the HIO and CSHIO algorithms after 100 attempts each consisting of 16000 iterations for varying levels of noise. The HIO algorithm completely fails in all 100 attempts and for all noise levels. The CSHIO algorithm improves on this by converging towards the correct solution in all cases. We found that the reconstruction success rate reduced with an increase in the noise level, as shown. Further improvements to the success rate might be possible through the inclusion of a shrink-wrapped support in the reconstruction [19].

IV. SUMMARY

By utilising an $\ell_1$-norm iterative reweighting scheme we have demonstrated the successful application of compressed sensing to the phase retrieval problem. The results presented herein show that when applied to the problem of *highly strained* material, the CSHIO algorithm is able to outperform other methods and recover phase information with a high degree of accuracy. We hope that this work will lead to a further understanding of the failure of conventional algorithms when reconstructing strong-phase structures. We also envisage the further use of compressed sensing in phase retrieval for reconstructing data obtained from advanced x-ray facilities.

---