

Phase retrieval with only a nonnegativity constraint

YAOCHENG TIAN*  AND JAMES R. FIENUP 

Institute of Optics, University of Rochester, 480 Intercampus Drive, Rochester, New York 14627, USA

*Corresponding author: ytian26@u.rochester.edu

Received 24 October 2022; revised 28 November 2022; accepted 3 December 2022; posted 5 December 2022; published 22 December 2022

We show that it may be possible to reconstruct a real-valued, nonnegative 2D object from the magnitude of its Fourier transform using only a nonnegativity constraint without the usual support constraint, even when significant noise is present in the Fourier intensity data. © 2022 Optica Publishing Group

<https://doi.org/10.1364/OL.478581>

In astronomy, wavefront sensing, X-ray coherent diffractive imaging, and in many other applications, one often encounters the phase retrieval problem. We consider here the Fourier phase retrieval problem of reconstructing a real-valued, nonnegative 2D object from its Fourier magnitude. That is, for a discrete Fourier transform pair $F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \exp[-i2\pi(ux/N + vy/M)] = |F(u, v)| \exp[i\psi(u, v)]$, we hope to reconstruct $f(x, y)$ given only $|F(u, v)|$ and the prior information that $f(x, y)$ is nonnegative ($f(x, y) \geq 0, \forall(x, y)$) and has finite support ($f(x, y) = 0, \forall(x, y) \notin S$), although the support, S , need not be known *a priori*. In astronomy, the solution of this problem enables diffraction-limited imagery of space objects despite atmospheric turbulence [1,2], and in X-ray coherent diffractive imaging, the solution gives images of objects at nanometer-scale resolution [3]. Further understanding of the problem and improvement in practical algorithms for solving the Fourier phase retrieval problem remains important.

One family of practical algorithms for solving the single-intensity Fourier phase retrieval problem is the iterative transform algorithms [1,2,4–6], illustrated in Fig. 1.

The most basic form is the error reduction (ER) algorithm, which is a descendant of the Gerchberg–Saxton algorithm [7] for solving the phase retrieval problem with two intensity measurements. The ER algorithm is equivalent to constrained optimization using steepest descent on the sum of squared error between the measured and the computed Fourier magnitude of the image, and it has the property of reducing errors in each iteration, until it stagnates [2]. Unfortunately, this algorithm often stagnates in local minima, and can be extremely slow to converge. A much more powerful hybrid input-output (HIO) algorithm was developed from the point of view of negative feedback from control theory, and it often converges to a solution much faster and can climb out of local minima, despite its lack of convergence proof [2,8,9]. When using both support and nonnegativity constraints, the next input after k th iteration for HIO is

$$g_{k+1}(x, y) = \begin{cases} g'_k(x, y), & g'_k(x, y) \geq 0 \ \& \ (x, y) \in S \\ g_k(x, y) - \beta g'_k(x, y), & \text{otherwise} \end{cases} \quad (1)$$

where S is the object support or an estimate of it, and β is a feedback parameter which works well between 0.5 and 1. Decades after the first introduction of the HIO algorithm, the continuous hybrid input-output (CHIO) algorithm was introduced as an improvement and modification of the popular HIO algorithm [10]. The CHIO sometimes gives reconstructions superior to those of HIO, presumably because at each pixel within the support S , the next input value is a continuous function of the output pixel value [10], where it is discontinuous when $g'_k(x, y) = 0$ for HIO. In this paper, a slight modification to the CHIO was made to ensure g_{k+1} is a continuous function also for pixels outside the support. Here for CHIO, the next input after the k th iteration is

$$g_{k+1}(x, y) = \begin{cases} g'_k(x, y), & \alpha g_k(x, y) \leq g'_k(x, y) \ \& \ (x, y) \in S \\ g_k(x, y) - (\frac{1-\alpha}{\alpha})g'_k(x, y), & 0 \leq g'_k(x, y) \leq \alpha g_k(x, y) \ \& \ (x, y) \in S \\ g_k(x, y) - \beta g'_k(x, y), & \text{otherwise.} \end{cases} \quad (2)$$

According to the analysis in [11], the phase retrieval problem using the support constraint could be extremely slow to converge due to the possible non-transversal intersection of the Fourier magnitude torus and the support condition, in which case, the algorithms take an exceedingly large number of iterations before reaching the solution at machine precision. Furthermore, it was discussed in Chs. 5 and 9 of [11] that the nonnegativity constraint alone uniquely defines an object of given Fourier magnitude, provided that the autocorrelation of the object has sufficiently small support. In most cases, the intersection between the Fourier magnitude torus and the nonnegativity constraint can be expected to be transversal, implying that the phase retrieval problem using only the nonnegativity constraint should be easier than that using only the support constraint, and even easier than the combination of the support constraint with the nonnegativity constraint. We have found that using nonnegativity alone, excellent reconstructions of certain objects (in particular, objects with noncentrosymmetric and nonconvex supports, which are relatively easy to reconstruct) can be achieved even with a large amount of noise present in the Fourier magnitude.

We found that when using only a nonnegativity constraint, certain algorithms work better than others. For this circumstance, we found that the CHIO algorithm works well, and converged to a lower error than the hybrid iterative map used in [11]. Note

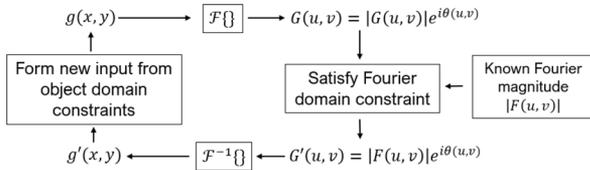


Fig. 1. Iterative transform algorithm.

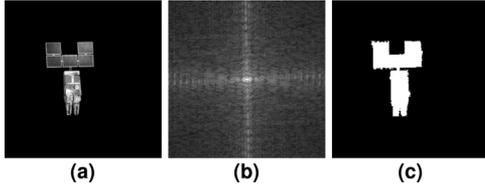


Fig. 2. (a) Object, (b) Fourier magnitude of the object raised to $1/5$ th power, and (c) the true object support.

that when $\alpha = 1/(1 + \beta)$, the CHIO algorithm is equivalent to the hybrid projection-reflection algorithm [4]. We found that convergence to a solution is more stable when $\alpha < 1/(1 + \beta)$. For the following digital experiment, we used the value of $\alpha = 0.4$ and $\beta = 0.7$ in Eq. (2) when running CHIO, and $\beta = 0.7$ in Eq. (1) when running HIO for comparison.

A satellite object [12] with a nonconvex and noncentrosymmetric support was embedded in an array of zeros more than twice as large as its width in either cardinal axis to ensure Nyquist sampling of the Fourier intensity. For comparison, we used two different algorithms (HIO or CHIO) on the Fourier magnitude data, shown in Fig. 2(b), with a true support, shown in Fig. 2(c), together with a nonnegativity constraint or with only a nonnegativity constraint. Using only a nonnegativity constraint is equivalent to setting S in Eqs. (1) and (2) to include all pixels within the computational window.

To improve convergence on this relatively easy-to-reconstruct object, we employed the following sequence: in the first 200 iterations, we used only HIO or CHIO to explore the solution space; after that, we alternated between 49 iterations of HIO or CHIO and 1 iteration of ER. We found that a small number of iterations of ER between HIO or CHIO iterations often gives superior reconstruction compared with using HIO or CHIO alone, when the object is relatively easy to reconstruct. We also found this to be true for the hybrid iterative map in [11], which gave results comparable to CHIO when combined with ER in this way. For objects with centrosymmetric supports, which are significantly more difficult to reconstruct and often prone to the twin-image stagnation mode [13], iterating between ER and HIO could, on the contrary, hinder convergence. In that case, a larger number of HIO iterations without any intervention from ER is effective for circumventing the twin-image stagnation mode [13–15].

Results are compared in Fig. 3. At each iteration, ϵ , the translation invariant normalized RMS error [16] defined by

$$\epsilon^2 = \min_{x_o, y_o} \frac{\sum_{x,y} |g'(x - x_o, y - y_o) - f(x, y)|^2}{\sum_{x,y} |f(x, y)|^2} \quad (3)$$

was calculated between the object estimate $g'(x, y)$ and the upright object, $f(x, y)$, or its twin image, $f(-x, -y)$, to an accuracy of $1/20$ pixels using an efficient algorithm [17]. The minimum of the two was taken as E_{true} at each iteration, since the twin image is also a valid solution.

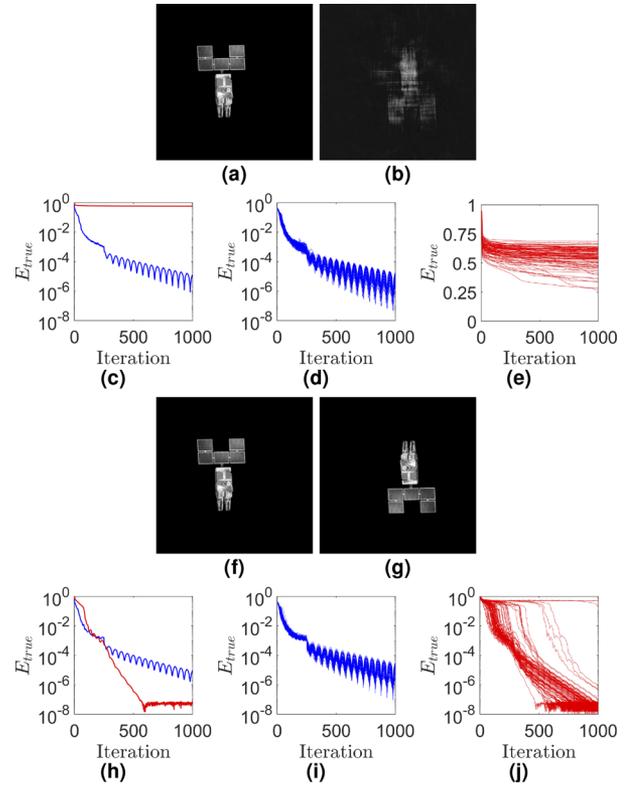


Fig. 3. Using HIO, reconstruction at the end of 1000 iterations using (a) both a true support and nonnegativity constraints, and (b) using only a nonnegativity constraint. (c) E_{true} over iterations using both constraints (blue) and only a nonnegativity constraint (red). For all 100 different random starting guesses, E_{true} over iterations with (d) both constraints, and (e) only a nonnegativity constraint. (f)–(j) Same set of figures but using CHIO.

Example reconstructions shown here are all from the same random starting guess, and registered to either the upright or the twin object, to whichever the E_{true} correspond, for display. When both a true support and nonnegativity constraints were employed with HIO and CHIO, excellent reconstructions were achieved [Figs. 3(a) and 3(f)]. For all 100 different random starting guesses, E_{true} values were consistently low at the end of reconstruction, as shown in Figs. 3(d) and 3(i). We observed that after one iteration of ER, E_{true} went down consistently. In this case, it seems that one iteration of ER helped reset the trajectory of the algorithm's search to a solution, and improved convergence.

Using only a nonnegativity constraint, HIO struggled to converge [Fig. 3(b)] and iterating with ER did not help convergence in any way, while CHIO converged (to the twin image for the case shown) with machine precision E_{true} (which is the square root of a summation with double precision, giving a minimum value on the order of 10^{-8}) in ≈ 600 iterations, shown in Fig. 3(g). Comparing the blue curve with the red curve in Fig. 3(h), we see that the convergence to machine precision was significantly more rapid using only a nonnegativity constraint compared with using both support and nonnegativity constraints. This suggests that when using both support and nonnegativity constraints, the intersection of constraints sets are less favorable compared to when using only a nonnegativity constraint, for this relatively easy-to-reconstruct object. In fact, when using CHIO with only

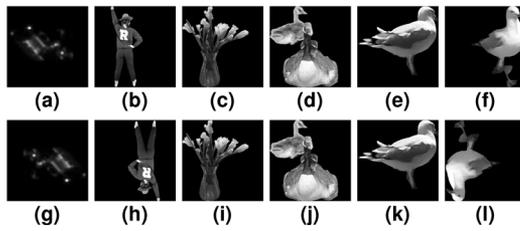


Fig. 4. (a)–(f) Six different objects and (g)–(l) their reconstructions from noiseless Fourier magnitudes using CHIO with only a nonnegativity constraint. Images are cropped to the central portion for display.

a nonnegativity constraint, 94 out of 100 starting guesses converged to machine precision E_{true} within 1000 iterations. Out of the remaining six starting guesses, three were close to machine precision (on the order of 10^{-7}), one was converging near the end with $E_{true} \approx 0.15$, and two stagnated with $E_{true} \approx 0.5$. Surprisingly, neglecting the support constraint altogether did not prevent convergence to the solution. Better reconstructions with lower E_{true} were achieved without the support constraint for the vast majority of the 100 different starting guesses.

This behavior is presumably because when a support constraint is not used, the geometry of the phase retrieval problem [11] is changed in a way that the convergence to the true intersection between constraint sets is more rapid. With a vastly greater number of trivial associates (any 2D rigid translation) when only a nonnegativity constraint is considered, the object estimate need no longer be inside a support constraint, and could appear anywhere in the computational window (including wrapping around the edges of the computational window), which dramatically increases the number of possible trivially ambiguous but perfectly acceptable solutions, and apparently decreases the likelihood of running into non-transversal intersection [11] between constraint sets.

Example reconstructions of other objects using only a nonnegativity constraint are shown in Fig. 4. Note that having a nonconvex and noncentrosymmetric support makes the reconstruction significantly less prone to the persistent twin-image stagnation mode and relatively easy to reconstruct. If the support of the autocorrelation of an object, which can be estimated directly from the Fourier magnitude using the autocorrelation theorem, is nonconvex, the object support is also nonconvex. Although the converse of the statement is not true. Other characteristics of a relatively easy-to-reconstruct object include a support with hard edges that diffract energy strongly in the Fourier plane, and isolated bright points in the object that are similar to those seen in off-axis Fourier transform holography. Each object in Fig. 4 contains at least one above-mentioned characteristic that increased the likelihood of successful reconstruction using only a nonnegativity constraint.

To show that the success of phase retrieval with only a nonnegativity constraint is not limited to only the noise-free case, we perform phase retrieval on noisy Fourier magnitude data. We show that using only a nonnegativity constraint is more stable than using both a true support and nonnegativity constraints. For the following experiment, shown in Fig. 5, different levels of Poisson-distributed noise were applied to the Fourier intensity data, shown in Fig. 6, by first scaling the noise-free Fourier intensity to the desired peak intensity in photon counts, then applying Poisson noise. Note that, even for the case of 10^6 peak

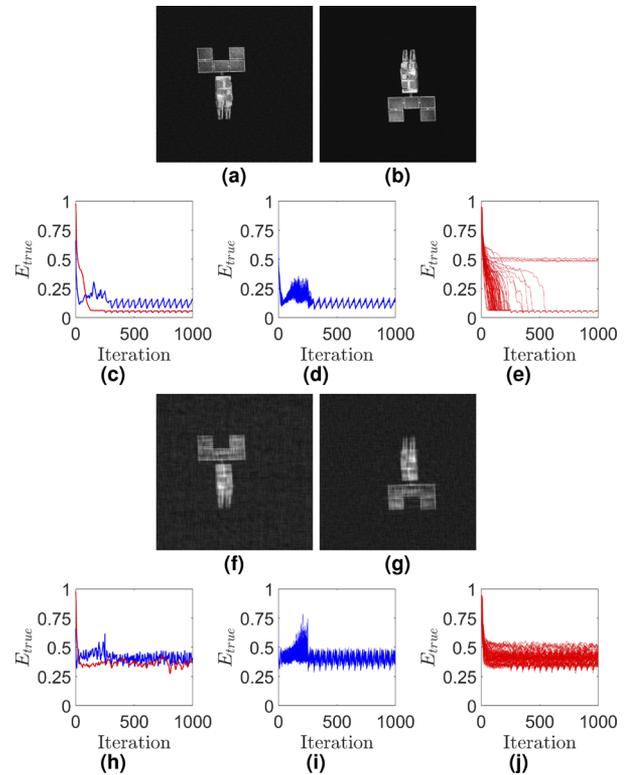


Fig. 5. With 10^6 photons in peak Fourier intensity, reconstruction at the end of 1000 iterations with (a) both a true support and nonnegativity constraints, and with (b) only a nonnegativity constraint. (c) E_{true} over iterations with both a true support and nonnegativity constraints (blue), and with only a nonnegativity constraint (red). For all 100 different random starting guesses, E_{true} over iterations with (d) both constraints, and (e) only a nonnegativity constraint. (f)–(j) Same set of figures but with peak Fourier intensities of 10^4 photons.

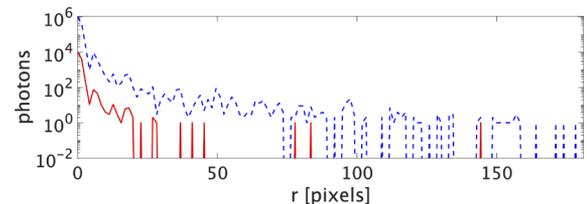


Fig. 6. Diagonal cut through of Fourier intensities from the origin to the bottom right corner, with peak photon counts of 10^6 (dashed blue line) and 10^4 (solid red line).

photons, due to the very high dynamic range of the Fourier data, there are numerous pixels at the higher spatial frequencies whose value becomes zero with photon noise. Reconstructions were attempted with both a true support, Fig. 2(c), and nonnegativity constraints or with only a nonnegativity constraint, using the CHIO algorithm with $\alpha = 0.4$ and $\beta = 0.7$. Due to the Poisson noise in Fourier intensity, $g'(x, y)$ could contain imaginary values after an inverse Fourier transform from $G'(u, v)$. In every iteration, the imaginary part of $g'(x, y)$ was set to zero.

Same as the noise-free case, only CHIO was used in the first 200 iterations, after which point, we iterated between 49 iterations of CHIO and one iteration of ER. We found that when a large amount of noise was present in the Fourier intensity

data, CHIO (and HIO) using a support constraint would produce $g_k(x, y)$ with decreasing mean value. When the mean value of $g_k(x, y)$ becomes negative, the phase of the DC term of $G_k(u, v)$ becomes π instead of 0, producing a $g'_k(x, y)$ with negative mean value after the Fourier magnitude projection. In this case, the algorithm would act in an unstable fashion, with E_{true} wildly oscillating. To avoid wildly oscillating E_{true} , one could either run ER when the mean value of $g_k(x, y)$ is less than some positive threshold value, or keep the phase of DC term of $G_k(u, v)$ to be 0, which ensures that the mean value of $g'_k(x, y)$ stays positive. For the experiment with noisy Fourier data, we kept the phase of the DC term of $G'_k(u, v)$ to be 0. Iterating between CHIO and ER after iteration 200 further stabilized convergence.

Note that E_{true} fluctuated over iterations by a modest amount when using both constraints, but remained much calmer and gave a reconstruction with lower E_{true} at iteration 1000 with only a nonnegativity constraint, comparing blue and red curves in Fig. 5(c). The magnitude of fluctuations when using both constraints increased with increasing levels of noise in the Fourier data. For the case with the greatest noise, when the peak Fourier intensity has 10^4 photons, using both constraints caused E_{true} to fluctuate in the first 200 iterations. When we started to iterate between CHIO and ER after iteration 200, better reconstructions with lower E_{true} were eventually achieved, as shown in Figs. 5(h) and 5(i).

The fluctuating behavior of E_{true} over iterations when using both a true support and nonnegativity constraints is likely caused by the noisy Fourier data. The closest object estimate, obtained using the correct Fourier phase imposed onto the corrupted Fourier magnitude, has energy outside of the support, in the form of zero-mean, independent random noise. However, with the support constraint imposed, the algorithm would attempt to drive the pixel values outside of the object support to zero, where there should be non-zero noise. This inconsistency between the Fourier data and the support constraint likely caused the algorithm to behave more violently than desired.

With only a nonnegativity constraint, the inconsistency problem was still present. Since the noise was not centrosymmetric, the noisy Fourier magnitude no longer preserved Hermitian symmetry. This was inconsistent with the object being real-valued. The closest possible object estimate, after setting the imaginary part of which to zero, contained negative values. However, the inconsistency problem with only a nonnegativity constraint was affecting the convergence much less. Comparing Fig. 5(d) with 5(e), we observed that using only a nonnegativity constraint could give somewhat better results with lower E_{true} .

For objects with centrosymmetric supports, such as squares or circles, the twin-image stagnation problem often hinders the convergence severely [13,14]. In that case, we found that using CHIO with only a nonnegativity constraint would often stagnate with the twin-image problem, although the reconstruction would appear to have the correct support even without employing a support constraint, presumably because the information about the object support is implicitly enforced by the Fourier magnitude data. It is possible that a better algorithm could be developed using only a nonnegativity constraint, which could circumvent the twin-image problem. So far, we have found using HIO with both constraints with many thousands of iterations, without any iteration of ER in between, can eventually circumvent the stagnation mode [15], provided that a good estimate of the centrosymmetric support is available.

In summary, for objects with a noncentrosymmetric and non-convex support, one can sometimes reconstruct the object with only a nonnegativity constraint, with results in this paper as a clear example. With no noise, this can sometimes be done to machine precision. In this case, the CHIO algorithm was superior to the more popular HIO algorithm. Alternating between CHIO and a small number of iterations of ER expedited convergence to machine precision. The ability for CHIO to converge when using only a nonnegativity constraint could be explained by the fact that the input pixel value is a continuous function of the previous output pixel values [10]. Its success persisted even when significant noise was present in the Fourier magnitude data. When using both a true support and nonnegativity constraints with highly noisy Fourier data, the algorithm would often quickly converge to an answer with relatively small error metric, but then quickly diverge to an estimate with no resemblance to the true object, and oscillate between the two cases. We hypothesize the cause of that behavior to be an inconsistency between the noisy Fourier data and the support constraint. The divergent behavior was eliminated by alternating with ER or by keeping the phase in the DC term of the Fourier transform to be 0. Using only the nonnegativity constraint in this case yielded much more stable reconstructions. We also found that using a support constraint along with nonnegativity for the initial iterations (which causes the reconstruction to converge faster initially), and then switching over to nonnegativity only (which is more stable) for the remaining iterations resulted in a faster, more stable reconstruction than using the same constraints for all iterations. These results are consistent with previous results showing that expanding the support constraint in later iterations is helpful [2].

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. J. R. Fienup, *Opt. Lett.* **3**, 27 (1978).
2. J. R. Fienup, *Appl. Opt.* **21**, 2758 (1982).
3. K. A. Nugent, *Adv. Phys.* **59**, 1 (2010).
4. H. H. Bauschke, P. L. Combettes, and D. R. Luke, *J. Opt. Soc. Am. A* **20**, 1025 (2003).
5. V. Elser, *J. Opt. Soc. Am. A* **20**, 40 (2003).
6. D. R. Luke, *Inverse Problems* **21**, 37 (2005).
7. R. W. Gerchberg and W. O. Saxton, *Optik* **35**, 237 (1972).
8. J. H. Seldin and J. R. Fienup, *J. Opt. Soc. Am. A* **7**, 412 (1990).
9. H. Takajo, T. Takahashi, R. Ueda, and M. Taniaka, *J. Opt. Soc. Am. A* **15**, 2849 (1998).
10. J. R. Fienup, *Appl. Opt.* **52**, 45 (2013).
11. A. H. Barnett, C. L. Epstein, L. Greengard, and J. Magland, *Geometry of the Phase Retrieval Problem: Graveyard of Algorithms, Cambridge Monographs on Applied and Computational Mathematics* (Cambridge University Press, 2022).
12. A. Hecht, "Spot4 satellite model," <https://3dwarehouse.sketchup.com/model/a101ec03303ac509e5c77f8d714940f4/SPOT4>.
13. J. R. Fienup and C. C. Wackerman, *J. Opt. Soc. Am. A* **3**, 1897 (1986).
14. M. Guizar-Sicairos and J. R. Fienup, *J. Opt. Soc. Am. A* **29**, 2367 (2012).
15. Y. Tian and J. R. Fienup, *Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, pcAOP)*, Technical Digest Series (Optica Publishing Group, 2022), paper CM2A.2.
16. J. R. Fienup, *Appl. Opt.* **36**, 8352 (1997).
17. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, *Opt. Lett.* **33**, 156 (2008).