Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Fourier phase retrieval with a single mask by Douglas–Rachford algorithms

Pengwen Chen [a], Albert Fannjiang [b,*]

[a] *Applied Mathematics, National Chung Hsing University, Taichung 402, Taiwan*
[b] *Department of Mathematics, University of California, Davis, CA 95616, USA*

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form 15 June 2016
Accepted 19 July 2016
Available online xxxx
Communicated by Radu Balan

Keywords:
Phase retrieval
Coded diffraction pattern
Douglas–Rachford algorithm
Geometric convergence
Spectral gap

ABSTRACT

The Fourier-domain Douglas–Rachford (FDR) algorithm is analyzed for phase retrieval with a single random mask. Since the uniqueness of phase retrieval solution requires more than a single oversampled coded diffraction pattern, the extra information is imposed in either of the following forms: 1) the sector condition on the object; 2) another oversampled diffraction pattern, coded or uncoded.
For both settings, the uniqueness of projected fixed point is proved and for setting 2) the local, geometric convergence is derived with a rate given by a spectral gap condition. Numerical experiments demonstrate global, power-law convergence of FDR from arbitrary initialization for both settings as well as for 3 or more coded diffraction patterns *without* oversampling. In practice, the geometric convergence can be recovered from the power-law regime by a simple projection trick, resulting in highly accurate reconstruction from generic initialization.

1. Introduction

X-ray crystallography has been the preferred technology for determining the structure of a biological molecule over the past hundred years. The method, however, is limited by crystal quality, radiation damage and phase determination [44,48]. The first two problems call for large crystals that yield sufficient diffraction intensities while reducing the dose to individual molecules in the crystal. The difficulty of growing large, well-diffracting crystals is thus the major bottleneck of X-ray crystallography – a necessary experimental step that can range from merely challenging to pretty much impossible, particularly for large macromolecular assemblies and membrane proteins.

By boosting the brightness of available X-rays by 10 orders of magnitude and producing pulses well below 100 fs duration, X-ray free electron lasers (XFEL) offer the possibility of extending structural studies to *single, non-crystalline* particles or molecules by using short intense pulses that out-run radiation damage,

* Corresponding author.
E-mail address: fannjiang@math.ucdavis.edu (A. Fannjiang).

http://dx.doi.org/10.1016/j.acha.2016.07.003
1063-5203/

thus circumventing the first two aforementioned problems [50]. In the so-called *diffract-before-destruct* approach [20,21,56], a stream of particles is flowed across the XFEL beam and randomly hit by a single X-ray pulse, forming a single diffraction pattern before being vaporized as a nano-plasma burst. Each diffraction pattern contains certain information about the planar projection of the scatterer along the direction of the beam which is to be recovered by phase retrieval techniques [11].

The modern approach to phase retrieval for non-periodic objects roughly starts with the Gerchberg–Saxton algorithm [34], followed by its variant, Error Reduction (ER), and the more powerful Hybrid-Input–Output (HIO) algorithm [32,33]. These form the cornerstones of the standard *iterative transform algorithms* (ITA) [8,43].

However, the standard ITA tend to stagnate and do not perform well without additional prior information, such as tight support and positivity. The reason is that the *plain* diffraction pattern alone does not guarantee uniqueness of solution (see [54], however, for uniqueness under additional prior information). On the contrary, many phase retrieval solutions exist for a given diffraction pattern, resulting in what is called the *phase* problem [37].

To this end, a promising approach is to measure the diffraction pattern with a *single* random mask and use the coded diffraction pattern as the data. As shown in [28], the uniqueness of solution is restored with a high probability given any scatterer whose value is restricted to a known sector (say, the upper half plane) of the complex plane (see Proposition 2.1).

Indeed, the sector constraint is a practical, realistic condition to impose on almost all materials as the imaginary part of the scatterer is proportional to the (positive) extinction coefficient with the upper half plane as the sector constraint [11]. For X-ray, the scatterers usually have positive real (except for resonance frequencies) and imaginary parts, making the first quadrant the sector constraint [18].

What happens if the sector condition is not met and consequently one coded diffraction pattern is not enough to ensure uniqueness? This question is particularly pertinent to the diffract-before-destruct approach as the particle can not withstand the radiation damage from more than one XFEL pulses.

A plausible measurement scheme is to guide the transmitted field (the transmission function [11]) from a *planar* illumination through a beam splitter [52], generating two copies of the transmitted field which are then measured separately as a coded diffraction pattern and a plain diffraction pattern. In this set-up, the object function is the transmitted field behind the particle and the phase retrieval problem becomes the wave-front reconstruction problem [11,36]. In practice beam splitters and the masks (or any measurement devices) should be used as sparingly as possible to avoid introducing excessive noises in the data.

As shown in [28], phase retrieval with two coded diffraction patterns has a unique solution, up to a constant phase factor, almost surely without the sector constraint (see Proposition 2.1).

With the uniqueness-ensuring sampling schemes (Section 1.1), *ad hoc* combinations of members of ITA (such as HIO and ER) can be devised to recover the true solution [30,31]. There is, however, no convergence proof for these algorithms, except for alternating projections, including ER (see [22] and references therein).

The main goal of the paper is to prove the *local, geometric convergence* of the Douglas–Rachford (DR) algorithm to a unique fixed point in the case of one or two oversampled diffraction patterns (Theorems 5.1, 6.3 and 4.2) and demonstrate *global convergence* numerically (Section 7).

DR has the following general form: Let P_1 and P_2 be the projections onto the two constraint sets, respectively. For phase retrieval, P_1 describes the projection onto the set of diffracted *fields* and P_2 the data fitting projector constrained by the measured diffraction patterns. Let $R_1 = 2P_1 - I$ and $R_2 = 2P_2 - I$ be the respective reflection operators. The Douglas–Rachford (DR) algorithm is defined by the *average alternating reflection* scheme [24,25,41]

$$\begin{aligned} y^{(k+1)} &:= \frac{1}{2}(I + R_1R_2)y^{(k)} \\ &= y^{(k)} + P_1(2P_2 - I)y^{(k)} - P_2y^{(k)}, \quad k = 1, 2, 3\cdots \end{aligned} \tag{1}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

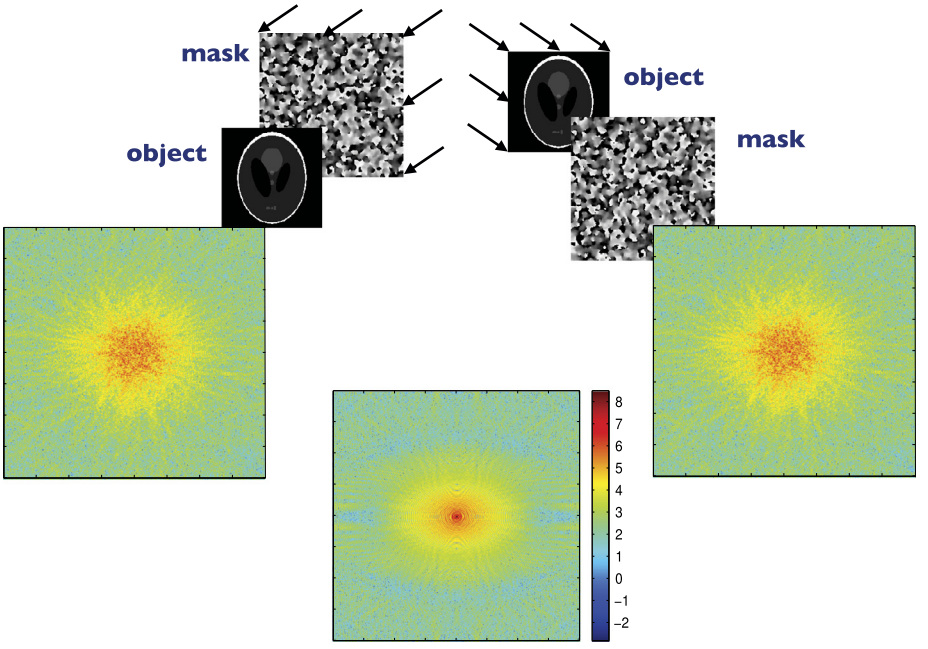


Fig. 1. Conceptual layout of coherent lensless imaging with a random mask (left) *before* (for random illumination) or (right) *behind* (for wavefront sensing) the object (middle). The diffraction pattern measured without a mask has a larger dynamic range and thus a higher chance of damaging the sensors. The color bar is on a logarithmic scale.

Closely related to HIO, DR also belongs to the ITA family (Section 3). ITA are computationally efficient thanks to the fast Fourier transform (FFT) and explicit nature of P_1, P_2 (see (14) below).

The focus of our analysis and implementation is on the Fourier-domain Douglas–Rachford algorithm (FDR) for which $y^{(k)}$ are the Fourier vectors (with phase). The FDR defined in (15) is a member of a general class of Fourier-domain fixed point algorithms formulated in [29].

1.1. Oversampled diffraction patterns

Next we describe our sampling schemes before we can properly introduce P_1, P_2 and the Douglas–Rachford algorithm for phase retrieval (Section 3).

Let $f(\mathbf{n})$ be a discrete object function with $\mathbf{n} = (n_1, n_2, \cdots, n_d) \in \mathbb{Z}^d$. Consider the object space consisting of all functions supported in $\mathcal{M} = \{0 \le m_1 \le M_1, 0 \le m_2 \le M_2, \cdots, 0 \le m_d \le M_d\}$. We assume $d \ge 2$.

Let

$$F(\mathbf{w}) = \sum_{\mathbf{m}\in\mathcal{M}} e^{-i2\pi\mathbf{m}\cdot\mathbf{w}} f(\mathbf{m}), \quad \mathbf{w} = (w_1, \cdots, w_d)$$

be the Fourier transform of f. Let $(\nu_1, \cdots, \nu_d)$ be the coordinates of the sensor plane and L the distance between the object plane and the sensor plane, cf. Fig. 1. Under the Fraunhofer approximation, the diffraction pattern is proportional to

$$I(\mathbf{w}) = |F(\mathbf{w})|^2, \quad \mathbf{w} = (w_j) \tag{2}$$

where λ is the wavelength and

$$w_j = \frac{\nu_j}{\lambda L}, \ j = 1, \cdots, d,$$

the spatial frequencies [11]. We can rewrite (2) as

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

$$I(\mathbf{w}) = \sum_{\mathbf{n}=-\mathbf{M}}^{\mathbf{M}} \sum_{\mathbf{m}\in\mathcal{M}} f(\mathbf{m}+\mathbf{n})\overline{f(\mathbf{m})}e^{-i2\pi\mathbf{n}\cdot\mathbf{w}}, \quad \mathbf{w} = (w_1, \cdots, w_d) \in [0,1]^d, \quad \mathbf{M} = (M_1, \cdots, M_d)$$

which is the Fourier transform of the autocorrelation

$$R_f(\mathbf{n}) = \sum_{\mathbf{m}\in\mathcal{M}} f(\mathbf{m}+\mathbf{n})\overline{f(\mathbf{m})}.$$

Here and below the over-line notation means complex conjugacy.

Note that R_f is defined on the enlarged grid

$$\widetilde{\mathcal{M}} = \{(m_1, \cdots, m_d) \in \mathbb{Z}^d : -M_1 \le m_1 \le M_1, \cdots, -M_d \le m_d \le M_d\}$$

whose cardinality is roughly 2^d times that of $\mathcal{M}$. Hence by sampling the diffraction pattern on the grid

$$\mathcal{L} = \left\{(w_1, \cdots, w_d) \mid w_j = 0, \frac{1}{2M_j+1}, \frac{2}{2M_j+1}, \cdots, \frac{2M_j}{2M_j+1}\right\} \tag{3}$$

we can recover the autocorrelation function by the inverse Fourier transform. This is the *standard oversampling* with which the diffraction pattern and the autocorrelation function become equivalent via the Fourier transform [45,46]. The oversampled diffraction pattern is measured by a denser array of sensors on the grid $\lambda L\mathcal{L}$.

We denote by Φ the oversampled discrete Fourier transform (DFT). Specifically $\Phi \in \mathbb{C}^{|\tilde{\mathcal{M}}|,|\mathcal{M}|}$ is the sub-column matrix of the standard DFT on the extended grid $\tilde{\mathcal{M}}$ where $|\mathcal{M}|$ is the cardinality of $\mathcal{M}$.

A coded diffraction pattern is measured with a mask whose effect is multiplicative and results in a *masked object* of the form $g(\mathbf{n}) = f(\mathbf{n})\mu(\mathbf{n})$ where $\{\mu(\mathbf{n})\}$ is a finite set of random variables representing the mask. In other words, a coded diffraction pattern is just the plain diffraction pattern of a masked object.

We will focus on the effect of *random phase* ϕ in the mask function $\mu(\mathbf{n}) = |\mu|(\mathbf{n})e^{i\phi(\mathbf{n})}$ where $\phi(\mathbf{n})$ are independent, continuous real-valued random variables. In other words, each $\phi(\mathbf{n})$ is independently distributed with a probability density function on $(-\pi, \pi]$ that may depend on $\mathbf{n}$. Continuous phase modulation can be experimentally realized with various techniques such as spread spectrum phase modulation [59].

We also require that $|\mu|(\mathbf{n}) \neq 0, \forall \mathbf{n} \in \mathcal{M}$ (i.e. the mask is transparent). This is necessary for unique reconstruction of the object as any opaque pixel of the mask where $\mu(\mathbf{n}) = 0$ would block the transmission of the information $f(\mathbf{n})$. By absorbing $|\mu(\mathbf{n})|$ into the object function we can assume, without loss of generality, that $|\mu(\mathbf{n})| = 1, \forall \mathbf{n}$, i.e. a *phase* mask.

With a proper choice of the normalizing constant c, a phase mask then gives rise to an *isometric* propagation matrix

$$\text{(1-mask)} \quad A^* = c\Phi \text{ diag}\{\mu\}, \tag{4}$$

i.e. $AA^* = I$.

When two phase masks μ_1, μ_2 are deployed and independent of each other, the propagation matrix A^* is the stacked coded DFTs, i.e.

$$\text{(2-mask case)} \quad A^* = c\begin{bmatrix} \Phi \text{ diag}\{\mu_1\} \\ \Phi \text{ diag}\{\mu_2\} \end{bmatrix}. \tag{5}$$

With proper normalization, A^* is again isometric.

In line with the spirit of simplifying measurement complexity discussed above, we remove the second mask (i.e. $\mu_2 \equiv 1$) and consider the propagation matrix [28,30,31]

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

$$(1\tfrac{1}{2}\text{-mask case}) \quad A^* = c \begin{bmatrix} \Phi\ \mathrm{diag}\{\mu\} \\ \Phi \end{bmatrix} \tag{6}$$

normalized to be isometric. In other words, one oversampled coded pattern and one oversampled plain pattern are used for reconstruction.

For convenience, we shall refer to this set-up as the $1\frac{1}{2}$-mask case to distinguish it from the one- and two-mask cases. Eq. (6) is the set-up for two-pattern experiments in Section 7.

The main objective of the paper is to prove local, geometric convergence of FDR (1) where $y^{(k)}$ are the Fourier vectors (with phase) with two diffraction patterns. To this end, we first give a concrete characterization of the fixed point set and a spectral decomposition of the gradient map of FDR. We show that the geometric rate is determined by the spectral gap which is positive as long as the measurement scheme contains at least one oversampled coded diffraction pattern. The characterization of the fixed point set and the spectral decomposition of the gradient map are the main ingredients of the proof of local, geometric convergence of FDR.

1.2. Comparison with other literature

For the optical spectrum [55], experiments with coded diffraction patterns are not new and can be implemented by computer generated holograms [12], random phase plates [1] and liquid crystal phase-only panels [27]. Recently, a phase mask with randomly distributed pinholes has been implemented for soft X-ray [42].

Coded-aperture phase retrieval was formulated as a convex trace-norm minimization problem in [13, 15,16,19] and the uniqueness of the global minimizer was proved in [15] under the assumption that the number of independently coded diffraction patterns is sufficiently large (polylogarithmic in $|\mathcal{M}|$). Moreover, the convex reformulation of phase retrieval increases the dimension from $\mathcal{O}(|\mathcal{M}|)$ to $\mathcal{O}(|\mathcal{M}|^2)$ and thus is prohibitively expensive for large problems (see also [7,35]).

Alternative non-convex minimization formulations were proposed and solved by various gradient methods [14,47]. In practice, these algorithms are locally convergent with a comparatively large number (≥ 6) of coded diffraction patterns.

An important difference between the measurement schemes in these papers and the present work (as well as [28,30,31]) is that their coded diffraction patterns are *not* oversampled. Another distinctive feature of the present setting is that the dimension $d \geq 2$ is required for the spectral gap (Theorem 6.3) and the uniqueness of fixed point (Theorem 4.2).

In this connection, we emphasize that reducing the number of coded diffraction patterns is crucial for the diffract-before-destruct approach and in comparison oversampling is a small price to pay with current sensor technologies.

Arguably a bigger price may be the loss of the robust injectivity property pursued in these works (also see [5,6]). Indeed, with at most two random masks, the phase retrieval map $|A^*f|$ with A^* given by (6) or (5) is injective only after certain finite set is excluded from the space of objects $\mathbb{C}^{|\mathcal{M}|}$ [28].

On the other hand, for any given f, the solution to the phase retrieval problem with A^* given by (6) or (5) is unique, up to a constant phase factor, with probability one [28]. In contrast, the injectivity theorems proved in [15] hold true with probability approaching 1 polynomially in $|\mathcal{M}|^{-1}$, instead of probability one, and require a large number of coded diffraction patterns.

In other words, both approaches exclude some small sets, ours from the space of objects and theirs from the space of random masks to achieve injectivity. On balance, neither approach carries over to the other setting.

Our numerical results show that the uniqueness framework of [28] suffices for most practical purposes and the numerical scheme proposed here, the Fourier domain DR (FDR), exhibits globally convergence behaviors from random initialization.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

To our knowledge, local geometric convergence is not known for any fixed point algorithms for Fourier phase retrieval except for alternating projections (see [22] and references therein). The present paper aims to fill this gap. It is noteworthy that the geometric rate of convergence was characterized as the cosine of the Friedrichs angle between the two constraint subspaces in [9,10] for the convex setting, it is characterized here for a nonconvex setting as the spectral gap condition. While we can not prove *global* convergence of FDR, we will present strong numerical evidence for it.

There is much more literature on phase retrieval with generic frames and independent random matrices [2–6,16,17,23,26,49,57,58] which is quite different from Fourier phase retrieval. There also is a growing body of work on phase retrieval under sparsity assumptions, see [39,40,51,53] and the references therein.

The rest of the paper is organized as follows. In Section 2, we simplify the notation for presenting the main results. In Section 3, we describe the DR algorithm widely used in convex optimization problems and formulate its two versions (FDR and ODR) for phase retrieval. In Section 4, we prove the uniqueness of the projected fixed point for the one- and two-pattern cases (Theorem 4.2) and give a complete characterization of the fixed point set in the Fourier domain (Corollary 4.5). In Section 5, we prove local convergence of FDR under the spectral gap assumption (Theorem 5.1). In Section 6, we prove the spectral gap condition with at least one oversampled coded diffraction pattern in the data (Theorem 6.3). In Section 7 we demonstrate global convergence of FDR by numerical experiments.

2. Set-up and notation

We simplify the notation as follows. The more elaborate notation of Section 1.1 will be needed only in the appendix.

First, we convert the d-dimensional grid into an ordered set of indices. The unknown object will now be denoted by $x_0 \in \mathbb{C}^n$ with $n = |\mathcal{M}|$. In other words, x_0 is the vectorized version of the object function f supported in $\mathcal{M} \subset \mathbb{Z}^d, d \geq 2$ (Section 1.1).

Rank ≥ 2 property: x_0 is rank ≥ 2 if the convex hull of $\text{supp}\{f\} \subset \mathbb{C}^n$ is a two or higher dimensional set.

Sector constraint: x_0 satisfies the sector constraint if the principal value (denoted by $\measuredangle x_0(j)$) of $\arg\{x_0(j)\}, \forall j$ is restricted to a *sector* $[-\alpha\pi, \beta\pi] \subsetneq (-\pi, \pi], \forall \mathbf{n}$. As mentioned above almost all scatterers f have a nonnegative imaginary part and hence satisfy the sector constraint with $\alpha = 0, \beta = 1$. The sector constraint serves as transition between the standard positivity constraint ($\alpha = \beta = 0$) and the null constraint ($\alpha = \beta = 1$).

The sector projection is explicitly given as follows: For $j \leq n$

$$[x]_{\mathcal{X}}(j) = \begin{cases} x(j) & \text{if } \measuredangle x(j) \in [-\alpha\pi, \beta\pi] \\ \Re[x(j)e^{-i\beta\pi}]e^{i\beta\pi} & \text{if } \measuredangle x(j) \in [\beta\pi, (\beta+1/2)\pi] \\ \Re[x(j)e^{i\alpha\pi}]e^{-i\alpha\pi} & \text{if } \measuredangle x(j) \in [-(\alpha+1/2)\pi, -\alpha\pi] \\ 0 & \text{else} \end{cases} \tag{7}$$

and $[x]_{\mathcal{X}}(j) = 0, j > n+1$.

Let $\mathcal{X}$ be a nonempty closed convex set in $\mathbb{C}^n$ and the space of objects. Denote the projection onto $\mathcal{X}$ by

$$[x]_{\mathcal{X}} = \arg\min_{x' \in \mathcal{X}} \|x' - x\|. \tag{8}$$

Phase retrieval problem. For a given unknown object x_0 of rank ≥ 2, let $A^* = [a_j^*] \in \mathbb{C}^{N \times n}$ be the propagation matrix given by (4), (5) or (6) where A^* is normalized to be isometric and $b = |A^* x_0| \in \mathbb{R}^N$ be the data vector. Phase retrieval is to find a solution x to the equation

$$b = |A^* x|, \quad x \in \mathcal{X}. \tag{9}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

We focus on two cases.

(i) One-pattern case: A^* is given by (4), $[x]_{\mathcal{X}}$ is given by (7).

(ii) Two-pattern case: A^* is given by (5) or (6), $\mathcal{X} = \mathbb{C}^n$.

Now we recall the following uniqueness theorem for Fourier phase retrieval.

Proposition 2.1 *([28]). Let x_0 be a rank ≥ 2 object and x a solution of the phase retrieval problem (9). Suppose that each $\phi(\mathbf{n})$ of the phase mask(s) is independently distributed with a probability density function on $(-\pi, \pi]$.*

(i) One-pattern case. *Suppose, in addition, that $\measuredangle x_0(j) \in [-\alpha\pi, \beta\pi]$, $\forall j$ with $\alpha + \beta \in (0, 2)$ and that the density function for $\phi(\mathbf{n})$ is a constant (i.e. $(2\pi)^{-1}$) for every $\mathbf{n}$.*

Then $x = e^{i\theta} x_0$ for some constant $\theta \in (-\pi, \pi]$ with a high probability which has a simple, lower bound

$$1 - n \left| \frac{\beta + \alpha}{2} \right|^{\lfloor\!\lfloor S/2 \rfloor\!\rfloor} \tag{10}$$

where S is the number of nonzero components in x_0 and $\lfloor\!\lfloor S/2 \rfloor\!\rfloor$ the greatest integer less than or equal to $S/2$.

(ii) Two-pattern case. *Then $x = e^{i\theta} x_0$ for some constant $\theta \in (-\pi, \pi]$ with probability one.*

The proof of Proposition 2.1 is given in [28] where more general uniqueness theorems can be found, including the $1\frac{1}{2}$-mask case.

Phase retrieval solution is unique only up to a constant of modulus one no matter how many coded diffraction patterns are measured. Thus the proper error metric for an estimate $\hat{x}$ of the true solution x_0 is given by

$$\min_{\theta \in \mathbb{R}} \|e^{-i\theta} x_0 - \hat{x}\| = \min_{\theta \in \mathbb{R}} \|e^{i\theta} \hat{x} - x_0\| \tag{11}$$

where the optimal phase adjustment $\hat{\theta}$ is given by

$$\hat{\theta} = \measuredangle(\hat{x}^* x_0).$$

Throughout the paper, we assume the canonical embedding

$$\mathbb{C}^n \subseteq \mathbb{C}^{\tilde{n}} \subseteq \mathbb{C}^N, \quad n \leq \tilde{n} \leq N.$$

For example, if $x \in \mathbb{C}^n$, then the embedded vector in $\mathbb{C}^{\tilde{n}}$ or $\mathbb{C}^N$, still denoted by x, has zero components $x(j) = 0$ for $j \geq n+1$. This is referred to as *zero padding* and $\tilde{n}/n$ is the *padding ratio*. Conversely, if $x \in \mathbb{C}^{\tilde{n}}$ or $\mathbb{C}^N$, then $[x]_n \in \mathbb{C}^n$ denotes the projected vector onto $\mathbb{C}^n$. Clearly, $[x]_{\mathbb{C}^n} = [x]_n$.

The vector space $\mathbb{C}^N = \mathbb{R}^N \oplus_{\mathbb{R}} i\mathbb{R}^N$ is isomorphic to $\mathbb{R}^{2N}$ via the map

$$G(v) := \begin{bmatrix} \Re(v) \\ \Im(v) \end{bmatrix}, \quad \forall v \in \mathbb{C}^N \tag{12}$$

and endowed with the real inner product

$$\langle u, v \rangle := \Re(u^* v) = G(u)^\top G(v), \quad u, v \in \mathbb{C}^N.$$

We say u and v are orthogonal to each other (denoted by $u \perp v$) iff $\langle u, v \rangle = 0$.

With a slight abuse of notation, we will use $G(u)$ to denote the conversion of a complex-valued vector u in $\mathbb{C}^n, \mathbb{C}^{\tilde{n}}$ or $\mathbb{C}^N$ to its real-valued version.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Phase vector: Let $y \odot y'$ and y/y' be the component-wise multiplication and division between two vectors y, y', respectively. For any $y \in \mathbb{C}^N$ define the phase vector $\omega \in \mathbb{C}^N$ with

$$\omega(j) = \exp(i\measuredangle y(j))$$

which, if $y(j) = 0$, is not uniquely defined and can be assigned an arbitrary value. Alternatively, we write $\omega(j) = y(j)/|y(j)|$ which, if $y(j) = 0$, is not uniquely defined and can be assigned an arbitrary value. Unless otherwise specified we set $\omega(j) = 1$ if $y(j) = 0$.

For the one-mask (4) and two-mask (5) cases, the mask function by assumption is a finite set of continuous random variables and so is $y_0 = A^* x_0$. Therefore y_0 vanishes nowhere almost surely, i.e.

$$b_{\min} = \min_j b_j > 0$$

On the other hand, a plain diffraction pattern may have zero components depending on the object.

3. Douglas–Rachford algorithms

Phase retrieval can be formulated as the following feasibility problem in the Fourier domain

$$\text{Find} \quad \hat{y} \in A^*\mathcal{X} \cap \mathcal{Y}, \quad \mathcal{Y} := \{y \in \mathbb{C}^N : |y| = b\}. \tag{13}$$

Let P_1 be the projection onto $A^*\mathcal{X}$ and P_2 the projection onto $\mathcal{Y}$:

$$P_1 y = A^*[Ay]_{\mathcal{X}}, \quad P_2 y = b \odot \frac{y}{|y|} \tag{14}$$

where the phase vector $y/|y|$ follows the convention discussed above.

Then DR (1) becomes $y^{(k+1)} = S_{\mathrm{f}}(y^{(k)})$ with

$$S_{\mathrm{f}}(y) = y + A^*\left[A\left(2b \odot \frac{y}{|y|} - y\right)\right]_{\mathcal{X}} - b \odot \frac{y}{|y|} \tag{15}$$

which we call the *Fourier-domain DR* (FDR) to contrast with the following object domain version. FDR is an example of Fourier-domain fixed point algorithms for phase retrieval [29]. Note that $S_{\mathrm{f}}(y)$ is differentiable at y if and only if $\mathrm{supp}(b) \subseteq \mathrm{supp}(y)$.

Let $\tilde{A}^* = [A^*, A_\perp^*] \in \mathbb{C}^{N,\tilde{n}}$ be a complex isometric extension of A^*, implying that $A_\perp A_\perp^* = I, AA_\perp^* = 0, A_\perp A^* = 0$. Then the phase retrieval problem can be more generally formulated as $|\tilde{A}^* x| = b, x \in \mathcal{X}$. Consider the feasibility problem

$$\text{Find} \quad \hat{x} \in \mathcal{X} \cap \tilde{\mathcal{X}}, \quad \tilde{\mathcal{X}} := \left\{x \in \mathbb{C}^{\tilde{n}} : |\tilde{A}^* x| = b\right\}. \tag{16}$$

Let P_1 be the projection onto $\mathcal{X}$, i.e. $P_1 x = [x]_{\mathcal{X}}$, and P_2 the projection onto $\tilde{\mathcal{X}}$. When $\tilde{n} = N$ (hence $\tilde{A}$ is unitary),

$$P_2 x = \tilde{A}\left(b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|}\right) \tag{17}$$

and (15) is equivalent to

$$S(x) = x + \left[\tilde{A}\left(2b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|}\right) - x\right]_{\mathcal{X}} - \tilde{A}\left(b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|}\right). \tag{18}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

In this case, we have

$$\tilde{A}^* S \tilde{A} = S_{\rm f}, \quad \text{for} \quad \tilde{n} = N. \tag{19}$$

In the 1-pattern case with the standard oversampling $N = \tilde{n} \approx 2^d n$, $\tilde{A} = A$ is unitary and (18) is also known as the Hybrid-Input–Output (HIO) algorithm (with the HIO parameter set to one) [8,32].

For $\tilde{n} < N$ (as with two oversampled diffraction patterns $N \approx 2^{d+1} n$), the precise form of P_2 is not known explicitly. For the purpose of contrasting with (15) and for lack of a better term we shall call (18) (with $\tilde{n} \le N$) the generalized Object-domain Douglas–Rachford algorithm (ODR for short). The ODR family is an interpolation between the HIO and FDR.

While ODR depends explicitly on $\tilde{n}$, FDR is independent of $\tilde{n}$ in the sense that

$$S_{\rm f}(y) = y + \tilde{A}^* \left[\tilde{A} \left(2b \odot \frac{y}{|y|} - y \right) \right]_{\mathcal{X}} - b \odot \frac{y}{|y|} \tag{20}$$

since $[\tilde{A} y]_{\mathcal{X}} = [Ay]_{\mathcal{X}} \in \mathbb{C}^n$ and $\tilde{A}^* [\tilde{A} y]_{\mathcal{X}} = A^* [Ay]_{\mathcal{X}}$.

The FDR map (15) is the main object of the subsequent analysis. What contributes to the superior numerical performance as well as the convergence guarantee is the existence of a large fixed point set $\mathcal{F}$ in the Fourier space. Moreover, under the proper projection depending on the measurement scheme, the entire fixed point set is mapped back to the true object, up to a constant phase factor.

4. Uniqueness of projected fixed point

To accommodate the arbitrariness of the phase of zero components as in [30], we call y_* a FDR **fixed point** if there exists

$$u \in U = \{u = (u(i)) \in \mathbb{C}^N : |u(i)| = 1, \ \forall i\}$$

satisfying

$$u \in U, \quad u(j) = 1, \quad \text{whenever } y_*(j) \neq 0 \tag{21}$$

such that the fixed point equation holds:

$$A^* \left[A \left(2b \odot \frac{y_*}{|y_*|} \odot u - y_* \right) \right]_{\mathcal{X}} = b \odot \frac{y_*}{|y_*|} \odot u. \tag{22}$$

Note that if the sequence $y^{(k)} = S_{\rm f}^{k-1}(y^{(1)})$ converges a limit y_∞ that has no zero component, then the limit y_∞ is a FDR fixed point with $u \equiv 1$.

Let $x_* = A y_*$ and $\omega_* = y_*/|y_*|$. Define

$$\hat{x} = [A\,(2b \odot \omega_* \odot u - y_*)]_{\mathcal{X}} = [2A(b \odot \omega_* \odot u) - x_*]_{\mathcal{X}}\,, \tag{23}$$

for some u satisfying (21) where $\mathcal{X}$ represents the sector condition in the 1-pattern case and $\mathcal{X} = \mathbb{C}^n$ in the 2-pattern case.

We have from (22)

$$A^* \hat{x} = b \odot \omega_* \odot u \tag{24}$$

which implies the following result.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Proposition 4.1.

$$|A^*\hat{x}| = |A^*x_0| \tag{25}$$

$$\measuredangle A^*\hat{x} = \measuredangle(\omega_* \odot u) \quad \textit{on } \operatorname{supp}(b). \tag{26}$$

Eq. (25) is related to phase retrieval while eq. (26) magnitude retrieval problem, both with coded diffraction patterns. The uniqueness theorem for the former is given in Proposition 2.1 while the uniqueness theorem for the latter is given in Proposition 6.1.

Theorem 4.2. *Under the assumptions of Proposition 2.1, the following statements hold.*
(i) One-pattern case. *With probability at least given in (10), $\hat{x} = e^{i\theta}x_0$ for some $\theta \in \mathbb{R}$.*
(ii) Two-pattern case. *Almost surely $\hat{x} = x_* = e^{i\theta}x_0$ for some constant $\theta \in \mathbb{R}$.*

Remark 4.3. See [29] for an extension of this result to a more general class of fixed point algorithms.

Proof. As mentioned in Sec. 2, $|\operatorname{supp}(b)| = N$ almost surely with the measurement schemes (4) and (5).

By Proposition 2.1 (25) implies that $\hat{x} = e^{i\theta}x_0$ for some constant $\theta \in \mathbb{R}$, with the only difference between case (i) and case (ii) being the probability with which this statement holds. To complete the proof, we only need to show $\hat{x} = x_*$ for (ii).

By (26) and the identity $\hat{x} = e^{i\theta}x_0$, we have

$$e^{i\theta}\omega_0 = \omega_* \odot u \tag{27}$$

since $b > 0$ almost surely. Substituting (27) into (23) we obtain

$$e^{i\theta}x_0 = 2e^{i\theta}x_0 - x_*$$

implying $e^{i\theta}x_0 = x_*$. In conclusion,

$$x_* = \hat{x} = e^{i\theta}x_0$$

as claimed. □

Next, we look into the FDR fixed point set (in the Fourier domain) more closely.

Define

$$\mathcal{F} = \cup_\theta \mathcal{F}_\theta \tag{28}$$

where

$$\mathcal{F}_\theta = \left\{e^{i\theta}(b+\eta) \odot \omega_0 : B\eta = 0,\ \eta \in \mathbb{R}^N\right\} \cap \left\{e^{i\theta}(b+\eta) \odot \omega_0 : b + \eta \geq 0\right\} \tag{29}$$

is convex for each θ since it is intersection of a real-affine set and a convex set. It is straightforward to check the following.

Proposition 4.4. *The elements of $\mathcal{F}$ are FDR fixed points if $\mathcal{X} = \mathbb{C}^n$.*

Proof. For any $y = e^{i\theta}(b+\eta) \odot \omega_0$ with $b + \eta \geq 0$, we have

$$\frac{y}{|y|} = e^{i\theta}\omega_0 \odot u$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

where u satisfies

$$u \in U, \quad u(j) = \begin{cases} 1, & b(j) + \eta(j) > 0 \\ e^{-i\theta}\overline{\omega_0(j)}, & b(j) + \eta(j) = 0 \end{cases} \quad \forall j.$$

Thus

$$A^*A\left(b \odot \frac{y}{|y|} \odot \bar{u}\right) = e^{i\theta}A^*A(b \odot \omega_0 \odot u \odot \bar{u}) = e^{i\theta}A^*Ay_0 = e^{i\theta}y_0. \tag{30}$$

On the other hand, in checking (22) we have the calculation

$$A^*Ay = e^{i\theta}A^*\left[A(b \odot \omega_0) + A(\eta \odot \omega_0)\right] = e^{i\theta}A^*A(b \odot \omega_0) = e^{i\theta}A^*Ay_0 = e^{i\theta}y_0 \tag{31}$$

and

$$b \odot \frac{y}{|y|} \odot \bar{u} = e^{i\theta}b \odot \omega_0 \odot u \odot \bar{u} = e^{i\theta}b \odot \omega_0 = e^{i\theta}y_0. \tag{32}$$

Now (22) is satisfied in view of (30)–(32) and the fact that

$$\bar{u} \in U, \quad \bar{u}(j) = 1, \quad \text{whenever } b + \eta > 0. \qquad \square$$

Corollary 4.5. *Under the assumptions of Proposition 2.1, any FDR fixed point y_* shares the same phase as y_0 up to a global constant, i.e.*

$$\measuredangle(\omega_* \odot u) = \theta + \measuredangle y_0 \tag{33}$$

for some constant $\theta \in \mathbb{R}$ where u satisfies (21).

In the two-pattern case, the fixed point set is identical to the set $\mathcal{F}$ defined in (28).

Proof. Eq. (33) follows immediately from (26) and Theorem 4.2. Hence for some $p \in \mathbb{R}^N$ with all nonnegative components, $y_* \odot u = e^{i\theta}p \odot y_0$ which is equivalent to

$$y_* = e^{i\theta}p \odot y_0 = e^{i\theta}(y_0 + \eta \odot \omega_0) \tag{34}$$

with

$$\eta = (p - 1) \odot b \in \mathbb{R}^N$$

since the value of u does not matter where y_* vanishes.

In the two-pattern case, we have by (23), (33) and Theorem 4.2 (ii)

$$A\left(2e^{i\theta}y_0 - y_*\right) = e^{i\theta}Ay_0$$

implying

$$e^{i\theta}Ay_0 = Ay_*. \tag{35}$$

From (35) it follows that

$$y_* = e^{i\theta}(y_0 + y'), \quad Ay' = 0,$$

which together with (34) yields $y' = (p - 1) \odot y_0$. $\square$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Note that $\mathcal{F}$ is contained in the much larger set $\mathcal{H} = \cup_\theta \mathcal{H}_\theta$ where

$$\mathcal{H}_\theta := \{e^{i\theta}(b+\eta) \odot \omega_0 : B\eta = 0,\ \eta \in \mathbb{C}^N\} \tag{36}$$

is the affine space of (real) dimension $2(N-n)$ and has the desirable property

$$Ay = e^{i\theta} Ay_0 + B\eta = e^{i\theta} x_0, \quad \forall y \in \mathcal{H}_\theta. \tag{37}$$

The high dimensionality of $\mathcal{H}$ and eq. (37) can be used to speed up the numerical convergence of FDR from a generic initialization (Section 7).

5. Local convergence

To prove local convergence of FDR, it is crucial to analyze the gradient map of (15) and understand its spectral properties.

We focus on FDR (15) with $\mathcal{X} = \mathbb{C}^n$:

$$S_{\mathrm{f}}(y) := y + A^*A\left(2b \odot \frac{y}{|y|} - y\right) - b \odot \frac{y}{|y|}. \tag{38}$$

ODR (18) becomes

$$S(x) = x + \left[\tilde{A}\left(2b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|}\right) - x\right]_n - \tilde{A}\left(b \odot \frac{\tilde{A}^* x}{|\tilde{A}^* x|}\right). \tag{39}$$

Let $y^{(k)} = S_{\mathrm{f}}^{k-1} A^* x^{(1)}$ and $x^{(k)} := Ay^{(k)}, k \in \mathbb{N}$. Define the optimal global phase adjustment at each iteration

$$\alpha^{(k)} := \arg\min_\alpha \{\|\alpha x^{(k)} - x_0\| : |\alpha| = 1, \alpha \in \mathbb{C}\}. \tag{40}$$

Indeed, we have

$$\begin{aligned} \alpha^{(k)} &= x^{(k)*} x_0 / |x^{(k)*} x_0| \\ &= y^{(k)*} y_0 / |y^{(k)*} y_0| \end{aligned} \tag{41}$$

and hence

$$\Im\left(y_0^* \alpha^{(k)} y^{(k)}\right) = 0. \tag{42}$$

Let P_0 denote the projection onto the convex set $\mathcal{F}_0$ defined in (29)

$$P_0 y = \arg\min_{z \in \mathcal{F}_0} \|y - z\| \tag{43}$$

and define

$$y_*^{(k)} = P_0 \alpha^{(k)} y^{(k)} = P_0 \alpha^{(k)} S_{\mathrm{f}}^{k-1} A^* x^{(1)}, \quad k = 1, 2, \cdots \tag{44}$$

Since $y_*^{(k)} \in \mathcal{F}_0$,

$$\Im\left(y_0^* y_*^{(k)}\right) = 0. \tag{45}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Define

$$v^{(k)} = \alpha^{(k)} y^{(k)} - y_*^{(k)}.$$

Combining (42) and (45) we have

$$\langle iy_0, v^{(k)} \rangle = \Re(-iy_0^* v^{(k)}) = \Im(y_0^* v^{(k)}) = 0.$$

In other words,

$$v^{(k)} \perp iy_0$$

or equivalently

$$\Omega_0^* v^{(k)} \perp i|y_0|. \tag{46}$$

Our main result is that the fixed point set $\mathcal{F}$ is geometrically attractive in the vicinity of $e^{i\theta} y_0$ for any $\theta \in \mathbb{R}$.

Theorem 5.1. *Let $x_0 \in \mathbb{C}^n$ and A^* any $N \times n$ isometric matrix with $N \geq 2n$. Let $y_0 = A^* x_0$, $b = |y_0|$ and suppose*

$$b_{\min} = \min_j b_j > 0.$$

Let

$$B := A \operatorname{diag}\left\{ \frac{y_0}{|y_0|} \right\}$$

and suppose

$$\lambda_2 = \max\{\|u\|^{-1} \|\Im(B^* u)\| : u \in \mathbb{C}^n,\ iu \perp x_0\} < 1. \tag{47}$$

Let $v^{(k)} = \alpha^{(k)} y^{(k)} - y_*^{(k)}$ where $\alpha^{(k)}$ and $y_*^{(k)}$ are given by (41) and (44), respectively. For any given $0 < \epsilon < 1 - \lambda_2$, if $\alpha^{(1)} x^{(1)}$ is sufficient close to x_0, then

$$\|v^{(k)}\| \leq (\lambda_2 + \epsilon)^{k-1} \|\alpha^{(1)} x^{(1)} - x_0\|, \quad k = 1, 2, \cdots \tag{48}$$

implying that

$$\|\alpha^{(k)} x^{(k)} - x_0\| \leq (\lambda_2 + \epsilon)^{k-1} \|\alpha^{(1)} x^{(1)} - x_0\|, \quad k = 1, 2, \cdots. \tag{49}$$

Remark 5.2. As mentioned+ in Section 2, y_0 under the measurement scheme (5) vanishes nowhere and hence $b_{\min} > 0$ almost surely.

Remark 5.3. In view of (19), the same error bound (49) holds for the ODR iterates $x^{(k)} = [\tilde{A} y^{(k)}]_n$ with $\tilde{n} = N$.

For $\tilde{n} < N$, however, we are unable to prove local convergence for ODR.

Remark 5.4. If A^* is not isometric, we can apply QR-decomposition to obtain $A^* = QR$, where Q is isometric, and treat Q as the new measurement matrix and Rx_0 as the new unknown.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

5.1. The gradient

First we derive a convenient expression for the gradient map.

Proposition 5.5. *Let $y \in \mathbb{C}^N$ and suppose that $|y|$ vanishes nowhere. Let $\omega = y/|y|$ and $\Omega = \mathrm{diag}(\omega)$. Define*

$$B = A\,\Omega, \quad \eta = \Omega^* v. \tag{50}$$

For a sufficiently small $\epsilon > 0$, we have

$$S_{\mathrm{f}}(y + \epsilon v) - S_{\mathrm{f}}(y) = \epsilon \Omega J_{\mathrm{f}} \eta + o(\epsilon) \tag{51}$$

where

$$J_{\mathrm{f}}\eta = (I - B^*B)\eta + i(2B^*B - I)\ \mathrm{diag}\left[\frac{b}{|y|}\right] \Im(\eta) \tag{52}$$

is the gradient map.

In particular, if $|y| = b$, then (52) becomes

$$J_{\mathrm{f}}\eta = (I - B^*B)\Re(\eta) + iB^*B\Im(\eta) \tag{53}$$

Proof. Let

$$\omega_\epsilon = \frac{y + \epsilon v}{|y + \epsilon v|}, \quad \Omega_\epsilon = \mathrm{diag}(\omega_\epsilon).$$

Reorganizing (38), we have

$$S_{\mathrm{f}}(y) = y - A^*Ay + (2A^*A - I)\Omega b, \tag{54}$$

and hence

$$\begin{aligned} S_{\mathrm{f}}(y + \epsilon v) - S_{\mathrm{f}}(y) &= \epsilon(I - A^*A)v + (2A^*A - I)(\Omega_\epsilon - \Omega)b \\ &= \epsilon(I - \Omega B^* B \Omega^*)v + (2\Omega B^* B \Omega^* - I)(\Omega_\epsilon - \Omega)b \end{aligned} \tag{55}$$

We next give a first order approximation to $(\Omega_\epsilon - \Omega)b$ in terms of v.

Using the first order Taylor expansion we have

$$\omega_\epsilon - \omega = i\Omega\Im\Big[\Omega^*(\omega_\epsilon - \omega)\Big] + o(\epsilon) = i\epsilon\Omega\Im\left[\Omega^*\frac{v}{|y|}\right] + o(\epsilon),$$

and hence

$$(\Omega_\epsilon - \Omega)b = i\epsilon\Omega\,\mathrm{diag}\left[\frac{b}{|y|}\right]\Im(\Omega^* v) + o(\epsilon). \tag{56}$$

Finally, substituting (56) into (55) we obtain

$$S_{\mathrm{f}}(y + \epsilon v) - S_{\mathrm{f}}(y) = \epsilon(I - \Omega B^* B \Omega^*)v + i\epsilon(2\Omega B^* B - \Omega)\,\mathrm{diag}(b/|y|)\Im(\Omega^* v) + o(\epsilon).$$

Multiplying Ω^* on both sides and using the definition of v we complete the proof. □

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Note that $J_{\rm f}$ is a *real*, but *not complex*, linear map since $J_{\rm f}(c\eta) \neq cJ_{\rm f}(\eta), c \in \mathbb{C}$ in general.

Define the real form of the matrix B:

$$\mathcal{B} := \begin{bmatrix} \Re[B] \\ \Im[B] \end{bmatrix} \in \mathbb{R}^{2n,N}. \tag{57}$$

Note that

$$\begin{bmatrix} \Re[B^\top] & \Im[B^\top] \\ -\Im[B^\top] & \Re[B^\top] \end{bmatrix} \tag{58}$$

is real isometric because B^* is complex isometric.

From (12) we have

$$G(B^*\xi) = \begin{bmatrix} \Re[B^\top]\Re[\xi] + \Im[B^\top]\Im[\xi] \\ \Re[B^\top]\Im[\xi] - \Im[B^\top]\Re[\xi] \end{bmatrix} = \begin{bmatrix} \mathcal{B}^\top G(\xi) \\ \mathcal{B}^\top G(-i\xi) \end{bmatrix}, \quad \xi \in \mathbb{C}^n. \tag{59}$$

For the rest of the paper, B denotes the matrix (50) with $\Omega = \Omega_0$, i.e.

$$B = A\Omega_0, \quad \Omega_0 = \mathrm{diag}[\omega_0], \quad \omega_0 = \frac{y_0}{|y_0|} \tag{60}$$

unless otherwise specified.

Next we give a spectral analysis of $J_{\rm f}$. As we will see below, $J_{\rm f}$ can be decomposed according to a sequence of 2-dimensional subspaces of descending eigenvalues. The leading eigenvalue is 1 and its eigenspace contributes only the ambiguity of constant phase factor. With at least one oversampled coded diffraction pattern in the data we can prove the spectral gap property that the second eigenvalue is strictly less than 1. The spectral gap property and the fixed point set $\mathcal{F}$ are the key to the proof of local convergence.

5.2. Eigen structure

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2n} \geq \lambda_{2n+1} = \cdots = \lambda_N = 0$ be the singular values of $\mathcal{B}$ with the corresponding right singular vectors $\{\eta_k \in \mathbb{R}^N\}_{k=1}^N$ and left singular vectors $\{\xi_k \in \mathbb{R}^{2n}\}_{k=1}^{2n}$. By definition, for $k = 1, \cdots, 2n$,

$$B\eta_k = \lambda_k G^{-1}(\xi_k), \tag{61}$$

$$\Re[B^* G^{-1}(\xi_k)] = \lambda_k \eta_k. \tag{62}$$

Proposition 5.6. *We have* $\xi_1 = G(x_0)$, $\xi_{2n} = G(-ix_0)$, $\lambda_1 = 1, \lambda_{2n} = 0$ *as well as* $\eta_1 = |y_0|$.

Proof. Since

$$B^* x_0 = \Omega_0^* A^* x_0 = |y_0|$$

we have by (59)

$$\Re[B^* x_0] = \mathcal{B}^\top \xi_1 = |y_0|, \quad \Im[B^* x_0] = \mathcal{B}^\top \xi_{2n} = 0 \tag{63}$$

and hence the results. □

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Corollary 5.7.

$$\begin{aligned}\lambda_2 &= \max\{\|\Im(B^*u)\| : u \in \mathbb{C}^n, u \perp ix_0, \|u\| = 1\} \\ &= \max\{\|\mathcal{B}^\top u\| : u \in \mathbb{R}^{2n}, u \perp \xi_1, \|u\| = 1\}\end{aligned} \tag{64}$$

Proof. By (59),

$$\Im[B^*u] = \mathcal{B}^\top G(-iu).$$

The orthogonality condition $iu \perp x_0$ is equivalent to

$$G(x_0) \perp G(-iu).$$

Hence, by Proposition 5.6 ξ_2 is the maximizer of the right hand side of (64), yielding the desired value λ_2. □

Proposition 5.8. *For $k = 1, \cdots, 2n$,*

$$\lambda_k^2 + \lambda_{2n+1-k}^2 = 1 \tag{65}$$

$$\xi_{2n+1-k} = G(-iG^{-1}(\xi_k)) \tag{66}$$

$$\xi_k = G(iG^{-1}(\xi_{2n+1-k})). \tag{67}$$

Proof. Since B^* is an isometry, we have $\|w\| = \|B^*w\|, \forall w \in \mathbb{C}^n$. On the other hand, we have

$$\|B^*w\|^2 = \|G(B^*w)\|^2 = \|\mathcal{B}^\top G(w)\|^2 + \|\mathcal{B}^\top G(-iw)\|^2$$

and hence

$$\|G(w)\|^2 = \|\mathcal{B}^\top G(w)\|^2 + \|\mathcal{B}^\top G(-iw)\|^2. \tag{68}$$

Now we prove (65), (66) and (67) by induction.

Recall the variational characterization of the singular values/vectors

$$\lambda_j = \max \|\mathcal{B}^\top u\|, \qquad \xi_j = \arg\max \|\mathcal{B}^\top u\|, \qquad \text{s.t. } u \perp \xi_1, \cdots, \xi_{j-1}, \quad \|u\| = 1 \tag{69}$$

$$\lambda_{2n+1-j} = \min \|\mathcal{B}^\top u\|, \quad \xi_{2n+1-j} = \arg\min \|\mathcal{B}^\top u\|, \quad \text{s.t. } u \perp \xi_{2n}, \cdots, \xi_{2n+2-j}, \|u\| = 1. \tag{70}$$

By Proposition 5.6, (65), (66) and (67) hold for $k = 1$. Suppose (65), (66) and (67) hold for $k = 1, \cdots, j-1$ and we now show that they also hold for $k = j$.

Hence by (68)

$$\lambda_j^2 = \max_{\|u\|=1} \|\mathcal{B}^\top u\|^2 = 1 - \min_{\|v\|=1} \|\mathcal{B}^\top v\|^2, \quad \text{s.t. } u \perp \xi_1, \cdots, \xi_{j-1}, \quad v = G(-iG^{-1}(u)).$$

The condition $u \perp \xi_1, \cdots, \xi_{j-1}$ implies $v \perp \xi_{2n}, \cdots, \xi_{2n+2-j}$ and vice versa. By (70), we have $\lambda_j^2 = 1 - \lambda_{2n+1-j}^2$ and $G(-iG^{-1}(\xi_j))$ is the minimizer, i.e. $\xi_{2n+1-j} = G(-iG^{-1}(\xi_j))$. □

Proposition 5.9. *For each $k = 1, \cdots, 2n$,*

$$B^*B\eta_k = \lambda_k(\lambda_k\eta_k + i\lambda_{2n+1-k}\eta_{2n+1-k}), \tag{71}$$

$$B^*B\eta_{2n+1-k} = \lambda_{2n+1-k}(\lambda_{2n+1-k}\eta_{2n+1-k} - i\lambda_k\eta_k). \tag{72}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Proof. By definition, $\mathcal{B}\eta_k = \lambda_k \xi_k$. Hence

$$B\eta_k = (\Re[B] + i\Im[B])\eta_k = \lambda_k(\xi_k^{\mathrm{R}} + i\xi_k^{\mathrm{I}})$$

where

$$\xi_k = \begin{bmatrix} \xi_k^{\mathrm{R}} \\ \xi_k^{\mathrm{I}} \end{bmatrix}, \quad \xi_k^{\mathrm{R}}, \xi_k^{\mathrm{I}} \in \mathbb{R}^n.$$

On the other hand, $\mathcal{B}^\top \xi_k = \lambda_k \eta_k$ and hence

$$\Re[B^\top]\xi_k^{\mathrm{R}} + \Im[B^\top]\xi_k^{\mathrm{I}} = \lambda_k \eta_k. \tag{73}$$

Now we compute $B^* B\eta_k$ as follows.

$$\begin{aligned} B^*B\eta_k &= \lambda_k B^*(\xi_k^{\mathrm{R}} + i\xi_k^{\mathrm{I}}) \\ &= \lambda_k(\Re[B^\top] - i\Im[B^\top])(\xi_k^{\mathrm{R}} + i\xi_k^{\mathrm{I}}) \\ &= \lambda_k(\Re[B^\top]\xi_k^{\mathrm{R}} + \Im[B^\top]\xi_k^{\mathrm{I}}) + i\lambda_k(\Re[B^\top]\xi_k^{\mathrm{I}} - \Im[B^\top]\xi_k^{\mathrm{R}}) \\ &= \lambda_k^2 \eta_k + i\lambda_k(\Re[B^\top]\xi_k^{\mathrm{I}} - \Im[B^\top]\xi_k^{\mathrm{R}}) \end{aligned} \tag{74}$$

by (73).

Notice that

$$\begin{aligned} \Re(B^\top)\xi_k^{\mathrm{I}} - \Im(B^\top)\xi_k^{\mathrm{R}} &= \mathcal{B}^\top \begin{bmatrix} \Re(-iG^{-1}(\xi_k)) \\ \Im(-iG^{-1}(\xi_k)) \end{bmatrix} \\ &= \mathcal{B}^\top G(-iG^{-1}(\xi_k)) \\ &= \mathcal{B}^\top \xi_{2n+1-k} \\ &= \lambda_{2n+1-k}\eta_{2n+1-k} \end{aligned} \tag{75}$$

by Proposition 5.8.

Putting (74) and (75) together, we have (71). (72) follows from a similar calculation. □

Corollary 5.10. *For $k = 1, 2, \cdots, 2n$, J_{f} leaves invariant the subspace $\mathrm{span}_{\mathbb{R}}\{\eta_k, i\eta_{2n+1-k}\}$ and has the 2×2 matrix representation*

$$J_{\mathrm{f}} = \lambda_{2n+1-k} \begin{bmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{bmatrix}, \quad \lambda_{2n+1-k} := \cos\theta_k, \quad \lambda_k := \sin\theta_k \tag{76}$$

in the basis of $\{\eta_k, i\eta_{2n+1-k}\}$. In particular,

$$J_{\mathrm{f}}\eta_1 = 0, \quad J_{\mathrm{f}} i\eta_1 = i\eta_1 \tag{77}$$

$$J_{\mathrm{f}}\eta_{2n} = \eta_{2n}, \quad J_{\mathrm{f}} i\eta_{2n} = 0 \tag{78}$$

where $\eta_1 = |y_0|$.

Proof. By Proposition 5.9, the span of η_k and $i\eta_{2n+1-k}$ is invariant under B^*B and hence under J_{f} for $k = 1, \cdots, 2n$. Moreover, (71) and (72) imply

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

$$B^*B = \begin{bmatrix} \lambda_k^2 & \lambda_k\lambda_{2n+1-k} \\ \lambda_k\lambda_{2n+1-k} & \lambda_{2n+1-k}^2 \end{bmatrix}$$

in the basis of $\eta_k, i\eta_{2n+1-k}$. Hence by the definition (53) and Proposition 5.8,

$$J_{\rm f} = \lambda_{2n+1-k}\begin{bmatrix} \lambda_{2n+1-k} & \lambda_k \\ -\lambda_k & \lambda_{2n+1-k} \end{bmatrix} = \lambda_{2n+1-k}\begin{bmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{bmatrix}, \quad \theta_k \in \mathbb{R}.$$

Hence $\lambda_{2n+1-k}(\lambda_{2n+1-k} \pm i\lambda_k)$ are eigenvalues of $J_{\rm f}$. □

The next two results completely characterize the eigenstructure of $J_{\rm f}$.

Proposition 5.11. *If $v^*\eta_k = 0, k = 1, 2, \cdots, 2n-1$, then*

$$Bv = 0, \quad J_{\rm f}v = \Re(v). \tag{79}$$

Proof. The condition $v^*\eta_k = 0$ is equivalent to $\eta_k^\top \Re(v) = \eta_k^\top \Im(v) = 0$. So we have

$$G(B\Re(v)) = \begin{bmatrix} \Re(B\Re(v)) \\ \Im(B\Re(v)) \end{bmatrix} = \begin{bmatrix} \Re(B)\Re(v) \\ \Im(B)\Re(v) \end{bmatrix} = \mathcal{B}\Re(v) = 0$$

implying $B\Re(v) = 0$. Likewise, $B\Im(v) = 0$. Hence $Bv = 0$.

By the definition of $J_{\rm f}$ and $Bv = 0$,

$$J_{\rm f}v = (I - B^*B)\Re(v) + iB^*B\Im(v) = \Re(v). \qquad \square$$

Corollary 5.12. *The fixed point set of $J_{\rm f}$ contains the subspace*

$$E_1 = \mathrm{null}_{\mathbb{R}}(\mathcal{B}) \subset \mathbb{R}^N$$

and the null space of $J_{\rm f}$ contains the subspace

$$E_0 = iE_1.$$

Moreover, if $\lambda_2 < 1$, then

$$E_2^\perp = E_0 \oplus_{\mathbb{R}} E_1$$

where

$$E_2 = \mathrm{span}_{\mathbb{R}}\{\eta_k, i\eta_k : k = 1, \cdots, 2n-1\}.$$

Proof. Note that η_{2n} and $i\eta_{2n}$ are excluded from E_2 because $\eta_{2n} \in E_1, i\eta_{2n} \in E_0$. On the other hand the null vector η_1 does not belong in E_0 and the fixed point $i\eta_1$ does not belong in E_1 for an obvious reason.

For any $v \in \mathbb{C}^N$, we can write $v = \Re(v) + i\Im(v)$. By Proposition 5.11, if $\Re(v), \Im(v) \in E_2^\perp$, then

$$\begin{aligned} \mathcal{B}\Re(v) = 0, \quad & J_{\rm f}(\Re(v)) = \Re(v) \\ \mathcal{B}\Im(v) = 0, \quad & J_{\rm f}(\Im(v)) = 0. \end{aligned}$$

In other words, $\Re(v) \in E_1$ and $\Im(v) \in E_0$.

On the other hand, if $\lambda_2 < 1$, then $\lambda_{2n-1} > 0$ and E_2 has no nontrivial intersection with either E_0 or E_1. Hence, $E_2^\perp = E_0 \oplus_{\mathbb{R}} E_1$. □

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

5.3. Proof of Theorem 5.1

Let $\mathcal{C}$ be the real subspace

$$\mathcal{C} = \left\{ \eta \odot \frac{y_0}{|y_0|} : \ \mathcal{B}\eta = 0, \ \eta \in \mathbb{R}^N \right\}, \quad \theta \in (-\pi, \pi].$$

Since every $y \in \mathcal{C}$ has the same phase vector $y/|y| = \pm \odot \omega_0$ with an arbitrary $\pm$ sign at each pixel where η does not vanish,

$$\langle y, iy_0 \rangle = \Re(y^* i y_0) = 0$$

and hence

$$\operatorname{span}_{\mathbb{R}}\{iy_0\} \subseteq \mathcal{C}^{\perp}. \tag{80}$$

Note that $\operatorname{span}_{\mathbb{R}}\{iy_0\}$ is invariant under both S_{f} and the gradient $\Omega_0 J_0 \Omega_0^*$ where

$$J_0 v := (I - B^*B)\Re(v) + iB^*B\Im(v),$$

cf. (53) and (60).

Let $\mathcal{D}_\theta$ be the affine space

$$\mathcal{D}_\theta := e^{i\theta}(y_0 + \mathcal{C}).$$

By (29),

$$\mathcal{F}_\theta \subset \mathcal{D}_\theta. \tag{81}$$

Let V be the open ball in $\mathbb{C}^N$ of radius $b_{\min}$ centered at y_0. Clearly,

$$V \cap \mathcal{F}_0 = V \cap \mathcal{D}_0 \tag{82}$$

and P_0 defined in (43) becomes

$$P_0 y = \arg\min_{z \in \mathcal{D}_0} \|z - y\|, \quad \forall y \in V. \tag{83}$$

Let

$$v^{(k)} = \alpha^{(k)} y^{(k)} - y_*^{(k)}.$$

Proposition 5.13. *If $\alpha^{(k)} y^{(k)} \in V$, then*

$$\Omega_0^* v^{(k)} \in (\operatorname{span}_{\mathbb{R}}\{i\eta_1\})^{\perp} \cap (E_0 \oplus_{\mathbb{R}} E_2) \tag{84}$$

cf. Corollary 5.12.

Proof. By (83) and the definition of $\mathcal{D}_0$, we have

$$v^{(k)} \perp \eta \odot \omega_0, \quad \forall \eta \in \operatorname{null}_{\mathbb{R}}(\mathcal{B}) \subset \mathbb{R}^N$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

or equivalently

$$\Omega_0^* v^{(k)} \perp \eta, \quad \forall \eta \in \mathrm{null}_{\mathbb{R}}(\mathcal{B}) \subset \mathbb{R}^N.$$

This and (46) imply (84) in light of Corollary 5.12. □

Next we prove a technical lemma.

Proposition 5.14. *Let $y_*^{(k)}$ be given by (44). For any $\epsilon_0 > 0$, there exist two positive numbers δ_1, δ_2, such that*

$$\|y_*^{(k)} - y_0\| < \delta_1, \ \|v^{(k)}\| < \delta_2 \tag{85}$$

implies

$$|\alpha^{(k+1)} - \alpha^{(k)}| < \epsilon_0 \|v^{(k)}\|.$$

Proof. With

$$z = \alpha^{(k)} y_0^* y^{(k+1)} = \alpha^{(k)} \overline{\alpha^{(k+1)}} |y_0^* y^{(k+1)}| \tag{86}$$

we can write

$$|\alpha^{(k+1)} - \alpha^{(k)}| = \left| \frac{z}{|z|} - 1 \right| \le \frac{|\Im(z)|}{|\Re(z)|}. \tag{87}$$

Substituting $y^{(k+1)} = S_{\mathrm{f}} y^{(k)}$ into (86) we have

$$\begin{aligned} z &= \alpha^{(k)} y_0^* \left(b \odot \frac{y^{(k)}}{|y^{(k)}|} \right) \\ &= y_0^* \left(b \odot \frac{y_0}{|y_0|} \right) + y_0^* \left[b \odot \left(\alpha^{(k)} \frac{y^{(k)}}{|y^{(k)}|} - \frac{y_*^{(k)}}{|y_*^{(k)}|} \right) \right] \\ &= \|b\|^2 + y_0^* \left[b \odot \left(\alpha^{(k)} \frac{y^{(k)}}{|y^{(k)}|} - \frac{y_*^{(k)}}{|y_*^{(k)}|} \right) \right]. \end{aligned} \tag{88}$$

By the linear approximation, we have

$$\frac{y}{|y|} - \frac{y_*^{(k)}}{|y_*^{(k)}|} = i\Omega_0 \Im \left(\Omega_0^* \left(\frac{y - y_*^{(k)}}{|y_*^{(k)}|} \right) \right) + \mathcal{O}(|y - y_*^{(k)}|^2),$$

and hence for any given ϵ_2 there exists $\delta_2 > 0$ such that

$$\left| y_0^* \left\{ b \odot \left[\alpha^{(k)} \frac{y^{(k)}}{|y^{(k)}|} - \frac{y_*^{(k)}}{|y_*^{(k)}|} - i\Omega_0 \Im \left(\Omega_0^* \frac{v^{(k)}}{|y_*^{(k)}|} \right) \right] \right\} \right| < \epsilon_2 \|b\|^2 \|v^{(k)}\| \tag{89}$$

whenever $\|v^{(k)}\| < \delta_2$. On the other hand, for any $\epsilon_1 > 0$, there exists $\delta_1 > 0$ such that

$$\left| y_0^* \left[b \odot i\Omega_0 \Im \left(\Omega_0^* \left(\frac{v^{(k)}}{|y_*^{(k)}|} - \frac{v^{(k)}}{|y_0|} \right) \right) \right] \right| < \epsilon_1 \|b\|^2 \|v^{(k)}\| \tag{90}$$

whenever $\|y_*^{(k)} - y_0\| < \delta_1$.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Using (42), (45) and the identity $y_0 = \Omega_0 b$, we have

$$y_0^* \left[b \odot i\Omega_0 \Im \left(\Omega_0^* \frac{v^{(k)}}{|y_0|} \right) \right] = i\Im(y_0^* v^{(k)}) = 0$$

and hence from (90)

$$\left| y_0^* \left[b \odot i\Omega_0 \Im \left(\Omega_0^* \frac{v^{(k)}}{|y_*^{(k)}|} \right) \right] \right| < \epsilon_1 \|b\|^2 \|v^{(k)}\|. \tag{91}$$

Combining (89) and (91) we have

$$\left| y_0^* \left[b \odot \left(\alpha^{(k)} \frac{y^{(k)}}{|y^{(k)}|} - \frac{y_*^{(k)}}{|y_*^{(k)}|} \right) \right] \right| < (\epsilon_1 + \epsilon_2) \|b\|^2 \|v^{(k)}\| \tag{92}$$

which implies that

$$|z - \|b\|^2| < (\epsilon_1 + \epsilon_2) \|b\|^2 \|v^{(k)}\|$$

whenever (85) holds. Therefore

$$|\alpha^{(k+1)} - \alpha^{(k)}| \le \frac{(\epsilon_1 + \epsilon_2)\|v^{(k)}\|}{1 - (\epsilon_1 + \epsilon_2)\|v^{(k)}\|} < \epsilon_0 \|v^{(k)}\|$$

for ϵ_1, ϵ_2 sufficiently small. □

We continue the proof with the induction argument. Suppose that (85) holds with $\delta_1 < b_{\min}$. Then the fixed point equation (22) holds with $u \equiv 1$.

By the projection property, we have

$$\begin{aligned} \|v^{(k+1)}\| &\le \|\alpha^{(k+1)} y^{(k+1)} - y_*^{(k)}\| \\ &\le \|y^{(k+1)} - \bar{\alpha}^{(k)} y_*^{(k)}\| + \|(\bar{\alpha}^{(k+1)} - \bar{\alpha}^{(k)}) y_*^{(k)}\| \\ &\le \|\alpha^{(k)} S_{\mathrm{f}}(y^{(k)}) - S_{\mathrm{f}}(y_*^{(k)})\| + |\alpha^{(k+1)} - \alpha^{(k)}| \|y_*^{(k)}\| \end{aligned} \tag{93}$$

since $S_{\mathrm{f}}(y_*^{(k)}) = y_*^{(k)}$. By Proposition 5.14,

$$|\alpha^{(k+1)} - \alpha^{(k)}| \|y_*^{(k)}\| \le \epsilon_0 \|v^{(k)}\| \|y_*^{(k)}\|. \tag{94}$$

On the other hand, by the linear approximation,

$$\|\alpha^{(k)} S_{\mathrm{f}}(y^{(k)}) - S_{\mathrm{f}}(y_*^{(k)})\| \le \|\Omega_0 J_k \Omega_0^* v^{(k)}\| + o(\|v^{(k)}\|) \tag{95}$$

where

$$J_k \eta = (I - B^*B)\eta + i(2B^*B - I) \operatorname{diag}\left[b/|y_*^{(k)}| \right] \Im(\eta) \tag{96}$$

as given in (52).

To bound the right hand side of (95), we infer from the continuity of the gradient (52) at y_0 that for any $\epsilon_1 > 0$,

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

$$\|J_k - J_0\| < \epsilon_1 \tag{97}$$

if δ_1 is sufficiently small. Moreover, as a result of Proposition 5.13 and Corollary 5.12,

$$\Omega_0^* v^{(k)} \in \mathrm{span}_{\mathbb{R}}\{\eta_1\} \oplus_{\mathbb{R}} \mathrm{span}_{\mathbb{R}}\{\eta_k, i\eta_k : k = 2, \cdots 2n-1\} \oplus_{\mathbb{R}} E_0,$$

and hence by Corollary 5.10,

$$\|\Omega_0 J_0 \Omega_0^* v^{(k)}\| \le \lambda_2 \|v^{(k)}\|. \tag{98}$$

The estimates (95), (97) and (98) then imply

$$\|\alpha^{(k)} S_{\mathrm{f}}(y^{(k)}) - S_{\mathrm{f}}(y_*^{(k)})\| \le \lambda_2 \|v^{(k)}\| + \epsilon_1 \|v^{(k)}\| + o(\|v^{(k)}\|). \tag{99}$$

It then follows from (93), (94) and (99)

$$\|v^{(k+1)}\| < \lambda_2 \|v^{(k)}\| + \epsilon_1 \|v^{(k)}\| + o(\|v^{(k)}\|) + \epsilon_0 \|y_*^{(k)}\| \|v^{(k)}\|. \tag{100}$$

Therefore for any ϵ with $\epsilon_1 + \epsilon_0 \|y_0\| < \epsilon < 1 - \lambda_2$ there exist δ_1, δ_2 such that

$$0 < \delta_2 < \sqrt{b_{\min}^2 - \delta_1^2}$$

and

$$\|v^{(k+1)}\| < (\lambda_2 + \epsilon) \|v^{(k)}\|. \tag{101}$$

To iterate (101), we claim that (85) holds with k replaced by $k+1$.

Let us postpone the proof of claim at the end and continue the proof of Theorem 5.1.

By choosing $\lambda_2 + \epsilon < 1$ and

$$\|\alpha^{(1)} y^{(1)} - y_*^{(1)}\| \le \|\alpha^{(1)} y^{(1)} - y_0\| < \delta = \min\{\delta_1, \delta_2\} < b_{\min}$$

we have from (101)

$$\|v^{(k)}\| < (\lambda_2 + \epsilon)^{k-1} \delta, \quad \forall k \in \mathbb{N}. \tag{102}$$

To prove the convergence of $\alpha^{(k)} x^{(k)}$ to x_0, consider the identities

$$\begin{aligned} A v^{(k)} &= A\alpha^{(k)} y^{(k)} - A y_*^{(k)} \\ &= \alpha^{(k)} A y^{(k)} - x_0 + A(\eta \odot \omega_0), \quad \text{some } \eta \in \mathrm{null}_{\mathbb{R}}(\mathcal{B}) \\ &= \alpha^{(k)} A y^{(k)} - x_0 \\ &= \alpha^{(k)} x^{(k)} - x_0 \end{aligned}$$

and hence

$$\|\alpha^{(k)} x^{(k)} - x_0\| \le \|v^{(k)}\| < (\lambda_2 + \epsilon)^{k-1} \delta, \quad \forall k \in \mathbb{N}.$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Now we return to the proof of the claim that if (85) holds then

$$\|y_*^{(k+1)} - y_0\| < \delta_1 \tag{103}$$

$$\|v^{(k+1)}\| < \delta_2 \tag{104}$$

for sufficiently small but fixed $\delta_1 > 0$ and $\delta_2 > 0$, respectively.

First note that if (103) holds then so does (104) by using (101) and setting

$$\|v^{(1)}\| < \delta_3 < \delta_2$$

where δ_3 is specified below.

By the projection property and repeating the calculation (93)–(101), we have

$$\|y_*^{(k+1)} - y_*^{(k)}\| \le \|\alpha^{(k+1)} y^{(k+1)} - y_*^{(k)}\| < (\lambda_2 + \epsilon)\|v^{(k)}\|.$$

To extend (103) from $\{1, 2, \cdots, k\}$ to $k+1$, we write

$$\|y_*^{(k+1)} - y_0\| \le \|y_*^{(k)} - y_0\| + \|y_*^{(k+1)} - y_*^{(k)}\|.$$

Iterating this inequality backward, we have

$$\begin{aligned}
\|y_*^{(k+1)} - y_0\| &\le \|y_*^{(1)} - y_0\| + \sum_{j=1}^{k} \|y_*^{(j+1)} - y_*^{(j)}\| \\
&< \|y_*^{(1)} - y_0\| + (\lambda_2 + \epsilon) \sum_{j=1}^{k} \|v^{(j)}\| \\
&< \|y_*^{(1)} - y_0\| + \|v^{(1)}\| \sum_{j=1}^{k} (\lambda_2 + \epsilon)^j
\end{aligned}$$

by using (101). Hence

$$\begin{aligned}
\|y_*^{(k+1)} - y_0\| &< \|y_*^{(1)} - y_0\| + \|v^{(1)}\| \times \frac{\lambda_2 + \epsilon}{1 - \lambda_2 - \epsilon} \\
&< \|y_*^{(1)} - y_0\| + \frac{\delta_3(\lambda_2 + \epsilon)}{1 - \lambda_2 - \epsilon}, \quad \forall k.
\end{aligned}$$

The proof of the claim is completed upon choosing δ_3 so small that

$$\frac{\delta_3(\lambda_2 + \epsilon)}{1 - \lambda_2 - \epsilon} < \delta_1/2, \quad \delta_3 < \delta_2$$

and $\alpha^{(1)} x^{(1)}$ so close to x_0 that

$$\|y_*^{(1)} - y_0\| \le \|\alpha^{(1)} y^{(1)} - y_0\| < \delta_1/2 < b_{\min}/2.$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

6. Spectral gap

In this section, we prove the spectral gap condition (47) with at least one oversampled coded diffraction pattern. This is the immediate consequence of the following two results.

Proposition 6.1. *Let A^* be isometric and $B = A\Omega_0$. Then $\|\Im(B^*x)\| = 1$ holds for some unit vector x if and only if x satisfies the equation*

$$\Re(a_j^*x)\Re(a_j^*x_0) + \Im(a_j^*x)\Im(a_j^*x_0) = 0,\ \forall j = 1, \cdots N, \tag{105}$$

where a_j are the columns of A, or equivalently

$$\frac{A^*x}{|A^*x|} = \sigma \odot \omega_0 \tag{106}$$

where the components of σ are either 1 or -1, i.e.

$$\sigma(j) \in \{1, -1\}, \quad \forall j = 1, \cdots N.$$

Proof. We have

$$\begin{aligned}\Im(B^*x) &= \Im\left(\frac{\overline{A^*x_0}}{|A^*x_0|} \odot A^*x\right)\\ &= \sum_{j=1}^{N} \frac{\Re(a_j^*x_0)\Im(a_j^*x) - \Im(a_j^*x_0)\Re(a_j^*x)}{(\Re^2(a_j^*x_0) + \Im^2(a_j^*x_0))^{1/2}}\end{aligned} \tag{107}$$

and hence

$$\|\Im(B^*x)\|^2 \le \sum_{j=1}^{N} \Re^2(a_j^*x) + \Im^2(a_j^*x) = \sum_{j=1}^{N} |a_j^*x|^2 = \|A^*x\|^2 = \|x\|^2$$

by the Cauchy–Schwartz inequality and the isometry of A^*.

In view of (107), the inequality becomes an equality if and only if (105) or (106) holds. □

Proposition 6.2 *(Uniqueness of Fourier magnitude retrieval). Let x_0 be a given rank ≥ 2 object and μ be continuously and independently distributed on the unit circle. Let A^* be given by (4).*

If the phase equation

$$\measuredangle A^*\hat{x} = \pm \measuredangle A^*x_0 \tag{108}$$

holds where the $\pm$ sign is arbitrary pixel-by-pixel, then almost surely $\hat{x} = cx_0$ for some constant $c \in \mathbb{R}$.

The proof of Proposition 6.2 is given in Appendix A.

Now we can prove the spectral gap theorem needed for geometric convergence of FDR.

Theorem 6.3. *Let Φ be the oversampled discrete Fourier transform. Let x_0 be a rank ≥ 2 object and at least one of $\mu_j, j = 1, \cdots, \ell \ge 2$, be independently and continuously distributed on the unit circle. Let*

$$A^* = c\begin{bmatrix}\Phi\ \mathrm{diag}\{\mu_1\}\\ \cdots\\ \Phi\ \mathrm{diag}\{\mu_\ell\}\end{bmatrix} \tag{109}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

be isometric with a proper choice of c and

$$B := A \operatorname{diag}\left\{\frac{A^* x_0}{|A^* x_0|}\right\}.$$

Then with probability one

$$\|\Im(B^* u)\| = 1, \quad \|u\| = 1 \quad \textit{iff} \quad u = \pm i x_0/\|x_0\| \tag{110}$$

and hence

$$\lambda_2 = \max\{\|\Im(B^* u)\| : u \in \mathbb{C}^n, u \perp i x_0, \|u\| = 1\} < 1. \tag{111}$$

Proof. Note that the proof of Proposition 6.1 depends only on the fact that A^* is isometric and hence holds for at least one coded diffraction pattern, oversampled or not.

Also, the uniqueness theorem, Proposition 6.2, clearly holds as long as there is at least one oversampled coded diffraction pattern.

Now Proposition 6.1 says that (110) holds if (108) has a unique solution up to a real constant and Proposition 6.2 says that (108) indeed has a unique solution up to a real constant. The proof is complete. □

We have the following corollary from Theorems 5.1 and 6.3.

Corollary 6.4. *Under the assumptions of Theorem 6.3, the geometric convergence (49) holds for phase retrieval with (109) as the propagation matrix.*

7. Numerical experiments

The performance metric is given by the relative error (RE) of the estimate $\hat{x}$ with the optimal phase adjustment $\hat{\alpha}$:

$$\frac{\|\hat{\alpha}\hat{x} - x_0\|}{\|x_0\|}, \quad \hat{\alpha} = \frac{\hat{x}^* x_0}{|\hat{x}^* x_0|}. \tag{112}$$

To identify the geometric convergence regime, we look for the straight-line portion of the *semi-log* plot of RE versus iteration. On the other hand, to identify the power-law convergence regime, we look for the straight-line portion of the *log–log* plot of RE versus iteration.

The stopping rule of the iteration is given by a thresholding rule based on the relative residual (RR)

$$\frac{\||A^* \hat{x}| - b\|}{\|b\|}. \tag{113}$$

7.1. Test images

For test images x_0 we consider the Randomly Phased Phantom (RPP) Fig. 2 (left) and the deterministic image, hereby called the Truncated Cameraman–Barbara (TCB), whose real part is the truncated cameraman, Fig. 2 (middle) and whose imaginary part is the truncated Barbara, Fig. 2 (right). The purpose of truncation is to create an (unknown) loose support (dark margins) which makes the image more difficult to recover. RPP has a loose support without additional truncation. Likewise, we randomize the original phantom in order to make its reconstruction more challenging. Based on our experience, a random object such as RPP is more difficult to recover than a deterministic object such as TCB (see, e.g. Fig. 8). The size n of both images is 256×256, including the margins.

The propagation matrix is primarily based on either (4) or (6) unless specified otherwise.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Fig. 2. The original phantom without phase randomization (left), the truncated cameraman (middle) and the truncated Barbara (right).

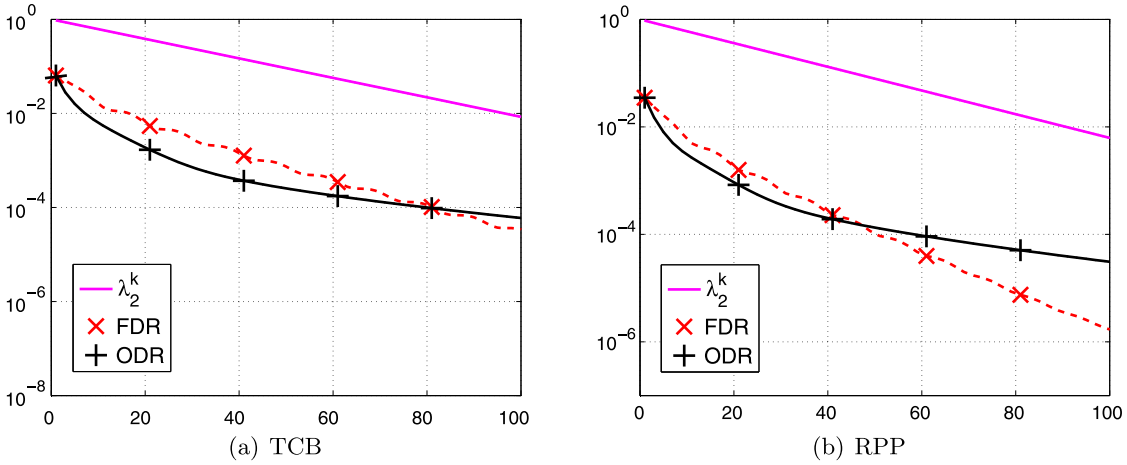


Fig. 3. Semi-log plot of RE versus iteration for (a) TCB and (b) RPP.

7.2. Convergence rate

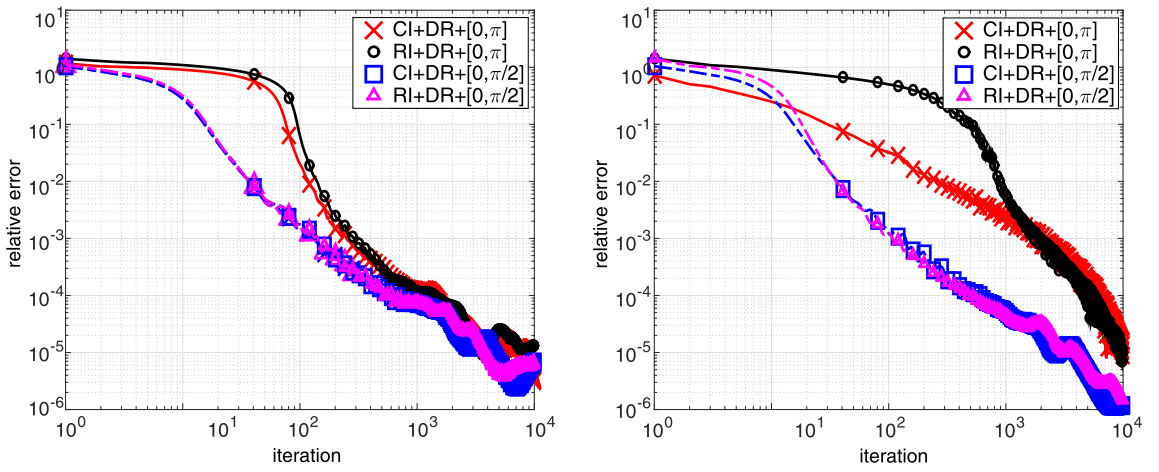
First we simulate the local convergence rate of the $1\frac{1}{2}$-mask case and compare them with λ_2.

The initial condition $x^{(1)}$ is chosen sufficiently close to the true object x_0, which is a unit vector. Fig. 3 shows the error $\|\alpha^{(k)}[x^{(k)}]_n - x_0\|$ on the log scale versus the iteration counter in the case of two oversampled diffraction patterns. The solid straight line represents the geometric sequence $\{\lambda_2^k\}_{k=1}^{100}$. The λ_2 value is computed via the power method, $\lambda_2 = 0.9505$ for TCB and $\lambda_2 = 0.9533$ for RPP. Note that the FDR curve decays slightly faster than the λ_2-curve, which decays still faster than the ODR curve (with $\tilde{n} \approx 4n$).

7.3. Initialization

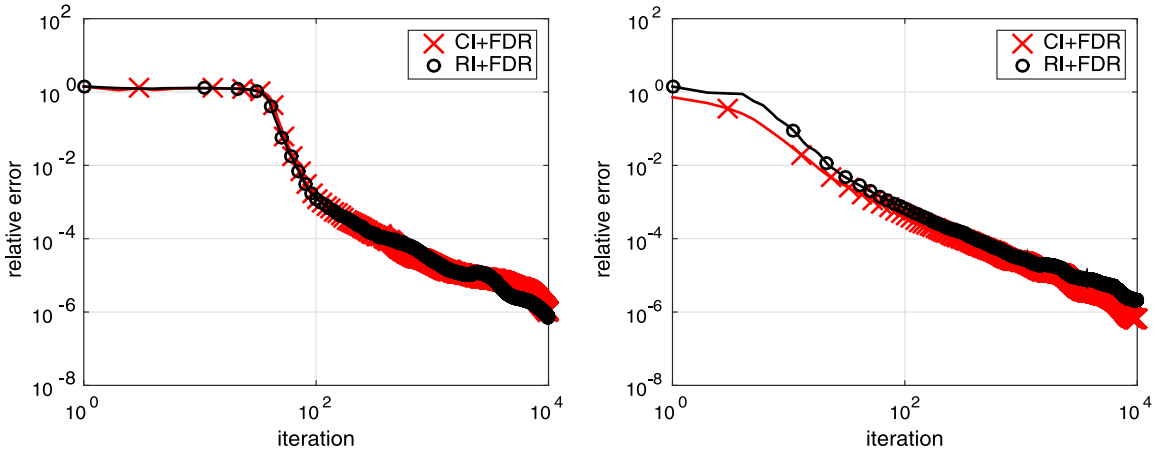
For *global* convergence behaviors, we test two different initializations: the Constant Initialization (CI), $x^{(1)} = c$, where c is a positive constant, and the Random Initialization (RI), $x^{(1)} = ce^{i\xi}$, where the components of $\xi \in \mathbb{R}^n$ are i.i.d. uniform random variables over $(-\pi, \pi]$. Note that the value c does not affect the end result of the iteration as $S_{\mathrm{f}}(cy) = S_{\mathrm{f}}(y)$ for any $y = A^*x$ and $c > 0$.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003



(a) RPP (b) TCB

Fig. 4. Log–log plot of RE versus iteration in the 1-pattern case with two different sector constraints: $[0, \pi/2]$ and $[0, \pi]$.



(a) RPP (b) TCB

Fig. 5. Log–log plot of RE versus iteration in the $1\frac{1}{2}$-mask case.

7.4. One-pattern case

Fig. 4 is the log–log plot of RE with one coded diffraction pattern and two different sector conditions (in this case FDR and ODR are equivalent as $N = \tilde{n}$).

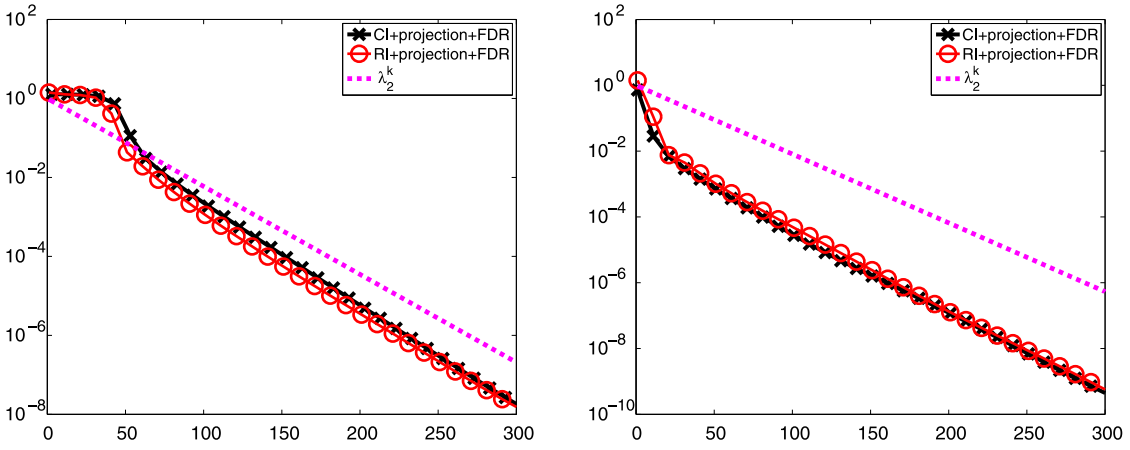
To test the effect of the sector constraint, the phase of RPP is uniformly distributed in two different intervals: $[0, \pi/2]$ and $[0, \pi]$. While FDR/ODR converges globally regardless of the initialization, the rate of convergence decreases as the sector enlarges. When the sector constraint is absent, the iteration ceases to converge in general. The nearly straight tail of the log–log plot suggests power-law decay between $k^{-1.5}$ and k^{-2}.

7.5. $1\frac{1}{2}$-Mask case

With two (or more) diffraction patterns, we let the phase of RPP be uniformly distributed in $(-\pi, \pi]$ (i.e. no sector constraint).

Like Fig. 4, the nearly straight tail of the log–log plot in Fig. 5 indicates a power-law decay roughly like $k^{-1.25}$.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003



(a) RPP (b) TCB

Fig. 6. Semi-log plot of RE versus iteration with one application of A^*A in the $1\frac{1}{2}$-mask case.

The power-law behavior corresponds to the convergence of $y^{(k)}$ to $\mathcal{H}_\theta$ given in (36). Because $A^*A\mathcal{H}_\theta = e^{i\theta}y_0$, we can enforce the geometric convergence regime, and thus speed up convergence, by applying the projection A^*A as in

$$y^{(k+1)} = S_{\mathrm{f}}(A^*Ay^{(k)}) \tag{114}$$

when, and only when, the relative residual of $x^{(k)} = Ay^{(k)}$ is sufficiently small. Fig. 6 bears this out nicely, showing RE less than 10^{-10} in less than 400 iterations on the semi-log scale, a vast improvement over the power-law convergence in Fig. 5.

7.6. Noisy data

When noise ϵ is present in the data we would like to know how RE varies with the noise-to-signal ratio (NSR)

$$\mathrm{NSR} = \frac{\|\epsilon\|}{\|A^*x_0\|}.$$

Since high-precision reconstruction is not possible with noisy data, we apply FDR without the extra projection A^*A.

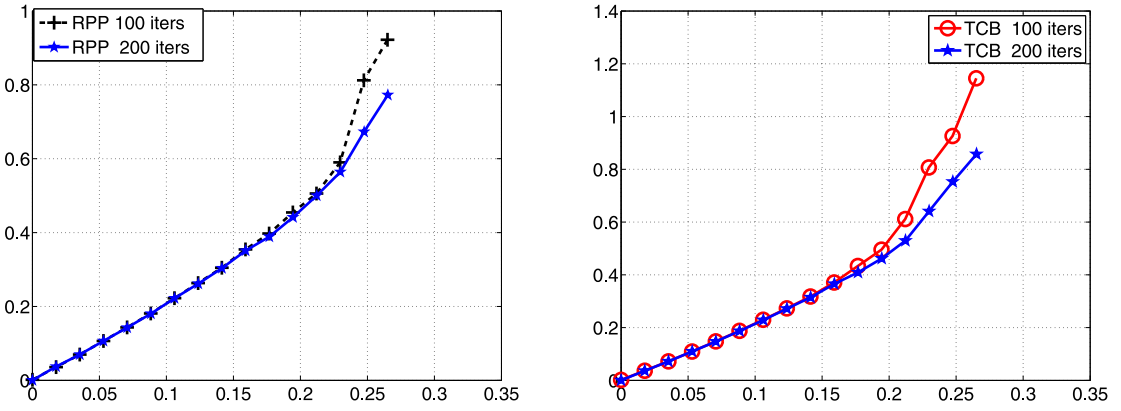
Fig. 7 shows the linear scale plot of RE versus NSR with the maximum number of iterations set to 100 and 200. For NSR$\leq$ 20% the result is approximately a straight line with slope $\approx$ 2.2, independent of the number of iterations. Increasing the number of iterations reduces the error and extends the straight line regime to higher NSRs.

7.7. Multi-mask case

To test how DR performs in the setting of multiple patterns without oversampling [15,16] we simulate the 3-pattern and 4-pattern cases with the propagation matrices given by

$$A^* = c\begin{bmatrix} \Phi\ \mathrm{diag}\{\mu_1\} \\ \cdots \\ \Phi\ \mathrm{diag}\{\mu_{\ell-1}\} \\ \Phi \end{bmatrix}, \quad \ell = 3, 4, \tag{115}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003



(a) RPP (b) TCB

Fig. 7. Linear scale plot of RE versus NSR in the $1\frac{1}{2}$-mask case with the maximum number of iterations set to 100 or 200.

where Φ is the standard (unoversampled) DFT.

Fig. 8 shows the semi-log plot of RE versus iteration with three patterns (a)–(b) and four patterns (c)–(d), both without oversampling. The projection A^*A is inserted into FDR once the RR falls below 1%, resulting in the geometric convergence regime. Clearly, ODR performs poorly with RPP while FDR performs well for both images, independent of the initialization. Going from three patterns ($N = 3|\mathcal{M}|$) to four patterns ($N = 4|\mathcal{M}|$) reduces the number of iterations by about half to achieve the same level of accuracy. Note that the number of data with four patterns ($N = 4|\mathcal{M}|$) is half of that with 2 oversampled patterns ($N \approx 8|\mathcal{M}|$) and Theorem 5.1 applies when the number of unoversampled patterns is at least two (i.e. $N \geq 2|\mathcal{M}|$) and the spectral gap condition (47) holds.

7.8. Padding ratio

Finally we test the effect of the padding ratio $\tilde{n}/n$ on the performance of ODR. For each $\tilde{n}/n \in [4, 8]$, we conduct 50 trials with independent, random initializations and average the REs. Recall that $\tilde{n}/n = 4$ is the standard padding rate and at $\tilde{n}/n = 8$ ODR is equivalent to FDR.

Fig. 9 shows the averaged RE versus the ratio $\tilde{n}/n$, demonstrating the phase transition from large RE at $\tilde{n}/n = 4$ to small RE at $\tilde{n}/n = 8$ (FDR). The phase transition depends on the number of iterations employed. As the number of iterations increases, the threshold ratio decreases.

8. Conclusion and discussion

FDR is a natural formulation of DR for phase retrieval. We have proved for the first time the local geometric convergence for FDR in the case of two (or more) oversampled diffraction patterns with the rate closely related to a spectral gap condition.

While we are unable to prove global convergence, the numerical experiments have strongly suggested the power-law convergence for both constant and randomly phased initializations. On the other hand, if the projection operator A^*A is inserted into the FDR iteration once the relative residual falls below a certain threshold indicating proximity to the set $\mathcal{H}$, the projected iterate enters the basin of attraction of $e^{i\theta}y_0$ and the geometric convergence regime is restored, giving rise to highly accurate reconstruction.

This scenario continues to hold for FDR with 3 and more coded diffraction patterns without oversampling, with each additional coded diffraction pattern resulting in a lower number of iteration for the same level of accuracy.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

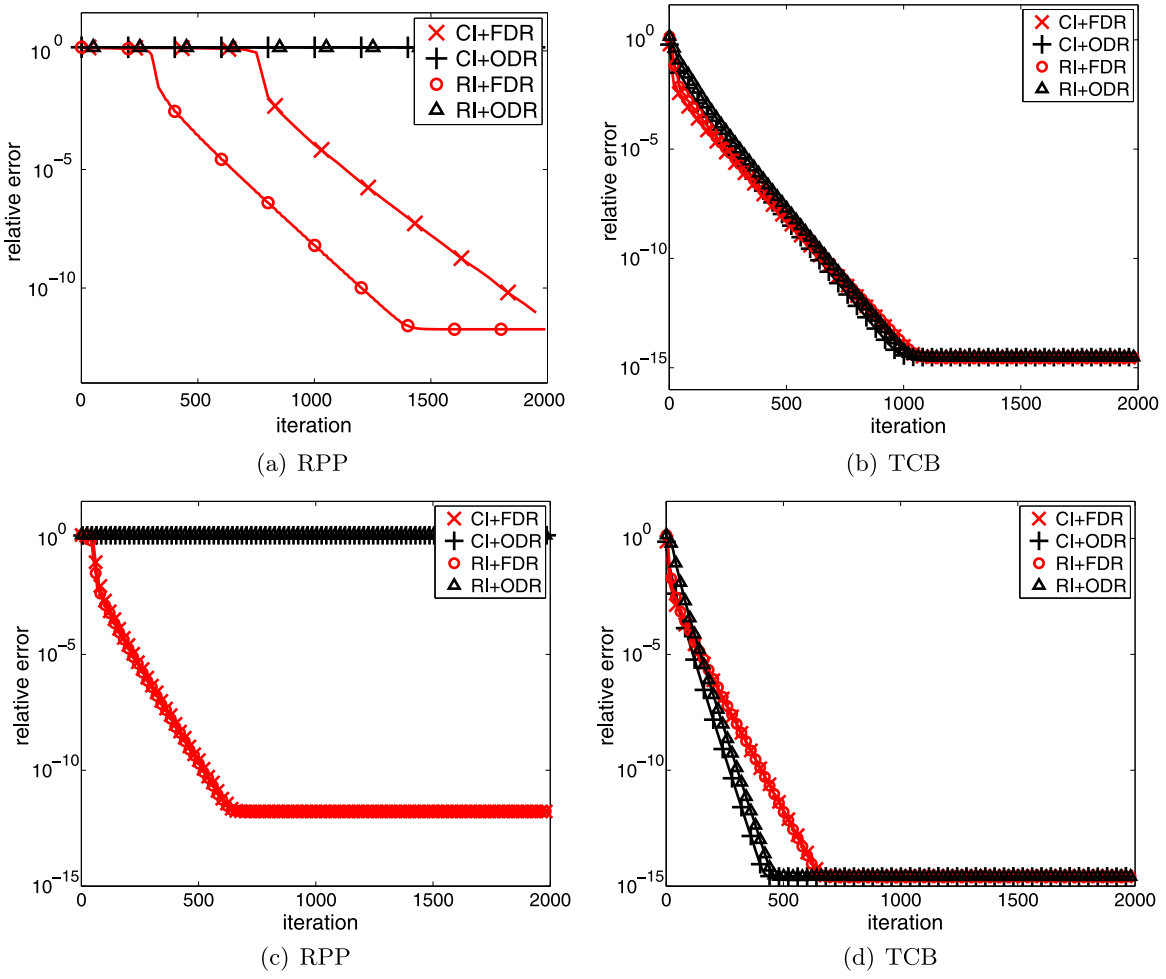


Fig. 8. Semi-log plot of RE with one application of A^*A. The number of diffraction patterns without oversampling is 3 for (a)–(b) and 4 for (c)–(d).

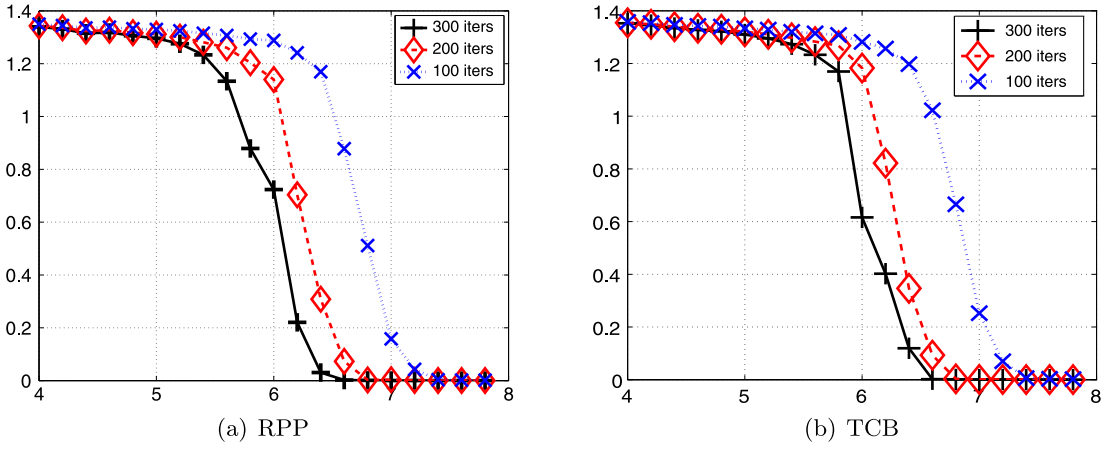


Fig. 9. Linear scale plot of RE versus $\tilde{n}/n$ for (a) RPP and (b) TCB with various numbers of iteration.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

Acknowledgments

The research of P. Chen is supported in part by the grant 103-2115-M-005-006-MY2 from Ministry of Science and Technology, Taiwan, and the grant U01-HL-114494 from National Heart, Lung, and Blood Institute, USA. The research of A. Fannjiang is supported in part by the US National Science Foundation grant DMS-1413373 and Simons Foundation grant 275037.

Appendix A. Proof of Proposition 6.2

In order to prove the uniqueness theorem for Fourier magnitude retrieval, we need to take up the more elaborate notation in Section 1.1.

Let

$$F(\mathbf{z}) = \sum_{\mathbf{n}} f(\mathbf{n})\mathbf{z}^{-\mathbf{n}}$$

be the z-transform of f. According to the fundamental theorem of algebra, $F(\mathbf{z})$ can be written uniquely as

$$F(\mathbf{z}) = \alpha \mathbf{z}^{-\mathbf{n}_0} \prod_{k=1}^{p} F_k(\mathbf{z}), \tag{116}$$

where $\mathbf{n}_0$ is a vector of nonnegative integers, α is a complex coefficient, and $F_k(\mathbf{z})$ are nontrivial irreducible polynomials in $\mathbf{z}^{-1}$.

Define the shift

$$f_{\mathbf{m}+}(\cdot) = f(\mathbf{m}+\cdot), \quad f_{\mathbf{m}-}(\cdot) = f(\mathbf{m}-\cdot).$$

Conjugate Symmetry. A polynomial $X(\mathbf{z})$ in $\mathbf{z}^{-1}$ is said to be conjugate symmetric if, for some vector $\mathbf{k}$ of positive integers and some $\theta \in (-\pi, \pi)$,

$$X(\mathbf{z}) = e^{i\theta}\mathbf{z}^{-\mathbf{k}}\overline{X(\bar{\mathbf{z}}^{-1})}.$$

In other words, the ratio between $X(\mathbf{z})$ and its conjugate inversion is a monomial in $\mathbf{z}^{-1}$ times a complex number of unit modulus.

A conjugate symmetric polynomial may be reducible, irreducible, trivial, or nontrivial. Any monomial $\mathbf{z}^{-\mathbf{k}}$ is conjugate symmetric.

Proposition A.1. *Suppose that the z-transform F of f has no conjugate symmetric factors. If the z-transform G of g satisfies $\measuredangle F(e^{2\pi i\mathbf{w}}) - \measuredangle G(e^{2\pi i\mathbf{w}}) \in \{0, \pi\}$, $\forall \mathbf{w} \in \mathcal{L}$ (defined in (3)) then $g = cf$ for some constant $c \in \mathbb{R}$.*

The real-valued version of the above proposition is given in [38]. For the reader's convenience, we provide the proof for the complex setting (see also [30], Appendix B, Proposition 5).

Proof. Let $\star$ denote the convolution operator. Consider

$$h = f \star \overline{g(-\cdot)}$$

whose z-transform is

$$H(\mathbf{z}) = F(\mathbf{z})\overline{G(\bar{\mathbf{z}}^{-1})}.$$

Note that h is defined on $\widetilde{\mathcal{M}}$, instead of $\mathcal{M}$, so $H(\mathbf{z})$ is completely determined by sampling H on $\mathcal{L}$.

Since

$$\measuredangle H(e^{2\pi i\mathbf{w}}) = \measuredangle F(e^{2\pi i\mathbf{w}}) - \measuredangle G(e^{2\pi i\mathbf{w}})$$

it follows $H(e^{2\pi i\mathbf{w}})$ is real-valued. By analytic continuation, we have

$$H(\mathbf{z}) = \overline{H(\bar{\mathbf{z}}^{-1})}$$

and

$$F(\mathbf{z})\overline{G(\bar{\mathbf{z}}^{-1})} = \overline{F(\bar{\mathbf{z}}^{-1})}G(\mathbf{z}). \tag{117}$$

Multiplying both sides of (117) by $\mathbf{z}^{-\mathbf{M}}$ results in the following polynomial equation in $\mathbf{z}^{-1}$:

$$F(\mathbf{z})\overline{G(\bar{\mathbf{z}}^{-1})}\mathbf{z}^{-\mathbf{M}} = \mathbf{z}^{-\mathbf{M}}\overline{F(\bar{\mathbf{z}}^{-1})}G(\mathbf{z}). \tag{118}$$

We observe $\mathbf{n}_0 = 0$ in view of (116) and the assumption that $F(\mathbf{z})$ has no conjugate symmetric factor. We also have

$$\mathbf{z}^{-\mathbf{M}}\overline{F(\bar{\mathbf{z}}^{-1})} = \tilde{\alpha}\mathbf{z}^{-\mathbf{n}_1}\prod_k \tilde{F}_k(\mathbf{z}), \tag{119}$$

where $\tilde{F}_k(\mathbf{z})$ are the nontrivial irreducible non-conjugate symmetric polynomials in $\mathbf{z}^{-1}$ of the form $\tilde{F}_k(\mathbf{z}) = \mathbf{z}^{-\mathbf{M}+\mathbf{p}_k}\overline{F_k(\bar{\mathbf{z}}^{-1})}$ for some vector $\mathbf{p}_k$ of positive integers.

Writing

$$G(\mathbf{z}) = \beta\mathbf{z}^{-\mathbf{m}_0}\prod_\ell G_\ell(\mathbf{z}), \tag{120}$$

where $G_\ell(\mathbf{z})$ are nontrivial irreducible polynomials in $\mathbf{z}^{-1}$, we have

$$\mathbf{z}^{-\mathbf{M}}\overline{G(\bar{\mathbf{z}}^{-1})} = \tilde{\beta}\mathbf{z}^{-\mathbf{m}_1}\prod_\ell \tilde{G}_\ell(\mathbf{z}), \tag{121}$$

where $\tilde{G}_\ell(\mathbf{z})$ are the nontrivial irreducible polynomials in $\mathbf{z}^{-1}$ of the form $\tilde{G}_\ell(\mathbf{z}) = \mathbf{z}^{-\mathbf{M}+\mathbf{q}_\ell}\overline{G_\ell(\bar{\mathbf{z}}^{-1})}$ for some vector $\mathbf{q}_\ell$ of positive integers.

Plugging (116), (119), (120) and (121) in (118) yields

$$\alpha\tilde{\beta}\mathbf{z}^{-\mathbf{m}_1}\prod_k F_k(\mathbf{z})\prod_\ell \tilde{G}_\ell(\mathbf{z}) = \tilde{\alpha}\beta\mathbf{z}^{-\mathbf{n}_1-\mathbf{m}_0}\prod_k \tilde{F}_k(\mathbf{z})\prod_\ell G_\ell(\mathbf{z}). \tag{122}$$

Each nontrivial irreducible factor $F_k(\mathbf{z})$ must be equal to some $\tilde{F}_{k'}(\mathbf{z})$ or some $G_{\ell'}(\mathbf{z})$. However, if $F_k(\mathbf{z}) = \tilde{F}_k(\mathbf{z})$, then $F_k(\mathbf{z})$ is a conjugate symmetric factor. If, on the other hand, $F_k(\mathbf{z}) = \tilde{F}_{k'}(\mathbf{z})$ for some $k' \neq k$, then $F_k(\mathbf{z})F_{k'}(\mathbf{z}) = \tilde{F}_{k'}(\mathbf{z})\tilde{F}_k(\mathbf{z})$ is a conjugate symmetric factor. Both cases, however, are excluded by the assumption that the z-transform of f does not have conjugate symmetric factors.

Hence each F_k (rest. $\tilde{F}_k$) must be equal to some G_ℓ (rest. $\tilde{G}_\ell$) and we can write

$$G(\mathbf{z}) = Q(\mathbf{z})F(\mathbf{z}) \tag{123}$$

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

where $Q(\mathbf{z})$ is a polynomial in $\mathbf{z}^{-1}$, i.e.

$$Q(\mathbf{z}) = \sum_{\mathbf{n}\geq 0} c_{\mathbf{n}}\mathbf{z}^{-\mathbf{n}}.$$

By the assumption that $\measuredangle F(e^{2\pi i\mathbf{w}}) - \measuredangle G(e^{2\pi i\mathbf{w}}) \in \{0,\pi\}$ we have $Q(e^{2\pi i\mathbf{w}}) \in \mathbb{R}, \forall \mathbf{w} \in \mathcal{L}$, and hence $\bar{c}_{\mathbf{n}} = c_{-\mathbf{n}} = 0$ except for $\mathbf{n} = 0$ in which case $c_0 \in \mathbb{R}$. Therefore, $Q = c_0 \in \mathbb{R}$ and this is what we start out to prove. □

Proposition A.2 *([28]). Let x_0 have rank ≥ 2. Let $\{\mu(\mathbf{n})\}$ be independent and continuous random variables on the unit circle of the complex plane. Then, the z-transform $F(\mathbf{z})$ of $f(\mathbf{n}) := \mu(\mathbf{n})x_0(\mathbf{n})$ is irreducible up to a power of $\mathbf{z}^{-1}$ with probability one.*

For the proof of Proposition A.2 see Theorem 2 of [28].

We next show that the z-transform of $\{\mu(\mathbf{n})x_0(\mathbf{n})\}$ is almost surely irreducible up to a power $\mathbf{z}^{-1}$ and not conjugate symmetric.

Proposition A.3. *Let $\{\mu(\mathbf{n})\}$ be independent and continuous random variables on the unit circle of the complex plane. Let $f(\mathbf{n}) := \mu(\mathbf{n})x_0(\mathbf{n})$. Then the z-transforms of both $f_{\mathbf{t}+}$ and $\overline{f_{\mathbf{t}-}}$ are almost surely not conjugate symmetric $\forall$ $\mathbf{t}$.*

Proof. The z-transform

$$F_{\mathbf{t}+}(\mathbf{z}) = \sum_{\mathbf{n}} f(\mathbf{t}+\mathbf{n})\mathbf{z}^{-\mathbf{n}}. \tag{124}$$

is conjugate symmetric if

$$F_{\mathbf{t}+}(\mathbf{z}) = e^{i\theta}\mathbf{z}^{-\mathbf{k}}\overline{F_{\mathbf{t}+}(\bar{\mathbf{z}}^{-1})} \tag{125}$$

for some vector $\mathbf{k}$ of positive integers and some $\theta \in (-\pi,\pi)$. Plugging (124) in (125) yields

$$\sum_{\mathbf{n}} f(\mathbf{t}+\mathbf{n})\mathbf{z}^{-\mathbf{n}} = e^{i\theta}\mathbf{z}^{-\mathbf{k}}\sum_{\mathbf{n}'}\overline{f(\mathbf{t}+\mathbf{n}')}\mathbf{z}^{\mathbf{n}'},$$

which implies

$$f(\mathbf{t}+\mathbf{n}) = e^{i\theta}\overline{f(\mathbf{t}+\mathbf{k}-\mathbf{n})},\ \forall \mathbf{n}. \tag{126}$$

However, x_0 is deterministic, and $\{\mu(\mathbf{n})\}$ are independent and continuous random variables on $\mathbb{S}^1$, so (126) fails with probability one for any $\mathbf{k}$. There are finitely many choices of $\mathbf{k}$, so the z-transform of $f_{\mathbf{t}+}$ is almost surely not conjugate symmetric.

Similarly, the z-transform of $\overline{f_{\mathbf{t}-}}$ is also almost surely *not* conjugate symmetric. □

References

[1] P.F. Almoro, S.G. Hanson, Random phase plate for wavefront sensing via phase retrieval and a volume speckle field, Appl. Opt. 47 (2008) 2979–2987.
[2] R. Balan, Stability of phase retrievable frames, Preprint, arXiv:1308.5465, 2013.
[3] R. Balan, B.G. Bodmann, P.G. Casazza, D. Edidin, Painless reconstruction from magnitudes of frame coefficients, J. Fourier Anal. Appl. 15 (2009) 488–501.
[4] R. Balan, P. Casazza, D. Edidin, On signal reconstruction without phase, Appl. Comput. Harmon. Anal. 20 (2006) 345–356.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

[5] R. Balan, Y. Wang, Invertibility and robustness of phaseless reconstruction, Preprint, arXiv:1308.4718, 2013.
[6] A.S. Bandeira, J. Cahill, D.G. Mixon, A.A. Nelson, Saving phase: injectivity and stability for phase retrieval, Appl. Comput. Harmon. Anal. 37 (2014) 106–125.
[7] A.S. Bandeira, Y. Chen, D. Mixon, Phase retrieval from power spectra of masked signals, Inf. Inference 3 (2014) 83–102.
[8] H.H. Bauschke, P.L. Combettes, D.R. Luke, Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization, J. Opt. Soc. Amer. A 19 (2002) 1334–1345.
[9] H.H. Bauschke, P.L. Combettes, D.R. Luke, Finding best approximation pairs relative to two closed convex sets in Hilbert spaces, J. Approx. Theory 127 (2004) 178–192.
[10] H.H. Bauschke, J.Y.B. Cruz, T.T.A. Nghia, H.M. Phan, X. Wang, The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle, J. Approx. Theory 185 (2014) 63–79.
[11] M. Born, E. Wolf, Principles of Optics, 7-th edition, Cambridge University Press, 1999.
[12] R. Br´auer, U. Wojak, F. Wyrowski, O. Bryngdahl, Digital diffusers for optical holography, Opt. Lett. 16 (1991) 1427–1429.
[13] E.J. Candès, Y.C. Eldar, T. Strohmer, V. Voroninski, Phase retrieval via matrix completion, SIAM J. Imaging Sci. 6 (2013) 199–225.
[14] E.J. Candes, X. Li, M. Soltanolkotabi, Phase retrieval via Wirtinger flow: theory and algorithms, IEEE Trans. Inform. Theory 61 (4) (2015) 1985–2007.
[15] E.J. Candes, X. Li, M. Soltanolkotabi, Phase retrieval from coded diffraction patterns, Appl. Comput. Harmon. Anal. 39 (2015) 277–299.
[16] E.J. Candès, T. Strohmer, V. Voroninski, Phaselift: exact and stable signal recovery from magnitude measurements via convex programming, Comm. Pure Appl. Math. 66 (2012) 1241–1274.
[17] J. Cahill, P. Casazza, J. Peterson, L. Woodland, Phase retrieval by projections, Preprint, arXiv:1305.6226, 2013.
[18] Center for X-Ray Optics, Lawrence Berkeley National Laboratory, http:henke.lbl.gov.
[19] A. Chai, M. Moscoso, G. Papanicolaou, Array imaging using intensity-only measurements, Inverse Probl. 27 (1) (2011).
[20] H.N. Chapman, et al., Femtosecond X-ray protein nanocrystallography, Nature 470 (2011) 73–77.
[21] H.N. Chapman, C. Caleman, N. Timneanu, Diffraction before destruction, Philos. Trans. R. Soc. Lond., B, Biol. Sci. 369 (2014) 20130313.
[22] P. Chen, A. Fannjiang, G.-R. Liu, Phase retrieval with one or two coded diffraction patterns by alternating projection with the null initialization, arXiv:1510.07379.
[23] L. Demanet, P. Hand, Stable optimizationless recovery from phaseless linear measurements, J. Fourier Anal. Appl. 20 (2014) 199–221.
[24] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Trans. Amer. Math. Soc. 82 (1956) 421–439.
[25] J. Eckstein, D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program. A 55 (1992) 293–318.
[26] Y.C. Eldar, S. Mendelson, Phase retrieval: stability and recovery guarantees, Available online: arXiv:1211.0872.
[27] C. Falldorf, M. Agour, C. v. Kopylow, R.B. Bergmann, Phase retrieval by means of a spatial light modulator in the Fourier domain of an imaging system, Appl. Opt. 49 (2010) 1826–1830.
[28] A. Fannjiang, Absolute uniqueness of phase retrieval with random illumination, Inverse Probl. 28 (2012) 075008.
[29] A. Fannjiang, Fourier-domain fixed point algorithms with coded diffraction patterns, arXiv:1406.2742.
[30] A. Fannjiang, W. Liao, Phase retrieval with random phase illumination, J. Opt. Soc. Amer. A 29 (2012) 1847–1859.
[31] A. Fannjiang, W. Liao, Fourier phasing with phase-uncertain mask, Inverse Probl. 29 (2013) 125001.
[32] J.R. Fienup, Phase retrieval algorithms: a comparison, Appl. Opt. 21 (1982) 2758–2769.
[33] J.R. Fienup, Phase retrieval algorithms: a personal tour, Appl. Opt. 52 (2013) 45–56.
[34] R.W. Gerchberg, W.O. Saxton, A practical algorithm for the determination of the phase from image and diffraction plane pictures, Optik 35 (1972) 237–246.
[35] D. Gross, F. Krahmer, R. Kueng, A partial derandomization of phaselift using spherical designs, arXiv:1310.2267, 2013.
[36] J.W. Hardy, Adaptive Optics for Astronomical Telescopes, Oxford University Press, New York, 1998.
[37] H.A. Hauptman, The phase problem of X-ray crystallography, Rep. Progr. Phys. 54 (1991) 1427–1454.
[38] M. Hayes, The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform, IEEE Trans. Acoust. Speech Signal Process. 30 (1982) 140–154.
[39] K. Jaganathan, S. Oymak, B. Hassibi, On robust phase retrieval for sparse signals, in: 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton, 2012, pp. 794–799.
[40] X. Li, V. Voroninski, Sparse signal recovery from quadratic measurements via convex programming, SIAM J. Math. Anal. 45 (5) (2013) 3019–3033.
[41] P.-L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979) 964–979.
[42] A.M. Maiden, G.R. Morrison, B. Kaulich, A. Gianoncelli, J.M. Rodenburg, Soft X-ray spectromicroscopy using ptychography with randomly phased illumination, Nat. Commun. 4 (2013) 1669.
[43] S. Marchesini, A unified evaluation of iterative projection algorithms for phase retrieval, Rev. Sci. Instrum. 78 (2007) 011301.
[44] J. Miao, P. Charalambous, J. Kirz, D. Sayre, Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens, Nature 400 (1999) 342–344.
[45] J. Miao, J. Kirz, D. Sayre, The oversampling phasing method, Acta Crystallogr. Sect. A 56 (2000) 1312–1315.
[46] J. Miao, D. Sayre, H.N. Chapman, Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects, J. Opt. Soc. Amer. A 15 (1998) 1662–1669.
[47] A. Migukin, V. Katkovnik, J. Astola, Wave field reconstruction from multiple plane intensity-only data: augmented Lagrangian algorithm, J. Opt. Soc. Amer. A 28 (2011) 993–1002.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003

[48] R.P. Millane, J.P.L. Chen, Aspects of direct phasing in femtosecond nanocrystallography, Philos. Trans. Roy. Soc. London B, Biol. Sci. 369 (2014), arXiv:2013.0498.
[49] P. Netrapalli, P. Jain, S. Sanghavi, Phase retrieval using alternating minimization, Preprint, arXiv:1306.0160v2, 2015.
[50] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, Potential for biomolecular imaging with femtosecond X-ray pulses, Nature 406 (2000) 753–757.
[51] H. Ohlsson, A.Y. Yang, R. Dong, S.S. Sastry, Compressive phase retrieval from squared output measurements via semidefinite programming, Preprint, arXiv:1111.6323, 2011.
[52] T. Osaka, M. Yabashi, Y. Sano, K. Tono, Y. Inubushi, T. Sato, S. Matsuyama, T. Ishikawa, K. Yamauchi, A Bragg beam splitter for hard X-ray free-electron lasers, Opt. Express 21 (2013) 2823–2831.
[53] S. Oymak, A. Jalali, M. Fazel, Y.C. Eldar, B. Hassibi, Simultaneously structured models with application to sparse and low-rank matrices, Preprint, arXiv:1212.3753, 2012.
[54] J. Ranieri, A. Chebira, Y.M. Lu, M. Vetterli, Phase retrieval for sparse signals: uniqueness conditions, Preprint, arXiv:1308.3058, 2013.
[55] Y. Shechtman, Y.C. Eldar, O. Cohen, H.N. Chapman, Jianwei Miao, M. Segev, Phase retrieval with application to optical imaging: a contemporary overview, IEEE Signal Process. Mag. 32 (3) (2015) 87–109.
[56] M.M. Seibert, et al., Single mimivirus particles intercepted and imaged with an X-ray laser, Nature 470 (2011) U78–U86.
[57] I. Waldspurger, A. d'Aspremont, S. Mallat, Phase recovery, maxCut and complex semidefinite programming, arXiv:1206.0102.
[58] P. Yin, J. Xin, Phaseliftoff: an accurate and stable phase retrieval method based on difference of trace and Frobenius norms, Commun. Math. Sci. 13 (2015).
[59] X. Zhang, J. Jiang, B. Xiangli, G.R. Arce, Spread spectrum phase modulation for coherent X-ray diffraction imaging, Opt. Express 23 (2015) 25034–25047.

Please cite this article in press as: P. Chen, A. Fannjiang, Fourier phase retrieval with a single mask by Douglas–Rachford algorithms, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.07.003