When Deep Denoising Meets Iterative Phase Retrieval

Yaotian Wang 1  Xiaohang Sun 1  Jason W. Fleischer 1

Abstract

Recovering a signal from its Fourier intensity underlies many important applications, including lensless imaging and imaging through scattering media. Conventional algorithms for retrieving the phase suffer when noise is present but display global convergence when given clean data. Neural networks have been used to improve algorithm robustness, but efforts to date are sensitive to initial conditions and give inconsistent performance. Here, we combine iterative methods from phase retrieval with image statistics from deep denoisers, via regularization-by-denoising. The resulting methods inherit the advantages of each approach and outperform other noise-robust phase retrieval algorithms. Our work paves the way for hybrid imaging methods that integrate machine-learned constraints in conventional algorithms.

1. Introduction

Phase is a fundamental component of complex signals and often carries more information than amplitude (Oppenheim & Lim, 1981). In many cases, however, only the amplitude or intensity can be measured (e.g. conventional cameras can only record the time-averaged light intensity). For such measurements, the recovery of the underlying signal requires reconstruction of the missing phase. More formally, phase retrieval (PR) is an inverse problem to recover an unknown signal $x \in \mathbb{R}^n$ or $\mathbb{C}^n$ from the phaseless measurement

$$y^2 = |Ax|^2 + w$$

where $A$ is a known linear transform and $w$ is the noise in the measurement.

In the past decade, the general PR problem has attracted much attention from the optimization and statistics community and various algorithms with provable convergence were developed (Candès et al., 2015a;b; Zhang & Liang, 2016; Zhang et al., 2016; Chen & Candès, 2017; Wang et al., 2018). Despite a solid theoretical foundation, generally provable algorithms have overly restrictive requirements (e.g. the statistics of measurement bases) that have limited their popularity. More progress has been made for Fourier phase retrieval (FPR), in which $A$ is the Fourier transform, i.e. the measurements are the Fourier intensities. This is also the most common type experimentally, mostly due to the fact that diffraction in the far field is the Fourier transform of the object (Goodman, 2005). Applications range from astronomy (Dainty & Fienup, 1987) to coherent diffraction (Miao et al., 1999; Chapman & Nugent, 2010) and speckle-correlation (Bertolotti et al., 2012; Katz et al., 2014) imaging.

The most broadly used algorithms for FPR are iterative methods, pioneered by Gerchberg-Saxton (Gerchberg & Saxton, 1972) and later developed by Fienup (Fienup, 1982). Though they lack theoretical proof of convergence, empirical use of Fienup algorithms and their variants (Bauschke et al., 2003; Elser, 2003; Luke, 2004; Martin et al., 2012; Rodriguez et al., 2013) has shown the avoidance of local minima and convergence to global solutions from random initialization. Together with the simplicity of their implementation, iterative phase retrieval methods have become the workhorse of FPR (Miao et al., 2005; Bertolotti et al., 2012; Katz et al., 2014).

It has been shown that applying the natural image prior to FPR can increase robustness to noise and improve reconstruction quality (Venkatakrishnan et al., 2013; Heide et al., 2016; Metzler et al., 2018; Işıl et al., 2019). However, such methods either have unsatisfying robustness when noise levels are high or are sensitive to initialization (thus relying on other algorithms to supply initial points). Both cases return us to the problem of poor reliability when the signal-to-noise ratio in measurements is low.

Our major contribution here is to combine the benefits of iterative FPR with natural image priors via Regularization-by-Denoising (RED) (Romano et al., 2017). The methods we propose deliver greater robustness to noise than other noise-robust FPR algorithms while relaxing the initialization requirements. The application of image priors also alleviates the stagnant mode issues in iterative phase retrieval (Fienup
We focus on two-dimension signals and assume the measure-
ment transform $A$ in (1) to be the discrete Fourier transform
(DFT)
\[
\hat{x}[k_1, k_2] = \frac{1}{\sqrt{n}} \sum_{n_1, n_2=0}^{\sqrt{n}} x[n_1, n_2] e^{-2\pi i (k_1 n_1 + k_2 n_2) / n}
\] (2)
denoted here as $\hat{x} = Fx$ (with $F^{-1}$ the inverse transform).
Below, we discuss the uniqueness of Fourier phase retrieval, common algorithms used, and their relation to more general optimization problems.

2.1. Uniqueness in FPR

If there is not enough sampling, the Fourier intensity may be insufficient to trace back to the input signal. For all $d$-dimensional signals with $d \geq 2$, except a set of measure 0 (Hayes & McClellan, 1982), it has been shown that if the Fourier intensity is oversampled by a factor no less than 2 in each dimension, then a signal is determined uniquely by its Fourier intensity up to the trivial ambiguities of translation, conjugate inversion and global phase (Hayes, 1982). Fortunately, in practice these ambiguities are often acceptable, since the geometrical transform, complex conjugate and global phase keep the characteristics of the object intact.

Oversampling in the Fourier domain is related to the so-called support constraint for FPR, a terminology that is used more often in iterative phase retrieval. For example, suppose the Fourier spectrum of $x \in \mathbb{C}^{\sqrt{m} \times \sqrt{n}}$ is oversampled twice uniformly at $k_i = \{0, 1/2, 1, \ldots, \sqrt{n} - 1/2\}$ for $i = 1, 2$. For simplicity, we write this oversampled spectrum as $\hat{x}^{(2)}$. By defining $\tilde{x} \in \mathbb{C}^{\sqrt{m} \times \sqrt{n}}$ with $m = 4n$ such that $\tilde{x}[n_1, n_2] = \sqrt{m/x}[n_1, n_2]$ if $n_i \in \mathbb{N} < \sqrt{n}$ and $\tilde{x}[n_1, n_2] = 0$ otherwise, we have
\[
\hat{x}^{(2)}[k_1, k_2] = \frac{1}{\sqrt{n}} \sum_{n_1, n_2=0}^{\sqrt{n}} x[n_1, n_2] e^{-i2\pi \frac{n_1 k_1 + n_2 k_2}{m}}
= \frac{1}{\sqrt{n}} \sum_{n_1, n_2=0}^{\sqrt{n}} \sqrt{m/x}[n_1, n_2] e^{-i2\pi \frac{n_1 k_1 + n_2 k_2}{m}}
= (Fx)[\tilde{k}_1, \tilde{k}_2]
\] (3)
where $\tilde{k}_i = \{0, 1, \ldots, 2\sqrt{n} - 1\} = 2k_i$. Therefore, there exists a supported signal $\tilde{x}$ by zero-padding $P_m$, and scaling $x$ by a factor of $\sqrt{m/n}$, such that its Fourier transform is the same as (uniform) oversampling in the Fourier space of $x$. If the vectorization order gives
\[
\tilde{x}^T = \sqrt{\frac{m}{n}} [x^T, 0_{m-n}^T]
\] (4)
then
\[
\hat{x}^{(2)} = F\tilde{x} = FO_mx
\] (5)
where $O_{mn} \in \mathbb{R}^{m \times n}$ is given by
\[
O_{mn} = \sqrt{\frac{m}{n}} [I_n, 0] = \sqrt{\frac{m}{n}} P_{mn}
\] (6)

Stated another way, oversampling FPR is equivalent to finding a supported signal $\tilde{x}$ from its DFT intensity, with the support constraint sometimes including the support of $x$ itself. To distinguish them, we denote the support for $x \in \mathbb{C}^n$ as $S = \{i \mid x_i \neq 0\}$ and the extended support for padded $\tilde{x}$ as $\tilde{S} = \{j \mid \tilde{x}_j \neq 0\}$.

2.2. ADMM

The Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2011; Hong et al., 2016) is a popular algorithm for solving the linearly constrained optimization problem
\[
\begin{align*}
\text{minimize} & \quad \ell(x_1, \ldots, x_N) = \sum_{i=1}^{N} f_i(x_i) \\
\text{subject to} & \quad \sum_{i=1}^{N} A_i x_i = b
\end{align*}
\] (7)

For each iteration, ADMM updates each $x_i$ sequentially and dual variable $u$ as
\[
\begin{align*}
x_i^{k+1} &= \arg\min_{x_i} f_i(x_i) + \frac{\rho}{2} \left\| \sum_{j \neq i} A_j x_j' + A_i x_i - b + u^k \right\|^2 \\
u^{k+1} &= u^k + \sum_{i=1}^{N} A_i x_i^{k+1} - b
\end{align*}
\] (8)

with $x_j' = x_j^{k+1}$ for $j < i$ and $x_i' = x_i^k$ otherwise for each update of $x_i^{k+1}$. Penalty parameter $\rho^k$ is constant or adaptive through iterations.

One often needs to evaluate the minimization problem of a form
\[
\text{prox}_f(z) := \arg\min_v f(v) + \frac{1}{2} \|v - z\|^2_2
\] (9)
which is defined as the proximal operator for $f$ at $z$ (Parikh & Boyd, 2014). The efficiency of ADMM generally depends on the complexity of evaluating the proximal operator for each $f$, while in return the functions can be non-differentiable. We show below that this latter property can be quite beneficial.
2.3. Hybrid-Input-Output Method

As possibly the most used iterative method in FPR, the Hybrid-Input-Output (HIO) (Fienup, 1982) is well known for its ability to converge to global minima from random initialization. HIO iterates on the padded and scaled signal \( \tilde{x} \) with the following step rules:

\[
\tilde{x}^{k+1} = F^{-1} \left( y \odot \frac{F\tilde{x}^k}{|F\tilde{x}|} \right)
\]

\( \forall i, \tilde{x}_i^{k+1} = \begin{cases} 
\tilde{x}_i^{k+1} & \text{if } i \in \hat{S} \\
\tilde{x}_i^{k} - \beta \tilde{x}_i^{k+1} & \text{otherwise}
\end{cases}
\]  \( (10) \)

where the product \( \odot \), the modulus \( | \cdot | \) and division are all performed element-by-element.

It was shown in (Bauschke et al., 2002) that HIO with \( \beta = 1 \) coincides with Douglas-Rachford splitting (DRS) (Douglas & Rachford, 1956; Lions & Mercier, 1979; Eckstein & Bertsekas, 1992). Since DRS is equivalent to ADMM updates on the feasibility problem with indicator functions (Boyd & Vandenberghe, 2004), one can find that HIO \( (\beta = 1) \) is equivalent to ADMM on the following minimization problem:

\[
\min_{x \in \mathbb{C}^n} \bar{I}_M(z) + \bar{I}_C(x)
\]

subject to \( z = O_{mn}x \)

where the indicator function for a subset \( S \) is defined as (Boyd & Vandenberghe, 2004)

\[
\bar{I}_S(x) = \begin{cases} 
0 & \text{if } x \in S \\
\infty & \text{otherwise}
\end{cases}
\]  \( (12) \)

and the set \( M \) is defined as the set of signals consistent with the measurement

\[
M := \{ x \in \mathbb{C}^n \mid |Fx| = y \}
\]  \( (13) \)

Here, \( C \) is the set of signals satisfying an additional constraint, such as inset support \( S \) and nonnegativity (which result in the Hybrid-Projection-Reflection algorithm (Bauschke et al., 2003)). More details of this mapping are given in the supplementary material.

The proximal operator of the indicator function \( (12) \) is given by the projection to the corresponding set

\[
\Pi_S(x) := \text{argmin}_{z \in S} \| z - x \| = \text{prox}_{\bar{I}_S}(x)
\]  \( (14) \)

In particular, the projection onto \( M \) can be written as

\[
F^{-1} \left( y \odot \frac{Fv}{|Fv|} \right) \in \Pi_M(v) = \text{prox}_{\bar{I}_M}(v)
\]  \( (15) \)

For the additional constraint \( C \) in \( (11) \), nonnegativity in the real part of the signal can be used. The projection onto the corresponding subset is

\[
\Pi_{Re_+}(x)_i = \begin{cases} 
x_i & \text{if } \Re(x_i) \geq 0 \\
i \Im(x_i) & \text{otherwise}
\end{cases}
\]  \( (16) \)

3. Related Works

In this section, we introduce the efforts to date for solving the PR problem in the presence of noise.

3.1. Iterative Phase Retrieval

Iterative phase retrieval methods commonly solve a feasibility problem, looking for a signal whose oversampled Fourier intensity is \( y^2 \) and simultaneously consistent with the other constraint \( C \). The problem occurs when noise levels increase in the measurement, resulting in oscillations and ambiguous solutions. To alleviate the degradation from corrupted data, efforts have been made to limit the effect of noise on iterative methods (Luke, 2004; Martin et al., 2012; Rodriguez et al., 2013). However, without further priors on the object space (e.g., image statistics), the denoising effect of these methods is often insufficient.

3.2. Deep Learning in PR

Deep neural networks (DNN) are well known for their capability to approximate complicated functions (given enough training data). In image processing, they have achieved significant improvements over traditional methods in areas such as denoising (Zhang et al., 2017a; 2018), deblurring (Nimisha et al., 2017), and superresolution (Dong et al., 2014; Lim et al., 2017). For solving PR, deep feedforward networks have shown some success in end-to-end predictions (Sinha et al., 2017; Rivenson et al., 2018), while network-assisted algorithms also have helped in support estimation (Kim & Chung, 2019), low-light (Goy et al., 2018) and compressive (Hand et al., 2018) situations.

However, using a feedforward neural network to approximate the inverse mapping is problematic for oversampling FPR. Such methodology relies on the assumption that forward mapping is one-to-one and well posed; this is not the case here, even with precise knowledge of the signal support, due to the existence of trivial ambiguities. Instead, the optimization method commonly adopted for solving FPR (e.g., in (Heide et al., 2016; Metzler et al., 2018)) minimizes the loss function

\[
\ell(x) := f(x; y) + \alpha R(x)
\]  \( (17) \)

where \( f \) is the data fidelity term and \( R \) is a regularizer involving prior belief, e.g., natural image statistics. This method is effectively a maximum a posteriori estimation (Venkatakrishnan et al., 2013).

3.3. Prior by Denoisers

Using a denoiser as the prior \( R \) in \( (17) \) has been proposed to boost image inference in inverse problems. There have been two major strategies to utilize the denoiser: Plug-and-Play (PnP) regularization (Venkatakrishnan et al., 2013).
and Regularization-by-Denoising (RED) (Romano et al., 2017). In PhP methods, the proximal operator for an implicit regularizer $R$ is approximated by an image denoiser. This approach provides promising results both empirically (Venkatakrishnan et al., 2013; Heide et al., 2014; 2016; Metzler et al., 2016; Meinhardt et al., 2017; Zhang et al., 2017b) and theoretically (Chan et al., 2017). Meanwhile, RED is a framework that constructs explicit regularizers with denoisers $D$ as the inner product between a signal and the noise it contains,

$$R(x) = \frac{\lambda}{2} \langle x, x - D(x) \rangle$$ (18)

It has been shown in (Romano et al., 2017) that if the denoiser $D$ has the properties of (local) homogeneity and Jacobian symmetry, then evaluation of the proximal operator in (18) at $x$ requires the solution to

$$z - x + \lambda(z - D(z)) = 0$$ (19)

Though these properties rarely hold for common denoisers, (19) can still be adopted either as an approximation or if certain conditions hold (Reehorst & Schniter, 2019). Recent applications of RED to PR have demonstrated a significant boost in noise robustness compared with bare iterative methods (Metzler et al., 2018; Wu et al., 2019).

4. Methodology

We aim to maintain the convergence benefits of HIO while alleviating the deleterious effects of noise. To this end, we adopt ADMM as the solver but modify the loss function used in HIO. More specifically, we eliminate the inconsistency from (11) by relaxation of the loss function and include natural image priors via RED, due to its explicit form and inherent flexibility.

For relaxation of the loss function, we consider two approaches: one in the Fourier constraint and one in the oversampling constraint. These give two algorithms, RED-ITA-F and RED-ITA-S, respectively.

In general, we refer to our algorithms as RED-ITA, and Deep-ITA for the specific choice of deep denoisers, such as DnCNN (Zhang et al., 2017a).

4.1. RED-ITA-F

We first consider substituting the indicator function on Fourier measurement with the data fidelity term. Following (Metzler et al., 2018), we seek to solve

$$\frac{1}{2} \|y - |FO_{mn}x||^2 + \frac{\lambda}{2} \langle x, x - D(x) \rangle$$ (20)

Similar to HIO, we transform (20) into a linearly constrained form as

$$\begin{align*}
\text{minimize} & \frac{1}{2} \|y - |Fz||^2 + R(x) \\
\text{subject to} & \quad z = O_{mn}x
\end{align*}$$ (21)

where $R$ contains RED and an additional constraint $\bar{I}_C(x)$:

$$R(x) = \bar{I}_C(x) + \frac{\lambda}{2} \langle x, x - D(x) \rangle$$ (22)

For $f(z) = \frac{1}{2} \|y - |Fz||^2$, the update rule of ADMM gives

$$\begin{align*}
x^{k+1} &= \text{argmin}_{x \in \mathbb{R}^n} R(x) + \frac{\rho}{2} \|z^k - O_{mn}x + u^k\|^2 \\
z^{k+1} &= \text{prox}_{\bar{I}_C}(O_{mn}x^{k+1} - u^k) \\
u^{k+1} &= u^k + z^{k+1} - O_{mn}x^{k+1}
\end{align*}$$ (23)

It remains to evaluate each update step. We note that for any $x \in \mathbb{R}^n, v = [v_n^T v_{m-n}^T]^T \in \mathbb{C}^m$, where $v_n = PT_{mn}v = \sqrt{\frac{m}{n}}O_{mn}v,$

$$\|v - O_{mn}x\|^2 = \|\Re(v) - O_{mn}x\|^2 + \|\Im(v)\|^2$$

$$= \frac{m}{n} \left\| \sqrt{\frac{m}{n}} \Re(v_n) - x \right\|^2 + \|\Re(v_m)\|^2 + \|\Im(v)\|^2$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex-valued signal. Therefore, in terms of $v = z^k + u^k$, the $x$-update step in (23) can be found as

$$\begin{align*}
x^{k+1} &= \text{argmin}_{x \in \mathbb{R}^n} R(x) + \frac{\rho}{2} \|v - O_{mn}x\|^2 \\
&= \text{argmin}_{x \in \mathbb{R}^n} R(x) + \frac{m\rho}{2n} \left\| \sqrt{\frac{m}{n}} \Re(v_n) - x \right\|^2 \\
&= \text{prox}_{\frac{m}{2n}}R\left( \sqrt{\frac{m}{n}} \Re(v_n) \right)
\end{align*}$$ (24)

which reduces to an evaluation of the proximal operator for $R$. For any $\tau > 0$, if $s^+ = \text{prox}_{\tau R}(s)$, we have

$$s^+ = \Pi_C\left( \frac{s + \lambda \tau D(s^+)}{1 + \lambda \tau} \right)$$ (25)

(a derivation is given in the supplementary material). Similar to RED in (Romano et al., 2017), the proximal operator in (25) can be evaluated by the fixed-point approach, updating

$$s^{(k+1)} = \Pi_C\left( \frac{s + \lambda \tau D(s^{(k)})}{1 + \lambda \tau} \right)$$ (26)

until convergence. In practice, the fixed point can be approximated by stopping after $p$ iterations, denoted as $\text{prox}_{\tau R}(v) = s^{(p)}$ with $p$ being a hyperparameter and
Figure 1. Test images used in the simulation. Top row: 6 commonly used “natural” test images (Zhang et al., 2017a). Bottom row: 6 “unnatural” images (Metzler et al., 2018). Images have been resized to 128 × 128.

Algorithm 1 RED-ITA-F

Input: Initialization \( x^0, u^0 \in \mathbb{C}^m \), positive real number \( \rho, \lambda > 0 \), oversampling transform \( O_{mn} \), Fourier measurement \( y \)

for \( k = 0, 1, 2, \cdots \) do

\[
\begin{align*}
& u^k = z^k + u^k \\
& \tau = (m\rho)^{-1} n \\
& z^{k+1} = \text{prox}_{\rho R}(\frac{n}{m} R(O_{mn}^*u^k)) \\
& x^{k+1} = O_{mn}z^{k+1} \\
& u^{k+1} = u^k + z^{k+1} - x^{k+1}
\end{align*}
\]

end for

\( s(0) = s \). Empirically, we found that \( p = 1 \) is efficient enough; therefore, \( p \) is set to 1 in all of our experiments.

For the \( z \)-update step, the proximal operator for \( f \) can be written as

\[
\text{prox}_{\rho f}(s) = \frac{1}{\tau + 1} s + \frac{\tau}{\tau + 1} \Pi_M(s)
\]

For solving oversampling FPR is shown in Algorithm 1.

4.2. RED-ITA-S

The second approach relaxes the oversampling constraint, instead of the Fourier measurement. Rather than assuming there exists \( x \in \mathbb{R}^n \) such that \( O_{mn}x = z \in \mathcal{M} \), we acknowledge that the difference \( \xi = z - O_{mn}x \) can be non-zero \( \forall z \in \mathcal{M} \) and minimize the norm of it. That is, an alternative to (21) is

\begin{equation}
\begin{aligned}
& \text{minimize} \quad I_M(z) + \frac{1}{2} \|\xi\|^2 + R(x) \\
& \text{subject to} \quad z = O_{mn}x + \xi
\end{aligned}
\end{equation}

Note that, given \( x \in \mathbb{R}^n \), the loss in (28) is an upper bound for that in (21), since \( \forall z \in \mathcal{M} \), Parseval’s theorem gives

\[
\|\xi\|^2 = \|z - O_{mn}x\|^2 = \|ye^{i\phi_z} - FO_{mn}x\|^2 \geq \|y - |FO_{mn}x|\|^2
\]

where \( \phi_z \) is the Fourier phase of \( z \).

A three-block ADMM is adopted to solve (28):

\[
\begin{aligned}
x^{k+1} &= \text{argmin}_{x \in \mathbb{R}^n} R(x) + \frac{\rho}{2} \|z^{k+1} - O_{mn}x - \xi^k + u^k\|^2 \\
z^{k+1} &= \text{prox}_{\frac{\lambda}{2} \Pi_M}(O_{mn}x^{k+1} + \xi^k - u^k) \\
\xi^{k+1} &= \text{prox}_{\frac{\rho}{\lambda} \Pi_M}(z^{k+1} - O_{mn}x^{k+1} + u^k) \\
u^{k+1} &= u^k + z^{k+1} - O_{mn}x^{k+1} - \xi^{k+1}
\end{aligned}
\]

where

\[
\text{prox}_{\frac{\lambda}{2} \Pi_M}(s) = \frac{\Pi_M(s)}{\lambda + \frac{\tau}{\tau + 1} s}
\]

This yields the RED-ITA-S shown in Algorithm 2.

4.3. Connection between PR Algorithms

(Metzler et al., 2018) proposed solving (20) with FASTA (Goldstein et al., 2014), a method known as prRED (the variant using deep denoisers like DnCNN is referred to as prDeep). Since FASTA is a forward-backward splitting method, if the stepsize \( \mu \) is fixed to be \( n/m \) and \( \lambda \to 0 \), prDeep reduces to (sub-)gradient descent on the squared loss on Fourier amplitude, which coincides (Marchesini, 2007) with the Error Reduction algorithm (Fienup, 1982).
Wang et al., 2018) are also excluded in comparison, since we compare Deep-ITA-F/S with other widely used algorithms on FPR, namely HIO (Fienup, 1982), Oversampling Smoothness (OSS) (Rodriguez et al., 2013), DnCNN-ADMM (Venkatakrishnan et al., 2013; Heide et al., 2016; Chan et al., 2017) and prDeep (Metzler et al., 2018). We did not include any post-reconstruction procedure to clean the results as in (I¸sil et al., 2019), which is not tested here since the algorithm performs worse than prDeep unless an additional DNN specifically trained to enhance the quality of reconstruction is used. Provable methods (Candès et al., 2015a;b; Zhang et al., 2016; Chen & Candès, 2017) and do not cover oversampled Fourier intensities. Besides, results of a typical provable method, Wirtinger Flow (Candès et al., 2015b), for oversampling FPR have been reported in (Metzler et al., 2018), where it significantly underperformed even HIO.

In principle, any denoiser can be adopted in RED. Here, we choose DnCNN (Zhang et al., 2017a), based on its competitive denoising performance and its flexibility on the input signal. DnCNN is stacked by Convolutional and Batch Normalization layers with Rectified Linear Unit (ReLU) activation functions. With a zero-padding of size 1 for $3 \times 3$ convolutional kernel size, the output dimension remains the same as that of the input. DnCNN models are trained on patches of natural images with mean-squared-error as the loss function, using Adam as the optimizer (Kingma & Ba, 2014).

The test images used in the simulations, shown in Figure 1, consist of 6 commonly used “natural” images and 6 “unnatural” ones. The images are resized to $128 \times 128$ and their Fourier intensities are oversampled uniformly by a factor of 2 in each dimension, yielding measurements of size $256 \times 256$. The signals used as ground truth are real-valued and have dynamic range of $[0, 255]$.

For simulation, shot noise is assumed to dominate the noise in the measurement. While this noise follows a Poisson distribution, it is commonly approximated as a Gaussian (Metzler et al., 2018; I¸sil et al., 2019). The noisy measurement $y$ on the oversampled Fourier amplitude $q = \hat{x}^{(2)}$ thus has the distribution

$$y^2 = |q|^2 + w \quad w \sim \mathcal{N}(0, \text{diag}(\alpha^2|q|^2))$$

(32)

It is worth noting that the (effective) SNR in the measurements scales roughly with $y/\alpha$, which is affected by $\alpha$ and any scaling in $|q|$. We define two metrics to characterize the SNR: MSNR$_1 = 10 \log_{10}(||q||^2/||y - |q|^2||^2)$ (I¸sil et al., 2019) and MSNR$_2 = 20 \log_{10}(||q||^2/||y - |q|^2||)$ (Luke, 2004).

Results from two experimental setups are reported here. In the first, we test the convergence of the competing phase retrieval algorithms with random initialization. All algorithms are initialized with the same random point and run for the

---

**Algorithm 2 RED-ITA-S**

**Input:** Initialization $x^0, z^0, u^0 \in \mathbb{C}^m$, $\rho, \lambda > 0$, oversampling transform $O_{mn}$, Fourier measurement $y$

**for** $k = 0, 1, 2, \cdots$ **do**

- $x^k = z^k + u^k - \xi^k$
- $\tau = (m\rho)^{-1} n$
- $x^{k+1} = \text{prox}_{\tau R}(\frac{n}{m} R(O_{mn}^T v^{k}))$
- $\xi^{k+1} = O_{mn} x^{k+1}$
- $z^{k+1} = \Pi_{\mathcal{M}}(z^{k+1} + \xi^k - u^k)$
- $\xi^k = \frac{1}{n+1} (z^{k+1} - x^{k+1} + u^k)$
- $y^k = u^k + z^{k+1} - \hat{x}^{k+1} - \xi^{k+1}$

**end for**

RED-ITA-F reduces to HIO with $\beta = 1$ when $\rho \to 0$ and $\lambda/\rho \to 0$. Similarly for DnCNN-ADMM if the denoising step is put first and the denoiser is the identity mapping.

**5. Experimental Results**

We compare Deep-ITA-F/S with other widely used algorithms on FPR, namely HIO (Fienup, 1982), Oversampling Smoothness (OSS) (Rodriguez et al., 2013), DnCNN-ADMM (Venkatakrishnan et al., 2013; Heide et al., 2016; Chan et al., 2017) and prDeep (Metzler et al., 2018). We did not include any post-reconstruction procedure to clean the results as in (I¸sil et al., 2019), which is not tested here since the algorithm performs worse than prDeep unless an additional DNN specifically trained to enhance the quality is used. Provable methods (Candès et al., 2015a;b; Zhang & Liang, 2016; Zhang et al., 2016; Chen & Candès, 2017; Wang et al., 2018) are also excluded in comparison, since their measurement models focus on Gaussian distributions and do not cover oversampled Fourier intensities. Besides, results of a typical provable method, Wirtinger Flow (Candès et al., 2015b), for oversampling FPR have been reported in (Metzler et al., 2018), where it significantly underperformed even HIO.

In principle, any denoiser can be adopted in RED. Here, we choose DnCNN (Zhang et al., 2017a), based on its competitive denoising performance and its flexibility on the input signal. DnCNN is stacked by Convolutional and Batch Normalization layers with Rectified Linear Unit (ReLU) activation functions. With a zero-padding of size 1 for $3 \times 3$ convolutional kernel size, the output dimension remains the same as that of the input. DnCNN models are trained on patches of natural images with mean-squared-error as the loss function, using Adam as the optimizer (Kingma & Ba, 2014).

The test images used in the simulations, shown in Figure 1, consist of 6 commonly used “natural” images and 6 “unnatural” ones. The images are resized to $128 \times 128$ and their Fourier intensities are oversampled uniformly by a factor of 2 in each dimension, yielding measurements of size $256 \times 256$. The signals used as ground truth are real-valued and have dynamic range of $[0, 255]$.

For simulation, shot noise is assumed to dominate the noise in the measurement. While this noise follows a Poisson distribution, it is commonly approximated as a Gaussian (Metzler et al., 2018; I¸sil et al., 2019). The noisy measurement $y$ on the oversampled Fourier amplitude $q = \hat{x}^{(2)}$ thus has the distribution

$$y^2 = |q|^2 + w \quad w \sim \mathcal{N}(0, \text{diag}(\alpha^2|q|^2))$$

(32)

It is worth noting that the (effective) SNR in the measurements scales roughly with $y/\alpha$, which is affected by $\alpha$ and any scaling in $|q|$. We define two metrics to characterize the SNR: MSNR$_1 = 10 \log_{10}(||q||^2/||y - |q|^2||^2)$ (I¸sil et al., 2019) and MSNR$_2 = 20 \log_{10}(||q||^2/||y - |q|^2||)$ (Luke, 2004).

Results from two experimental setups are reported here. In the first, we test the convergence of the competing phase retrieval algorithms with random initialization. All algorithms are initialized with the same random point and run for the
When Deep Denoising Meets Iterative Phase Retrieval

Table 1. PSNRs and SSIMs of reconstructions initialized with random noise with varying noise level in the measurements. For $\alpha = 0$, no noise is added to the Fourier intensity. For $\alpha = 4$, averaged MSNR$_1 = 32.09$dB, MSNR$_2 = 33.36$dB.

<table>
<thead>
<tr>
<th>$\alpha = 0$</th>
<th>AVERAGE PSNR</th>
<th>AVERAGE SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NATURAL</td>
<td>UNNATURAL</td>
</tr>
<tr>
<td>HIO</td>
<td>48.88</td>
<td>56.01</td>
</tr>
<tr>
<td>OSS</td>
<td>24.27</td>
<td>44.31</td>
</tr>
<tr>
<td>prDeep</td>
<td>13.70</td>
<td>18.27</td>
</tr>
<tr>
<td>DnCNN-ADMM</td>
<td>29.11</td>
<td>27.94</td>
</tr>
<tr>
<td>DEEP-ITA-F</td>
<td>65.06</td>
<td>57.88</td>
</tr>
<tr>
<td>DEEP-ITA-S</td>
<td>64.94</td>
<td>57.93</td>
</tr>
</tbody>
</table>

Table 2. PSNRs and SSIMs of reconstructions initialized from HIO with varying noise level in the measurements. For $\alpha = 8$, the averaged MSNR$_1 = 29.09$dB, MSNR$_2 = 27.54$dB. For $\alpha = 12$, averaged MSNR$_1 = 27.38$dB, MSNR$_2 = 24.49$dB. For $\alpha = 16$, averaged MSNR$_1 = 25.84$dB, MSNR$_2 = 22.52$dB.

<table>
<thead>
<tr>
<th>$\alpha = 8$</th>
<th>AVERAGE PSNR</th>
<th>AVERAGE SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NATURAL</td>
<td>UNNATURAL</td>
</tr>
<tr>
<td>HIO (INIT.)</td>
<td>20.78</td>
<td>23.03</td>
</tr>
<tr>
<td>OSS</td>
<td>22.02</td>
<td>27.58</td>
</tr>
<tr>
<td>prDeep</td>
<td>28.50</td>
<td>30.75</td>
</tr>
<tr>
<td>DnCNN-ADMM</td>
<td>26.95</td>
<td>27.76</td>
</tr>
<tr>
<td>DEEP-ITA-F</td>
<td>32.90</td>
<td>31.36</td>
</tr>
<tr>
<td>DEEP-ITA-S</td>
<td>33.31</td>
<td>32.78</td>
</tr>
</tbody>
</table>

same total number of 1200 iterations. In the second, we follow the initializing strategy used in (Metzler et al., 2018; Işıl et al., 2019): first, make 50 runs of randomly initialized HIO (giving $\hat{x}_i$ for $i = 1, \cdots, 50$), each with 50 iterations; next, pass the one with the lowest residual $\hat{x} = \text{argmin}_i f(\hat{x}_i)$ to initialize another HIO run of 1000 iterations. The output is then used as initialization for other algorithms. For both experiments, the whole procedure is repeated three times and the one most matched with the measurement is selected as the final output for each algorithm.

The parameters in the algorithms were as follows: for HIO and OSS, $\beta = 0.9$. The regularization parameter $\lambda$ is found best set at $0.01 \sigma^2$ for DnCNN-ADMM, $0.025 \sigma^2$ for both Deep-ITA-S/F, and $0.05 \sigma^2$ for prDeep, where $\sigma$ is the stan-
When Deep Denoising Meets Iterative Phase Retrieval

Figure 3. Reconstructions from random initialization with $\alpha = 0$. The stripes in the HIO reconstruction are artifacts from stagnation (Fienup & Wackerman, 1986); they are resolved in our method.

The standard deviation of noise in the Fourier amplitude (or set to 0.1 if no noise is added). Similar to the practice in (Metzler et al., 2018), prDeep and Deep-ITAs sequentially use DnCNN models that are trained with noise standard deviations of 60, 40, 20 and 10, each with 300 iterations for a total of 1200 iterations. The penalty parameter $\rho$ used in Deep-ITAs is set to $\frac{1}{2}\lambda$. We notice that reducing $\lambda$ and $\rho$ when using the DnCNNs for high noise levels can increase the stability of our methods. For prDeep and Deep-ITAs, we use the nonnegativity in (16) as the additional constraint in the regularizer.

For quantitative evaluation of the reconstructions, we characterize the output by its Peak Signal-to-Noise Ratio (PSNR) compared to the ground truth as well as the Structure Similarity (SSIM) Index (Wang et al., 2004). The PSNR computed for each reconstruction is capped at 80 dB, in case an outlier has a high value that adversely affects the estimation of mean reconstruction quality (which could happen, e.g., in the noise-free case $\alpha = 0$).

5.1. Random Initialization

Results of the experiments with random initialization are shown in Figure 2 and Table 1. Our methods outperform every other PR algorithm by large margins, in both PSNR and SSIM. Significantly, this includes HIO even when noise is absent (Figure 3). (This is probably due to stagnation in HIO, which is hard to overcome in a limited number of iterations (Fienup & Wackerman, 1986).) prDeep has issues with random initialization, which is not surprising considering its connection with Error Reduction, which has been shown to have slow convergence in practice (Fienup, 1982). On the contrary, DnCNN-ADMM and Deep-ITAs have the ability to work with random initial points, since all of them use ADMM as the solver. Our methods are more effective, as we integrate the denoiser in the update via RED, rather than apply it in a Plug-and-Play manner.

5.2. Initialization by HIO

Table 2 shows the performance of test algorithms with different level of noise in Fourier intensity when initialized with HIO. Deep-ITAs exhibit higher robustness to noise for every level of noise added. Figure 4 shows a visual comparison between PR algorithms for $\alpha = 12$, where Deep-ITA-S provides the best reconstruction. For the other methods, artifacts appear in the reconstructions and many details are lost.

6. Conclusion

Phase retrieval is part of a more general class of algorithms that has (to date) resisted full, end-to-end solutions from machine learning. While an admirable goal, such approaches often apply machine learning in situations where it is ill-suited. It also neglects traditional algorithms and their corresponding strengths, viz. convergence benefits.

The approach advocated here is to build algorithms in the fashion of traditional methods but with added priors utilizing...
deep neural networks. In the problem of Fourier phase retrieval, we added the object-space regularizer of image statistics and improved noise robustness. More generally, the results pave the way for hybrid methods that integrate machine-learned constraints in conventional algorithms.

Acknowledgements

This material is based upon work supported by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-18-1-0219 and by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00111890042.

References


When Deep Denoising Meets Iterative Phase Retrieval


