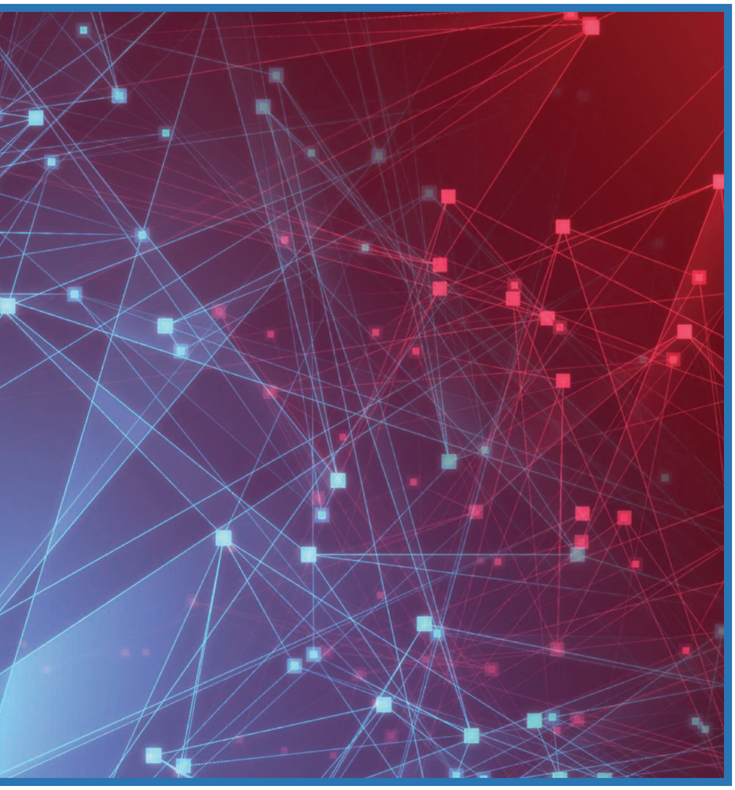
Jonathan Dong, Lorenzo Valzania, Antoine Maillard,
Thanh-an Pham, Sylvain Gigan, and Michael Unser

Phase Retrieval: From Computational Imaging to Machine Learning

A tutorial



©SHUTTERSTOCK.COM/PAPAPIG

Phase retrieval consists in the recovery of a complex-valued signal from intensity-only measurements. As it pervades a broad variety of applications, many researchers have striven to develop phase-retrieval algorithms. Classical approaches involve techniques as varied as generic gradient descent routines or specialized spectral methods, to name a few. However, the phase-recovery problem remains a challenge to this day. Recently, however, advances in machine learning have revitalized the study of phase retrieval in two ways: 1) significant theoretical advances have emerged from the analogy between phase retrieval and single-layer neural networks, and 2) practical breakthroughs have been obtained thanks to deep learning regularization. In this tutorial, we review phase retrieval under a unifying framework that encompasses classical and machine learning methods. We focus on three key elements: applications, an overview of recent reconstruction algorithms, and the latest theoretical results.

Introduction

Phase retrieval is a longstanding computational problem that is simple to define yet difficult to solve. It consists of the d-dimensional search for $\mathbf{x}^* \in \mathbb{C}^d$ such that

$$\mathbf{y} = |\mathbf{A}\mathbf{x}^*|^2. \tag{1}$$

There, one assumes that the measurements $\mathbf{y} \in \mathbb{R}^n$ and the matrix $\mathbf{A} \in \mathbb{C}^{n\times d}$ are known, while $|\cdot|$ is the elementwise modulus operator [Figure 1(a)]. Compared to the linear equation $\mathbf{y} = \mathbf{A}\mathbf{x}^*$, the missing phase in (1) makes the computational problem more challenging.

Phase retrieval is a problem that is encountered in many fields. In the context of imaging with electromagnetic waves, it arises whenever one wants to image an object from intensity-only measurements [Figure 1(b)]. It has been described in the physics literature since the 1950s, first in crystallography [1], followed by intense studies in astronomy since the 1980s [2]. More recently, it has been intimately linked with the next generation of image reconstruction algorithms, which solve this computational problem to push the limits of conventional imaging [3], [4], [5], [6]. Today, phase retrieval encompasses vastly different

Digital Object Identifier 10.1109/MSP.2022.3219240
Date of current version: 29 December 2022

imaging settings, from X-ray to optical or terahertz (THz) imaging, to observe objects from molecular to astronomical scales. This equation is also encountered outside of computational imaging, for example, in computer-generated holography [7] or optical computing [8], to name a few. Despite the wide range of applications, the underlying physics-based models lead to the same universal phase-retrieval problem.

The phase-retrieval problem has stimulated the development of a variety of novel imaging modalities. Likewise, a plethora of reconstruction algorithms have been designed [9], from the first alternating-projection algorithms to convex relaxations [10] and spectral methods [11], [12]. Given the rich palette of new reconstruction strategies, it may be unclear for practitioners which one to choose for a particular application. A description of their performance and range of applicability would help in this regard.

Phase retrieval raises theoretical questions, such as the uniqueness of the solution [13] and the performance guarantees of algorithms [12], [14]. For instance, strong results have recently been obtained for random models where the elements of $\mathbf{A}$ are drawn as independent identically distributed (i.i.d.) random variables. In this case, a clear picture has emerged on regions of solvability [15], [16] and the best algorithm to use. While recent theoretical progress has pushed our understanding to more structured matrices $\mathbf{A}$ [16], tackling other models related to real-world settings remains a challenge.

Phase retrieval is also deeply linked with machine learning. The underlying forward model can be interpreted as a single-layer neural network with a quadratic activation function [Figure 1(c)]. Thus, the solution to the phase-retrieval problem amounts to the learning of the network weights, which is a typical machine learning optimization problem. While phase retrieval certainly does not expose one to the full complexity of deep learning, it is already rich enough to provide an entry point to the difficult questions debated in the machine learning community today. In particular, it exhibits local minima, strongly relies on the quality of the initialization, and behaves differently depending on the amount of training data. As a result, it has sparked a strong interest in the theoretical machine learning community. In particular, phase retrieval could help us better understand the nonconvex optimization procedures typical of real-world machine learning. Ultimately, phase retrieval represents a transdisciplinary topic that impacts fields ranging from imaging applications to fundamental mathematics.

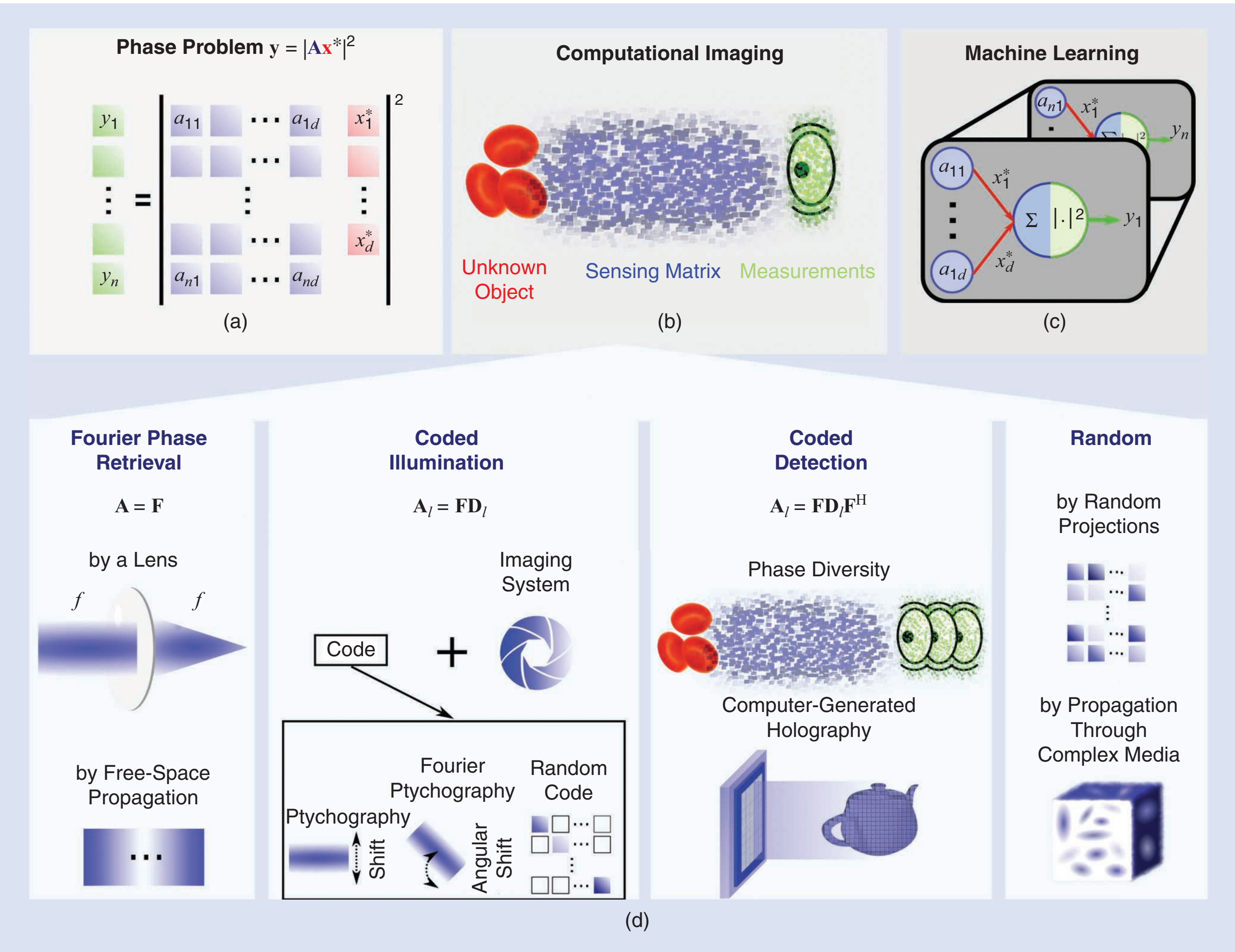


FIGURE 1. The unifying framework for phase-retrieval applications. (a) The nonlinear phase-retrieval equation is encountered in (b) many computational imaging settings and (c) corresponds to single-layer neural network optimization. (d) Depending on the measurement matrix $\mathbf{A}$, four different classes of phase-retrieval problems can be encountered in imaging applications.

In this tutorial, we present an overview of phase retrieval today, from imaging applications to algorithms and theory. We introduce a unifying framework for the various problems in which a phase-retrieval equation arises: diverse and scattered applications can actually be grouped into four main categories depending on the matrix **A**. We also present the latest algorithmic and theoretical advances, in particular, the recent results developed by the theoretical machine learning community during the past few years. The intuition and principles behind these concepts are described without diving into the technical points. These are described in previous works or dedicated reviews, which we reference for interested readers. We hope that our unified overview will foster interaction among research fields and lead to a better understanding of the phase-retrieval equation.

Background

Mathematical model

The phase-retrieval equation (1) relates three mathematical objects. In the context of imaging, one wants to reconstruct an object of interest $\mathbf{x}^*$ from the knowledge of the captured intensities $\mathbf{y}$ and the imaging system modeled by the linear operator **A**. Beyond the field of imaging, this equation often arises over the estimation of a physical quantity $\mathbf{x}^*$ from energy measurements on a sensor, as detectors are often sensitive to the squared modulus of a field. It will be useful to rewrite this equation as

$$y_i = \left|\langle \mathbf{a}_i, \mathbf{x}^* \rangle\right|^2 \quad \text{for } i = 1, \ldots, n \tag{2}$$

where **A** is the vertical concatenation of the rows $\mathbf{a}_i^{\mathrm{H}}$ (with H denoting Hermitian conjugation), and $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathrm{H}}\mathbf{y}$ is the complex inner product. Thus, each measurement is obtained from the squared modulus of a scalar product with a sampling vector $\mathbf{a}_i \in \mathbb{C}^d$. The phase-retrieval problem can then be restated as follows: how many phaseless scalar products are required to reconstruct the unknown vector $\mathbf{x}^*$?

To make algorithms robust to experimental measurements, the model can be enriched by incorporating a general noise channel p_{out} such that $y_i \sim p_{\text{out}}(\cdot||\mathbf{a}_i^{\mathrm{H}}\mathbf{x}^*|^2)$, which typically is additive Gaussian or Poissonian. The noise statistics can, indeed, be leveraged in the maximum-likelihood and Bayesian algorithms.

An important quantity to determine the difficulty of the problem is the oversampling ratio $\alpha = n/d$. It compares the number of measurements to the number of parameters we want to estimate. The greater the oversampling ratio α, the easier the computational problem becomes. This sampling ratio also plays a role in the simpler problem of linear regression $\mathbf{y} = \mathbf{A}\mathbf{x}^*$. There, the solution is not unique—the set of linear equations is underdetermined—for $\alpha < 1$, and it can be solved exactly for $\alpha \geq 1$ in the absence of noise and assuming **A** to be full rank. Similar considerations apply to phase retrieval, although with a more complicated behavior. In this tutorial, we define and present the latest results regarding injectivity (see "Injectivity: When Is the Solution Unique?"), weak recovery (see "Weak Recovery Threshold: When Do Reconstructions Start to Be Possible?"), full recovery (see "Perfect Recovery: When Can We Reconstruct 'Almost Exactly'?"), and asymptotic algorithmic performance (see "Asymptomatic Guarantees: Can We Characterize the Performance of Reconstruction Algorithms?"). These discussions are highlighted in stand-alone theory boxes.

Let us emphasize that phase retrieval is, at first sight, a nonconvex problem. Indeed, if some $\mathbf{x} \neq \mathbf{0}$ is in the solution set $\mathcal{S}$ of the phase-retrieval equation (1), then the whole set $\{e^{i\theta}\mathbf{x}, \theta \in [0, 2\pi)\}$ is included in $\mathcal{S}$, while the middle point $(\mathbf{x} + (-\mathbf{x}))/2 = \mathbf{0}$ is not in $\mathcal{S}$. This global phase ambiguity can be removed, for instance, by a quotient of the solution space or by considering instead the estimation of invariant quantities,

Injectivity: When Is the Solution Unique?

To build a mathematically sound theory of phase retrieval, it is important to understand if the solution to the (noiseless) phase-retrieval problem is unique. Stated differently, for a given measurement matrix $\boldsymbol{A} \in \mathbb{C}^{n\times d}$, is the map $\mathcal{A}_A : \mathbf{x} \in \mathbb{C}^d \to |\boldsymbol{A}\boldsymbol{x}|^2 \in \mathbb{R}^n$ injective? While we do not mention it explicitly, the reader should remember that injectivity is always meant up to a global phase. Since 2014, the following picture has gradually emerged on a complete characterization of injectivity.

Conjecture S1

For any $d \geq 2$ and any full-rank $\mathbf{A} \in \mathbb{C}^{n\times d}$, the following holds:

i. If $n < 4d - 4$, then the map $\mathcal{A}_{\mathbf{A}}$ is noninjective.
ii. If $n \geq 4d - 4$, , then the map $\mathcal{A}_{\mathbf{A}}$ is injective for a "generic" $\mathbf{A}$. We refer to the description of the conjecture in [13] for a more detailed definition of the notion of a "generic" $\mathbf{A}$.

First described as the "$(4d-4)$ conjecture" [13], point ii was later proven in [S1]. Meanwhile, point i has been proven for small dimensions $d \leq 3$ in [13], disproved for $d = 4$, and refined for large dimensions in a probabilistic sense [S2].

In particular, for large $d \gg 1$, the threshold $n/d \simeq 4$ should indeed distinguish between "typically injective" and "typically noninjective" behaviors. In practice, to efficiently recover in the not-too-oversampled regime $n/d \lesssim 4$, this suggests that one should leverage a priori information on the object (e.g., regularization methods).

References

[S1] A. Conca, D. Edidin, M. Hering, and C. Vinzant, "An algebraic characterization of injectivity in phase retrieval," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 2, pp. 346–356, Mar. 2015, doi: 10.1016/j.acha.2014.06.005.

[S2] C. Vinzant, "A small frame and a certificate of its injectivity," in *Proc. 2015 Int. Conf. Sampling Theory Appl. (SampTA)*, pp. 197–200, doi: 10.1109/SAMPTA.2015.7148879.

such as $\mathbf{x}^*(\mathbf{x}^*)^{\mathrm{H}}$. Notwithstanding, the solution set might still be nonconvex. Thus, the general phase-retrieval problem cannot be analyzed directly through the prism of convex optimization.

Neural network representation

As (1) describes a linear operation followed by an elementwise nonlinearity, phase retrieval is equivalent to the training of a single-layer neural network with a quadratic activation function. More precisely, for each $i = 1, \ldots, n$, the input data $\mathbf{a}_i$ are presented to a network with weights $\mathbf{x}$, with a target output $\mathbf{y}$, as depicted in Figure 1(c). Training the network for this regression task amounts to the recovery of $\mathbf{x}^*$ in the phase-retrieval equation.

This corresponds to the simplest neural network architecture with a nontrivial polynomial nonlinearity. Still, it captures the difficulties of training a shallow neural network with a nonconvex loss functional. In particular, convergence is not assured with gradient descent methods, as they may get trapped into local minima or saddle points. More generally, single-layer neural networks belong to the class of generalized linear models. This simple yet rich architecture has attracted the attention of theoreticians, which explains the recent profusion of studies around phase retrieval. This parallel enriches the discussion between physics and machine learning, as we further illustrate throughout this tutorial.

Applications

Unifying framework for phase-retrieval applications

The phase-retrieval problem is central to a large variety of applications in science and technology. Although many of these approaches were developed independently over the years, they share the same phase-retrieval equation (1). Diversity comes notably from the set of computations performed by the matrix $\mathbf{A}$ that depends on the physical model of the imaging system. This domain knowledge sometimes makes it difficult for nonexperts to navigate the landscape of phase-retrieval applications.

To offer a unifying perspective, we are proposing classifying phase retrieval into four major groups: Fourier, coded-illumination, coded-detection, and random models [Figure 1(d)]. Targeted toward imaging, this classification has been designed to facilitate interactions between fields. It may be used by practitioners to

Weak Recovery Threshold: When Do Reconstructions Start to Be Possible?

Complementary to injectivity studies, many recent works have focused on the question of weak recovery—in other words, when does it become possible to nontrivially estimate the signal $\mathbf{x}^*$? In mathematical terms, this means that there exists a procedure that is polynomial time in d and that returns an estimator $\hat{\mathbf{x}}$ satisfying

$$\lim_{d\to\infty} |\langle \hat{\mathbf{x}}, \mathbf{x}^* \rangle| > 0. \tag{S1}$$

Note that an $\hat{\mathbf{x}}$ uniformly sampled on the sphere (i.e., completely random) satisfies with high probability $|\langle \hat{\mathbf{x}}, \mathbf{x}^* \rangle| = \mathcal{O}(1/\sqrt{d}) \to 0$ for $d \to \infty$.

A series of precise results on the weak recovery of phase retrieval under a random-design assumption was obtained in recent years. In particular, under R.1 or R.2, [15] and [16] showed that, if $n, d \to \infty$ with $\alpha = n/d$, there exists a sharp threshold α_{WR} distinguishing between impossible ($\alpha < \alpha_{\mathrm{WR}}$) and possible ($\alpha > \alpha_{\mathrm{WR}}$) weak recovery. For example, in noiseless complex phase retrieval with a Gaussian measurement matrix, this yields $\alpha_{\mathrm{WR}} = 1$, as seen in Figures 2 and 3. Note that the "optimal" spectral method we described in the "Spectral Methods" section has been shown to reach the asymptotic weak recovery threshold; i.e., it satisfies (S11) for any $\alpha > \alpha_{\mathrm{WR}}$ [15], [33].

Perfect Recovery: When Can We Reconstruct "Almost Exactly"?

Crucially, injectivity studies inherently consider "worst case" scenarios on the choice of the signal $\mathbf{x}^*$ to recover and the measurement matrix $\mathbf{A}$, and they do not leverage the high dimensionality of the problem. To go beyond this framework, the perfect recovery (or full recovery) phenomenon has also attracted interest in past years. This question is fundamentally high-dimensional and can be formulated as follows: for $d \gg 1$, when is $\alpha = n/d$ large enough so that, for any typical $\mathbf{x}^*$ and $\mathbf{A}$, there exists an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}^*$ (with $\|\hat{\mathbf{x}}\| = 1$) that satisfies $|\hat{\mathbf{x}}, \mathbf{x}^*| \simeq 1$?

By leveraging prior knowledge on the structure of the signal $\mathbf{x}^*$ (i.e., giving a sense to the typicality), this question has been answered in the very large class of random models R.1 and R.2 [16]. In particular, perfect recovery can be achieved much before the typical injectivity threshold $\alpha = 4$, e.g., at $\alpha \simeq 2$ for Gaussian matrices $\mathbf{A}$ and a uniformly sampled $\mathbf{x}^*$ as seen in Figure 3.

Studies of perfect recovery turn out to be more relevant to practical phase-retrieval setups than injectivity analyses. In these setups, the measurement matrix and signal are not the worst case but, rather, constrained by the properties of the physical problem (e.g., a natural image). Then, one is often willing to accept vanishingly small errors.

draw links between applications and benefit from the expertise of other fields. For theoreticians, it may give directions to follow when defining new theoretical models to study.

Fourier phase retrieval

Fourier phase retrieval was, historically, the first occurrence of this nonlinear equation. Wave propagation over a long distance, referred to as *far-field propagation*, is modeled by a Fourier transform [17], as in $\mathbf{A} = \mathbf{F}$, where $\mathbf{F}$ is a 2D Fourier transform matrix. The first phase-retrieval problems appeared in modalities where optical components are difficult to manufacture, most notably for X-rays or electron imaging [3]. In these cases, the imaging system is often minimalist and only involves far-field propagation. From the determination of molecular structures using X-rays to aberration correction in astronomy [2], Fourier phase retrieval has been associated with major scientific discoveries in the past century.

Since the Fourier transform is a unitary operation (i.e., $\mathbf{F}^{\mathrm{H}} = \mathbf{F}^{-1}$), we necessarily have $n = d$, leading to the oversampling ratio $\alpha = 1$. This problem is fundamentally ill posed, as we only know amplitudes, with no phase information about the Fourier transform of $\mathbf{x}^*$. Additional information on $\mathbf{x}^*$ is required to lift this degeneracy. The famous work of Gerchberg [18] uses another intensity measurement of $\mathbf{x}^*$ directly so that, in (1), $\mathbf{A}$ becomes a vertical concatenation of the identity and Fourier operators. Another common assumption exploits a support and nonnegativity constraint on $\mathbf{x}^*$ [2], which is often valid in astronomy. Other more involved assumptions may be invoked, and they are described in the “Regularization in the Deep Learning Era” section. New applications of Fourier phase retrieval have recently appeared, such as

Asymptotic Guarantees: Can We Characterize the Performance of Reconstruction Algorithms?

A considerable effort has been devoted to characterizing the performance of the diverse algorithms presented in the “Reconstruction Algorithms” section, with a particular emphasis on the theoretical analysis of the error for large n and d, often under random-design assumptions.

For instance, the Bayesian algorithms of the “Bayesian Algorithms” section and, in particular, asymptomatic message passing (AMP) (which conjecturally reaches the optimal polynomial-time performance) can be analytically studied in the high-dimensional limit with a set of tools known as *state evolution* [16], [35]. On one hand, these allow us to establish the theoretical solid lines in Figure 3, which agree very well with finite-size simulations of AMP. On the other hand, the spectral methods described in the “Spectral Methods” section can be tackled more directly using tools of random-matrix theory, and the authors of [12] derive analytical formulas for the asymptotic performance achieved by a wide class of spectral methods. Finally, gradient-based algorithms (see the “Gradient-Based Optimization” section) can also be analyzed in high dimension. For instance, [S3] strikingly explained how overparameterization greatly enhances their performance in phase retrieval. Another avenue to understand the trajectories of (stochastic) gradient descent in general settings is dynamical mean field theory [14], [S4].

We summarize in Table S1 these performance analyses in the specific case of noiseless complex-valued phase retrieval with the Gaussian/independent identically distributed model R.1. Note that many of these results have been extended to more general noise and measurement models.

Table S1. Asymptotic guarantees for the performance of recovery algorithms in the high-dimensional limit for the Gaussian independent identically distributed model R.1 in noiseless complex-valued phase retrieval.

Algorithm	Weak Recovery	Perfect Recovery	Analytic Characterization of the Performance
(Stochastic) gradient descent	—	$n \gtrsim \mathcal{O}(d)^*$	Yes [14], [50], [49]
PhaseLift and PhaseCut	—	$n \gtrsim d \log d$ [10], [30]	—
Optimal spectral method $M(\mathcal{T}^*)$	$n/d \geq 1$	$n/d \to \infty$	Yes [12]
$M(\mathcal{T}^*)$ + gradient descent	$n/d \geq 1$	$n/d \gtrsim 4$ (Figure 2)	Yes (combination of [12], [14], and [50])
AMP	$n/d \geq 1$	$n/d \gtrsim 2.027$ (Figure 3)	Yes [16]

*The result requires informed initialization.

Despite missing pieces and several simplifying assumptions, theoreticians have reached a very precise understanding of the performance of phase-retrieval algorithms. Nevertheless, almost all of these studies rely on high-dimensional random models that, even though they can be quite refined, fail to describe many practical setups, such as coded diffractive imaging with structured Fourier matrices or low-dimensional configurations. In particular, the theory of Bayesian methods and of spectral methods has been empirically shown to fail when the eigenspaces of $\mathbf{A}$ are very structured (e.g., with vanilla Fourier matrices). In this sense, theoretical asymptotic guarantees for structured matrices are still an open problem.

References

[S3] S. Sarao Mannelli, E. Vanden-Eijnden, and L. Zdeborová, “Optimization and generalization of shallow neural networks with quadratic activation functions,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 13,445–13,455.

[S4] M. Celentano, C. Cheng, and A. Montanari, “The high-dimensional asymptotics of first order methods with random data,” 2021, *arXiv:2112.07572*.

speckle correlation imaging for nonline-of-sight imaging or complex media imaging [19], [20]. Ultimately, while the Fourier phase was the first physical model of phase retrieval, it remains a challenging computational problem.

Coded-illumination phase retrieval

More advanced imaging techniques were developed to overcome the overly restrictive assumptions of Fourier phase retrieval. A solution to lift the phase ambiguity consists of shining different illuminations onto the object of interest to capture multiple images after far-field propagation. Each image corresponds to a slightly different view of the same object. This provides the algorithm with complementary and partially redundant information and improves its convergence at the expense of a more involved experimental setup and longer measurement time. This is the general principle underlying coded-illumination phase retrieval. Here, the linear operator $\mathbf{A}$ is a vertical concatenation of operators $\mathbf{A}_l$:

$$\mathbf{A}_l = \mathbf{F}\mathbf{D}_l \quad \text{for } l = 1, \ldots, L \tag{3}$$

where $\mathbf{D}_l$ is a diagonal matrix that corresponds to the coded illumination, $n = Ld$, and the oversampling ratio α is equal to the number L of images. A variety of imaging applications based on coded illuminations are listed as follows:

- Coded-diffraction imaging relies on a modulation near the object plane [21]. This modality allows ample flexibility on the choice and implementation of the modulation scheme, which explains its application in many different imaging settings. Gratings, masks, and tilted wavefronts all represent solutions for structured illumination.
- Ptychography is arguably the most successful implementation of coded-illumination phase retrieval [22]. It records the diffraction pattern of the object while it is scanned by an illumination probe with finite support. Oversampling is linked with the overlap between two consecutive illuminated areas—the higher the overlap, the easier it is to solve the phase-retrieval problem. It has been successfully used with a broad range of electromagnetic waves, from X-rays to the THz regime, as well as for electron microscopy.
- Fourier ptychography is a variant of ptychography in Fourier space where a sequence of tilted plane waves is sent on the sample [4], [5]. In this case, the object $\mathbf{x}^*$ is defined as the Fourier transform of the object transmission function, and the diagonal matrix $\mathbf{D}_i$ corresponds to a finite-support probe that is translated in Fourier space. This technique is used in optical microscopy to achieve high resolution on a large field of view.

Imaging scientists have developed specialized algorithms for each of these applications. For example, the ptychographic iterative engine [22] is a popular projection algorithm, while gradient-based methods are also routinely used. However, the structure present in these sensing matrices prevents a thorough establishment of theoretical results for now.

Coded-detection phase retrieval

Acting on the illumination in the proximity of the object may not always be feasible, for instance, in astronomical observations. Nevertheless, multiple images can be obtained upon modulation on the detection side, leading to coded-detection phase retrieval. This can be done by taking several shots at different defocused positions—a technique derived, again, from astronomy and called *phase diversity* [23] but applied to other imaging fields as well, such as lensless imaging [6].

There are many ways to describe defocus via free-space wave propagation, and all rely on various degrees of approximation of Maxwell's equations. Using the angular-spectrum method, the operator $\mathbf{A}$ is now the vertical concatenation of

$$\mathbf{A}_l = \mathbf{F}\mathbf{D}_l\mathbf{F}^{\mathrm{H}} \quad \text{for } i = 1, \ldots, L, \quad \text{with } \mathbf{D}_l = \mathrm{Diag}\left(e^{iz_l\sqrt{1-|\mathbf{u}|^2}}\right) \tag{4}$$

where $\mathbf{u}$ denotes normalized coordinates in Fourier space, and the unitless curvature z_l is proportional to the defocus distance. Interestingly, this class of phase retrieval also appears outside of imaging for computer-generated holography [7]. There, one wants to produce a 3D hologram—a volumetric intensity pattern. This target defines a set of equations at multiple planes, such that a phase mask is optimized to satisfy these constraints and applied with a wavefront-shaping device.

More general coded-detection strategies may be designed to simplify the phase recovery by using the diversity of the diagonal matrices $\mathbf{D}_l$ in Fourier space. For example, one can add higher order optical aberrations beyond defocus or a physical aperture. Once again, a thorough theoretical understanding of this phase-retrieval class is still lacking. It would especially benefit from more results regarding solvability and algorithm choice.

Random phase retrieval

In random phase retrieval, $\mathbf{A}$ is a random matrix. A common setting is the Gaussian model, in which each element of $\mathbf{A}$ is drawn from an i.i.d. complex Gaussian distribution. Such an operator $\mathbf{A}$ may be seen as a generic linear operation: each sampling vector is independent from the others and adds a little bit of information on the unknown $\mathbf{x}^*$. Contrary to expectations, random phase retrieval arises in practical applications, too, such as imaging in complex media [24]. There, the random linear operator comes from multiple light scattering or random input patterns synthesized by a programmable device.

Distinct from the previous models, randomness greatly facilitates theoretical studies and allows one to derive asymptotic results on recovery guarantees and to characterize the typical behavior of algorithms. These advances are recent and will be described in more details later in the article. The bottom line is that phase retrieval in the random setting is quite well understood, with active research still ongoing.

Some imaging applications exhibit pseudorandom models, such as coded-diffraction imaging with random masks or ptychography with a random probe [22]. They emulate random models, but residual correlations are present between rows of $\mathbf{A}$. This variable amount of randomness makes it challenging to derive precise theoretical guarantees on performance but, to some extent, motivates the use of the latest algorithms designed for the fully random setting.

These observations motivate further work to bridge the gap between experiments and theory. Indeed, to illustrate current efforts, a few theoretical results have recently been generalized beyond the i.i.d. random setting [16]. We can distinguish between two classes of random models, R.1 and R.2, for which strong theoretical results have been reported:

- ***R.1** (i.i.d. random)*: The matrix components of $\mathbf{A}$ satisfy $[\mathbf{A}]_{i,k} \overset{i.i.d.}{\sim} P$, with P a complex (centered) probability distribution with finite moments of all orders (for simplicity).
- ***R.2** (right unitarily invariant)*: $\mathbf{A}$ is a random matrix with probability distribution $P(\mathbf{A})$, and $P(\mathbf{A}) = P(\mathbf{AU})$ for every unitary matrix $\mathbf{U}$. Equivalently, the right eigenvectors of $\mathbf{A}$ are completely delocalized.

As a final remark, the observed robustness of random phase retrieval may inspire new ideas to introduce randomness in experimental settings.

Reconstruction algorithms

A large number of reconstruction algorithms have been proposed to tackle the diversity of phase-retrieval models. We present them here, together with some intuitive explanation and indications on when to use them in practice. The purpose of this overview is to illustrate the inspiring creativity that researchers deployed while designing phase-retrieval algorithms. Dedicated surveys cover more technical descriptions of the recent developments [9], [27].

Projection algorithms

Projection algorithms were the first strategy proposed to solve the phase-retrieval problem. They provide a direct and intuitive way to exploit prior information or several intensity measurements. Thanks to their ease of implementation and flexibility, projection algorithms are still widely used today.

More precisely, each intensity image or prior knowledge corresponds to a constraint labeled from 1 to N. The earliest implementation is the Gerchberg–Saxton algorithm [18], which uses the intensity at two measurement planes as constraints in the object and Fourier planes. The error-reduction algorithm resorts to a modulus constraint in the Fourier plane and a support constraint in the object plane [2]. Other techniques, such as ptychography [22], [28], Fourier ptychography [4], [5], or phase diversity perform projections onto more than two sets to exploit even more information to recover the missing phase.

Each constraint is satisfied by a set of vectors $\mathbf{x} \in \mathbb{C}^d$, and one is aiming to find the intersection of such sets. One, thus, projects on all of the sets sequentially to hopefully converge to a suitable solution. The initial estimate is $\hat{\mathbf{x}}^{(0)}$ and the iterations proceed according to

$$\hat{\mathbf{x}}^{(k+1)} = \Pi_N \circ \cdots \circ \Pi_1(\hat{\mathbf{x}}^{(k)}) \tag{5}$$

where Π_l denotes the projection operators on the solution set of constraint $l = 1, \ldots, N$.

For example, we can write the projector of a modulus constraint for an intensity image $\mathbf{y}_l = |\mathbf{A}_l \mathbf{x}^*|^2$ that corresponds to an invertible linear operator $\mathbf{A}_l$ as

$$\Pi_l(\mathbf{u}) = \mathbf{A}_l^{-1} \mathrm{Diag}\left(\frac{\sqrt{\mathbf{y}_l}}{|\mathbf{A}_l \mathbf{u}|}\right) \mathbf{A}_l \mathbf{u}. \tag{6}$$

This operator effectively replaces the square modulus of $\mathbf{A}_l \mathbf{u}$ by the one obtained from the measurements $\mathbf{y}_l$. Similarly, the projector of a support constraint γ sets all of the values outside the support to zero: $\Pi(\mathbf{u}) = \mathrm{Diag}(\mathbb{1}_\gamma)\mathbf{u}$, where $\mathbb{1}_\gamma$ denotes the indicator function of γ.

While this class of algorithms is widely used in practice, it unfortunately does not necessarily converge to the set of solutions. More advanced algorithms have been suggested for better convergence, such as the hybrid input–output algorithm [2] that uses relaxed projections for Fourier measurements with support constraint. Another strategy replaces the sequential projection with an average to avoid being stuck in loops. When applied to ptychographic algorithms, these concepts gave rise to the difference map and relaxed averaged alternating reflections algorithms [22].

Gradient-based optimization

Gradient-based iterative methods are the methods of choice to optimize nonlinear objective functions in machine learning. This class of algorithms is very flexible, with many variations and acceleration strategies. They have been used to solve the phase-retrieval problem [11], [29], with applications ranging from ptychography and coded-diffraction imaging to defocus-based implementations.

One wants to find an estimate $\hat{\mathbf{x}}$ that minimizes a loss function $\mathcal{L}(\mathbf{x}, \mathbf{y})$, for instance, the square loss $\mathcal{L}_2(\mathbf{x}, \mathbf{y}) = \| \mathbf{y} - |\mathbf{Ax}|^2 \|^2$. At each iteration k, a step of size η_k toward the negative gradient of $\mathcal{L}$ is taken to refine the solution, leading to the process

$$\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} - \eta_k \nabla_{\hat{\mathbf{x}}} \mathcal{L}(\hat{\mathbf{x}}^{(k)}, \mathbf{y}). \tag{7}$$

The choice of the loss function is of crucial importance since it determines the performance of the algorithm. An educated guess of a suited loss function can be made if the noise statistics are known. Typical noise models are Poissonian (for shot-noise limited measurements) or additive Gaussian [5]. In this case, the loss function corresponds to the negative log-likelihood function $-\log(p_{\mathrm{out}}(\mathbf{y} \,||\mathbf{Ax}^*|^2))$. Typically, the loss function $\mathcal{L}_2$ corresponds to the negative log-likelihood of the additive Gaussian noise case.

As the optimization is nonconvex, gradient descent may get stuck in local minima or saddle points. Many acceleration strategies are at our disposal to circumvent these difficulties, from conjugate gradient and Nesterov acceleration to second-order methods. Another classical technique in machine learning to escape local minima is stochastic gradient descent, in which the gradient of the loss function at each iteration is computed using a partial dataset only. Intensity images are often used one at a time, akin to sequential projection algorithms, although this introduces noise in the gradient descent updates.

Gradient descent represents a preferred tool to solve the phase-retrieval equation. Nevertheless, the theoretical bases of gradient-based algorithms still represent an active research topic. For example, the choice of the initial estimate greatly impacts the performance: in the "Spectral Methods" section, we describe algorithms that yield informed initializations.

Convex relaxations

In contrast to gradient descent, which is a universal optimization technique, convex relaxations exploit the quadratic form of the phase-retrieval equation. Convex relaxations recast the phase-retrieval problem using auxiliary variables. We seek a slightly different, or relaxed, formulation of the problem, which makes it convex. Convex optimization then provides a large variety of algorithms, with stronger convergence guarantees. For instance, PhaseLift [10] replaces the unknown vector $\mathbf{x}^* \in \mathbb{C}^d$ by the matrix $\mathbf{X}^* = \mathbf{x}^*(\mathbf{x}^*)^{\mathrm{H}} \in \mathbb{C}^{d\times d}$. The quadratic equation (2) then becomes linear in the components of $\mathbf{X}^*$ since one now has

$$y_i = \mathbf{a}_i^{\mathrm{H}} \mathbf{X}^* \mathbf{a}_i \quad \text{for } i = 1, \ldots, n. \tag{8}$$

In its original form, the phase-retrieval equation would require one to seek a rank-one solution. As this rank constraint is itself nonconvex, PhaseLift looks for a minimal-trace solution instead, leading to the convex program

$$\begin{aligned} &\text{minimize } \mathrm{Tr}(\mathbf{X}), \quad \text{subject to} \quad \mathbf{X} \geq 0 \quad \text{and} \\ &y_i = \mathbf{a}_i^{\mathrm{H}} \mathbf{X} \mathbf{a}_i \quad \text{for } i = 1, \ldots, n. \end{aligned} \tag{9}$$

Written in this form, the problem falls into the category of semidefinite programming, where one can readily use an off-the-shelf solver. The output is then given by the leading eigenvector of the final estimate $\mathbf{X}$. Other convex relaxations have been developed [30], [31]; a detailed list is reviewed in [9].

The issue associated with semidefinite programming lies in its computational complexity, as one needs to build a matrix in $\mathbb{C}^{d\times d}$. To illustrate this issue, consider a 100×100 image, which would correspond to $d = 10^4$. Standard convex-relaxation techniques would then require one to perform optimization on a $d \times d = 10^8$-dimensional space. For this reason, semidefinite-programming strategies have only been used as a proof of concept in real imaging settings, although promising sketching methods to keep the computation tractable have been proposed recently [32].

Spectral methods

To enhance the performance of the possibly expensive methods described in the previous sections, practitioners often require an "informed" initialization $\hat{\mathbf{x}}^{(0)}$, preferably one that is close to the optimum. Recently proposed, algorithms in this privileged class to obtain such initializations at a low computational cost are spectral methods [11]. These methods build an estimate $\hat{\mathbf{x}}$ as the principal eigenvector of a weighted covariance matrix, defined from the intensities $\{y_i\}$ and the sampling vectors $\{\mathbf{a}_i\}$ as

$$\mathbf{M}(\mathcal{T}) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \mathbf{a}_i \mathbf{a}_i^{\mathrm{H}}. \tag{10}$$

Intuitively, since $y_i = |\mathbf{a}_i^{\mathrm{H}} \mathbf{x}^*|^2$, this scalar measurement is minimal when $\mathbf{a}_i$ and $\mathbf{x}^*$ are orthogonal, while it increases when the two vectors are correlated. For any increasing preprocessing function $\mathcal{T}$, the matrix in (10), thus, gives "more weight" to those sampling vectors $\mathbf{a}_i$ that are more aligned with $\mathbf{x}^*$

The performance of a given spectral method heavily depends on the choice of the preprocessing function $\mathcal{T}$. Various choices have been considered in the literature, for example, the trimming scheme $\mathcal{T}(y) = y\mathbb{1}\{|y| \leq t\}$ (with possibly $t = \infty$, meaning, in this case, that $\mathcal{T}$ is the identity). Interestingly, an optimal preprocessing method has been derived for the random models R.1 and R.2 and any type of noise [12], [33]. In particular, in a noiseless setting, it reads $\mathcal{T}^\star(y) = 1 - y^{-1}$ assuming the scalings $\|\mathbf{x}^*\|^2 = 1$ and $(1/n)\mathrm{Tr}[\mathbf{A}\mathbf{A}^{\mathrm{H}}/d] = 1$.

In practice, the simplicity of spectral methods ensures that their computational cost remains very low. The leading eigenvector can basically be recovered with power iterations or its accelerated variants. For instance, [12] and [33] show that the largest eigenvalue of $\mathbf{M}(\mathcal{T}^\star)$ concentrates on $\lambda_{\max} = 1$ when recovery is possible, so that one can use very efficient inverse iterations, i.e., power iterations on $(\mathbf{M}(\mathcal{T}^\star) - \mathrm{I})^{-1}$. Importantly, the convergence toward the leading eigenvector is ensured, a guarantee that is missing in many other nonlinear optimization strategies. Unlike convex-relaxation methods, spectral methods do not involve a lifting process either, thus simplifying their adoption in imaging settings such as ptychography [34]. The solution estimated via spectral methods is then typically refined with an iterative optimization algorithm, such as gradient descent.

In Figure 2, we illustrate the performance of some implementations of spectral methods in the noiseless phase-retrieval problem with $\mathbf{A}$ being a complex Gaussian matrix. The optimal spectral method $\mathbf{M}(\mathcal{T}^\star)$ yields almost perfect correlation with the true solution $\mathbf{x}^*$, even without subsequent refinement by gradient descent. Results are also compared with bare gradient descent reconstructions. They show how all of these theoretical findings transfer very well to the recovery of a real image.

Bayesian algorithms

Bayesian estimators are a widely used and powerful ensemble of algorithms that leverage the properties of the posterior distribution, which is given by Bayes' law:

$$\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{A}) = \frac{p_{\mathrm{out}}(\mathbf{y}||\mathbf{A}\mathbf{x}|^2) p_0(\mathbf{x})}{\mathbb{P}(\mathbf{y}|\mathbf{A})}. \tag{11}$$

Here, $\mathbf{y}$ are the observations; p_{out} is the likelihood; and p_0 is a prior distribution, typically Gaussian if one does not want to enforce any particular structure on the reconstructed image. Bayesian algorithms are not based on the minimization of a loss function but, rather, on the computation of estimates defined from the posterior distribution. Thus, they can potentially exploit more efficiently the available information in the measurements. A typical Bayesian estimator is the minimal mean-square error estimator $\hat{\mathbf{x}}_{\mathrm{MMSE}} = \mathrm{argmin}_{\mathbf{x}} \mathbb{E}_{\mathbf{x}'}[\|\mathbf{x}' - \mathbf{x}\|^2 | \mathbf{y}, \mathbf{A}] = \mathbb{E}[\mathbf{x}|\mathbf{y}, \mathbf{A}]$. However, the computation of such estimators—defined on the posterior distribution—is a hard problem in general since sampling from the posterior is often an exponential-time procedure with classical Monte Carlo–Markov chain approaches.

In this regard, an important class of iterative Bayesian algorithms is approximate message passing (AMP) [35], which is based on high-dimensional asymptotics of the belief-propagation algorithm [36]. The power of this class of algorithms arises from several insights. First, AMP is known to be optimal among a large class of polynomial-time algorithms for many random models

[37] and is conjectured to be so for many more. Moreover, it is well understood, in the sense that one can analytically track its performance in the high-dimensional limit. However, AMP is computationally heavy compared to simpler algorithms, such as spectral methods or gradient-based optimization procedures. More generally, the main limitation of Bayesian algorithms is that they assume the knowledge of the generative model (i.e., of the model of the noise present in the observations and of the distribution of the random sensing matrix). While algorithmic procedures such as AMP can be used off the shelf beyond these

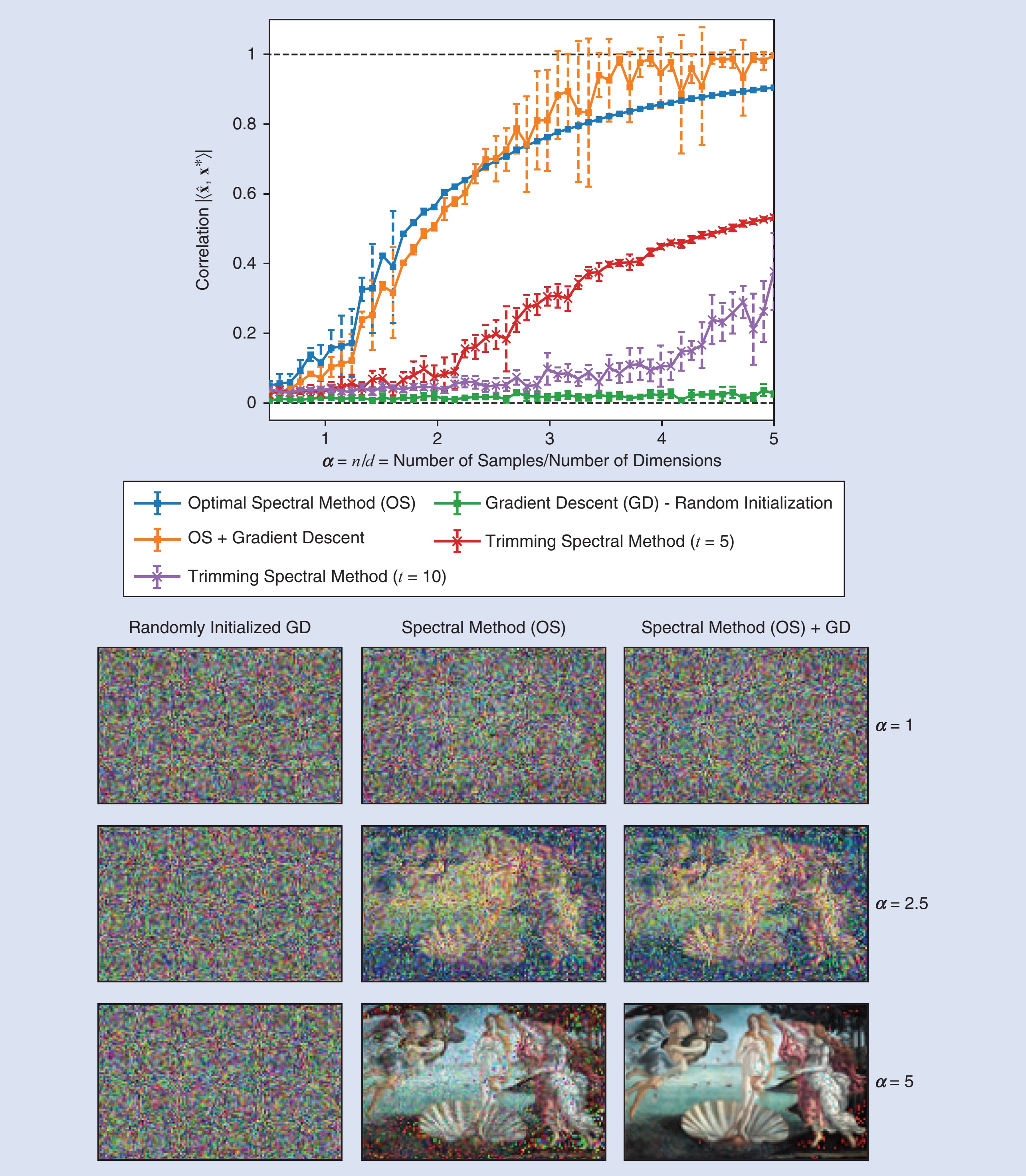


FIGURE 2. The correlation $|\langle \hat{\mathbf{x}}, \mathbf{x}^* \rangle|$ achieved by spectral methods in noiseless phase retrieval with a complex Gaussian matrix for the recovery of an image. The optimal spectral method (OS) is compared with the trimming method with different values of t. Gradient descent is initialized with the OS, showing zero error already at $\alpha \gtrsim 4$. Finally, a comparison with randomly initialized gradient descent (GD) is shown, which does not achieve a positive correlation at any $\alpha \leq 5$. In the last two columns, the optimal spectral method was chosen.

assumptions (e.g., for a nonrandom $\mathbf{A}$), their performance and convergence properties are no longer guaranteed. For example, while AMP achieves optimal polynomial-time performance in phase retrieval with random unitary matrices $\mathbf{A}$, it performs very poorly when $\mathbf{A}$ is a vanilla Fourier matrix [16].

We show in Figure 3 the performance of AMP in noiseless phase retrieval with a complex Gaussian matrix. We obtain a very good agreement with the theory (see "Asymptomatic Guarantees: Can We Characterize the Performance of Reconstruction Algorithms?") and, as for the spectral

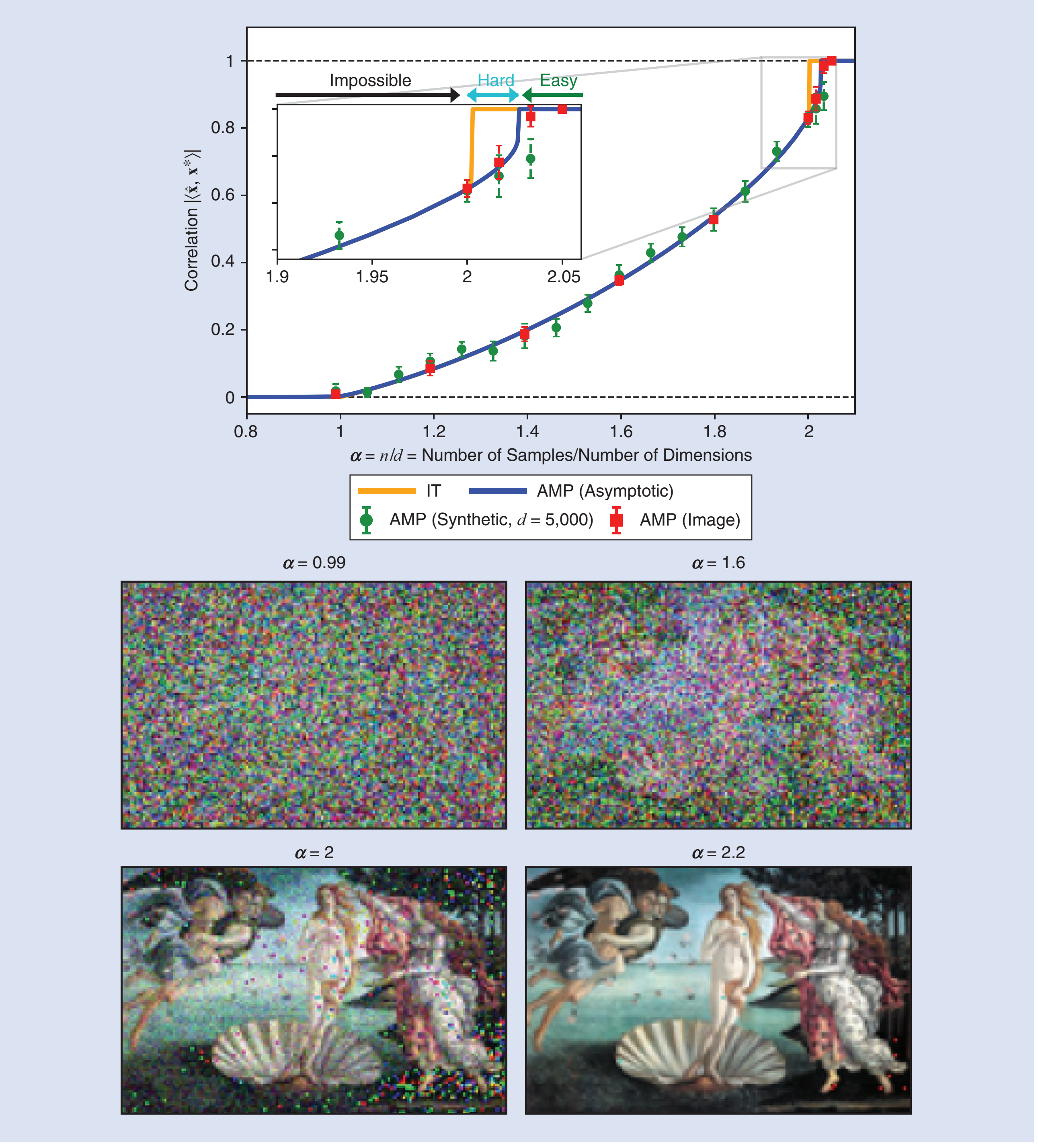


FIGURE 3. The correlation $|\langle \hat{\mathbf{x}}, \mathbf{x}^* \rangle|$ achieved by AMP in noiseless phase retrieval with a complex Gaussian matrix. We compare the theoretical prediction, known as state evolution (blue curve), with actual runs of the algorithm. The results are very consistent whether the signal $\mathbf{x}^*$ is synthetic (i.e., random) or an actual image, achieving perfect recovery for $\alpha \gtrsim 2.03$. The inset of the left column compares the performance of AMP with the information theoretic (IT) performance, which is the optimal error that any algorithm can reach, no matter its running time. We uncover a computational gap (or "hard phase") for $2 < \alpha \lesssim 2.03$, for which AMP is not able to reach the IT performance [16]. We use an efficient modular implementation of AMP in a wide class of inference procedures (including phase retrieval), publicly available in [38].

methods [33], a very consistent performance for the recovery of real images.

Regularization in the deep learning era

Conventional regularization

When faced with a difficult phase-retrieval problem, an appropriate regularization can be an efficient solution to enhance recovery of the desired images. It introduces prior information on the class of images to recover, which will help the reconstruction algorithm. This concept has long been associated with phase retrieval since the first projection algorithms were based on support and positivity constraints.

Most regularization approaches can be formulated as the optimization problem

$$\underset{\mathbf{x}\in\mathbb{C}^d}{\operatorname{argmin}}\,\{\mathcal{L}(\mathbf{x},\mathbf{y})+\mathcal{R}(\mathbf{x})\}. \tag{12}$$

The first term enforces consistency with the intensity measurements $\mathbf{y}$ (the data-fidelity term). It corresponds to the metric in the gradient-based reconstruction algorithms discussed in the "Gradient-Based Optimization" section. The second term penalizes unrealistic estimates to favor suitable properties on $\mathbf{x}$. In the Bayesian setting of (11) with maximum-likelihood estimation, the two terms correspond to $p_{\text{out}}(\mathbf{y}||\mathbf{Ax}|^2)$ and $p_0(\mathbf{x})$, respectively. In the "Bayesian Algorithms" section, the prior was set to a generic Gaussian distribution, but it can be modified to introduce other prior information on the recovered image.

Indicator functions may be used to enforce nonnegativity or support constraints. Another frequently studied regularization term is $\|\mathbf{x}\|_1$, the L_1 norm of $\mathbf{x}$. It promotes sparsity in the final reconstruction and bears deep links with the field of compressed sensing. Sparsity in a transform domain can often be interesting for images, such as sparsity of $\mathbf{Lx}$, with $\mathbf{L}$ the discrete gradient operator for total-variation regularization. To tackle these potentially nondifferentiable composite-optimization problems, proximal algorithms alternatively minimize the loss $\mathcal{L}$ and the regularization $\mathcal{R}$ using gradients or projection operators. A large variety of algorithms have been designed for this task. We refer readers to the tutorial [27] on conventional regularization and sparsity for more details.

Machine learning algorithms and, in particular, convolutional neural networks (CNNs), have recently emerged as powerful new tools to analyze natural images. Remarkably, they achieve state-of-the-art performance on many image benchmarks, such as classification, object detection, denoising, or image generation. Not surprisingly, CNNs have also been deployed for image reconstruction, as they can provide a very strong prior for natural images. This prior consistently outperforms classical regularization on natural images. There are several ways to introduce it in our phase-retrieval algorithms.

Plugging in a denoiser

CNNs are remarkably effective for denoising—in other words, for the retrieval of an image from noisy observations. They are typically trained on a large dataset made of ground-truth images along with their corrupted versions, which corresponds to a self-supervised learning scheme.

Denoising networks can then be deployed at each iteration of an optimization framework. This applies in particular to plug-and-play (PnP) methods, which are a class of proximal algorithms commonly used to solve (12). These solvers alternate between two distinct steps: 1) decrease data-fidelity loss and (b) correct the current guess by enforcing regularization. This class of algorithms is very flexible. For example, step 1 can represent a proximal operator of the data-fidelity term or a gradient descent iteration, leading to the alternating-direction method of multipliers and fast iterative shrinkage-thresholding algorithm, respectively. As discussed in the "Conventional Regularization" section, step 2 has traditionally been a proximal operator of a well-defined regularization term $\mathcal{R}$. PnP methods replace this step by the denoising CNN.

Even though convergence is, in general, not assured and can only be proven in restricted settings, PnP with a denoising network is extremely flexible and has been used with great success in many applications. For instance, such an approach enables computationally efficient large-scale phase retrieval with state-of-the-art performance [39].

Likewise, regularization by denoising [40] also exploits inner-loop denoisers to improve the quality of the reconstructions. However, the regularization term is now

$$\mathcal{R}(\mathbf{x})=\mathbf{x}^{\mathrm{H}}(\mathbf{x}-f(\mathbf{x})) \tag{14}$$

where $f:\mathbb{C}^d\to\mathbb{C}^d$ is the denoiser. This formulation makes the regularization term differentiable, which offers an interesting alternative to PnP. See [41] for a more in-depth discussion.

Generative models

Generative adversarial networks (GANs) oppose two networks: a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$. The goal of the generator is to generate realistic images that follow the distribution of a dataset of natural images. It receives as input a random vector $\mathbf{z}$, also called the *latent variable*, and outputs $\mathbf{x}=\mathcal{G}(\mathbf{z})$. Then, the task of the discriminator is to choose whether a given image is from the real dataset or some output of the generator. In this adversarial scheme, the two networks are trained together. Throughout training, the generator learns a manifold of realistic images, while the discriminator gets better and better at discerning real and fake images.

To use such a generator for phase retrieval, one constrains the image $\mathbf{x}$ to be an output of the generator $\mathbf{x}=\mathcal{G}(\mathbf{z})$ [42]. Gradient descent is performed directly on the latent variable $\mathbf{z}$. This reparameterization constrains the search space to the learned manifold of generator outputs. With this point of view, one can not only produce a point estimate of the unknown signal but also perform a Bayesian sampling of the posterior distribution. This enables the computation of statistical estimates, such as the spatial distribution of the variance of the final image [43].

Deep image prior

The machine learning approaches of the "Plugging in a Denoiser" and "Generative Models" sections rely on a dataset of images for training. This raises questions about overfitting: by design, the final image will resemble a realistic image from the training set, while crucial distinctive information may be discarded. Deep image prior (DIP) has been proposed to remove any dependence on a training set. Here, one assumes that the estimates are generated from a CNN $\mathbf{x} = \mathcal{G}_{\mathbf{w}}(\mathbf{z})$, with $\mathbf{w}$ the weights of the network and $\mathbf{z}$ a random input. Optimization is then directly performed on the weights $\mathbf{w}$ of the CNN. DIP has been implemented in several phase-retrieval applications [44], [45].

While both the GAN and DIP approaches rely on a generative network, their reparameterizations differ considerably. The former optimizes a low-dimensional latent vector $\mathbf{z}$, while the second tunes the CNN weights $\mathbf{w}$. The dimension of $\mathbf{w}$ is typically larger than the dimension of $\mathbf{x}$. As a result, DIP can overfit any input $\mathbf{x}$, even random noise. Nevertheless, it has been observed that it naturally converges at first toward smooth natural images. Regularization comes from the architecture of the CNN itself but requires early stopping to avoid fitting noise. In the end, the exact principle behind DIP regularization remains an open question.

Discussion

Historically, the first attempts at the application of deep learning targeted a direct inversion of the phase-retrieval problem. From the intensity measurements $\mathbf{y}$, they trained a network to output the desired image $\mathbf{x}$ and applied it to ptychography [46] or holography [25]. This strategy corresponds to an usual supervised learning scheme. However, the algorithms following this strategy are specific to one particular imaging application with a specific forward model, noise level, or camera sampling. This greatly limits their range of applicability. The methods described in the "Plugging in a Denoiser," "Generative Models," and "Deep Image Prior" sections belong to a second generation of algorithms with increased robustness and flexibility.

All of these various strategies demonstrate how deep learning methods offer a powerful alternative to classical regularization. They open new perspectives for solving the difficult nonlinear equation of phase retrieval. While our understanding is progressing fast, many of these methods remain black boxes that still need to be deciphered. In practice, one also needs to be aware of the notorious instabilities of neural networks [26], which can sometimes create artifacts (hallucinations) or remove certain features of interest. These issues are particularly troublesome and difficult to detect. The final result may look deceptively realistic (by reason of the regularization inherent with deep learning) even though the algorithm has actually failed.

From their inception, phase-retrieval algorithms have relied on prior information about the object to reconstruct. Regularization has gone a long way and evolved substantially, from nonnegativity and support constraints to deep learning, but the full potential and limits of the most recent tools at our disposal are not yet clearly established. The majority of the network-based proposed works require a training dataset, which may impede certain applications, typically in medical setups. DIP and other unsupervised schemes offer an interesting alternative and may pave the way to powerful and controlled regularization strategies based on learning.

Conclusion

We have described the many facets of phase retrieval, a problem that is central to a broad range of applications in computational optics. Various experimental configurations and reconstruction algorithms have been proposed to solve the missing-phase problem. To classify such a diversity of applications, we have outlined a unifying framework of many phase-retrieval models. We have also provided a concise overview of the latest phase-retrieval algorithms to help practitioners choose one that fits their own application.

Viewed through the prism of machine learning, the links between phase retrieval and neural networks are twofold. First, the parallel between the defining equation (1) and single-layer neural networks has enabled the application of many theoretical machine learning studies, pushing forward our understanding of this nonlinear equation. Second, the unrivaled performance of neural networks when dealing with images brings strong promises for future applications but also comes with challenges on how to counteract unstability issues.

Regarding future developments, the introduction of more randomness in the illumination and/or detection could facilitate phase recovery in experiments by benefiting from the strong theoretical results obtained for the random setting. However, many realistic phase-retrieval models still remain unexplored from a theoretical viewpoint, calling for more studies. With this document, we hope to promote a fruitful exchange between different disciplines, from applied physics to theoretical computer science.

Acknowledgment

Jonathan Dong, Thanh-an Pham, and Michael Unser acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement 101020573 FunLearn). Lorenzo Valzania acknowledges funding from Swiss National Science Foundation (Grant P400P2_199329). Sylvain Gigan acknowledges funding from the ERC under the European Union's Horizon 2020 research and innovation program (Grant Agreement 724473 SMARTIES).

Authors

Jonathan Dong (jonathan.dong@epfl.ch) is with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland.

Lorenzo Valzania (lorenzo.valzania@lkb.ens.fr) is with Laboratoire Kastler Brossel, École Normale Supérieure–Paris Sciences et Lettres Research University, Sorbonne Université, Centre National de la Recherche Scientifique Unité Mixte de Recherche 8552, Collège de France, 75005 Paris, France.

Antoine Maillard (antoine.maillard@math.ethz.ch) is with the Department of Mathematics and the Institute for Mathematical Research, Eidgenössische Technische Hochschule Zürich, 8092 Zürich, Switzerland.

Thanh-an Pham (thanh-an.pham@epfl.ch) was with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne 1015, Switzerland. He is now with the 3D Optical Systems Group, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Sylvain Gigan (sylvain.gigan@lkb.ens.fr) is with Laboratoire Kastler Brossel, École Normale Supérieure–Paris Sciences et Lettres Research University, Sorbonne Université, CNRS UMR 8552, Collège de France, 75005 Paris, France.

Michael Unser (michael.unser@epfl.ch) is with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland.

References

[1] D. Sayre, "Some implications of a theorem due to Shannon," *Acta Crystallographica*, vol. 5, no. 6, pp. 843–843, Nov. 1952, doi: 10.1107/S0365110X52002276.

[2] J. R. Fienup and C. Dainty, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery, Theory and Application*, vol. 231, H. Stark, Ed. Orlando, FL, USA: Academic, Jan. 1987, p. 275.

[3] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature*, vol. 400, no. 6742, pp. 342–344, Jul. 1999, doi: 10.1038/22498.

[4] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nature Photon.*, vol. 7, no. 9, pp. 739–745, Jul. 2013, doi: 10.1038/nphoton.2013.187.

[5] L.-H. Yeh et al., "Experimental robustness of Fourier ptychography phase retrieval algorithms," *Opt. Exp.*, vol. 23, no. 26, pp. 33,214–33,240, Dec. 2015, doi: 10.1364/OE.23.033214.

[6] Y. Wu and A. Ozcan, "Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring," *Methods*, vol. 136, pp. 4–16, Mar. 2018, doi: 10.1016/j.ymeth.2017.08.013.

[7] J. Zhang et al., "3D computer-generated holography by non-convex optimization," *Optica*, vol. 4, no. 10, pp. 1306–1313, Oct. 2017, doi: 10.1364/OPTICA.4.001306.

[8] J. Chang et al., "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, Aug. 2018, doi: 10.1038/s41598-018-30619-y.

[9] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," *Acta Numer.*, vol. 29, pp. 125–228, May 2020, doi: 10.1017/S0962492920000069.

[10] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Aug. 2013, doi: 10.1002/cpa.21432.

[11] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015, doi: 10.1109/TIT.2015.2399924.

[12] W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2347–2356, May 2019, doi: 10.1109/TSP.2019.2904918.

[13] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, "Saving phase: Injectivity and stability for phase retrieval," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 106–125, Jul. 2014, doi: 10.1016/j.acha.2013.10.002.

[14] F. Mignacco, P. Urbani, and L. Zdeborová, "Stochasticity helps to navigate rough landscapes: Comparing gradient-descent-based algorithms in the phase retrieval problem," *Mach. Learn., Sci. Technol.,* vol. 2, no. 3, p. 035029, Jul. 2021, doi: 10.1088/2632-2153/ac0615.

[15] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *Found. Comput. Math.*, vol. 19, no. 3, pp. 703–773, Jun. 2019, doi: 10.1007/s10208-018-9395-y.

[16] A. Maillard, B. Loureiro, F. Krzakala, and L. Zdeborová, "Phase retrieval in high dimensions: Statistical and computational phase transitions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 1–12.

[17] J. W. Goodman, *Introduction to Fourier Optics*. Englewood, CO, USA: Roberts and Company Publishers, 2005.

[18] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, Jan. 1972.

[19] O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nature Photon.*, vol. 8, no. 10, pp. 784–790, Oct. 2014, doi: 10.1038/nphoton.2014.189.

[20] C. A. Metzler et al., "Deep-inverse correlography: Towards real-time high-resolution non-line-of-sight imaging," *Optica*, vol. 7, no. 1, pp. 63–71, Jan. 2020, doi: 10.1364/OPTICA.374026.

[21] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 2, pp. 277–299, Sep. 2015, doi: 10.1016/j.acha.2014.09.004.

[22] J. Rodenburg and A. Maiden, "Ptychography," in *Springer Handbook of Microscopy*, P. W. Hawkes and J. C. H. Spence, Eds. Cham, Switzerland: Springer-Verlag, 2019, pp. 819–904.

[23] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Amer. A*, vol. 9, no. 7, pp. 1072–1085, Jul. 1992, doi: 10.1364/JOSAA.9.001072.

[24] C. A. Metzler et al., "Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset," in *Proc. 2017 IEEE Int. Conf. Comput. Photography (ICCP)*, pp. 1–16, doi: 10.1109/ICCPHOT.2017.7951483.

[25] Y. Rivenson et al., "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light, Sci. Appl.*, vol. 7, no. 2, p. 17,141, Feb. 2018, doi: 10.1038/lsa.2017.141.

[26] N. M. Gottschling, V. Antun, B. Adcock, and A. C. Hansen, "The troublesome kernel: Why deep learning for inverse problems is typically unstable," 2020, *arXiv:2001.01258*.

[27] Y. Shechtman et al., "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015, doi: 10.1109/MSP.2014.2352673.

[28] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy*, vol. 109, no. 10, pp. 1256–1262, Sep. 2009, doi: 10.1016/j.ultramic.2009.05.012.

[29] M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: A nonlinear optimization approach," *Opt. Exp.*, vol. 16, no. 10, pp. 7264–7278, May 2008, doi: 10.1364/OE.16.007264.

[30] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *Math. Program.*, vol. 149, nos. 1–2, pp. 47–81, Feb. 2015, doi: 10.1007/s10107-013-0738-9.

[31] T. Goldstein and C. Studer, "PhaseMax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, Apr. 2018, doi: 10.1109/TIT.2018.2800768.

[32] A. Yurtsever, M. Udell, J. Tropp, and V. Cevher, "Sketchy decisions: Convex low-rank matrix optimization with optimal storage," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, vol. 54, pp. 1188–1196.

[33] A. Maillard, F. Krzakala, Y. M. Lu, and L. Zdeborová, "Construction of optimal spectral methods in phase retrieval," in *Proc. 2nd Math. Sci. Mach. Learn.*, 2021, vol. 145, pp. 1–28.

[34] L. Valzania, J. Dong, and S. Gigan, "Accelerating ptychographic reconstructions using spectral initializations," *Opt. Lett.*, vol. 46, no. 6, pp. 1357–1360, Mar. 2021, doi: 10.1364/OL.406156.

[35] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in *Proc. 2016 50th Asilomar Conf. Signals, Syst. Comput.*, pp. 1525–1529, doi: 10.1109/ACSSC.2016.7869633.

[36] M. Mézard and A. Montanari, *Information, Physics, and Computation*. London, U.K.: Oxford Univ. Press, 2009.

[37] M. Celentano, A. Montanari, and Y. Wu, "The estimation error of general first order methods," in *Proc. 33rd Conf. Learn. Theory*, 2020, vol. 125, pp. 1078–1141.

[38] A. Baker, B. Aubin, F. Krzakala, and L. Zdeborová, "Tree-AMP: Compositional inference with tree approximate message passing," 2021, *arXiv:2004.01571*.

[39] X. Chang, L. Bian, and J. Zhang, "Large-scale phase retrieval," *eLight*, vol. 1, no. 1, pp. 1–12, Sep. 2021, doi: 10.1186/s43593-021-00004-w.

[40] C. Metzler et al., "prDeep: Robust phase retrieval with a flexible deep network," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, vol. 80, pp. 3501–3510.

[41] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging," 2022, *arXiv:2203.17061.*

[42] P. Hand, O. Leong, and V. Voroninski, "Phase retrieval under a generative prior," in *Proc. 32nd Int. Conf. Adv. Neural Inf. Process. Syst.*, 2018, vol. 31, pp. 9154–9164.

[43] P. Bohra, T.-A. Pham, J. Dong, and M. Unser, "Bayesian inversion for nonlinear imaging models using deep generative priors," 2022, *arXiv:2203.10078*.

[44] F. Wang et al., "Phase imaging with an untrained neural network," *Light, Sci. Appl.*, vol. 9, no. 1, pp. 1–7, May 2020, doi: 10.1038/s41377-020-0302-3.

[45] E. Bostan et al., "Deep phase decoder: Self-calibrating phase microscopy with an untrained deep neural network," *Optica*, vol. 7, no. 6, pp. 559–562, Jun. 2020, doi: 10.1364/OPTICA.389314.

[46] A. Kappeler et al., "Ptychnet: CNN based Fourier ptychography," in *Proc. 2017 IEEE Int. Conf. Image Process. (ICIP)*, pp. 1712–1716, doi: 10.1109/ICIP.2017.8296574.

SP