Yonina C. Eldar, Nathaniel Hammen, and Dustin G. Mixon

Recent Advances in Phase Retrieval

In many applications in science and engineering, one is given the modulus squared of the Fourier transform of an unknown signal and then tasked with solving the corresponding inverse problem, known as phase retrieval. Solutions to this problem have led to some noteworthy accomplishments, such as identifying the double helix structure of DNA from diffraction patterns, as well as characterizing aberrations in the Hubble Space Telescope from point spread functions. Recently, phase retrieval has found interesting connections with algebraic geometry, low-rank matrix recovery, and compressed sensing. These connections, together with various new imaging techniques developed in optics, have spurred a surge of research into the theory, algorithms, and applications of phase retrieval. In this lecture note, we outline these recent connections and highlight some of the main results in contemporary phase retrieval.

Relevance

Phase retrieval finds applications in areas such as optics, X-ray crystallography, astronomical imaging, speech processing, and computational biology. Every application of this inverse problem encounters several fundamental questions: How do we reconstruct the desired signal from the measurements? To what extent is the reconstruction unique and/or stable? Can we develop new measurement devices that allow for robust signal recovery? These questions form the basis for the notes presented below.

Prerequisites

We assume the reader has a basic understanding of linear algebra, optimization, and probability. Some familiarity with compressed sensing is helpful but not necessary.

Problem statement

In standard applications of phase retrieval, we receive measurements of the modulus squared of the Fourier transform of an unknown signal x_0, $y = |Fx_0|^2$, where the magnitude is taken component-wise. (For simplicity, we model x_0 as a vector in $\mathbb{C}^n$ so that the Fourier operator F may be represented by a matrix.) Observe that, without additional information, this inverse problem is terribly ill posed. For example, if y is the vector of all ones, then x_0 can be any standard basis element multiplied by an arbitrary phase factor. In general, the set of solutions can have as many as n real degrees of freedom. Examples of these degrees of freedom include the so-called trivial ambiguities: $|Fz|^2 = |Fx|^2$ if $z = e^{i\phi}x$ for some $\phi \in [0, 2\pi)$, if z is a translation of x, or if z is the conjugate reversal of x.

Phase retrieval finds applications in areas such as optics, X-ray crystallography, astronomical imaging, speech processing, and computational biology.

Example 1

Suppose that $x = (2, i, 0, 0)$. *Then* $|Fx|^2 = (5, 9, 5, 1)$. *Let* z *be any of* $(2i, -1, 0, 0)$, $(0, 2, i, 0)$, *or* $(0, 0, -i, 2)$. *These are trivial ambiguities for which* $|Fz|^2 = (5, 9, 5, 1)$. *However, trivial ambiguities are not the only solutions. For example,* $z = (\sqrt{5} + 1/2, i, \sqrt{5} - 1/2, -i)$ *is not a trivial ambiguity, but still* $|Fz|^2 = (5, 9, 5, 1)$.

To obtain a well-posed problem, we must acquire additional information. Historically, this has been accomplished by imposing structure on the signal. For example, one might assume that x_0 is real and has compact support. In fact, this uniquely determines almost every x_0 up to trivial ambiguities when taking a two-dimensional (2-D) Fourier transform [or three-dimensional (3-D) or higher] [5]. However, unfortunately, in 1-D there is no uniqueness even if the signal support is bounded. Furthermore, even in settings in which uniqueness is guaranteed, there is no known general algorithm to find the unknown signal from its Fourier magnitude. A more recent class of signal structures used in phase retrieval relies on sparsity [4]. A vector is defined to be *k-sparse* if it has at most k nonzero entries. Sparsity priors have been used extensively in many fields of engineering and statistics and

Digital Object Identifier 10.1109/MSP.2016.2565061
Date of publication: 2 September 2016

are known to closely approximate various classes of images and signals.

Another way to obtain a well-posed problem is to collect additional intensity measurements. For example, in diffractive imaging, one may implement multiple structured illuminations of the form $|FD_jx_0|^2$, where each D_j is a known diagonal matrix. In other applications, we can take redundant measurements via the short-time Fourier transform. This approach has been used in speech and audio processing, in measurements of ultrashort laser pulses via frequency resolved optical gating, and in ptychographical diffractive imaging, among others.

To account for the apparent multitude of plausible intensity measurements (such as structured illuminations or the short-time Fourier transform), we consider a general phase retrieval setting in which we receive $y = |Ax_0|^2$ for some known matrix $A \in \mathbb{C}^{m\times n}$. We then seek to solve the following program:

$$\text{find } x \text{ subject to } |Ax|^2 = y, x \in S, \quad (1)$$

where $S \subseteq \mathbb{C}^n$ corresponds to the imposed structure. In this lecture note, we focus on cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse vectors. For both settings, we discuss transformations A that allow for (1) to uniquely determine x_0 and consider algorithms that were recently designed to solve (1) for various choices of A. The results we present throughout are surveyed in [6], [8], and [10], unless indicated otherwise.

Uniqueness

For a fixed A, we are interested in whether (1) has a "unique" solution for every $x_0 \in S$ (or for most $x_0 \in S$). We focus on the cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse vectors. Note that in both cases, $x \in S$ if and only if $e^{i\phi}x \in S$ for every $\phi \in [0, 2\pi)$, which means (1) never has a unique solution in the literal sense. To account for this technicality, we say (1) has a unique solution (up to a global phase factor) if every solution lies in the set $[x_0] := \{e^{i\phi}x_0 : \phi \in [0, 2\pi)\}$.

Notice that the set $[x_0]$ is determined by the outer product $x_0x_0^*$ (and vice versa). Let a_i^* denote the ith row of A (here, $a_i \in \mathbb{C}^n$ is a column vector, and a_i^* denotes its conjugate transpose). Then

$$y_i = |a_i^*x_0|^2 = a_i^*x_0 \cdot \overline{a_i^*x_0} = a_i^*x_0x_0^*a_i. \quad (2)$$

Consider the case where $S = \mathbb{C}^n$. In this setting, (2) implies that (1) has a unique solution for every $x_0 \in \mathbb{C}^n$ precisely when the mapping $x_0x_0^* \mapsto \{a_i^*x_0x_0^*a_i\}_{i=1}^m$ is one to one. Recent research has investigated the number of measurements m that are necessary or sufficient for this map to be one to one. For example, it has been shown that the inequality $m \geq 4n - O(\log n)$ is a necessary condition. Conversely, for almost every $A \in \mathbb{C}^{m\times n}$ with $m \geq 4n - 4$, (1) has a unique solution for every $x_0 \in \mathbb{C}^n$. Whether such A exist when $m < 4n - 4$ remains an open problem for general n. We know this is impossible when n has the form $n = 2^k + 1$ and yet possible when $n = 4$ [11]. Alternatively, for almost every $A \in \mathbb{C}^{m\times n}$ with $m \geq 2n$, (1) has a unique solution for almost every $x_0 \in \mathbb{C}^n$, and no such A exists when $m < 2n - 1$.

In many real-world applications, A exhibits some sort of Fourier structure. For example, in the classical setting in which x_0 is compactly supported, A may be viewed as an oversampled Fourier matrix. As mentioned before, 1-D uniqueness from Fourier measurements cannot be guaranteed in general. To achieve uniqueness beyond trivial ambiguities, consider the model in which A is a $kn \times n$ matrix composed of k different $n \times n$ blocks of the form FD_j, where F is the $n \times n$ discrete Fourier transform matrix and each D_j is some diagonal matrix. This model is called the structured illumination model, and each D_j is referred to as a mask. While four such masks are required to determine every possible x_0 (by the aforementioned discussion), we currently only know how to do so with $O(\log n)$ masks. On the other hand, we do know how to determine almost every possible x_0 with only two masks, which matches the above theory; in particular, the two masks may be taken to be the identity matrix and $\text{diag}(0, 1, \ldots, 1)$, as illustrated in the following example.

Example 2

Suppose that x_0 is a vector of length $n = 4$ and we measure it using the two masks $D_0 = I$ and $D_1 = D = \text{diag}(0, 1, 1, 1)$. The resulting structured illumination matrix has the form

$$A = \begin{bmatrix} FI \\ FD \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^9 \\ 0 & 1 & 1 & 1 \\ 0 & \omega^1 & \omega^2 & \omega^3 \\ 0 & \omega^2 & \omega^4 & \omega^6 \\ 0 & \omega^3 & \omega^6 & \omega^9 \end{bmatrix},$$

where $\omega = e^{-i\frac{2\pi}{n}} = -i$. The two-mask structured illumination model measures the discrete Fourier transform of the signal and of the signal minus the first component (or any other desired element). A larger dimensional example of this idea is given in Figure 1.

Another structured example of A is the short-time Fourier transform, which can be thought of as a special case of the structured illumination model: The diagonal entries of each D_j come from a different translation of a common window function of width w, and we only consider every lth translation of this window. This particular measurement model finds applications in cross-correlation frequency-resolved optical gating (XFROG), in which one measures ultrafast laser pulses by optically producing a spectrogram; another application is ptychography, a diffractive imaging method where different overlapping patches of the unknown object are measured. For the short-time Fourier transform model, we know that x_0 is not uniquely determined if it has w consecutive zeros, but (1) does uniquely determine most nonvanishing signals when $l \ll w \ll n$.

Example 3

Consider the short-time Fourier transform of a signal of length $n = 6$. We choose the measurement window as a rectangular function of length $w = 3$

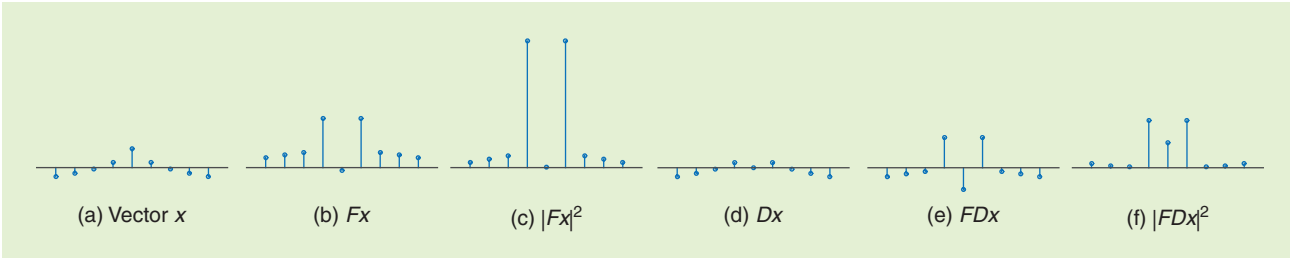

FIGURE 1. An illustration of two deterministic masks. (a) A vector x is chosen with indices from –4 to 4. We take the Fourier transform of x to get (b) and then square it to get (c). Applying the mask with a zero in the zero-indexed component, we obtain (d) a new vector Dx. Note that the Fourier transform (e) and squared Fourier transform (f) of Dx are quite different than those of x. With the data $|Fx|^2$ and $|FDx|^2$ it is possible to recover almost every vector x. Clearly, if x already has a zero in the zero-indexed component; then D will apply no change to x, and it will not be possible to recover x.

and select a step size $l = 2$. *The measurements are then the discrete Fourier transforms of vectors that are equal to the signal of interest in the given windows, and 0 elsewhere. With this choice of parameters, the short-time Fourier transform is equivalent to structured illumination using every other translation of* $\mathrm{diag}(1,1,1,0,0,0)$. *The resulting measurement matrix has the form*

$$A = \begin{bmatrix} FD_1 \\ FD_2 \\ FD_3 \end{bmatrix} = \begin{bmatrix} F\mathrm{diag}(1,1,1,0,0,0) \\ F\mathrm{diag}(0,0,1,1,1,0) \\ F\mathrm{diag}(1,0,0,0,1,1) \end{bmatrix}.$$

In the case where S is the set of k-sparse signals, a simple argument gives that for almost every $A \in \mathbb{C}^{m\times n}$, Ax_0 uniquely determines every k-sparse x_0 provided $m \geq 2k$ [4]. When the phases are discarded, the number of measurements increases by a factor of four: It can be shown that for almost every $A \in \mathbb{C}^{m\times n}$ with $m \geq 8k-4$, (1) uniquely determines every k-sparse x_0. When k is much smaller than n, one may wonder whether the classical phase retrieval problem $|Fx_0|^2 \mapsto [x_0]$ is plausible (we will certainly have $m = n \geq 8k-4$ intensity measurements, as is typically sufficient). Unfortunately, trivial ambiguities are still present in this case, but we can nonetheless uniquely determine most k-sparse signals with $k = O(n^{1/2-\epsilon})$ up to trivial ambiguities.

Algorithms

In this section, we describe several methods for solving the phase retrieval problem (1) in the special cases where S is either all of $\mathbb{C}^n$ or the set of k-sparse signals.

The most popular class of phase-retrieval algorithms is based on alternating projections, pioneered by the work of Gerchberg and Saxton and extended by Fienup. These methods consist of iteratively imposing the constraints in time/space and in the Fourier domain, namely, consistency with the measurements. Adapting to our generalized phase-retrieval setup, the basic steps consist of choosing an initial guess, and then alternating between projecting onto the sets $\{x: |Ax|^2 = y\}$ and S:

$$z_n := y^{1/2} \circ \frac{Ax_n}{|Ax_n|}, \quad x_{n+1} := P_S(A^\dagger z_n),$$

where $y^{1/2}$ denotes the entrywise square root of y, $\circ$ denotes entrywise product, P_S denotes the nearest-point projection onto S, and $A^\dagger$ denotes the pseudoinverse of A. Unfortunately, convergence to the true solution is not guaranteed since the sets are not convex. In what follows, we review a few successful alternatives to this approach.

PhaseLift

As before, we let a_i^* denote the ith row of A, and take $\mathcal{A}$ to be the function that maps Hermitian matrices X to vectors $y \in \mathbb{R}^m$ such that $y[i] = a_i^* X a_i$. It is easy to verify that $\mathcal{A}$ is linear, and considering (2), we also have $\mathcal{A}(xx^*) = |Ax|^2$ for every $x \in \mathbb{C}^n$. As such, the following program is equivalent to (1) when $S = \mathbb{C}^n$:

$$\text{find } X \text{ subject to } \mathcal{A}(X) = y, \quad X \succeq 0, \ \mathrm{rank}(X) = 1. \tag{3}$$

Discarding the rank constraint produces a convex relaxation of the phase retrieval problem:

$$\text{find } X \text{ subject to } \mathcal{A}(X) = y, \ X \succeq 0. \tag{4}$$

If this relaxation is tight (i.e., every solution satisfies $\mathrm{rank}(X) = 1$), then the relaxation solves the phase retrieval problem. If the relaxation is not tight, then we might instead minimize the trace of X subject to $\mathcal{A}(X) = y$ and $X \succeq 0$ so as to encourage X to have low rank. Both options are members of a family of convex relaxations of (3) called PhaseLift. An alternative relaxation, referred to as PhaseCut, is obtained by separating the measurements into an amplitude and phase component, and optimizing only the phase [12].

Amazingly, PhaseLift (4) is typically tight and robust to noise whenever A is "sufficiently random" and the number of measurements m is appropriately large. To see this, we first note that the set of complex Hermitian matrices is an n^2-dimensional vector space over the real numbers, by using a basis consisting of $n(n+1)/2$ real Hermitian matrices $E_{jk} + E_{kj}$ and $n(n-1)/2$ imaginary Hermitian matrices $i(E_{jk} - E_{kj})$. Thus, every Hermitian X is uniquely determined by $\mathcal{A}(X)$, if $\{a_i a_i^*\}_{i=1}^m$ is a spanning set; this typically occurs when $m \geq n^2$ and the a_is are drawn at random. To get away with fewer measurements, we leverage the fact that the solution we seek satisfies $X \succeq 0$. Then, it can be shown that (4) is typically tight when $m = \Omega(n \,\mathrm{polylog}\, n)$, provided $A \in \mathbb{C}^{m\times n}$ is drawn from an appropriate random distribution. For example, the entries of A may be drawn independently

from a complex Gaussian distribution, or A may be composed of $n \times n$ blocks of the form FD_j, where F is the $n \times n$ discrete Fourier transform matrix and D_j is a diagonal matrix with random diagonal entries from an acceptable distribution (the latter case models the structured illumination application of phase retrieval).

Wirtinger flow

While PhaseLift allows solving the phase retrieval problem in polynomial time (say, with an interior-point method), such methods scale poorly with the problem size due to the lifting operation, leading one to seek alternative solvers. To this end, we consider a different program:

$$\text{minimize} \quad \sum_{i=1}^{m} \left(|a_i^* x|^2 - y_i\right)^2. \tag{5}$$

Observe that (5) is equivalent to the phase retrieval problem (1) when $S = \mathbb{C}^n$. Unfortunately, since (5) is not convex, we expect to encounter local minima when attempting to solve it. In addition, this particular objective function has a continuum of global optimizers: The true solution x_0 induces a circle of global optimizers $[x_0] = \{e^{i\phi} x_0 : \phi \in [0, 2\pi)\}$. Perhaps surprisingly, (5) admits a fast initialization of gradient descent that allows for convergence to this circle provided A is sufficiently random. This gradient descent iteration is called *Wirtinger flow* because the gradient is conveniently expressed in terms of Wirtinger derivatives [3].

The convergence of Wirtinger flow is established by first showing that initializations sufficiently close to $[x_0]$ yield convergence by verifying a local convexity-type property. Next, a good initialization is found. Suppose the rows $\{a_i^*\}_{i=1}^m$ of A are complex Gaussian. Then a simple moment calculation reveals that

$$\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} y_i a_i a_i^*\right] = I + 2x_0 x_0^*,$$

meaning the true solution x_0 is a leading eigenvector of the expected matrix. Furthermore, $1/m \sum_{i=1}^m y_i a_i a_i^*$ is typically spectrally close to its expectation, and so its leading eigenvector (suitably scaled) is close to $[x_0]$. With this initialization, gradient descent converges linearly to $[x_0]$ when $m = \Omega(n \log n)$ and A has complex Gaussian entries; a variant of the gradient descent iteration exhibits similar performance when $m = \Omega(n \log^4 n)$ and A is composed of $n \times n$ blocks of the form FD_j (again, following the structured illumination model). Figure 2 shows a comparison between the computation time required for PhaseLift and for Wirtinger flow.

Sparse phase retrieval

Now suppose that the unknown vector x_0 is known to be k-sparse. If we were given Ax_0 instead of $|Ax_0|^2$, then we could leverage the now-rich theory of compressed sensing to reconstruct x_0 provided A is an $m \times n$ random matrix with $m = \Omega(k \text{ polylog } n)$; see [1] for a short introduction to this theory. We aspire to reconstruct $[x_0]$ from $|Ax_0|^2$ with similar requirements on A. We discuss two algorithms along these lines but note that their performance is strictly worse than the desired $m = \Omega(k \text{ polylog } n)$. Indeed, achieving this performance with complex Gaussian or sufficiently random matrices A remains an open problem.

In compressed sensing, one of the most popular reconstruction algorithms given $Ax_0 = y$ minimizes $\|x\|_1$ subject to $Ax = y$ [4]. In phase retrieval, we receive linear measurements of $x_0 x_0^*$, and since $x_0 x_0^*$ is sparse, it makes sense to minimize $\|X\|_1$ subject to $\mathcal{A}(X) = y$. We also want to encourage X to be rank-1, leading to the following variation of PhaseLift:

$$\begin{aligned} &\text{minimize} \quad \|X\|_1 + \lambda \text{Tr}[X] \\ &\text{subject to} \quad \mathcal{A}(X) = y, X \succeq 0. \end{aligned} \tag{6}$$

If the entries of A are complex Gaussian, then for an appropriate choice of λ, (6) typically recovers $X = x_0 x_0^*$ provided $m = \Omega(k^2 \log n)$. Drawing from compressed sensing-based intuition, the k^2 here comes from the fact that $x_0 x_0^*$ is k^2-sparse, and so the barrier to improving this sample complexity is perhaps an artifact of the lifting approach.

Following the motivation of Wirtinger flow, we seek a faster alternative to this semidefinite program. Let us reformulate (1) in the case where S is the set of k-sparse vectors:

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} \left(|a_i^* x|^2 - y_i\right)^2 \\ &\text{subject to} \quad \|x\|_0 \le k. \end{aligned} \tag{7}$$

Here, $\|x\|_0$ denotes the number of nonzero entries in x; note the similarity to (5). When $\{a_i\}_{i=1}^m$ are complex Gaussian, the solution to (7) typically

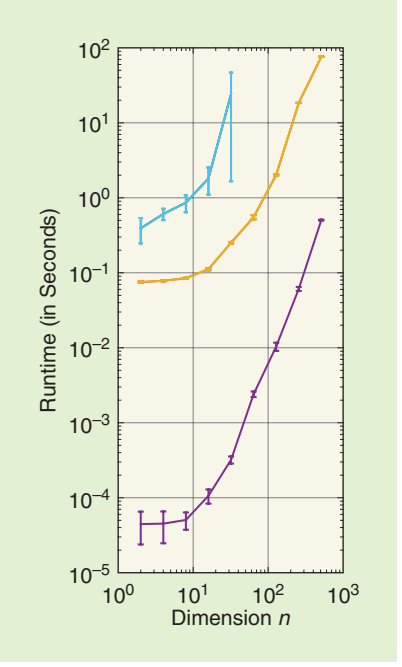

FIGURE 2. The runtime comparison between PhaseLift (blue), Wirtinger flow (orange), and least squares (purple) assuming the known phase. For each dimension $n \in \{2^1, \ldots, 2^9\}$, we perform 20 iterations of the following experiment: Draw a 4.5 $n \times n$ matrix A with independent entries with complex Gaussian distribution $\mathcal{N}(0, 1/2) + i\mathcal{N}(0, 1/2)$ and a signal $x \in \mathbb{C}^n$, also with independent and identically distributed complex Gaussian entries. Compute $z = Ax$ and $y = |z|^2$. Then reconstruct x from z by MATLAB's built-in implementation of least squares, and reconstruct x from y up to global phase using Wirtinger flow and PhaseLift. (For PhaseLift, we solve the semidefinite program using TFOCS v1.3 release 2 [2]; the runtime was prohibitively long for $n > 2^5$.) After conducting all 20 iterations, we plot the average runtime along with error bars that illustrate one standard deviation. As expected, the least-squares solver (which enjoys a phase "oracle") is faster than the phase retrieval solvers. For larger dimensions, Wirtinger flow appears to be about 100 times slower than least-squares, whereas PhaseLift is even slower.

reconstructs the true solution x_0 provided $m = \Omega(k \log(n/k) \log k)$, and furthermore, the reconstruction is robust to noise. To solve this program, we may apply an algorithm called Greedy Sparse Phase Retrieval (GESPAR), which iteratively improves a guess $K \subseteq \{1,\ldots,n\}$ of the support of x_0. For each guess, the objective function reduces to an instance of (5). Therefore, GESPAR optimizes locally to produce an estimate x such that $\text{supp}(x) \subseteq K$, and then updates K by swapping the member of K that contributes least to x with the index outside of K that contributes most to the (negative) gradient of the objective function. This iteration terminates when the swap fails to produce an improvement. See [9] for an implementation of GESPAR. In practice, GESPAR is prone to local minima, and so one must attempt multiple trials with different initializations before succeeding. Still, GESPAR is much faster than the semidefinite programming alternative, and it empirically performs well when $m = \Omega(k^3)$; for comparison, the Fienup-type alternative is about twice as fast, but only performs well when m is much larger. Recently, GESPAR has been used to solve phase retrieval problems in coherent diffraction imaging and in ankylography.

Understanding conditions under which recovery from Fourier phase retrieval measurements is possible, as well as developing practical measurement systems from which the signal can be recovered efficiently are some of the important directions for future research.

Conclusions

There have been several interesting developments in phase retrieval over the past decade from both the optics and signal processing communities. We discussed various settings in which signals are uniquely determined by intensity measurements, as well as new algorithms for reconstructing signals from such measurements. Understanding conditions under which recovery from Fourier phase retrieval measurements is possible, as well as developing practical measurement systems from which the signal can be recovered efficiently are some of the important directions for future research. We believe that this field will continue to grow and have a significant impact on optical imaging.

Acknowledgments

The work of Yonina C. Eldar was funded by the European Unions Horizon 2020 research and innovation programme under grant agreement ERC-BNYQ, by the Israel Science Foundation under grant 335/14, and by ICore: the Israeli Excellence Center Circle of Light. The work of Nathaniel Hammen and Dustin G. Mixon was funded by an Air Force Office of Scientific Research (AFOSR) Young Investigator Research Program award, National Science Foundation grant DMS-1321779, and AFOSR grant F4FGA05076J002. Nathaniel Hammen was also supported in part by an appointment to the Postgraduate Research Participation Program at the U.S. Air Force Institute of Technology, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Air Force Institute of Technology. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

Authors

Yonina C. Eldar (yonina@ee.technion.ac.il) is a professor in the Department of Electrical Engineering at the Technion–Israel Institute of Technology, Haifa, where she holds the Edwards Chair in Engineering. Her research interests are in the areas of signal processing and optimization and their applications to communications, radar, biomedical imaging, and optics. She is an IEEE Fellow and a member of the Young Israel Academy of Science and the Israel Committee for Higher Education.

Nathaniel Hammen (nhammen@gmail.com) is a postdoctoral researcher in the Department of Mathematics and Statistics at the Air Force Institute of Technology, Wright-Patterson, Air Force Base, Ohio. His research interests include applied harmonic analysis and signal processing.

Dustin G. Mixon (dustin.mixon@afit.edu) is an assistant professor of mathematics in the Department of Mathematics and Statistics at the Air Force Institute of Technology, Wright-Patterson, Air Force Base, Ohio. His research interests include applied harmonic analysis, mathematical signal processing, and mathematical data science.

References

[1] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[2] S. Becker, E. J. Candès, and M. Grant. (Sept. 2011). Templates for convex cone problems with applications to sparse signal recovery. *Math. Program. Comput.* [Online]. *3*(*3*), 165–218. Available: http://cvxr.com/tfocs/

[3] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1985–2007, Feb. 2015.

[4] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, UK: Cambridge Univ. Press, 2015.

[5] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 30, no. 2, pp. 140–154, Apr. 1982.

[6] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," submitted for publication.

[7] R. Juri, A. Chebira, Y. M. Lu, and M. Vetterli. (2013). Phase retrieval for sparse signals: Uniqueness conditions. [Online]. Available: arXiv: 1308.3058

[8] D. G. Mixon. (2014, Mar. 7). Phase transitions in phase retrieval. [Online]. Available: arXiv: 1403.1458

[9] Y. Shechtman, A. Beck, and Y. C. Eldar, GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Processing*, vol. 62, no. 4, pp. 928–938, Jan. 2014.

[10] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," *IEEE Signal Processing Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[11] C. Vinzant, "A small frame and a certificate of its injectivity," in *Proc. Int. Conf. SampTA*, Washington, DC, 2015, pp. 197–200.

[12] 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.

SP