Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing

Yoav Shechtman,[1] Yonina C. Eldar,[2] Alexander Szameit,[1] and Mordechai Segev[1,*]

[1]*Physics Department and Solid State Institute, Technion, Haifa 3200, Israel*
[2]*Department of Electrical Engineering, Technion, Haifa 3200, Israel*
*msegev@tx.technion.ac.il

Abstract: We demonstrate that sub-wavelength optical images borne on partially-spatially-incoherent light can be recovered, from their far-field or from the blurred image, given the prior knowledge that the image is sparse, and only that. The reconstruction method relies on the recently demonstrated sparsity-based sub-wavelength imaging. However, for partially-spatially-incoherent light, the relation between the measurements and the image is quadratic, yielding non-convex measurement equations that do not conform to previously used techniques. Consequently, we demonstrate new algorithmic methodology, referred to as quadratic compressed sensing, which can be applied to a range of other problems involving information recovery from partial correlation measurements, including when the correlation function has local dependencies. Specifically for microscopy, this method can be readily extended to white light microscopes with the additional knowledge of the light source spectrum.

©2011 Optical Society of America

OCIS codes: (180.0180) Microscopy; (110.4980) Partial coherence in imaging; (100.3010) Image reconstruction techniques.

References and links

1. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
2. E. A. Ash and G. Nicholls, "Super-resolution aperture scanning microscope," Nature **237**(5357), 510–512 (1972).
3. A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, "Development of a 500 A spatial resolution light microscope: I. light is efficiently transmitted through λ⁄16 diameter apertures," Ultramicroscopy **13**(3), 227–231 (1984).
4. E. Betzig, J. K. Trautman, T. D. Harris, J. S. Weiner, and R. L. Kostelak, "Breaking the diffraction barrier: optical microscopy on a nanometric scale," Science **251**(5000), 1468–1470 (1991).
5. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Opt. Lett. **19**(11), 780–782 (1994).
6. Z. Jacob, L. V. Alekseyev, and E. Narimanov, "Optical hyperlens: far-field imaging beyond the diffraction limit," Opt. Express **14**(18), 8247–8256 (2006).
7. Z. Liu, H. Lee, Y. Xiong, C. Sun, and X. Zhang, "Far-field optical hyperlens magnifying sub-diffraction-limited objects," Science **315**(5819), 1686 (2007).
8. I. I. Smolyaninov, Y. J. Hung, and C. C. Davis, "Magnifying superlens in the visible frequency range," Science **315**(5819), 1699–1701 (2007).
9. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary optical transmission through subwavelength hole arrays," Nature **391**(6668), 667–669 (1998).
10. F. M. Huang and N. I. Zheludev, "Super-resolution without evanescent waves," Nano Lett. **9**(3), 1249–1254 (2009).
11. S. Gazit, A. Szameit, Y. C. Eldar, and M. Segev, "Super-resolution and reconstruction of sparse sub-wavelength images," Opt. Express **17**(26), 23920–23946 (2009).
12. E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).
13. A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," SIAM Rev. **51**(1), 34–81 (2009).
14. A. Szameit, Y. Shechtman, H. Dana, S. Steiner, S. Gazit, T. Cohen-Hyams, E. Bullkich, O. Cohen, Y. C. Eldar, S. Shoham, E. B. Kley, and M. Segev, "Far-field microscopy of sparse subwavelength objects," arXiv:1010.0631 [physics.optics] (2010).

15. Y. Shechtman, S. Gazit, A. Szameit, Y. C. Eldar, and M. Segev, "Super-resolution and reconstruction of sparse images carried by incoherent light," Opt. Lett. **35**(8), 1148–1150 (2010).
16. J. W. Goodman, *Statistical Optics* (Wiley, 1985).
17. M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in Proceedings of Am. Control Conf. (2003).
18. E. J. Candes (Departments of Mathematics and of Statistics, Stanford University, Stanford CA 94305), Y. C. Eldar, T. Strohmer, and V. Voroninski are preparing a manuscript to be called "Phase retrieval via matrix completion."
19. D. Gross, Y. K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," Phys. Rev. Lett. **105**(15), 150401 (2010).
20. A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, "Efficient measurement of quantum dynamics via compressive sensing," Phys. Rev. Lett. **106**(10), 100401 (2011).

Introduction

The resolution of an optical imaging system has traditionally been bounded by Abbe's limit – setting the smallest resolvable feature in an imaged object to be about the size of $\lambda/2$, with λ being the optical wavelength of the light illuminating the object [1]. Indeed, when imaging an object containing sub-wavelength features, these features become blurred and unresolvable. The reason for this blurring is that only waves containing spatial frequencies lower than $1/\lambda$ propagate through the medium, while higher spatial frequencies become evanescent and decay exponentially in a homogeneous medium. Such waves are greatly attenuated over a propagation distance larger than several λ, hence the nature of Maxwell's equation in homogeneous media acts practically as an ideal low-pass filter [1]. Over the years, there have been many attempts to recover information with a resolution exceeding the diffraction limit. Some of these techniques have become standard commercial tools, e.g. the scanning near-field optical microscope [2–4], and methods based on light-emitting molecules [5]. Other approaches offer great promise but their technology still requires major improvements, e.g., devices based on negative-index materials [6–8], or scanning with nanoholes [9] or with a sub-wavelength hot-spot [10]. However, one common feature limits almost all of these "hardware solutions" for sub-wavelength imaging: they are not single-exposure experiments, hence they cannot work in real time, because they either require scanning [2–4,9,10], or multiple experiments [5]. Needless to say, it would be highly desirable to have sub-wavelength imaging technology operating at ultrafast scales, which is the time scale of many important chemical reactions.

During the past decades, there have been attempts to recover sub-wavelength information through bandwidth extrapolation algorithms. These techniques utilize the fact that the 2D Fourier transform of a spatially bounded function (the electromagnetic field) is an analytic function, hence measuring the far-field at very high precision could, in principle, allow recovery of the information contained in the evanescent waves, which has never reached the far field. However, these attempts were not successful, as they are extremely sensitive to noise in the measured data and to the assumptions made on the prior knowledge about the sub-wavelength information (see discussion in [1], page 165).

Recently, we have demonstrated a method to overcome the diffraction limit [11], by using prior knowledge that the imaged object is sparse in a known basis (say, in the near-field). The methodology is based on bandwidth extrapolation, where the extrapolating function is chosen such that it yields the sparsest solution (image) that is consistent with the measurements. Mathematically, the problem was posed in matrix representation, and solved using efficient tools borrowed from the realm of signal processing, specifically – methods for finding sparse solutions to underdetermined systems of equations [12,13]. Experimentally, using coherent laser light, we were recently able to recover sub-wavelength features as small as 100nm using 532nm wavelength [14]. However, it is highly desirable to develop such techniques also for the incoherent case – simply because most microscopes work with incoherent light. Consequently, we have demonstrated sparsity-based super-resolution with completely spatially-incoherent quasi-monochromatic light [15]. However, taking the technique of [15] into the sub-wavelength regime (as we did for coherent light [14]) has limited applicability,

because it uses the Optical Transfer Function (OTF) which assumes that the correlation distance is much smaller than any feature in the optical object. The idea would work when light emitting molecules are distributed on the object itself, but would not work in microscopes using incoherent light. This is because microscopes use lenses (or mirrors) to collect the incoherent illumination; hence the correlation distance at the object plane cannot be smaller than the diffraction limit of the collecting lens. Consequently, when using incoherent light to illuminate sub-wavelength optical information, the optical features can be smaller, comparable, or larger than the correlation distance of the light in the object plane, implying that the OTF cannot represent the propagation of the light adequately. In such cases, which are most common in microscopy, the light incident upon sub-wavelength optical objects must be treated as partially-spatially-incoherent.

Here, we demonstrate theoretically the reconstruction of objects containing sub-wavelength features, by using the prior knowledge that the objects are sparse in real space, i.e. containing only a few non-zero pixels. Unlike the two extreme cases of coherence that we previously dealt with [11,15], the partially-incoherent case cannot be represented by a transfer function, which translates into linear measurements in the electric-field [11] or in the intensity [15], for fully coherent and for completely incoherent light, respectively. Instead, data resulting from the partially-incoherent case is represented by a more complicated relation, depending on the mutual intensity [16]. As such, the signal processing theory and methods, which are well established for sparse solutions to a set of linear equations, cannot be directly applied here. We therefore develop the underlying theory for the reconstruction of sub-wavelength images borne on partially-spatially-incoherent light, and show that the resulting problem calls for finding a sparse solution to a set of quadratic equations. We consequently propose a heuristic method to find a sparse solution to this set of quadratic equations, a problem which we name Quadratic Compressed Sensing, and demonstrate (via simulations) the reconstruction of sparse sub-wavelength objects illuminated by partially incoherent light using this approach.

Our method can be extended to work under broad-band illumination ("white light"), provided that the spectral composition of the light is known. Hence, the techniques demonstrated here can be generalized to work with white light microscopes, which are the most commonly used microscopes. Finally, this concept of reconstructing information from a small subset of measurements of its correlation function is universal, having a broad variety of applications in optics and beyond. Examples range from Fourier transform spectroscopy (FTIR), Frequency Resolved Optical Gating (FROG) and other methods relying on interference, to equivalent examples with any other kind of signals. Perhaps most intriguingly, the concept established here could pave the way to the recovery of quantum states from a small subset of measurements of their correlation function, given only that the original quantum information is sparse in a known basis.

Propagation of images borne on partially-spatially-coherent light

An optical imaging system is typically designed to image the intensity structure of an EM field (henceforth the "object") placed at some input plane, to some other place ("output plane") where the output intensity structure ("image") is measured with a camera (or displayed on a screen). Here, we are interested in the case where the object contains sub-wavelength features. In the extreme case where the transverse correlation distance (sometimes called "transverse coherence length") of the light is much smaller than the smallest feature of the object, the light can be assumed to be completely spatially incoherent, and therefore the imaging process can be described by the simple OTF for the intensity [1]. In that case, one can recover the information carried by the spatial frequencies that are smaller than the cut-off frequency by simply dividing the Fourier transform of the image by the OTF. This is a standard deconvolution procedure, which works perfectly well when the OTF is nonzero. Similarly, for the fully-coherent case, i.e. when the transverse coherence length of the light is much larger than the largest features in the object, the imaging process is described by the amplitude transfer function (ATF) of the system, sometimes also called the Coherent Transfer

Function [1]. One can then recover the field amplitude of the object – up to the cut-off frequency - by dividing the Fourier transform of the image-amplitude by the ATF. In both the coherent and incoherent cases, all information about the frequencies higher than the cut-off is lost.

Recently, we suggested and demonstrated a method to recover the information carried by high frequencies that were cut-off by the imaging process – for both the coherent and incoherent cases [11,14,15], by exploiting the object's sparsity in the near field.

Here, we suggest a way to recover sub-wavelength information carried by light that is neither completely coherent nor completely incoherent – but with some partial spatial coherence. We utilize prior knowledge that the information (the object intensity) is sparse in real space, and only that. As we shall show, the case of partially-incoherent illumination leads to an interesting mathematical problem that is fundamentally different than the extreme coherence cases – and requires a novel method for recovering the information.
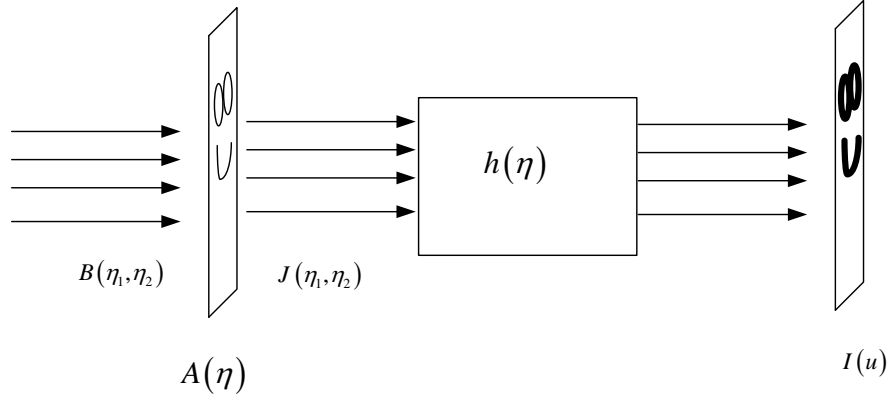

Fig. 1. Setting for imaging using partially spatially coherent light.

Consider the setting sketched in Fig. 1. A thin optical object with complex transmittance function A, placed at plane $z = 0$, is illuminated by quasi-monochromatic partially-spatially-coherent light. For simplicity, we analyze here one dimensional objects, hence A is a function of η, the transverse coordinate at the input plane; however, the extension to two transverse dimensions is straightforward. The object is imaged using an optical system with a coherent impulse response $h(u;\eta)$, u being the transverse coordinate at the image plane. Note that h is a function of the coordinates of both the input and the output planes. In other words, an object consisting of a single narrow peak $A(\eta) = \delta(\eta - \eta_0)$ would yield the image $A_{image}(u) = h(u;\eta_0)$, had it been illuminated by perfectly spatially-coherent illumination. We measure the light intensity distribution $I(u) = \left\langle \left| A_{image}(u,t) \right|^2 \right\rangle$ at the image plane, where $\langle \cdot \rangle$ denotes the time-average, and the averaging is carried out over the response time of the camera (or the eye), which is assumed to be much longer than the characteristic fluctuation time of the incoherent light. From the measurements of $I(u)$, we wish to retrieve the original intensity distribution of the object, and, if possible, also its phase.

For most physical systems, the transmittance function $A(\eta)$ defining the object does not vary with time. The light incident upon the object is quasi-monochromatic and partially-spatially-incoherent, that is, its complex electric field $u(\eta,t)\big|_{z=0^-}$ is fluctuating randomly with some characteristic time that is much longer than one cycle of the optical frequency. The light at the input plane ($z = 0$) is neither fully spatially coherent nor completely spatially

incoherent. Hence, the field amplitudes at two different points η_1 and η_2 in plane $z = 0$ exhibit some correlation, described by the mutual intensity function $B(\eta_1,\eta_2) := \langle u(\eta_1,t)u^*(\eta_2,t)\rangle\big|_{z=0^-}$. Note that the ability to recover the phase, even when the impulse response is hypotetically a delta function, is limited due to the incomplete spatial coherence of the light. Namely, the partially incoherent light is characterized by a transverse correlation distance L_c between any two points in the transverse plane; for two points separated by $\eta < L_c$ the relative phase is well defined, whereas any two points separated by $\eta > L_c$ are not phase-correlated. Hence, phase information that varies over distances much shorter than L_c is lost, and cannot be recovered.

As explained above, the light travels through the optical system, until reaching the image plane at which the intensity measurements are conducted. The instantaneous electric field of the light travels as a coherent field, and is therefore described by the coherent impulse response of the system, $h(u;\eta)$. When the impulse response of the optical system is space-invariant, with proper coordinate scaling we can write $h(u;\eta) = h(u-\eta)$, so that the intensity of light reaching the image plane can be written as (see page 313 in [16]):

$$I(u) = \iint h(u-\eta_1)h^*(u-\eta_2)J(\eta_1,\eta_2)d\eta_1 d\eta_2 \qquad (1)$$

where $J(\eta_1,\eta_2)$ is the mutual intensity function immediately after the optical object ($z = 0^+$). Since the optical object $A(\eta)$ does not vary in time, one can write:

$$\begin{aligned}J(\eta_1,\eta_2) &= \langle A(\eta_1,t)u(\eta_1,t)A^*(\eta_2,t)u^*(\eta_2,t)\rangle = \\ &= A(\eta_1)A^*(\eta_2)\langle u(\eta_1,t)u^*(\eta_2,t)\rangle = A(\eta_1)A^*(\eta_2)B(\eta_1,\eta_2).\end{aligned} \qquad (2)$$

For most sources emitting partially-spatially-incoherent light (e.g., the sun, thermal sources, a laser beam passing through a rotating diffuser), $B(\eta_1,\eta_2)$ is a function of the coordinate difference $|\eta_1 - \eta_2|$ only. Under the above circumstances, Eq. (1) can be rewritten as:

$$I(u) = \iint h(u-\eta_1)h^*(u-\eta_2)A(\eta_1)A^*(\eta_2)B(|\eta_1-\eta_2|)d\eta_1 d\eta_2. \qquad (3)$$

An example illustrating the effect of the mutual intensity function is depicted in Fig. 2, showing a 1D object comprising of two peaks spaced by 0.6λ (Figs. 2(a) and 2(b)) and by λ (Figs. 2(c) and 2(d)), respectively. The impulse response function $h(u)$ of the optical system is taken as $h(x) = \frac{1}{x}h_0 \sin\left(2\pi x / \lambda\right)$, corresponding to an ideal low-pass-filter with a spatial cutoff frequency of $1/\lambda$. The peaks have either the same phase (Figs. 2(a) and 2(c)) or opposite phases (Figs. 2(b) and 2(d)). Figure 2 shows the intensity at the image plane, $I(u)$, for three cases: fully coherent (dotted line), completely incoherent light (dashed line), and partially-incoherent light (dash-dotted) with the mutual intensity function being a Gaussian with 0.55λ Full Width Half Maximum (FWHM), i.e., transverse correlation distance is $L_c = 0.55\lambda$.

Three key points should be emphasized in Fig. 2. First, the image is a smeared (blurred) version of the object, regardless of the coherence of the light, demonstrating the low-pass nature of the system. Second, the effect of the coherence of the light on the image is much more pronounced when the peaks are close together, specifically having sub-wavelength spacing. Third, when the optical information varies on a sub-wavelength scale (Figs. 2(a) and 2(b)), the partially-incoherent image (dash-dotted) is considerably different than both the coherent and incoherent images. This implies that using either the fully coherent or the fully incoherent approximations to recover the object from its measured image would not yield

good recovering. Hence, in physical settings where the optical information varies on a sub-wavelength scale, it is essential to employ a method dealing specifically with partially-spatially-coherent light.
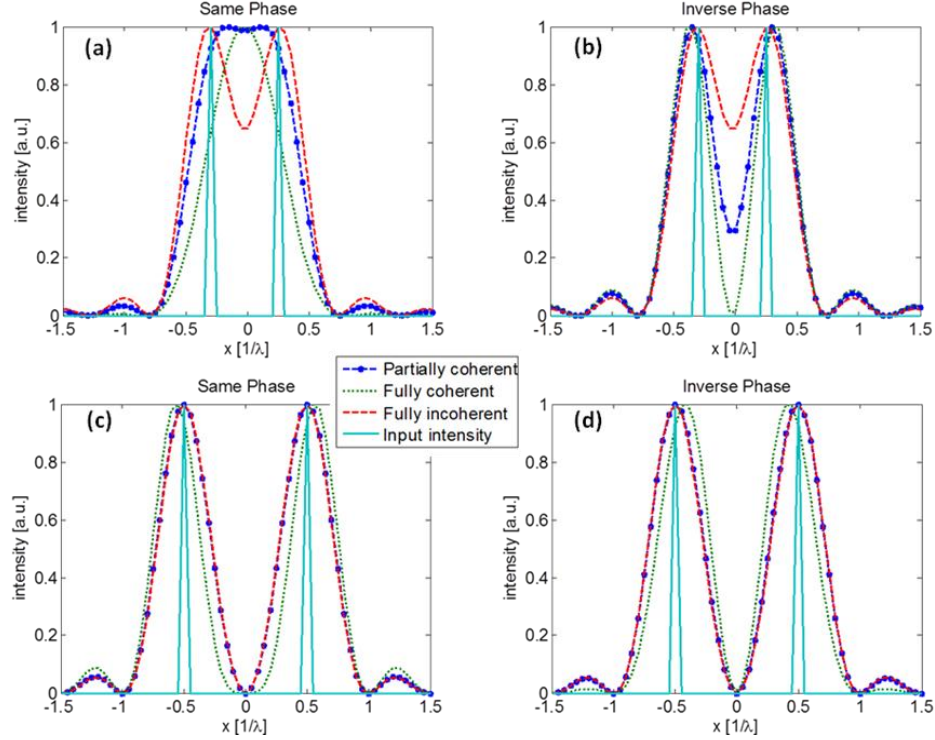

Fig. 2. Illustration of the effect of mutual-intensity width on output intensity, for 2 equal phase (a) or inverse phase (b) spikes, separated by 0.55λ (same/inverse phase). In (c) and (d) the spikes are 1λ apart. Coherent light corresponds to a mutual-intensity width of infinity; incoherent light corresponds to mutual-intensity width of 0. The partially coherent mutual-intensity width here is 0.55λ FWHM. The resolution is better for equal phase spikes when using incoherent light (a), and for inverse phase spikes – using coherent light (b). Note that the effect of the coherence of the light is much stronger for closely (sub-λ) separated peaks, and that the partially-incoherent image in that case (dash-dotted in (a) and (b)) is considerably different than both the coherent and incoherent images.

Matrix representation

Our goal is to solve Eq. (3). Namely, to find the transmittance function $A(\eta)$ given $h(\eta)$, $B(|\eta_1 - \eta_2|)$ and the measured intensity $I(u)$. In order to use our sparsity-based reconstruction algorithm, we first discretize the problem and rewrite Eq. (3) in matrix form. This yields

$$I(u) = \sum_i \sum_j h_u(\eta_i) h_u^*(\eta_j) B(\eta_i - \eta_j) A(\eta_i) A^*(\eta_j) \tag{4}$$

using $h_u(\eta) := h(u - \eta) \cdot d\eta$, where $d\eta$, the discretization parameter, is selected to be smaller than the features in the image, with the purpose of making the discrete problem similar enough to the true analog problem from which it originated. Equation (4) can be written in a compact form as:

$$y_u = a^* M_u a \tag{5}$$

where we define $M_{u,ij} := h_u(\eta_i) h_u^*(\eta_j) B(\eta_i - \eta_j)$. The vector y_u represents the measured intensity, where the index u refers to the location of the uth measurement in the image plane. We denote the set of all measurements by U.

Equation (5) describes a quadratic relation between the object a and the uth measurement, related by a matrix M_u, for all $u \in U$. The size of y_u is determined by the number of measurements, and the size of M_u is determined by the discretization step.

Our goal is now to find a vector a that includes the optical information at the input plane, or to "invert" Eq. (5).

Lack of uniqueness and Sparsity prior

Unfortunately, under even small amounts of noise, there is no unique vector a that solves Eq. (5). In other words, there are many different vectors a that would yield the same measurements y_u (within the noise level), after being 'smeared' by the optical system. Our goal is to find the true one. One can notice immediately that uniqueness can never be guaranteed. For example, if $\tilde{a}$ is a solution to Eq. (5), then so is $-\tilde{a}$, or in fact any $\tilde{a} \cdot e^{i\varphi}$ for any phase φ. We therefore limit ourselves to reconstruction of $\tilde{a}$ up to a global phase constant. However, even when allowing a global phase ambiguity, there is still no uniqueness, under a small amount of noise. The reason for this is the low-pass nature of the system: It can be shown (based on [1], chapter 7.2.3), that the spectrum of the output image intensity is bounded in k-space by $2\nu_c$, where ν_c is the cutoff frequency of the coherent transfer function of the system, irrespective of the spectrum of the input image or of the coherence properties of the impinging light. This causes inherent loss of information, of the same sort that was dealt with previously in the two extreme cases of completely coherent and completely incoherent imaging [11,15].

The low-pass nature of the system can also be inferred by examining the form of the M_u matrices, as will be shown below. Note that the cut-off frequency of the image implies a theoretical upper bound for the sampling rate (or a lower bound for the distance between two samples), given by the sampling theorem, beyond which no new frequency information will be obtained. This is referred to as the Nyquist rate and is equal in our case to $4\nu_c$, which in an ideal diffraction-limited optical system corresponds to a spacing between samples of λ/4. We shall see later that, indeed, increasing the sampling rate above this rate does not improve the results significantly.

In order to overcome the lack of uniqueness - in the spirit of the fully coherent and fully incoherent extreme cases - we assume a sparse model of the input field (in real space), meaning that the input field comprises of only a few nonzero values in space. Sparsity-based algorithms are known to handle linear problems very well, that is, problems of the form $y = Ma$ where y is the measurement vector, a is the information we wish to recover, and M is the transformation relating the two. Here, our problem is not linear, as manifested by the matrix representation of Eq. (5). Hence, we cannot directly harness the theoretical knowledge about linear sparse problems to solve the current problem. Nevertheless, we can still use the prior knowledge about the sparsity of a to aid in the reconstruction process.

Optimization problem: formulation and solution method

We seek the sparsest input field a that is consistent with the measurements y. In addition, in a passive optical system with no gain, the overall output power cannot be larger than the overall input power. This translates mathematically to solving the following optimization problem:

$$(P) \min_{a} \|a\|_0$$

$$s.t. \quad \left| a^* M_u a - y_u \right|^2 \leq \varepsilon \quad u \in U \tag{6}$$

$$a^* a \geq y^* y$$

where $\|a\|_0$ denotes the 0 'norm' of a, or simply the number of nonzero elements in a, and U is the set of measurements taken at the image plane. The consistency parameter ε is determined by the typical amount of noise added by the system.

Problem (6) is not convex, due to both the objective function (which represents the sparsity requirements) and to the quadratic consistency and energy constraints. In general, when seeking a sparse solution that is consistent with a set of measurements, the sparsity requirement is never convex. When the measurement vector is related linearly to the signal vector (so-called "linear measurements" in signal processing), many methods have been developed to approximate the sparsest solution (for a review of these methods, see [13]). However, in our case the relation between the measurement and signal vectors is quadratic (Eq. (5)). To the best of our knowledge, there are no prior results addressing such a problem, hence we have to find a new method to approximate the sparsest solution under "quadratic measurements".

Our approach for solving Eq. (6) is based on expressing quadratic measurements as linear matrix equalities and relaxing the still combinatorial problem into semi-definite programming. There are many ways of relaxing a combinatorial problem. Here we have chosen a variation on the trace norm heuristic, namely, the log det heuristic proposed in [17], with modifications to incorporate the sparsity requirement. Expressing quadratic constraints as linear matrix equalities and relaxing the combinatorial problem via the trace norm or log det heuristic has also been proposed in [18] for phase retrieval from intensity measurements. Here, we have the additional requirement that a is sparse, which is not present in [18]. On the other hand [18], uses optical preprocessing (via a grating or structured illumination) that is impossible in our context of sub-wavelength imaging, which helps improve the recovery.

We solve the quadratic problem posed by Eq. (6) as follows. We define a matrix $X := aa^*$, such that the non-convex consistency constraint in Eq. (6) may be re-written as a convex constraint on the matrix X : $a^* M_u a = Trace(M_u X)$. The energy constraint becomes convex, since $a^* a = Trace(X)$. Note that we must also add the requirement that X can be written as $X = aa^*$ for some vector a – which is really what we are after. This requirement means that X is positive-semi-definite (PSD), and a rank 1 matrix. The rank 1 requirement is not convex, so instead of adding it as a constraint, we attempt to minimize the rank, subject to the remaining constraints, as shown below in Eq. (7).

Next, we deal with the sparsity constraint - sparsity of a implies row-sparsity of X, i.e. if a has k nonzero elements (k-sparse), then X has k rows that contain nonzero elements, and each of these rows is also k-sparse. The row-sparsity of X is promoted by adding a constraint on the upper bound of the mixed-1,2 norm, i.e. $\sum_a \left(\sum_b |X_{ab}|^2 \right)^{1/2} < \zeta$, with ζ being a parameter derived from the degree of assumed sparsity of the object. This requirement promotes a solution X that is sparse in rows, and furthermore each row in itself is a sparse vector, due to the fact that X is Hermitian.

We can now express our problem (P) as a minimization problem in the matrix X, which should be PSD, row-sparse and rank 1. This leads to the optimization problem:

$$\underset{X}{\text{argmin}} \ Rank\left(X\right)$$

$$s.t. \quad \sum_a \left(\sum_b |X_{ab}|^2\right)^{1/2} \leq \zeta$$

$$\left|Trace\left(M_u X\right) - y_u\right|^2 \leq \varepsilon \ \forall u \in U \tag{7}$$

$$Trace\left(X\right) \geq y^* y$$

$$X \geq 0$$

The constraints in Eq. (7) are all convex, however the objective is non-convex. The algorithm we develop to try to solve Eq. (7) is based on an iterative rank minimization heuristic method [17], that attempts to find a low-rank matrix consistent with a set of convex constraints, by minimizing the log-det (the logarithm of the determinant) of the matrix. In other words, the low-rank objective is replaced by the function $\log\det(X + \delta I)$ with δ being a small regularization parameter. This function is not convex, however the method suggested in [17] is shown to converge to a local minimum. The requested properties of our solution are therefore incorporated as convex constraints into the log-det algorithm.

Summarizing, the optimization problem we actually solve is the following:

$$\underset{X}{\text{argmin}} \ \log\det\left(X + \delta I\right)$$

$$s.t. \quad \sum_a \left(\sum_b |X_{ab}|^2\right)^{1/2} \leq \zeta$$

$$\left|Trace\left(M_u X\right) - y_u\right|^2 \leq \varepsilon \tag{8}$$

$$Trace\left(X\right) \geq tr\left(y \cdot y^T\right)$$

$$X \geq 0$$

The basic iteration of the log-det heuristic method is performed by solving the following convex optimization problem:

$$X_{k+1} = \underset{X \in C}{\text{argmin}} \ Trace\left(X_k + \delta I\right)^{-1} X \tag{9}$$

where C is the set of convex constraints written in Eq. (8). In addition, to further promote a sparse solution, at each iteration a thresholding step is performed, in which the smallest elements in the diagonal of X are located, and the intersecting rows and columns are taken out of the possible support in the next iteration. This is added as a convex support constraint to the next iteration. Each iteration can be solved using standard convex optimization tools.

Once a low-rank X is found (empirically, X is approximately rank 1 at the end of the process), aa^* is determined as the best rank-1 approximation of X, using the Singular Value Decomposition (SVD). If the rank minimization process succeeds, i.e. the solution matrix is close to rank-1, then its rank-1 approximation will also be row-sparse as required. From there, the approximation of a is obtained. Figure 3 describes the method in detail.

The following parameters are used in the algorithm:

- ε - Determined by the noise added by the optical system. This could, in principle, be calibrated using a known reference object.

- ζ - Determined by the assumed sparsity of the object.

- δ - A small regularization parameter – in practice the algorithm's performance is not very sensitive to it.

- T, ΔT - Thresholding strength, determined by the system's noise.

The actual values used in the simulations are given in the algorithm-evaluation section.
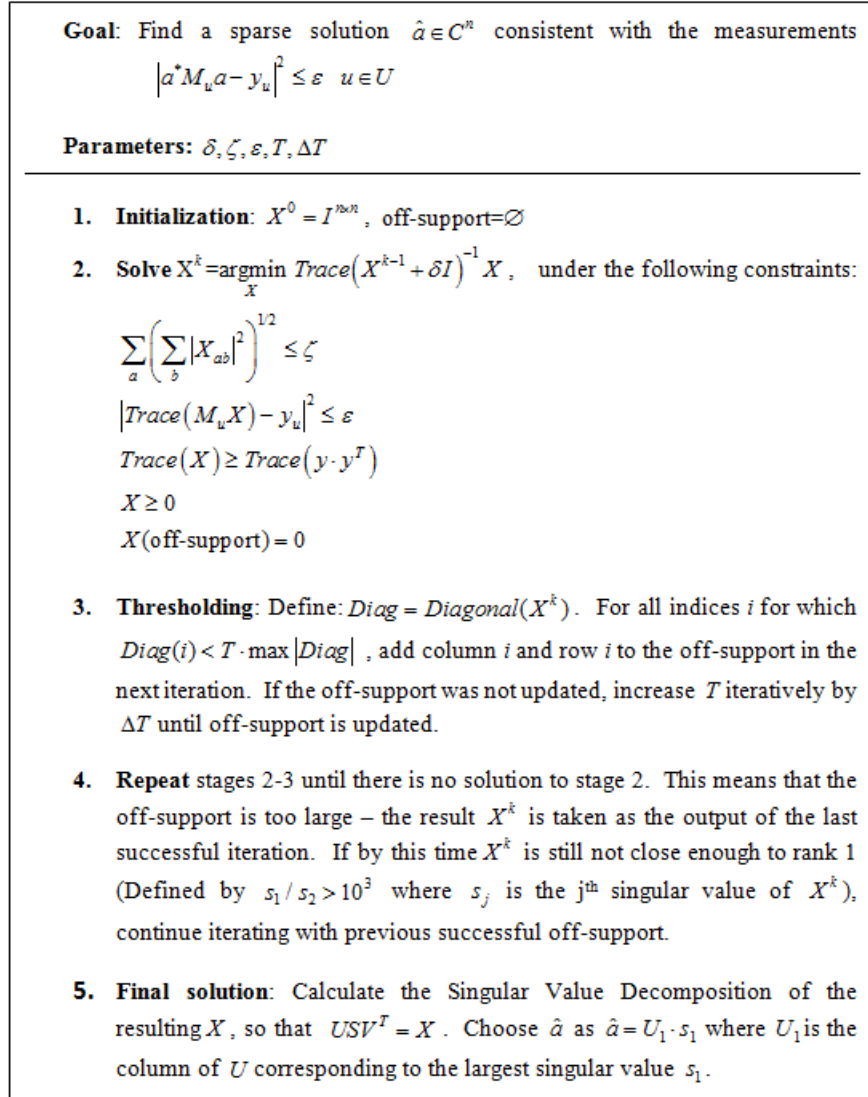
Goal: Find a sparse solution $\hat{a} \in C^n$ consistent with the measurements

$$|a^* M_u a - y_u|^2 \leq \varepsilon \quad u \in U$$

Parameters: $\delta, \zeta, \varepsilon, T, \Delta T$

1. **Initialization**: $X^0 = I^{n \times n}$, off-support$= \varnothing$

2. **Solve** $X^k = \underset{X}{\text{argmin}} \; Trace\left(X^{k-1} + \delta I\right)^{-1} X$, under the following constraints:

$$\sum_a \left(\sum_b |X_{ab}|^2 \right)^{1/2} \leq \zeta$$

$$\left| Trace\left(M_u X\right) - y_u \right|^2 \leq \varepsilon$$

$$Trace(X) \geq Trace\left(y \cdot y^T\right)$$

$$X \geq 0$$

$$X(\text{off-support}) = 0$$

3. **Thresholding**: Define: $Diag = Diagonal(X^k)$. For all indices i for which $Diag(i) < T \cdot \max |Diag|$, add column i and row i to the off-support in the next iteration. If the off-support was not updated, increase T iteratively by ΔT until off-support is updated.

4. **Repeat** stages 2-3 until there is no solution to stage 2. This means that the off-support is too large – the result X^k is taken as the output of the last successful iteration. If by this time X^k is still not close enough to rank 1 (Defined by $s_1 / s_2 > 10^3$ where s_j is the jth singular value of X^k), continue iterating with previous successful off-support.

5. **Final solution**: Calculate the Singular Value Decomposition of the resulting X, so that $USV^T = X$. Choose $\hat{a}$ as $\hat{a} = U_1 \cdot s_1$ where U_1 is the column of U corresponding to the largest singular value s_1.

Fig. 3. Algorithm 1.

The transfer matrices M_u

The set of transfer matrices M_u describing the imaging process are determined by both the mutual intensity function of the light incident on the object plane, and the coherent impulse response of the optical system. These are defined as follows: $M_{u,ij} := h_u(\eta_i) h_u^*(\eta_j) B(\eta_i - \eta_j)$, where $h_u(\eta) := h(u-\eta) \cdot d\eta$, and $h(u-\eta)$ is the mirrored and discretized coherent impulse response of the system, shifted by u. Each matrix M_u is created digitally by:

$$M_u = h_u h_u^* .* B \tag{10}$$

where B is the Toeplitz matrix representing the mutual intensity right before the object plane, and the symbol '.*' denotes element-wise multiplication. For example, for a Gaussian mutual intensity distribution, each row in B is a shifted Gaussian.

As an example, we show a subset of M_u matrices in Fig. 4. Each matrix M_u defines the relation between the unknown object a and the measurement y_u (related through Eq. (5)). Observing Fig. 4, it is apparent that each measurement is a function of the different values of a, centered around the measurement point. There are two physical quantities that affect this dependence. The first quantity is the coherence of the incident light, which affects the effective width of each row in M_u. In the extreme (physically unattainable) case where the light is completely incoherent, each M_u is a diagonal matrix, and then each measurement is a function only of $|a|^2$, i.e. of the intensity at each point in the object. As the coherence of the light is increased, each measurement becomes a function of terms mixing pairs of values of a farther and farther apart (i.e. terms of the form $a_i a_j^*, i \neq j$). The other quantity affecting M_u is the coherent impulse response function of the system which affects the effective spreading of each M_u along the *diagonal* direction. The more spread-out the coherent impulse response of the system is, the more spread out each M_u will be along the diagonal, therefore each measurement y_u will be a mixture by values of a farther away from it. Since the width of the coherent impulse response is at best bounded from below by ~$\lambda/2$ which corresponds to the diffraction limit, each measurement y_u will always be a weighted sum of values in a region of a, which would always lead to smearing of the signal.

This 'matrix' viewpoint is indeed consistent with the physical cutoff frequency acting on the output signal: the output field of the partially-coherent light propagating in the system must be bounded in k-space, due to the low-pass nature of light propagation. An example of this fact can be seen in Fig. 5(c), where the output image is bounded in k-space by $2/\lambda$.
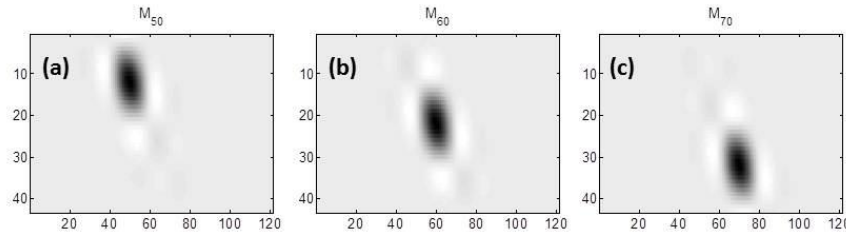

Fig. 4. Example subset of M_u matrices (dark indicates high values), corresponding to a Gaussian mutual-intensity function, and an impulse response of the form of sinc(x). Shown are the matrices M_{50} (a), M_{60} (b), and M_{70} (c).

Numerical reconstruction

We test the proposed reconstruction algorithm numerically, using sparse objects in real space, namely – a few spikes at varying distances and amplitudes. A typical reconstruction simulation is depicted in Fig. 5. In Figs. 5(a) and 5(b) an object comprising of 4 spikes distanced 0.5λ apart is illuminated by partially-incoherent light, with a Gaussian mutual intensity function that is 0.55λ wide (FWHM). The light is then propagating through a diffraction-limited imaging system, and the intensity of the blurred image is measured. The coherent impulse response of the system $h(x)$ is approximated as a sinc function, corresponding to an ideal low-pass filter, yielding: $h(x) = \frac{1}{x} h_0 \sin\left(2\pi x/\lambda\right)$. [In reality, the impulse response of an optical imaging system, especially at high numerical apertures, is considerably more complicated, accounting for aberrations etc.; experimentally, one would actually measure the coherent impulse response of the optical system (amplitude and phase), and substitute it numerically into the algorithm]. To demonstrate the robustness of our technique, we add noise at the level of 40dB (1%) to the measured intensity distribution of the blurred image, i.e. $y_{Noisy} = y + \nu$ where ν is random Gaussian noise and $\|\nu\|_2 / \|y\|_2 = 10^{-2}$ (reasonable for an optical system). Using the "measured" data, we attempt reconstructing the object, using only the detected (smeared) image, and knowledge that the object is sparse in real space. Figure 5(a) shows the reconstructed object using the sparsity prior. For comparison, we show in Fig. 5(b) the reconstructed object without using the knowledge about the sparsity of the input image. Namely, the thresholding procedure is removed from the log-det heuristic iterations, and the small mixed 1-2 constraint is removed as well. The sparsity-based reconstructed image of Fig. 5(a) is virtually identical to the true sparse object.

The sparsity prior knowledge is essential in order to obtain successful reconstruction. This is illustrated in Fig. 5(b), where sparsity is not used; in this case, the reconstructed object is very different than the input optical information (the object we wish to recover).
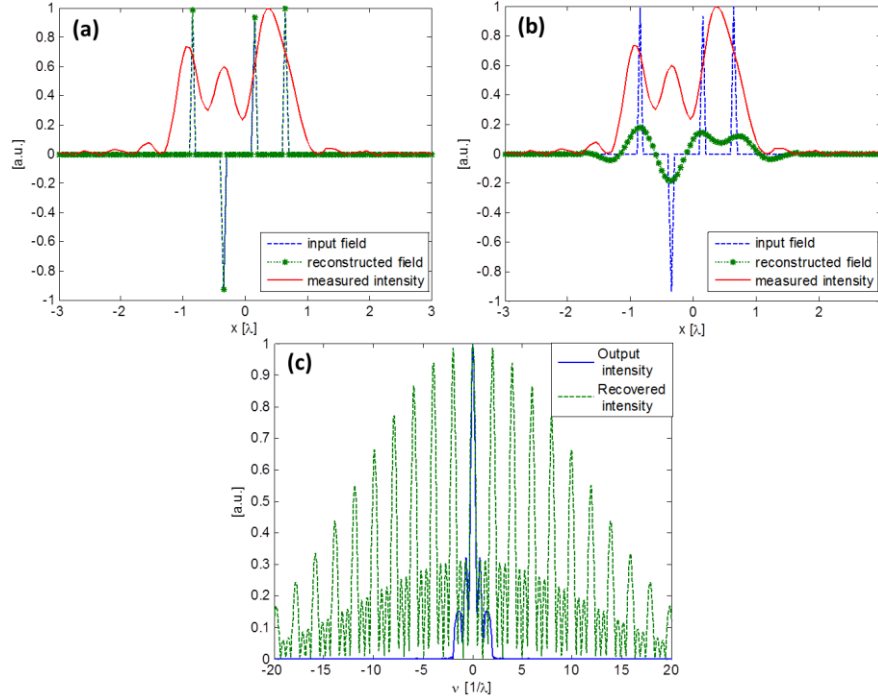
Fig. 5. Reconstruction of a sparse object consisting of 4 spikes at different phases ($\pm\,\pi$), distanced 0.5λ apart. The mutual intensity function of the light impinging on the object is a Gaussian with FWHM of 0.55λ. The coherent impulse response of the optical system is taken

as $h(x) = \frac{1}{x} h_0 \sin\left(\frac{2\pi x}{\lambda}\right)$. (a) Sparsity-based reconstruction, which yields reconstruction that is

practically identical to the true input field (dashed-blue). (b) Reconstruction under the same assumptions, without using the sparsity prior Both reconstructions (dotted green, marked with *) are consistent with the measurements (red). The sparsity-based method of (a) yields a reconstruction virtually identical to the initial object, whereas the reconstruction that does not use sparsity, shown in (b), cannot retrieve any of the fine features of the object. The corresponding bandwidth extrapolation is shown in (c) – the Fourier spectrum of the recovered intensity (dashed green) contains information much higher than the output intensity, which is limited by the physical cutoff frequency at 2/λ (in blue).

Algorithm evaluation

Several aspects of the algorithm's performance are tested numerically. First of all, the necessity of the sparsity assumption is verified. Figure 6 shows the reconstruction error as a function of different noise levels, both for the sparsity-based reconstruction, using the proposed algorithm, and for a reconstruction that does not assume sparsity. The latter is obtained by using the same algorithm, but omitting the sparsity requirements (i.e. the mixed norm bound) and the thresholding iterations. The recovery error is defined as follows:

$$\varepsilon_{\text{rec}} = \frac{\|a_{\text{rec}} - a_{\text{true}}\|_2}{\|a_{\text{true}}\|_2} \tag{11}$$

where a_{rec} is the recovered input signal, and a_{true} is the true input signal, after they have been convolved with a Gaussian of FWHM of 0.1λ (the results are fairly robust to this parameter), in order to allow a small deviation from the true locations, and yet to penalize for larger deviations. The input signal is comprised of 3 peaks, distanced 0.35λ from each-other, each with a random phase (of $\pm\,\pi$), and amplitude of $10 \pm v$ where v is a normally distributed

random variable, i.e. $v \sim N(0,1)$. The number of measurements taken is 121. Due to the low-pass filter nature of the system, the measured signal is bounded in k-space by $2/\lambda$. Therefore, increasing the number of measurements above the Nyquist rate (in this example 25 samples), does not add more frequency information, although it does have the effect of improving reconstruction in the presence of noise.

Each point in the plot is the average over 50 reconstructions. In all simulations the mutual intensity function of the light incident on the object is a Gaussian with FWHM of 0.55λ. Both reconstruction methods perform better as noise is reduced, but the results clearly show the advantage of exploiting the sparsity of the optical information at the input plane. Another aspect that is tested is the effect of the number of peaks on the reconstruction error. This time, the input signal is comprised of several peaks, equally spaced from one another, within a range of 1.3λ. The other parameters are identical to the ones used to produce Fig. 5. Figure 7 shows the average reconstruction error as a function of the number of peaks. The results show that as the signal becomes less sparse (higher peak density), the error increases.

Finally, although the main purpose of this work is not to reduce the number of samples necessary for reconstruction, the effect of the number of samples taken in the image plane is also examined. This is shown in Fig. 8, where the reconstruction error is plotted as a function of the number of equally-spaced measurements in the image plane. The tested scenario is that of 3 peaks, separated by 0.45λ, with other parameters identical to those of Fig. 6. Above a number of measurements comparable to the Nyquist rate (~25 measurements), the reconstruction does not improve dramatically. The problem of bandwidth extrapolation is not solved by adding more samples, since the physical cutoff frequency of the optical system obviously does not depend on this quantity.

In all of the simulations described above, the parameter values (see Fig. 3) are selected as follows: $\delta = 10^{-3}$, $\zeta = 1.1 \cdot \sum_a \left(\sum_b \left| \tilde{X}_{ab} \right|^2 \right)^{1/2}$, $\varepsilon = \max_u \left| \tilde{y}_u - y_u \right|$, $T = 10^{-1}$, $\Delta T = 10^{-2}$ where $\tilde{X} := \tilde{a}\tilde{a}^*$ with $\tilde{a}$ being the true input signal. $\tilde{y}_u$ are the measurements without noise, and y_u are the noisy measurements.
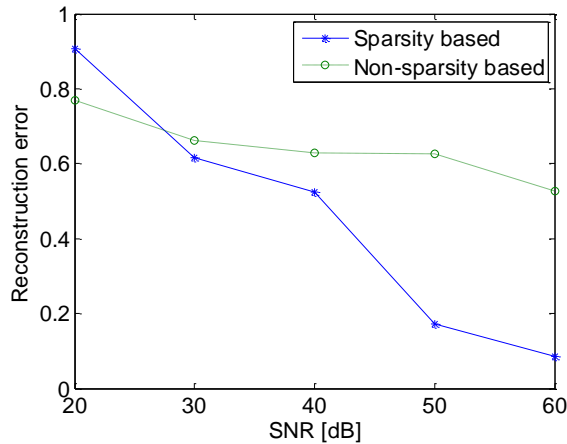

Fig. 6. Reconstruction error as a function of noise in the measurements, for sparsity-based reconstruction (marked with *) and for non-sparsity based reconstruction (dashed). The advantage of sparsity based reconstruction is clear, especially under low noise levels.
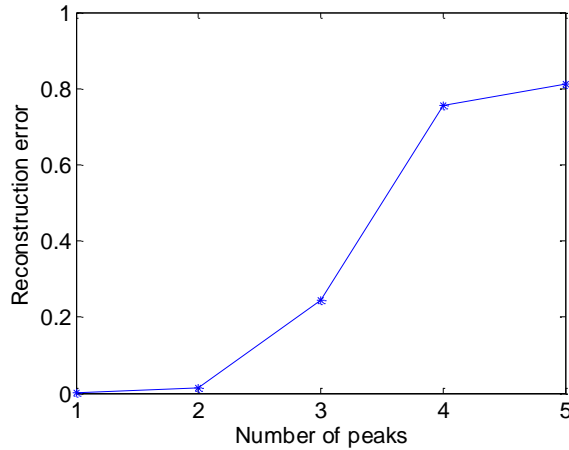
Fig. 7. Reconstruction error as a function of the number of peaks. The peaks are equally spaced inside a range of 1.3λ. Each point is averaged over 100 reconstructions.
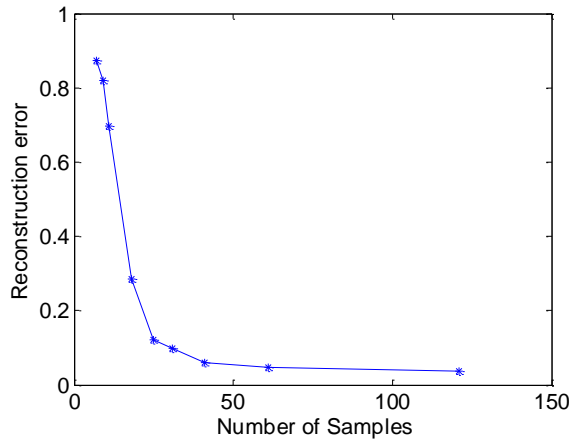

Fig. 8. Reconstruction error as a function of the number of samples in the image plane. The samples are equally spaced. Each point in the plot is averaged over 100 reconstructions.

We presented a sparsity-based technique to recover sub-wavelength features of optical objects illuminated by partially-incoherent light, by using intensity measurements of the blurred image, and exploiting the knowledge that the object is sparse (and only that). The method is based on finding a sparse solution to a system of quadratic equations. Mathematically, this is done by minimizing the rank of a positive semi-definite matrix, under the constraints arising from the physical imaging system.

In a more general scope, the techniques described here show how to recover information from partial (truncated) measurements of the correlation function. The immediate application of this general idea for sub-wavelength optical imaging with partially-spatially-incoherent light can be readily extended to include light with partial temporal coherence, with the knowledge of the spectral distribution of the light. However, the concept of sparsity-based reconstruction of data from partial correlation measurements holds the promise for other areas as well. An immediate example that comes to mind is for FTIR spectroscopy, where our group has indeed demonstrated the ability to distinguish between adjacent atomic lines using measurements of the truncated autocorrelation function. Perhaps most intriguing is the possibility of implementing these ideas on a quantum information system where the measurements are always correlation measurements. Recent work on Compressed Sensing of quantum information has demonstrated signal recovery from sub-Nyquist sampling [19,20],

but has not treated the specific issues of signal recovery from low-pass filtered measurements, or truncated correlation function. Would it be possible to recover information on entangled states from partial (truncated) measurements of the correlation function? Would these ideas also work for higher-order correlation functions? Can these ideas be applied to quantum tomography, quantum photolithography, and quantum cryptography? We expect much future research on sparsity-based techniques to recover information from partial / truncated correlation measurements.