Super-resolution and reconstruction of sparse sub-wavelength images

Snir Gazit[1], **Alexander Szameit**[1], **Yonina C. Eldar**[2], **Mordechai Segev**[1]

[1] *Department of Physics and Solid State Institute, Technion, Haifa 32000, Israel*
[2] *Department of Electrical Engineering, Technion, Haifa 32000, Israel*

msegev@techunix.technion.ac.il

Abstract: We show that, in contrast to popular belief, sub-wavelength information can be recovered from the far-field of an optical image, thereby overcoming the loss of information embedded in decaying evanescent waves. The only requirement is that the image is known to be sparse, a specific but very general and wide-spread property of signals which occur almost everywhere in nature. The reconstruction method relies on newly-developed compressed sensing techniques, which we adapt to optical super-resolution and sub-wavelength imaging. Our approach exhibits robustness to noise and imperfections. We provide an experimental proof-of-principle by demonstrating image recovery at a spatial resolution 5-times higher than the finest resolution defined by a spatial filter. The technique is general, and can be extended beyond optical microscopy, for example, to atomic force microscopes, scanning-tunneling microscopes, and other imaging systems.

OCIS codes: (100.6640) Superresolution; (170.0180) Microscopy; (100.3010) Image reconstruction techniques.

References and links

1. E. Hecht, *Optics* (Addison-Wesley, 1998).
2. M. Saleh and B. Teich, *Fundamentals of Photonics* (Wiley, New York, 1991).
3. E. A. Ash and G. Nicholls, “Super-resolution aperture scanning microscope,” Nature **237**, 510–512 (1972).
4. A. Lewis, M. Isaacson, A. Harotunian, and A. Muray, “Development of a 500å spatial-resolution light-microscope: I. light is efficiently transmitted through l/16 diameter apertures,” Ultramicroscopy **13**, 227–232 (1984).
5. 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**, 1468–1470 (1991).
6. T. W. Ebbesen, H. G. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolf, “Extraordinary optical transmission through subwavelength hole arrays,” Nature **391**, 667–669 (1998).
7. F. M. Huang and N. I. Zheludev, “Super-resolution without evanescent waves,” Nano Lett. **9**, 1249–1254 (2009).
8. J. B. Pendry, “Negative refraction makes a perfect lens,” Phys. Rev. Lett. **85**, 3966–3969 (2000).
9. N. Fang, H. Lee, C. Sun, and X. Zhang, “Sub-diffraction-limited optical imaging with a silver superlens,” Science **308**, 534–537 (2005).
10. Z. Jacob, L. V. Alexeyev, and E. Narimanov, “Optical hyperlens: far-field imaging beyond the diffraction limit,” Opt. Express **14**, 8247–8256 (2006).
11. A. Salandrino and N. Engheta, “Far-field subdiffraction optical microscopy using metamaterial crystals: Theory and simulations,” Phys. Rev. B **74**, 075103 (2006).
12. Z. Liu, H. Lee, Y. Xiong, C. Sun, and X. Zhang, “Far-field optical hyperlens magnifying sub-diffraction-limited objects,” Science **315**, 1686 (2007).
13. I. I. Smolyaninov, Y. J. Hung, and C. C. Davis, “Magnifying superlens in the visible frequency range,” Science **315**, 1699–1701 (2007).

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

14. A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, "Myosin v walks hand-over-hand: Single fluorophore imaging with 1.5nm localization," Science **300**, 2061–2065 (2003).
15. S. W. Hell, R. Schmidt, and A. Egner, "Diffraction-unlimited three-dimensional optical nanoscopy with opposing lenses," Nat. Photon. **3**, 381–387 (2009).
16. N. I. Zheludev, "What diffraction limit?" Nat. Mater. **7**, 420–422 (2008).
17. J. W. Goodman, *Introduction to Fourier optics* (Englewood, CO: Roberts & Co. Publishers, 2005), 3rd ed.
18. A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **22**, 735–742 (1975).
19. R. W. Gerchberg, "Super-resolution through error energy reduction," J. Mod. Opt. **21**, 709–720 (1974).
20. E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**, 489–509 (2006).
21. E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" IEEE Trans. Inf. Theory **52**, 5406–5425 (2006).
22. E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag. **25**, 21–30 (2008).
23. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).
24. Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," IEEE Trans. Signal Process. **57**, 2986–2997 (2009).
25. M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," IEEE Trans. Signal Process. **57**, 993–1009 (2009).
26. A. Ashok, P. K. Baheti, and M. A. Neifeld, "Compressive imaging system design using task-specific information," Appl. Opt. **47**, 4457–4471 (2008).
27. O. Katz, Y. Bromberg, and Y. Silberberg, "Ghost imaging via compressed sensing," in "Frontiers in Optics (FiO)," (2009).
28. Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Near-oracle performance of basis pursuit under random noise," IEEE Trans. Signal Process. (submitted).
29. S. S. Chen, D. L. Donoho, , and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput. **20**, 33–61 (1998).
30. M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," IEEE Trans. Sig. Proc. **50**, 1417–1428 (2002).
31. V. A. Mandelshtam, "FDM: the Filter Diagonalization Method for data processing in NMR experiments," Prog. Nucl. Mag. Res. Sp. **38**, 159–196 (2001).
32. M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," arXiv [0902.4291v1] (2009).
33. D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," Proc. Natl. Acad. Sci. **100**, 2197–2201 (2003).
34. Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," IEEE Signal Proc. Mag. **26**, 48–68 (2009).
35. T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," IEEE Signal Process. Mag. **25**, 31–40 (2008).
36. D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," Proc. Natl. Acad. Sci. **102**, 9446–9451 (2005).
37. A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," IEEE Trans. Inf. Theory **54**, 4813–4820 (2008).

1. Introduction

Back in 1873, Ernst Abbe formulated the theory of imaging, stating that the maximal resolution recoverable in a perfect optical imaging system is determined by the numerical aperture of the lenses involved [1]. Some decades later, it became clear that the true limit on imaging arises from the optical wavelength λ and the best recoverable resolution is $\lambda/2$. This is because the propagation of EM waves in bulk media behaves as a low-pass filter, for distances much larger than the wavelength, rendering spatial frequencies larger than $1/\lambda$ evanescent [2]. Therefore, such spatial frequencies decay rapidly, on a distance scale of several wavelengths. Hence, the observation of sub-wavelength features is essentially impossible using conventional imaging methods. Throughout the years, there have been many attempts to bypass the $\lambda/2$ limit on imaging. One approach is the Near-field Scanning Optical Microscope (NSOM): a very narrow tip, which samples the electromagnetic field at a single point at very close proximity ("near field") to the sub-wavelength specimen, is scanned across the object. The NSOM has nowadays

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

become a frequently used commercial product, but it does have some major disadvantages: (I) the tip must be placed in the near-field, hence it cannot be used, for example, to look into living cells, and (II) acquiring an image requires scanning the specimen point-by-point [3–5], which implies that real-time imaging is impossible. These are severe limitations particularly when studying objects that vary in time (e.g., living objects like bacteria). In the last few years, several genuine ideas have been proposed to allow for more effective sub-wavelength imaging. One approach is based on probing the information with sub-wavelength holes made from thin film of plasmonic metals [6]. A more recent idea relies on constructing super-oscillatory wavepackets to sample at sub-wavelength resolution [7]. However, both of these methods still require scanning, either in the near-field [6] or in the plane where the super-oscillations are generated [7]. Another intriguing avenue is constructing an imaging system made of negative-index materials. The early version of such a system is the "superlens", where the sub-wavelength object is imaged 1:1 to another plane [8, 9]. Hence, this superlens cannot yield any magnification of the object features. The more advanced version is the "hyperlens", where the sub-wavelength information is magnified such that its smallest feature is larger than $\lambda/2$, thereby transforming all its evanescent waves into propagating waves - which can subsequently be imaged with an ordinary microscope [10–13]. Both the superlens and the hyperlens, albeit offering much promise, also have various shortcomings: the heavy loss involved in all current optical negative-index materials, and the stringent requirement to fabricate metamaterial structures at nanometer precision, to name only a few. All of these are nontrivial issues, posing serious challenges before such negative-index structures can become viable technology. Other ideas for sub-wavelength imaging rely on distributing smaller-than-wavelength fluorescing items on the object and repeating the experiments multiple times. In this way, the ensemble-average fluorescent light together with prior knowledge on the size and shape of the items facilitate acquiring sub-wavelength information on the object [14, 15]. However, these methods are once again not real-time, and in addition in many cases attaching external items is undesirable, especially when dealing with biological specimen. Altogether, in spite of the major progress recently accomplished with sub-wavelength optical imaging (for a recent review see [16]), having a far-field method that could do real-time imaging with sub-wavelength resolution is still a long-standing goal.

In parallel to the attempts to obtain sub-wavelength imaging through "hardware", there have been several attempts to achieve this goal through theoretical tools, such as bandwidth extrapolation and related techniques [17]. The key ideas in all of these methods are (quoting from Ref. [17]) (1) the far-field of a spatially-bounded 2D image is described by an analytic function, and (2) if an analytic function is known exactly in an arbitrarily small (but finite) region of the far field, then the entire function can be found uniquely by means of analytic continuation. These concepts and the extrapolation methods arising from them theoretically allow the recovery of sub-wavelength information [18, 19]. However, all of these algorithms are known to be extremely sensitive to noise in the measured data. As summarized by Goodman's 2005 book [17], "all methods for extrapolating bandwidth beyond the diffraction limit are known to be extremely sensitive to both noise in the measured data and the accuracy of the assumed a priori knowledge" and "it is generally agreed that the Rayleigh diffraction limit represents a practical frontier that cannot be overcome with a conventional imaging system."

Here, we show theoretically and provide an experimental proof of concept that sub-wavelength information can be recovered robustly from the far-field of an optical image, overcoming the loss of information embedded in decaying evanescent waves. The only requirement is that the image is known to be sparse, a specific but very general and wide-spread property of signals which occur almost everywhere in nature. Our approach is based on theoretical tools from the emerging field of Compressed Sensing (CS), which is being used to reduce sampling rates in information processing [20–25]. Recently, CS has been suggested in the context of

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

optics [26] and was actually used for ghost imaging [27]. Our purpose is different: to recover the information contained in spatial frequencies that were cut off by diffraction limit, which acts as a low pass filter. We reformulate sub-wavelength imaging as a sparse sampling problem. We provide several examples demonstrating reconstruction of 1D and 2D images of sub-wavelength resolution, and discuss the interrelation between the three parameters controlling the system: sparsity, resolution, and the optical wavelength. In addition, we extend the standard basis-pursuit algorithm [20, 22, 28] commonly used in CS for sparse signal recovery, in order to enable reconstruction of optical images including non-uniform phase, which is an essential attribute in optical image recovery. As we show in the Theory Appendix, basis-pursuit alone is not able to recover features close in space that have opposite phase. Therefore, this novel feature is crucial to the extraction of physically relevant data from the image, through the phase of the electromagnetic field. We then provide an experimental proof-of-principle of our approach, by demonstrating image recovery at a spatial resolution greatly exceeding the finest resolution defined by a spatial filter. Finally, we discuss the general concept and broad applicability of our technique, and its possible extension to other, non-optical, microscopes, such as Atomic Force Microscopes, Scanning Tunneling Microscopes, Magnetic Microscopes, and more.

2. Theoretical considerations

Consider the time-harmonic EM field at some initial plane $z = 0$:

$$E(x,y,z=0) = \mathfrak{Re}\left\{ f(x,y)e^{i\omega t} \right\} \tag{1}$$

where ω is the optical frequency of the wave. The spatial function $f(x,y)$ can be expanded as a function of plane waves $f(x,y) = \iint F(k_x,k_y)e^{-i(k_x x+k_y y)}dk_x dk_y$, where k_x, k_y are the transverse wave numbers, related to the spatial frequencies by $k_x = 2\pi\nu_x$, $k_y = 2\pi\nu_y$ etc. The propagation of the field at all planes $z > 0$, in a homogeneous isotropic and linear medium, can be described through

$$g(x,y) = \iint F(k_x,k_y)H(k_x,k_y,z)e^{-i(k_x x+k_y y)}dk_x dk_y \tag{2}$$

where $H(k_x,k_y,z)e^{ik_z z}$ is the optical coherent transfer function, with $k_z = (k^2 - k_x^2 - k_y^2)^{-\frac{1}{2}}$. Here $k = \omega/c = 2\pi/\lambda$ is the wave number, and c and λ are the speed of light and the wavelength in the medium, respectively. The limits of the integral in Eq. (2) are determined by the numerical aperture of the system. However, even if the system in principle has infinite width (as in free space), the transfer function always acts as a low-pass filter. Since for sufficiently large spatial frequencies k_z becomes imaginary, such waves decay exponentially with propagation. Thus, for propagation distances z much larger than the wavelength, $\left|H(k_x,k_y,z)e^{ik_z z}\right|$ has a cutoff at $k^2 = k_x^2 + k_y^2$, and all waves with spatial frequencies beyond the cutoff are evanescent. This is the reason why sub-wavelength information imprinted on EM fields cannot be observed by conventional imaging techniques.

To explain our technique, consider an EM wave that has propagated a distance z much larger than the wavelength λ. Since the transfer function acts as a low-pass filter, all spatial frequencies larger than $1/\lambda$ are lost. We will now show that the information detected in the far-field can be used to recover the sub-wavelength features, in a robust fashion, provided they are sparse in an appropriate basis.

Let us first explain our approach on intuitive grounds. The idea can be elegantly illustrated by first pinpointing why other extrapolation methods fail: they are not robust to noise in the measured data. As a simple example, consider Taylor expansion as a means for analytic extension from some region in the far-field, close to the cutoff spatial frequency. Taylor expansion fails when the value of some term in the expansion (some higher derivative) is comparable

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

to the noise in the measured data. Other, more advanced, extrapolation methods [18, 19] fail for similar reasons [17]. An illustrative comparison between the performance of the technique following [18] and our CS-based approach is shown in the Theory Appendix. The comparison was carried out on our experimental data - through which our technique gives excellent reconstruction, as shown in Figs. 4 and 5 below. As demonstrated in that Appendix, the extrapolation method of [18], performed on the same experimentally measured data, fails to reconstruct the information. A comparison was also carried out with Taylor expansion, and the result is even worse. Let us now discuss why these methods fail. All extrapolation methods rely on projecting the measured data on some set of orthogonal functions (a basis) spanning the space of solutions. The noise in most physical systems is uncorrelated, hence it is distributed uniformly on the basis functions. The extrapolation methods fail when the value of some projection on the basis functions is comparable to (or lower than) the noise in the measured data, which obviously introduces large errors. Over the years, various ideas for rectifying this problem have been studied [18, 19], all relying on additional a priori knowledge on the information. But this introduces another problem: the extrapolation methods now become very sensitive to the a priori assumptions, where small inaccuracies can introduce large errors. This is why generally extrapolation methods have failed in optical sub-wavelength imaging, exactly as stated in [17].

Let us now explain the intuition underlying CS, and the reason why it succeeds where other extrapolation methods have always failed. CS relies on a single assumption: that the information is sparse, in some basis spanning the space of solutions. If the information is sparse, it is possible to find a proper basis, where we could identify a sharp separation into two sub-spaces: a sub-space where the projections of the measured data is much larger than the noise, and another sub-space where the projections are very small, and can be set to zero without losing much information. If we could somehow identify these basis functions, which of course depend on the actual data, we could use only the sub-space where the projections are large, and completely ignore the other sub-space. Such method will not suffer from noise, because we do not use the sub-space where the projections are small and susceptible to noise. The CS technique does exactly that: it automatically identifies the first sub-space, and ignores the second. To do that, CS uses prior knowledge that the information is sparse (and just that), which implies that the information can be represented in a very compact way in some (mathematical) basis spanning only a (preferably small) sub-space of all possible solutions. Then, since the uncorrelated noise is distributed uniformly on all basis functions, a large fraction of the noise lies in unoccupied basis functions (the second sub-space which we ignore). This is the main idea behind CS and its robustness to noise. We use this idea in reverse logic, to recover sparse high-bandwidth information that was low-pass filtered, in a highly robust fashion. As explained below and in the Theory Appendix, CS has a tradeoff between two parameters: sparsity - the fraction of degrees of freedom occupied by the sparse information, and the desired signal extrapolation ratio - the ratio between all degrees of freedom (known + missing) in the system, and the known (measured) degrees of freedom. In our case of sub-wavelength features, sparsity is the ratio between the non-zero features and the total field of view, whereas the signal extrapolation ratio is the ratio between the full bandwidth of the sub-wavelength information and the measured spatial bandwidth determined by the numerical aperture of the system.

On more mathematical grounds, the ability to reconstruct sparse signals from a limited number of measurements has become feasible with the emergence of a new signal processing technique called Compressed Sensing (CS) [20–23, 25, 28]. This method challenges the traditional limits on the signal reconstruction and measurement process. The underlying logic behind this approach is that sparsely represented signals hold a very limited number of degrees of freedom, since only a small fraction of their coefficients in a particular mathematical basis representation are non-zero. Hence, sparsity is extremely powerful and useful prior informa-

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

tion, enabling considerable reduction in the number of measurements required to reconstruct the signal. In what follows, we demonstrate the results of the CS technique applied to reconstructing sub-wavelength optical information from the measured far-field of a signal, which is the optical analogue to the Fourier transform of the signal after passing through a low-pass filter. Theoretical background on CS and the new recovery techniques we develop here are provided in the Theory Appendix.

The problem of reconstructing sub-wavelength images is equivalent to that of recovering a signal from its low spatial frequencies only. Clearly this is impossible without additional information. Here we exploit the knowledge that the signal is sparse (and nothing else!) to resolve the fine sub-wavelength features. To see how we benefit from sparsity, note that sparse signals can be represented very compactly in a given basis, meaning that only a small fraction of their projections on the basis functions are non-zero. This feature significantly restricts the number of degrees of freedom the signal possesses. More specifically, each non-zero coefficient holds exactly two degrees of freedom: one for the amplitude of the projection and the other for the choice of the basis function. If we knew in advance which functions are chosen, then the degrees of freedom would be reduced in half. Given that the relative fraction of occupied basis functions is β (< 1), we only need to determine β samples of the signal in an alternative basis expansion. However, we must choose the measurement basis wisely such that the combined matrix describing the signal and measurement bases is (left-) invertible, to ensure the existence of a solution. This follows from standard linear algebra considerations and is well known.

We now turn to the more interesting setting, in which we know that the signal is sparse, but we do not know the location of the basis elements comprising the signal. In this case, the degrees of freedom are doubled. We therefore expect that at least a fraction of 2β measurements of the total number of possible measurements are required. However, since the chosen basis functions are unknown, it is now less clear how to choose the measurement basis and how to recover the signal. An essential result of CS is that ***we need to choose a measurement basis such that is satisfies the requirement of invertibility***, obtained in the case in which the locations are known, ***for every possible set of locations***. This mathematical condition is quite difficult to verify in practice; however, it can be shown that a sufficient condition is that the measurement basis is uncorrelated with the signal basis. To understand this requirement intuitively, suppose first that the signal basis is orthonormal, and we choose as a measurement basis the signal basis itself. In this case, the majority of the measurements will yield zero, and contain no information about the true signal. We would have to acquire almost all of the projections to make sure we have not lost any information. Instead, we would like to choose the measurement basis such that that a measurement of any projection in this particular basis contains information about the signal. This can be achieved by ***requiring that each measurement basis function has low correlation with each signal basis function***. A highly uncorrelated pair of bases obeys a specific mathematical condition. This important theorem, similar to the uncertainty principle in quantum mechanics, prevents a signal from being sparse in both bases, and ensures that, if the signal is sparse in one of the bases, it will be very spread in the other. Therefore almost each projection will yield a non-zero informative measurement. Classical examples of maximally uncorrelated bases are the spatial and Fourier domains: A highly sparse signal, e.g. a single Dirac delta function is Fourier-transformed into a spread function that covers the entire spectrum. In our sub-wavelength optical setting, the measurement basis is fixed as the low spatial frequencies in the Fourier domain. According to the discussion above, measuring these will be sufficient to recover the signal if it is sparse in a real-space basis that is uncorrelated with the low-pass Fourier basis. This is in particular the case if the signal is highly localized in real-space.

In the next section we will address how the signal can be recovered in practice from measurements that are all contained within the low-pass filter window. If the correlation be-

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

tween the measurement and signal bases is low enough, then one can prove that a combinatorial search over all sets of basis functions will recover the true underlying signal. However, clearly this approach exhibits high complexity. Instead, a variety of different recovery algorithms have been proposed that run in polynomial time, and are aimed at seeking a sparse signal that is consistent with the given measurements. One of the most common techniques is the basis-pursuit method [29], which amounts to solving an l_1 optimization problem involving minimization of a l_1 norm, and can be implemented quickly and efficiently. A key result of CS is that at the expense of slightly increasing the number of measurements, or in turn, a more demanding condition on the signal sparsity, these polynomial-time algorithms can recover the true sparse signal ***in a robust fashion***, provided that the measurement and signal bases have sufficiently low correlation. ***Consequently, noise in the measurement (which is always present in any physical system) can be tolerated by a slight increase in the required sparsity of the object***. In the context of optical imaging, an important feature is the ability to detect signals with nonuniform phase. As we illustrate in the Theory Appendix, the standard basis-pursuit approach is not able to resolve fine details with different phase. Therefore, we extend this technique to account for nonuniform phase by adding an iterative nonlocal thresholding step. This new method is described in detail in the Appendix, and is referred to as Non Local Hard Thresholding (NLHT).

Our technique is demonstrated theoretically in Fig. 1(a), showing the ability to recover phase and amplitude information that is $(\beta/2)$-times smaller than the wavelength, in a robust fashion. The sub-wavelength information is represented by a one-dimensional optical image with alternating phase. Such a signal, with alternating phase, can be reconstructed via CS techniques, but not by the standard basis-pursuit algorithm when the non-zero information is close in space. In order to recover signals containing arbitrary phases, we develop the NLHT algorithm described in the Theory Appendix. The data in Fig. 1 represents, e.g., a sequence of 1D items with different amplitudes (grey levels) and phases [Fig. 1(a)]. The spatial frequency spectrum of this image is shown in Fig. 1(b), where the red lines mark the cutoff boundaries of the low-pass filter $|H(k_x)|$ at $k_x = \pm 2\pi/\lambda$. In conventional optical imaging systems, the contents at all frequencies beyond the cutoff are lost [Fig. 1(d)]. Hence, the observed optical image is strongly deteriorated [Fig. 1(c)]. For example, the loss of information beyond the cutoff renders the two peaks around $x/\lambda \approx 6$ indistinguishable, in the observed image. Using CS (our NLHT algorithm), we are able to achieve perfect recovery of both the image [Fig. 1(e)] and its spatial spectrum [Fig. 1(f)].

To demonstrate the robustness of NLHT, we add noise to the system; evidently, the reconstruction is robust and the noise has a very small effect on the recovered image. In order to meet physical relevant conditions, we use uniformly-distributed noise in Fourier space, amounting to 1% of the image power. Since the sparsity of this particular image is $\beta = 0.03$ in real space, the recoverable spatial frequencies (as we explain below) are in the range $k_x = \pm 2\pi/(2\beta\lambda)$, greatly exceeding the initial low pass window. ***Clearly, the results displayed in Fig. 1 demonstrate that indeed CS methods facilitate robust and accurate recovery of sub-wavelength optical amplitude and phase information***. Comparing the robustness to noise of CS techniques to other bandwidth extrapolation methods yields overwhelming results. See the discussion in the Theory Appendix. As shown there, CS techniques are robust to noise even at low SNR, where other bandwidth extrapolation methods completely fail even in the presence of weak noise, as discussed in [17] and references therein.

As demonstrated in Fig. 1 theoretically, CS can facilitate the recovery of sub-wavelength information, based on a-priori knowledge that the information is sparse. The key idea is to exploit sparsity. We point out that there are other modern information processing techniques that are also sparsity-based, such as Finite Rate of Innovation (FRI) [30], which was also applied for calculating fine spectral lines in nuclear magnetic resonance (under the name FDM) [31].

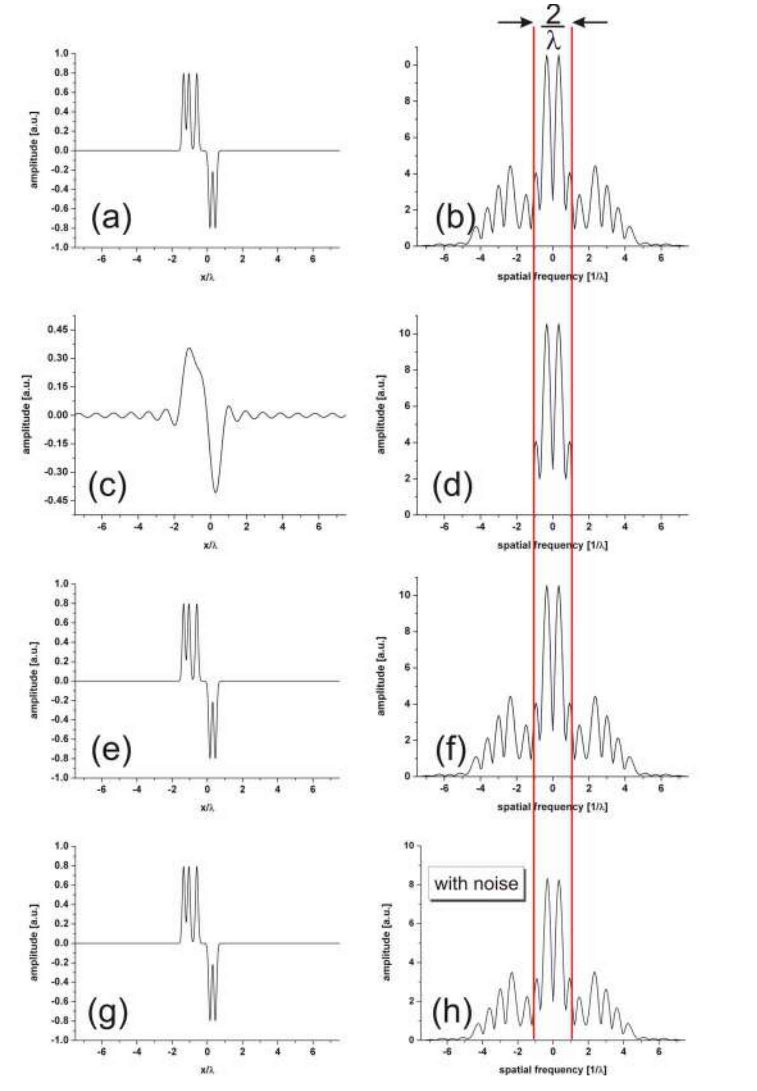


Fig. 1. **Theoretical reconstruction of one-dimensional sub-wavelength information (amplitude and phase).** (a) The original function, which we want to reconstruct. (b) The Fourier (plane-wave) spectrum of the original information shown in (a). The vertical red lines indicate the width of the low-pass filter, which for sub-wavelength information is $2/\lambda$. (c) The distorted image obtained by an inverse Fourier transform on the filtered spectrum; the features are highly blurred. (d) The low-pass-filtered spectrum; a large fraction of the frequency contents is lost. (e,f) Reconstructed image (e) and its spectrum (f) using CS-methods based on the sparsity of the original information. The function is reconstructed perfectly in both real space and Fourier space, including the phase information. Our algorithm is robust against noise. (g,h) Adding 1% noise to the filtered spectrum (not shown here), we are still able to reconstruct the original information at high quality in both real space (g) and Fourier space (h). Amplitude and intensity are given in arbitrary units (a.u.), because the system does not depend on the light intensity.

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

In principle, any method supporting sparsity can be used. However, as we show in the Theory Appendix, some of these approaches may be highly sensitive to noise in some settings. In the Appendix, we briefly compare between our CS-based method and FRI, showing that in the presence of noise CS often performs much better. In addition, as we also demonstrate there, our method tends to be more robust at recovering information with closely-spaced phase variations, while standard CS and FRI techniques have difficulties resolving such details. Our point here is to bring forth the advantage of exploiting sparsity in sub-wavelength imaging, and to suggest one possible method that in our examples appears to be robust and provide superior recovery in comparison with other techniques, to the extent that it can be used for imaging of sub-wavelength information. However, certainly, this calls for a more detailed and careful comparison with other techniques, and possibly even coming up with new sparsity-based ideas that are tailored to optical imaging. Such algorithms can address the specific issues related to optics (e.g., noise that could be partially correlated, etc.). This is beyond the scope of the current work, but an interesting and important direction for future pursuit.

The ideas presented in Fig. 1 can be extended to reconstructing two-dimensional sub-wavelength features. Figure 1 depicts an example containing 2D sub-wavelength amplitude information. However, the 2D case is physically more challenging, because the scalar relation of Eq. (2) requires a modification to describe inevitable polarization effects. That is, EM waves containing sub-wavelength 2D optical images cannot be linearly polarized. This implies that using CS for 2D sub-wavelength imaging should contain vectorial mapping between real space and the plane-wave spectrum [a unit vector should be added in the integral of Eq. (2)]. Nonetheless, extending the CS techniques described here to 2D sub-wavelength images will require some further sophistication, but it is not a major obstacle in any way. In this sense, Fig. 1 describes a scalar version of the physical reality, simply to demonstrate the ability to recover 2D sub-wavelength images.

3. Experimental proof-of-principle

In what follows we provide proof-of-principle experiments, demonstrating image recovery at a spatial resolution greatly exceeding the finest resolution defined by a spatial filter. The experimental setting (Fig. 3) is the simplest optical imaging system: the so-called 4-f system, with an adjustable slit placed at the common focal plane of the lenses, where it acts as a 1D low-pass spatial filter. It is important to note that our setup does not contain sub-wavelength objects, but rather paraxial objects where the features are much larger than the wavelength. The aperture of the adjustable slit defines the highest resolution in the image recovered optically at the output plane (image plane). As such, our system contains exactly the same physical impact of low-pass filtering as naturally done by $\left|H(k_x,k_y)\right|$ in free space, only that the cutoff spatial frequency in our experiment is controlled by the aperture of our filter, whereas for the transfer function in free space the cutoff is set by the wavelength. Our adjustable filter facilitates precise control over the resolution of the imaging system, since, in contrast to the fixed and symmetric filter window of the optical transfer function in free space, the window size and position can be tuned in our experimental setup. As we explain in details in the Theory and the Experimental Appendices, the data for the reconstruction via-CS can be taken in the Fourier space, or in the (spatially-filtered) image plane, and/or at any plane between the Fourier plane and the image plane. Of course, taking the data at multiple planes constitutes over-sampling, and increases the performance of our CS reconstruction.

We first demonstrate the recovery of a generic amplitude-only picture: 3 stripes at uneven spacing [Fig. 4(a)]. The input information is generated by passing a broad Gaussian beam ("plane wave") through 3 transparent elongated rectangles drawn on an opaque slide (acting as 3 rectangular slits). We emphasize that Fig. 4(a) is the actual input information: it is photographed

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

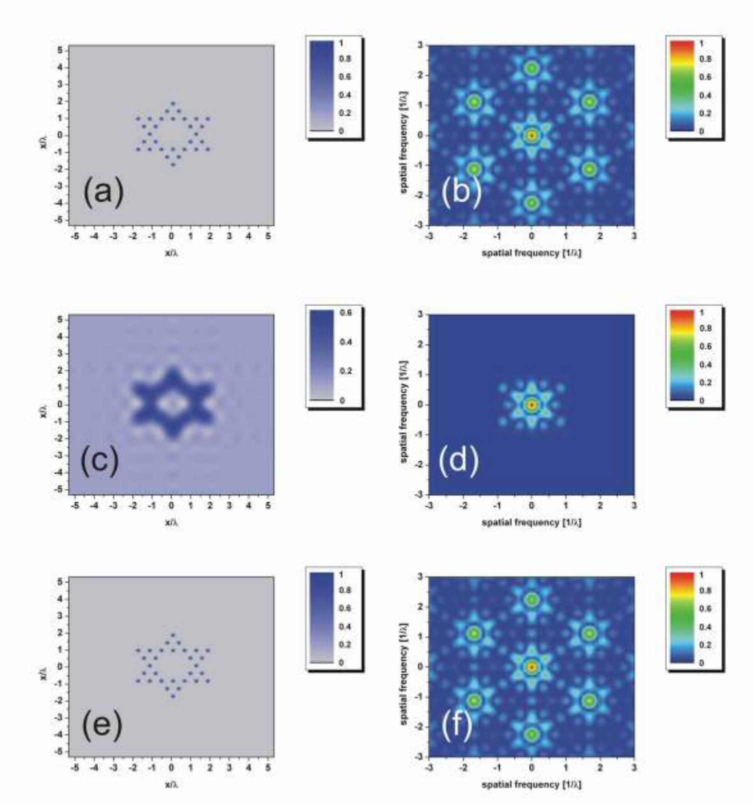


Fig. 2. **Theoretical reconstruction of two-dimensional sub-wavelength information.** (a,b) The original information consists of an arrangement of circles, forming the Star of David (a), and its respective Fourier transform (b). (c,d) After some propagation distance, all spatial frequencies above $1/\lambda$ are lost (d), so that the actual observed image is strongly blurred and the fine features cannot be resolved anymore (c). (e,f) Applying our CS algorithm reveals the underlying sub-wavelength structure in the real space (e), since the Fourier spectrum is fully restored (f).

right after (1mm) the input plane. As such, the horizontal cross-section [Fig. 4(c)] contains 3 almost-perfect square-waves with sharp edges, in contrast to the best optically-recoverable output image generated in our system when the slit is completely open, which has wiggles on each square-wave. These wiggles occur because of the finite aperture of the lenses, which act as a low-pass filter even with opened slit. [This effect is known in information processing as the Gibbs effect]. Our input information passes through the first lens, which generates the Fourier transform of the information, at the focal plane. Figure 4b, showing that plane, depicts the full spatial spectrum of the input information. We then adjust the spatial filter (slit) to cut off a large fraction of the spectrum - leaving practically only the central lobe [Fig. 4(e)]. The output image recovered via direct optical imaging (additional Fourier-transform by the second lens) is now only a single, very broad, intensity peak containing practically only low-frequency information [see Fig. 4(d) for the experimental image and Fig. 4(f) for the horizontal cross-section]. Comparing Fig. 4(d) to Fig. 4(a), and Fig. 4(f) to Fig. 4(c), demonstrates nicely the impact of low-pass filtering on an optical picture, due to the loss of information caused by the filter.

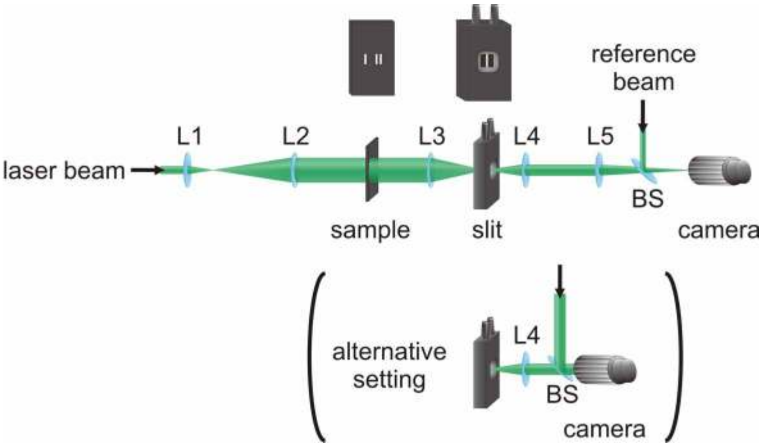


Fig. 3. **Experimental setup for the proof-of-concept experiments.** The laser beam is collimated using lenses L1 and L2, before the sample is illuminated. The signal is then Fourier transformed using lens L3, low-pass filtered by the slit and again Fourier transformed into the real plane by lens L4. Another lens L5 performs an additional Fourier transform, which is recorded by a camera. In order to measure the phase distribution, a probe beam is superimposed (using the beam splitter BS) on the signal in order to create interference fringes. In an alternative setup, the information can be directly taken in the real plane, so that the camera is positioned directly behind lens L4. .

We now employ our CS techniques, on the measured Fourier spectrum acquired after the low-pass filter (which has cut off a large fraction of the spectrum). The simplest CS technique (called basis pursuit; see Theory Appendix) facilitates the reconstruction of 3 accurate stripes, with the appropriate amplitudes and spacing [Fig. 4(g)], thereby circumventing the loss of information caused by the low-pass filter. Importantly, the cross-section shown in Fig. 4(k) reveals that the wiggles are absent. This shows that actually our CS technique performs better than any optical direct-imaging system, by removing the wiggles caused by the finite apertures of the lenses. In this vein, the CS-reconstructed spectrum is almost identical to the original (uncut) spectrum [Fig. 4(h)].

Moving on to an optical image containing phase information, we perform measurements with the structure depicted in Fig. 5(a): two closely-spaced in-phase stripes and a single stripe further away with an opposite phase. The spectrum of this image is shown in Fig. 5(b), the cross section in Fig. 5(c). The low-pass spatial filter is set to pass just the central region of the spectrum [Fig. 5(e)]. Consequently, the image obtained via direct optical imaging (4f system) has two very broad peaks [Fig. 5(d), 5(f)]. Note that, because the input phase information is basically zero and π, any low-pass filtered image always has at least two peaks. We then use CS to recover the image, including both amplitude and phase. To do that, we employ the NLHT algorithm (see Theory Appendix). The CS-recovered image, depicted in Figs. 5(g) and 5(k), is in excellent agreement with the input image, in all of its features. Likewise, its CS-recovered spectrum is very similar to the spectrum of the original image [Fig. 5(h)].

Figures 1–5 demonstrate the ability to recover optical information at a resolution greatly exceeding the maximum resolution (defined by a low-pass filter in Fourier space), that can be recovered by direct optical imaging. Our CS techniques compensate for the loss of information by taking advantage of the sparsity of the input information. It is therefore instructive to estimate the highest resolution recoverable via CS, given the sparsity of the input information β, and the width of the pass-band of the low-pass filter Δk. In principle, in a noise-free scenario,

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

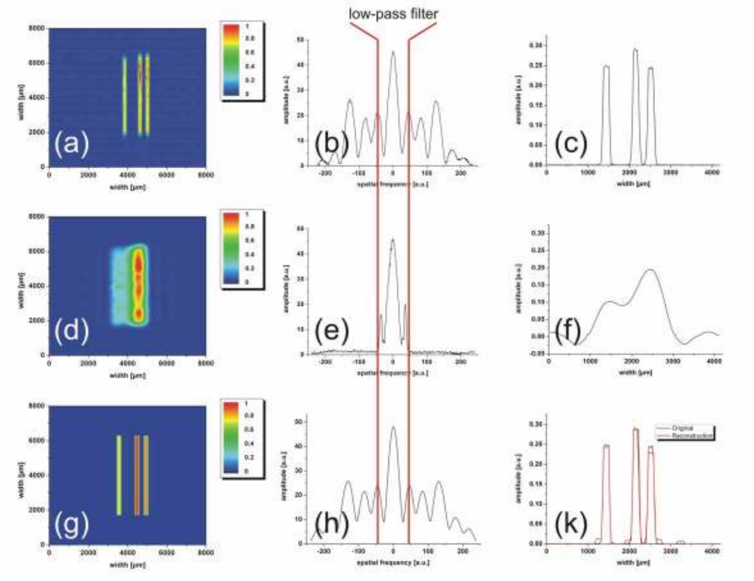


Fig. 4. **Experimental proof-of-concept: reconstruction of amplitude information.** (a,b,c) The original information consisting of three vertical stripes (a), its Fourier spectrum (b), and a horizontal cross-section of the amplitude, taken through the real-space information (c). (d,e,f) Using the optical slit, the signal is low-pass filtered at the vertical red lines, yielding a highly blurred image (d). The Fourier spectrum now contains now only the lowest frequencies (e), which cause the mergence of the three stripes (in real-space) into one, as seen in the horizontal cross section (f). (g,h,k) Reconstruction using CS methods yields a high quality recovered information (g) and its respective Fourier spectrum (h). The strong correspondence between original and recovery is clearly visible in the horizontal cross section (k).

the CS techniques could act by extending the pass-band up to $\Delta k/(2\beta)$. [As explained in the Theory Appendix, CS techniques cannot yield an improvement of $1/\beta$, because there is always a penalty of factor 2 for finding the proper basis]. This would amount to extending the pass-band of the transfer function of free-space $H(k_x,k_y)$, from $\Delta k = 4\pi/\lambda$ to $\Delta k = 4\pi/(2\beta\lambda)$. In the particular example of Fig. 1 ($\beta = 0.03$), the recoverable feature can be as small as $\lambda/16$. In optical microscopy of sparse objects such as living bacteria (where β can be 0.01 and smaller), the resolution is even much higher. Apart from sparsity, another physical limitation is noise, which can never be eliminated. As demonstrated in Fig. 1, CS techniques are rather robust to noise, although noise does reduce their performance. However, the detriment effects of noise can be minimized using over-sampling to increase the precision of the measurements. Using a beam-splitter in the optical system, one could measure simultaneously both the Fourier spectrum and the output image (both after low-pass filtering), and in principle - measure the field distribution in any plane between those. Hence, even though noise will still affect the results somewhat, its detriment effects could be minimized. Finally, the system analyzed in this article assumes coherent illumination (as used in many modern sub-wavelength imaging techniques [8–12]). However, CS techniques are general and can be extended also to imaging with incoherent light.

The key ingredient of all CS techniques is sparsity of the input information. In fact, $\beta = 0.5$ poses a stringent fundamental condition on all CS techniques; that is, without sparsity, CS cannot provide any improvement. It is therefore important to note that the vast majority of natural

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

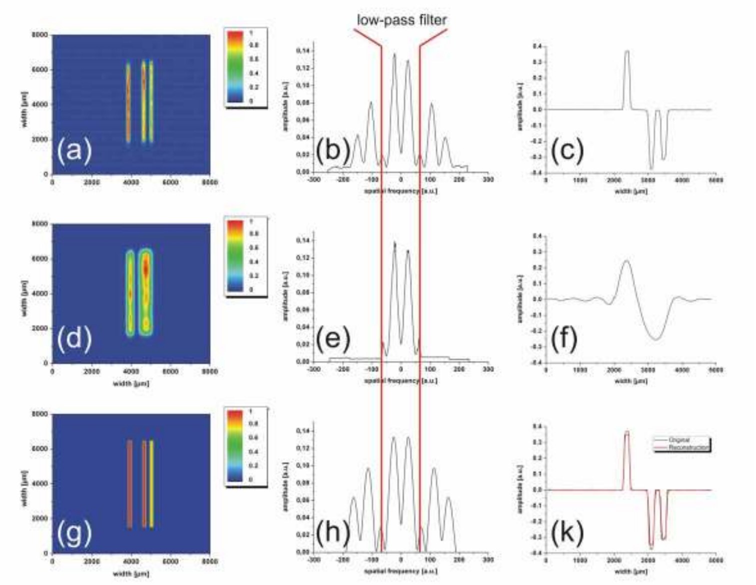


Fig. 5. **Experimental proof-of-concept: reconstruction of amplitude + phase information.** An important feature of our proposed algorithm is the ability to recover both amplitude and phase, which is essential for pictorial information carried upon electromagnetic waves. (a,b,c) The original information consisting of three vertical stripes (a), its Fourier spectrum (b), and a horizontal cross-section of the amplitude, taken through the real-space information, revealing that the two stripes on the right are π-phase shifted with respect to the stripe on the left (c). (d,e,f) Using the optical slit, the signal is low-pass filtered at the vertical red lines, yielding a highly blurred image consisting of two distinct lobes (d). The Fourier spectrum now contains now only the lowest frequencies (e), which cause the mergence of the two stripes on the right, as seen in the horizontal cross section (f). (g,h,k) Reconstruction using CS methods yields a high quality recovered information (g) and its respective Fourier spectrum (h). The strong correspondence between original and recovery is clearly visible in the horizontal cross section (k).

objects, as well as artificial objects, are sparse. Notwithstanding that, the information does not necessarily have to be sparse in real space: it can be sparse in any mathematical basis that is sufficiently incoherent with the Fourier basis. Moreover, one can use a mask with random phase (speckles) in the near field right after the object, which projects more information from the original signal into the low-frequency range, thereby increasing the amount of measurable data [32]. An excellent example for naturally-sparse information is the interior of a living bacterium, which occupies only a small fraction of the area of the cross sections, being therefore highly sparse. Another example of sparse objects, this time from the man-made world, is liquid crystals consisting of giant molecules with lengths slightly below the visible wavelength. In both of these examples, CS can provide a major improvement of ”looking beyond the resolution limit”. Of course, there are objects that are not sparse, for example, electronic chips. However, it is clear that sparse objects are not esoteric, but are rather common in very many systems, especially in biological specimen. ***Finally, we emphasize that our approach can be applied to every optical microscope as a simple computerized image processing tool, delivering results in almost real time with practically no additional hardware***. Our technique is very general, and can be extended also to other, non-optical, microscopes, such as atomic force microscope,

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

scanning-tunneling microscope, magnetic microscopes, and other imaging systems. The main idea presented here holds the promise to revolutionize the world of microscopy with just minor adjustments to current technology: sparse sub-wavelength images could be recovered by making efficient use of their available degrees of freedom.

A. Appendix: Experiment

The objective of the experimental setup is to provide a proof-of-concept that CS techniques can be used to recover pictorial optical information at a spatial resolution greatly exceeding the finest resolution defined by a spatial filter. Our experiments demonstrate exactly that: image reconstruction in spite of a major loss of information caused by a low-pass filter placed in the Fourier space. The basic principles underlying this proof of concept are identical to the ability to reconstruct information that was lost due to the optical transfer function, which acts as a fixed low-pass filter, because it renders evanescent all waves carrying sub-wavelength information. The ability to recover information that was lost, either because it is cut off by a filter (as in our experiments) or because the waves carrying it are exponentially-decaying, is basically identical. Hence, our experiments are indeed a proof of concept for recovering sub-wavelength information via CS techniques.

For this purpose, we use a 4-f-setup, as shown in Fig. 3. A laser beam (Verdi 5W, Coherent Inc.) at λ=532 nm is collimated using a telescope of lenses L1 and L2, and passed through a mask. The information upon the mask is imprinted on the beam, and serves as the input information. To facilitate filtering in the Fourier domain, we optically Fourier-transform the information using the lens L3. At the focal plane of this lens (where the Fourier spectrum is obtained) we place a tunable slit, acting as a low-pass filter. It is important to note that this setup covers all physical features of the low-pass filtering due to the optical transfer function, with the important difference that transmission window is arbitrary in both size and symmetry. The Fourier transform back into the real domain is then accomplished by another lens L4. The measurements are carried out such that they take maximum advantage of the sparsity of the input information. To do this, it is essential that the input information has a maximum attainable field of view: the largest possible with lens L3 (the limiting factor is of course the numerical aperture of L3). This is accomplished by a proper choice of the focal lengths of lenses L3 and L4.

Our measurements are carried out with a conventional CCD camera (Cohu 3400), placed either at the Fourier plane, where it measures the cut spectrum, or at the image plane - at the output of the 4-f system, where it measures the filtered-information. [In fact, using beam-splitters and two cameras, one could perform the measurements at both planes simultaneously, which offers over-sampling, thereby improving the robustness to noise]. Note that the Fourier spectrum is measured in a Fourier plane which is created by another lens L5. The actual number of measurements in each frame is of course determined by the finite number of pixels in the camera. The camera provides direct measurements of the power-spectrum (or the intensity). The phase information, either in the Fourier plane or in the filtered-image plane, is provided by interference with a plane wave propagating at a known angle. Finally, optical information is inherently 2D, whereas our current experiments are dedicated to 1D information. In order to extract the 1D information from our 2D images, we average over the direction along which the information is uniform, and take a cross-section through the averaged image.

B. Appendix: Theory

This appendix provides an overview of the compressed sensing (CS) framework and techniques, in relation to the underlying objective of recovering subwavelength optical information from measurements conducted in the far-field of a sparse high-resolution image. We also detail our

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

new algorithm Non-Local Hard Thresholding (NLHT), for recovery of signals with nonuniform phase.

We begin in Section B.1 with a brief overview of CS, which complements the description in the paper. In Section B.2 we introduce the signal and sampling model on which our developments are based. Recovery of a signal from its low frequency content is discussed in Section B.3. Some remarks on implementation issues and subwavelength imaging are provided in Section B.4.

B.1. Compressed Sensing

Compressed sensing addresses the problem of reconstructing a signal from a limited number of measurements, when the signal is known to be represented very compactly in a given basis [20, 21, 23]. We consider a finite, discrete-time signal $x[\ell]$ of length N. We wish to reconstruct the signal from $K \ll N$ linear measurements of the form $y_k = \langle \psi_k, x\rangle$ where

$$\langle p,q\rangle = p^H q = \sum_i p_i^* q_i, \tag{3}$$

is the usual inner product, and ψ_k are a given set of measurement vectors. Here $(\cdot)^H$ denotes the conjugate transpose of the corresponding vector or matrix, and $(\cdot)^*$ is the complex conjugate. Clearly, if no prior information is given on x, then we would need N measurements to ensure perfect recovery. In addition, if we construct the measurement matrix Ψ whose rows are equal to ψ_k, then Ψ must be invertible. Performing less than N measurements results in an underdetermined set of equations, leading to ambiguity in the solution. To compensate for the lost information we must add additional priors on x. CS treats the scenario in which x is assumed to be sparse. Such signals can be represented very compactly in a specific basis Φ, so that x can be written as $x = \Phi d$ where only a small number $S \ll N$ of the coefficients d_k are nonzero.

Given a sparsity prior and the measurements y_k, we may attempt to recover x by seeking the sparsest representation $x = \Phi d$ that is consistent with the given measurements. Denoting by $\|d\|_0$ the ℓ_0 pseudo-norm of the vector d which counts the number of nonzero coefficients of d, we may solve

$$(P_0) \quad \min_d \|d\|_0 \quad \text{subject to} \quad y = \tilde{\Psi}x = \tilde{\Psi}\Phi d = Wd. \tag{4}$$

The tilde notation in $\tilde{\Psi}$ is a reminder of the fact that we only preform K measurements, so that $\tilde{\Psi}$ is of size $K \times N$. We define $W = \tilde{\Psi}\Phi$ as the measurement matrix. The recovery process described in (4) is useful only if it produces the exact and unique recovery of the original signal x. To study the properties of (4), we first examine the simple setting where the support of the vector d, namely the location of the nonzero values, is known. In this case, uniqueness is guaranteed as long as the columns of the matrix W corresponding to the support of d are linearly independent. If this condition is satisfied, then we can invert the reduced linear set of equations and obtain exact recovery. Consequently, knowledge of the signal support allows recovery from only S measurements as long as the corresponding S columns of W are linearly independent.

Reconstruction becomes significantly more complicated when the support is unknown. To guarantee the uniqueness of the solution for any signal, without knowledge of the signal support, we have to ensure that *every* $2S$ columns of W are linearly independent. To see this, suppose that we have two solutions d_1 and d_2 that are S-sparse, and satisfy

$$y = Wd_1 = Wd_2. \tag{5}$$

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

Then, taking the difference we have

$$W(d_1 - d_2) = Wz = 0, \tag{6}$$

where we denoted $z = d_1 - d_2$. It is easy to see that the ℓ_0 norm satisfies the triangle inequality:

$$\|d_1\|_0 + \|d_2\|_0 \geq \|d_1 - d_2\|_0. \tag{7}$$

Therefore, $\|z\|_0 \leq 2S$. If every $2S$ columns of W are linearly independent, then $Wz = 0$ implies $z = 0$ for any z such that $\|z\|_0 \leq 2S$. Consequently, under this condition there is a unique S-sparse signal d that satisfies $y = Wd$.

A useful definition in this context is that of the spark of a matrix [33]: Spark(A) is defined as the minimal number of linearly dependent columns of the matrix A. An important virtue of Spark(A) is that if x is in the null space of A, namely $Ax = 0$, then $\|x\|_0 \geq \mathrm{Spark}(A)$. From our discussion above we can therefore conclude that $y = Wd$ has a unique S-sparse solution if $\mathrm{Spark}(A) > 2S$. Calculating the Spark of a matrix is a combinatorial process that becomes computationally intractable as the matrix size grows (one needs to examine every subset of the matrix columns). A simpler approach that relaxes spark computation and ensures uniqueness, is to consider the mutual coherence [33]. The mutual coherence of a matrix is defined as:

$$\mu(A) = \max_{j,i \neq j} = \frac{|a_i^H a_j|}{\|a_i\|_2 \|a_j\|_2}, \tag{8}$$

where a_i is the ith column of A. The mutual coherence measures the largest correlation between the columns of A. Low coherence implies that the measurements are informative and uncorrelated. One can use the mutual coherence in order to bound the Spark [33]:

$$\mathrm{Spark}(A) \geq 1 + \frac{1}{\mu}. \tag{9}$$

From (9) we conclude that if

$$\|d\|_0 \leq \frac{1}{2}(1 + 1/\mu(W)), \tag{10}$$

then there is a unique S-sparse solution to $y = Wd$.

Under condition (10) we are guaranteed that there is a unique sparse solution to (4). However, unfortunately, this problem is an NP (Non-Polynomial) hard combinatorial optimization problem, that becomes computationally intractable as the signal length grows. Instead, the CS literature offers a variety of different relaxation schemes that provide computationally efficient alternatives to (4). In this paper we focus mainly on the basis pursuit (BP) algorithm [29], in which the ℓ_0 norm is relaxed by the sparsity promoting ℓ_1 norm $\|x\|_1 = \sum_i |x_i|$. This transforms (4) into:

$$(P_1) \quad \min_d \|d\|_1 \quad \text{subject to} \quad \tilde{\Psi}x = \tilde{\Psi}\Phi d = Wd. \tag{11}$$

The main advantage of $(P1)$ is that it can be written as a standard linear programming problem, which can be solved efficiently in polynomial time using any one of the many well-known standard software packages for solving such problems. However, the solution of $(P1)$ is not guaranteed to be the sparsest one. If (10) is satisfied, then it can be shown that the solutions to $(P1)$ and $(P0)$ coincide. It is worth noting that this bound is very loose and numerical simulations usually give better average performance rate.

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

B.2. Signal Model

Our goal is to show how we can transform the subwavelength imaging problem into a CS counterpart. To this end, we first introduce the class of signals we treat.

We consider the following 1D spatial information model, which describes the electric field of the EM wave we would like to recover:

$$g(x) = \sum_{\ell=0}^{N-1} d_\ell a(x - \Delta\ell). \tag{12}$$

Here $a(x)$ is a generator describing the image, and d_ℓ are the unknown image coefficients. The generator is chosen according to the specific imaging problem, and should capture as accurately as possible the high frequency content of the image (12) beyond $\nu = 1/(2\Delta)$ (which is beyond our resolution target). A common example is the $\mathrm{sinc}(x)$ function used to describe bandlimited functions; other commonly used functions are splines and interpolating wavelets [34]. In our experiments, the optical mask was composed of rectangular slits corresponding to $a(x) = \mathrm{rect}(x)$. However, it is important to note that the actual choice of the generating function is usually of little importance, since the image is assumed to vary slowly on a length scale which is smaller than Δ. In practice, the value of N is determined according to experimental considerations, as we detail in the following sections. Our goal is to recover the expansion coefficients d_ℓ from the lowpass regime of the image.

To proceed, we define the continuous-spatial Fourier transform (CSFT) of a function $a(x)$ as:

$$A(\nu) = \int_{-\infty}^{\infty} a(x) e^{-j2\pi\nu x} dx, \tag{13}$$

which describes the spectrum of $g(x)$ given by (eq. 1) in the paper. We further define the discrete-spatial Fourier transform of a sequence $d_\ell, 1 \leq \ell \leq N$ by

$$D(e^{j2\pi\Delta\nu}) = \sum_{\ell=0}^{N-1} d_\ell e^{-j2\pi\Delta\ell\nu}. \tag{14}$$

The discrete Fourier transform (DFT) of a length-N sequence d_ℓ is given by

$$D[k] = \sum_{\ell=0}^{N-1} d_\ell e^{-j2\pi k\ell/N}. \tag{15}$$

The sequence d_ℓ can be recovered from $D[k]$ using the inverse DFT:

$$d_\ell = \frac{1}{N} \sum_{k=-(N-1)/2}^{(N-1)/2} D[k] e^{j2\pi k\ell/N}. \tag{16}$$

Here, and throughout, we assume that N is odd; similar results hold for the case of N even with appropriate modifications.

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

Taking the CSFT of $g(x)$ given by (12), we have

$$\begin{aligned} G(\nu) &= \sum_{\ell=0}^{N-1} d_\ell \int a(x-\Delta\ell)e^{-j2\pi\nu x}dx \\ &= \sum_{\ell=0}^{N-1} d_\ell e^{-j2\pi\Delta\ell\nu} \int a(x)e^{-j2\pi\nu x}dx \\ &= D(e^{j2\pi\Delta\nu})A(\nu). \end{aligned} \tag{17}$$

As explained in the paper [Eq. (2)] the optical transfer function is modeled as a perfect low-pass filter (LPF) with cutoff frequency $\nu_c = 1/\lambda$:

$$H(\nu) = \begin{cases} 1 & |\nu| \leq \nu_c \\ 0 & \text{else.} \end{cases} \tag{18}$$

Therefore, in the far field the Fourier transform of the image is given by:

$$G^{\mathrm{FF}}(\nu) = G(\nu)H(\nu) = \begin{cases} D(e^{j2\pi\Delta\nu})A(\nu) & |\nu| \leq \nu_c \\ 0 & \text{else.} \end{cases} \tag{19}$$

Our problem then is to recover d_ℓ from $G^{\mathrm{FF}}(\nu)$.

B.3. Signal Reconstruction

B.3.1. Fourier Sampling

In order to measure $G^{\mathrm{FF}}(\nu)$, we need to sample it. The most straightforward approach is to measure it directly in the Fourier plane. Specifically, the samples are obtained by sampling $G^{\mathrm{FF}}(\nu)$ in the interval $|\nu| \leq \nu_c$ with a uniform spacing of $1/N$. The spectral resolution $\eta = 1/(N\Delta)$ is determined by the resolution of the sensing device and sets the value of $N = 1/(\eta\Delta)$.

In order to make the derivations generic, we rescale the frequency axis by $1/\Delta$ which transforms the cut off frequency to $\nu_c = \alpha = \Delta/\lambda$, the spatial resolution Δ becomes equal to 1, and the first replica of $D(e^{j2\pi\Delta\nu})$ is contained in the interval $[-1/2, 1/2]$. Denoting:

$$k_{\max} = \lfloor \alpha N \rfloor, \tag{20}$$

we may write the samples explicitly as

$$c_k = D(e^{j2\pi k/N})A(k/N) = D[k]A(k/N), \quad |k| \leq k_{\max}. \tag{21}$$

We now distinguish between two different cases:

1. $\alpha \geq (1-1/N)/2$;
2. $\alpha < (1-1/N)/2$.

Case $\alpha \geq (1-1/N)/2$: We begin with the simple case in which $\alpha \geq (1-1/N)/2$, corresponding to $k_{\max} \geq (N-1)/2$. In this regime, we do not lose any information embedded in $D(e^{j2\pi f})$. Indeed, assuming that $A(k/N) \neq 0$ for $|k| \leq k_{\max}$, we can define the modified measurements

$$b_k = \frac{c_k}{A(k/N)} = D[k], \quad |k| \leq k_{\max}. \tag{22}$$

Since $k_{\max} \geq (N-1)/2$, we have enough DFT coefficients $D[k]$ in order to recover d_ℓ via the inverse DFT formula (16). Once these coefficients are known we can construct the image $g(x)$

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

using (12).

Case $\alpha < (1-1/N)/2$*:* When $\alpha < (1-1/N)/2$, the modified coefficients (22) correspond to only a segment of the DFT of d_ℓ since $k_{\max} < (N-1)/2$. Therefore, taking the inverse DFT on these coefficients will not yield the correct sequence d_l, and we have lost some of the information embedded in $D(e^{j2\pi\Delta f})$ due to the LPF effect.

B.3.2. Spatial Domain Sampling

The far-field image can also be measured in the spatial domain. In fact, it can be sampled in any basis as long as there is an invertible transformation between the measurement basis and the Fourier basis. This is an important virtue since it allows flexibility in the physical realization of the measurement process which can be preformed in the Fourier domain, spatial domain, or even in the Fresnel zone. Below we describe sampling in the spatial plane. As explained in the experimental part of the supplementary notes, for practical reasons, in our experiments the measurements were preformed in the spatial plane of the filtered image.

As in the case of fourier sampling, the value of N is determined by the sensing devise; however, here the camera's field of view (FOV) plays the role of the spectral resolution. Since $\eta = 1/\mathrm{FOV}$, we get $N = \mathrm{FOV}/\Delta$, which is the number of features with resolution Δ that fit into the FOV. As before, we rescale the spatial domain resolution so that $\Delta = 1$.

In the spatial domain the far field image takes on the following form:

$$g^{\mathrm{FF}}(x) = \sum_{\ell=0}^{N-1} d_\ell (a_\ell \star s)(x) \tag{23}$$

where

$$a_\ell(x) = a(x-\ell), \tag{24}$$

$s(x)$ is the LPF convolution kernel

$$s(x) = 2\alpha \mathrm{sinc}(2\alpha x), \tag{25}$$

and $(f_1 \star f_2)$ denotes the continuous-time convolution:

$$(f_1 \star f_2)(x) = \int_{-\infty}^{\infty} f_1(x-\chi) f_2(\chi) d\chi. \tag{26}$$

We sample $g^{\mathrm{FF}}(x)$ uniformly with critical sampling rate $T = \Delta = 1$ over the FOV of the camera; this rate follows from noting that the relevant frequency content is limited to the first replica of $D(e^{j2\pi\Delta v})$. In practice, one can increase the sampling rate in the spatial domain in order to improve robustness to noise. The resulting samples are given by

$$g^{\mathrm{FF}}(x=k) = \tilde{d}[k] = \sum_{\ell=0}^{l=N-1} d_\ell \int_{-\infty}^{\infty} a(k-\ell-\chi) s(\chi) d\chi, \quad 0 \le k \le N-1. \tag{27}$$

In order to model the jitter in the measurements we shift the sampling points by τ, resulting in:

$$g^{\mathrm{FF}}(x=k-\tau) = \tilde{d}[k] = \sum_{\ell=0}^{\ell=N-1} d_\ell \int_{-\infty}^{\infty} a(k-\ell-\tau-\chi) s(\chi) d\chi. \tag{28}$$

We can rewrite (27) in matrix form as

$$\tilde{d} = Md \tag{29}$$

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

where $\tilde{d}, d$ are the vectors with elements $d_\ell, \tilde{d}_\ell$ respectively, and

$$M_{k\ell} = \int_{-\infty}^{\infty} a(k-\ell-\tau-\chi)s(\chi)d\chi. \tag{30}$$

It is easy to see that we can recover the DFT coefficients $D[k]$ for $|k| \leq k_{\max}$ from the measurements $\tilde{d}$. Indeed, let $\tilde{F}$ be the Fourier matrix with elements $\tilde{F}_{r\ell} = \frac{1}{\sqrt{N}} e^{-j2\pi r\ell/N}$. Noting that $D = \sqrt{N}\tilde{F}d$ and $\tilde{F}^H\tilde{F} = I$, we have from (29)

$$\sqrt{N}\tilde{F}\tilde{d} = \tilde{F}M\tilde{F}^H D. \tag{31}$$

Since M is an approximately circulant matrix, it is (approximately) diagonalized by $\tilde{F}$. Therefore, $P = \tilde{F}M\tilde{F}^H$ is a diagonal matrix with diagonal values

$$P_{kk} \quad = \quad \begin{cases} A(k/N)e^{-j2\pi\tau k/N}, & |k| \leq k_{\max} \\ 0, & |k| > k_{\max}. \end{cases} \tag{32}$$

(Note the similarity to (21)). We may then take the pseudoinverse of $\tilde{F}M\tilde{F}^H$ in (31) to recover the Fourier coefficients $D[k]$ for $|k| \leq k_{\max}$.

Clearly the matrix P depends on the value of the jitter τ, which is unknown. In the simulations, we tested a few values of $0 \leq \tau \leq 1$ until we obtained the desired results.

B.3.3. Reconstruction with Sparsity Priors

We have seen in the previous sections that regardless of whether we sample in the Fourier domain or the spatial plane, we can recover the frequency content below the LPF cutoff. However, we still need to overcome the loss of high frequency information. To do so we assume a sparsity prior on the series d_ℓ. Let $\beta = S/N$ be the relative fraction of the nonzero elements of d_ℓ, where S is the number of nonzero values. Our goal is to reconstruct the time series d_ℓ from the partial Fourier measurements

$$d_k = D(e^{j2\pi k/N}), \quad |k| \leq k_{\max}, \tag{33}$$

under the assumption that at most S values are nonzero.

To set up our problem within the framework of CS, let F be the $M \times N$ partial Fourier matrix with elements $e^{-j2\pi kl/N}$ for $1 \leq \ell \leq N$ and $-(M-1)/2 \leq k \leq (M-1)/2$. Here

$$M = 2k_{\max} + 1 = 2\lfloor \alpha N \rfloor + 1 \tag{34}$$

is the number of measurements. When $M = N$, F is equal to the full Fourier matrix and is an orthogonal invertible matrix. In this case, the elements of the vector Fd are the DFT coefficients of d_ℓ. When $M < N$, the matrix has more rows than columns and cannot be (left) inverted. In the context of CS, $F = W$ is the measurement matrix.

Denote by d the vector with elements $d_\ell, 1 \leq l \leq N$, and by b the measurement vector with elements $b_k, |k| \leq k_{\max}$ with b_k defined by (22). Our goal then is to find the sparsest vector d that is consistent with the measurements. Similarly to (4) this can be written as

$$(P_0) \quad \min_d \|d\|_0 \quad \text{subject to} \quad b = Fd. \tag{35}$$

When N is prime, it can be shown that every M columns of F are linearly independent [20]. Therefore, (35) has a unique solution if:

$$\frac{M}{N} \geq 2\beta. \tag{36}$$

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

To solve (35) we can use known algorithms from the CS literature such as BP (11). However, these methods are not guaranteed to recover the true d. In the special case where the measurement matrix is a partial Fourier matrix as in (35), the unknown d can be recovered exactly (in the noiseless setting) using the annihilating filter method [35] whenever (36) is satisfied; we will discuss this technique in Section B.3.3. The drawback of this approach is that in many settings it is highly sensitive to noise and is therefore often not reliable. In contrast, BP is more robust, but requires an increase in the number of measurements to recover d. As we will show, BP works well with uniform phase, or in the nonuniform phase setting when the nonzero elements are far enough apart. Unfortunately, when the details to be resolved are closely spaced and with different phases, the BP algorithm tends to fail. To overcome this limitation of the BP technique, we propose a new method, referred to as NLHT, which leads to very good recovery results in the presence of noise and closely spaced elements with nonuniform phase.

In the next subsections we briefly discuss each one of these approaches and point out their limitations in our context of subwavelength imaging. Further we'll provide a comparison with the reconstruction obtained by Gerchberg-Papoulis algorithm on the experimental data.

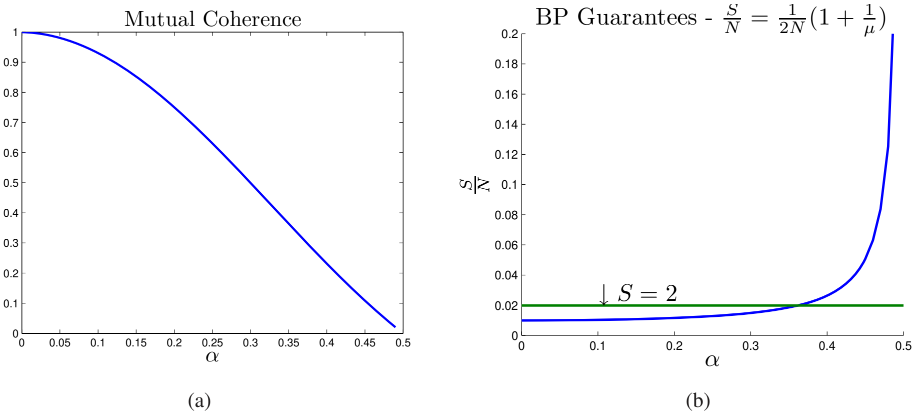


Fig. 6. (a) Mutual coherence of the lowpass Fourier matrix. (b) Reconstruction guarantees for BP. The maximal sparsity level that ensures exact reconstruction remains very low even for relatively high values of α. The case $S = 2$ corresponding to two spikes is illustrated by the horizontal line.

Basis Pursuit The BP algorithm relaxes P_0 by replacing the non-convex ℓ_0 norm in (35) by the sparsity promoting ℓ_1 norm:

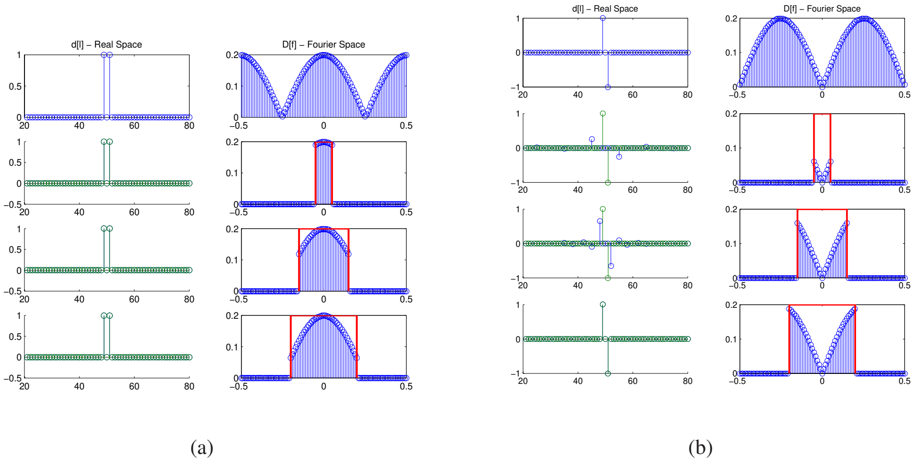
$$(P_1) \quad \min_d \|d\|_1 \quad \text{subject to} \quad b = Fd. \tag{37}$$

As we have seen, if the coherence of F is low enough, then (P_1) and (P_0) will yield identical results. Unfortunately, the partial Fourier matrix consisting of the lowpass frequencies has high coherence even for relatively large values of α, as can be seen in Fig. 6(a). In Fig. 6(b) we plot the BP bound for the maximal degree of sparsity that enables exact recovery. Even for relatively large values of α, we cannot guarantee the reconstruction of two spikes.

Experiments demonstrate that (37) yields good recovery results when the coefficients d_ℓ are in-phase (e.g. $d_\ell \geq 0$). However, in the presence of multiple phases, the method fails to reconstruct d_ℓ when two spikes with different phase are distanced below $2/\alpha$ (corresponding to $\lambda/2$). Figures 7(a) and 7(b) illustrate this phenomena in the case of two spikes. One can clearly

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

see that when the spikes are positive perfect reconstruction is possible far below the diffraction limit; however, in the multi-phase scenario, the method fails. These results have been discussed theoretically in [36, 37].



(a) (b)

Fig. 7. (a) Reconstruction of an in-phase signal. The first row corresponds to the original information in real-space and in Fourier space. The following rows are reconstruction using Basis Pursuit with different cutoff frequencies, as indicated by the red LPF. Green corresponds to the original signal while blue is the reconstructed signal. In this example, these sequences overlap completely matrix. (b) Reconstruction of a multi-phase signal. The first row corresponds to the original sampled information in real space and Fourier space. The following rows are reconstruction using Basis Pursuit with different cutoff frequencies, as indicated by the red LPF. Green corresponds to the original signal while blue is the reconstructed signal. In this example, a high cutoff frequency is needed in order to obtain good recovery.

Annihilating Filter The annihilating filter method, also known as "Finite Rate of Innovation"(FRI), can be used to solve our problem exactly in the noiseless case, under the condition (36). This method is described in [30, 35], and is summarized below.

The algorithm consists of two stages. In the first stage, the locations of the spikes are determined, by finding the roots of an annihilating filer $H(z) = \sum_{n=1}^{S} h_n z^{-n}$ that annihilates the Fourier domain measurements b_k. The filter is of length S as the number of nonzero coefficients of the series d_ℓ. This can be achieved by solving the following set of equations:

$$\begin{bmatrix} b_{-1} & b_{-2} & \dots & b_{-k_{\max}} \\ b_o & b_{-1} & \dots & b_{-k_{\max}+1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k_{\max}-2} & b_{k_{\max}-3} & \dots & b_{-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_S \end{bmatrix} = - \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{k_{\max}-1} \end{bmatrix}. \qquad (38)$$

Once these locations are known, we can solve for the amplitudes of d_ℓ by inverting the system $b = Fd$ over the location set S.

Although, in principle, this approach can yield perfect recovery, the process of root finding is very sensitive to noise. In order to overcome the noise sensitivity it was suggested in [35] to use Cadzow's algorithm that de-noises the signal by imposing self consistency conditions

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

on the convolution matrix of (38): The matrix should have a maximal rank of βN (because the original signal is βN sparse) and also have a Toeplitz structure. Iterating between these two requirements leads to denoising of the measurements. Although this process reduces the noise, the resulting method is still often very sensitive and nonrobust as we demonstrate below.

We tested the algorithm on the following simulation setting. We considered a signal of length $N = 101$, and sparsity levels $\beta = 0.06, 0.07, 0.08, 0.09$. The locations of the spikes were drawn randomly between $0.1N - 0.9N$, with randomly chosen amplitudes drawn from a uniform distribution ranging between $5 - 10$, with a random sign pattern. Figure 8 shows the probability to reconstruct the support of the sparse signal for different signal-to-noise ratios (SNRs) and sparsity levels, where the noise is additive white Gaussian noise. The empirical probability was calculated from 1000 runs; Cadzow's algorithm was used with 10 iterations.

As can been seen from Fig. 8 in the scenario tested, this method is extremely sensitive to noise especially for signals which have adjacent spikes. This point is noted in [35], in which it is stated that in order to ensure exact support recovery the spikes should be distanced enough from one another.

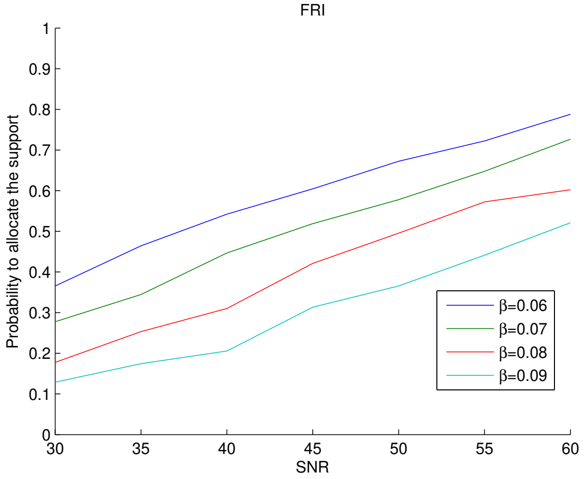


Fig. 8. Probability of support recovery as a function of the SNR using the annihilating filter method.

NLHT: Non Local Hard Thresholding Standard BP is a robust to noise and efficient, but cannot resolve closely-spaced spikes with different phases. Below, we propose a new algorithm, referred to as NLHT, which appears to be both robust and capable of resolving closely-spaced spikes with arbitrary phases.

The algorithm attempts to allocate the off-support of the sparse signal in an iterative fashion, by performing a thresholding step that depends on the values of the neighboring locations (in real space). In each iteration, we use a BP step which takes into account noise with level ε (this algorithm is referred to in the literature as BP denoising [29]):

$$(P_1) \quad \min_d \|d\|_1 \quad s.t \quad \|b - Fd\|_2 \leq \varepsilon. \tag{39}$$

Based on the solution we try to allocate the off-support of the signal by performing a nonlocal thresholding step. Each element of $\hat{d}$ which is below a fixed threshold *along with its neighbors*

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

is zeroed out and considered as off-support. In the next iteration, we repeat the BPDN step (39) with the additional constraint that the locations corresponding to the off-support are set to zero. Table **(Algorithm 1)** provides a more detailed version of the algorithm.

Algorithm 1 Non Local hard Thresholding.

```
Require:
    • $S$ - off support
    • $\mu$ - Nearest neighbor window size
    • $\zeta_0$ - Threshold
    • $\Delta\zeta$ - Increment in the Threshold
    • $\varepsilon$ - Noise level
  Initialize: $S = \emptyset, \quad \mu = \mu_0, \quad \zeta = \zeta_0.$
  Repeat
    Solve:
      $\min_{\hat{d}} \|\hat{d}\|_1 \quad \text{subject to} \quad \left\|b - F\hat{d}\right\|_2 \le \varepsilon, \quad \hat{d}[\ell] = 0, \ \forall \ell \in S.$   (40)
    Allocate the off support:
      1. Find all $\tilde{\ell}$ such that $\hat{d}[\ell] \le \zeta \cdot \max(\hat{d})$ for all $\ell$'s which are distanced from $\tilde{\ell}$ to the
         right or left by $\mu$ or distanced from both sides by $\mu/2$.
      2. Add $\tilde{\ell}$ to $\tilde{S}$.
    Update $S = S \cup \tilde{S}$.
    If the support was not updated increase $\zeta$ by $\Delta\zeta$ and decrease $\mu$ by 1.
  Until $|S| \le \|d\|_0$
Return $\hat{d}$
```

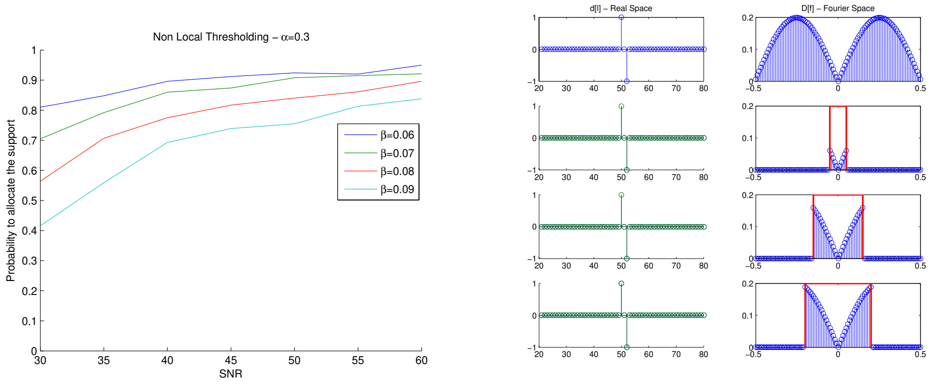
The NLHT method was tested on the same setting as the annihilating filter method. The parameter values were chosen to be $\mu = 9$, $\lambda = 0.025$, $\Delta\lambda = 0.025$, ε as the noise level. In choosing the parameter values one should make sure that the initial window size is significantly larger than $1/\alpha$.

Figure 9(a) plots the probability of support recovery as a function of the SNR. Comparing Figs. 9(a) and 8 reveals a significant improvement with respect to the annihilating filter approach with Cadzow's algorithm.

Figure 9(b) illustrates the reconstruction of multi-phase adjacent spikes using the NLHT algorithm. This is in contrast to the failure of the BP method illustrated in Fig. 7(b).

Gerchberg-Papoulis Algorithm Finally, we provide a comparison between our reconstruction method, which is based on compressed sensing techniques, and the Gerchberg-Papoulis extrapolation algorithm [18, 19]. The comparison is carried out with our actual experimental data: the measured data presented in Fig. 4(d-f). Recall that in the paper we demonstrate successful reconstruction using CS. As we show in this section, using the Gerchberg-Papoulis algorithm on the very same measured data fails in reconstructing the correct information. The noise in the measured data arises from the actual physical noise in our experimental system.

The Gerchberg-Papoulis algorithm [18, 19] attempts to extrapolate the frequency content



(a)

(b)

Fig. 9. (a) Probability of support recovery as a function of the SNR using the NLHT algorithm. (b) Reconstruction results of the NLHT algorithm for multiple-phase spikes.

beyond the cut off frequency by iteratively imposing the measured data under the assumption that the image is space-limited. We start by performing an inverse Fourier transform on the low-pass Fourier data, leading to a spread image in the spatial domain which cannot be space limited. Next, we reinforce the assumption that image is space-limited simply by truncating the image. The resulting space-limited image possesses frequency contents beyond the cut off frequency, leading to a different frequency content than the original data. In addition, the frequency content inside the low-pass is also altered. We therefore force the known measured data on those frequencies. This procedure is repeated in an iterative fashion by employing the above steps, until the frequency and spatial domain images conform to one another. The algorithm is implemented and tested on the same experimental data of Fig. 4 in the article. The image was assumed to be limited to the first lobe in the spatial domain. Hence in the iteration process all values beyond the first lobe were zeroed out. Figure 10 of this section compares between the performance of the Gerchberg-Papoulis algorithm and the algorithm described in the article, which is based on CS. The first row shows the experimental data: the actual blurred image 10(a), the respective Fourier transform 10(b) and a cross section of the amplitude in real space 10(c). Applying the Gerchberg-Papoulis method yields the reconstruction shown in 10(d), with the corresponding Fourier transform 10(e). In particular the cross section of the reconstructed image 10(f) reveals that this algorithm, based on the assumption of space limitation, completely fails, being unable to recover the true image. In contrast, the reconstruction task is easily done by of our CS-based technique. Its advantage in precision and in robustness to noise, based solely on the sparsity of the object, yields the reconstruction depicted in 10(g), which clearly shows the three stripes of the original sample. The respective Fourier transform in 10(h) is almost identical to the original shown in Fig. 4(b). In the cross section plot, the reconstructed amplitudes are directly compared to the original data, proving the superiority of our CS-based technique, in precision and in robustness to noise with sparsity being the only assumption, over other extrapolation methods and their underlying assumptions.

B.4. Implementation Considerations and Subwavelength Imaging

Naturally, when trying to recover information from an incomplete set of measurements, one would like to use all the data at hand, without any additional loss of information. When treating true subwavelength optical information, the loss of information occurs due to the decay of

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

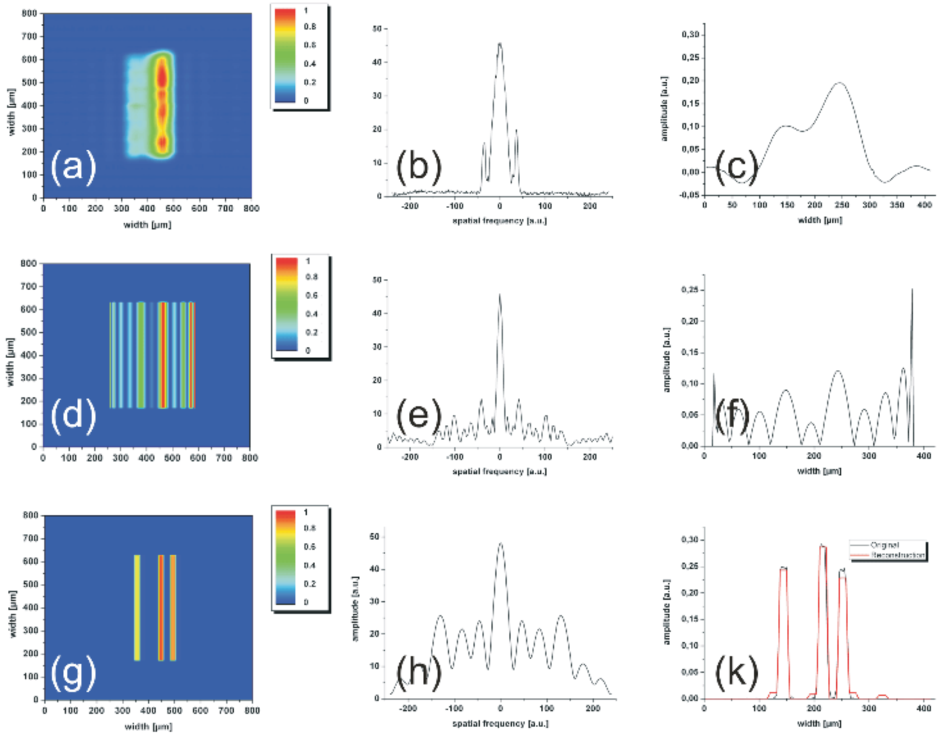


Fig. 10. Comparison between the performance of the Gerchberg-Papoulis extrapolation algorithm and our CS approach. The comparison is made on our experimental data (of Fig.4 of the paper) (a,b,c) The filtered information, blurred to a single stripe (a), its cut Fourier spectrum (b), and a horizontal cross-section of the amplitude, taken through the real-space information (c). (d,e,f) Reconstruction using GP-extrapolation methods yields a distorted recovery with little resemblance to the original data (d) and an incorrect Fourier spectrum (e). The recovery error is most apparent in the horizontal cross section (f). (g,h,k) Reconstruction using CS methods yields a high quality recovered information (g) and its respective Fourier spectrum (h). The strong correspondence between original and recovered image is clearly visible in the horizontal cross section (k).

evanescent waves with spatial frequencies $\nu \geq 1/\lambda$. Any optical element placed at some distance z after the evanescent waves have already decayed ($z > \lambda$) will cause additional loss of information, because all such elements have a finite extent (i.e., their numerical aperture is smaller than unity). Therefore, for true subwavelength imaging, one should perform the measurements at any plane $z > \lambda$, but without passing the waves through any additional lenses. This means that the measurements should be taken as close as possible to the subwavelength object.In that region the low-pass filter $H(\nu)$ also contains phase. The recovery of information would be somewhat more complicated, but not by much. Alternatively, of course, one could use another lens and perform the measurements either in the Fourier plane or at the low-pass-filtered plane, where measuring the data is more convenient. However, the finite numerical aperture of the lens causes further loss of information (low-pass filtering).

The experiments presented in our paper are proof-of-concept, not containing subwavelength information. However, the considerations are identical: it would be best to perform the

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009

measurements immediately after the low-pass filter, and avoid using additional lenses. Unfortunately, in that plane the optical intensity distribution is concentrated in a small region, and taking the data with a camera whose pixels (detectors) are typically 10 microns wide. would cause under-sampling of the data. For this reason, we used another lens, and performed the measurements in the low-pass-filtered image plane (see Fig. 3). Alternatively, we could use a pair of lenses, magnify the Fourier plane information, and take the data where the finest resolution in Fourier plane can be sampled properly by the pixels in the camera. The results after CS reconstruction are practically identical. The experimental results presented in Figs. 4 and 5 in the paper were obtained for measurements taken at the low-pass-filtered image plane. For this case, one can use $|H(\nu)|$ (instead of $H(\nu)$) because the phase is anyway compensated by the imaging system.

Acknowledgments

AS acknowledges support by the Leopoldina - German Academy of Sciences (grant LPDS 2009-13). This work was supported by an Advanced Grant of the European Research Council (ERC)

#120729 - $15.00 USD Received 30 Nov 2009; revised 15 Dec 2009; accepted 15 Dec 2009; published 16 Dec 2009