Vrije Universiteit Brussel

Vrije Universiteit Brussel

Regularized non-convex image reconstruction in digital holographic microscopy

Schretter, Colas; Blinder, David; Bettens, Stijn; Ottevaere, Heidi; Schelkens, Peter

Document Version:
Final published version

Link to publication

Check for updates

Regularized non-convex image reconstruction in digital holographic microscopy

COLAS SCHRETTER,[1,2,*] DAVID BLINDER,[1,2] STIJN BETTENS,[1,2] HEIDI OTTEVAERE,[3] AND PETER SCHELKENS[1,2]

[1]VrijeVrije Universiteit Brussel, Dept. of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium
[2]imec, Kapeldreef 75, B-3001 Leuven, Belgium
[3]Vrije Universiteit Brussel, Dept. of Applied Physics and Photonics, Brussels Photonics Team (B-PHOT), Pleinlaan 2, B-1050 Brussels, Belgium
*cschrett@vub.ac.be

Abstract: Inverse problem approaches for image reconstruction can improve resolution recovery over spatial filtering methods while reducing interference artifacts in digital off-axis holography. Prior works implemented explicit regularization operators in the image space and were only able to match intensity measurements approximatively. As a consequence, convergence to a strictly compatible solution was not possible. In this paper, we replace the non-convex image reconstruction problem for a sequence of surrogate convex problems. An iterative numerical solver is designed using a simple projection operator in the data domain and a Nesterov acceleration of the simultaneous Kaczmarz method. For regularization, the complex-valued object wavefield image is represented in the multiresolution CDF 9/7 wavelet domain and an energy-weighted preconditioning promotes minimum-norm solutions. Experiments demonstrate improved resolution recovery and reduced spurious artifacts in reconstructed images. Furthermore, the method is resilient to additive Gaussian noise and subsampling of intensity measurements.

References and links

1. M. K. Kim, "Principles and techniques of digital holographic microscopy," SPIE Rev. **1**, 8005 (2010).
2. E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of fresnel off-axis holograms," Appl. Opt. **38**, 6994–7001 (1999).
3. N. Pavillon, C. S. Seelamantula, J. Kühn, M. Unser, and C. Depeursinge, "Suppression of the zero-order term in off-axis digital holography through nonlinear filtering," Appl. Opt. **48**, H186–H195 (2009).
4. C. S. Seelamantula, N. Pavillon, C. Depeursinge, and M. Unser, "Exact complex-wave reconstruction in digital holography," J. Opt. Soc. Am. A **28**, 983–992 (2011).
5. S. Sotthivirat and J. A. Fessler, "Reconstruction from digital holograms by statistical methods," Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers **2**, 1928–1932 (2003).
6. S. Sotthivirat and J. A. Fessler, "Penalized-likelihood image reconstruction for digital holography," J. Opt. Soc. Am. A **21**, 737–750 (2004).
7. J. A. Fessler and S. Sotthivirat, "Simplified digital holographic reconstruction using statistical methods," 2004 International Conference on Image Processing (ICIP'04) **4**, 2435–2438 (2004).
8. A. Bourquard, N. Pavillon, E. Bostan, C. Depeursinge, and M. Unser, "A practical inverse-problem approach to digital holographic reconstruction," Opt. Express **21**, 3417–3433 (2013).
9. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Phys. D: Nonlin. Phenom. **60**, 259–268 (1992).
10. Y. Rivenson, A. Stern, and B. Javidi, "Compressive Fresnel holography," J. Displ. Technol. **6**, 506–509 (2010).
11. D. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52**, 1289–1306 (2006).
12. E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," J. Fourier Anal. Appl. **14**, 877–905 (2008).
13. M. M. Marim, M. Atlan, E. Angelini, and J.-C. Olivo-Marin, "Compressed sensing with off-axis frequency-shifting holography," Opt. Lett. **35**, 871–873 (2010).
14. A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," Commun. Pure Appl. Math. **45**, 485–560 (1992).

15. S. Bettens, H. Yan, D. Blinder, H. Ottevaere, C. Schretter, and P. Schelkens, "Studies on the sparsifying operator in compressive digital holography," Opt. Express (2017, submitted).
16. M. Liebling, T. Blu, and M. Unser, "Fresnelets: New multiresolution wavelet bases for digital holography," IEEE Trans. Image Process. **12**, 29–43 (2003).
17. T.-C. Poon, *Digital holography and three-dimensional display: Principles and Applications* (Springer Science & Business Media, 2006).
18. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2004), 3rd ed.
19. S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Anal. Mach. Intel. **11**, 674–693 (1989).
20. A. A. Petrosian and F. G. Meyer, *Wavelets in signal and image analysis: from theory to practice*, vol. 19 (Springer Science & Business Media, 2013).
21. P. Schelkens, A. Skodras, and T. Ebrahimi, *The JPEG 2000 Suite* (Wiley Publishing, 2009).
22. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
23. A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," Ultrason. Imag. **6**, 81–94 (1984).
24. S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen," Bulletin International de l'Académie Polonaise des Sciences et des Lettres **35**, 355–357 (1937).
25. B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," SIAM J. Contr. Optim. **30**, 838–855 (1992).
26. D. Needell, N. Srebro, and R. Ward, "Stochastic gradient descent and the randomized Kaczmarz algorithm," arXiv preprint arXiv:1310.5715 (2013).
27. K. Wei, "Phase retrieval via Kaczmarz methods," arXiv preprint arXiv:1502.01822 (2015).
28. Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$," Doklady an SSSR **269**, 543–547 (1983).
29. Y. Nesterov, "Smooth minimization of non-smooth functions," Math. Program. **103**, 127–152 (2005).
30. D. Kim and J. A. Fessler, "Optimized first-order methods for smooth convex minimization," Math. Program. **159**, 81–107 (2016).
31. C. Schretter and H. Niederreiter, "A direct inversion method for non-uniform quasi-random point sequences," Monte Carlo Meth. Appl. **19**, 1–9 (2013).
32. W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," Appl. Comput. Harmon. Anal. **3**, 186 – 200 (1996).
33. I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," J. Fourier Anal. Appl. **4**, 247–269 (1998).

1. Introduction

Digital holography is a technique to capture simultaneously the amplitude and the phase of an incoming wavefront using a digital image sensor. Various configurations for holographic setups exist [1], such as Gabor holography, Fourier holography, phase-shifting holography and off-axis holography. In off-axis holography, the object wavefront is retrieved by using a high-frequency off-axis carrier wave to separate the desired first order from the unwanted zero order and twin image terms in the Fourier domain.

The off-axis geometry allows users to capture the wavefront with a single recording, unlike *e.g.* in phase-shifting holography, in which stability is crucial for a realization of at least three exposures. Single exposure holography is mandatory when the recorded object is dynamic. However, the drawback is that diffraction orders must be well-separated in the Fourier domain for subsequent filtering. Stringent constraints have to be put on the placement and bandwidth of the first order term. This will cause much of the digital bandwidth to be underused and thereby, discarding information about the object. Advanced reconstruction methods are therefore essential for resolution recovery.

Classical approaches for extracting a complex wavefield from intensity measurements relies on direct manipulations in the Fourier domain [2]. Assuming that the object under investigation is band-limited, the Fourier filtering technique may become optimal using clever non-linear filters that are implemented using a log-transform on the intensity measurements [3, 4]. Unfortunately, strict bandlimitation is not a realistic hypothesis in lens-free setups.

Inverse problem approaches have been successfully applied for resolution recovery in a wide range of modalities. The seminal works in digital holography by Sotthivirat and Fessler [5–7]

propose to reconstruct the complex wavefront using a penalized likelihood objective function with an edge-preserving filter for regularizing the solution image. They solved a surrogate problem using the optimization transfer principle and demonstrated, on horizontally-modulated artificial data, the feasibility of reconstructing complex images from arbitrary reference beams.

Subsequently, the work of Bourquard et al. [8] devised a practical method for the reconstruction of acquired data. They represented the solution complex image with a dense array of pixels and used a dual-loop method where data similarity is ensured up to a given precision, then the solution is regularized in the image space, using a total variation minimization prior [9].

In principle, this method is only able to match the measurements approximately and the interleaving of two separate objective functions in the data and solution spaces leads to slow convergence. Indeed, an exhaustive line search was used for solving the two optimization problems. Nonetheless, they demonstrated encouraging results as the method reduced greatly border artifacts and slightly increased resolution recovery compared to classical direct Fourier filtering. Furthermore, their inverse approach was robust to data undersampling and measurement noise could be mitigated by tuning the tolerance bound in the data fidelity objective.

Reconstruction from undersampled data has been initially investigated by Riverson et al. [10], in the case where complex measurements are recorded, for example through multiple acquisitions with a controlled global phase shift. In this case, the reconstruction problem is convex but underdetermined if fewer measurements than image pixels are recorded. This draws a link with the compressive sensing framework [11, 12] that relies on sparse regularization. Marim et al. [13] have used compressive sensing in the hologram space for explicitly recovering missing values.

In this work, we took a different stance on the inverse problem approach for reconstructing a regularized image from off-axis intensity measurements. We propose a numerical method that is able to converge to a strictly compatible solution to the non-convex image reconstruction problem using a simple consistent projection operator in data domain. We use as solver a fast Nesterov acceleration of the simultaneous Kaczmarz iterative method. No explicit regularization is implemented in image space, but instead we select the minimum-norm representation of the solution in a multiresolution wavelet domain. Experiments conducted on a synthetic resolution chart demonstrate drastic resolution recovery improvements in comparison to the best compromise we can achieve with direct Fourier reconstruction. Moreover, robustness to noise on the intensity measurements can be obtained with a modified projection operator.

We observed that if the detector would be able to measure the complete complex values, *i.e.* both the amplitude and phase information of the incident light, then the inverse reconstruction problem would be convex and could be solved using well established numerical methods. We create a sequence of intermediate surrogate convex problems using inferred complete complex data. This inference step uses simple data-consistency conditions and shifts the complexity away from the numerical solver. Our method is also robust to data undersampling and implicitly interpolates the missing intensity data using both wavelet decompositions and an accurate physically-based forward propagation model using the angular spectrum operator.

For regularization, the real and imaginary parts of the complex sample image to reconstruct are represented in the Cohen-Daubechies-Feauveau (CDF) 9/7 wavelet space [14]. Our prior studies [15] demonstrated that combining CDF wavelets with the angular spectrum light transport operator was a more appropriate choice over Fresnelets [16]. This indirect minimum norm objective in coefficient space yields global multiresolution image regularization that is very different than direct processing in image space.

The remainder of the paper is structured as follows. We first present the geometry of off-axis holography and the image formation model in Section 2. Our regularized inverse reconstruction method is derived in Section 3. Experiments on noisy and undersampled synthetic data are shown in Section 4, as well as results from resolution charts and biological samples. Finally, we conclude with an outreach on future works in Section 5.
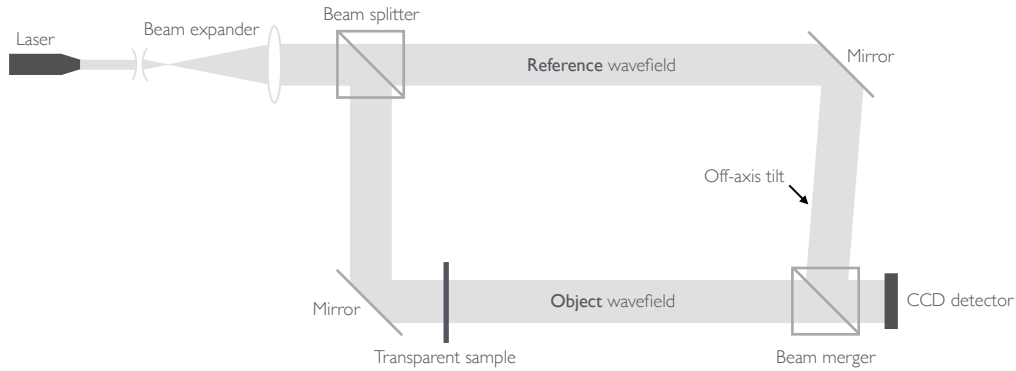
Fig. 1. Geometry of an off-axis holographic acquisition setup. The intensity of the interference between a slightly tilted reference wavefield and the object wavefield is recorded by the digital CCD sensor. The phase information of the observed sample is then encoded in slight local displacements of fringe patterns appearing on the detector plane.

2. Off-axis holography

Following the illustration in Fig. 1, an off-axis intensity hologram is measured on the digital detector as an interference pattern I formed by the superimposition of an object beam O containing the object information and a tilted incident carrier planar wave, called the reference beam R. For more details, the fundamentals of holography have been studied extensively [17, 18].

Let $I \in \mathbb{R}_+^N$ be a rectangular recorded intensity image defined with $N = N_x \times N_y$ pixels that are vectorized in lexicographical order. In off-axis holography, we express intensities as energy measurements of the interference between a reference beam $R \in \mathbb{C}^N$ and the sought object wavefield $O \in \mathbb{C}^N$ illuminating the sample under investigation. Although the waveform is complex-valued, we acquire only an intensity image

$$I = |O + R|^2 \tag{1}$$

$$= |O|^2 + |R|^2 + R \circ O^* + O \circ R^* \tag{2}$$

where $*$ denotes the complex conjugate and $\circ$ is the Hadamard product.

2.1. Image formation

The distance from the detector to the sample is relatively short and therefore in this work, we rely on the angular spectrum method [18] for the light transport operator from the object wavefield $O_d \in \mathbb{C}^N$ at the sample plane towards the detector. The detector plane is located at the origin and the sample is displaced by a distance d along the z-axis. The angular spectrum transforms point sources to apodized chirp functions on the detector plane. The following forward and adjoint backward relations hold for small propagation distances d:

$$O = \Psi_d O_d = F^{-1} (K_d \circ F(O_d)) \quad \text{and} \quad O_d = \Psi_{-d} O = F^{-1} (K_{-d} \circ F(O)) \ . \tag{3}$$

The multiplication before inverse Fourier transform F^{-1} is equivalent to a convolution in the spatial domain. Specifically, the Hadamard product of K_d and the spectrum of O_d, followed by the inverse Fourier transformation, is equivalent to the convolution of O_d with the point spread function $F^{-1}(K_d)$, with the optical transfer function

$$K_d = \frac{1}{K_{-d}} = \exp\left(-2i\pi \frac{d}{\lambda} D\right) \tag{4}$$
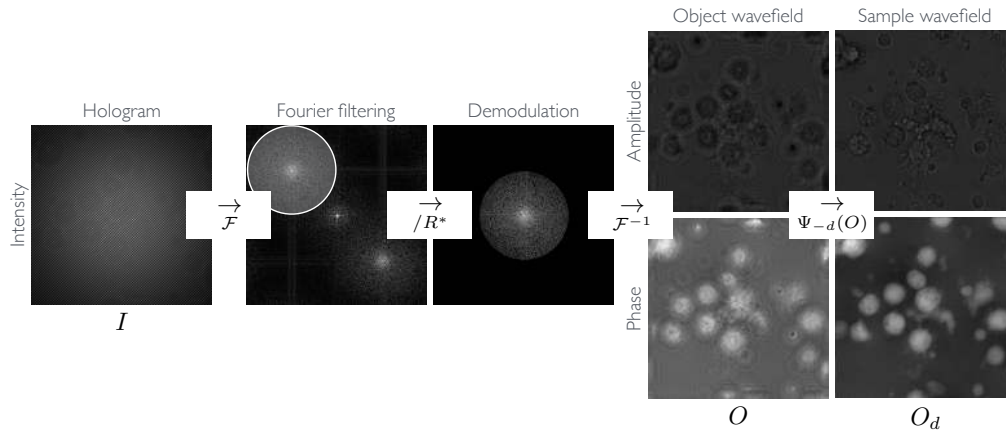
Fig. 2. Direct image reconstruction in the Fourier domain. The complex-valued object wavefield image O is recovered from the intensity hologram $I = |O + R|^2$ by demodulation and band-pass filtering of higher frequencies for removing the fringe modulation pattern of the reference beam R. Finally, the wavefield is refocused on the sample plane.

where the distance field image $D \in \mathbb{R}_+^N$ defined by

$$D(x, y) = \sqrt{1 - \left(\frac{\lambda}{p}\frac{x - N_x/2 - \frac{1}{2}}{N_x}\right)^2 - \left(\frac{\lambda}{p}\frac{y - N_y/2 - \frac{1}{2}}{N_y}\right)^2}, \tag{5}$$

with the wavelength λ of the coherent illumination source, the distance d from the sample plane to the detector, the pixel pitch p and the integer coordinates $x \in [1 \ldots N_x]$ and $y \in [1 \ldots N_y]$.

In off-axis geometry, the reference beam R is a slightly slanted planar wave such that the time delay of arrival on the detector draws a modulating carrier signal that shifts the spectrum of O to obtain $O \circ R^*$. Prior to acquiring I, the reference wavefield R is determined from recording a calibration intensity hologram $I_R \in \mathbb{R}_+^N$ of the interference of R and the blank object wavefield $O_R \in \mathbb{C}^N$ that is is obtained without sample in the field of view. Therefore, both the object beam amplitude attenuations and phase shifts at the sample plane are known to be zero.

The first image in Fig. 2 shows a typical intensity hologram and its characteristic Fourier power spectrum. The vertical and horizontal off-axis angles of incidence have been calibrated such that the modulated object wavefield fills one full quadrant of the Fourier spectrum. This configuration leads to an optimal use of the available image bandwidth.

2.2. Direct Fourier reconstruction

From the recorded intensity hologram I, an approximation of the focused complex object wavefield O_d can be reconstructed through simple manipulations in the Fourier domain. Intensities from I are decomposed into the -1, 0 and +1 diffraction orders: the first two terms in (2) correspond to the zero order, the third term is the so-called twin image and the last term contains the sought object beam O modulated by the complex conjugate of the reference beam R.

Since R is known, and assuming noise-free measurements, the possible solutions for $R + O$ in every pixel lie on a circle in the complex plane. The right subfigure shows the corona of acceptable solutions, within an interval of confidence. The width of that interval depends on the assumed noise variance. As Fourier reconstructions do not allow to model noise, we consider in this section, solely the noise-free case.

First, we remove the non-modulated intensity information from the known reference beam:

$$\tilde{I} = I - |R|^2 \tag{6}$$

$$= |O|^2 + R \circ O^* + O \circ R^*, \tag{7}$$

then, we demodulate the intensity signal by dividing with the complex conjugate of R:

$$\tilde{O} = \frac{\tilde{I}}{R^*} \tag{8}$$

$$= \frac{|O|^2}{R^*} + \frac{R \circ O^*}{R^*} + O \tag{9}$$

and finally, we crop the Fourier spectrum with a smoothed circular cropping disc for explicitly removing the two first demodulated terms and thereby isolating O. This filtering step is crucial for removing interference artifacts from the remaining zero order term $|O|^2/R^*$ and the unwanted twin image depending on O^*.

We define a circularly-symmetric function with its center at the origin of the frequency domain as follow. First, we generate a smoothed circular mask $C_H \in \mathbb{R}^N_{[0,1]}$, starting from a disc indicator function with linear ramp on the edge:

$$C(x, y) = \min\left(1, \max\left(0, r - \sqrt{(x - r - 0.5)^2 + (y - r - 0.5)^2}\right)/w\right), \tag{10}$$

with the quadrant cropping radius $r = N_x/4 - 1$, the width w of the linear ramp at the window border and the pixel coordinates x and y defined above. Finally, we remove discontinuities in the first derivative by applying the Hermite polynomial *"smoothstep"* function:

$$C_H(x, y) = C(x, y)^2 \, (3 - 2 \, C(x, y)). \tag{11}$$

After masking the frequency response for selecting the object wavefield O in the Fourier domain, the result is refocused by reversing the effect of light transport from the detector plane to the sample position. All in one, the final propagated object wavefield is

$$\tilde{O}_d = \mathrm{F}^{-1}\left(K_{-d} \circ (\tilde{O} \circ C_H)\right). \tag{12}$$

where K_{-d} is a diagonal matrix which applies a phase delay based on the propagation distance d.

3. Regularized inverse reconstruction

From the recorded intensity hologram I, an improved approximation of the propagated complex object wavefield O_d can be reconstructed through iterative refinement of a regularized complex solution in the CDF 9/7 wavelet domain. One important property of the sought propagated object beam in sample plane O_d is that not all valid wavefronts O on the detector plane satisfying Eq. 2 are statistically equally likely: neighboring pixels will generally be highly correlated in the focused image O_d. These correlations can be reduced significantly by using a sparsifying multiresolution transform [19]. By doing so, the signal will be well-approximated with only a few coefficients that are selected according to their energy in the decomposition.

In the forward model shown in Fig. 3, we chose a discrete Wavelet transform (DWT) for the image transform. The DWT is a multiscale transform that allows for the simultaneous spatial and frequency analysis of signals [20]. It separates a discrete signal into a low-pass and a high-pass signal by means of a dyadic transform, followed by a downsampling. This operation is applied recursively on the ensuing low-pass signal, resulting in a multiscale representation. For this application, we chose the biorthogonal CDF 9/7 wavelet transform. This wavelet was chosen for
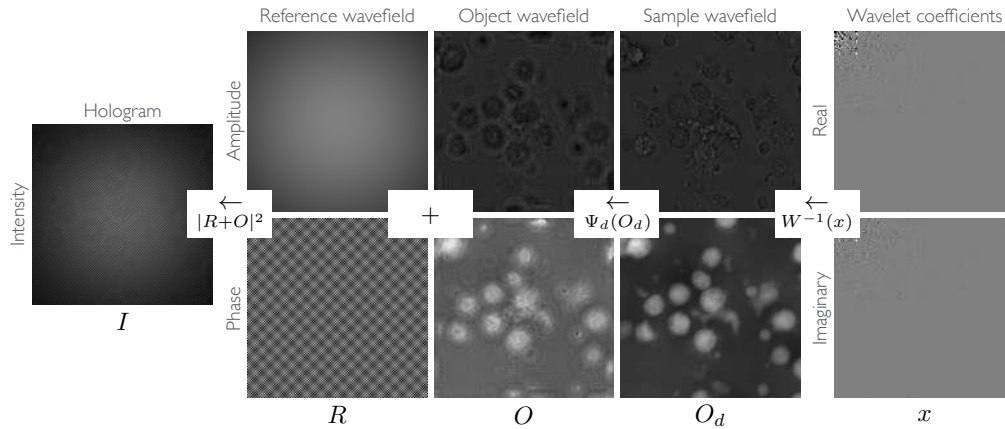
Fig. 3. Forward image formation model used in inverse image reconstruction. CDF 9/7 wavelet coefficients represent the amplitude and phase components of the sample wavefield O_d. Angular spectrum light transport and the interference between the propagated wavefield O and the near-uniform off-axis reference beam R generates a fringe pattern in I.

its symmetry, superior energy compaction and compression performance [21]. Since the CDF wavelets are not orthogonal, the transpose operator will not be equal to the inverse one. For more details on the efficient implementation of the transpose, we refer the reader to Appendix 5.

Reconstructing the two channels of the complex sample wavefield O_d from intensity measurements I only is ill-posed in two ways. First, the problem is *physically ill-posed* since many different sets of complex values for O_d may be compatible with the intensity measurements of I, regardless of the optics and detector definition, *i.e.*, the number of intensity samples. Secondly, the problem may be *algebraically ill-posed* if we record $M < N$ samples. In that case, the number of measurements is smaller than the number of unknowns leading to an underdetemined inverse problem with more solution elements than data constrains.

For the aforementioned reasons, regularization is needed to select a unique solution among all compatible realizations. More importantly, the solution should be plausible as well. We propose in this work to use a minimum-norm regularization prior in the CDF 9/7 wavelet domain for selecting a unique solution among all combinations of amplitude and phase pairs that are compatible with the M recorded samples from the full intensity hologram I.

More formally, let $x \in \mathbb{C}^{N_x \times N_y}$ be the unique minimum-norm solution representing the object wavefield $O_d = W^{-1}x$ where the matrix $W \in \mathbb{C}^{N_x \times N_y}$ is a complete dictionary of complex wavelet basis functions such that the vector-matrix product can be implemented with a fast transform. One could aim first at solving the following non-convex problem:

$$\underset{x}{\operatorname{argmin}} \|x\| \quad \text{subject to} \quad I = |R + \Psi_d W^{-1}x|^2 \,, \tag{13}$$

in which the data compatibility constraint only applies to squared magnitudes from complex values and therefore the phase information of the solution is implicitly lost. There are no globally convergent numerical algorithms for solving this problem and some sort of heuristic search method should be used instead. Since the problem is large, we will pursue along another path.

Let's assume for a moment that the detector could measure both the real and imaginary parts of the complex object wavefield instead of the squared magnitudes only. From these complex-valued measurements of O, we could solve the simpler convex problem

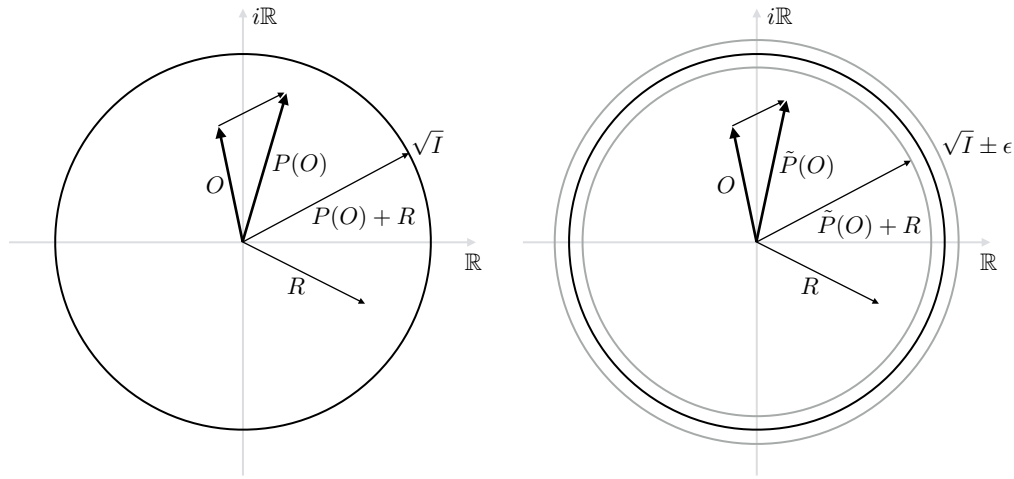$$\underset{x}{\operatorname{argmin}} \|x\| \quad \text{subject to} \quad O = \Psi_d W^{-1}x \,, \tag{14}$$

Fig. 4. Exact (left) and relaxed (right) projection operators infering complex-valued measurements from the current solution O. The closest value from O is selected such that addition to the reference beam R intersects the circular region of all compatible intensity values.

in which the data compatibility constraint applies to both the amplitude and phase information and a convergent convex numerical method may be used to find this unique regularized solution.

3.1. Projection operators

In this work, we define two projection operators that infer new complex measurements O using a pilot estimate of wavelet coefficients x. At each step of a convex iterative solver, a projection will be applied such that data consistency conditions are respected. This procedure is similar in spirit to the Gerchberg-Saxton method for phase retrieval [22]. This method alternately modifies the amplitude of the solution in the spatial and Fourier domain, while we update both channels simultaneously in a non-alternating scheme.

Starting from a zero vector of coefficients $x_0 = 0$, we first infer, at each iteration $t > 0$, the complex-valued measurements from the current object wavefield estimate with

$$O = \Psi_d W^{-1} x_t , \tag{15}$$

then, we use an orthogonal projection for recovering the phase and amplitude information from data consistency constraints.

In off-axis holography, we have intensity measurements that set the squared magnitude of complete complex values. When the measurements are assumed to be noise-free and unbiased, we can plug-in an approximation of the phase information into the current complex data estimate O using the following geometric non-linear data projection operator

$$P(O) = \sqrt{I}\frac{(O + R)}{|O + R|} - R . \tag{16}$$

This exact projection operator is illustrated in the left panel in Fig. 4. The current solution will subsequently be updated using the minimum-norm difference vector such that data consistency is ensured and therefore, we have

$$I = |P(O) + R|^2 . \tag{17}$$

When intensity measurements are corrupted by Gaussian noise or some mismatch with the forward model remains, then a relaxed data consistency condition is more practical. Assuming a

tolerance margin with respect to the expected noise variance ϵ^2, we use

$$
\tilde{P}(O) = \sqrt{I_\epsilon} \frac{(O+R)}{|O+R|} - R \quad \text{with} \quad I_\epsilon = \begin{cases} \max(I - \epsilon^2, 0) & \text{if} \quad |O+R|^2 < \max(I - \epsilon^2, 0) \\ I + \epsilon^2 & \text{if} \quad |O+R|^2 > I + \epsilon^2 \\ |O+R|^2 & \text{otherwise} \end{cases}
$$
(18)

A relaxed version of the projection operator is updating the current solution using the minimum-norm difference vector such that only approximate data consistency is ensured and thus, complex values are restricted to lie inside the corona of radius ϵ^2 around $\sqrt{I}$:

$$
\sqrt{\max(I - \epsilon^2, 0)} \leq |\tilde{P}(O) + R| \leq \sqrt{I + \epsilon^2} \,.
$$
(19)

This relaxed approximate projection operator is illustrated in the right panel in Fig. 4.

For initialization, we assume a blank object wavefield $O = \emptyset$ and therefore we only use the reference beam inside the projection. This gives us the initial complex measurements

$$
P(O) = R \frac{\sqrt{I} - |R|}{|R|} \,.
$$
(20)

The resulting initial wavefield does obviously not correspond to the ground truth image. A particularity is that fringes will appear both in the amplitude and the phase images. Therefore, the wavelet decomposition of this complex image will require lots of high-frequency coefficients to encode the fringes. The energy evaluated as the sum of squared wavelet coefficients will be considerably high. An energy-minimization prior will tend to suppress fringes from the solution, as this information is already present in the reference beam R.

3.2. Iterative solver

From the projected data-compatible complex measurements $y = P(O)$, we seek for the minimum-norm regularized solution x by using the Moore-Penrose pseudo-inverse system matrix

$$
x = (A^* A)^{-1} A^* y \,.
$$
(21)

with the linear system matrix $A = \Psi_d W^{-1}$. In practice, the square Gram matrix $(A^* A)$ is not explicitly computed when inverted, but the minimum-norm solution x is found by an iterative method, given a pair of functions computing the forward operator $y = Ax$ and its Hermitian adjoint $x = A^* y$. The inverse wavelet transform and the conjugate transpose of the inverse wavelet transform are computed with fast lifting schemes that are detailed in Appendix 5.

3.2.1. Simultanous Kaczmarz method

At first, the simultaneous algebraic reconstruction technique (SART) [23] is adapted to our problem. SART is a parallel acceleration of the classical sequential row-action Kaczmarz updates [24] that can be seen as an averaged stochastic gradient algorithm [25, 26]. The convergence is monotonous as the norm of the solution increases from iteration to iteration while the mean squared data residuals decrease accordingly.

Starting from an empty initial image at $t = 0$, the norm of the solution is incrementally increased at each iteration with a simple incremental update rule:

$$
x_{t+1} = x_t + \Delta_t
$$
(22)

and the following simultaneous incremental update

$$
\Delta_t = \frac{2}{M} \sum_{i=1}^{M} \frac{[P(O)]_i - [A]_i x_t}{\|[A]_i\|^2} [A^*]_i
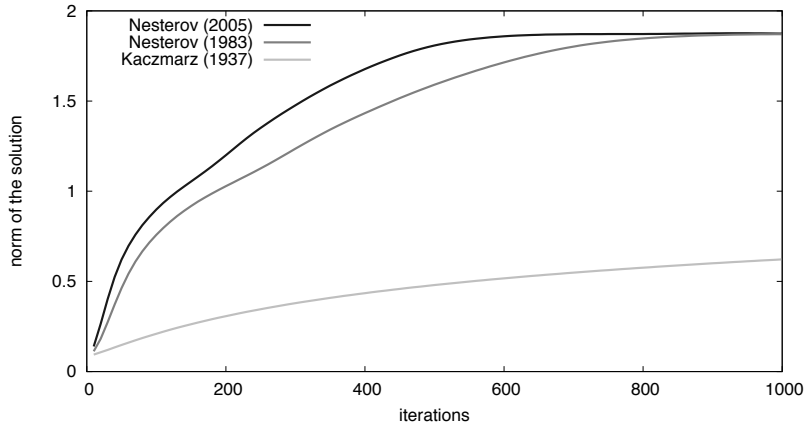$$
(23)

Fig. 5. Comparison of the overall increasing norm as a function of the iteration number for the reference Kaczmarz method and the two Nesterov accelerated methods discussed in this work. Nesterov methods reach quadratic convergence speed instead of the much slower linear convergence of the Kaczmarz method plotted in light gray.

where the sum runs over any subset of measurements $1 \leq M \leq N$. This update is a simultaneous *row-action*, as only the ith row $[A]_i$ is needed for computing the corresponding term in the sum.

The structure of the update contributions inside the sum is simple and interesting: The difference term on the numerator is the discrepancy between the complex measurement $[P(O)]_i$ that is inferred from the current solution and the footprint of the current solution $[A]_i x$, after propagation towards the detector plane. The differences are normalized by the energies $\|[A]_i\|^2$ in data space before backprojection in the wavelet space using the adjoint operator A^*.

The energy normalization values do not depend on the solution nor on data and are precomputed before reconstruction. Note that detector sampling and rows of the system matrix are not explicitly stored; rather, results of multiplications with all matrix row vectors $[A]_i$ and $[A^*]_i$ are implemented using fast transforms. The Kaczmarz method has been recently used for the closely related phase retrieval problem [27]. However, convergence is extremely slow and accelerations are needed in practice, without resorting to hazardous numerical over-relaxation.

3.2.2. Nesterov accelerations

The 1983 vintage of Nesterov accelerations [28] uses the difference with wavelet coefficients at the previous iteration for increasing convergence speed. This variant is proven to be convergent, even though the norm of the solution may decrease after subsequent iterations. This first accelerated gradient update proposed by Nesterov introduces an initial relaxation value $\lambda_0 = 1$ and a temporary solution vector initialized as $y_0 = x_0$. This quantity is incrementally updated as in the Kaczmarz method using

$$y_{t+1} = x_t + \Delta_t , \tag{24}$$

then, the solution is updated using a supplemental additive term as

$$x_{t+1} = y_{t+1} + \frac{\lambda_t - 1}{\lambda_{t+1}}(y_{t+1} - y_t) \quad \text{with} \quad \lambda_{t+1} = \frac{1 + \sqrt{1 + 4\lambda_t^2}}{2} . \tag{25}$$

This second step is akin to updating the solution with a finite difference approximation of the gradient. This update is therefore related to the Newton method and inherits from the quadratic $O(1/t^2)$ convergence speed instead of the much slower linear $O(1/t)$ convergence of

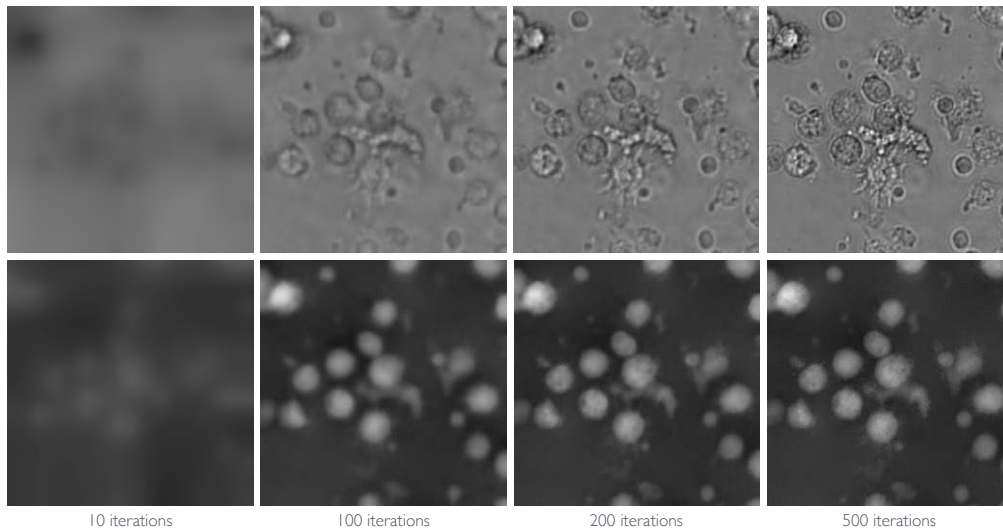| 10 iterations | 100 iterations | 200 iterations | 500 iterations |

Fig. 6. Progressive sharpening of the solution until convergence. The regularization prior in the inverse reconstruction method selects the minimum-norm solution that is compatible with the intensity measurements. Stopping early the iterative process yields a blurry intermediate solution that may already be of sufficient quality for the operator.

the Kaczmarcz method. A remarkable property of the Nesterov acceleration, is that no extra calculations are needed beside a simple difference with the previous solution. At almost the same computational cost, we get a stable method with a guaranteed quadratic convergence regime.

The 2005 version of Nesterov's methods [29] further improves convergence speed by using an accumulated history of all previous incremental updates since the beginning of the iterative process. In addition to the above procedure, we keep track of the accumulated weighted updates:

$$u_{t+1} = u_t + \lambda_t \Delta_t \tag{26}$$

with $u_0 = x_0$. Then, the final update is replaced by the linear combination

$$x_{t+1} = \left(1 - \frac{1}{\lambda_{t+1}}\right) y_{t+1} + \frac{1}{\lambda_{t+1}} u_{t+1} \quad \text{with} \quad \lambda_{t+1} = \frac{1 + \sqrt{1 + 4\lambda_t^2}}{2}. \tag{27}$$

and we iterate until the convergence is satisfactory. Note that the recent work of Kim and Fessler [30] has shown that the increment in (26) may be doubled while preserving convergence. We therefore used that variant in our numerical experiments. Figure 5 compares the evolution of the norm of the solution in function of the iteration number. Nesterov accelerations are essential for reaching reasonable computing performances.

4. Experiments

We implemented the described reconstruction methods in MATLAB. The source code is available upon request to the authors. We conducted experiments on both simulated and optically acquired data of a USAF-1951 resolution chart and two biological samples. The simulated resolution chart used the very same geometry for matching closely the experimental conditions of the acquisition setup. For the biological samples, we started from a real amplitude and phase image of a blood sample acquired on an iLine F holographic microscope from Ovizio Imaging Systems. We then generated artificial intensity holograms using the complex image as the ground truth. We used this realistic phantom for evaluating the impact of noise and subsampling on the image quality.
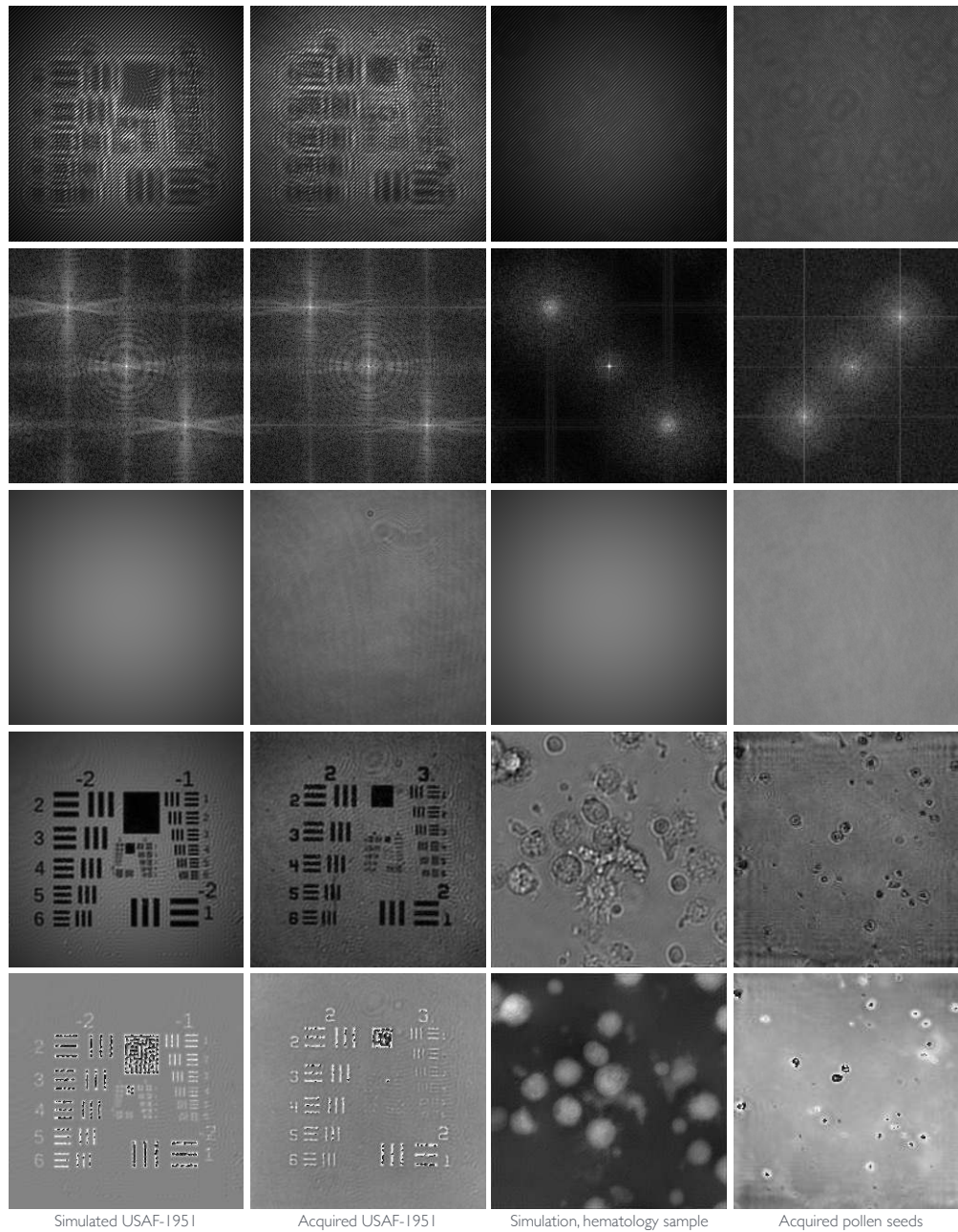
Fig. 7. The four simulated and optically acquired intensity hologram data used in experiments (top row) and their frequency power spectrum (second row). The blank scan for recording the reference beam amplitude are shown as well (third row) along with the amplitude (fourth row) and phase shift (bottom row) reference results with direct Fourier reconstruction.

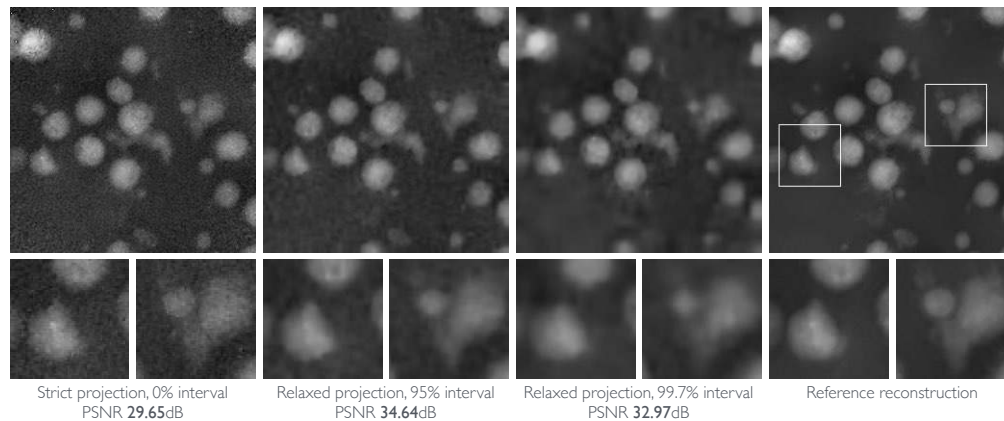| Strict projection, 0% interval | Relaxed projection, 95% interval | Relaxed projection, 99.7% interval | Reference reconstruction |
| PSNR **29.65**dB | PSNR **34.64**dB | PSNR **32.97**dB | |

Fig. 8. Impact of the relaxed data projection operator for avoiding the introduction of noise artifacts into the solution. A bias-variance trade-off is driven by the radius of the intensity tolerance region, expressed as a fraction of the standard deviation of the noise realization.

We also reconstructed an acquisition of pollen seeds from the hologram used in the work of Seelamantula et al. [4]. The optics used for that acquisition used bandlimiting filters, and in this case theoretically exact reconstruction is possible using their specific direct Fourier filtering technique. In practice, slight crosstalk between the zero order terms and the modulated complex wavefield is still present in the data whereas regularization alleviates artifacts appearing in the direct Fourier filtering image reconstructions.

Figure 7 shows the two simulated and acquired intensity holograms and their Fourier power spectra. Additional rows in this array show the amplitude images of the reference beam used for inverse reconstruction. the best result we could obtain using the direct Fourier reconstruction approach is shown as well. For simulations, we implemented a slight Gaussian fall-off intensity profile for modeling the effect of laser beam expansions to an approximate planar wave.

4.1. Data simulations and acquisitions

In our simulations and laboratory experiments on the USAF-1951 test target we built a digital holographic setup in transmission mode as shown in Fig. 1. The sample is illuminated using a JDSU Uniphase 1135P He-Ne laser of 20 mW operating at a wavelength of $\lambda = 632.8$ nm. The planar sample was placed at a distance of 125 mm from the CCD camera: The Ximea MD120MU-SY, which is a 12-bit monochrome camera with a resolution of 4242×2830 pixels and a pixel pitch of $\delta x = 3.1$ μm. The exposure time was optimized for using the full dynamic range of the camera without signal clipping from saturation.

We recorded both the amplitude of the reference beam R and the off-axis hologram formed by the interference pattern. For the phase of the reference beam, a square crop was taken from the camera, followed by a four-fold downsampling to 256×256 pixels. We chose to modulate R as to have the carrier frequency of the first order to lie in the center of a quadrant of the Fourier domain. This configuration optimally uses the detector bandwidth in the off-axis geometry. The exact same geometry and procedure were used for the simulated hematology sample.

4.2. Results

We first evaluated the effective monotonous convergence of the iterative inverse method. As the norm of the solution increases, the complexity of the solution increases as well. From iteration to iteration, we observe a progressive sharpening until convergence. The convergence at early iterations is faster, suggesting that relaxation strategies could possibly be introduced for later

10% quasi-random subsampling
PSNR **33.51**dB

25% quasi-random subsampling
PSNR **37.09**dB

50% quasi-random subsampling
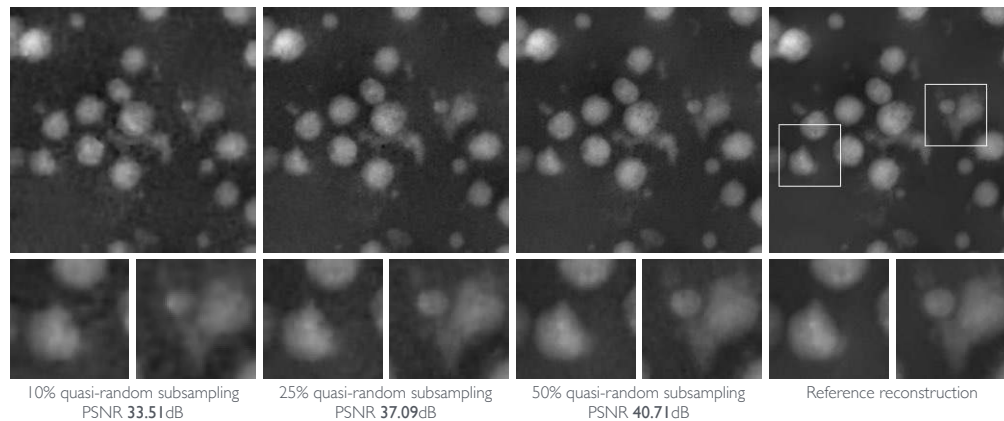PSNR **40.71**dB

Reference reconstruction

Fig. 9. Progressive quality improvement with increasing amount of intensity measurements used for constraining the solution. In comparison to a full-data reference reconstruction, a high quality phase image is recovered from one quarter of all detector pixels only.

iterations. Figure 6 shows the progressive improvement of the solution on the phase image channel. The amplitude image is jointly sharpened in a similar way.

In terms of running time, the process converges to a high-quality solution after about 200 iterations. The required number of iterations depends on the complexity of the image content. For each iteration, the forward model and its adjoint are evaluated. Each of these operations is roughly equivalent to the computing cost of a direct Fourier extraction plus a wavelet transform, hence the required computing budget is about 600 fold in comparison to simple refocusing in Fourier space. A key advantage of the inverse problem approach is its resolution gain with flexible regularization priors, its ability to model noise and missing measurements. The phase shifting pattern also may be arbitrary instead of being restricted to a planar modulation wave. When implemented on GPU, the whole iterative image reconstruction task requires less than a second of computation.

For assessing the robustness to measurement noise, we corrupted measurements with 10% Gaussian noise. A 10% standard deviation from the maximum peak intensity is much higher than what is accustomed in current holographic exposures. The possibility to model noise allows however for new ultra-fast acquisitions protocols, using less sensitive detectors and/or low-light.

Figure 8 shows the impact of the tolerance radius in the relaxed data projection operator in the left part of Fig. 4. We obtained a pleasant visual result for a confidence interval of 95%, corresponding to a radius of 1.96 times the standard deviation of the assumed Gaussian noise realization. Quantitative evaluations in terms of PSNR confirmed the 95% rule for trading-off optimally between blur and variance in the final image.

Our inverse method may use subsets of intensity samples in detector space for updating the complex solution wavefront. We selected quasi-randomly detector pixels using a mask retaining 10%, 25% and 50% of all available data. The low-discrepancy image sampling technique was designed for avoiding regularities that could interfere with the regular fringe pattern and introduce aliasing artifacts in the reconstruction [31].

Figure 9 shows the progressive image improvement as more data are collected. As expected, the data loss translates into loss of resolution that manifests as blur artifacts instead of high frequency artifacts that could be dangerously interpreted as small image features. The norm of the solution is systematically lower as less data are processed.

For illustrating experiments on the simulated and acquired USAF test targets, we extracted crops in the image region showing the boundary of resolution capabilities. Figure 10 shows

Simulated, direct reconstruction Simulated, inverse reconstruction Acquired, direct reconstruction Acquired, inverse reconstruction
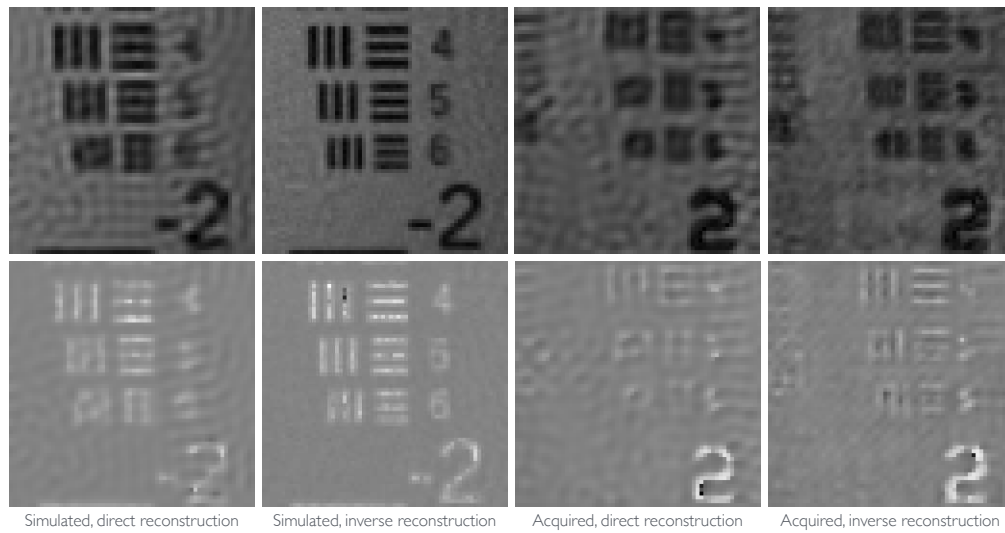
Fig. 10. Side-by-side comparison of resolution recovery in a similar region of interest in both the simulated and acquired USAF-1951 test targets. Parallel bars are crisper with the inverse reconstruction technique, while spurious low-frequency background noise is reduced.

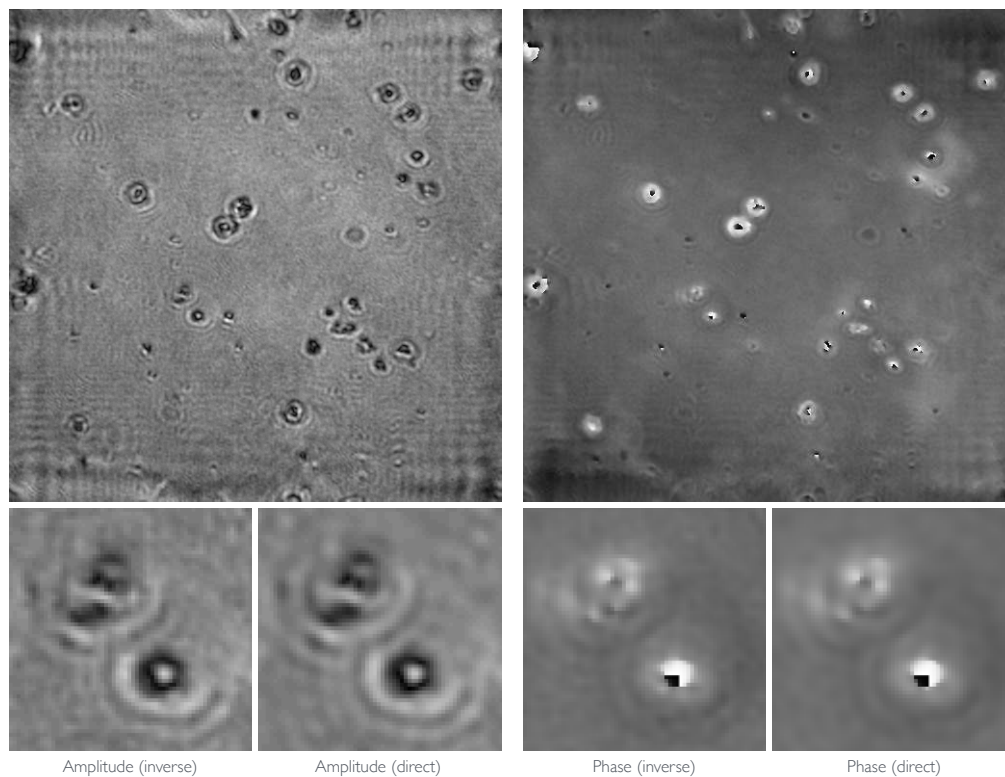Amplitude (inverse) Amplitude (direct) Phase (inverse) Phase (direct)

Fig. 11. Amplitude and phase shift images recovered by iterative reconstruction (top row). A side-by-side comparison in a region of interest (bottom row) shows a slightly crisper recovery in comparison to a Fourier reconstruction. Note that in this case, a bandlimiting filter was used in the optical line for optimizing quality of the direct reconstruction solution.

the side-by-side comparison between the best results we could obtain using direct Fourier reconstruction and the regularized inverse reconstruction method. On simulation, the effective resolution gain was more prominent than on the optical acquisition.

However, we observed a slight discrepancy between the acquisition of the reference beam intensity image and the hologram intensities. This issue has been encountered as well in the work of Bourquard et al. [8] and we used a first order gain correction factor as described in Appendix 5. The discrepancy may be explained by detector non-linearities and possible optical aberrations. We expect further gain with a more precise modeling of the detector response.

Finally, we reconstructed the pollen seeds sample images that were acquired in a setup with a bandlimited numerical aperture and magnification lenses. Figure 11 shows the whole image result with the iterative approach and makes the comparison with the direct Fourier method in regions of interest. The effect of non-exact lens rectification can be observed by the wave patterns at the border of the image domain. We do not observe any discrepancy between the acquisition of the blank scan for the reference beam intensity and the actual hologram, therefore no adaptive gain correction was used for this reconstruction.

The resolution improvement is more noticeable in the amplitude image, suggesting that the surface of pollens is relatively smooth. Note that the optical bandlimitation may also be responsible for limited resolution gain in the phase image. Moreover, not any parameter tweaking was needed for using the reconstruction method on the two different acquisition setups. The sole regularization through norm minimization is generic enough for selecting a plausible single complex solution image among all possible object beams that are compatible with measurements.

5. Conclusion

With transmissive off-axis holographic acquisition setups, it is possible to capture both the amplitude and phase information of light that is encoded implicitly in fringe interferences of a known modulating reference beam and the sought object wavefield. This work presented a convergent iterative method for image reconstruction of the complex wavefront from intensity holograms only. The reconstruction problem is ill-posed and a minimum-norm regularization prior selects a low-energy unique solution. Unfortunately, this objective function is non-convex and we solve instead a sequence of surrogate convex linear problems using a data projection operator at each iteration. The projection steps infer complex measurements that are consistent with recorded intensities. The method progressively sharpens the reconstructed image and does not require any tunable parameter except for the number of iterations. Arbitrary reference beam may be used beside off-axis modulation. Furthermore complete measurement data are not required and sparse sampling may be applied for reducing input data. Experiments demonstrate improvements in terms of resolution recovery, compared to direct Fourier reconstruction.

Appendix A. Lifting schemes for the CDF 9/7 wavelet transforms and their adjoints

The DWT may be implemented as a pair of convolution filters, followed by downsampling operations. However, it is more practical to use a lifting scheme representation [32]. Lifting schemes are generic methods for constructing "second-generation" wavelets, and can be seen as a special case of a filter bank. They have more flexibility, as they allow to define wavelet bases on intervals, irregularly sampled grids and even non-linear transforms. Additionally, lifting steps have reduced computational complexity and memory constraints w.r.t. the convolutional approach. The generic lifting scheme is summarized on Fig. 12.

Lifting methods can be explained by polyphase matrix representations. The z-transform of a signal is defined as
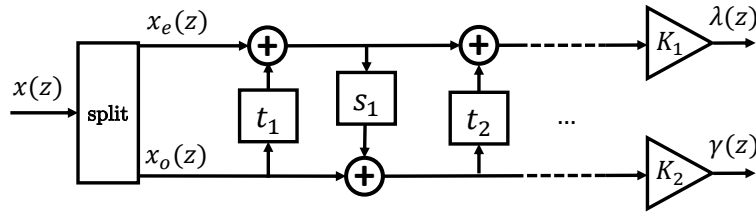
$$x(z) = \sum_k x_k z^{-k} \, , \tag{28}$$

Fig. 12. Lifting scheme block diagram where an input signal x is split into even and odd samples (x_e and x_o). Then a series of convolution-accumulate operations is applied alternately on the two divided signals, using prediction (s_i) and update (t_i) filters. Finally, the channels are scaled with constants K_i, resulting in the final approximation signal λ and detail signal γ.

thus, the z-transform of a finite impulse response filter $h = \{h_{k_b}, ..., h_{k_e}\}$ where $h_{k_b}, h_{k_e} \neq 0$ is

$$h(z) = \sum_{k=k_b}^{k_e} h_k z^{-k} . \tag{29}$$

Filtering a signal x by a filter h is simply described in the z-domain as an ordinary multiplication: $y(z) = h(z)x(z)$. Using this notation we can decompose a signal $x(z)$ into even ($x_e(z)$) and odd ($x_o(z)$) components as $x(z) = x_e(z^2) + z^{-1}x_o(z^2)$, where

$$x_e(z^2) = \frac{x(z) + x(-z)}{2} \quad \text{and} \quad x_o(z^2) = \frac{z}{2}[x(z) - x(-z)]. \tag{30}$$

After splitting the signal into odd and even parts, the signal can be transformed by a filter pair (h, g) represented by the polyphase matrix $P(z)$:

$$\begin{pmatrix} \lambda(z) \\ \gamma(z) \end{pmatrix} = P(z) \begin{pmatrix} x_e(z) \\ x_o(z) \end{pmatrix} . \tag{31}$$

Any polyphase matrix $P(z)$ representing a wavelet transform with finite impulse response filters can be factored into a product of a diagonal matrix and multiple unit upper and lower triangular matrices [33], matching the primal and dual lifting steps:

$$P(z) = \begin{pmatrix} h_e(z) & h_o(z) \\ g_e(z) & g_o(z) \end{pmatrix} = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} \prod_{i=1}^{m} \begin{pmatrix} 1 & s_i(z) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t_i(z) & 1 \end{pmatrix} . \tag{32}$$

Furthermore, we can directly compute the inverse transform as

$$P(z)^{-1} = \left(\prod_{i=m}^{1} \begin{pmatrix} 1 & 0 \\ -t_i(z) & 1 \end{pmatrix} \begin{pmatrix} 1 & -s_i(z) \\ 0 & 1 \end{pmatrix} \right) \begin{pmatrix} 1/K_1 & 0 \\ 0 & 1/K_2 \end{pmatrix} . \tag{33}$$

For efficiently solving the gradient descent problem using Nesterov's method, we not only need the inverse transform but its transpose as well. The inverse transpose transform will only be equal to the forward transform when the DWT is orthogonal. Using the formalism of the lifting scheme we can construct the transpose transform of any non-orthogonal wavelet, or more generally any linear lifting-based filter bank, as follows:

$$(P(z)^{-1})^\top = \begin{pmatrix} 1/K_1 & 0 \\ 0 & 1/K_2 \end{pmatrix} \prod_{i=1}^{m} \begin{pmatrix} 1 & 0 \\ -s_i(z) & 1 \end{pmatrix} \begin{pmatrix} 1 & -t_i(z) \\ 0 & 1 \end{pmatrix} . \tag{34}$$

In particular, we have implemented our system using the CDF 9/7 wavelet transform. The specific prediction, update and scaling constants are given in Table 1.

Table 1. Numerical values of the prediction and update filters for the CDF 9/7 DWT.

Filter	Value
$t_1(z)$	$-0.052980118572961 \, (1 + z^{-1})$
$s_1(z)$	$-1.586134342059924 \, (1 + z)$
$t_2(z)$	$+0.443506852043971 \, (1 + z^{-1})$
$s_2(z)$	$+0.882911075530934 \, (1 + z)$
$K_1 = 1/K_2$	$+1.230174104914001$
$K_2 = 1/K_1$	$+0.812893066115961$

Appendix B. Gain estimation of the reference wavefield

For acquired intensity holograms, it is essential that the overall brightness remains identical for both the recording of the reference beam amplitude as well as the generated interference of the reference beam and the object wavefield. In our setup, we often observe a slight discrepancy in the intensity scales: A practical issue that was also pointed out in the work of Bourquard et al. [8]. Instead of a spatially variant approach, we used a single global gain calibration factor that was estimated by equaling the mean intensity of the input hologram with the predicted intensities using the current solution. Formally, we computed the gain

$$\alpha = \sqrt{\frac{\sum_{i=1}^{N} I_i}{\sum_{i=1}^{N} |R_i + [\Psi_d W^{-1} x]_i|^2}} \approx 1 \,, \tag{35}$$

and the calibrated data projection in (16) is generalized by

$$P(O) = \sqrt{I} \frac{(O + \alpha R)}{|O + \alpha R|} - \alpha R \,. \tag{36}$$

The multiplicative gain correction factor α is estimated at each iteration and its value gradually decreases until convergence. Since the mean estimates may be statistically significant with fewer samples than pixels in the recorded hologram, one could evaluate the sum over a much smaller random subset of pixels for decreasing the computational run-time. In this work, we used the gain calibration only for the reconstructions from the acquired USAF 1951 resolution chart.