

## Convex optimization-based blind deconvolution for images taken with coherent illumination

Doelman, Reinier; Verhaegen, Michel

**DOI**

[10.1364/JOSAA.36.000678](https://doi.org/10.1364/JOSAA.36.000678)

**Publication date**

2019

**Document Version**

Final published version

**Published in**

Journal of the Optical Society of America A: Optics and Image Science, and Vision

**Citation (APA)**

Doelman, R., & Verhaegen, M. (2019). Convex optimization-based blind deconvolution for images taken with coherent illumination. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 36(4), 678-685. <https://doi.org/10.1364/JOSAA.36.000678>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Convex optimization-based blind deconvolution for images taken with coherent illumination

REINIER DOELMAN\*  AND MICHEL VERHAEGEN 

Delft Center for Systems and Control, Delft University of Technology, The Netherlands

\*Corresponding author: r.doelman@tudelft.nl

Received 7 December 2018; revised 28 February 2019; accepted 4 March 2019; posted 4 March 2019 (Doc. ID 354788); published 28 March 2019

**A rank-constrained reformulation of the blind deconvolution problem on images taken with coherent illumination is proposed. Since in the reformulation the rank constraint is imposed on a matrix that is affine in the decision variables, we propose a novel convex heuristic for the blind deconvolution problem. The proposed heuristic allows for easy incorporation of prior information on the decision variables and the use of the phase diversity concept. The convex optimization problem can be iteratively re-parameterized to obtain better estimates. The proposed methods are demonstrated on numerically illustrative examples. © 2019 Optical Society of America**

<https://doi.org/10.1364/JOSAA.36.000678>

Provided under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. INTRODUCTION

In application areas such as coherent diffraction imaging (CDI) [1], long-range horizontal imaging [2], imaging of layered metamaterials [3], or ptychography [4], the image formation process can be described using the expressions for imaging with coherent illumination [5]

$$\mathbf{y} = |\mathbf{g}_o \star \mathbf{h}|^2, \quad (1)$$

where  $\mathbf{y}$  denotes the (noiseless) measurement (recorded discretized image), and  $|\cdot|^2$  denotes the element-wise squared absolute value of the complex-valued argument, in this case the complex field in the imaging plane.  $\mathbf{g}_o$  is the object-plane complex amplitude, and  $\mathbf{h}$  denotes the amplitude impulse response.  $\star$  is the (discrete) convolution operator. The amplitude impulse response is sometimes called the *coherent point spread function* (coherent PSF). In ptychography, for example, the quantities of interest are the Fourier transforms of  $\mathbf{g}_o$  or  $\mathbf{h}$ .

If either  $\mathbf{g}_o$  or  $\mathbf{h}$  is known, and the other is to be estimated based on the measurements  $\mathbf{y}$ , then this problem is called a *phase retrieval problem*. If both quantities are to be estimated based on  $\mathbf{y}$ , then the problem is called a *blind deconvolution problem*. The method proposed in this paper can be seen as an extension of a method we proposed in [6] for phase retrieval, where that algorithm is compared to other standard phase retrieval methods.

In this paper we consider the blind deconvolution problem for images taken with coherent illumination, that is,

$$\begin{aligned} \text{find} \quad & \mathbf{g}_o, \mathbf{h} \\ \text{subject to} \quad & \mathbf{y} = |\mathbf{g}_o \star \mathbf{h}|^2, \\ & \mathbf{g}_o \in \mathcal{K}_{\mathbf{g}_o}, \mathbf{h} \in \mathcal{K}_{\mathbf{h}}, \end{aligned} \quad (2)$$

where  $\mathcal{K}$  denotes a set of (convex) constraints on the variables that encode the available prior information.

This blind deconvolution problem is different from what is typically encountered in the literature, the blind deconvolution problem for images taken with *incoherent* illumination [5]

$$\begin{aligned} \text{find} \quad & \mathbf{f}, \mathbf{s} \\ \text{subject to} \quad & \mathbf{y} = \mathbf{f} \star \mathbf{s}, \\ & \mathbf{f} \in \mathcal{K}_{\mathbf{f}}, \mathbf{s} \in \mathcal{K}_{\mathbf{s}}, \end{aligned} \quad (3)$$

where  $\mathbf{f} = |\mathbf{g}_o|^2$  is the (real and positive valued) intensity of the object in the object plane, and  $\mathbf{s} = |\mathbf{h}|^2$  is the *intensity impulse response*, more often called the point spread function. For the incoherent illumination case, there are several categories of blind deconvolution methods in the literature. Classic iterative projection methods [7–11] use alternating projections of the estimates and their Fourier transforms on their respective constraints in the constraint sets. A second group is that of (non-convex), gradient-based optimization approaches [12], including Bayesian estimation approaches [13–15]. A downside of gradient-based approaches is that the initial guess is often crucial for performance. Recently, a third group of algorithms is being developed based on convex optimization of a “lifted” problem [16–19]. The “lifting” of the problem hinders the use of *phase diversity* [20], a powerful type of prior information, which can be described as a linear constraint on  $\mathbf{h}$  for different images with different phase diversities.

For the coherent illumination blind deconvolution problem, we can make the same classifications.

In the first category, there is [21,22], where the *extended Ptychographical Iterative Engine* (ePIE) is proposed, an iterative transform algorithm for ptychography. Other iterative Fourier transform-based techniques are [23–29].

In the second category, there are the methods proposed in [30,31] for the estimation of wavefront errors in CDI and in ptychography [32–34]. Reference [35] compares the performance of several gradient descent schemes showing superior robustness to noise for amplitude-based metrics. Refinement of a guessed object and wavefront aberration in a maximum likelihood context can be found in [36]. Related to this, gradient-descent schemes are also popular in ptychography for compensating for positioning errors [37].

A convex optimization-based approach has, to the best of our knowledge, not been applied to the coherent blind deconvolution problem in the literature. For example, in ptychography, [38] only solves the deconvolution problem, not the blind deconvolution problem, using convex optimization-based heuristic methods.

In this paper we propose a blind deconvolution method for the coherent illumination case based on a rank-constrained reformulation of Eq. (2). The reformulation is such that the use of multiple images and phase diversity is easily incorporated into the reformulation and subsequent optimization problem. To attempt to find a solution with rank constraints satisfied, we propose to use the nuclear norm as a convex heuristic for the rank constraint. An iterative extension of the subsequent convex optimization problem is proposed to possibly improve the convex heuristic approximation. This iterative extension has been shown in the validation studies to improve the results. To anticipate the problem that the convex optimization problem results in an unsatisfactory solution, we propose an iterative scheme of convex optimization problems that produces, in our experience, iteratively improved results.

The organization of this paper is as follows. In Section 2 we formulate the blind deconvolution problem as a problem to estimate a complex-valued object and the affinely parameterized pupil function of the optical system with unknown phase aberration. Section 3 explains how to reformulate the blind deconvolution into a rank-constraint problem with constraints on matrices affinely parameterized in the object and amplitude impulse response. Section 3.D describes the convex heuristic for the problem, and Section 3.F describes how to incorporate several types of prior information. In Section 4 we demonstrate the algorithm on an illustrative numerical example and compare our method to a gradient descent scheme.

## A. Notation

The operation  $x = \text{vect}(X)$  stacks the columns from the left to right of matrix  $X$  on top of each other to obtain the vector  $x$ .  $\otimes$  denotes the Kronecker product.  $I_n$  denotes an  $n \times n$  identity matrix.  $X = \text{d}(x)$  is the diagonal matrix with the values of the vector  $x$  on its diagonal. The Hermitian transpose of  $X$  is denoted by  $X^H$ . The nuclear norm is denoted as  $\|X\|_*$ , and the Frobenius norm as  $\|X\|_F$ .

## 2. PROBLEM DESCRIPTION

The generalized pupil function (GPF) characterizing an optical system [5] is the complex-valued function

$$P(\rho, \theta) = A(\rho, \theta) \exp(j\phi(\rho, \theta)), \quad (4)$$

where  $(\rho, \theta)$  are the radius and angle of the polar coordinates, respectively.  $|\rho| \leq 1$  and  $\theta \in [0, 2\pi)$ .  $A(\rho, \theta)$  is the amplitude apodization function, and  $\phi(\rho, \theta) \in \mathbb{R}$  is the phase aberration function of the optical system.

To obtain more measurements of the same object  $\mathbf{g}_o$  with different PSFs, a phase diversity  $\phi_d$  may be introduced into the system by means of, for example, a deformable mirror. The GPF then becomes

$$P_d(\rho, \theta) = A(\rho, \theta) \exp(j\phi(\rho, \theta)) \exp(j\phi_d(\rho, \theta)). \quad (5)$$

In this paper we consider the problem in modal form: we assume that the GPF can be well-approximated with a weighted sum of basis functions. We use real-valued radial basis functions and complex coefficients to approximate the GPF [39]. Switching from polar coordinates  $(\rho, \theta)$  to Cartesian coordinates  $(x, y)$ , the radial basis functions and approximate GPF are given by

$$G_i = \chi(x, y) \exp(-\lambda_i((x - x_i)^2 + (y - y_i)^2)),$$

$$P(x, y) \approx \tilde{P}(x, y, \mathbf{v}) = \sum_{i=1}^s \mathbf{v}_i G_i(x, y), \quad (6)$$

with  $(x_i, y_i)$  being the centers of the radial basis functions and  $\mathbf{v}_i \in \mathbb{C}$ .  $\chi(x, y)$  is the aperture support function,  $\lambda_i$  is the spread of the radial basis function, and  $\mathbf{v} \in \mathbb{C}^s$  is the complex-valued vector of coefficients  $\mathbf{v}_i$ . Including an introduced diversity  $\phi_d(x, y)$ , the approximate pupil function reads

$$\tilde{P}_d(x, y, \mathbf{v}) = \sum_{i=1}^s \mathbf{v}_i G_i(x, y) \exp(j\phi_d(x, y)). \quad (7)$$

The amplitude impulse response, also called the *coherent point spread function* (coherent PSF),  $\mathbf{h}_d(u, v)$  is the (two-dimensional) inverse Fourier transform of the GPF

$$\mathbf{h}_d(u, v) = \sum_{i=1}^s \mathbf{v}_i \mathcal{F}^{-1}\{G_i(x, y) \exp(j\phi_d(x, y))\}$$

$$= \sum_{i=1}^s \mathbf{v}_i \mathbf{B}_{d,i}(u, v). \quad (8)$$

Here the coordinates  $(u, v)$  are the Cartesian coordinates in the image plane of the optical system.

A complex amplitude in the object plane  $\mathbf{g}_o$ , imaged through this optical system, in the case of coherent illumination, gives the complex amplitude  $\mathbf{g}_i$  in the image plane [5] as follows:

$$\mathbf{g}_i(u, v) = \mathbf{g}_o(u, v) \star \mathbf{h}_d(u, v). \quad (9)$$

In the noise-free case, the intensity of the complex field  $\mathbf{g}_i$  is recorded to produce the measurements  $\mathbf{y}$  as follows:

$$\mathbf{y}(u, v) = |\mathbf{g}_i(u, v)|^2. \quad (10)$$

We now drop the notation for the dependency on coordinates and assume the signals  $\mathbf{y}$ ,  $\mathbf{g}_i$ ,  $\mathbf{g}_o$ , and  $\mathbf{h}_d$  are sampled on square grids of sizes  $r \times t$ ,  $r \times t$ ,  $m \times n$ , and  $p \times q$ , respectively, such that we obtain matrices of the corresponding size. Throughout this paper we assume that the edges of  $\mathbf{g}_o$  and  $\mathbf{h}_d$  are zero padded, which for the discrete two-dimensional convolution results in the relation  $r = m + p - 1$ ,  $t = n + q - 1$ .

With a slight abuse of notation, the blind deconvolution problem in Eq. (2) has now turned into the problem to identify  $\mathbf{g}_o$  and  $\mathbf{v}$  from measurements  $\mathbf{y}$  as follows:

$$\begin{aligned}
&\text{find} \quad \mathbf{g}_o, \mathbf{v}, \mathbf{h}, \mathbf{g}_i \\
&\text{subject to} \quad \mathbf{y} = |\mathbf{g}_i|^2, \\
&\quad \mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}, \quad \mathbf{h} = \mathbf{B}\mathbf{v}, \\
&\quad \mathbf{g}_o \in \mathcal{K}_{\mathbf{g}_o}, \quad \mathbf{h} \in \mathcal{K}_{\mathbf{h}}.
\end{aligned} \tag{11}$$

Here the constraints  $\mathbf{y} = |\mathbf{g}_i|^2$  and  $\mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}$  are non-convex constraints. The prior information in  $\mathcal{K}_{\mathbf{g}_o}$  and  $\mathcal{K}_{\mathbf{h}}$  is assumed to be a convex constraint on the decision variables.

### 3. BLIND DECONVOLUTION AS A RANK-CONSTRAINED FEASIBILITY PROBLEM

The aim of this section is to rewrite Eq. (11) into a feasibility problem with rank constraints: one rank constraint to replace  $\mathbf{y} = |\mathbf{g}_i|^2$ , and one rank constraint to replace  $\mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}$ .

In the following two subsections, we use the following lemma:

**Lemma 1** [40] *Define the matrix*

$$\begin{aligned}
&M(C, A, B, Q, X, Y, W_1, W_2) \\
&= \begin{pmatrix} W_1 & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} C + AQY + XQB + XQY & (A + X)Q \\ Q(B + Y) & Q \end{pmatrix} \\
&\quad \times \begin{pmatrix} W_2 & 0 \\ 0 & I \end{pmatrix},
\end{aligned} \tag{12}$$

where  $I$  is the identity matrix.

For any  $X$  of the same size as  $A$ , for any  $Y$  of the same size as  $B$ , for any invertible matrices  $W_1, W_2$  of a size corresponding to the sizes of matrix  $C$ , and for nonzero  $Q$ , it holds that the equality

$$\text{rank}(M(C, A, B, Q, X, Y, W_1, W_2)) = \text{rank}(Q)$$

is equivalent to the equality

$$C = AQB. \tag{13}$$

Note that variables  $A$  and  $B$  appear in a product in Eq. (13), but they do not appear in a product in the matrix  $M$  in Eq. (12).

#### A. Convolution Constraint $\mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}$

The two-dimensional (discrete) convolution of  $\mathbf{g}_o$  and  $\mathbf{B}\mathbf{v}$  gives the matrix  $\mathbf{g}_i$ . The elements of the matrix  $\mathbf{g}_i$  are given by the summation of products of individual elements of  $\mathbf{g}_o$  with individual elements of  $\mathbf{B}\mathbf{v}$ . Lemma 2 states how this can be cast into a bilinear matrix equality as follows:

**Lemma 2** *The constraint  $\mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}$  is equivalent to the bilinear equality*

$$\text{vect}(\mathbf{g}_i) = (\text{vect}(\mathbf{g}_o)^T \otimes I_{rt}) V \text{vect}(\mathbf{B}\mathbf{v}) \tag{14}$$

for a matrix of zeros and ones  $V \in \mathbb{B}^{mrt \times pq}$ .

*Proof.* See Appendix A.

In Eq. (14), the general bilinear form  $C = AQB$  of Lemma 1 shows through, with

$$\begin{aligned}
C &= \text{vect}(\mathbf{g}_i), \quad A = (\text{vect}(\mathbf{g}_o)^T \otimes I_{rt}), \\
Q &= V, \quad B = \text{vect}(\mathbf{B}\mathbf{v}).
\end{aligned} \tag{15}$$

We can therefore replace the constraint  $\mathbf{g}_i = \mathbf{g}_o \star \mathbf{h}$  with the rank constraint

$$\begin{aligned}
&\text{rank}(M(\text{vect}(\mathbf{g}_i), \text{vect}(\mathbf{g}_o)^T \otimes I_{rt}, \text{vect}(\mathbf{B}\mathbf{v}), V, X, Y, W_1, W_2)) \\
&= \text{rank}(V).
\end{aligned} \tag{16}$$

The matrices  $X, Y, W_1$ , and  $W_2$  are here further specified to

$$\begin{aligned}
X &= -\text{vect}(\hat{\mathbf{g}}_o)^T \otimes I_{rt}, \quad Y = -\text{vect}(\mathbf{B}\hat{\mathbf{v}}), \\
W_1 &= I, \quad W_2 = I,
\end{aligned} \tag{17}$$

where  $\hat{\mathbf{g}}_o$  and  $\hat{\mathbf{v}}$  are—for the moment—some guesses for  $\mathbf{g}_o$  and  $\mathbf{v}$ , respectively. The expression for the matrix-valued function  $M$  we now abbreviate for notational convenience and call this specific abbreviation  $M_c$ :

$$\begin{aligned}
&M_c(\mathbf{g}_i, \mathbf{g}_o, \mathbf{v}, V, \hat{\mathbf{g}}_o, \hat{\mathbf{v}}) \\
&= M(\text{vect}(\mathbf{g}_i), \text{vect}(\mathbf{g}_o)^T \otimes I_{rt}, \text{vect}(\mathbf{B}\mathbf{v}), V, X, Y, W_1, W_2),
\end{aligned} \tag{18}$$

where  $M_c \in \mathbb{C}^{(rt+mrt) \times (1+pq)}$ .

#### B. Measurement Constraint $\mathbf{y} = |\mathbf{g}_i|^2$

The treatment of the measurement constraints is similar to that in [6]. The constraint  $\mathbf{y} = |\mathbf{g}_i|^2$  uses the element-wise operator  $|\cdot|$ . To obtain the relation between  $\mathbf{y}$  and  $\mathbf{g}_i$  in matrix format, we place the values on a matrix diagonal as follows:

$$\mathbf{y} = |\mathbf{g}_i|^2 \Leftrightarrow \text{d}(\text{vect}(\mathbf{y})) = \text{d}(\text{vect}(\mathbf{g}_i))^H \text{d}(\text{vect}(\mathbf{g}_i)). \tag{19}$$

The related rank constraint is

$$\begin{aligned}
&\text{rank}(M(\text{d}(\text{vect}(\mathbf{y})), \text{d}(\text{vect}(\mathbf{g}_i))^H, \\
&\quad \text{d}(\text{vect}(\mathbf{g}_i)), I_{rt}, X, Y, W_1, W_2)) = rt.
\end{aligned} \tag{20}$$

We further specify here that

$$\begin{aligned}
X &= -\text{d}(\text{vect}(\hat{\mathbf{g}}_i))^H, \quad Y = -\text{d}(\text{vect}(\hat{\mathbf{g}}_i)), \\
W_1 &= I, \quad W_2 = I,
\end{aligned} \tag{21}$$

where  $\hat{\mathbf{g}}_i$  is a guess for  $\mathbf{g}_i$ . Furthermore, we abbreviate the arguments of  $M$  as

$$\begin{aligned}
&M_m(\mathbf{y}, \mathbf{g}_i, \hat{\mathbf{g}}_i) = M(\text{d}(\text{vect}(\mathbf{y})), \text{d}(\text{vect}(\mathbf{g}_i))^H, \\
&\quad \text{d}(\text{vect}(\mathbf{g}_i)), I_{rt}, X, Y, W_1, W_2),
\end{aligned} \tag{22}$$

where  $M_m \in \mathbb{C}^{2rt \times 2rt}$ .

#### C. Rank-Constrained Blind Deconvolution Problem

Using Eqs. (16) and (20), the blind deconvolution problem in Eq. (11) can be expressed as

$$\text{find} \quad \mathbf{g}_o, \mathbf{v}, \mathbf{g}_i, \tag{23a}$$

$$\text{subject to} \quad \text{rank}(M_m(\mathbf{y}, \mathbf{g}_i, \hat{\mathbf{g}}_i)) = rt, \tag{23b}$$

$$\text{rank}(M_c(\mathbf{g}_i, \mathbf{g}_o, \mathbf{v}, V, \hat{\mathbf{g}}_o, \hat{\mathbf{v}})) = \text{rank}(V), \tag{23c}$$

$$\mathbf{g}_o \in \mathcal{K}_{\mathbf{g}_o}, \quad \mathbf{h} = \mathbf{B}\mathbf{v} \in \mathcal{K}_{\mathbf{h}}. \tag{23d}$$

The caveat is that rank-constrained problems are in general NP-hard, that is (informally), in general there do not exist algorithms that can compute a guaranteed feasible solution within a time that is bounded by a polynomial in the number of variables. However, we can attempt to compute a solution  $\{\mathbf{g}_o^*, \mathbf{g}_i^*, \mathbf{v}^*\}$  and check whether the matrices  $M_m(\mathbf{y}, \mathbf{g}_i^*, \hat{\mathbf{g}}_i)$  and  $M_c(\mathbf{g}_i^*, \mathbf{g}_o^*, \mathbf{v}^*, V, \hat{\mathbf{g}}_o, \hat{\mathbf{v}})$  have the correct rank.

## D. Convex Heuristic for Blind Deconvolution

Even though the reformulated problem with its rank constraints is still non-convex, we propose to use a convex heuristic, the nuclear norm [41], to attempt to minimize the ranks of the matrices involved. The nuclear norm of a matrix is defined as the sum of the singular values of a matrix

$$\|X\|_* = \sum_i \sigma_i(X), \quad (24)$$

where  $\sigma_i(X)$  is the  $i$ th largest singular value of  $X$ . We can therefore use the nuclear norm as a convex heuristic for the blind deconvolution problem to attempt to find a solution, but success is not guaranteed. The convex optimization approach for Eq. (23) is

$$\begin{aligned} \min_{\mathbf{g}_o, \mathbf{v}, \mathbf{g}_i} \mu \|M_m(\mathbf{y}, \mathbf{g}_i, \hat{\mathbf{g}}_i)\|_* + \|M_c(\mathbf{g}_i, \mathbf{g}_o, \mathbf{v}, V, \hat{\mathbf{g}}_o, \hat{\mathbf{v}})\|_*, \\ \mathbf{g}_o \in \mathcal{K}_{\mathbf{g}_o}, \quad \mathbf{h} = \mathbf{B}\mathbf{v} \in \mathcal{K}_{\mathbf{h}}, \end{aligned} \quad (25)$$

where the parameter  $\mu > 0$  is a tuning parameter that weighs the nuclear norm of matrix  $M_m$  with the nuclear norm of matrix  $M_c$ .

The optimization problem in Eq. (25) is parameterized in Eqs. (17) and (21) by  $\hat{\mathbf{g}}_o, \hat{\mathbf{v}},$  and  $\hat{\mathbf{g}}_i$ . The interpretation is that, given some guess  $\{\hat{\mathbf{g}}_o, \hat{\mathbf{v}}, \hat{\mathbf{g}}_i\}$ , Eq. (25) produces a new estimate  $\{\hat{\mathbf{g}}_o^+, \hat{\mathbf{v}}^+, \hat{\mathbf{g}}_i^+\}$ . Motivated by [6,40], Eq. (25) can be used in an iterative update scheme; see Algorithm 1.

**Algorithm 1.** Convex optimization-based blind deconvolution for images taken with coherent illumination

```

1: procedure
2:    $k = 0$ 
3:   while not converged do
4:     Let  $\{\hat{\mathbf{g}}_i^{k+1}, \hat{\mathbf{g}}_o^{k+1}, \hat{\mathbf{v}}^{k+1}\}$  be the arguments that minimize Eq. (25).
5:      $k = k + 1$ 
6:   end while
7: end procedure

```

Such an iterative scheme gives rise to three questions. First, do the estimates converge to a fixed point? Second, are the resulting estimates correct solutions to the blind deconvolution problem? Third, if they converge, how fast do they converge? Unfortunately, all three questions are very difficult to answer, and we cannot provide a theoretical proof of convergence. We do notice, however, that correct solutions of the blind deconvolution problem are fixed points of Algorithm 1. For solutions  $\{\mathbf{g}_o^*, \mathbf{v}^*, \mathbf{g}_i^*\}$  of the blind deconvolution problem, we verify that by substitution

$$\begin{aligned} \mu \|M_m(\mathbf{y}, \mathbf{g}_i^*, \hat{\mathbf{g}}_i^*)\|_* + \|M_c(\mathbf{g}_i^*, \mathbf{g}_o^*, \mathbf{v}^*, V, \hat{\mathbf{g}}_o^*, \hat{\mathbf{v}}^*)\|_* \\ = \mu \left\| \begin{pmatrix} 0 & 0 \\ 0 & I_{rt} \end{pmatrix} \right\|_* + \left\| \begin{pmatrix} 0 & 0 \\ 0 & V \end{pmatrix} \right\|_* = \mu rt + \|V\|_*, \end{aligned} \quad (26)$$

which does not depend on any of the variables. So if  $\{\hat{\mathbf{g}}_o, \hat{\mathbf{v}}, \hat{\mathbf{g}}_i\} = \{\mathbf{g}_o^*, \mathbf{v}^*, \mathbf{g}_i^*\}$  in Eq. (25), the optimal parameters for Eq. (25) are  $\{\mathbf{g}_o^*, \mathbf{v}^*, \mathbf{g}_i^*\}$ .

The convergence speed properties and success rate of Algorithm 1 depend on the initialization  $\{\hat{\mathbf{g}}_o^0, \hat{\mathbf{v}}^0, \hat{\mathbf{g}}_i^0\}$  and

tuning of  $\mu$  and the matrices  $W_1, W_2$  in Eqs. (17) and (21). To show the difference that tuning of  $W_1$  and  $W_2$  in Eqs. (17) and (21) can make, we solve a small, one-dimensional blind deconvolution problem with three different sets of tuning parameters (see <https://bitbucket.org/rdoelman/blinddeconvolution>). We set  $W_1 = W_2 = m_1 I$  in Eq. (21),  $W_1 = c_1 I$ ,  $W_2 = c_2 I$  in Eq. (17), and  $\mu = 1$ . The three sets of parameters  $(m_1, c_1, c_2)$  are (1, 1, 1), (2, 1, 4), and (0.6, 1, 0.6). The different convergence speeds can be seen in Fig. 1. It can be seen that the effect of tuning on the convergence speed can be very large. Unfortunately, we cannot provide general tuning rules that optimize convergence speed.

## E. Computational Complexity of Eq. (25)

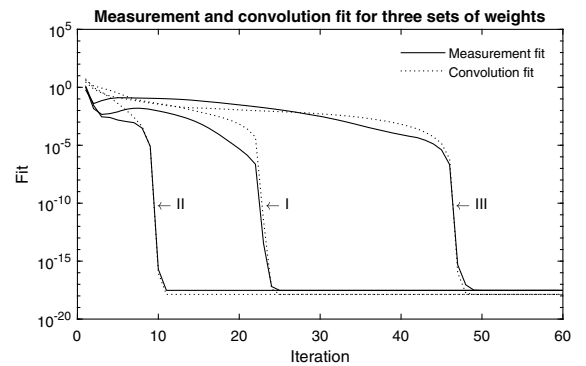
The computational complexity of the nuclear norm minimization in Eq. (25) can be estimated as follows. If we assume that minimizing the nuclear norm of a matrix with  $n$  variables is of approximately  $\mathcal{O}(n^6)$  when using a standard semidefinite programming (SDP) solver [42], then solving Eq. (25) is of complexity  $\mathcal{O}((rt + mn + pq)^6)$ , which is very unfavorable for practical applications. An alternating direction method of multipliers (ADMM, [43,44]) solution for Eq. (25) consists of the singular value decomposition of the matrices  $M_c \in \mathbb{C}^{(rt+mnrt) \times (1+pq)}$  and  $M_m \in \mathbb{C}^{2rt \times 2rt}$  that are of  $\mathcal{O}(rt(1+mn)(1+pq)^2 + (1+pq)^3)$  and  $\mathcal{O}((rt)^3)$ , respectively.

Exploiting parallelization opportunities similar to [6], this can be reduced to  $rt$  singular value decompositions (SVDs) with complexity  $\mathcal{O}(\max(mn, pq)^3)$  and  $rt$  SVDs of matrices of size 2 with complexity  $\mathcal{O}(1)$ , which can be computed in parallel.

## F. Including Prior Information and Regularization

The optimization in Eq. (25) is a *convex* optimization problem in the decision parameters  $\mathbf{g}_i, \mathbf{g}_o,$  and  $\mathbf{v}$ . This makes the addition of prior information and regularization very simple, if these can be expressed as convex constraints or convex penalty functions. The convex optimization-based blind deconvolution (for incoherent illumination) techniques such as [17] are based on directly estimating  $|\mathbf{g}_i|^2$  and  $|\mathbf{h}|^2$ , making it difficult to apply constraints on  $\mathbf{g}_i$  and  $\mathbf{h}$ .

We here list some examples of prior information that can be included.



**Fig. 1.** Convergence of the constraint violations  $\|\mathbf{y} - |\hat{\mathbf{g}}_i|^2\|_F^2$  (measurement fit) and  $\|\hat{\mathbf{g}}_i - \hat{\mathbf{g}}_o \star \mathbf{B}\hat{\mathbf{v}}\|_F^2$  (convolution fit) through updates in Algorithm 1 for three different sets of tuning parameters.

1. The imaged object has a known support (known non-zero-valued pixels). This can be expressed as the constraint  $\text{Qvect}(\mathbf{g}_o) = 0$  for a selection matrix  $\mathbf{Q}$ .

2. The imaged object is sparse, in the sense that many pixels of  $\mathbf{g}_o$  have value 0. This can be expressed through the addition of a penalty term  $\tau \sum_i |\mathbf{g}_{oi}|$  with regularization parameter  $\tau$  and  $i$  denoting the  $i$ th pixel.

3. The extension to the use of multiple images, taken with different phase diversities, can be done by adding additional terms to the objective function corresponding to the different images and with addition of the constraints  $\mathbf{h}_n = \mathbf{B}_n \mathbf{v} \in \mathcal{K}_{h_n}$  for the  $n$ th image.

4. In ptychography, overlapping parts of an object positioned in the pupil plane are imaged with the same “probe” or amplitude transfer function. If we write the Fourier transform as a linear mapping with a matrix  $\mathbf{F}$ ,  $\text{vect}(\mathcal{F}\{x\}) = \mathbf{F} \text{vect}(x)$ , then a shift in the position of the illuminated object can be represented by the constraint  $\mathbf{F}_1 \text{vect}(\mathbf{g}_{o1}) = \mathbf{F}_2 \text{vect}(\mathbf{g}_{o2})$ , where  $\mathbf{F}_1$  and  $\mathbf{F}_2$  are those parts of the Fourier transform matrices that correspond to the overlapping part of the object. This constraint addresses the problem that a phase aberration of the probe can be attributed to the phase of the object and the other way around.

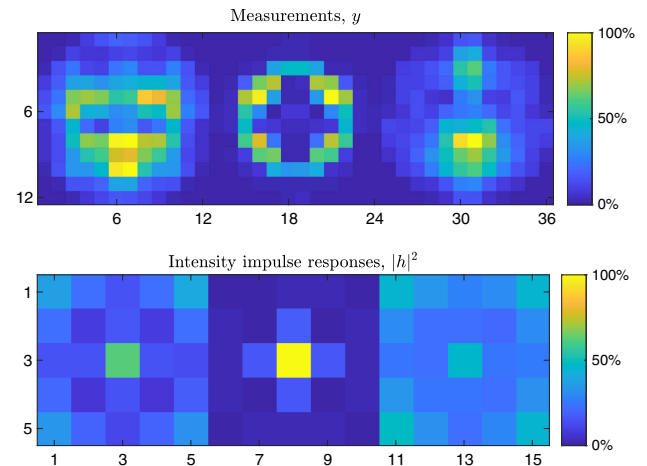
#### 4. NUMERICAL EXPERIMENTS

We implemented an ADMM algorithm in MATLAB to compute the updates in Algorithm 1. Although this allows for parallel computations of *rtcd* SVDs with complexity  $\mathcal{O}(\max(mm, pq)^3)$ , where  $d$  is the number of images taken, and *rtcd* SVDs with complexity  $\mathcal{O}(1)$ , we computed these in series. Due to the computational complexity, we tested the algorithm for two cases with small dimensions. Furthermore, the ADMM algorithm that iteratively finds the optimal solution to Eq. (25) is terminated after only 10 iterations.

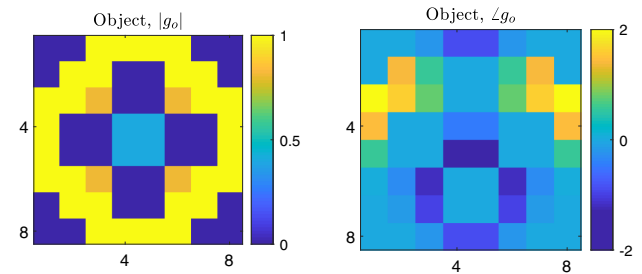
For comparison, we implemented a gradient descent method comparable to [30,31,36], but adapted to our formulation with decision variables defined in the focal plane. An accelerated gradient descent scheme, ADAM [45], is used to speed up the procedure and automatically determine step size. The step size  $\eta$  is tuned once up front to ensure convergence. The settings are:  $\beta_1 = 0.8, \beta_2 = 0.999, \varepsilon = 1 \cdot 10^{-8}, \eta = 2 \cdot 10^{-4}$ .

The experiment models an unknown aberration consisting of eight basis functions, as in Eq. (6), that approximate a small defocus  $\phi = 0.2Z_2^0$ , where  $Z_2^0$  is the defocus Zernike polynomial. We take three images with phase diversities that are defoci with coefficients  $\approx -2, 0$  and  $2$ . Due to the computational complexity, the aperture is undersampled when the amplitude impulse response is computed, and the resulting matrix is cut to a size of  $5 \times 5$ . The object  $\mathbf{g}_o$  is a complex-valued matrix of dimensions  $8 \times 8$ , and the resulting measurements  $\mathbf{y}$  are of size  $12 \times 12$ ; see Figs. 2 and 3. The value of  $\mu$  in Eq. (25) is tuned to 0.55.

Both Algorithm 1 and the gradient descent method are tested on a noiseless case and the same case where measurement noise has been added with a signal-to-noise ratio (SNR) of 20 dB. Both algorithms in both cases are initialized with the same initial guess, where the pixels of the initial object estimate are randomly drawn from a Gaussian distribution and the



**Fig. 2.** Top: the three  $12 \times 12$  (noiseless) measured intensities  $\mathbf{y} = |\mathbf{g}_o \star \mathbf{h}|^2$ . Bottom: the three  $5 \times 5$  intensity impulse response functions (point spread functions)  $\mathbf{s} = |\mathbf{h}|^2$  corresponding to the three different diversities that generate the different  $\mathbf{h}$ .



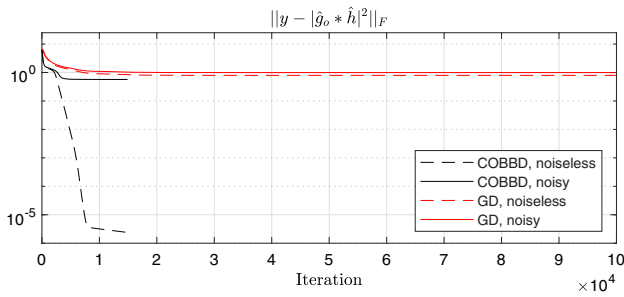
**Fig. 3.** Left: the amplitude of the object. Right: the phase in radians of the object.

initial guesses for the coefficients are those that best approximate zero aberration. The computation time for the proposed method, implemented without taking advantage of parallel computation of SVDs, is approximately 10 h for the 15,000 iterations as shown here. The computation time consists of roughly 5 h for computation of SVDs, 40 min for solving least-squares problems, and the rest is overhead. The gradient descent method is much faster, with approximately 18 min for 1,00,000 iterations. The resulting norms of the residual between measurements and convolution of the estimated object and amplitude impulse response are plotted in Fig. 4. As can be seen in this figure, the gradient descent method gets stuck in the noiseless case, whereas the proposed method converges to a feasible solution.

The estimated values  $\hat{\mathbf{v}}$  and  $\hat{\mathbf{g}}_o$  have an ambiguity, since for a complex scalar  $c$ ,  $c\hat{\mathbf{B}}\hat{\mathbf{v}} \star \hat{\mathbf{g}}_o = \hat{\mathbf{B}}\hat{\mathbf{v}} \star c\hat{\mathbf{g}}_o$ . We can remove the ambiguity from  $\hat{\mathbf{v}}$ , for example, when reporting the estimation error by computing

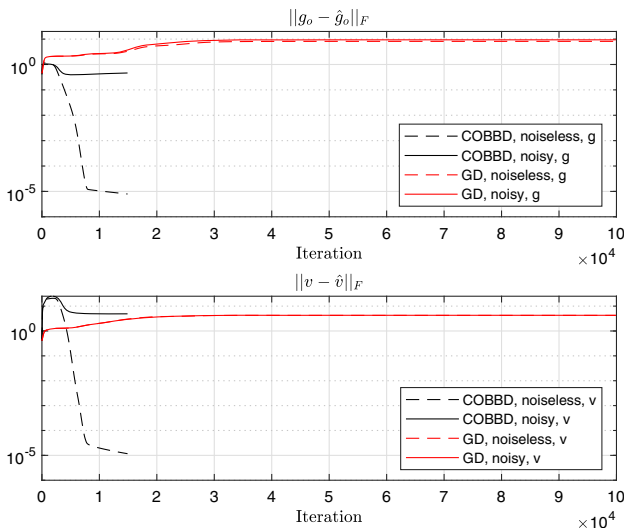
$$\min_{c \in \mathbb{C}} \|c\hat{\mathbf{v}} - \mathbf{v}\|_2. \quad (27)$$

After removal of the ambiguity of the estimated values of  $\mathbf{g}_o$  and  $\mathbf{v}$ , we plot in Fig. 5 the norms of the residuals between the

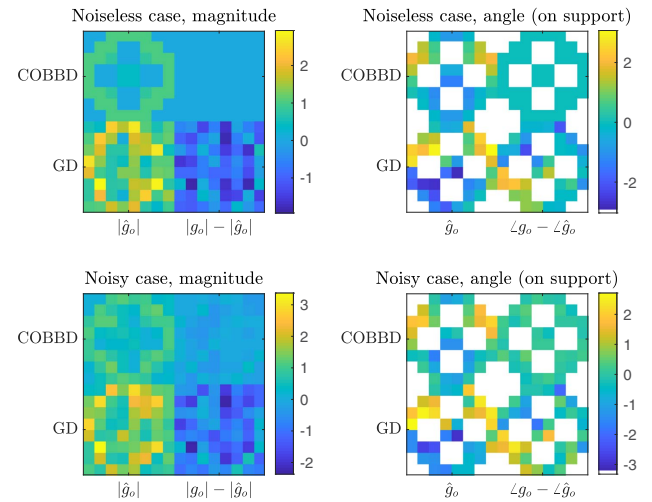


**Fig. 4.** Measurement fit generated by the two algorithms. Black: Algorithm 1. Red: gradient descent. Solid lines show the case with noisy measurements (SNR: 20 dB); dashed lines show the noiseless case.

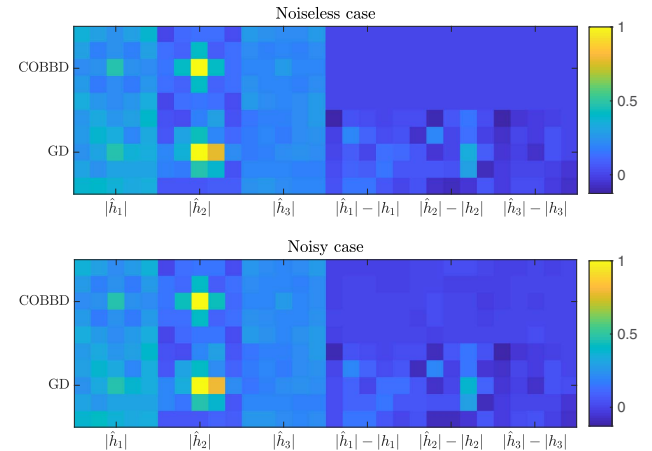
actual complex amplitude of the object  $\mathbf{g}_o$  and coefficients  $\mathbf{v}$  and their estimated values. As can be seen in this figure, the proposed method converges in the noiseless case not just to a feasible solution but to the correct solution, whereas the gradient descent method stops progressing towards the solution. In the noisy case, Algorithm 1 computes the solution with the best estimated measurements (see Fig. 4) and with the best estimated object (Fig. 5, top). The coefficients  $\mathbf{v}$  have an estimate that is further from the real coefficients than the estimate of the gradient descent method, but given the measurement fit and fit of  $\mathbf{g}_o$ , the effect of this error is small. The estimates resulting from Algorithm 1 and from the gradient descent method of the object  $\mathbf{g}_o$  are displayed in Fig. 6. From Fig. 6 it becomes clear that even though in the noisy case the proposed method does not converge to the exact solution, it converges to a solution that resembles the original object quite well. Inspecting Fig. 5 shows that the gradient descent method provides estimates of



**Fig. 5.** Frobenius norm of the residual between the true variables  $\mathbf{g}_o$ , the complex-valued object, and  $\mathbf{v}$ , the radial basis function coefficients, and the (ambiguity removed) estimated variables  $\hat{\mathbf{g}}_o$  and  $\hat{\mathbf{v}}$ . Top figure: residuals for  $\mathbf{g}_o$ . Bottom figure: residuals for  $\mathbf{v}$ . Black: Algorithm 1. Red: gradient descent. Solid lines show the case with noisy measurements (SNR: 20 dB); dashed lines show the noiseless case.



**Fig. 6.** Estimates and residuals of  $\mathbf{g}_o$  using Algorithm 1 and gradient descent with identical initialization. Top left: the estimated amplitude of  $\mathbf{g}_o$  and the residual for the noiseless case. Top right: the estimated angle of  $\mathbf{g}_o$  and the residual for the noiseless case. Bottom left: the estimated amplitude of  $\mathbf{g}_o$  and the residual for the noisy case. Bottom right: the estimated angle of  $\mathbf{g}_o$  and the residual for the noisy case. Since estimates of very small complex values can have a radically different complex angle, the angle is only plotted for the nonzero pixels in the original object of Fig. 3.



**Fig. 7.** Estimates of  $|\mathbf{h}|$ . The maximum absolute value of  $\hat{\mathbf{h}}$  is scaled to 1.

$\mathbf{g}_o$  that are far from it. The resulting estimates of  $|\mathbf{h}|$  are shown in Fig. 7.

## 5. CONCLUSION AND FUTURE RESEARCH

We derived a convex heuristic for the blind deconvolution problem for images taken with coherent illumination that is also able to incorporate the concept of phase diversity. We suggested an update scheme and demonstrated on a numerically illustrative example that it is capable of retrieving the object and PSF from a random initialization, thereby overcoming local minima. At the moment, the method is computationally burdensome, but we expect computational improvements similar to [6] by

fully exploiting parallelization opportunities and the structure in the optimization problems. Apart from the nuclear norm heuristic, there are also other methods that attempt to find low rank results, like the difference of convex programming (e.g., [46]) or application of the truncated nuclear norm (e.g., [47]), but we leave the evaluation of their performance for future research. Several questions still remain open concerning optimal tuning rules for the different parameters in the optimization, the performance of other non-convex low-rank-inducing norms, bounds on convergence speed, and the computational speed-up by exploiting parallelization opportunities.

## APPENDIX A: PROOF OF LEMMA 2

The vector  $z = \text{vect}(\text{vect}(\mathbf{g}_o)\text{vect}(\mathbf{B}\mathbf{v})^T)$  lists all possible products between elements of  $\mathbf{g}_o$  and elements of  $\mathbf{B}\mathbf{v}$ . Since the elements of  $\mathbf{g}_o$ , the result of a discrete convolution, are sums of specific elements of  $z$ , it is possible to construct a matrix of zeros and ones  $L \in \mathbb{B}^{rt \times mpq}$ , such that

$$\text{vect}(\mathbf{g}_i) = L \text{vect}(\text{vect}(\mathbf{g}_o)\text{vect}(\mathbf{B}\mathbf{v})^T). \quad (\text{A1})$$

Application of the identity [48]

$$\text{vect}(AXB) = (B^T \otimes A)\text{vect}(X) \quad (\text{A2})$$

allows us to rewrite Eq. (A1) as

$$L(I_{pq} \otimes \text{vect}(\mathbf{g}_o)^T)\text{vect}(\mathbf{B}\mathbf{v}). \quad (\text{A3})$$

Let  $l_i$  be the  $i$ th row of  $L$ . Then, applying Eq. (A2) on the  $i$ th row of Eq. (A3) gives

$$l_i(I_{pq} \otimes \text{vect}(\mathbf{g}_o)^T) = \text{vect}(\mathbf{g}_o)^T L_i, \quad (\text{A4})$$

where  $L_i^T = \text{vect}(L_i)$  and  $L_i \in \mathbb{B}^{mn \times pq}$ . Combining the expressions for all rows, the result is

$$\text{vect}(\mathbf{g}_i) = (I_{rt} \otimes \text{vect}(\mathbf{g}_o)^T) \begin{pmatrix} L_1 \\ \vdots \\ L_{rt} \end{pmatrix} \text{vect}(\mathbf{B}\mathbf{v}). \quad (\text{A5})$$

Using a row reordering of the matrix with blocks  $L_i$  ([48], Eq. 2.14) gives us  $V$ , the expression in Eq. (14).

**Funding.** FP7 Ideas: European Research Council (IDEAS-ERC) (FP7/2007-2013) (339681).

## REFERENCES

1. J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micro-metre-sized non-crystalline specimens," *Nature* **400**, 342–344 (1999).
2. A. MacDonald, "Blind deconvolution of anisoplanatic images collected by a partially coherent imaging system," Ph.D. thesis (Air Force Institute of Technology, Wright-Patterson Air Force Base Ohio, 2004).
3. D. Pastor, T. Stefaniuk, P. Wróbel, C. J. Zapata-Rodríguez, and R. Kotyński, "Determination of the point spread function of layered meta-materials assisted with the blind deconvolution algorithm," *Opt. Quantum Electron.* **47**, 17–26 (2015).
4. J. M. Rodenburg and H. M. Faulkner, "A phase retrieval algorithm for shifting illumination," *Appl. Phys. Lett.* **85**, 4795–4797 (2004).
5. J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 2008).
6. R. Doelman, N. H. Thao, and M. Verhaegen, "Solving large-scale general phase retrieval problems via a sequence of convex relaxations," *J. Opt. Soc. Am. A* **35**, 1410–1419 (2018).
7. G. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.* **13**, 547–549 (1988).
8. M. Tofighi, O. Yorulmaz, K. Köse, D. C. Yildirim, R. Çetin-Atalay, and A. E. Cetin, "Phase and TV based convex sets for blind deconvolution of microscopic images," *IEEE J. Sel. Top. Signal Process.* **10**, 81–91 (2016).
9. D. Wilding, O. Soloviev, P. Pozzi, G. Vdovin, and M. Verhaegen, "Blind multi-frame deconvolution by tangential iterative projections (tip)," *Opt. Express* **25**, 32305–32322 (2017).
10. D. Fish, A. Brincombe, E. Pike, and J. Walker, "Blind deconvolution by means of the Richardson-Lucy algorithm," *J. Opt. Soc. Am. A* **12**, 58–65 (1995).
11. P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Process. Mag.* **23**(3), 32–45 (2006).
12. G. Liu, S. Chang, and Y. Ma, "Blind image deblurring using spectral properties of convolution operators," *IEEE Trans. Image Process.* **23**, 5047–5056 (2014).
13. T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *J. Opt. Soc. Am. A* **10**, 1064–1073 (1993).
14. T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach," *J. Opt. Soc. Am. A* **9**, 1052–1061 (1992).
15. R. Mourya, L. Denis, J.-M. Becker, and E. Thiébaud, "A blind deblurring and image decomposition approach for astronomical image restoration," in *23rd European Signal Processing Conference (EUSIPCO)* (IEEE, 2015), pp. 1636–1640.
16. D. Stöger, P. Jung, and F. Krahmer, "Blind deconvolution and compressed sensing," in *4th International Workshop on Compressed Sensing Theory and Its Applications to Radar, Sonar and Remote Sensing (CoSeRa)* (IEEE, 2016), pp. 24–27.
17. A. Ahmed, A. Cosse, and L. Demanet, "A convex approach to blind deconvolution with diverse inputs," in *6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)* (IEEE, 2015), pp. 5–8.
18. S. Ling and T. Strohmer, "Regularized gradient descent: a non-convex recipe for fast joint blind deconvolution and demixing," *Information and Inference: A Journal of the IMA* **8**, 1–49 (2019).
19. P. Jung, F. Krahmer, and D. Stöger, "Blind demixing and deconvolution at near-optimal rate," *IEEE Trans. Information Theory* **64**, 704–727 (2018).
20. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng.* **21**, 215829 (1982).
21. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy* **109**, 1256–1262 (2009).
22. A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg, "Superresolution imaging via ptychography," *J. Opt. Soc. Am. A* **28**, 604–612 (2011).
23. R. Horstmeyer, X. Ou, J. Chung, G. Zheng, and C. Yang, "Overlapped Fourier coding for optical aberration removal," *Opt. Express* **22**, 24062–24080 (2014).
24. F. Jian and L. Peng, "A general phase retrieval algorithm based on a ptychographical iterative engine for coherent diffractive imaging," *Chin. Phys. B* **22**, 014204 (2013).
25. M. Foreman, C. Giusca, P. Török, and R. Leach, "Phase-retrieved pupil function and coherent transfer function in confocal microscopy," *J. Microsc.* **251**, 99–107 (2013).
26. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," *Opt. Express* **22**, 4960–4972 (2014).
27. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," *Ultramicroscopy* **109**, 338–343 (2009).
28. R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam, "Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging," *SIAM J. Imaging Sci.* **8**, 426–457 (2015).
29. H. Chang, P. Enfedaque, and S. Marchesini, "Blind ptychographic phase retrieval via convergent alternating direction method of multipliers," *SIAM J. Imaging Sci.* **12**, pp. 153–185 (2019).
30. G. R. Brady, M. Guizar-Sicairos, and J. R. Fienup, "Optical wavefront measurement using phase retrieval with transverse translation diversity," *Opt. Express* **17**, 624–639 (2009).

31. M. Guizar-Sicairos and J. R. Fienup, "Measurement of coherent x-ray focused beams by phase retrieval with transverse translation diversity," *Opt. Express* **17**, 2670–2685 (2009).
32. Y. S. Nashed, T. Peterka, J. Deng, and C. Jacobsen, "Distributed automatic differentiation for ptychography," *Procedia Comput. Sci.* **108**, 404–414 (2017).
33. M. Odstrčil, A. Menzel, and M. Guizar-Sicairos, "Iterative least-squares solver for generalized maximum-likelihood ptychography," *Opt. Express* **26**, 3108–3123 (2018).
34. Y. Zhang, W. Jiang, and Q. Dai, "Nonlinear optimization approach for Fourier ptychographic microscopy," *Opt. Express* **23**, 33822–33835 (2015).
35. L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, "Experimental robustness of Fourier ptychography phase retrieval algorithms," *Opt. Express* **23**, 33214–33240 (2015).
36. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," *New J. Phys.* **14**, 063004 (2012).
37. A. Tripathi, I. McNulty, and O. G. Shpyrko, "Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods," *Opt. Express* **22**, 1452–1466 (2014).
38. R. Horstmeyer, R. Y. Chen, X. Ou, B. Ames, J. A. Tropp, and C. Yang, "Solving ptychography with a convex relaxation," *New J. Phys.* **17**, 053044 (2015).
39. P. Piscaer, A. Gupta, O. Soloviev, and M. Verhaegen, "Modal-based phase retrieval using Gaussian radial basis functions," *J. Opt. Soc. Am. A* **35**, 1233–1242 (2018).
40. R. Doelman and M. Verhaegen, "Sequential convex relaxation for convex optimization with bilinear matrix equalities," in *European Control Conference (ECC)* (IEEE, 2016), pp. 1946–1951.
41. B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.* **52**, 471–501 (2010).
42. Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM J. Matrix Anal. Appl.* **31**, 1235–1256 (2009).
43. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.* **3**, 1–122 (2011).
44. J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.* **20**, 1956–1982 (2010).
45. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
46. H. Tuy, "DC optimization: theory, methods and algorithms," in *Handbook of Global Optimization* (Springer, 1995), pp. 149–216.
47. Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 2117–2130 (2013).
48. D. A. Turkington, *Generalized Vectorization, Cross-Products, and Matrix Calculus* (Cambridge University, 2013).