PROCEEDINGS OF SPIE

Bilinear solution to the phase diversity problem for extended objects based on the Born approximation

Raluca M. Andrei
Rufus Fraanje
Michel Verhaegen
Visa A. Korkiakoski
Christoph U. Keller
Niek Doelman

SPIE.

Bilinear solution to the phase diversity problem for extended objects based on the Born approximation

Raluca M. Andrei*[a], Rufus Fraanje[a], Michel Verhaegen[a], Visa A. Korkiakoski[b], Christoph U. Keller[b] and Niek Doelman[c]

[a]Delft Center for Systems and Control, Mekelweg 2, 2628CD Delft, The Netherlands;
[b]Leiden Observatory, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands;
[c]TNO Science and Industry, Stieltjesweg 1, 2628CK Delft, The Netherlands

ABSTRACT

We propose a new approach for the joint estimation of aberration parameters and unknown object from diversity images with applications in imaging systems with extended objects as astronomical ground-based observations or solar telescopes. The motivation behind our idea is to decrease the computational complexity of the conventional phase diversity (PD) algorithm and avoid the convergence to local minima due to the use of nonlinear estimation algorithms. Our approach is able to give a good starting point for an iterative algorithm or it can be used as a new wavefront estimation method. When the wavefront aberrations are small, the wavefront can be approximated with a linear term which leads to a quadratic point-spread function (PSF) in the aberration parameters. The presented approach involves recording two or more diversity images and, based on the before mentioned approximation estimates the aberration parameters and the object by solving a system of bilinear equations, which is obtained by subtracting from each diversity image the focal plane image. Moreover, using the quadratic PSFs gives improved performance to the conventional PD algorithm through the fact that the gradients of the PSFs have simple analytical formulas.

Keywords: point-spread function, wavefront reconstruction, born approximation, conjugate gradients

1. INTRODUCTION

All optical instruments are subject to optical aberrations either intrinsic to the instrument or due to the atmospheric turbulence when imaging through the Earth's atmosphere. Inhomogeneities of the temperature induce inhomogeneities of the refractive index of air which lead to phase variations in the pupil plane and severely reduce the optical transfer function of the instrument. If the aberrations can be estimated, they can be at least partially compensated either through real-time techniques during image acquisition or post-processing. We will investigate here a way to adapt a certain wavefront sensing (WFS) technique, called phase diversity, used mostly for post-processing, to real-time adaptive optics systems.

PD directly uses image data for the estimation of the aberrations, and it is thus sensitive to all aberrations degrading the quality of the optical instrument, contrary to wavefront sensors as the Shack-Hartmann. The method consists of collecting two or more images. One of them is the focal plane image that has been degraded by unknown aberrations. Additional images are obtained by introducing known aberrations into the system. This method was first proposed by Gonsalves[1] for two diversity images, and it was based on the minimization of a least-squares (LS) criterion. Since the paper of Gonsalves, the method has been extended in numerous papers[2–5] . Paxman[2] , formulated it as a maximum-likelihood estimation problem and generalized it to include an arbitrary number of diversity measurements. Also, considering known aberrations, a closed-form expression was obtained for the optimum object and introduced in the initial problem. The modified problem to be solved has the aberrations as the only unknowns. If prior distributions on the object and aberrations are included in the estimation, then the problem becomes a joint maximum a-posteriori estimation (MAP)[4] . One disadvantage of the MAP is that it introduces hyper parameters - parameters of a prior distribution used in the Bayesian

Further author information: (Send correspondence to Raluca Andrei)
Raluca Andrei: E-mail: r.m.andrei@tudelft.nl

estimation theory to distinguish them from parameters of the model for the underlying system under analysis -, that have to be tuned for each object or turbulent condition. This problem is solved by integrating the object out of the problem, or marginalization[5] , and estimating the hyper parameters together with the aberrations.

PD has a number of advantages over other WFS techniques, like the fact that it does not require a bright, point-like reference beacon, that may not be always available, or the fact that it can estimate high order aberrations. The disadvantages are given by the fact that it requires more extensive post-processing to estimate the aberration parameters and also the computational time required is considerable compared to the evolution time of the turbulence. Research efforts are looking at different ways to use this type of WFS technique in a real-time adaptive optics system. Demonstrations of real-time correction have been obtained for very few corrected aberrations[6] . Improvements to PD have been made by better numerical algorithms[7,8] , by using object independent error metrics to estimate the aberrations from the data[9] or by different linearization techniques[10–12] .

The approach we use here is to linearize the wavefront when the aberrations are small. We give an analytic formula for the PSF which is quadratic in the aberration parameters using a finite series of Zernike polynomials. This approximation can speed up algorithms like the one proposed in[2] by using simple analytical formulas for gradients of the PSFs. Our approach is to subtract two or more diversity images with quadratic PSFs in order to end up with linear expressions in both the aberration parameters and the object, in other words a bilinear problem. By solving this bilinear problem, we estimate the aberration parameters and the unknown object. Using this approach, we show that the computational time is reduced compared to conventional PD[2] .

Systems of bilinear equations are not as well understood as linear systems, and few results are reported in the literature[13–15] . The estimation of the bilinear parameters is usually formulated as a nonlinear least squares problem. In this paper, the solution of the bilinear system is obtained using the so-called Two-Stage Algorithm (TSA)[16] . In a first step, the TSA uses an over-parametrization of the bilinear system and transforms it into a linear system before reducing the estimated linear parameters back to the bilinear ones in a second step. The linear system appearing in the first step is a large-scale, sparse system that can be solved using conjugate gradient type iterative algorithms. The second step consists of an 'economic' singular value decomposition (SVD) of a fat matrix (more columns than rows).

The remaining part of this paper is structured as follows. Section 2 starts with the analytical expression for the PSF, which is written as a quadratic function in terms of the Zernike coefficients. Section 3 summarizes the phase diversity algorithm[3] and shows how the gradients of the analytical PSFs can be computed. In Section 4 we define and give the solution to the wavefront estimation problem formulated as a system of bilinear equations. Section 5 is dedicated to numerical simulations that validate the results presented in the previous sections. Concluding comments are placed in Section 6.

2. ANALYTICAL FORMULA FOR THE PSF

In this section we derive analytical formulas for the complex amplitude U (or spatial impulse response) of the PSF using a linearization of the wavefront. Once we compute this quantity, the intensity PSF can be easily derived as $j = |U|^2$. To this end, we use the Nijboer-Zernike theory of diffraction integrals[17,18] containing small aberrations. These diffraction integrals apply to optical systems where the pupil is large compared to the wavelength of the light used. The choice of Zernike polynomials for the representation of the wavefront is motivated by the fact that they have analytical expressions and can be easily included in the computation of the PSFs. The derivations here follow the lines of[19–21] . We consider a point source of monochromatic light in the object plane of a centered optical system. The diffraction integral that gives the spatial impulse response $U(x, y)$ in image space is

$$
\begin{aligned}
U(x, y) \equiv U(r, \varphi) &= \frac{1}{\pi} \iint_{\nu^2 + \mu^2 \leq 1} \exp\left[i\Phi(\nu, \mu)\right] \exp\left[2\pi i\left(\nu x + \mu y\right)\right] d\nu d\mu \\
&= \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \exp\left[i\Phi(\rho, \theta)\right] \exp\left[2\pi i\rho r \cos(\theta - \varphi)\right] \rho d\rho d\theta,
\end{aligned}
\tag{1}
$$

where the normalized coordinates in the exit pupil are $(\nu, \mu) = (\rho\cos\theta, \rho\sin\theta)$, the normalized coordinates in the image plane are $(x, y) = (r\cos\varphi, r\sin\varphi)$ and $\Phi(\cdot, \cdot)$ is the aberration function. We assume here that the

optical system introduces a wavefront aberration only and that the amplitude distribution over the wavefront is uniform. The aberration function Φ is approximated with a Zernike series

$$\Phi(\rho,\theta) = \sum_{n,m} Z_n^m(\rho,\theta)\,\alpha_n^m + Z_n^{-m}(\rho,\theta)\,\alpha_n^{-m} = \sum_{n,m} R_n^m(\rho)\left(\alpha_n^m \cos m\theta + \alpha_n^{-m}\sin m\theta\right), \tag{2}$$

where $Z_n^m(\rho,\theta) := R_n^m(\rho)\cos m\theta$ and $Z_n^{-m}(\rho,\theta) := R_n^m(\rho)\sin m\theta$. $Z_n^{\pm m}$ are the even and odd Zernike polynomials, which form an orthonormal basis on the unit disk, $R_n^m(\rho)$ are called radial polynomials, $n \geq 0$ is called the radial degree of the corresponding Zernike polynomial, $m \geq 0$ is called the azimuthal frequency, and $n - m \geq 0$ and even. Under the assumption that Φ is sufficiently small, we can linearize the wavefront

$$\exp\left[i\Phi(\rho,\theta)\right] \approx 1 + i\Phi(\rho,\theta). \tag{3}$$

It is easily seen that the Born approximation (3) is a first order Taylor approximation of the wavefront around 0. This approximation is meaningful for wavefronts with less than ~ 0.5 rad of phase variance and has been used in previous studies for phase retrieval (see[10,11]). With Eq. (2) we get

$$\exp\left[i\Phi(\rho,\theta)\right] \approx 1 + i\sum_{n,m} R_n^m(\rho)\left(\alpha_n^m \cos m\theta + \alpha_n^{-m}\sin m\theta\right). \tag{4}$$

Introducing Eq. (4) in Eq. (1), we obtain the complex amplitude U

$$
\begin{aligned}
U(r,\varphi) &\approx \frac{1}{\pi}\int_0^1 \rho \int_0^{2\pi}\left[1 + i\sum_{n,m} R_n^m(\rho)\left(\alpha_n^m \cos m\theta + \alpha_n^{-m}\sin m\theta\right)\right] \\
&\quad \times \exp\left[2\pi i\rho r\cos(\theta-\varphi)\right]\mathrm{d}\rho\mathrm{d}\theta \\[4pt]
&\approx \frac{1}{\pi}\int_0^1 \rho\left\{\int_0^{2\pi}\exp\left[i2\pi\rho r\cos(\theta-\varphi)\right]\mathrm{d}\theta\right. \\
&\quad + i\sum_{n,m} R_n^m(\rho)\,\alpha_n^m \int_0^{2\pi}\exp\left[i2\pi\rho r\cos(\theta-\varphi)\right]\cos m\theta\,\mathrm{d}\theta \\
&\quad \left. + i\sum_{n,m} R_n^m(\rho)\,\alpha_n^{-m}\int_0^{2\pi}\exp\left[i2\pi\rho r\cos(\theta-\varphi)\right]\sin m\theta\,\mathrm{d}\theta\right\}\mathrm{d}\rho.
\end{aligned}
\tag{5}
$$

In order to reduce Eq. (5), we use the following relation[20] , valid for an integer m

$$\int_0^{2\pi}\exp\left[iz\cos(\theta-\varphi)\right]\exp(im\theta)\,\mathrm{d}\theta = 2\pi i^m \exp(im\varphi)\,J_m(z). \tag{6}$$

Here, J_m is the Bessel function of the first kind of order m

$$J_m(x) = \sum_{k=0}^{m}\frac{(-1)^k}{k!\,\Gamma(k+m+1)}\left(\frac{x}{2}\right)^{2k+m}, \tag{7}$$

where $\Gamma(n) = (n-1)!$. Using Eq. (6), we can carry out the integration over θ in Eq. (5) and obtain

$$U(r,\varphi) \approx 2T_0^0(r) + 2i\sum_{n,m} i^m T_n^m(r)\left(\alpha_n^m \cos m\varphi + \alpha_n^{-m}\sin m\varphi\right), \tag{8}$$

where

$$T_n^m(r) = (-1)^{\frac{n-m}{2}}\frac{J_{n+1}(2\pi r)}{2\pi r}. \tag{9}$$

In the case when no aberrations are present, $2T_0^0(r)$ in Eq. (8) is the only nonzero term and is the amplitude corresponding to the Airy function. Also, in terms of intensity, $4\left(T_0^0(r)\right)^2$ is exactly the Airy function and the

PSF is the summation of this quantity plus/minus the other terms depending on $\boldsymbol{\alpha}$. In Eq. (9) we have used a well-known result in the Nijboer-Zernike theory (see Chapter 9 of[17]), namely that

$$\int\limits_0^1 \rho R_n^m\left(\rho\right) J_m\left(2\pi r\rho\right) \mathrm{d}\rho = \left(-1\right)^{\frac{n-m}{2}} \frac{J_{n+1}\left(2\pi r\right)}{2\pi r}. \tag{10}$$

Finally, the PSF is $j = |U|^2$

$$
\begin{aligned}
j\left(r,\varphi\right) = & \ 4\left[T_0^0\left(r\right)\right]^2 - 8\sum_{n,m} \mathrm{Im}\left(i^m\right) T_n^m\left(r\right) T_0^0\left(r\right) \left(\alpha_n^m \cos m\varphi + \alpha_n^{-m} \sin m\varphi\right) \\
& +4 \sum_{n_1,m_1;n_2,m_2} \alpha_{n_1}^{m_1} \alpha_{n_2}^{m_2} \mathrm{Re}\left(i^{m_1-m_2}\right) T_{n_1}^{m_1}\left(r\right) T_{n_2}^{m_2}\left(r\right) \cos m_1\varphi \cos m_2\varphi \\
& +4 \sum_{n_1,m_1;n_2,m_2} \alpha_{n_1}^{m_1} \alpha_{n_2}^{-m_2} \mathrm{Re}\left(i^{m_1-m_2}\right) T_{n_1}^{m_1}\left(r\right) T_{n_2}^{m_2}\left(r\right) \cos m_1\varphi \sin m_2\varphi \\
& +4 \sum_{n_1,m_1;n_2,m_2} \alpha_{n_1}^{-m_1} \alpha_{n_2}^{m_2} \mathrm{Re}\left(i^{m_1-m_2}\right) T_{n_1}^{m_1}\left(r\right) T_{n_2}^{m_2}\left(r\right) \sin m_1\varphi \cos m_2\varphi \\
& +4 \sum_{n_1,m_1;n_2,m_2} \alpha_{n_1}^{-m_1} \alpha_{n_2}^{-m_2} \mathrm{Re}\left(i^{m_1-m_2}\right) T_{n_1}^{m_1}\left(r\right) T_{n_2}^{m_2}\left(r\right) \sin m_1\varphi \sin m_2\varphi.
\end{aligned} \tag{11}
$$

The rather involved Eq. (11) can be rewritten into a simpler quadratic form

$$j\left(r,\varphi\right) = c_0 + \boldsymbol{c_1}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T Q\boldsymbol{\alpha}, \tag{12}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0^0 & \alpha_1^{-1} & \alpha_1^1 & \alpha_2^0 & \alpha_2^{-2} & \alpha_2^2 & \dots \end{bmatrix}^T, \tag{13}$$

$$c_0 := 4\left[T_0^0\left(r\right)\right]^2,$$

$$\boldsymbol{c_1}^T := -8 \begin{bmatrix} 0 & T_1^1 T_0^0 \sin\varphi & T_1^1 T_0^0 \cos\varphi & 0 & 0 & 0 & \dots \end{bmatrix},$$

$$Q := 4\times$$

$$
\begin{bmatrix}
\left(T_0^0\right)^2 & 0 & 0 & T_0^0 T_2^0 & -T_0^0 T_2^2 \sin 2\varphi & -T_0^0 T_2^2 \cos 2\varphi \dots \\
0 & \left(T_1^1\right)^2 \sin^2\varphi & \left(T_1^1\right)^2 \sin\varphi\cos\varphi & 0 & 0 & 0 \dots \\
0 & \left(T_1^1\right)^2 \cos\varphi\sin\varphi & \left(T_1^1\right)^2 \cos^2\varphi & 0 & 0 & 0 \dots \\
T_0^0 T_2^0 & 0 & 0 & \left(T_2^0\right)^2 & -T_2^0 T_2^2 \sin 2\varphi & -T_2^0 T_2^2 \cos 2\varphi \dots \\
-T_2^2 T_2^0 \sin 2\varphi & 0 & 0 & -T_2^2 T_2^0 \sin 2\varphi & \left(T_2^2\right)^2 \sin^2 2\varphi & \left(T_2^2\right)^2 \sin 2\varphi\cos 2\varphi \dots \\
-T_2^2 T_2^0 \cos 2\varphi & 0 & 0 & -T_2^2 T_2^0 \cos 2\varphi & \left(T_2^2\right)^2 \cos 2\varphi\sin 2\varphi & \left(T_2^2\right)^2 \cos^2 2\varphi \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}, \tag{14}
$$

$\boldsymbol{\alpha} \in \mathbb{R}^{N_\alpha}$, $c_0 \in \mathbb{R}$, $\boldsymbol{c_1} \in \mathbb{R}^{N_\alpha}$, and $Q \in \mathbb{R}^{N_\alpha \times N_\alpha}$.

Eq. (12) gives a way to compute the intensities in the focal planes as functions of the Zernike coefficients $\boldsymbol{\alpha}$ corresponding to the aberration function Φ. Given a grid of size $M \times N$ and using Eq. (12) for each point of this grid, the intensity PSF can be written as

$$
t_0\left(\boldsymbol{\alpha}\right) = \begin{bmatrix}
j\left(r_{11},\varphi_{11}\right) & j\left(r_{12},\varphi_{12}\right) & \dots & j\left(r_{1N},\varphi_{1N}\right) \\
j\left(r_{21},\varphi_{21}\right) & j\left(r_{22},\varphi_{22}\right) & \dots & j\left(r_{2N},\varphi_{2N}\right) \\
\vdots & \vdots & \vdots & \vdots \\
j\left(r_{M1},\varphi_{M1}\right) & j\left(r_{M2},\varphi_{M2}\right) & \dots & j\left(r_{MN},\varphi_{MN}\right)
\end{bmatrix}. \tag{15}
$$

The PSF in Eq. (15) can be rewritten as a function of $\boldsymbol{\alpha}$

$$t_0\left(\boldsymbol{\alpha}\right) = C_0 + C_1\left(I_N \otimes \boldsymbol{\alpha}\right) + Q_t\left(I_N \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha}\right), \tag{16}$$

where $\otimes$ denotes the Kronecker product, I_N is the identity matrix of order N, and the coefficients $C_0 \in \mathbb{R}^{M \times N}$, $C_1 \in \mathbb{R}^{M \times (N_\alpha N)}$, $Q_t \in \mathbb{R}^{M \times (N_\alpha{}^2 N)}$ are given below

$$
C_0 := \begin{bmatrix}
c_0\,(r_{11}) & c_0\,(r_{12}) & \ldots & c_0\,(r_{1N}) \\
c_0\,(r_{21}) & c_0\,(r_{22}) & \ldots & c_0\,(r_{2N}) \\
\vdots & \vdots & \ddots & \vdots \\
c_0\,(r_{M1}) & c_0\,(r_{M2}) & \cdots & c_0\,(r_{MN})
\end{bmatrix},
$$

$$
C_1 := \begin{bmatrix}
\boldsymbol{c_1}^T\,(r_{11},\varphi_{11}) & \boldsymbol{c_1}^T\,(r_{12},\varphi_{12}) & \ldots & \boldsymbol{c_1}^T\,(r_{1N},\varphi_{1N}) \\
\boldsymbol{c_1}^T\,(r_{21},\varphi_{21}) & \boldsymbol{c_1}^T\,(r_{22},\varphi_{22}) & \ldots & \boldsymbol{c_1}^T\,(r_{2N},\varphi_{2N}) \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{c_1}^T\,(r_{M1},\varphi_{M1}) & \boldsymbol{c_1}^T\,(r_{M2},\varphi_{M2}) & \cdots & \boldsymbol{c_1}^T\,(r_{MN},\varphi_{MN})
\end{bmatrix}, \tag{17}
$$

$$
Q_t := \begin{bmatrix}
\text{vec}\,(Q\,(r_{11},\varphi_{11}))^T & \text{vec}\,(Q\,(r_{12},\varphi_{12}))^T & \ldots & \text{vec}\,(Q\,(r_{1N},\varphi_{1N}))^T \\
\text{vec}\,(Q\,(r_{21},\varphi_{21}))^T & \text{vec}\,(Q\,(r_{22},\varphi_{22}))^T & \ldots & \text{vec}\,(Q\,(r_{21},\varphi_{21}))^T \\
\vdots & \vdots & \ddots & \vdots \\
\text{vec}\,(Q\,(r_{M1},\varphi_{M1}))^T & \text{vec}\,(Q\,(r_{M1},\varphi_{M1}))^T & \ldots & \text{vec}\,(Q\,(r_{MN},\varphi_{MN}))^T
\end{bmatrix}.
$$

Here $\text{vec}\,(Q)$ is a vector obtained by stacking together all the columns of the matrix Q from left to right. The coefficients in Eq. (17) do not have to be computed in real-time, but are calculated in advance.

The PSF in Eq. (16) is given by a quadratic, analytical formula in the aberration parameters $\boldsymbol{\alpha}$. This expression can be used to improve the efficiency of phase diversity algorithms as will be shown in the following two sections. Because the approximation in Eq. (3) does not conserve the energy of the wave, as this is problematic when taking differences of images, we have to use normalized PSFs in our implementations.

3. PHASE DIVERSITY AND THE ANALYTICAL COMPUTATION OF THE PSF GRADIENTS

In this section we assume that we have recorded one focal plane image and one diversity image (using a known defocus). If d_0, d_1 are the observed focused and defocused images and t_0, t_1 are the corresponding PSFs at the moment of exposure, then their relations with the object f are

$$
\begin{aligned}
d_0 &= t_0\,(\boldsymbol{\alpha}) * f + n_0 \\
d_1 &= t_1\,(\boldsymbol{\alpha}) * f + n_1,
\end{aligned} \tag{18}
$$

where $*$ denotes convolution, t_0 is given by Eq. (16), t_1 is given by

$$
t_1\,(\boldsymbol{\alpha}) = t_0\,(\boldsymbol{\alpha} + \boldsymbol{\alpha}_{d1}), \tag{19}
$$

$\boldsymbol{\alpha}_{d1}$ is a known quantity, in most cases a given defocus, and n_0, n_1 represent measurement noise. The error metric to be minimized in the PD algorithm (see for example[1-3]) measures the sum of the square errors in the difference between the observed images and the ones obtained from the reconstructions

$$
L = \sum_{u,v} |D_0 - \hat{F}\hat{T}_0|^2 + |D_1 - \hat{F}\hat{T}_1|^2, \tag{20}
$$

where $\hat{\cdot}$ denotes an estimated value, u and v spatial frequencies and F, D_0, D_1, T_0, T_1 are the Fourier transforms of f, d_0, d_1, t_0, t_1. The optimum restored object, F_M, is found by minimizing L considering known aberrations

$$
F_M = \left(|\hat{T}_0|^2 + |\hat{T}_1|^2 \right)^{-1} \left(D_0\hat{T}_0^* + D_1\hat{T}_1^* \right) \tag{21}
$$

Substituting the object from Eq. (21) in Eq. (20) we obtain the modified error metric

$$
L_M = \sum_{u,v} \left(|\hat{T}_0|^2 + |\hat{T}_1|^2 \right)^{-1} \left(D_1\hat{T}_0 - D_0\hat{T}_1 \right). \tag{22}
$$

We will call this minimization problem the PD algorithm.

One important step in the minimization of the metric in Eq. (22) is the computation of the gradients of the PSFs in Eq. (16) and Eq. (19). It is stated in[3] that the computation time of the algorithm is completely dominated by the time required to obtain these gradients. Using our approach, the gradients can be easily computed and included in the PD algorithm

$$
\begin{aligned}
\frac{\partial t_0(\boldsymbol{\alpha})}{\partial \alpha_i} &= \frac{\partial}{\partial \alpha_i} \left[C_0 + C_1 \left(I_N \otimes \boldsymbol{\alpha} \right) + Q_t \left(I_N \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha} \right) \right] \\
&= C_1 \left(I_N \otimes e_i \right) + Q_t \frac{\partial}{\partial \alpha_i} \left(I_N \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha} \right) \\
&= C_1 \left(I_N \otimes e_i \right) + Q_t \left[\frac{\partial (I_N \otimes \boldsymbol{\alpha})}{\alpha_i} \otimes \boldsymbol{\alpha} + \left(I_N \otimes \boldsymbol{\alpha} \right) \otimes \frac{\partial \boldsymbol{\alpha}}{\partial \alpha_i} \right] \\
&= C_1 \left(I_N \otimes e_i \right) + Q_t \left(I_N \otimes e_i \otimes \boldsymbol{\alpha} + I_N \otimes \boldsymbol{\alpha} \otimes e_i \right),
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
\frac{\partial t_1(\boldsymbol{\alpha})}{\partial \alpha_i} &= \frac{\partial t_0(\boldsymbol{\alpha})}{\partial \alpha_i} + Q_t \frac{\partial}{\partial \alpha_i} \left(I_N \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha}_{d1} + I_N \otimes \boldsymbol{\alpha}_{d1} \otimes \boldsymbol{\alpha} \right) \\
&= \frac{\partial t_0(\boldsymbol{\alpha})}{\partial \alpha_i} + Q_t \left(I_N \otimes e_i \otimes \boldsymbol{\alpha}_{d1} + I_N \otimes \boldsymbol{\alpha}_{d1} \otimes e_i \right),
\end{aligned}
\tag{24}
$$

where $e_i = \begin{bmatrix} \dots & 0 & 1_i & 0 & \dots \end{bmatrix}^T$. We can then compute the gradients in the Fourier domain as

$$
\frac{\partial}{\partial \alpha_i} \mathcal{F} \left\{ t_j \left(\boldsymbol{\alpha} \right) \right\} = \mathcal{F} \left\{ \frac{\partial t_j \left(\boldsymbol{\alpha} \right)}{\partial \alpha_i} \right\} \quad j = 1, 2, \quad i = 1, \dots, N_\alpha.
\tag{25}
$$

4. BILINEAR PROBLEM

For this part, we assume that we have obtained $N_\alpha + 2$ images, one focus plane image d_0, and diversity images d_k, $k = 1 \dots N_\alpha + 1$, with t_0 in Eq. (16) and t_k, $k = 1 \dots N_\alpha + 1$, such that

$$
t_k \left(\boldsymbol{\alpha} \right) = t_0 \left(\boldsymbol{\alpha} + \boldsymbol{\alpha}_{dk} \right), \ k = 1, \dots, N_\alpha + 1,
\tag{26}
$$

where $\boldsymbol{\alpha}_{dk}$ are known quantities. As in the previous section, the images are obtained as a convolution of the PSF with the object f in the presence of additive Gaussian noise n

$$
\begin{aligned}
d_0 &= t_0 \left(\boldsymbol{\alpha} \right) * f + n_0 \\
d_k &= t_k \left(\boldsymbol{\alpha} \right) * f + n_k, \ k = 1 \dots N_\alpha + 1.
\end{aligned}
\tag{27}
$$

We note that the assumption of Gaussian noise is adequate for solar and horizontal path imaging that have extended, relatively low-contrast objects, and not so good for night-time astronomical applications with extremely high-contrast images such as stars. For the latter, the space-varying photon noise with a Poisson distribution would be a more adequate assumption.

The approach we use here is to subtract from each diversity image the focus plane image and end up with linear relations both in the aberration parameters and the object. When we subtract the pairs of images we obtain

$$
\begin{aligned}
d_{dk} &= d_k - d_0 \\
&= \left[t_k \left(\boldsymbol{\alpha} \right) - t_0 \left(\boldsymbol{\alpha} \right) \right] * f + \left(n_k - n_0 \right) \\
&= \left[t_0 \left(\boldsymbol{\alpha} + \boldsymbol{\alpha}_{dk} \right) - t_0 \left(\boldsymbol{\alpha} \right) \right] * f + n_{k0},
\end{aligned}
\tag{28}
$$

where $n_{k0} := n_k - n_0$. We further use Eq. (16) to write expressions for the two PSFs in Eq. (28) and obtain

$$
\begin{aligned}
d_{dk} &= \left[C_1 \left(I_N \otimes \boldsymbol{\alpha}_{dk} \right) + Q_t \left(I_N \otimes \boldsymbol{\alpha}_{dk} \otimes \boldsymbol{\alpha}_{dk} + I_N \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\alpha}_{dk} + I_N \otimes \boldsymbol{\alpha}_{dk} \otimes \boldsymbol{\alpha} \right) \right] * f + n_{k0} \\
&= \left[C_1 \left(I_N \otimes \boldsymbol{\alpha}_{dk} \right) + Q_t \left(I_N \otimes \boldsymbol{\alpha}_{dk} \otimes \boldsymbol{\alpha}_{dk} \right) + 2 \sum_{i=1}^{N_\alpha} \alpha_i v_{ki} \right] * f + n_{k0},
\end{aligned}
\tag{29}
$$

where $v_{ki} = Q_t \left(I_N \otimes e_i \otimes \boldsymbol{\alpha}_{dk} \right)$, $i = 1 \ldots N_\alpha$ and $e_i = \begin{bmatrix} \ldots & 0 & 1_i & 0 & \ldots \end{bmatrix}^T$. With

$$p_k := C_1 \left(I_N \otimes \boldsymbol{\alpha}_{dk} \right) + Q_t \left(I_N \otimes \boldsymbol{\alpha}_{dk} \otimes \boldsymbol{\alpha}_{dk} \right), \tag{30}$$

we obtain

$$d_{dk} = \left(p_k + 2 \sum_{i=1}^{N_\alpha} \alpha_i v_{ki} \right) * f + n_{k0}. \tag{31}$$

We take the Fourier transform in Eq. (31) and obtain

$$\mathcal{F} \{d_{dk}\} = \left(\mathcal{F} \{p_k\} + 2 \sum_{i=1}^{N_\alpha} \alpha_i \mathcal{F} \{v_{ki}\} \right) \mathcal{F} \{f\} + \mathcal{F} \{n_{k0}\}. \tag{32}$$

Using the notations $D_{dk} := \mathcal{F} \{d_{dk}\}$, $P_k := \mathcal{F} \{p_k\}$, $V_{ki} := \mathcal{F} \{v_{ki}\}$, $F := \mathcal{F} \{f\}$, $N_{k0} := \mathcal{F} \{n_{k0}\}$ (of dimension $M_\mathcal{F} \times N_\mathcal{F}$), we obtain

$$D_{dk} = \left(P_k + 2 \sum_{i=1}^{N_\alpha} \alpha_i V_{ki} \right) \odot F + N_{k0}, \tag{33}$$

where $\odot$ denotes point-wise multiplication. We would like to rewrite Eq. (33) in a form where the linear variables are $\begin{bmatrix} 1 & \boldsymbol{\alpha}^T \end{bmatrix}^T$ and $\mathrm{vec}\,(F)$. To this end, we vectorize Eq. (33) (we stack together all the columns) in order to obtain

$$\mathrm{vec}\,(D_{dk}) = \left[\mathrm{vec}\,(P_k) + 2 \sum_{i=1}^{N_\alpha} \alpha_i \mathrm{vec}\,(V_{ki}) \right] \odot \mathrm{vec}\,(F) + \mathrm{vec}\,(N_{k0}). \tag{34}$$

In Eqs. (33) and (34), D_{dk} contain the measured data and P_k and V_{ki} depend on the diversities we impose and can be precomputed. The quantities that we want to estimate are α_i, $i = 1, \ldots, N_\alpha$ and F. It is easy to see that each equation of the system in Eq. (34) can be expressed as a bilinear equation in the unknowns $\begin{bmatrix} 1 & \boldsymbol{\alpha}^T \end{bmatrix}^T$ and $\mathrm{vec}\,(F)$ as mentioned above. We show this for the equation corresponding to the element number j

$$
\begin{aligned}
\mathrm{vec}\,(D_{dk})_j &= \left[\mathrm{vec}\,(P_k)_j + 2 \sum_{i=1}^{N_\alpha} \alpha_i \mathrm{vec}\,(V_{ki})_j \right] \mathrm{vec}\,(F)_j + \mathrm{vec}\,(N_{k0})_j \\
&= \begin{bmatrix} 1 & \boldsymbol{\alpha}^T \end{bmatrix} \begin{bmatrix} \mathrm{vec}\,(P_k)_j \\ 2\mathrm{vec}\,(V_{k1})_j \\ 2\mathrm{vec}\,(V_{k2})_j \\ \vdots \\ 2\mathrm{vec}\,(V_{kN_\alpha})_j \end{bmatrix} \mathrm{vec}\,(F)_j + \mathrm{vec}\,(N_{k0})_j, \ j = 1, \ldots, M_\mathcal{F} N_\mathcal{F}.
\end{aligned}
\tag{35}
$$

Let $A(j)$, $j = 1, \ldots, M_\mathcal{F} N_\mathcal{F}$ be the $(N_\alpha + 1) \times M_\mathcal{F} N_\mathcal{F}$ sparse matrix with non-zero elements only in column j

$$A(j) = \begin{bmatrix} 0 & \cdots & 0 & \mathrm{vec}\,(P_k)_j & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 2\mathrm{vec}\,(V_{k1})_j & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 2\mathrm{vec}\,(V_{k2})_j & 0 & \cdots & 0 \\ & & & \vdots & & & \\ 0 & \cdots & 0 & \underbrace{2\mathrm{vec}\,(V_{kN_\alpha})_j}_{j} & 0 & \cdots & 0 \end{bmatrix}. \tag{36}$$

Using Eq. (36) in Eq. (35) we obtain

$$\mathrm{vec}\,(D_{dk})_j = \begin{bmatrix} 1 & \boldsymbol{\alpha} \end{bmatrix}^T A(j) \mathrm{vec}\,(F) + \mathrm{vec}\,(N_{k0})_j. \tag{37}$$

With the notations $\boldsymbol{d} := \mathrm{vec}\,(D_{dk})$, $\boldsymbol{y} := \begin{bmatrix} 1 & \boldsymbol{\alpha}^T \end{bmatrix}^T$, $\boldsymbol{x} := \mathrm{vec}\,(F)$, $\boldsymbol{n} := \mathrm{vec}\,(N_{k0})$, we obtain the bilinear system of equations

$$d_j = \boldsymbol{y}^{*T} A(j) \boldsymbol{x}^* + \boldsymbol{n}_j, \ j = 1, \ldots, M_\mathcal{F} N_\mathcal{F}, \tag{38}$$

equivalent to the one appearing in[14] , where the superscript * denotes the true values. We obtain the same type of equations for the $N_\alpha + 1$ pairs of focus plane and diversity images. Stacking together all these bilinear equations we obtain a system of bilinear equations. The bilinear equations can be written in the following form

$$d_j = \boldsymbol{y}^T A\left(j\right) \boldsymbol{x}, \; j = 1, \ldots, \left(N_\alpha + 1\right) M_\mathcal{F} N_\mathcal{F}, \tag{39}$$

and the goal is to find a solution in the least squares (LS) sense

$$\left(\boldsymbol{x}, \boldsymbol{y}\right) = \arg\min_{\boldsymbol{x}, \boldsymbol{y}} \frac{1}{\left(N_\alpha + 1\right) M_\mathcal{F} N_\mathcal{F}} \sum_{i=1}^{\left(N_\alpha + 1\right) M_\mathcal{F} N_\mathcal{F}} \left[d_i - \boldsymbol{y}^T A\left(i\right) \boldsymbol{x}\right]^2, \tag{40}$$

with $\left(\boldsymbol{x}, \boldsymbol{y}\right)$ as close as possible to the true values $\left(\boldsymbol{x}^*, \boldsymbol{y}^*\right)$.

In[14] , a comparison of three methods to solve systems of bilinear equations is presented. The first is an iterative algorithm whose convergence is not proved in the general case, the second is a so-called over-parametrization method or TSA, which implies solving as a first step a least-squares estimate using an over-parametrized vector and then projecting the solution onto the class of bilinear equations via the SVD, and the third, a method which solves the nonlinear LS problem (40) using a direct optimization technique. We choose to solve this problem using the second approach. The TSA was initially proposed in[16] to identify a bilinear Hammerstein-Wiener system. It is an optimal two-stage algorithm.

In the following we use the notations introduced in Eq. (38), which can be written in the following scalar form

$$\sum_{i=1}^{N_\alpha+1} \sum_{k=1}^{M_\mathcal{F} N_\mathcal{F}} a_{ik}\left(j\right) y_i x_k = d_j \text{ for } j = 1, \ldots, \left(N_\alpha + 1\right) M_\mathcal{F} N_\mathcal{F}. \tag{41}$$

The first step in the proposed algorithm is to write the bilinear system as an over-parametrized linear system. We do this by defining a new vector consisting of all the products of the unknowns $\boldsymbol{x}$ and $\boldsymbol{y}$. Thus, we define

$$\begin{aligned}
\boldsymbol{\theta} &= \left(y_1 x_1, \ldots, y_1 x_{M_\mathcal{F} N_\mathcal{F}}, y_2 x_1, \ldots, y_2 x_{M_\mathcal{F} N_\mathcal{F}}, \ldots, y_{N_\alpha+1} x_1, \ldots, y_{N_\alpha+1} x_{M_\mathcal{F} N_\mathcal{F}}\right) \\
&= \left(\theta_1, \theta_2, \ldots, \theta_{\left(N_\alpha+1\right)\left(M_\mathcal{F} N_\mathcal{F}\right)}\right)
\end{aligned} \tag{42}$$

as the over-parametrized unknown vector and Θ_{yx} as the $\left(N_\alpha + 1\right) \times \left(M_\mathcal{F} N_\mathcal{F}\right)$ matrix form of $\boldsymbol{\theta}$

$$\Theta_{yx} = yx^T = \left[\begin{array}{cccc}
y_1 x_1 & y_1 x_2 & \cdots & y_1 x_{M_\mathcal{F} N_\mathcal{F}} \\
y_2 x_1 & y_2 x_2 & \cdots & y_2 x_{M_\mathcal{F} N_\mathcal{F}} \\
\vdots & \vdots & \ddots & \vdots \\
y_{N_\alpha+1} x_1 & y_{N_\alpha+1} x_2 & \cdots & y_{N_\alpha+1} x_{M_\mathcal{F} N_\mathcal{F}}
\end{array}\right]. \tag{43}$$

The matrix Θ_{yx} is a rank 1 matrix. We also define the data matrix

$$\Phi_N = \left[\begin{array}{cccc}
a_{11}\left(1\right) & a_{12}\left(1\right) & \cdots & a_{\left(N_\alpha+1\right)\left(M_\mathcal{F} N_\mathcal{F}\right)}\left(1\right) \\
a_{11}\left(2\right) & a_{12}\left(2\right) & \cdots & a_{\left(N_\alpha+1\right)\left(M_\mathcal{F} N_\mathcal{F}\right)}\left(2\right) \\
\vdots & \vdots & \ddots & \vdots \\
a_{11}\left(N\right) & a_{12}\left(N\right) & \cdots & a_{\left(N_\alpha+1\right)\left(M_\mathcal{F} N_\mathcal{F}\right)}\left(N\right)
\end{array}\right].$$

where $N := \left(N_\alpha + 1\right) M_\mathcal{F} N_\mathcal{F}$.

The system of equations in Eq. (41) can now be written as a linear system

$$\Phi_N \boldsymbol{\theta} = \boldsymbol{d}. \tag{44}$$

We use the following TSA[14] to obtain an estimate of $\boldsymbol{y}^*$ and $\boldsymbol{x}^*$:

Step 1. Compute the LS estimate

$$\hat{\boldsymbol{\theta}} = \left(\Phi_N^T \Phi_N\right)^{-1} \Phi_N^T \boldsymbol{d}; \tag{45}$$

Step 2. Construct $\hat{\Theta}_{yx}$ from $\hat{\boldsymbol{\theta}}$ and let

$$\hat{\Theta}_{yx} = \sum_{i=1}^{\min(N_\alpha+1, M_\mathcal{F} N_\mathcal{F})} \sigma_i \mu_i \nu_i^T \tag{46}$$

be its SVD, where μ_i, $i = 1, \ldots, N_\alpha + 1$ and ν_i, $i = 1, \ldots, M_\mathcal{F} N_\mathcal{F}$ are $N_\alpha + 1$, $M_\mathcal{F} N_\mathcal{F}$ -dimensional orthonormal vectors, respectively;

Step 3. Let s_μ denote the sign of the first nonzero element of μ_1. Define the estimates as

$$\hat{\boldsymbol{x}} = s_\mu \sigma_1 \nu_1, \ \hat{\boldsymbol{y}} = s_\mu \mu_1. \tag{47}$$

The LS solution $\hat{\boldsymbol{\theta}}$ in the first step of the above algorithm gives an estimate of the over-parametrized unknown vector $\boldsymbol{\theta}$. The second step computes the optimum $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ as solutions of

$$\arg \min_{\boldsymbol{x} \in \mathbb{R}^{M_\mathcal{F} N_\mathcal{F}}, \boldsymbol{y} \in \mathbb{R}^{N_\alpha+1}} \left\| \hat{\Theta}_{yx} - \boldsymbol{y} \boldsymbol{x}^T \right\|_F^2. \tag{48}$$

The estimates $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are provided by the SVD of $\hat{\Theta}_{yx}$, which is proved to be optimal in lemma A.1 in[16] . Because the second step is globally optimal over all vector space, the convergence of this algorithm depends on the convergence of the over-parametrized LS problem in the first stage. The system in step 1 is a large-scale sparse linear system and can be solved using specific methods that take advantage of this structure (see for example the LSQR algorithm in[22]). LSQR is the implementation of a conjugate-gradient type method for solving sparse linear equations and sparse least-squares problems

$$\min_{\boldsymbol{\theta}} \|\Phi_N \boldsymbol{\theta} - \boldsymbol{d}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_2^2, \tag{49}$$

where the matrix Φ_N may be square or rectangular (over-determined or under-determined), and may have any rank. The scalar λ is a damping parameter. If $\lambda > 0$, the solution is "regularized" in the sense that a unique solution always exists, and $\|\boldsymbol{\theta}\|_2$ is bounded. Theoretically, LSQR converges in N steps, but in practice a solution is found in a considerably smaller number of iterations. Such an iterative method can also be implemented in a parallel configuration as shown in[23] in order to speed up computations. If the first stage converges, then $\hat{\boldsymbol{x}} \to \boldsymbol{x}^*$, $\hat{\boldsymbol{y}} \to \boldsymbol{y}^*$ in the second stage.

5. SIMULATIONS

The simulations presented here consider isoplanatic image formation, namely the PSF is considered space invariant over the entire field of view. We give simulation results for two algorithms: the bilinear algorithm presented in Section 4 (bilinear) and the PD algorithm in Eq. (22). The image formation process in Eq. (27) is considered without additive noise. We assume that the aberration function in Eq. (2) is sampled on a 32×32 grid. Two cases are presented, were the rms error of the wavefronts is 0.5 rad (w_1) and 0.3 rad (w_2). We further approximate the wavefront with a series as in Eq. (2) using 14 Zernike coefficients. As stated in Section 2, the coefficients in Eq. (17) of the PSF in Eq. (16) are precomputed using a grid of 32×32 points over a pupil of diameter $D = 1$ m. The wavelength considered is 550 nm. We present simulations considering a 64×64 pixel scene (which is a low-contrast synthetic image of the solar surface) (d_1) and a 64×64 pixel point source (we will denote it by d_2) and .

We simulate our data by first generating wavefronts with Kolmogorov statistics with $r_0 = 0.2$ m. Then we perturb each Zernike mode, starting with defocus, by a known quantity, more exactly we use 12 diversities. The values of the diversity shapes used are 0.3 rad corresponding to each Zernike mode. Next, we compute the corresponding PSFs in each case and perform the convolution (see Eq. (27)) to obtain the images. In order to remove the edge effects that appear because of the fast Fourier transforms, we use a Hanning window to apodize the images. We also subtract a constant intensity from the image such that the average intensity of the apodized image is zero. We simulate 50 realizations of the wavefront and compute the estimates (we present here only

the wavefront estimates, although the computation of object estimates is inherent to the method). The bilinear solution is biased, so we computed this bias and subtract it from the estimates. The results are presented as follows. We solve the system of bilinear equations and the PD algorithm using as stopping criteria for LSQR a tolerance of $1e-2$ or a maximum of 10 iterations. Fig. 1 plots the mean and variance of the absolute values of the differences between the real parameters and the estimated ones in four cases (from left to right): (w_2, d_1), (w_1, d_1), (w_2, d_2), and (w_1, d_2). In Fig. 2 we plot the computational times and ratios of computational times in the four cases.

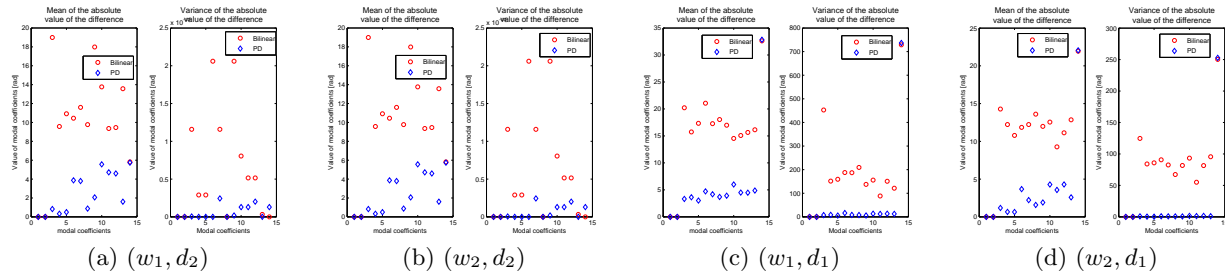| (a) (w_1, d_2) | (b) (w_2, d_2) | (c) (w_1, d_1) | (d) (w_2, d_1) |

Figure 1: Mean (left) and variance (right) of the absolute values of the differences between estimated and real Zernike coefficients (in radians)

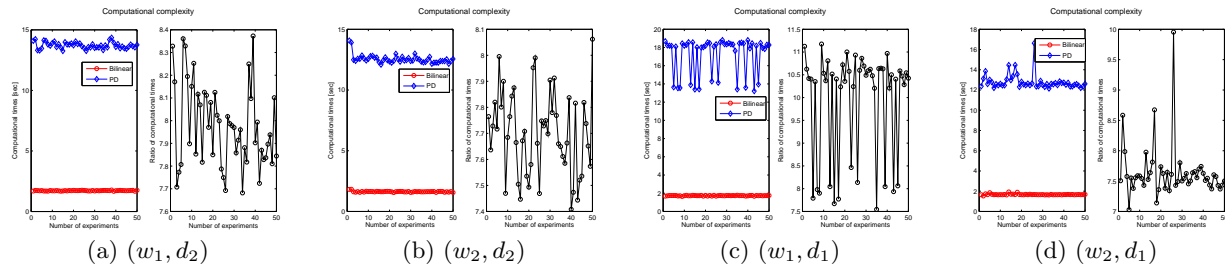| (a) (w_1, d_2) | (b) (w_2, d_2) | (c) (w_1, d_1) | (d) (w_2, d_1) |

Figure 2: (left) Computational times; (right) Ratios of computational times

The solution of the bilinear problem can be used as an initial estimate for the PD algorithm. In Fig. 3 we can see computational times for PD when using as initial estimate zero and the bilinear solution (more precisely, on the left we plot the computational time of PD with initial value zero and on the right the sum of the computational times corresponding to computing the bilinear solution and to the PD algorithm having as initial value the previously computed solution).
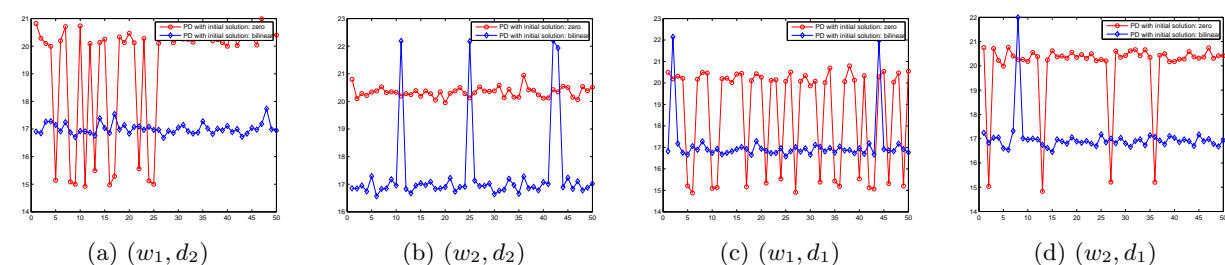

| (a) (w_1, d_2) | (b) (w_2, d_2) | (c) (w_1, d_1) | (d) (w_2, d_1) |

Figure 3: Computational times [sec] for PD with initial value zero (red), bilinear (blue) for 50 experiments

6. CONCLUSIONS

We have developed a new method for estimating aberration parameters and an unknown object estimation in case of imaging through turbulence. This method is based on the linearization of the wavefront for small values of the aberrations. A systems of bilinear equations is solved to obtain the estimates. Because this method is

explicit in the unknown object and it involves solving a large-scale linear system of equation, it is very well suited either for wavefront estimation with small objects or post-processing. In the case of small objects, the method shows potential to be used as a wavefront sensor in real time control schemes in adaptive optics.

REFERENCES

[1] Gonsalves, R. A., "Phase retrieval and diversity in adaptive optics," *Optical Engineering* **21**, 829–832 (1982).

[2] Paxman, R. G., Schultz, T. J., and Fienup, J. R., "Joint estimation of object and aberrations by using phase diversity," *Journal of Optical Society of America A* **9**, 1072–1085 (1992).

[3] Lofdahl, M. and Scharmer, G. B., "Wavefront sensing and image restoration from focused and defocused solar images," *Astronomy & Astrophysics Supplement Series* **107**, 243–264 (1994).

[4] Thelen, B., Paxman, R., Carrara, D., and Seldin, J., "Maximum a *posteriori* estimation of fixed aberrations, dynamic aberrations, and the object from phase-diverse speckle data," *Journal of Optical Society of America A* **16**(5), 1016–1025 (1999).

[5] Blanc, A., Mugnier, L., and Idier, J., "Marginal estimation of aberrations and image restoration by phase diversity," *Journal of Optical Society of America A* **20**(6), 1035–1045 (2003).

[6] Dolne, J. J., Menicucci, P., Miccolis, D., Widen, K., Seiden, H., Vachss, F., and Schall, H., "Real time phase diversity advanced image processing and wavefront sensing," *SPIE* **6712** (2007).

[7] C.R. Vogel, Chan, T., and Plemmons, R., "Fast algorithms for phasediversity-based blind deconvolution," *Adaptive Optical System Technologies* **3353**, 994–1005 (1998).

[8] Lofdahl, M., Duncan, A., and Scharmer, G., "Fast phase diversity wavefront sensor for mirror control," *Adaptive Optical System Technologies* **3353**, 952–963 (1998).

[9] Scharmer, G., "Object-independent fast phase-diversity," *High Resolution Solar Physics: Theory, Observations and Techniques* **183**, 330–341 (1999).

[10] Gonsalves, R., "Small-phase solution to the phase-retrieval problem," *Optics letters* **26**(10), 684–685 (2001).

[11] Wild, W. J., "Linear phase retrieval for wave-front sensing," *Optics letters* **23**(8), 573–575 (1998).

[12] Mocœur, I., Mugnier, L., and Cassaing, F., "Analytical solution to the phase-diversity problem for real-time wavefront sensing," *Optics letters* **34**(22), 3487–3489 (2009).

[13] Cohen, S. and Tomasi, C., "Systems of bilinear equations," tech. rep., Department of Computer Science, Stanford University (1992).

[14] Bai, E.-W. and Liu, Y., "On the least squares solution of a system of bilinear equations," *Proc. 44th CDC and ECC* (December 2005).

[15] Wang, J., Zhang, Q., and Ljung, L., "Revisiting hammerstein system identification through the two-stage algorithm for bilinear parameter estimation," *Automatica* **45**, 2627–2633 (2009).

[16] Bai, E., "An optimal two-stage identification algorithm for hammerstein-wiener nonlinear systems," *Automatica* **14**, 333–338 (1998).

[17] Born, M. and Wolf, E., [*Principles of Optics*], Pergamon, Oxford, UK (1985).

[18] Noll, R., "Zernike polynomials and atmospheric turbulence," *Journal of Optical Society of America* **66**(3), 207–211 (1976).

[19] Braat, J., Dirksen, P., and Janssen, A., "Assessment of an extended nijboer-zernike approach for the computation of optical point-spread functions," *Journal of the Optical Society of America A* **19**(5), 858–870 (2002).

[20] Janssen, A., "Extended nijboer-zernike approach for the computation of optical point-spread functions," *Journal of the Optical Society of America A* **19**(5), 849–857 (2002).

[21] van der Avoort, C., Braat, J., Dirksen, P., and Janssen, A., "Aberration retrieval from the intensity point-spread function in the focal region using the extended nijboer-zernike approach," *Journal of Modern Optics* **52**(12), 1695–1728 (2005).

[22] Paige, C. and Saunders, M., "Lsqr: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software* **8**(1), 43–71 (1982).

[23] Jing-Song, L., Fu-Tian, L., Jun, L., and Tian-Yao, H., "Parallel lsqr algorithms used in seismic tomography," *Chinese Journal of Geophysics* **49**(2), 483–488 (2006).