

Iterative method for wave front reconstruction from adaptively formed images of an arbitrary extended source

G.L. Degtyarev, A.V. Makhan'ko, S.M. Chernyavskii, and A.S. Chernyavskii

A.N. Tupolev Kazan State Technical University

Received June 26, 2003

An iterative method is proposed for reconstruction of a wave phase at the exit pupil of an optical system from the intensity distributions in the images of an unknown extended incoherent source in some planes parallel to the focal plane. It is assumed that at each of the iterations the phase is corrected for the value obtained in accordance with this method.

Introduction

The problem of restoring the phase of a field in front of a lens from the field amplitude and intensity in several planes parallel to the focal plane was investigated by many authors.¹ Seeking a solution to this problem can be reduced to the solving the following equation:

$$F[GG_0(z)] = g(x,y,z), \quad |g(x,y,z)|^2 = I(x,y,z), \quad (1)$$

where

$$G(\xi,\eta) = A(\xi,\eta)e^{i2\pi\Phi(\xi,\eta)}$$

is the wave function of the field in front of the lens with the known amplitude $A(\xi,\eta)$, $(\xi,\eta) \in \Omega$ is the lens area, and unknown function of wave front aberrations Φ ; F denotes the two-dimensional Fourier transform in terms of the variables (ξ,η) ; z is the axial coordinate of the image plane counted off from the focal plane of the lens, and $G_0 = e^{-iz(\xi^2+\eta^2)/2}$ is the phase factor of defocusing. The function $I(x,y,z)$ is the known intensity within the area ω of the plane z_s of the image recording $s = \overline{1,S}$.

From the theoretical point of view, the knowledge of the amplitude A is excessive. The wave function G can be reconstructed from the image within the volume that includes the focal plane.² From the viewpoint of practice, this theoretical result is important, because it makes, based on the solution of the wave problem for Eq. (1) the ground for seeking ways of developing constructively simple wave front sensors.

The numerical experiments on recovering the aberration function Φ from the wave function G determined by Eq. (1) from the noisy intensity in a limited area ω show that the recovered function Φ can significantly differ from the "real" aberration function for this experiment.

An adaptive-optics system (AOS) intended for compensating for the wave front aberrations provides for iteratively recovering wave front from the

solution of phase problem for Eq. (1) [Ref. 3]. The function $\tilde{\Phi}$ found from the solution of the phase problem is taken as an estimate of the function Φ , for which the wave front is being corrected. Then the amplitude A and the intensity I are measured repeatedly, the phase problem (1) is solved for them, the wave front correction is performed for these new values, and so on. This method will be referred to as the method of wave front reconstruction from adaptively formed images of the source.

The possibility of applying this approach to wave front reconstruction by Eq. (1) from only the intensity is considered in Section 1.

As applied to astronomic objects, the problem (1) is the problem of wave front reconstruction from images of a point source. When observing an arbitrary portion of the sky, the probability of finding the natural reference source is low.⁴ That is why the application of laser guide stars for wave front measurements is actively investigated.⁵ An alternative approach is the wave front measurements using images of an unknown extended source. In this case, Eq. (1) should be replaced by the convolution equation

$$h(x,y,z) * I_0(x,y) = I(x,y,z), \quad h = |g|^2, \quad z = z_s, \quad (2)$$

where I_0 is the unknown intensity distribution over the source. The source is assumed to be in the isoplanatic region of the optical system.

The traditional method to exclude I_0 from Eq. (2) consists in formation of the equalities in the frequency region:

$$\frac{H(\xi,\eta,z_s)}{H(\xi,\eta,z_1)} = \frac{J(\xi,\eta,z_s)}{J(\xi,\eta,z_1)}, \quad s = \overline{2,S}, \quad (3)$$

where H and J are the inverse two-dimensional Fourier transforms of the functions h and I in terms of the variables x and y .

Equalities (3) play the role of equalities (1) in the problem of wave front reconstruction from the corresponding images. The transition from Eq. (2) to Eq. (3) is possible, when the images of the source are

known. In reality, this condition is not fulfilled, and therefore one faces the problem of extending the image in the area outside the image at unknown l_0 and h . Section 2 considers the method of wave front reconstruction from adaptively formed images based on the solution of Eq. (2).

1. Wave front reconstruction from adaptively formed images of a point source

The problem of determination of the wave function G from Eq. (1) will be considered in the geometric interpretation as a problem on determining common point of the given sets in the Hilbert space. The Hilbert space H is taken to be the set of complex-valued functions $g(x,y,z)$ defined on the direct product $Oxy \times \{z_1, \dots, z_s\}$ with the summable square in terms of the variables x and y and the norm

$$\|g\|^2 = \sum_{s=1}^S \|g(z_s)\|^2 = \sum_{s=1}^S \int_{-\infty}^{+\infty} \int |g(x,y,z_s)|^2 dx dy.$$

Introduce two sets in this space:

$$V_1 = \{g : g = F[GG_0(z)], \text{ supp } G = \Omega\}$$

and

$$V_2 = \{g : |g(x,y,z_s)| = a_2(x,y,z_s) = l^{1/2}(x,y,z_s); (x,y) \in \omega, s = \overline{1,S}\}.$$

The functions $g \in V_1$ and G are related by a continuous biunique correspondence

$$G = F^{-1}[g(z)]G_0^*(z),$$

where the asterisk $*$ denotes complex conjugation. Therefore, the wave problem can be formulated as finding the function

$$g \in V_0 = V_1 V_2. \tag{4}$$

The restriction on the form of V_1 was introduced for the first time by Fienup⁶ in the problem on a source reconstruction from its carrier and the amplitude of its Fourier spectrum.

Any point in expression (4) is connected with the point of minimum of the approach functional

$$J(g, g_1, g_2) = \alpha_1 \|g - g_1\|^2 + \alpha_2 \|g - g_2\|^2, \tag{5}$$

$$\alpha_1 + \alpha_2 = 1, \alpha_1, \alpha_2 > 0.$$

The minimum of the functional is achieved at $g_1 = g_2 = g \in V_0$, therefore the functions providing for the minimum of the functional (5) determine the wave function G of the problem (1).

If the functional (5) is minimized by the coordinate descent method first in terms of g_1 and g_2 and then in terms of g , then we obtain the iterative algorithm⁷:

$$g_{10} \in V_1, g_{20} \in V_2, g_0 = \alpha_1 g_{10} + \alpha_2 g_{20},$$

$$g_{1n+1} = P_1 g_n, g_{2n+1} = P_2 g_n, \tag{6}$$

$$g_{n+1} = g_n + \lambda(\alpha_1 g_{1n+1} + \alpha_2 g_{2n+1} - g_n), 0 < \lambda < 2.$$

Here P_1 and P_2 are the operators of projection onto the sets V_1 and V_2 , defined according to the condition

$$\|g - P_k g\| = \inf_{g_k \in V_k} \|g - g_k\|, k = 1, 2.$$

Since

$$\|g - g_1\|^2 = \sum_{s=1}^S \|g(z_s) - g_1(z_s)\|^2 =$$

$$= \sum_{s=1}^S \|F^{-1}g(z_s) - GG_0(z_s)\|^2 = \sum_{s=1}^S \|G_0^*(z_s)F^{-1}g(z_s) - G\|^2,$$

$$g_{1n+1} = P_1 g_n = F^{-1}[GG_0(z)],$$

where

$$G = \frac{1}{S} \sum_{s=1}^S G_0^*(z_s)F^{-1}g_n(z_s)$$

at Ω and $G = 0$ outside Ω .

The approximation

$$g_{2n+1} = P_2 g_n =$$

$$= \begin{cases} \frac{a_2(x,y,z)g_n(x,y,z)}{|g_n(x,y,z)|} & \text{for } g_n(x,y,z) \neq 0, (x,y) \in \omega, \\ g_{2n}(x,y,z) & \text{for } g_n(x,y,z) = 0, (x,y) \in \omega, \\ g_n(x,y,z) & \text{for } (x,y) \notin \omega. \end{cases}$$

If we take the approach functional in the form

$$J_1(g_1, g_2) = \|g_1 - g_2\|,$$

then its minimum is achieved at $g_1 = g_2 \in V_0$.

The method of coordinate descent first in terms of g_1 and then in terms of g_2 leads to the Gershberg–Saxton algorithm:

$$g_{10} \in V_1, g_{20} \in V_2,$$

$$g_{2n+1} = g_{2n} + \lambda_2(P_2 g_{1n} - g_{2n}), \tag{7}$$

$$g_{1n+1} = g_{1n} + \lambda_1(P_1 g_{2n+1} - g_{1n}), \lambda_1, \lambda_2 \in (0,2).$$

2. Numerical simulation

The phase of the unknown field was defined by a segment of the series over Zernike polynomials⁸:

$$2\pi\Phi = 2\pi \sum R_n^m(\rho)[A_n^m \cos(m\theta) + B_n^m \sin(m\theta)], \tag{8}$$

which accounted for the following modes: tilt (A_1^1, B_1^1), coma ($A_3^1, B_3^1, A_3^3, B_3^3$), defocusing and

spherical aberration (A_2^0, A_4^0) , and astigmatism $(A_2^2, B_2^2, A_4^2, B_4^2)$; (ρ, θ) are the polar coordinates of the beam aperture, ρ is the polar radius related to the beam radius. The values of the modes were simulated by the generator of normal random numbers and restricted in the absolute value to the value of α . The amplitude of the unknown field was simulated by the function $A = Ce^{-\rho^2/2}$, where C is the normalization coefficient, at which the integral of the squared amplitude equals unity.

The position of the planes, at which the intensity was recorded, was determined by the axial coordinate z , varying within the range $|z| < 8$. In this range, defocusing in the absence of other aberrations does not distort significantly the image of a point source.⁸ The intensity corresponding to the function G chosen was calculated by Eq. (1) and distorted by the additive noise with the variance σ^2 .

The geometry of the beam aperture and the image intensity give rise to the two sets: V_1 , V_2 , and their intersection V_0 . To find the point $g \in V_0$, a combined algorithm was used: a half of all iterations (20–30 iterations) were performed by the algorithm (6) with the initial approximation $g_{10} = F[GG_0(z)]$,

$$G = \frac{1}{\sqrt{\pi}} \text{ and } g_{20} = a_2(x, y, z) \text{ inside } \omega \text{ and } g_{20} = g_{10}$$

outside ω . Other iterations were continued by the algorithm (7). This combined algorithm provided for a closer approach to the minimum of the norm $\|g_1 - g_2\|$, than each of the algorithms (6) and (7) did at the same total number of iterations. The point g_1 obtained at the final iteration by the combined algorithm was taken to be an estimate of the point from V_0 . The components entering into Eq. (8) were separated from the phase of the function G_1 corresponding to the function g_1 .

Then the phase (8) was corrected for by the value of the components separated from the phase G_1 (partial correction). The intensity is calculated according to Eq. (1) from the corrected phase (8). The new intensity was distorted by the additive noise with the same variance σ^2 . This new intensity determined the set V_2 . The combined algorithm consisting of the algorithms (6) and (7) was used to determine the new estimate $g_1 \in V_0$, and so on.

The quality of compensating for the modes was characterized by the maximum of intensity $\max I = \max_{x,y,z} I(x, y, z)$ in all planes. The described method of wave front reconstruction and compensation in AOS is similar to the method of external tuning of AOS using the functional of image quality. The difference is that in the latter case the modes are determined from the intensity distributions in the images of the source in different planes $z = z_s$, $s = \overline{1, S}$.

Consider an example of wave front reconstruction from images of a point source in three

planes $z = 0, -5$, and 5 and the parameters $\alpha = 0.2$, $\sigma = 0.06$. The initial values of the even and odd modes and their values in the process of partial correction are tabulated below.

Odd modes (tilt and comas) – six modes

-0.0865	-0.2000	0.0251	0.0575	-0.2000	0.2000
-0.0435	0.0011	-0.0069	-0.0350	-0.0274	0.1017
-0.0143	-0.0147	-0.0103	-0.0043	-0.0459	-0.0055
-0.0059	-0.0075	0.0121	0.0024	-0.0019	-0.0050

Even modes (defocusing, spherical aberration, and astigmatism) – six modes

0.2000	-0.0075	0.0655	0.0349	-0.0373	0.1452
0.0709	0.0652	0.0181	0.0764	-0.0463	0.1042
0.0155	0.0356	-0.0053	0.0079	-0.0325	-0.0181
0.0103	0.0153	-0.0135	-0.0134	-0.0286	0.0113

The initial $\max I$ value and the values changed in the process of partial correction of the modes are respectively: 1.2742, 2.4042, 2.8437, and 2.8792. The examples presented below correspond to the case with the increased noise: $z = 0, -5$, and 5 ; $\alpha = 0.2$, $\sigma = 0.1$.

Odd modes (tilt and comas) – six modes

-0.0865	-0.2000	0.0251	0.0575	-0.2000	0.2000
-0.0253	-0.0034	0.0061	-0.0597	-0.0368	0.1915
-0.0037	-0.0219	-0.0064	-0.0637	0.0150	-0.0528
0.0058	-0.0012	-0.0040	-0.0083	0.0142	0.0411

Even modes (defocusing, spherical aberration, and astigmatism) – six modes

0.2000	-0.0075	0.0655	0.0349	-0.0373	0.1452
0.0713	0.0730	0.0437	0.0584	0.0016	0.1121
0.0126	0.0600	-0.0318	0.0022	0.0014	-0.0088
0.0061	0.0070	0.0131	-0.0343	-0.0375	-0.0141

In this case, the initial $\max I$ value and the values changed in the process of partial correction for the modes are: 1.3568, 2.2109, 2.7676, and 2.8556.

The same example in the absence of noise: $z = 0, -5$, and 5 ; $\alpha = 0.2$, $\sigma = 0$, leads to the following change of the modes in the process of their correction:

Odd modes (tilt and comas) – six modes

-0.0865	-0.2000	0.0251	0.0575	-0.2000	0.2000
-0.0172	-0.0004	0.0082	-0.0256	-0.0009	0.0082
0.0001	-0.0004	-0.0004	-0.0003	0.0003	0.0008
-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0000

Even modes (defocusing, spherical aberration, and astigmatism) – six modes

0.2000	-0.0075	0.0655	0.0349	-0.0373	0.1452
0.0298	0.0187	0.0180	0.0063	-0.0262	-0.0085
0.0006	0.0017	-0.0011	0.0000	0.0010	-0.0001
0.0001	0.0001	0.0000	-0.0000	-0.0001	-0.0000

The initial $\max I$ value and the values changed in the process of partial correction of the modes are respectively: 1.2004, 2.8250, 2.8899, and 2.8900.

These examples correspond to the distributions of the initial modes, which can be successfully

compensated for by AOS. Such distributions of the initial modes were most frequent in simulation. However, some distributions of the initial modes could not be completely compensated for, but the value of $\max I$ always increased markedly.

3. Wave front reconstruction from adaptively formed images of an unknown extended source

It is proposed to consider equality (2) as an equation connecting three unknowns: $I(x, y, z)$, $I_0(x, y)$, and $g(x, y, z)$. Equality (2) defines the set

$$V = \{(I, I_0, g) : I(z) = I_0 * |g(z)|^2\}.$$

The measurements and the *a priori* data on the functions define another one set

$$V_1 = \{(I, I_0, g) : I = I_{\text{meas}} + n \text{ in } \omega \times \{r_1, \dots, r_S\}, n \in N; \\ I_0 \geq 0 \text{ and } \text{supp } I_0 = \omega_0; \\ g = F[GG_0(z)], \text{ supp } G = \Omega\},$$

where N is the set determining the *a priori* statistical properties of noise in the intensity measured in the image; ω_0 is the area in the object plane, in which the source intensity affects the image in the area ω . In the geometric interpretation, the problem of wave front reconstruction from incomplete images of an unknown source can be reduced to determination of a point

$$(I, I_0, g) \in VV_1. \tag{9}$$

Introduce three Hilbert spaces: spaces of real functions and complex-valued functions defined in $oxy \times \{z_1, \dots, z_S\}$ and real functions in oxy . All the functions are modulo square-integrable. On the direct product of these spaces we define the Hilbert space H with the norm

$$\|(I, I_0, g)\|^2 = \sum_{s=1}^S \|I(z_s)\|^2 + \|I_0\|^2 + \sum_{s=1}^S \|g(z_s)\|^2 = \\ = \sum_{s=1}^S \iint_{-\infty}^{\infty} I^2(x, y, z_s) dx dy + \iint_{-\infty}^{\infty} I_0^2(x, y) dx dy + \\ + \sum_{s=1}^S \iint_{-\infty}^{\infty} |g(x, y, z_s)|^2 dx dy.$$

Any point in Eq. (9) is connected with the point of the minimum of the approach functional

$$\mathcal{J}_1(I, I_0, g, I_1, I_{01}, g_1) = \|(I - I_1, I_0 - I_{01}, g - g_1)\|^2,$$

where $(I, I_0, g) \in V$ and $(I_1, I_{01}, g_1) \in V_1$. The minimum of the functional is achieved at $(I, I_0, g) = (I_1, I_{01}, g_1) \in VV_1$.

The minimization of the functional \mathcal{J}_1 will be performed by the method of coordinate descent in terms of the variables $(I, I_0, g) \in V$, then in terms of the variables $(I_1, I_{01}, g_1) \in V_1$, and so on. Let the n th approximation be defined at the points $(I, I_0, g)_n \in V$ and $(I_1, I_{01}, g_1)_n \in V_1$. If the $(n + 1)$ th approximation from V_1 is defined as

$$(I_1, I_{01}, g_1)_{n+1} = P_{V_1}(I, I_0, g)_n,$$

then the approximation

$$I_{1\ n+1} = \begin{cases} I_{\text{meas}} + n_* & \text{in } \omega \times \{z_1, \dots, z_S\}, \\ I_n & \text{in } \omega \times \{z_1, \dots, z_S\}, \end{cases}$$

where n_* is the solution of the problem

$$\sum_{s=1}^S \|I_n(z_s) - I_{\text{meas}}(z_s) - n\|_{\omega}^2 = \\ = \min_{n \in N} \sum_{s=1}^S \|I_n(z_s) - I_{\text{meas}}(z_s) - n\|_{\omega}^2.$$

The index ω indicates that the integration in the equation for the norm in terms of the variables (x, y) is performed over the area ω . The approximation

$$I_{01\ n+1} = \begin{cases} I_{0n} & \text{at the points of } \omega_0, \text{ at which } I_{0n} \geq 0; \\ 0 & \text{at the other points.} \end{cases}$$

The approximation $g_{1\ n+1}$ is determined, with the allowance for what has been said in Section 2, by the equality

$$g_{1\ n+1} = F[GG_0(z)],$$

where

$$G = \begin{cases} 1/S \sum_{s=1}^S G_0^*(z_s) F^{-1} g_n(z_s) & \text{in } \Omega, \\ 0 & \text{outside } \Omega. \end{cases}$$

Now consider the $(n + 1)$ th approximation for the variables from the set V . The set V relates three functions I , I_0 , and g by same nonlinear equality, and therefore it is impossible to obtain the analytical equation for the approximation $(I, I_0, g)_{n+1}$. Taking this into account, we shall seek this approximation from the condition of descent of the approach functional:

$$\mathcal{J}_1[(I, I_0, g)_{n+1}, (I_1, I_{01}, g_1)_{n+1}] \leq \\ \leq \mathcal{J}_1[(I, I_0, g)_n, (I_1, I_{01}, g_1)_{n+1}].$$

Using the equality $I = I_0 * |g|^2$, we can exclude the variable I from the functional \mathcal{J}_1 :

$$\mathcal{J}_1 = \mathcal{J}_1(I_0, g) = \sum_{s=1}^S \|I_0 * |g(z_s)|^2 - I_1\|^2 + \|I_0 - I_{01}\|^2 + \\ + \sum_{s=1}^S \|g(z_s) - g_1(z_s)\|^2.$$

Only the third term of the functional depends on the phase of the function $g(z_s)$. From the condition of minimum of the functional in terms of $\arg g(z_s)$, we can find that $\arg g(z_s) = \arg g_1(z_s)$. The minimization of the functional in terms of the variables (I, I_0, g) from the set V is reduced to the unconditional minimization, in I_0 and $a(z_s) = |g(z_s)|$, of the functional

$$\begin{aligned} \mathcal{J}_1(I_0, a) = & \sum_{s=1}^S \|I_0 * a^2(z_s) - I_1\|^2 + \|I_0 - I_{01}\|^2 + \\ & + \sum_{s=1}^S \|a(z_s) - |g_1(z_s)|\|^2. \end{aligned} \quad (10)$$

The sign of the variable $a(z_s)$ affects only the third term, whose minimum is achieved at $a(z_s) \geq 0$, and therefore the restriction on the sign $a(z_s) \geq 0$ can be neglected.

To realize the gradient descend of the functional (10) in terms of the variables I_0 and a , we should know the equation for the antigradient vector in these variables. Determine the gradient based on the variation of the functional. The variation of the functional (10) corresponding to the variation δI_0 is

$$\begin{aligned} \delta \mathcal{J}_1(I_0, a) = & \delta \left(\sum_{s=1}^S \|I_0 * a^2(z_s) - I_1(z_s)\|^2 + \|I_0 - I_{01}\|^2 \right) = \\ = & 2 \sum_{s=1}^S [I_0 \cdot a^2(z_s) - I_1(z_s), \delta I_0 \cdot a^2(z_s)] + 2(I_0 - I_{01}, \delta I_0) = \\ = & 2 \left(\sum_{s=1}^S [I_0 * a^2(z_s) - I_1(z_s)] \cdot \bar{a}^2(z_s) + I_0 - I_{01}, \delta I_0 \right), \end{aligned}$$

where

$$\bar{a}(z_s) = a(-x, -y, z_s).$$

The variation of the functional (10) corresponding to the variation δa is

$$\begin{aligned} \delta \mathcal{J}_1(I_0, a) = & \delta \left(\sum_{s=1}^S \|I_0 * a^2(z_s) - I_1(z_s)\|^2 + \|a(z_s) - |g_1(z_s)|\|^2 \right) = \\ = & 2 \sum_{s=1}^S [I_0 * a^2(z_s) - I_1(z_s), I_0 * 2a(z_s) \delta a(z_s)] + \\ & + [a(z_s) - |g_1(z_s)|, \delta a(z_s)] = \end{aligned}$$

$$\begin{aligned} = & 2 \left\{ \sum_{s=1}^S [(I_0 * a^2(z_s) - I_1(z_s)) \cdot \bar{I}_0] 2a(z_s) + \right. \\ & \left. + a(z_s) - |g_1(z_s)|, \delta a(z_s) \right\}, \end{aligned}$$

where $\bar{I}_0(x, y) = I_0(-x, -y)$. The right-hand sides of the equations for the variations $\delta \mathcal{J}(I_0, a)$ are the scalar products of variations of the variables $I_0, a(z_s)$ and the functional derivatives of the approach functional with respect to these variables. These derivatives, taken with the negative sign, define the antigradient, in whose direction, from the point $(I_0, a)_n$, it is possible to decrease the value of the approach functional.

4. Discussion and conclusions

We have described and demonstrated, through numerical simulation, the method, which allows the phase function to be calculated in the plane of the output pupil from the incomplete intensity distributions in adaptively formed images of the source in several parallel planes.

The images in several planes can be identified with the images in one plane with different phase modulations of the wave at the output pupil. In this sense, other modulations are also acceptable. Thus, in AOS with the Shack–Hartman wave front sensor, the image of an optical system and sensor spots can be considered as different images of a point source, corresponding to different modulations of the wave field. Therefore, the proposed method can be used to reconstruct the phase at the pupil from these intensity distributions.

References

1. M.A. Vorontsov, A.V. Koryabin, and V.I. Shmal'gauzen, *Controlled Optical Systems* (Nauka, Moscow, 1988), 268 pp.
2. S.M. Chernyavskii, *Atmos. Oceanic Opt.* **9**, No. 1, 51–55 (1996).
3. G.L. Degtyarev, A.V. Makhan'ko, S.M. Chernyavskii, and A.S. Chernyavskii, *Proc. SPIE* **4678**, 144–153 (2001).
4. P.A. Bakut and V.E. Kirakosyants, *Atmos. Oceanic Opt.* **11**, No. 11, 1023–1027 (1998).
5. V.P. Lukin and B.V. Fortes, *Adaptive Beaming and Imaging in the Turbulent Atmosphere* (SPIE Press, 2002).
6. H. Stark, ed., *Image Recovery, Theory and Application* (Academic Press, 1987).
7. S.M. Chernyavskii, *Proc. SPIE* **3583**, 282–287 (1998).
8. M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, New York, 1959).